RIIO-GD2 and T2: BSC and CAI assessment methodology

Methodology Paper

7 May 2020

Submitted to Ofgem by:

Economic Consulting Associates

Economic Consulting Associates Limited
41 Lonsdale Road, London NW6 6RA, UK
tel: +44 20 7604 4546, fax: +44 20 7604 4547
www.eca-uk.com

Contents

Tables and figures

Tables

Figures

Boxes

Abbreviations and acronyms

BPDT	Business Plan Data Template
BSC	Business Support Costs
CAI	Closely Associated Indirects
CMA	Competition and Markets Authority
COLS	Corrected Ordinary Least Squares
CSV	Composite Scale Variable
DEA	Data Envelopment Analysis
DNO	Distribution Network Operator
DPCR	Distribution Price Control Review
ECA	Economic Consulting Associates
EoE	East of England (Cadent GDN)
ET	Electricity Transmission
FE	Fixed Effects
GD	Gas Distribution
GDN	Gas Distribution Network
GDPCR	Gas Distribution Price Control Review
GT	Gas Transmission
IT&T	Information Technology & Telecommunications
Lon	North London (Cadent GDN)
MEAV	Modern Equivalent Asset Value
NGET (SO)	National Grid Electricity Transmission (System Operator)
NGET (TO)	National Grid Electricity Transmission (Transmission Operator)
NGGT (SO)	National Grid Gas Transmission (System Operator)
NGGT (TO)	National Grid Gas Transmission (Transmission Operator)
NGN	Northern Gas Networks
NW	North West England (Cadent GDN)
POLS	Pooled Ordinary Least Squares
RE	Random Effects
RESET	Ramsey Regression Equation Specification Error Test
RIIO	Revenue = Incentives + Innovation + Outputs
Sc	Scotland (SGN GDN)
SCF	Special Cost Factor
SFA	Stochastic Frontier Analysis
SHET	Scottish Hydro Electric Transmission
So	Southern (SGN GDN)
SPT	Scottish Power Transmission
T	Transmission

TPCR	Transmission Price Control Review
UQ	Upper Quartile
VIF	Variance Inflation Factor
WM	West Midlands (Cadent GDN)
WWU	Wales and West Utilities

Executive summary

Ofgem commissioned a partnership of CEPA, Economic Consulting Associates (ECA) and AFRY Management Consulting (AFRY) to provide economic advice for RIIO-2. . Under this Economic Strategic Partner contract for RIIO-2,, and building on ECA's *RIIO-GD2 Cost Assessment: Business Support Costs* paper from June 2019, ECA was tasked with assisting Ofgem in its development of an assessment methodology for Business Support Costs (BSCs) and Closely Associated Indirects (CAIs).

The primary task for this assignment was to develop a well-reasoned methodology for assessing BSCs and CAIs in RIIO-2. There were **multiple questions to consider in building up a robust methodology**:

- How should BSCs / CAIs be evaluated?
  - In aggregate or by sub-category?
  - At a group level or individual network level?
  - At a gross or net (after allocations) level?
- How can fixed costs be accounted for in a model?
- Should any costs be excluded from the analysis?
- What sample should be used?
  - What sectors can / should be pooled?
  - Should external comparators or Electricity Distribution (ED) data be included?
  - Should historical or forecast data (or both) be used?
- What are the most robust cost drivers for these activities? Are they economically / technically logical, statistically robust, and reasonably exogenous, i.e. outside of the networks' control?
- What estimation methodology is most appropriate?
  - An econometric approach may allow for more sophisticated approaches that account for multiple cost drivers and panel datasets, which may be first-order preferable, but inherently small sample sizes limit modelling options
  - Alternatively, ratio benchmarks can be simplistic and crude, but they also produce transparent and intuitive results.

Acknowledging the **trade-offs and imperfections of any assessment model**, Ofgem has not requested that we strictly adhere to a regulatory precedent set by the final assessment methodologies for BSCs / CAIs in RIIO-1 or prescribed an analytical approach or model. Instead, we have considered **a 'toolkit' of potential assessment options** that have previously been used by Ofgem and other regulatory bodies, seeking to take **a pragmatic approach that best represents the data**.

BSC and CAI categories and trends

Consistency in cost reporting is vital to any regulatory benchmarking exercise and we note that in the Business Plan Data Templates (BPDT), BSC categories have been consistently defined for gas distribution (GD), gas transmission (GT), and electricity transmission (ET). CAIs are also consistently defined for GT and ET.

However, given its different nature, NGGT (SO) only reports two CAI categories, which means it cannot be separately included in an aggregate CAI benchmark. It would either need to be separately assessed or combined with the costs of NGGT (TO). The latter approach raises the question of whether ET TOs can be compared to a combined GT TO and SO. **Hence, we have excluded the GT SO from our benchmarking analysis.** Furthermore, GT SO CAIs are primarily IT&T, which are being separately assessed.

We also compared these categories of indirect costs to those used for RIIO-ED1, which was considered for inclusion in the benchmarking. **We note in this regard that ED BSC categories are consistent but there are some marked differences between CAI categories for ET / GT and for ED.**

An examination of sector-wide BSC trends reveals a divergence in RIIO-1 between GD and ET / GT, with the former declining and the latter increasing. This does not on its own suggest pooling GD and ET / GT is inappropriate, as the benchmarking analysis will discern whether these trends can be explained by appropriate cost drivers. All three sectors' BSCs are forecast to be relatively flat for RIIO-2.

There are diverging trends for CAIs across ET and GT, which may signal caution about pooling the two sectors. However, the apparent trends for RIIO-1 and RIIO-2 also align with the expectation that activity on GT will be flat given stagnant gas demand, while investment and activity for the ETs can be expected to rise in future as electrification, new offshore wind, etc. trends continue. Hence, a robust cost driver may be able to capture these expected workloads.

Pre-modelling considerations

We explored both cross-sector and sector-specific approaches to determine their appropriateness. Our analysis began with the (strong) assumption that all of GD, GT, and ET can be pooled into the same benchmarking approach (i.e. a cross-sector examination), which would maximise the available sample size. However, given general concerns about the true comparability of costs between the sectors (given their different operating environments, for example), we signal that within-sector analysis might need to be relied on if cross-sector models do not appear to be robust.

We have ruled out the use of external benchmarks, whose comparability to networks is inherently questionable. **We also exclude using ED data** as this would require significant data normalisations, and would not on its own resolve any apparent unreliability in a cross- or within-sector model, despite the advantage of increasing sample size.

We do not *a priori* rule out the use of forecast data, which can be useful in explaining future costs, especially where the historical costs are not a reliable guide. In the end, **our baseline models use only historical data** given concerns that for this analysis forecast data will be subject to company bias and may be mistaken in hindsight. However, we do conduct **model sensitivity checks including forecast data**.

To better ensure comparability of costs, **we have assessed costs at a gross rather than net level**. Otherwise, a model's assessment may be influenced by differing cost allocation policies between networks rather than actual efficiency. If Ofgem assesses other aspects of the price control in net terms, a consistent methodology will be needed for reallocating gross costs to net. Ofgem previously did this on a proportional basis, which is a reasonable and practical approach.

The cost allocation issue is also partially resolved by our preference to **assess at the group level** as assessing at the individual network level may distort results due to differing cost allocation policies within groups. However, **a group level assessment was not practical for the transmission sectors given the sample size**. Network level assessments were also conducted as a sensitivity check for gas distribution.

Our assessments have been conducted from **a top-down basis rather than at an activity level** to further set aside differences in cost allocations and to reduce the risk of being seen to 'cherry pick' results. Furthermore, potential inconsistencies in cost allocation reporting between RIIO-1 and RIIO-2 necessitated evaluating CAIs in aggregate.

We have not explicitly modelled fixed cost adjustments, but we have been conscious of whether ratio benchmarks differ significantly from regression models (for which a positive intercept term should to some extent capture fixed costs) for standalone networks. Our **baseline models have excluded IT&T costs**, for which the potential presence of fixed / overhead costs is most obvious. Sensitivity checks where IT&T costs were included, and insurance costs excluded (as has been recommended by multiple companies), gave broadly similar results for BSCs and CAIs for transmission companies, but the result with IT&T costs included for GD suggested they may need to be assessed separately. On this note, we are aware that Ofgem is already assessing networks' IT&T costs as part of a separate expert review.

Cost driver selection and model specifications

We considered several potential cost drivers. For BSCs, we looked to scale-related drivers, but we also recognised that many BSCs are a combination of fixed and semi-variable factors. Hence, **Modern Equivalent Asset Value (MEAV) is our preferred cost driver** given it simultaneously reflects the scale, complexity, characteristics, and composition of the network asset base. However, **Composite Scale Variables (CSV) were also tested** in order to control for other cost drivers: FTEs for HR costs and Total Spend / Totex for Procurement costs.

Relative to more scale-driven BSCs, CAIs, being 'closely associated', can be expected to be linked to network workloads. **We considered both scale- and workload-related drivers**.

However, the pooling of ET and GT presented a difficulty in cost driver comparability, e.g. Network Length cannot be considered comparable between the two sectors. We considered both Total Capex, as a workload driver, and MEAV, as a scale driver, and multivariate regressions of the two. Potential inconsistency in CAI reporting between RIIO-1 and RIIO-2 prevented confidently splitting CAIs into 'fixed' and 'flex' components, as was considered for RIIO-ED1.

Our econometric approach began with using Pooled Ordinary Least Squares (POLS) estimators given their relative simplicity, transparency, and favourable small sample properties. However, we considered panel-based estimators – Random Effects and Fixed Effects – as sensitivity checks. **If econometric models did not prove to be robust, we turned to ratio benchmarks**, which can be simplistic and crude, but the results can be easily visualised and understood; Ofgem has used ratio analysis for past price controls.

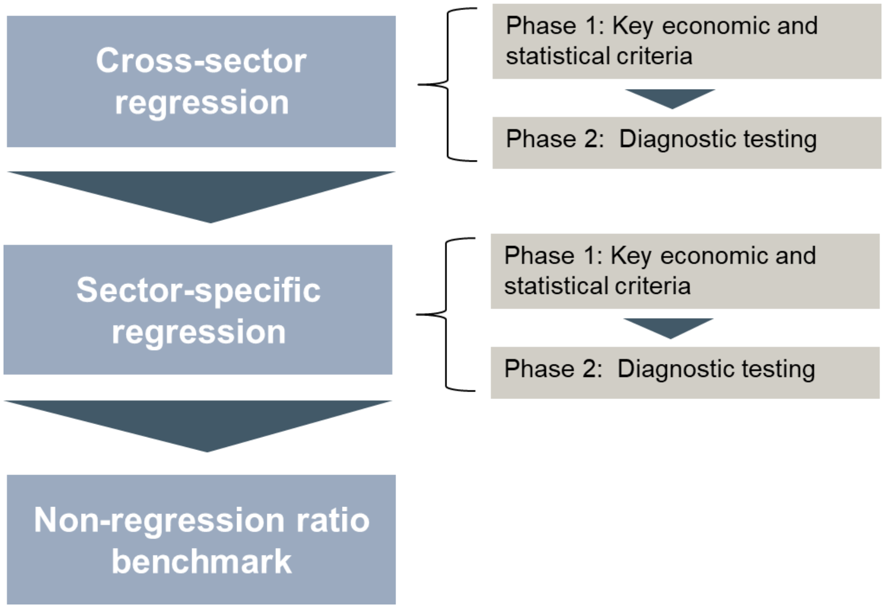
Models were first evaluated at the average (for regressions) or median (for ratio benchmarks) model fit. If the models were considered sufficiently robust, we considered the implications of a more stringent benchmark, e.g. setting the benchmark at an upper quartile.

Model selection

This entailed two analytical phases. **Phase 1 assessed the model's general statistical fit, the robustness of the chosen cost drivers, and whether the modelled results appeared plausible**. All models were first run via POLS, but panel estimators were considered as sensitivity checks, or if the POLS results were not found to be promising. Other modelling adjustments, such as time trends or sector dummy variables, were also considered to improve model fits, while being conscious of small sample sizes limiting degrees of freedom. If a regression model was found to meet these criteria, we turned to **Phase 2 where a range of diagnostic tests were run to further test the model's robustness**. The table below summarises the tests conducted in the two analytical phases.

Phase	Category of assessment	Test/criterion
Phase 1	Economic and technical logic	Estimated coefficients are of a sign, size, and significance that agrees with economic logic
		Generates plausible estimates of modelled costs (efficiency scores)
	Statistical robustness	F-test for joint significance of coefficients
		Goodness of fit (Adjusted R-squared)
		t-test for significance of individual coefficients
		Avoidance of multicollinearity (where relevant)
	Estimator selection	Hausman test
		Breusch-Pagan test
Phase 2	Diagnostic tests	RESET test for misspecification caused by non-linearities
		White test for heteroscedasticity
		Jarque-Bera test for normality

If regression models, whether POLS or panel estimators, were not found to be robust, we considered ratio benchmarks as an alternative approach. We have used the easy visualisation of ratios to help provide intuition for the resulting benchmarks. The model assessment process employed is summarised in the figure below.



BSC models

We began with a cross-sector approach for BSCs but found the econometric results to be of middling strength and the resulting modelled costs led to implausible results between sectors that could not be resolved by modelling adjustments, i.e. including sector dummy variables. Hence, we then took sector-specific approaches for GD and ET / GT.

For GD, we found regression models, whether using MEAV or a CSV, to have limited statistical power, so then examined a simple MEAV ratio benchmark, which provided intuitive results as to which groups / networks were found to be efficient / inefficient. The MEAV ratio benchmark results were also generally consistent with those of both the group and network level regression models. Furthermore, given concerns about the consistency of the historical time series due to Cadent's change in ownership in the middle of RIIO-1, we have more confidence in using a ratio benchmark based only on 2019 data, which may better reflect the *current* state of the sector, than a regression that uses the full historical time series. An upper quartile benchmark was considered as well, but considering the uncertain measurement of a relatively simplistic ratio analysis, we settled on **using a MEAV 2019 median ratio benchmark for assessing GD BSCs**.

For the transmission companies, both MEAV and CSV regression models appeared to have a strong statistical fit, but the implied modelled costs led to concerning outlier results. Ratio analysis did not help illuminate the issue nor did panel-based estimators or the inclusion of time effect variables. There were few alternative cost drivers that could be used, given limited transferability between ET and GT for other scale-related drivers, e.g. Customer Numbers or Throughput.

Our proposed solution is the addition of a GT sector dummy variable, which materially improves the model's statistical fit and helps address a general concern about the true comparability of ET and GT. This adjustment lessened the dispersion of the efficiency assessments and the inclusion of a GT dummy variable helps focus the efficiency discussion to within the ET sector. The GT dummy variable was combined with a CSV as this was found to give a stronger model fit than a MEAV-only regression. **Our BSCs model of choice for ET and GT was therefore a CSV regression that included a GT sector dummy variable.** Outlier results remain, particularly for SPT, for which we have considered several sensitivity checks to confirm the model's general robustness. Given the presence of outliers, we have not considered setting a more stringent benchmark than the average model fit.

CAI models

Both Total Capex and MEAV were found to be robust cost drivers for ET / GT CAIs. A multivariate regression model including both variables has a strong statistical fit. However, when converting to modelled costs, a significant dispersion emerged in efficiency scores, particularly for NGET and SHET. We considered a number of different specifications, including Random Effects, Fixed Effects, time trends, and simple ratio analyses.

An investigation of both ratio analysis and a Fixed Effects specification, which had opposite results for SHET, helped explain the underlying dynamics: SHET appears to be efficient *relative to the sector*, but its RIIO-2 submission appears to be inefficient *relative to SHET's RIIO-1 actuals*. Hence, a simple ratio analysis concludes that SHET is highly efficient while a Fixed Effects model argues that SHET is inefficient. For NGET, all models consistently conclude that its RIIO-2 submission is inefficient, only disagreeing with the extent of the inefficiency. We considered the potential for a mixed model approach, for example, the application of a 50 / 50 split of the Total Capex ratio and Fixed Effect models. However, the use of multiple models in combination can be justified when it addresses or ameliorates the defects of the individual models. This was not the case here, as the relevant model options do not directly address the trade-offs apparent to either modelling approach.

We therefore conclude that a multivariate regression that includes both MEAV and Total Capex, which has attractive statistical properties and intuitive reasoning, is our model of choice for assessing CAIs for the transmission companies. However, we include a discussion of each network's results, which require further scrutiny. We consider the model appropriate for forming the *basis* of an efficiency challenge, but further investigation (outside of the modelling process) is needed by Ofgem before it takes its decision on where to set the allowances, particularly for NGET and SHET.

1 Introduction

Ofgem commissioned a partnership of CEPA, Economic Consulting Associates (ECA) and AFRY Management Consulting (AFRY), to provide economic advice for RIIO-2. Under this Economic Strategic Partner contract for RIIO-2,, ECA was tasked with assisting Ofgem in its development of an assessment methodology for Business Support Costs (BSCs) and Closely Associated Indirects (CAIs).

The scope of this work comprises the following main elements:

- A methodology paper evaluating the key factors that determine whether Ofgem should pursue a regression-based or non-regression based approach to assessing BSCs / CAIs in RIIO-2, including identification and testing of different drivers, discussion of the strengths and weaknesses of a given approach, and development and application of selection criteria to choose the preferred approach.

- A recommendation on the appropriateness of undertaking a combined BSC assessment for Gas Distribution (GD), Gas Transmission (GT) and Electricity Transmission (ET), and a combined CAI assessment for GT and ET, taking into account differences in the availability of data from the Business Plans.

- Justification of the choice of the level of aggregation (aggregated BSCs / CAIs *vs* single subcategories) or of their combination if more than one level of aggregation is used.

- Detailed discussion of the relative advantages and disadvantages of assessing BSCs / CAIs at group or individual network level, and of using net or gross level data to undertake BSCs / CAIs assessment.

- Modelling to explore a range of alternative cost drivers for regressing BSCs and CAIs.

- A recommendation of the model(s), which are best suited for calculating allowances for the cost categories and sectors with justification of the modelling outcome, specifically providing statistical test outputs and associated justifications for model selection.

- Preparation of a final suite of models, using December Business Plan data, for use by Ofgem in their modelling suite, along with appropriate levels of supporting documentation. These models and supporting material accompany this methodology paper and have been separately submitted to Ofgem.

We were conscious of how our modelling approach aligned with the regulatory precedent set by Ofgem across RIIO-1. However, it is important to note that Ofgem has sought to maintain a consistent ‘toolkit’ of approaches rather than keeping to any specific methodology. This provides scope for adjusting the approach to ensure it is fit for purpose in the RIIO-2 context.

Accordingly, **our approach has concentrated on pre- and within-modelling adjustments** in order to ensure data comparability and goodness of fit to thereby set a reasonable efficiency benchmark, while highlighting the trade-offs and intuition of each modelling approach. Any post-modelling adjustments would need to be made outside of the top-down modelling framework employed in this paper.

The paper is structured as follows:

- Section 2 provides a brief introduction to BSCs and CAIs and compares how their sub-categories overlap between the GD, GT, ET, and ED sectors;
- Section 3 looks at overall BSC and CAI trends across RIIO-1 actuals and RIIO-2 forecasts, which provides some initial context for the benchmarking exercise;
- Section 4 details numerous considerations that need to be made with respect to data normalisations and comparability, which inform the modelling approach;
- Section 5 discusses the merits of different cost drivers and modelling specifications;
- Section 6 describes our model selection process, as informed by the preceding sections, and the modelling results that lead us to our assessment models of choice;
- Section 7 concludes and compares our resulting approach to RIIO-1.

The paper also contains two Annexes:

- Annex **Error! Reference source not found.** reviews network level BSC and CAI trends, which provide further context to the benchmarking results; and
- Annex A1 provides a brief overview of the approaches taken to assessing BSCs and CAIs across RIIO-1. agreed

2 Description of BSCs and CAIs

Box 1 BSC and CAI category consistency

- BSC categories are consistent across GD, GT, and ET, as well as for ED, which may favour a cross-sector approach
- CAI categories overlap for GT and ET, favouring pooling the two sectors in the CAI assessment
- There is limited overlap between CAI categories for ED and for transmission, which casts some doubt on pooling ED data with ET and GT for the CAI assessment

We begin by briefly describing BSCs and CAIs, setting out how these are broken down into cost categories in RIIO-2 and discussing how well these categories overlap across sectors, which has implications for their comparability in a cross-sector benchmark.

2.1 Business support costs

Business Support Costs (BSCs) are incurred supporting companies' general business activities and are one component of network companies' indirect operating expenditure (opex).

In Table 1, we compare the overlap of BSC categories for GT, ET, and GD, as well as ED (the potential inclusion of which as a comparator is discussed in Section 4.1). The categories are consistent across the three sectors of GD, GT (TO and SO roles) and ET, with the exception of Stores and Logistics, which are included in CAIs for GT and ET. Other than minor terminology differences, ED BSC categories for RIIO-ED1 are also consistent.

Table 1 Business Support Cost categories by sector

BSC category	GT	ET	GD	ED
IT & Telecoms (IT&T)	Yes	Yes	Yes	Yes
Property management	Yes	Yes	Yes	Yes
Audit, finance, and regulation	Yes	Yes	Yes	Yes[1]
HR and non-operational training	Yes	Yes	Yes	Yes[2]
Insurance	Yes	Yes	Yes	Yes
Procurement	Yes	Yes	Yes	Yes
CEO and group management	Yes	Yes	Yes	Yes[3]
Stores and logistics	CAI	CAI	Yes	CAI[4]
Fines and penalties (outside of street works)	No	No	No	Yes

Notes: Compiled from Ofgem's final RIIO-2 business plan data templates and compared to the assessment of BSCs for RIIO-ED1. [1]Categorised as "Finance and regulation" for ED. [2]HR and Non-operational Training are categorised separately for ED. [3]Categorised as "CEO" for ED. [4]Only stores mentioned in ED under CAI

For consistency in aggregate benchmarking, either Stores and Logistics could be moved from CAI to BSC in ET and GT, or it could be removed from BSC in GD and separately assessed. **We have chosen to exclude Stores and Logistics (which make up less than 1% of networks' BSCs / CAIs) from GD costs in our BSC analysis** given concerns about their comparability across GD and transmission and some apparent inconsistencies in the values reported by GDNs.

2.2 Closely associated indirect costs

Closely Associated Indirect (CAI) costs comprise other indirect opex that supports operational activities.

Ofgem separately considered CAI operating costs in its transmission and electricity distribution RIIO-1 price controls, but not in gas distribution. In gas distribution, most of the CAI costs are within direct costs, as part of "work management". Hence, **gas distribution CAIs are not considered in our analysis**.

Consistent categories for CAIs in the TO roles in ET and GT helps facilitate cross-sector CAI benchmarking. Table 2 illustrates the overlap between CAI categories for ET, GT (TO), and GT (SO), as well as for ED. Unlike for BSCs, there are significant differences in CAI categories between ED and the transmission sector.s.

Table 2 Closely Associated Indirects categories by sector

CAI category	ET	GT (TO)	GT (SO)	ED
Operational IT & Telecoms	Yes	Yes	Yes	No
Project management	Yes	Yes	No	Yes
Network design and engineering	Yes	Yes	No	Yes
System mapping	Yes	Yes	No	Yes
Engineering management and clerical support	Yes	Yes	No	Yes
Network policy (including R&D)	Yes	Yes	No	Yes[2]
Health, safety, and environment (HSE)	Yes	Yes	Yes	No
Operational training[1]	Yes	Yes	No	Yes
Store and logistics	Yes	Yes	No	Yes[3]
Vehicles and transport	Yes	Yes	No	Yes
Market facilitation	Yes	Yes	No	No
Network planning	Yes	Yes	No	No
Call centre	No	No	No	Yes
Control centre	No[4]	No[4]	No	Yes
Wayleaves	No[5]	No[5]	No	Yes

Notes: [1]CAIs are not specifically evaluated for GD. However, GD does separately report training and apprentices. [2]R&D is not mentioned for ED. [3]Only Stores is mentioned for ED. [4]Included in network planning. [5]Included in engineering management and clerical support.

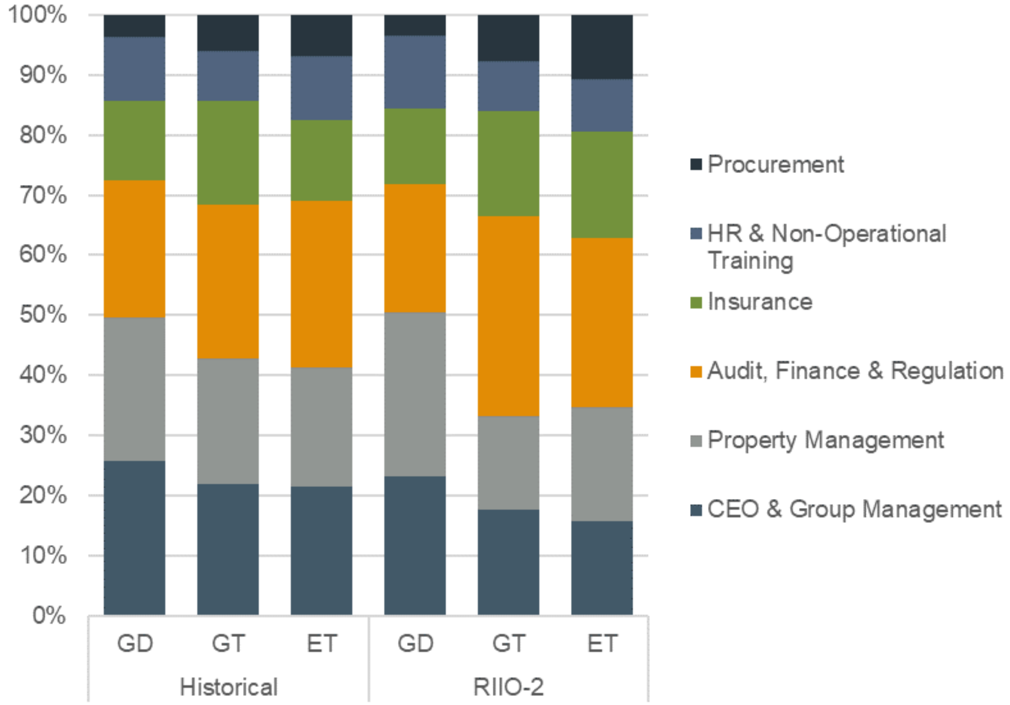
Given its different nature, there are just two categories of CAIs for the SO role in GT (Operational IT & Telecoms, which we understand is already being separately assessed by Ofgem, and Health, Safety & Environment). Consequently, the GT SO role cannot be separately included in any aggregated benchmarking of CAIs. Rather, those costs would either need to be separately assessed or combined with the costs for the GT TO role. This latter approach raises the question as to whether it is appropriate to compare ET TOs to a combined GT TO and SO. **Hence, we have excluded the GT SO from our benchmarking analysis**.

2.3 Sector-level composition of BSCs and CAIs

As an initial indication of comparability, we compare the composition of BSC and CAI categories in percentage terms across sectors for RIIO-1 actuals and RIIO-2 forecasts. This analysis is limited given it does not account for scale or underlying network activity, i.e. cost drivers, but it provides a first look at comparability across sectors.

Looking at BSCs (Figure 1), **there appears to be general consistency in BSC allocations across GD, ET, and GT, particularly between ET and GTs, and also between RIIO-1 actuals and RIIO-2 forecasts for each sector**. There are some minor deviations between GD and ET/GT with respect to CEO & Group Management and Property Management. This provides some justification for our initial BSC models (Section 6.2.2), where we consider a cross-sector assessment of BSCs, and particularly with respect to jointly modelling ET and GT BSCs.

Figure 1 Comparison of BSC cost composition for GD, GT and ET, RIIO-1 actuals and RIIO-2 forecast

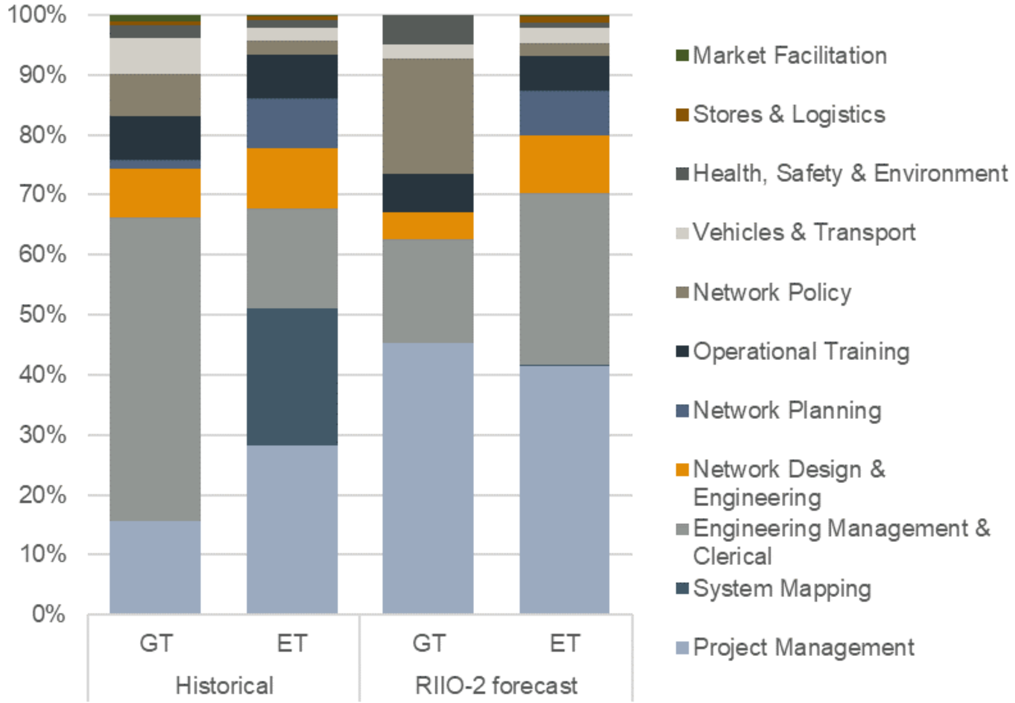


Source: ECA analysis of Ofgem data from December business plans. Excluding IT&T and Stores and Logistics.

More significant deviations are present in the composition of CAIs for ET and GT (Figure 2). The differences are less pronounced for RIIO-2, largely due to ET System Mapping CAIs falling to almost zero for RIIO-2.[1] Significant differences are also present *within* the ET sector between RIIO-1 and RIIO-2, suggesting it is not necessarily only an issue of ET and GT being incomparable.

[1] [REDACTED]

Figure 2 Comparison of ET and GT CAI cost composition, RIIO-1 actuals and RIIO-2 forecast



Source: ECA analysis of Ofgem data from December business plans. Excluding IT&T.

These differences reflect the limits of CAI category boundaries and potential inconsistencies in CAI allocation reporting, particularly for NGET. This **stresses the need for top-down benchmarking of CAIs** rather than attempting CAI activity-level benchmarking. As discussed in Section 4.4, aggregated rather than activity-level benchmarking should at least partially address the issue of differences in how companies allocate costs. Differences in overall operating environments / workloads can ultimately be controlled for at an aggregate level by a robust workload cost driver.

3 BSCs and CAIs in the RIIO-1 and RIIO-2 submissions

Box 2 Summary of BSC and CAI trends and implications

- There are diverging BSC trends between the GD and transmission sectors, which may reflect different underlying cost drivers.
- ET and GT exhibit similar trends for both BSCs and CAIs across both RIIO-1 and RIIO-2.
- Diverging BSC trends *within* the GD sector seem incongruous with the generally flat activity across the sector in aggregate.
- A high-level look at the overall sector trends does not provide *a priori* grounds for ruling out pooling sectors or using forecast data for statistical purposes.

To provide some high-level context, we summarise some trends of sectors' BSCs and CAIs for RIIO-1 and in their RIIO-2 business plans. These trends provide a cursory look at whether the sectors have been on similar trajectories, which has implications for the appropriateness of cross-sector benchmarking, and what may be expected in the benchmarking analysis. See Annex **Error! Reference source not found.** for network-level BSC and CAI trends.

3.1 BSCs

Across the five years of RIIO-2, companies forecast £1,235m (2018/19 prices) of BSCs (Table 3).[2] Both ET and GT show increases in average annual BSCs in RIIO-2 from RIIO-1. **Comparing average annual BSCs for RIIO-1 (actuals plus forecast) to RIIO-2 forecasts shows increases of 8.3% in GT and 3.4% in ET**. **By contrast, GD shows a reduction for RIIO-2 of 5%.** These are divergent trends between GD and the transmission companies, but it also suggests relative consistency between RIIO-1 and RIIO-2 at an aggregate level, which does not immediately rule out pooling both historical and forecast data.

Table 3 Sector level BSCs (2018/19 prices)

	GD	GT	ET
Average annual values			
RIIO-1 actuals (2013/14 – 2018/19)	£121m	[REDACTED]	£96m
RIIO-1 forecast (2019/20 – 2020/21)	£121m	[REDACTED]	£107m
RIIO-2 forecast (2021/22 – 2025/26)	£115m	[REDACTED]	£102m
Total for period			
RIIO-2 forecast (2021/22 – 2025/26)	£573m	[REDACTED]	£511m

[2] As discussed in Section 4.6, we excluded IT&T costs from our baseline analysis for both BSCs and CAIs, but they are included as a model sense-check. All costs are expressed in 2018/19 prices.

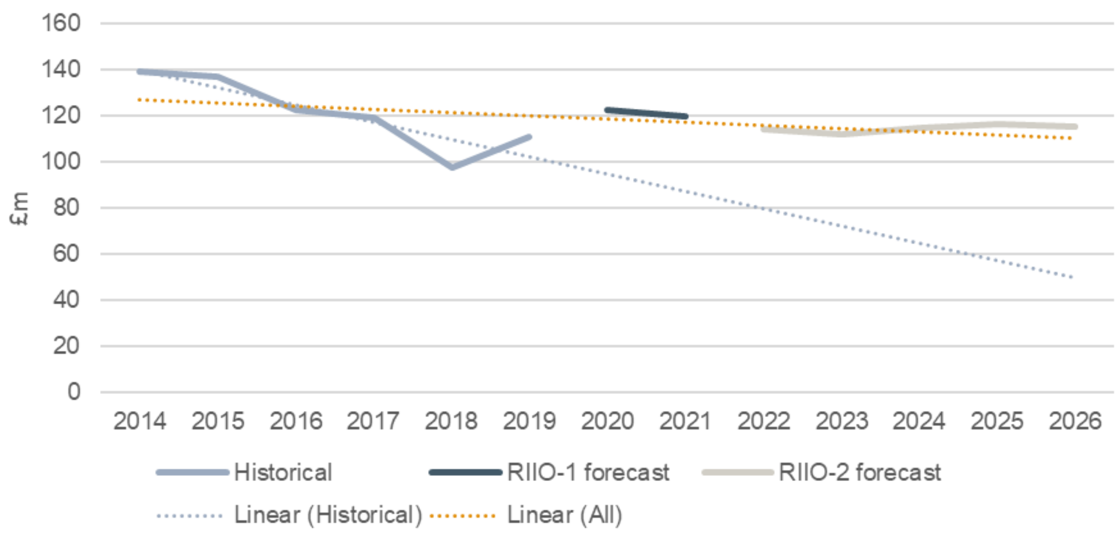
Source: ECA analysis of Ofgem data from December business plans. Gross BSCs. Excluding GT SO and IT&T.

3.1.1 BSC trends

The following charts show the trend in BSCs for each sector in aggregate, comparing RIIO-1 actuals to the companies' RIIO-1 and RIIO-2 forecast data.

In GD, costs have reduced over RIIO-1 to date, but forecasts show a generally flat trend. A naïve trendline benchmark of GD would expect RIIO-1's downward trend to continue through RIIO-2.

Figure 3 Gas Distribution BSC trend, RIIO-1 and RIIO-2



Source: ECA analysis of Ofgem data from December business plans. Excluding IT&T.

Looking at network-level trends (Annex **Error! Reference source not found.**), we see that the flat sector-wide trend masks major divergences within GD:

- The Cadent networks (East of England, North London, North West, and West Midlands) all exhibit a significant downward trend through RIIO-1, with a levelling off forecasted for RIIO-2
- Northern Gas Networks shows a flat trend
- The SGN networks, Scotland and Southern, show an upward trend through both RIIO-1 and RIIO-2
- Wales and West Utilities appears to forecast an upward step change, starting with its forecast for the remaining years of RIIO-1.

In both GT ([REDACTED]) and ET ([REDACTED]

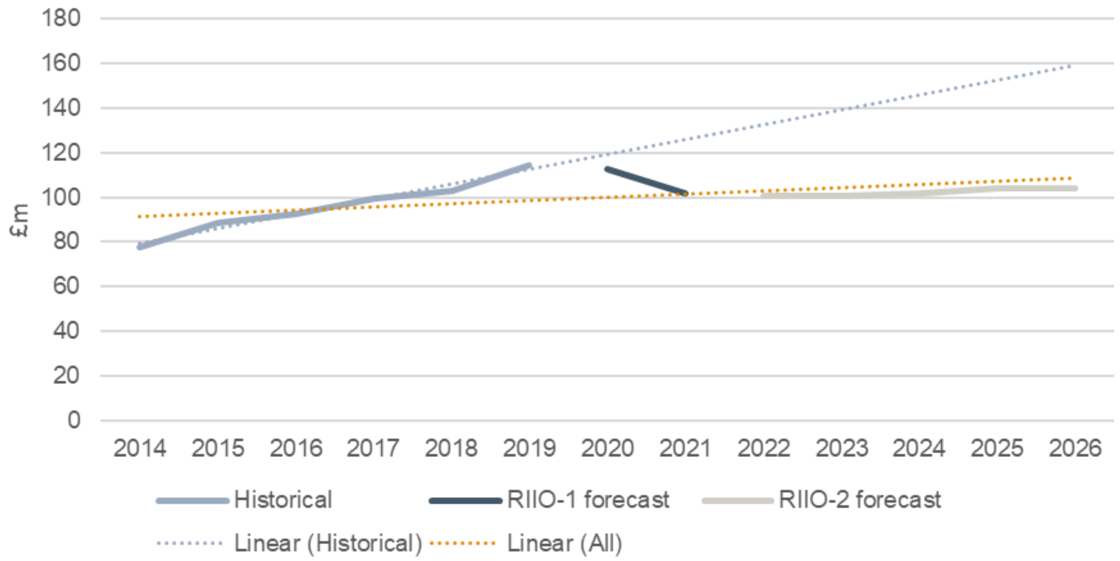
Figure 4), costs have generally risen in RIIO-1 to date, but forecasts have BSCs declining and then flat for RIIO-2. Examined at the network level (Annex **Error! Reference source**

not found.), the ET networks all exhibit a similar upward trend through RIIO-1, which is projected to flatten out for RIIO-2.

[REDACTED]

[REDACTED]

Figure 4 Electricity Transmission BSC trend, RIIO-1 and RIIO-2



Source: ECA analysis of Ofgem data from December business plans. Excluding IT&T.

3.2 CAIs

Across the five years of RIIO-2, companies' forecasts of CAIs are £1,647m (Table 4), with ET accounting for 84% [REDACTED]. Both ET and GT sectors forecast an increase in CAIs in RIIO-2 from RIIO-1. **Comparing RIIO-1 (actuals plus forecast) to RIIO-2 forecasts shows increases of** [REDACTED] **in GT and 7% in ET**.

Table 4 Sector level CAIs (2018/19 prices)

	GT	ET
Average annual values		
RIIO-1 actuals (2013/14 – 2018/19)	[REDACTED]	£258m
RIIO-1 forecast (2019/20 – 2020/21)	[REDACTED]	£265m
RIIO-2 forecast (2021/22 – 2025/26)	[REDACTED]	£275m
Total for period		
RIIO-2 forecast (2021/22 – 2025/26)	[REDACTED]	£1,377m

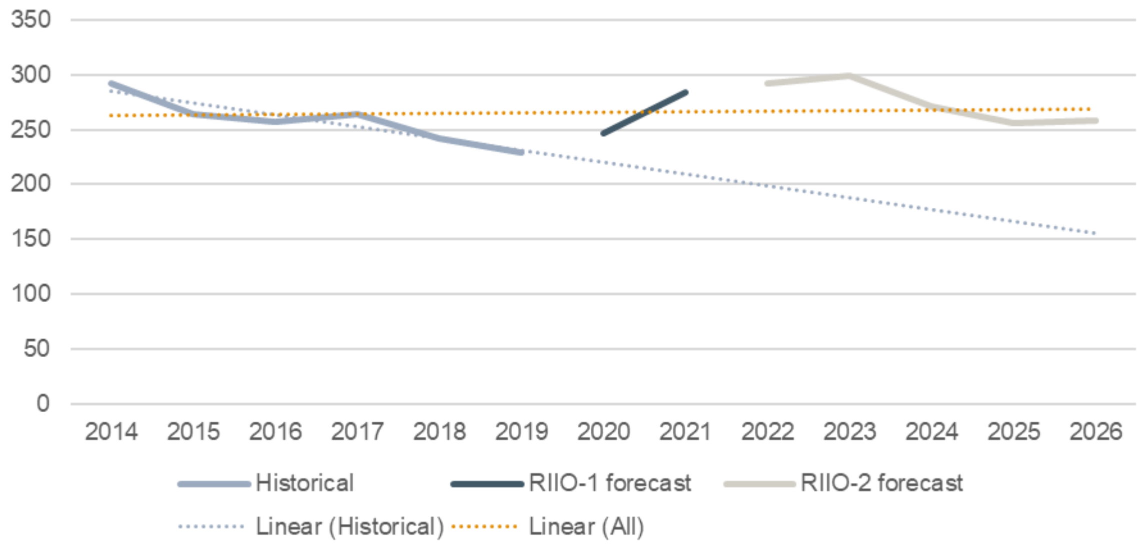
Source: ECA analysis of Ofgem data from December business plans. Excluding Operational IT&T and GT SO.

3.2.1 CAI trends

The marked increase in CAIs forecast for GT in RIIO-2 compared to RIIO-1 reflects the continuation of a rising trend in RIIO-1, as shown in [REDACTED]. By contrast (see Figure 5), in RIIO-1 to date, CAIs in ET have been on a downward trend. A large increase is forecast in ET CAIs ahead of RIIO-2 (roughly back to the value at the start of RIIO-1) after which costs rise, before falling back below the level at the start of RIIO-2.

[REDACTED]

Figure 5 Electricity Transmission CAI trend, RIIO-1 and RIIO-2



Source: ECA analysis of Ofgem data from December business plans

Within ET, there are significant diverging trends (see Annex **Error! Reference source not found.** for network-level charts). NGET CAIs have steadily declined through RIIO-1, but a sharp upward tick is forecast for the rest of RIIO-1, and then flattening out for RIIO-2. SHET CAIs have been steadily increasing for RIIO-1, with further upward step changes for the remaining RIIO-1 years and RIIO-2 also being forecast. SPT has seen a sharp decline in CAIs in the last few years of RIIO-1, but this is forecast to flatten out for RIIO-2.

3.3 Implication of BSC / CAI trends for modelling approaches

GD BSCs are forecast to be flat at the sector level, and this lines up with a general impression of flat cost drivers and limited new activity on the networks. Scale-related cost drivers, which tend to have flatter trends, may not adequately explain year-to-year cost variations, but this appears to be less of an issue given relatively flat BSC costs across the GD sector. However, given diverging BSC patterns exhibited at the group / network level, our benchmarking analysis will need to discern whether this can be explained by group / network level cost drivers.

BSCs have shown an upward trend for both ET and GT in RIIO-1, which is then forecast to flatten out for RIIO-2 for both sectors. This provides some confidence in pooling ET and GT for BSC benchmarking given similar aggregate trends. In contrast, GD exhibited a downward trend throughout RIIO-1, broadly suggesting an inherent difference with the transmission sectorss. However, a similar flat trend for GD in RIIO-2 suggests that pooling GD and transmission may still be appropriate, subject to underlying cost drivers.

There are diverging trends for CAIs across ET and GT, which may signal caution about pooling the two sectors. However, the apparent trends for RIIO-1 and RIIO-2 also align with the expectation that activity on GT will be flat given stagnant gas demand, while investment and activity for the ETs can be expected to rise in future as electrification, new offshore wind, etc. trends continue. Hence, a robust cost driver may be able to capture these expected workloads.

Finally, there does not appear to be an obvious 'step change' in BSCs or CAIs between RIIO-1 and RIIO-2, which may suggest pooling historical and forecast data is not an issue, but we discuss this further in Section 4.1.

4 Pre-modelling considerations

Box 3 Summary of pre-modelling decisions and adjustments

- Our analysis started with cross-sector analyses of BSCs, but, given comparability concerns, we turned to within-sector analysis if cross-sector models appeared unreliable. There is less scope for within-sector analysis (separating ET and GT) for CAIs given the limited sample size. We ruled out the use of external benchmarks or ED data.
- Costs were evaluated at a gross, rather than net, level to ensure better comparability of costs. A consistent process for reallocating gross back to net would be needed if other parts of the price control are assessed on a net basis.
- Our preference was to conduct modelling on a group, rather than network, basis to avoid inconsistencies in cost allocations. Network level assessments were still considered as a sense-check. Group-level assessments were not practical for the transmission companiess given the small sample size and cross-sector implications. Therefore, the assessments for ET and GT (for both BSCs and CAIs) were conducted at a network level.
- In the end, our baseline models only used historical data as forecast data will be subject to company bias and may be inaccurate in hindsight. We conducted model sensitivity checks that include forecast data.
- We have not explicitly modelled fixed cost adjustments, but we have factored in whether ratio benchmarks differ significantly from regressions (for which a positive intercept term should to some extent capture fixed costs) for standalone networks.
- We have excluded IT&T costs from our baseline models given concerns about its predictability and its fixed cost components. Sensitivity checks were conducted with IT&T costs included.

This section details a range of issues that raise choices that need to be made with respect to data normalisations in the assessment and benchmarking of BSCs and CAIs. Data normalisations aim to adjust costs to make them more comparable over time and / or between entities.

4.1 Comparability of costs

In determining a benchmark, Ofgem can potentially use comparators from:

- *outside* the network sectors (i.e. external benchmarks);
- *across* different network sectors, such as GD, GT, ET, and ED (i.e. cross-sector benchmarks); or
- *within* a network sector, such as GD only (i.e. a within-sector benchmark).

Across RIIO-1, Ofgem used or considered all three of these options. In RIIO-GD1/T1, Ofgem used external benchmarks to set BSC allowances, but also compiled cross-sector benchmarks, while for RIIO-ED1 it benchmarked DNOs' BSCs against each other. In broad terms, there is a trade-off across these options between the **number of comparators** and the **comparability of costs**, both of which can affect the robustness of results.

There may be reason to suspect BSC categories have more overlap between sectors than for CAIs. As discussed in Section 2.2, CAI categories overlap for ET and GT, but overlap with ED is rather limited. Any pooling of ET/GT and ED CAI data would require significant data normalisations, which may only stress the incomparability of ED and ET/GT CAIs.

Ofgem relied on within-sector benchmarking in RIIO-ED1, noting the number of available comparators, i.e. 14 DNOs managed by six companies / groups. This enabled benchmarking to be conducted with a higher degree of confidence in the comparability of companies, their operating environment and size, and across a reasonable number of comparators. However, the benchmarking of ET, GT and GD is less fortunate in this regard.

In the case of GT, within-sector benchmarking is not an option as there is only one company: NGGT (TO). Therefore, benchmarking would need to involve either cross-sector or external comparators.[3] In the case of ET, there are three companies and in GD eight networks in four companies (and three ownership groups).

Provided costs are defined and normalised in similar ways, cross-sector benchmarking could provide greater confidence in the resulting benchmark as a higher sample size would make econometric estimates more precise. Using external comparators further expands the number of comparators but as discussed below, the comparability of activities of the external comparators to the network companies is more challenging to demonstrate.

Furthermore, while adding comparators from other sectors may help improve the precision of model estimates, if the benchmark significantly *changes* with the inclusion of other sectors or external comparators, this could either signal that they are not appropriate comparators, or a detailed argument would need to be made that the networks need to match the efficiency of the additional comparators.

Our model selection process for this report started with the (strong) assumption that data from the GD, ET, and GT networks can be pooled together for BSCs, and ET pooled with GT for CAIs, maximising the available sample size. However, it was understood by Ofgem that this depended on the robustness of the model. If the cross-sector model proved not to be robust, we turned to GD-only and T-only (still pooling ET and GT) models. We discuss the potential inclusion of external comparators or ED DNO data below.

External benchmarks

The usefulness of any benchmark depends on the comparability of activities underlying the costs. As BSCs relate to activities which support general business, rather than network operations, in principle they have *some* comparability across different network sectors. However, in the case of external comparators, there is a greater likelihood of differences in

---

[3] In principle, international gas transmission companies could be used, if (similarly categorised) BSC data were available, but differences in their operating and regulatory environments would still need to be recognised and accounted for.

costs arising from genuine differences between sectors and their operating environments than there is for network comparators.

Identifying and understanding the extent of comparability of activities is, therefore, an important element of using external benchmarks to set BSC / CAI allowances. For example, for RIIO-GD1/T1, in recognition of the possible higher regulatory burden of network companies compared to the external comparators, the benchmark for CEO and group management was set as a composite of an external benchmark and the network companies' own costs. We also note that, whilst at an aggregate level there was very little difference between the external benchmark and the networks benchmark, there were material differences in some *individual* BSC categories. This highlights the potential for inconsistencies in assessing costs at an activity level rather than a top-down approach (as discussed in Section 4.4).

External benchmarks, particularly at a disaggregated level, tend to rely on proprietary data and the underlying methodology, sample, and data might not be publicly available. This may limit the understanding of the data and the corresponding ability to make appropriate adjustments to allowances. A lack of transparency in the data can also undermine the transparency of decisions. This issue arose in RIIO-GD1/T1, where companies protested that the external benchmark's underlying data was not made available to them.[4]

External benchmarks also limit the choice of cost drivers, as those that are not common across all the comparators cannot be used. This would include candidate cost drivers like MEAV and network length. In turn, the choice of cost drivers in cross-sector benchmarking is more limited than within-sector, e.g. network length could not be used as a meaningful driver when comparing GD, GT, and ET.

Given significant concerns about comparability and transparency, we did not include external comparators in our analysis.

Selecting the sample period (historical and / or forecast data)

External benchmarks are typically limited to historical cost data, whereas network companies submitted forecast BSCs as part of their business plans. Both cross-sector and within-sector benchmarking can incorporate forecast data. Combining historical and forecast data also increases the sample size. In the case of RIIO-ED1 CAIs, which relied on only DNO data, the benchmark was calculated from forecast data (2014-15 to 2022-23). This was in recognition of an industry-wide reduction in costs across DPCR5. Only using historical data resulted in modelled costs significantly greater than submitted costs and models combining historical and forecast data did not pass statistical tests.[5]

Forecast data can be helpful to inform appropriate allowances, particularly where the past is not a good guide to the future. For example, historical data quality may be considered weak if companies were asked to report historical data in ways that their reporting

[4] Ofgem, GD1 Final Proposals – supporting document – cost efficiency, 5.9.
[5] Ofgem, RIIO-ED1: Final determinations for the slow-track electricity distribution companies – Business plan expenditure assessment, 10.16.

systems did not allow, which could lead to a large number of allocation assumptions.[6] Or, as may have been the case for RIIO-ED1 CAIs, unit costs could be expected to significantly change between the historical and forecast periods. Hence, it may be better to set a baseline cost assessment with forecast data, using networks' forecasts to further scrutinise and understand their expected costs, and then use Special Cost Factor claims to explain any variations.

However, forecast data is inherently uncertain and reliant on company views (which might be subject to bias). Company bias may be an even greater concern when drawing forecasts from a small sample of companies. For BSCs, the exact relationship between costs and cost drivers can be unclear. Hence, it may be better to rely on what has been shown to be technically achievable by historical data. For example, in CEPA's guidance to the UR for NIE's RP6 price control, it advised to focus on actual data when benchmarking Northern Ireland Networks against GB DNOs. With some 60 observations (4 years x 15 DNOs), this was considered a sufficient sample to rely on historical data and it was argued that using historical data meant allowances were set according to *currently technically achievable efficiency levels* rather than using forecasts that may be mistaken in hindsight.[7]

Alternatively, *further* historical data could be utilised, e.g. data from GD/TPCR. However, including older data raises a number of issues: the operating environment of past price control periods may significantly differ from today, any subsequent ownership changes would need to be accounted for, historical data extends the time series but it does not expand the number of comparators in a panel dataset, older data may require significant data normalisations (and readily normalised data was not available for this analysis), and questions would arise if including older data significantly changes the benchmark compared to a model that only uses RIIO-1 data.

Our modelling approach focused on using historical data from RIIO-1 given concerns about relying on uncertain forecast data that is potentially biased by company views and may be inaccurate in hindsight. However, if there was reason to believe the nature or level of BSCs may change in the future given companies' RIIO-2 submissions, we did not rule out considering a model with forecast data in our model selection process.

Consistency with the rest of the price control regime is another consideration in the choice of sample. We were conscious that Ofgem may wish to consider the use of forecast data in other areas of the RIIO-2 price control. Using forecast data may also be a reasonable compromise when faced with limited sample sizes and / or concerns about the robustness or relevance of historical data. **Hence, we still conducted sensitivity checks on our regression models with added forecast data to confirm general consistency with the historical data-only results.**

Including ED data

One consideration is the inclusion of ED data. Adding ED data would increase the sample size, which may improve the precision of benchmarking estimates, but, as shown in

[6] Uncertainties about cost allocations relates to our preference for assessing BSCs and CAIs at a group and aggregated level, as discussed below in Sections 4.3 and 4.4.
[7] UREGNI, Northern Ireland Electricity Networks Ltd – Transmission & Distribution 6th Price Control (RP6): Final determination, 30 June 2017, 5.77.

Section 2.2 for CAIs, there may be significant inconsistencies in the cost categories of ED and those of GD/GT/ET (see Table 1 and Table 2).[8] The ED price control is also on a different timeframe, so its data may need significant normalisations to ensure time consistency with GD/GT/ET.

Furthermore, as discussed above, questions would arise if adding ED data significantly changed the benchmark. Suppose a case where the statistical fit of a GD-only data model is poor. If adding ED data improves the model fit, it does not necessarily give confidence that the model now better represents GD costs. The reason for any difficulties in modelling the costs of a specific sector should first be investigated within that sector. Adding potentially incomparable data from another sector to the model does not resolve the issue. If the inclusion of ED data lowered the benchmark, a strong argument would need to be made that DNOs are, like-for-like, more efficient than the GDNs, i.e. that the data has been sufficiently normalised and GD and ED can be considered to have sufficiently comparable operating environments, in addition to questions as to the extent the other networks can 'catch up' to the efficiency of DNOs in RIIO-2.

Given these issues, **we ruled out the inclusion of ED data for our analysis**, but Ofgem may still wish to consider their inclusion as a sense check if sufficiently normalised ED data can be feasibly developed.

4.2 Gross versus net costs

Some networks have previously allocated elements of BSCs elsewhere, e.g. to other indirect opex, to direct opex, or to capex. In RIIO-GD1/T1, Ofgem assessed BSCs on a gross basis by adding back elements of BSCs previously allocated "*to direct opex, capex, or repex, or to non-network businesses*".[9] The gross to net ratios across the four groups, reported at initial proposals, ranged from 4.4% (for National Grid) to 20.1% (for SSE).[10] Using this 'gross' basis provides more consistent data as networks may have different policies and approaches to allocating BSCs. **We recommend continuing to assess costs on a gross basis**.

Ofgem then set BSC allowances by converting the assessed gross costs back into net terms "*in the same proportion as in the companies / groups submitted forecasts*".[11] We think this remains a reasonable and practical approach for any required gross to net readjustments. For CAIs, there is no explicit reference in RIIO-T1 as to whether they were assessed on a net or gross basis.

[8] For BSCs, Procurement is not a category for ED, HR and Non-Operational Training are separate categories for ED yet are jointly reported for GD, GT, and ET, and Fines & Penalties are separately reported for ED, but not in GD, GT, and ET. For CAIs, ED has categories for Call Centre, Control Centre, and Wayleaves, which are not present for transmission CAIs. Unlike transmission, ED does not have separate CAI categories for Health, Safety, and Environment, Market Facilitation, Network Planning, and Operational IT&T.
[9] Ofgem, RIIO-GD1: Initial proposals – Supporting document – Cost efficiency, Appendix 6, July 2012, 1.12.
[10] Ofgem, RIIO-T1: Initial Proposals for National Grid Electricity Transmission and National Grid Gas, July 2012, Table A4.2.
[11] Ibid.

From our review of the BPDTs:

- For GT and GD, BSCs were requested both pre- and post-capitalisation, but only pre-capitalisation data were collected for ET
- Allocation of any BSCs to direct opex were not collected in any sector.[12]

As a consequence, if Ofgem *did* want to assess BSCs in net terms (for ET) it would need to separately request the data and further data would also need to be requested to understand the extent of any allocations to other cost categories (i.e. repex and direct opex). However, if Ofgem assesses costs in gross terms, as we recommend, these differences do not matter.

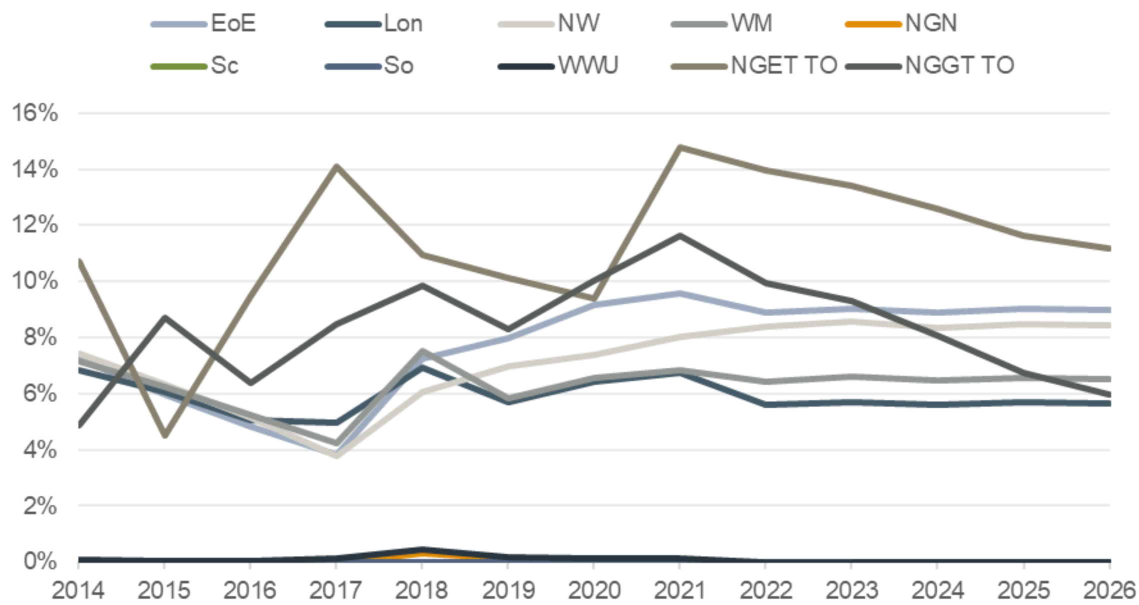
Evidence from business plans

Figure 6 shows the difference between gross and net BSCs, expressed as a percentage of gross costs over time and by network, i.e. the proportion of gross costs allocated elsewhere, as submitted in the companies' business plans. For NGGT (TO), transmission owner (TO) data is presented separately, whilst no data are available for the Scottish electricity transmission networks.[13]

This chart shows material differences between networks (ranging from no allocations to 15%) as well as within sectors (e.g. allocations for GDNs range from zero to 5-10%). There are also material variations in allocations for some networks over time (e.g. 5% to 15% for NGET). These call into question relying on net data.

[12] The draft GD BPDT provided for reporting of BSCs by capex, repex, direct opex, and other, but these categorisations were removed for the final BPDTs. More generally, we are unaware of the size of any reallocations to direct opex in setting RIIO-1 allowances but note that Ofgem stated BSC allocations were to "*to direct opex, capex, or repex, or to non-network businesses*", and not just capex.
[13] We understand that separate gross and net data were not requested for ET, but net and gross BSCs for National Grid ET are available from National Grid GT's business plan.

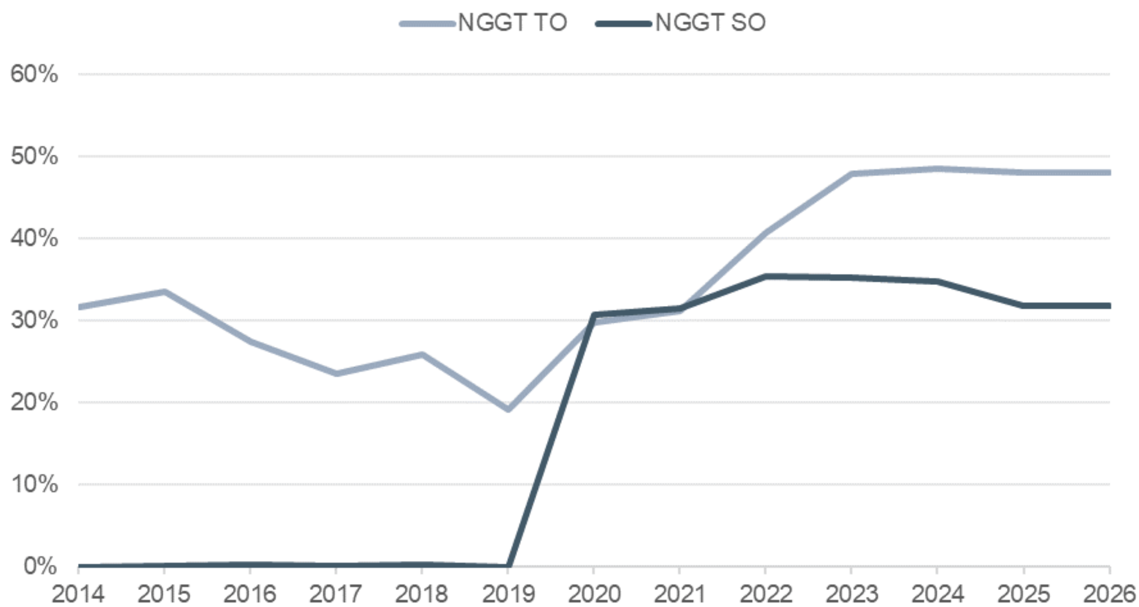
Figure 6 Proportion of gross BSCs that are "allocated"



Source: ECA analysis of Ofgem data from companies' business plans

Figure 7 shows the difference between gross and net CAIs, expressed as a percentage of gross costs over time for the GT TO and SO, i.e. the proportion of gross costs allocated elsewhere (no equivalent CAI data was made available to us for ET). The variation both in the values of each over time and between the two series, particularly into the forecast period, further calls into question the use of net data for any benchmarking.

Figure 7 Proportion of gross CAIs that are "allocated" (NGGT)



Source: ECA analysis of Ofgem data from companies' business plans

Stakeholder views

In response to Ofgem's June 2019 *RIIO-2 tools for cost assessment* consultation, three GDNs commented in support of a gross basis for BSCs, but issues were also highlighted:

- SGN stated that a net to gross adjustment is "*essential to the robust assessment of BSCs, and is worth re-visiting to ensure it is done as accurately as possible.*"

- Cadent noted that if gross BSCs are used, then those costs previously allocated elsewhere need to be stripped out if they are not to be assessed twice (i.e. once in gross BSCs and once in the area they were allocated to) and that this would be challenging; and

- NGN and SGN linked the issue of the use of gross costs to issues arising from (respectively) a GDN's proportion of contractors or whether it operates a partnership model. GDNs "*that make more use of contractors may have more overheads allocated within contractor costs, rather than within BSCs.*"[14] And "*Separating out and reallocating overheads from contractor costs can be challenging, but is important as different business model choices shouldn't impact benchmarking results other than the extent to which they have a genuine impact on efficiency.*"[15]

Considerations as to whether to assess net or gross indirect costs

The importance of whether to assess costs in net or gross terms partially depends on the overall approach to assessing indirect costs (BSCs / CAIs). An aggregated benchmarking approach covering *all* BSC or CAI expenditure categories will not be affected by networks' cost allocations between activities *within* these categories of cost. However, where the costs of different BSC / CAI activities are separately assessed (i.e. under a bottom-up approach) then networks' decisions as to how to allocate costs between them may distort the separate assessments.

In a bottom-up approach, whilst the outcome of assessments in each separate activity may be distorted by networks' different cost allocation decisions, the consequence for a network across activities is unclear. For example, differences in cost allocation may result in a network looking more efficient in one activity and correspondingly less efficient in another but with no material effect on the overall assessment across both activities.

An assessment in gross or net terms will only likely result in different outcomes if network companies have different cost allocation approaches which result in meaningful differences between networks' net or gross costs. In this regard, a review of cost allocation methodologies could establish whether there are material differences between companies that could undermine cost comparisons. **Data from the RIIO-1 and the RIIO-2 business plans show that the differences between net and gross costs can be material and that there are significant differences in allocation between networks.** This means, as part of a bottom-up approach, the choice over whether to assess BSCs and CAIs in net or gross terms may affect the outcomes of the separate assessments.

[14] SGN, *RIIO 2 Tools for Cost Assessment Consultation SGN Response*, August 2019, p. 15-16.
[15] *Ibid.*

Given these apparent differences, **conducting the assessment of BSCs and CAIs on a gross basis helps minimise distortions arising from differences in cost allocations**, thereby allowing for a more like-for-like comparison of costs between networks. As already noted, respondents to Ofgem's cost assessment document generally supported the use of gross BSCs, which was also the approach in RIIO-GD1/T1. A related question is whether to assess at a group or network level.

4.3 Group or network level

Differences between networks' BSCs can arise from differences in cost allocations. A benefit of benchmarking group level costs, in principle, is that any differences arising from cost allocations are avoided, enabling a more consistent comparison of BSCs.

A review of network-level data for GDNs reveals some BSC trends within groups that need to be understood if they were to be used rather than group-level costs. In particular, [REDACTED] shows the cumulative change[16] (from 2014) in the BSCs of Cadent's four networks. Through to 2018, the four are on the same downward trend. However, marked differences emerge between the networks in 2019. Whilst these may be the result of fundamental changes in the individual performance of the networks in that year, it appears likely that these differences came about from a change in cost allocation policy, post the change in Cadent's ownership.

[REDACTED]

Given the trends noted above, **we consider group-level benchmarking as our default approach for GDNs**. Otherwise, for example, if allowances were set at the network level, a network within the Cadent group could be unfairly docked for the cost reallocation decisions apparently made in 2019 despite an apparent overall reduction in BSCs across the Cadent networks for RIIO-1.
However, we still consider network-level benchmarking as a sensitivity check in our modelling process. We would generally expect that group and network level assessments provide similar results. Any significant differences between group and network level assessments may require further investigation, justification, and/or adjustments.

For transmission , there is less scope for group level assessments as there are only four companies to assess: NGGT (TO), NGET, SHET, and SPT. **We therefore have focused on network-level benchmarking for the transmission companies**.

We considered sensitivity tests where NGGT (TO) and NGET were grouped under National Grid, and SHET and SGN (33% owned by SSE) were grouped under SSE in a cross-sector regression. However, if ED data were readily available, SSE's ownership of two DNOs (Scottish Hydro Electric Power Distribution and Southern Electric Power Distribution) and Iberdrola's ownership of both SPT and one DNO (SP Energy Networks), would also need to be considered. This necessarily entails a cross-sectoral approach, which has incomparability risks (as discussed in Section 4.1). We therefore had little confidence in the cross-sector, group-level regressions that we checked, which for some

[16] Cumulative change is used as a control for scale.

companies gave very different (and implausible) results from our sector-specific, network- and group-level regressions.[17]

Post-modelling adjustments could be considered to control for this, as Ofgem did with SSE for RIIO-1. Ofgem set allowances for SGN based 50% on its costs and 50% on those of SSE's (in recognition of SSE's, at the time, 50% ownership of SGN). However, the scope for such an adjustment is less clear for RIIO-2 with SSE's ownership stake in SGN since dropping to 33%.

If a group level benchmarking of BSCs is pursued for the final determination, there are a number of practical considerations and decisions to be made:

- While we have applied groups in GD[18] and assessed the transmission companies at a network level, the issues laid out above regarding cross-sector ownership need to be considered by Ofgem for the final determination, as was performed for SSE and SGN for RIIO-GD1/T1.[19] Ofgem will need to decide on what constitutes a group and which networks should be assessed (in part or full) together. For example, in addition to the SSE example, the electricity system operator is now legally separate from (but still owned by) National Grid Group. WWU and NGN have some common ownership,[20] but one owner having financial stakes in multiple companies does not necessarily imply that costs should be considered as being pooled.[21] More relevant is if they are considered to have a common *operator*. These decisions can be informed by data on group accounting and cost allocation policies.

- Where historical data are used, they may need to be interpreted in the context of any group changes (e.g. Cadent's change in ownership in 2017 could limit the time series of consistent data).

- A group level approach means there are fewer observations than under a network level approach. This may have implications for the robustness of results.

- Group level assessments need to be converted into network level allowances. In RIIO-1, Ofgem allocated group allowances to networks in the same proportions as in the groups' submitted costs. Figure 8 shows how BSC

---

[17] For example, the cross-sector, group-level MEAV ratio benchmark suggested that SPT's BSCs should be cut by almost 70%.

[18] Our GD groups (and standalone networks) being: Cadent (East of England, North London, North West England, and West Midlands), Northern Gas Networks, SGN (Scotland and Southern), and Wales & West Utilities.

[19] We note that this adjustment was made in response to SGN disagreeing that it should be treated as part of the SSE group, while Ofgem noting that ~25% of SGN's BSCs are allocated from SSE: Ofgem, RIIO-GD1: Final proposals – Supporting document – Cost efficiency, Appendix 6, December 2012, 1.17.

[20] WWU is owned by CK Infrastructure Holdings Limited, which also holds a stake in NGN, as well as a DNO (UK Power Networks).

[21] For example, for RIIO-ED1, the CAIs assessment was conducted at a network level due to insufficient evidence of shared costs within groups.

allowances were split within groups for RIIO-1. We believe this remains a reasonable and practical approach.

Figure 8 Network split of group business support costs

Network		Group							
		National Grid	NGN	SSE	WWU	National Grid	NGN	SSE	WWU
		RIIO-T1/GD1 Submitted cost % split				RIIO-T1/GD1 calculated total allowance % allocation			
T	NGET TO	24.29%				23.78%			
T	NGET SO	18.74%				19.42%			
T	NGGT TO	8.59%				8.41%			
T	NGGT SO	9.31%				10.14%			
T	SHETL			8.65%				10.45%	
GD	East of England	13.11%				12.83%			
GD	London	8.43%				8.25%			
GD	North West	10.13%				9.91%			
GD	West Midlands	7.41%				7.26%			
GD	Northern		100.00%				100.00%		
GD	Scotland			14.31%				14.03%	
GD	Southern			26.08%				25.57%	
GD	Wales & West				100.00%				100.00%
ED	SSE Hydro			21.39%				20.96%	
ED	SSE Southern			29.58%				28.99%	

Source: Ofgem, RIIO-T1: Initial Proposals for National Grid Electricity Transmission and National Grid Gas, July 2012, Table A4.6. For the final determination, Ofgem changed its approach to SGN's group costs, separating it from SSE for initial benchmarking. Ofgem then set SGN's baseline allowance half on SGN's baseline and half on SSE's (in recognition of SSE's, at the time, 50% ownership of SGN).

As some respondents to Ofgem's cost consultation noted (see below), to the extent that there are fixed costs and Ofgem is minded to consider taking these into account in setting allowances, a group level assessment will be beneficial (as it is within the group that the fixed costs are more likely to be incurred). However, in RIIO-1, Ofgem rejected the notion of applying fixed cost adjustments, arguing this would imply that Ofgem would need to reassess costs every time a network's ownership changes. We further discuss fixed costs in Section 4.5.

Stakeholder views

Respondents to Ofgem's June 2019 *RIIO-2 tools for cost assessment* consultation were largely supportive of assessing BSCs at a group level. Reasons for this included that it helps address distortions that may arise from different cost allocation approaches between groups, that some costs are incurred centrally, and that it may help understand economies of scale. Whilst supporting assessment of group level costs, one respondent (SGN) observed that a group level assessment means fewer comparators, which could affect any regression benchmarking. Networks also linked this issue to that of fixed costs (which are discussed below in Section 4.5).

4.4 Activity-level or aggregated

BSCs can be assessed at either aggregated or disaggregated levels. Assessment at the disaggregated level involves separately assessing each of the individual categories (e.g. HR, procurement, IT&T, etc.), before they are summed to obtain total allowances for

BSCs/CAIs. Alternatively, BSCs / CAIs can be assessed in aggregate, i.e. a 'top-down' assessment.

In principle, by abstracting away from individual cost items, an aggregated approach can help avoid issues arising from differences in cost data reporting (which necessitates assessing CAIs at an aggregate level), business strategies, activity definitions, etc., which are more acute in disaggregated assessments.

However, it should be noted that aggregate assessments can only help for differences that are the result of allocating costs *between* BSC / CAI categories. To the extent that, for example, different operating models result in higher BSCs (i.e. opex) but lower capex, this will not be addressed simply by using an aggregate approach to BSCs. For example, a network that outsources some IT&T costs may have higher opex but lower capex than one which retains all IT&T activity in-house. It was for this reason that in RIIO-ED1 Ofgem included non-operational capex for IT&T (the most material BSC category) in its assessment of BSCs. **This report does not consider the inclusion of costs outside of BSCs/CAIs in our assessment**, but this trade-off needs to be broadly kept in mind by Ofgem for its decision on the BSC / CAI assessment approach.

A clear understanding of these cost trade-offs is required if any cost adjustments are to be applied to facilitate benchmarking. Such understanding could be developed through a disaggregated assessment, even if a disaggregated approach is not used in benchmarking BSCs. Disaggregated models do have one advantage in that their interpretation is usually clearer compared to top-down models. For example, we noted a step-change in insurance costs for SHET and SPT in RIIO-1, which motivated a sensitivity check where insurance costs were excluded from total BSCs.

One challenge with a disaggregated benchmarking approach is the choice of benchmark level (discussed in Section 5.4). Choosing, say, an upper quartile level of performance for each category of BSCs would likely result in an *above* upper quartile performance at the aggregate level. Such concerns were voiced by respondents to Ofgem's initial proposals in RIIO-GD/T1, which was one of the reasons Ofgem switched from disaggregated benchmarking of BSCs to an aggregated approach for final proposals.[22]

In deciding which business support activities to assess together and which activities may need individual assessments, regulators need to be mindful of the risks (and accusations) of inconsistency across activities and 'cherry picking'. However, a lack of comparators can limit confidence in top-down models. In its RIIO-ET2 Sector Specific Methodology Decision (SSMD), Ofgem said that it was likely to rely on bottom-up benchmarking (supplemented by engineering and other expert knowledge) more heavily than in the distribution sectors due to there being fewer comparators and the specific nature of projects.[23]

Our modelling for this report was limited by the lack of some cost driver data that has been used for previous activity-level benchmarking:

[22] Although it is noted that SP Energy Networks disagreed with this assertion in the summer consultation.
[23] Ofgem, RIIO-2 Sector Specific Methodology Decision, Annex: Electricity Transmission, 5.2.

- We have not been provided with networks' Revenue forecasts. For RIIO-GD1, Revenue was used as the cost driver for Finance, Audit, and Regulation; Property Management; and CEO and Group Management.

- We lacked IT end-user data for multiple companies. IT end-users was considered as a cost driver for IT&T costs for both RIIO-GD1/T1 and RIIO-ED1.[24]

MEAV was proposed as an alternative cost driver for all the above cost categories for RIIO-ED1 and was ultimately used as the cost driver for the final determinations. Given we use MEAV as our primary cost driver for our top-down BSC modelling (see Section 5.1 below), cost driver data limitations, and our general concerns about activity-level assessments, we did not consider activity-level benchmarking to be a necessary sensitivity check. Given available cost drivers, activity-level benchmarking would have largely amounted to adding up individual MEAV-based regressions, which would have provided limited new information. Alternatively, our CSV specification (Section 6.2.1) controlled for the impact of FTEs (weighted by HR costs) and Totex (weighted by Procurement costs) in our top-down analyses.

In summary, we conducted our benchmarking from an aggregated, top-down perspective given:

- an inconsistency over time in NGET's reporting of the CAI category System Mapping;

- the advantage of aggregate benchmarking avoiding issues of differing cost allocations between BSCs;

- the need to avoid 'cherry picking';

- missing cost driver data; and

- the regulatory precedent of Ofgem's most recent price controls.

4.5 Fixed cost normalisations

It has been argued that elements of BSCs, particularly IT&T, which Ofgem is already assessing as part of a separate expert review, have a large proportion of fixed / overhead costs. Where this is the case, networks that are part of a group may benefit from economies of scale, with the fixed costs spread across several networks. In none of RIIO-1 price controls did Ofgem make normalisations for fixed BSCs, although it considered them in developing its BSC approach for the ED1 slow track determinations. Ofgem ultimately adopted a ratio benchmark approach for its ED1 slow-track determination and argued that by assessing at a group level, it accounts to some degree for shared costs

[24] However, we note that Ofgem is undertaking a separate expert review of IT&T.

within a group.[25] Ofgem has also argued that allowing for fixed cost adjustments would mean that adjustments would need to be made every time a network was purchased or divested from a group.[26]

Taking a regression approach to modelling would to some degree account for fixed costs. A regression equation's 'constant' (or 'intercept') should capture the fixed cost element of an activity. Furthermore, the interpretation of the constant would be complicated by the use of a panel data model, such as Random Effects or Fixed Effects, where each network would effectively have its own constant, or the inclusion of dummy variables, which also apply a 'fixed' effect to regression equations.

Stakeholder views

A number of respondents commented on the issue of fixed costs / economies of scale. National Grid do not support adjusting BSC benchmarks for fixed costs. WWU and NGN both noted that a group analysis would help understanding economies of scale and WWU stated that stand-alone GDNs will always have a level of fixed overhead that will look more inefficient than a group. WPD thought Ofgem should consider some form of scale driver in assessing BSCs at a group level. NPG noted that if theory and evidence supports the presence of fixed costs then it is important for Ofgem to ensure allowances are adequate for the fixed costs of smaller groups, and that Ofgem need to consider how to do this irrespective of whether it benchmarks at the group or network level.

We did not make any explicit fixed cost normalisations in our modelling. However, we reviewed whether our ratio benchmarks appeared to be biased against smaller / standalone networks relative to groups in comparison to our regression-based approaches, where the presence of a constant should partly account for fixed costs. Furthermore, our preference for group-level assessments should account for fixed cost sharing between networks within a group.

4.6 Cost exclusions: IT&T and insurance costs

Excluding IT&T costs, for which the presence of fixed costs is most apparent, from our baseline modelling should help ameliorate the problem caused by fixed costs. Removing IT&T costs from our baseline models is an example of removing a 'non-controllable' cost, which networks may find difficult to 'efficiently' control and setting a benchmark may therefore not be appropriate.

Our default approach was to exclude IT&T costs from our benchmarking for both BSCs and CAIs. This was in acknowledgement of the difficulty of forecasting IT&T costs, which in past regulatory decisions have often been subject to expert review. Hence, Ofgem is conducting a separate expert review of IT&T costs. However, **we also conducted sensitivity checks on our models where we included IT&T, while also excluding**

[25] Ofgem, RIIO-ED1: Draft determinations for the slow-track electricity distribution companies – Business plan expenditure assessment – Supplementary annex to RIIO-ED1 overview paper, 10.33-10.36.

[26] Ofgem, RIIO-ED1: Final determinations for the slow-track electricity distribution companies – Business plan expenditure assessment – Supplementary annex to RIIO-ED1 overview paper, 10.47.

insurance costs for BSCs. The reason for the latter being that ththat SHET and in particularl SPT experienced a step-change increase in insurance costs in RIIO-1 that is forecast to continue into RIIO-2, and it has been argued that differing insurance costs may better reflect different risk appetites and / or appropriate insurance coverage levels rather than being an indicator of 'efficiency'.[27]

Ofgem may also consider excluding 'atypical' / 'exceptional' costs, which are not expected to recur.[28] This helps enable a more meaningful comparison over time and focuses the analysis on controllable (or more easily 'predictable') costs. This can be supported by requiring companies to explain / justify atypical costs in their business plans. Our analysis of the BPDT data did not identify any 'atypical' cost exclusions and excluding any costs from the final benchmark would need to be well-justified by evidence of either high uncertainty in forecasting a cost category or an expected step-change that may be difficult to account for with a benchmark based on historical data. Hence, for the exclusions we did consider – IT&T and insurance – our modelling process considered sensitivity checks where these costs were all included. As discussed in Section 6, we concluded that a separate IT&T assessment (as Ofgem is doing) may be appropriate for GD BSCs, while excluding IT&T and insurance costs does not appear to explain SPT's inefficient result for BSCs for the transmission companies. Including IT&T costs does not materially change the modelling result for CAIs for the transmission companies

[27] Multiple companies have commented in favour of excluding both IT&T and insurance costs.
[28] Atypical costs may arise from activities such as reorganisations and head office moves or one-off regulatory and legal projects. If any such atypical costs were considered excludable, Ofgem would ideally need to provide clear and consistent *ex ante* guidance to companies as to what qualifies as atypical.

5 Cost driver selection and model specifications

Box 4 Model specifications and identified cost drivers

- Our cost drivers of choice were MEAV for BSCs and both Total Capex and MEAV for CAIs. However, we tested other cost drivers and constructed Composite Scale Variables (CSVs) for BSCs.
- Our model selection approach began with Pooled Ordinary Least Squares (POLS) regression models, but consideration was also given to panel-based estimators.
- Ratio benchmarks were considered if regression-based models were not deemed to be reliable.
- Models were first evaluated at the average (for regressions) or median (for ratio benchmarks) model fit. If the models were considered sufficiently robust, we considered a more stringent benchmark, e.g. setting the benchmark at an upper quartile.

We discuss the various modelling specifications and choices that need to be made when conducting benchmarking. Namely, these choices are what cost driver(s) should be used, what econometric specifications should be employed, what options exist for non-regression-based benchmarking, and what benchmark to ultimately set for the network companies.

5.1 Cost drivers

For any benchmarking approach, whether regression-based (Section 5.2) or ratio benchmarking-based (Section 5.3), the chosen cost driver is a key factor. This section describes the many considerations in choosing a cost driver for BSCs and CAIs.

5.1.1 Principles for a 'good' cost driver

Ofgem has set out that a good cost driver should:[29]

- **Make economic and/or engineering sense**, so that it can be interpreted and understood as reasonable and relevant.
- **Be accurately and consistently measurable.**
- **Have a relatively stable relationship with costs over time and incorporate as much relevant information as possible:** helps distinguish between costs which are explained by differences in exogenous conditions and costs which are explained by differences in efficiency.

[29] Ofgem, RIIO-2 tools for cost assessment: Consultation, June 2019, 2.26-2.32.

- **Be beyond the control of the network company**, as far as reasonably practical, to avoid distorting company incentives in ways which may increase inefficiency.

However, Ofgem acknowledges that **the choice of cost driver will involve trade-offs**. Not all cost drivers will necessarily perform well against all these principles.

A key concern with selecting cost drivers is limiting the potential for perverse incentives. Ideally, the network will have a limited ability to manipulate the chosen cost driver and (inefficiently) increase its allowance, i.e. the driver can be considered exogenous rather than endogenous. Hence, Ofgem has sought in recent price controls to use scale or workload cost drivers rather than direct expenditure.

For any chosen cost driver, a 'plausible casual narrative' should be set out as to what might be expected of chosen cost drivers. A cost driver may also be rejected if its coefficient violates economic / technical rationale, i.e. a negative coefficient or a coefficient greater than 1, which would imply diseconomies-of-scale.

The inclusion of year dummy variables or a time trend variable were tested for all models. These would control for real expenditure changes over time relative to cost drivers and explanatory variables. This could be expected due to ongoing efficiency or Real Price Effects (RPEs). However, if a time trend variable's coefficient is significant and sizeable, caution is needed in projecting its effect forward into RIIO-2 as the observed time trend may only be a temporary effect and a simple linear extrapolation may not be appropriate. This is not a concern with year dummy variables, but their inclusion limits degrees of freedom given small sample sizes.

5.1.2 Composite Scale Variables

One option when multiple cost drivers are under consideration, but there are concerns about including more than one in a regression - whether due to limited degrees of freedom (due to a low sample size), difficulties in interpreting multivariate models (i.e. the CMA's criticism of Ofwat's PR14 modelling), or multicollinearity when cost drivers are highly correlated - is to combine multiple cost drivers into a single Composite Scale Variable (CSV). CSVs can serve to capture the impact of multiple cost drivers while maintain degrees of freedom. CSVs also have regulatory precedent within Ofgem, including for RIIO-1.

CSVs have drawbacks: their interpretation is not necessarily more intuitive than a multivariate regression model. The determination of the weights of a CSV also require pre-modelling decisions that may be seen as arbitrary. Nevertheless, they were considered as an alternative to single-variable models for our top-down analyses. We discuss our construction of CSVs for our BSC regression analyses in Section 6.2.1.

5.1.3 Identifying BSC cost drivers

As discussed in Section 4.4, we have focused on a top-down assessment of BSCs due to its apparent advantages over activity-level assessments and the unavailability of some cost driver data. In aggregate, BSCs reflect a mix of both semi-variable and 'fixed' costs that will increase by step changes in response to both size / volume and the complexity of an organisation. Hence, while scale-related drivers are the starting point for BSCs, an ideal cost driver will suitably reflect both scale *and* activity.

Therefore, **Modern Equivalent Asset Value (MEAV) was our preferred cost driver given it simultaneously reflects the scale, complexity, characteristics, and composition of the network asset base**. MEAV is effectively a weighted asset scale index, with weights determined by the assumed unit costs of the asset base. MEAV was used as a cost driver for RIIO-ED1 BSCs, where it was selected as the most appropriate driver for aggregated BSCs,[30] and it should be familiar to the networks.

MEAV may in principle be under the control of networks, which choose the assets they invest in and the asset standards they target. Using MEAV may then incentivise networks to (inefficiently) increase MEAV to increase the regulatory benchmark or subsequently run a relatively poor-quality network in order to outperform the benchmark. However, it is questionable how materially networks can impact their MEAV in the short run. Networks would need to spend significantly to influence benchmarking results relative to the historical value of the network asset base. Network assets are also long-lived assets that may more reflect the investment decisions of past management, making their *current value* relatively exogenous with respect to *current operations*. Another concern is that benchmarking based on 'efficient' MEAV today is not necessarily an indicator of what will be efficient in future but addressing this criticism would require a view of the efficient value of future networks, which is highly uncertain.[31]

Other scale-related cost drivers were considered. Customer Numbers can be expected to track scale at least within a sector, i.e. between comparable networks, and so Customer Numbers was tested as an alternative scale-related cost driver within the GD sector. Throughput similarly does not translate across sectors, and it does not necessarily directly relate to scale for infrastructure that is designed to meet peak demand rather than aggregate throughput. Peak Demand may therefore be a better measure of scale given it will track system capacity, but it is a difficult variable to accurately forecast, particularly in the gas sector. Network Length may reflect scale, but the perverse incentive to install more (lengthy) network assets in order to appear more efficient is obvious and network length only accounts for one dimension of scale.

As an alternative to a MEAV-only model, we constructed a CSV for BSCs using available cost driver data (see Section 6.2.1). Our CSV is heavily weighted toward MEAV given it was the (available) cost driver of choice for most BSC categories (Finance,

[30] Ofgem, RIIO-ED1: Draft determinations for the slow-track electricity distribution companies – Business plan expenditure assessment – Supplementary annex to RIIO-ED1 overview paper, 10.37.
[31] We understand that Ofgem will be subsequently applying some sensitivity checks to the company-submitted MEAV values, such as different unit cost assumptions, in order to confirm its robustness as a cost driver.

Insurance, IT&T, Property, and CEO & Group Management)[32], but it also includes FTEs, weighted by HR, and Totex / Total Spend[33], weighted by Procurement.

5.1.4 Identifying CAI cost drivers

Relative to scale driven BSCs, CAIs, being 'closely associated', can be expected to relate more to actual workload on the network. **Given CAIs were not evaluated via regression for RIIO-T1, we turned to the approach taken for CAIs in RIIO-ED1, which considered both scale- and workload-related drivers**.

Workload drivers can be used to control for different network workloads that are outside of networks' control, such as variations in asset condition that drive variations in year-to-year costs, maintenance and repair costs, or repex work due to the inherited state of the network. They should also better capture year-to-year cost fluctuations than scale variables. If a model struggles to capture year-to-year cost fluctuations, it may risk a need for post-modelling adjustments, with companies submitting 'Special Cost Factor' (SCF) requests to the regulator, calling into question the initial model specification.

However, in certain specifications, workload drivers can lead to perverse incentives as they are typically within the control of the company. A company may then be able to artificially / inefficiently inflate its workload projections in order to increase its allowance. Companies may also be incentivised to put forward high workload forecasts, which are not actually delivered.[34]

Pooling ET and GT for the CAIs assessment presents a difficulty in identifying comparable cost drivers. Potential cost drivers such as Customer Numbers, Throughput[35], Peak Demand, and Network Length cannot be considered directly comparable between ET and GT.[36]

Hence, **our top-down analysis focused on cost drivers that were considered to be reasonably comparable between ET and GT: MEAV, which should reflect both scale and activity but may struggle to capture year-to-year cost fluctuations, and Total Capex, which should better reflect actual activity on the network. These cost drivers were considered individually and together** (as part of a multivariate regression).

Asset Additions was considered as a CAI cost driver. It was one of the cost drivers chosen for RIIO-ED1 CAIs (along with MEAV). Its rationale as a cost driver being that while it may not reflect all direct activity, it will reflect both load- and non-load-related activity, and thus does not favour asset replacements over reinforcement or vice versa. Ofgem

---

[32] We also considered Customer Numbers instead of MEAV as a proxy for Revenue in our regressions for GD as Customer Numbers was found to have the highest correlation with Revenue for GDNs in RIIO-1.

[33] We did not have complete Total Spend data to use as the cost driver for Procurement for the transmission companies, so Totex was used instead as a proxy for Total Spend in our cross-sector and transmission sector regressions.

[34] One mitigation to this is that networks' capex forecasts ultimately need to be approved by Ofgem as an efficient view of network's investment needs.

[35] We only received incomplete Throughput data from the Ts.

[36] As detailed above with respect to BSCs, there are nevertheless potential technical, economic, and exogeneity issues with these cost drivers.

produced an estimate of the value of the Asset Additions incurred and forecast by the ETs,[37] but the resulting dataset proved to be very 'lumpy', which led to implausible benchmarked results. This was not unexpected as, compared to ED where new asset additions by the DNOs will be more regular, smaller in scale, and distributed across a network, the value of new transmission assets are likely to be concentrated large-scale, long-term investment projects.[38] **However, in finalising its decision on the CAI assessment approach, Ofgem may wish to explore multivariate regressions that include Asset Additions in addition to a general scale cost driver, such as MEAV, or consider 'smoothing' Asset Additions, such as taking a multi-year average**. While there may be endogeneity concerns with using Total Capex, like Asset Additions, it should reflect both asset replacements and reinforcements.

Separating CAIs into Fixed and Flex components

One option for CAIs is to split the activities into those that may be considered as 'fixed' and less responsive to changes in activity and those that are 'flex' and more directly related to activity on the network. This can allow for cost drivers to be as closely aligned to the activity as possible. This approach was considered for the assessment of CAIs in RIIO-ED1.

[REDACTED][39] This issue highlights the potential for inconsistencies in networks' cost allocation policies, and hence our preference for assessing CAIs (and BSCs) at an aggregated level.

5.2 Regression-based approaches

5.2.1 Choice of estimator

A key question is **whether the econometric approach should explicitly recognise the panel structure of the dataset, which consists of multiple firms over multiple years, while still ultimately having a limited sample size**. Fixed or random effects estimates reinterpret time invariant effects as measures of inefficiency. At the frontier, all firms are then benchmarked against the firm with the lowest fixed or random effect in a cost function. However, Dr Andrew Smith, Professor of Transport Performance and Economics at the University of Leeds, argued in Ofgem's summer 2019 consultation on RIIO-2 cost assessment tools that whether the chosen estimator recognises the panel structure (i.e. random or fixed effects) or not is primarily about estimating the parameters in the cost function rather than the approach to estimating inefficiency.[40]

---

[37] As of the time of writing this report, an asset additions value for NGGT had not been calculated. However, given a generally lower level of activity on the GT network compared to the ET networks, we would expect that an asset additions cost driver for NGGT may be even lumpier than the ET numbers.

[38] [REDACTED]

[39] Nevertheless, we considered this as a sensitivity check for our CAI modelling, but the results were little different from our aggregated modelling.

[40] Andrew Smith, 2019, 'Note for Ofgem on Diagnostic Tests in Efficiency Benchmarking Studies', Ofgem, RIIO-2 tools for cost assessment: Appendix 5, Page 11.

In our discussions with Ofgem, we agreed that our econometric modelling for BSCs and CAIs would focus on techniques that have precedent in previous regulatory decisions and strong small sample properties. Therefore, our preferred specification was Pooled Ordinary Least Squares (POLS), which is relatively simple, transparent, and replicable. The benchmark for a chosen POLS model can be adjusted to a higher benchmark, such as an upper quartile, if deemed appropriate, i.e. a Corrected Ordinary Least Squares (COLS) model (see Section 5.4).

As recommended by CEPA in a June 2019 report for Ofgem on GD cost assessment, we consider Random Effects (RE) models as a sensitivity check to control for the data's panel structure.[41] Other techniques, such as Stochastic Frontier Analysis (SFA) or Data Envelopment Analysis (DEA), were considered too complex and data-intensive given small sample sizes and the need for regulatory transparency.[42]

Fixed Effects (FE) models can control for company characteristics that are unobservable or cannot be sufficiently controlled for in a cost driver model. Fixed Effects effectively applies a dummy variable for each company. A company's time invariant cost effect will then be captured in the regression's intercept term. FE models have advantages, as they take account of the dataset's panel structure and produce unbiased and consistent parameter estimates in the presence of correlation between company effects and cost drivers. Despite these attractive properties, CEPA commented that in limited time series a company's fixed effect may to some degree capture differences in efficiency rather than only time invariant factors and it is a relatively data intensive exercise given small samples.[43] Professor Smith also suggested that fixed effects results may be found to be 'implausible' in a regulatory setting.[44] Nevertheless, we ran FE models as a sensitivity check to gauge the extent to which inconsistency arising from panel effects may affect results. Where there were no OLS models that satisfied our minimum criteria, we considered FE models as an alternative.

5.2.2 Testing for model robustness

Ofgem sets out the following tests to apply when determining (econometric) model robustness:[45]

- **Statistical significance of the coefficients (elasticities)**

[41] CEPA, 'Econometric modelling & regional factors', Ofgem, RIIO-GD2 cost assessment: Annex 1, June 2019, 4.1.3.
[42] Complexity should be avoided where possible to ensure regulatory decisions are clearly communicated and well-understood. For example, some DNOs criticised Ofgem's fast-track assessment business support modelling for RIIO-ED1, which used Monte Carlo simulations, as overly complex: Ofgem, RIIO-ED1: Draft determinations for the slow-track electricity distribution companies – Business plan expenditure assessment – Supplementary annex to RIIO-ED1 overview paper, 10.32.
[43] CEPA, 'Econometric modelling & regional factors', Ofgem, RIIO-GD2 cost assessment: Annex 1, June 2019, Text Box 1: Company FE models.
[44] Andrew Smith, 'Note for Ofgem on Diagnostic Tests in Efficiency Benchmarking Studies', Ofgem, RIIO-2 tools for cost assessment: Annex 5, June 2019, Page 12.
[45] Ofgem, RIIO-2 tools for cost assessment: Consultation, June 2019, Table 2.3.

- **RESET test**: testing for non-linearity with a translog specification
- **Normality of errors**: has implications for finite sample inference
- **Correlation/heteroscedasticity**: can use robust standard errors when assessing statistical significance
- **Testing for panel effects**: valid question as to whether models should explicitly control for the panel structure of the data
- **Endogeneity**: needs to be considered on a case-by-case basis.

Starting from this list, we considered a range of diagnostic tests for our chosen econometric models. These were ranked by importance in our model selection process (Section 6). As with evaluating cost drivers, an econometric model may fail to pass every one of the diagnostic tests. Hence, some judgement is needed as to whether any single diagnostic test can be considered 'fatal' to a model on its own. A cost driver coefficient being statistically insignificant or negative (i.e. violating economic/technical rationale) or greater than 1 (i.e. implying diseconomies of scale) would likely cast doubt over an entire model. However, more specific tests may be less crucial.

For example, Professor Andrew Smith's summer consultation report suggested it would be overly cautious to reject a model based on a failed RESET test alone, arguing its usage may be best for distinguishing between two similar models rather than rejecting a model entirely.[46] Similarly, violations of the normality of errors assumption does not directly affect the properties of OLS estimators. OLS remains the best linear unbiased estimator (BLUE). However, it does have implications for finite small sample inference, calling into question the statistical significance of parameters for small samples. If this assumption is violated, caution may need to be attached to OLS estimates of small sample GD- or T-only models.

5.2.3 Interpretation

All our econometric models have taken a log-log specification. This means that the resulting coefficients can be directly interpreted as elasticities, i.e. a 1% increase in the cost driver translates to an X% increase in costs.

The CMA criticised the large number of explanatory variables used for Ofwat's PR14 modelling given the relatively small sample size, arguing this contributed to a risk of inaccurate results.[47] Of relevance for our modelling of GDN BSCs, where network activity has been relatively flat, the CMA commented on the difficulty of applying multivariate models when the explanatory variables are highly correlated with each other and / or show little variation over time. Hence, we have considered multivariate regression

[46] Andrew Smith, 'Note for Ofgem on Diagnostic Tests in Efficiency Benchmarking Studies', Ofgem, RIIO-2 tools for cost assessment: Annex 5, June 2019, Page 12.
[47] CMA (2015): 'Bristol Water plc – A reference under section 12(3)(a) of the Water Industry Act 1991', p. 72.

models in some cases, but in general **we have sought to use univariate models in order to maintain both ease of interpretation and degrees of freedom**.

Statistically significant coefficients and a strong model fit cannot be the only considerations in determining a cost model. Including squared or interaction terms may improve model fit, but an additional consideration is the need to communicate and explain models in a regulatory setting. The CMA criticised the translog models used by Ofwat for PR14 as being difficult to interpret.[48] Small sample sizes also limit the degrees of freedom for more complex model specifications. We have thus ruled out the use of interaction terms or polynomials.

CSVs also have interpretation difficulties. The coefficients of a log-log CSV econometric model will provide the elasticity of the CSV. However, it is not immediately obvious how to interpret a 1% increase in a CSV. A CSV will be a weighted index of multiple cost drivers, which may also be standardized to prevent the CSV being unduly weighted by its components' scale and units. A 1% increase in a CSV may then only be interpreted as an inevitably vague 1% increase in 'activity'.

5.3 Non-regression-based approaches

5.3.1 Trend analysis

Trend analysis was part of Ofgem's regulatory toolkit for assessing BSCs in RIIO-1. A narrow interpretation is that it involves analysing historical cost trends as a basis for forecasting costs. Under a broader interpretation, in the context of a price control regime, it can comprise a range of analyses, for example:

- analysis of historical trends and reasons for changes / movements;
- analysis of historical performance against price control allowances and reasons for differences; and
- analysis of forecast costs, to understand potential step changes and the network companies' justification for these.

The above analyses can be conducted at both a disaggregated (i.e. per BSC component) level and aggregated (i.e. total BSCs) level, as well as at the individual network level, in network groups, within sectors, and across sectors.

Such analysis can help inform both the appropriate cost normalisation and the overall assessment approach for BSCs. For example, it may help:

- identify step changes in costs (say as a result of new lines of work that need to be undertaken), which may mean that historical benchmarks are less useful

[48] CMA (2015): 'Bristol Water plc – A reference under section 12(3)(a) of the Water Industry Act 1991', p. 72-73.

and greater emphasis being placed on reviewing companies' justification for changes;

- in understanding whether differences in operating models result in differences in reported costs, and what cost normalisations may be required as a result to support meaningful benchmarking; or

- in understanding the confidence that can be placed in particular benchmarks, e.g. if there are strongly divergent trends over time in costs of potential comparators (controlled for cost drivers) or unusual 'lumpy' expenditure, then this may reduce confidence in their use as benchmarks.

As a minimum, trend analysis serves as a useful first step in understanding BSCs/CAIs and informing other analysis. For example, see our examination of BSC trends within the Cadent group in [REDACTED] or our concerns about the 'lumpiness' of the Asset Additions variable constructed for the ETs (Section 5.1.4).

For setting or informing allowances, trend analysis cannot be considered useful if there is reason to believe past trends are not good indicators for the future, and network companies are forecasting increased BSCs/CAIs. In these circumstances, companies' justification for the additional activities and / or increased costs need to be understood and critically reviewed. Hence the need for robust cost drivers to help explain trends via benchmarking.

5.3.2 Ratio benchmarking

Ratio benchmarking involves the comparison of cost ratios, e.g. a cost per unit of cost driver. For example, at an activity-level, for IT&T costs this could be £/IT end-user, or for HR costs, £/FTE. Ratios can also be benchmarked at an aggregated level, e.g. total BSC, as has been done for this report. For example, ratios for total BSCs could be expressed as £/MEAV (as for RIIO-ED1) or £/composite cost driver (as for RIIO-GD1/T1). The choice of cost driver is, therefore, clearly important, as discussed in Section 5.1.

Ratio benchmarking is less sophisticated than regression-based approaches. However, we consider it **a viable option given its simplicity, transparency, and replicability**. It has distinct advantages in a regulatory setting and where econometric approaches can prove unstable with small samples. Ratio benchmarks can transparently guide discussions with networks about their efficiency. Furthermore, ratio benchmarks have regulatory precedence, having been commonly used by Ofgem in past price controls.

5.3.3 Expert review

Expert review is usually used for specific activities (e.g. IT&T or insurance) and is conducted by a specialist consultant in those areas. The analysis often involves an amalgam of the above techniques, i.e. reviewing historical cost trends, analysing business plans and future investment needs, cost benchmarking, etc.

Expert review can be closely associated with conducting business case analyses, which may be undertaken to evaluate major investments. For BSCs, this may be the case for proposed major IT&T investments. Network companies would be required to demonstrate (in a quantitative manner) that the solution they have chosen is the most appropriate and the forecast expenditure is expected to be efficient. Businesses should propose a counterfactual to their proposed investment and expert review may then be needed to evaluate the appropriateness of the counterfactual and the investment itself.

Expert reviews are most likely to be of benefit for more material BSC/CAI categories and where it is challenging to reasonably benchmark activities, either due to the difficulty of comparing costs across networks or other comparator groups or if there is reason to believe that the cost environment for these activities may significantly change in the future. For example, in RIIO-ED1 Ofgem used an expert review of IT&T (covering non-operational capex, as well as opex), as well as a ratio benchmark.

Expert review is not a panacea in cost assessments. Such reviews place a high amount of trust in the judgement of the specialist consultants and are typically costly and time-consuming (as contentious debate can arise between the expert reviewer and the network companies). Expert review was not an option explored for this paper, but we report its potential use here for completeness.

5.4 Assessment and benchmarking options

Once a model has been chosen, consideration needs to be given to the actual benchmark that is applied. In recent price controls, Ofgem has benchmarked at the upper quartile, third, or average. Benchmarks have not been set at the actual 'frontier' due to Ofgem acknowledging it does not have perfect information. There may be measurement error in the data or systemic differences, besides efficiency, between networks that cannot be explained by the available cost drivers and models. Hence, the modelled fit can be 'shifted' to, say, an upper quartile rather than using the efficiency frontier as the benchmark. This adjustment transforms a POLS model into a COLS model. Ofgem's use of an upper quartile has not been challenged by the CMA and it does appear to be a generally accepted starting point for setting benchmarks.[49]

What benchmark is set may depend on the confidence that Ofgem has in both the data inputs and the statistical fit of the model. For example, in DPCR5 Ofgem set a less stringent benchmark of the upper third for network operating costs "*due to greater variability in the data*".[50] For GD1, 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 than GDNs' relative efficiency (i.e. there are statistical errors).*"[51]

For our ratio benchmark models, we have considered taking medians, upper quartiles, and averages of ratios (whether cross-sector or within-sector). Our starting point for our

[49] It was recommended by CEPA as the starting point for evaluating GD2 during the summer consultation: CEPA, 'RIIO-GD2 cost assessment – econometric modelling & regional factors', Ofgem, RIIO-GD2 cost assessment: Annex 1, June 2019, 4.1.3.

[50] Ofgem, 'Electricity Distribution Price Control Review: Final Proposals - Allowed revenue - Cost assessment', December 2019, p. 4.

[51] Ofgem (2012): 'RIIO-GD1 Initial Proposals – Step-by-step guide for the cost efficiency assessment methodology', p. 13.

ratio benchmarks was to use the median, which will be more applicable if there is an asymmetric distribution of efficiencies while still close to the average if the distribution is symmetric, and values for 2019, which should reflect the most up to date view of networks' operations, with consideration of an upper quartile if the modelled result appeared to be robust. For our regression models, networks / groups were first evaluated at the predicted model fit, i.e. the average. If the model appeared to be robust, we considered the result of shifting the benchmark to an upper quartile, i.e. COLS.

A final consideration is if networks are to be given time to 'close the gap' with the set target. For example, for RIIO-GD1, Ofgem decided that GDNs could only be expected to close 75% of the gap with the upper quartile benchmark in recognition that the target will be affected by measurement error. A regulator may also conclude it is not realistic for a network to make up the efficiency gap within the price control period. This decision may come down to regulatory judgement and whether a network has met past efficiency targets. Our understanding is that Ofgem has excluded the possibility of allowing for a 'glide path' for RIIO-2.

6 Model selection

Box 5 Results of model selection process

- Cross-sector BSC models were determined to be unreliable.
- We recommend the use of a MEAV 2019 ratio benchmark model for GD for its simplicity, transparency, potential consistency issues for RIIO-1 data, and intuitive results (which are also supported by consistent, if potentially unreliable, regression results).
- For BSCs for GT and ET, we recommend a CSV regression model that includes a GT sector dummy, but further consideration is needed of the result for SPT.
- We recommend using a MEAV + Total Capex multivariate regression model for CAIs (ET and GT only), although we note that the results of this model require further scrutiny.

We set out our model selection process below, which guided us to our assessment models of choice for BSCs and CAIs. Our approach began with regression-based analyses. An econometric approach allows for relatively sophisticated modelling and controlling for multiple factors within one model. However, given the relatively small number of comparator networks, especially outside of the cross-sector network-level specification, there were concerns about the reliability of econometric results given small sample sizes. We therefore sought to be transparent about any model's statistical limitations and considered ratio benchmarks, which are simple, transparent, replicable, and have regulatory precedent, as a 'fallback' option to help intuitively explain modelled results.

Models were evaluated on their statistical properties, their economic and technical logic, and whether their resulting 'efficiency scores' (calculated as actual / forecast costs divided by modelled costs) appeared to be sensible and intuitive. Numerous sensitivity checks and diagnostic tests were then run on our models of choice and apparent model outliers are discussed in detail. All models were first run with only historical data, but forecast data was included as a sensitivity check.

6.1 Model selection process

Our model selection process explored the following approaches in turn:

- **Cross-sector regression**: This was our starting point as it utilises the largest sample size, which may allow for more sophisticated econometric analysis and a more standardised approach rather than having to shape sector-specific regressions. We tested the viability of this approach by using a suite of modelling tests and criteria.

- **Sector-specific regression**: If cross-sector regressions were found to not be robust, we turned to sector-specific regressions. We again applies a suite of tests and criteria to assess our candidate models.

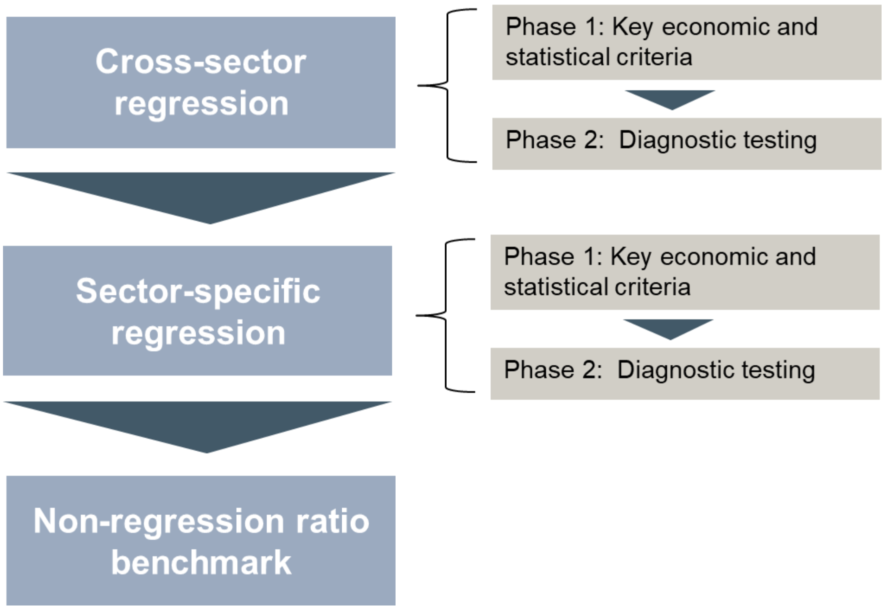
- **Non-regression ratio benchmark**: Where regressions did not appear to be appropriate in explaining the data, we turned to ratio benchmarking, which gives a simple and transparent overview of which companies appear to be efficient.

Following CEPA's recommendation in the GD2 summer consultation, **we used a two-phase approach in assessing each regression model** (Figure 9):

- **Phase 1**: Tests the model against key initial criteria to see if it is viable enough to proceed to Phase 2 and determines the chosen estimator.

- **Phase 2**: Diagnostic tests and sensitivity checks to assess the robustness of the identified models.

Figure 9 Model assessment process



Source: ECA

The assessments we made in each phase are shown in Table 5 overleaf. Following from our discussion in Section 5.2.1, we considered the following estimators:

- **The pooled OLS (POLS) estimator**: Data was 'pooled' (data from each of the networks or groups was grouped together and treated as individual data points) and the coefficients were estimated with OLS.

- **The RE estimator**: This estimator accounts for an error structure which incorporates company effects and thus can be a more efficient estimator, provided sufficient sample size.

- **The FE estimator**: Uses only within-company variation to estimate the coefficient. When company effects are correlated with the explanatory variables the FE estimator remains consistent whereas the RE and POLS estimators will be inconsistent. The Hausman test (discussed below) can be used to test if this is the case. Because the FE estimator only makes use of within-company variation it is less efficient than the RE estimator and therefore RE should be preferred to FE where the Hausman test does not show RE to be inconsistent.

We prioritised the use of the OLS estimator where possible because of its relative simplicity, transparency, and replicability, and of POLS given its good small sample properties. POLS has consistently been used over RE by Ofgem in the past and it will be more familiar to stakeholders. Hence, models were first run with a POLS specification, while RE and FE were considered as a sensitivity check.

During diagnostic testing of our OLS models, we tested for the presence of panel effects using the Breusch-Pagan LM test (discussed below). If this test indicates that panel effects are present, this provides evidence that RE and FE models may be informative, and this should be considered when examining the results of the RE and FE sensitivity tests. Similarly, we used the Hausman test to check for inconsistency in our OLS and RE estimates. However, as discussed in Section 6.3.2, we did not automatically choose the FE estimator based on the results of the Hausman test as it can be imprecise in small samples and a greater concern may be if the resulting FE coefficient violates economic or technical logic. Therefore, we used this test to provide context to the results of the FE sensitivity test.

Table 5 Two-phase approach for assessing regression models

Phase	Category of assessment	Test/criterion
Phase 1	Economic and technical logic	Estimated coefficients are of a sign, size, and significance[1] that agrees with economic logic
		Generates plausible estimates of modelled costs (efficiency scores)
	Statistical robustness	F-test for joint significance of coefficients
		Goodness of fit (Adjusted R-squared[52])
		t-test for significance of individual coefficients
		Avoidance of multicollinearity[2]
	Estimator selection	Hausman test
		Breusch-Pagan test
Phase 2	Diagnostic tests	RESET test for misspecification caused by non-linearities
		White test for heteroscedasticity
		Jarque-Bera test for normality

Source: ECA. [1]We report whether coefficients are significant at the 1%, 5%, and 10% levels (marked by ***, **, and *, respectively). We considered 5% to be the primary significance threshold, while noting if a result is only significant at the 10% level. [2]Note that multicollinearity is not a relevant issue for univariate regressions.

We ranked our regression modelling tests and criteria in order of importance. This ranking, alongside a discussion of each item used for modelling assessment is presented in Table 6. Note that the ranking of importance corresponds to some degree, but not entirely, with the ordering of the two-phase approach.

[52] This paper reports Adjusted R-squared values throughout. Adjusted R-squared attempts to account for the phenomenon of R-squared automatically, and potentially spuriously, increasing whenever any explanatory variable is added to a model. Adjusted R-squared adjusts for the number of explanatory terms in a model relative to the sample size. We deemed this to be an appropriate metric given general concerns about sample sizes throughout this report.

Table 6 Ranking of importance for model assessment criteria

Importance	Criterion	Comments
High	F-test for joint significance of coefficients	A model which fails this test would not have any explanatory power and therefore would be rejected.
	Goodness of fit (adjusted R squared)	The model should be able to predict the dependent variable with a good degree of accuracy.
	Estimated coefficients are of a sign, size, and significance that agrees with technical / economic logic	Cost drivers with negative coefficients or coefficients greater than 1 (implying diseconomies-of-scale) would be rejected. Explanatory variables should generally have statistically significant coefficients (as determined by a t-test) but if the threshold for statistical significance is not reached, explanatory variables may still be retained on the basis of economic or technical logic.
	Generates plausible estimates of modelled costs	While excessive focus on this metric may lead to overfitting[53], modelled costs should correspond to expectations and be considered plausible and reasonable.
Medium	Hausman test for fixed effects	If the Hausman test shows the Random Effects (and by implication OLS) estimates to be inconsistent, Fixed Effects should be considered. However, before the choice is made, consideration should be given to the economic sense and statistical significance of the Fixed Effect coefficients, and whether the sample size is sufficient to use the Fixed Effects estimator.
	Variable Inflation Factor (VIF) test for excessive correlation between explanatory variables	Also known as multicollinearity. Moderate levels of correlation are not a major concern, but very high correlation can cause instable coefficients. Model options where correlation between explanatory variables is 0.9 or above were rejected.
	Sensitivity of coefficients to adding a time trend	Coefficient estimates should be relatively robust against minor changes to specification and dataset to give us confidence in our cost modelling. However, in small datasets, removing part of the sample may be expected to significantly change the coefficient so some degree of instability should be tolerated.
	Sensitivity of coefficients to removing a year	
Low	RESET test for misspecification caused by non-linearities	If non-linearities are identified, we can try alternative functional forms such as the translog specification. However, this must be balanced against the need for a parsimonious and transparent model.
	White test for heteroscedasticity	The presence of heteroscedasticity does not affect coefficient estimates but makes standard errors unreliable
	Jarque-Bera test for normality	A breakdown of the normality assumption does not affect coefficient estimates, but it reduces the reliability of tests of coefficient significance in small sample sizes
	Breusch-Pagan test for individual fixed effects	Tests whether inclusion of random effects improves the model. If so, we would consider using RE over OLS. However, if the sample size is small, OLS may be more efficient than RE.

Source: ECA

An important model 'sense check' is the calculation of the resulting 'efficiency scores' from a model. Efficiency scores are calculated as actual / submitted costs divided by modelled costs. Hence, a score above 1 indicates inefficiency (and setting a network's allowance at the benchmark means setting it lower than its submission), while a score

[53] Overfitting a model is a condition where a statistical model begins to describe the random errors in the data rather than the actual relationship between variables. This is a concern with limited sample sizes, which means there is limited degrees of freedom for additional model controls.

below 1 indicates efficiency. There is no rule-of-thumb as to what indicates a 'reasonable' efficiency score. From a pure theoretical point of view, a very high or low efficiency score merely indicates a network's high or low efficiency. However, conscious of the trade-offs involved in any model, whether due to imperfect data, unobserved heterogeneity, and / or measurement error, we sought to sufficiently explore any 'outlier' efficiency scores to confirm our confidence in the models.

6.2 BSC modelling

We first describe our approach in selecting the models of choice for assessing BSCs. As discussed in Section 6.2.3, we first constructed CSVs as an alternative to MEAV for our BSC regressions.

6.2.1 Constructing composite scale variables for BSCs

Standardising variable weights

There is much regulatory precedent in using Composite Scale Variables (CSVs) to capture the impact of multiple cost drivers, avoiding the complexity and potential multicollinearity of multivariate regressions, and maintaining degrees of freedom when sample sizes are relatively small. However, CSVs risk being criticised for arbitrary input weighting. This issue is obvious if, say, a simplistic 50/50 variable weighting has been applied to the inputs. A more subtle issue is the potential impact of combining cost drivers with different scales and units, which may then give undue weighting to an included cost driver.

This issue can be addressed by standardising variable weights. A variable is standardised by subtracting the average of the variable from each observation and dividing by the standard deviation. The new standardised variable then has an average of 0 and a standard deviation of 1. This allows for evaluating the effect of cost drivers *independent of the size of their average*. **We therefore applied standardised variable weights for our CSVs.**

Our CSVs were composed of the following:

- For the cross-sector CSV: MEAV, FTE (for HR), and Totex[54] (for Procurement)
- For the GD-only CSV: Customer Numbers,[55] FTE, and Total Spend
- For the T-only CSV: MEAV, FTE, and Totex.

[54] Total Spend data was not available from NGGT, so Totex was used as a proxy driver for Procurement in our cross-sector and ET/GT CSVs.

[55] MEAV and Throughput were also considered as proxies for Revenue (for which RIIO-2 forecast data was not provided) for the GD CSV, but we found that Customer Numbers had the highest correlation with RIIO-1 Revenue.

Top-down regression-based weighting

After standardising the input variables, the next step was to run a regression that includes each standardised variable. The coefficients of each variable were then used to determine their weight in the CSV.

One risk when applying this approach is that the coefficient for a cost driver may turn out to be negative, implying a *negative* weight should be applied, which lacks intuitive sense. This issue may arise due to high multicollinearity between cost drivers. For an example, see this issue occurring when using a standardised CSV for econometric benchmarking in the Irish water sector.[56]

This issue arose when applying this approach to each of our CSVs – for the cross-sector and GD-only samples the FTE weight was negative, and for the T-only sample the Totex weight was negative (see Table 7). Hence, likely due to multicollinearity, the top-down regression-based weighting approach did not lead to a practical CSV.

Bottom-up cost weighting

While one solution to the negative weighting issue above would be to exclude the negatively weighted cost driver from the CSV, it was important to distinguish the CSV from MEAV as much as reasonably possible. Otherwise, MEAV's already high weighting in the CSV would be accentuated, which would limit the new evidence gained from running a CSV-based regression over a MEAV-only regression.

We therefore reverted to weighting each (standardised) cost driver in the CSV by the weight of its associated cost category (Table 7). This weighting approach was used for RIIO-GD1 and for the bottom-up totex analysis for RIIO-ED1.

---

[56] NERA, IW IRC2 (2017-18) Assessment – ANNEX Econometric Benchmarking, Prepared for the Commission for Energy Regulator (CER), Section 4.3.1. Accessed online here: https://www.cru.ie/wp-content/uploads/2016/07/CER16270-NERA-IRC2-Econometric-Benchmarking-Report.pdf

Table 7 Calculated CSV weights: top-down approach

Cost driver	Cross-sector	GD	ET/GT
Top-down approach			
FTE	-0.087	-1.151	1.088
MEAV	0.520		0.268
Customer numbers		1.327	
Totex	0.567		-0.356
Total Spend		0.824	
Bottom-up approach			
FTE	0.104	0.106	0.102
MEAV	0.841		0.827
Customer numbers		0.856	
Totex	0.055		0.071
Total Spend		0.038	

Source: ECA analysis. The calculated CSV weights are at the group-level for GD. Weights were calculated at the network-level for the network CSV, and the same issue arose of a negative coefficient for FTE.

6.2.2 Cross-sector BSC models

We first attempted a model with all sectors – GD, ET, and GT - 'pooled' together. This maximises the sample size, but there are obvious concerns about cost comparability across sectors. Our cost driver choice was limited to MEAV given other potential scale cost drivers, such as customer numbers or throughput, cannot be considered comparable cross-sector.

The resulting R-squared in the POLS model was middling at 0.50 and poor at only 0.20 for the RE model, but the coefficient for MEAV was positive and significant in both cases. A secondary concern was that the coefficient on MEAV turned greater than 1 in the FE specification, implying diseconomies-of-scale, suggesting MEAV may not coherently explain within-network BSC changes at a cross-sector level. A CSV specification, which is heavily weighted to MEAV, did not alleviate these concerns.

Mixed econometric results aside, of particular concern was that implied modelled cost results, both for RIIO-1 actuals and the RIIO-2 forecast, were highly inconsistent. Implied efficiency scores (calculated as actual / forecast costs divided by modelled costs) for the transmission companies reached as high as two (2) in a couple of cases, i.e. the model suggested allowances be reduced by *half* relative to both RIIO-1 actuals and RIIO-2 submissions. Conversely, implied efficiency scores for GDNs appeared to be excessively low (i.e. highly efficient) in a few cases. These variances were even more pronounced when we tried applying a simple ratio benchmark. 'Outlier' efficiency scores do not on their own necessarily imply an unreliable model, but given overriding concerns about cross-sector comparability, such results motivate investigating within-sector results.

We tried a sensitivity check where sector dummy variables were included in order to try controlling for inherent sectoral differences by allowing for differing regression constants, which improved the R-squared to 0.62 (although the R-squared for the RE model was still low at 0.23), but the coefficients for the sector dummy variables were only weakly significant (at the 10% level) and the contrast in efficiency assessments between the GDNs and the transmission companies remained. The dispersion of efficiency scores was dampened by the inclusion of sector dummy variables, but major variations persisted between the GDNs and the transmission companiess, particularly for SPT. Hence, the sector-level dummies did not appear to address our concerns about incomparability between the sectors.

As discussed in Section 4.3, we deem group level assessments to be preferable as they can control for differing cost allocation practices within groups. However, for cases like SSE, which holds stakes in both SHET and the SGN GDNs, as well as two DNOs, this necessitates a cross-sectoral approach and the use of DNO data. With multiple sectors in one group, it is also no longer possible to apply sector dummy variables. An attempt at a cross-sector, group-level regression (grouping SHET and SGN under SSE and NGGT and NGET together under National Grid) similarly produced extreme outlier results. Moreover, the proposed grouping can only be considered incomplete given it ignores how Iberdrola owns both SPT and the DNO SP Energy Networks and that CKI Infrastructure Holdings has stakes in both NGN and WWU. Therefore, we concluded that a cross-sector, group-level regression was simply impractical.

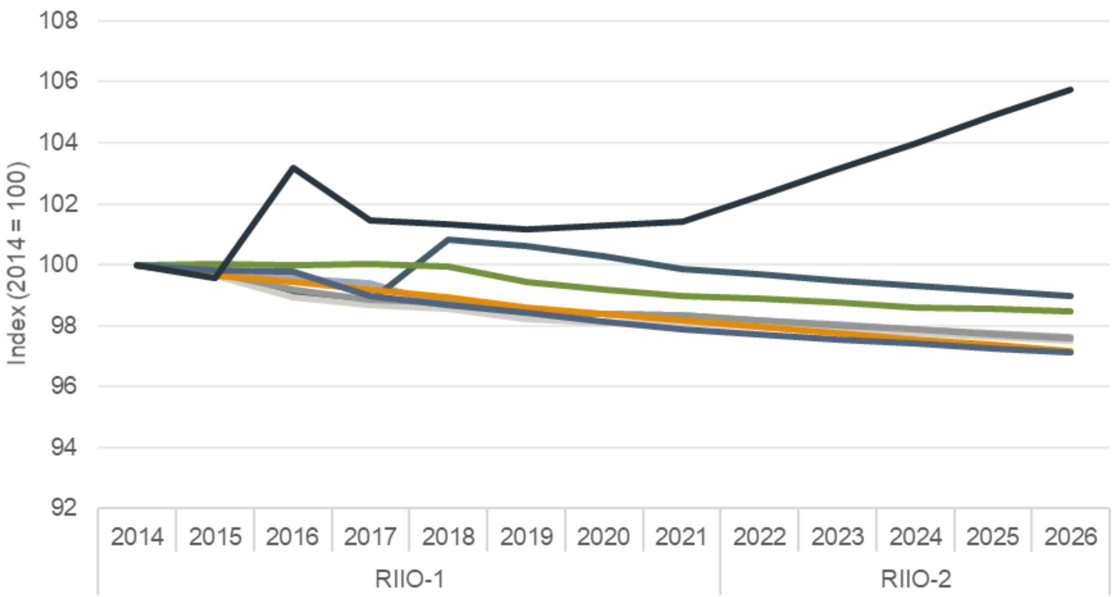
Overall, the combination of initial concerns about comparability, mixed econometric results, and extreme outliers when checking implied modelled costs against RIIO-1 actuals and RIIO-2 forecasts (for network-level, group-level, and sector dummy specifications), **led us to conclude that a cross-sectoral model that pools GD, GT, and ET could not be considered a credible starting point for setting BSC allowances.**

6.2.3 GD-only BSC models

With the cross-sector model appearing to lack credibility, we turned to sector-specific models where concerns about comparability are less pronounced. Assessing at the sector-level helps focus on which companies may be considered efficient / inefficient relative to their true peers.

We first examined a couple of cost driver trends within GD, which provided some initial insight as to what to expect for benchmarking. Looking at submitted MEAVs across the GDNs (Figure 10), we identified a general lack of growth in terms of the historical and projected network asset base. [REDACTED] is the only GDN projecting an increase in MEAV, but the increase is relatively modest: its value for 2026 is only 1.5% higher than the 2019 value. All other GDNs exhibit a steady decline, which is expected to continue through RIIO-2.

Figure 10 GDN MEAV trends, 2014 – 2026



Source: ECA analysis of Ofgem data from December business plans.

Regression models: Phase 1

We first conducted regressions on MEAV and a CSV (Table 8). As discussed in Section 6.2.1, the CSV consisted of customer numbers, total spend, and FTEs.[57] The coefficients on MEAV and the CSV were both positive and statistically significant. However, a major concern was the low R-squared, at only 0.20-0.24. This suggests that the model does not sufficiently explain the networks' costs, which reduces overall confidence in its robustness. The inclusion of a time trend to control for time effects had little impact on the model fit.[58] We also considered a multivariate specification with Customer Numbers and Network Length, but it had similarly poor explanatory power and inconsistent coefficients.

[57] A CSV specification that replaced customer numbers with MEAV was also considered, but it was essentially indistinguishable from MEAV given its high weight in the CSV.
[58] We also considered including year dummy variables and combining both a time trend and year dummy variables, but none of these approaches significantly improved the model fit, and we were cautious about adding variables to the model given the small sample size.

Table 8 GD BSC regression models, network level

	MEAV (Model 1)	MEAV + time trend (Model 2)	CSV (Model 3)	CSV + time trend (Model 4)
MEAV coefficient	0.611***	0.609***		
CSV coefficient			0.192***	0.193***
Time trend		-0.045		-0.047*
Constant	-3.039*	-2.860	2.633***	2.798***
Observations	48	48	48	48
Adjusted R-squared	0.199	0.223	0.236	0.267

Notes: *** p<0.01, ** p<0.05, * p<0.1. All regressions were run with POLS.

Given the poor model fit of the network-level regressions, we turned to group-level regressions (Table 9), where the Cadent GDNs were grouped together and the Scotland and Southern GDNs were grouped under SGN. The group-level regressions exhibit greatly improved R-squared values of 0.79-0.80 for our MEAV and CSV models. However, the MEAV coefficient being greater than 1, implying diseconomies-of-scale, is a concern.[59] This is not the case for the CSV regression, which suggests a more coherent model.

Table 9 GD BSC regression models, group level

	MEAV (Model 5)	CSV (Model 6)
MEAV coefficient	1.062***	
CSV coefficient		0.608***
Constant	-7.326***	3.132***
Observations	24	24
Adjusted R-squared	0.792	0.786

Notes: *** p<0.01, ** p<0.05, * p<0.1. Both regressions were run with POLS.

Ratio benchmark

The stronger model fit for the group-level benchmark was encouraging but outliers were present when converting to efficiency scores, namely NGN and WWU, which were judged to be very inefficient relative to the average model fit, and SGN's RIIO-1 actuals being judged as highly efficient. WWU was a particular outlier for RIIO-2. [REDACTED]

We therefore turned to a simple MEAV ratio benchmark, evaluating groups against the 2019 median BSC-MEAV ratio. The resulting efficiency scores, and how they compare to the group-level CSV regression, are presented in Table 10. The efficiency scores are relatively consistent across the two models, suggesting that the differences between the ratio and regression models is immaterial. With respect to the potential presence of fixed costs, it is notable that the regression approach, which has a positive constant, draws the same conclusion for NGN and deems WWU to be even more inefficient than the ratio

[59] However, note that a t-test *retains* the null hypothesis that the coefficient on MEAV is no greater than 1.

benchmark.[60] Our analysis therefore suggests that, whether through a CSV regression or a MEAV ratio benchmark, WWU's forecast uplift in BSCs does not appear to be justified by its underlying cost drivers, nor by the presence of fixed costs.

We also report the results if an upper quartile were applied for the ratio benchmark. However, we recommend using the median as a ratio benchmark can be a relatively simplistic and uncertain measure of efficiency. Applying an upper quartile may not symmetrically correct mismeasurement errors, so we prefer a cautious approach to setting the benchmark. Furthermore, we are uncertain if it would be appropriate to expect further efficiency gains from the Cadent networks (the group's efficiency score rises from 0.98 to 1.03 when an upper quartile benchmark is applied) after the significant drop in BSCs observed across RIIO-1 after the ownership change.

Table 10 Implied efficiency scores from the GD BSC regression and ratio benchmarking models

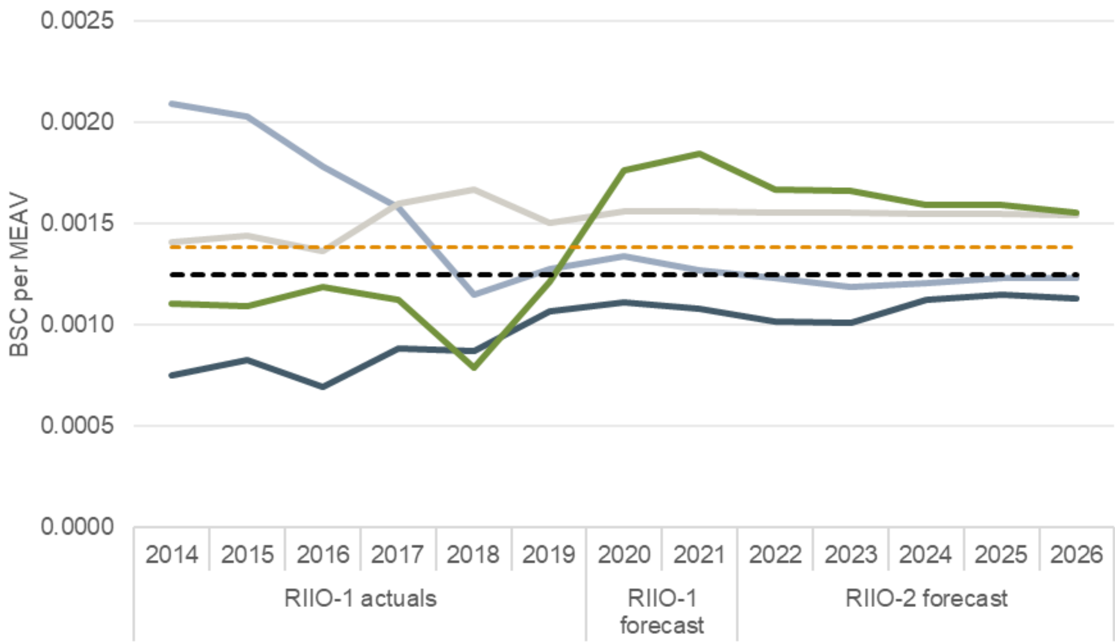
Group / network	CSV regression		MEAV median ratio benchmark		MEAV upper quartile ratio benchmark	
	RIIO-1 actuals	**RIIO-2 forecast**	**RIIO-1 actuals**	**RIIO-2 forecast**	**RIIO-1 actuals**	**RIIO-2 forecast**
Cadent	1.33	0.96	1.33	0.98	1.41	1.03
NGN	1.19	1.19	1.20	1.24	1.27	1.32
SGN	0.66	0.81	0.68	0.87	0.72	0.92
WWU	1.01	1.48	0.87	1.29	0.92	1.37
Benchmark:	Average model fit		2019 median		2019 upper quartile	

Source: ECA analysis.

We can visualise the GD groups' ratio trends across RIIO-1 and RIIO-2 in Figure 11 in comparison to the 2019 median and the forecast average ratio for RIIO-2. [REDACTED]

[60] As a further sensitivity check, a network-level CSV regression gave efficiency scores of 1.14 and 1.42 to NGN and WWU, respectively. The network-level regression results were also consistent with the group level results for the Cadent and SGN networks.

Figure 11 GD BSC-MEAV ratio trends for RIIO-1 and RIIO-2



Source: ECA analysis.

Sensitivity checks

We conducted the following sensitivity checks for the GD-only BSC assessment:

- **Random Effects and Fixed Effects**: When controlling for the dataset's panel structure with a Random Effects specification, the coefficient for the CSV was almost identical and remained significant at the 1% level. The CSV coefficient became insignificant under a Fixed Effects specification, but a Hausman test did not reject the null hypothesis, i.e. the Random Effects model was *not* found to be inconsistent.

- **Forecast data:** Given the apparent disconnect between RIIO-1 actuals and RIIO-2 forecasts for some of the GDNs (see the network level trends in Annex **Error! Reference source not found.**) and the poor statistical fit of the network-level models, this may be a case where RIIO-2 forecasts exhibit a 'step change' from RIIO-1 actuals. As was done for RIIO-ED1 CAIs, where an industry-wide downward step in CAIs was observed in RIIO-2 forecasts, a model with forecast data may better fit the data. However, including RIIO-2 data did not significantly improve the regression model fit (R-squared of 0.30). Alternatively, we considered *only* using RIIO-2 forecast data, but the modelled fit was not very strong (R-squared of 0.47) and the CSV coefficient was widely inconsistent between estimators, turning insignificant under an RE specification and significantly *negative* under an FE specification. Given the BSC trends, as shown in Annex **Error! Reference source not found.**, do not suggest a *sector-wide* shift in GD BSCs, the poor model results, and our general misgivings about using forecast data (as discussed in Section 4.1), we do not think that using forecast data can better explain GD BSC trends.

- **Including IT&T costs**: A sense check where IT&T costs were included found that the resulting benchmark was lower than RIIO-2 forecasts for *every* GD group / network. IT&T costs are an exception to the general downward trend in BSCs across the GD sector, exhibiting an upward trend across both RIIO-1 actuals and RIIO-2 forecasts. IT&T also constitutes a significant share of GD BSCs, averaging 41% of BSCs over RIIO-1. It is not surprising that available cost drivers, which are relatively flat across the GD sector, struggle to benchmark this proposed increase. This suggests the dynamics of IT&T within GD may require a specific assessment by Ofgem (which it is in any case undertaking).

GD-only BSC model of choice

The reasonably strong econometric results may suggest that the CSV regression model could be the default assessment model of choice. However, as with all regressions for this report, the relatively small sample size (24 observations of RIIO-1 actuals at the group-level) may call into doubt the reliability of the regression. In contrast, the ratio benchmark allows for an easy visual of the results (e.g. Figure 11), which helps one intuitively understand each group / network's results, and the ratio benchmark results are consistent with the regression results.

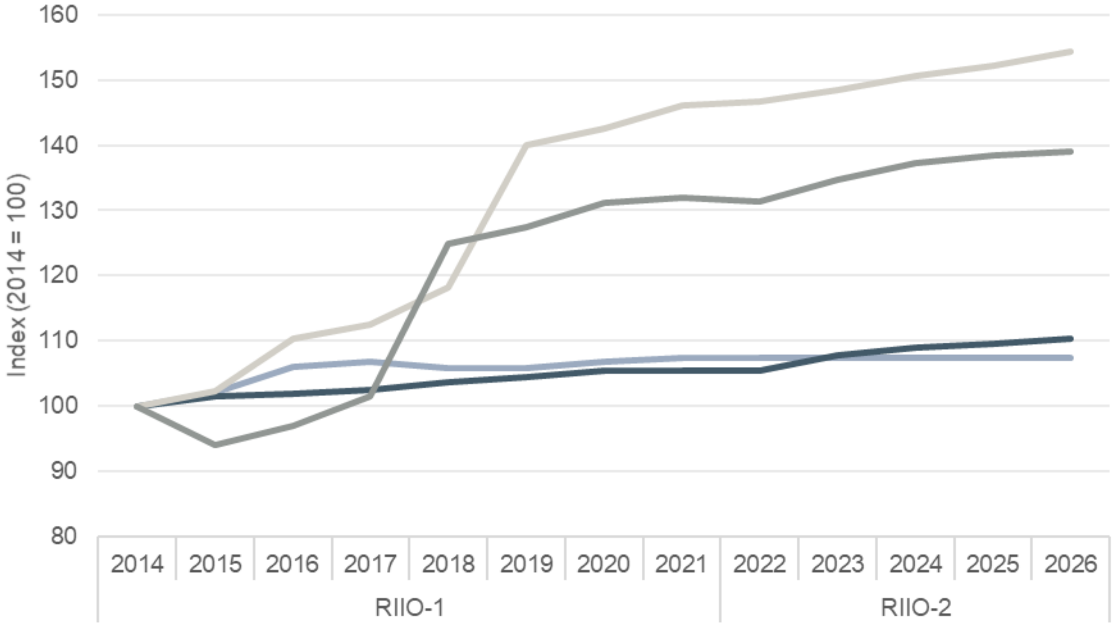
The use of a 2019 median ratio benchmark also avoids a potential inconsistency in the GD data time series. The Cadent networks uniformly exhibit a downward trend in BSCs from 2014 to 2018, i.e. up to the change in Cadent ownership, and then a flat trend from 2019 onwards (see Annex **Error! Reference source not found.**). A regression approach using all historical data may unreasonably expect the Cadent RIIO-1 trend to continue through RIIO-2 (and impose such a benchmark on the rest of the sector), so setting a benchmark using 2019 data alone, like the ratio benchmark, while not exploiting the entire available sample size, may be more representative of the *current* state of the sector.

We therefore propose that the ratio benchmark model be used instead, which has distinct advantages given its simplicity and transparency. The general consistency of results across the regression and ratio benchmark models is also reassuring, providing further evidence of the robustness of the ratio benchmark. We report the results for both 2019 median and upper quartile benchmarks in Table 10, which do not substantially differ, but we prefer using a median as the assessment for Cadent seems more intuitive given their significant decline in BSCs during RIIO-1. Ratio benchmarks also have regulatory precedent, having been regularly used across RIIO-1. Hence, **our recommended GD assessment model is a BSC-MEAV ratio benchmark.**

6.2.4 T-only BSC models

[REDACTED]

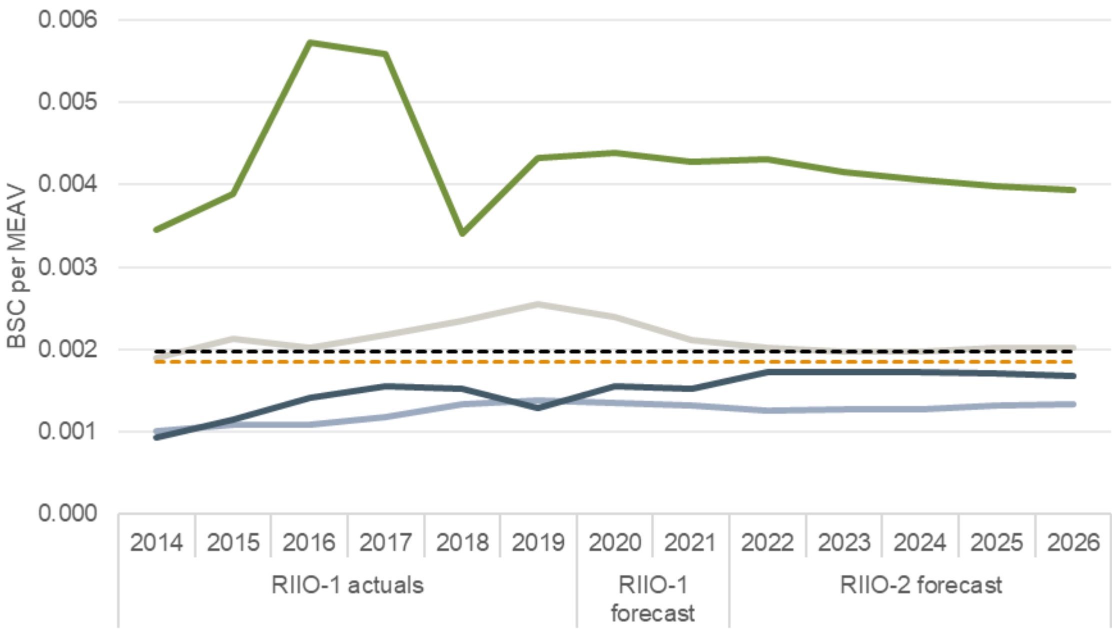
Figure 12 MEAV trends for transmission companies, 2014 - 2026



Source: ECA analysis of Ofgem data from December business plans.

[REDACTED]

Figure 13 BSC-MEAV ratio trends for transmission companies for RIIO-1 and RIIO-2



Source: ECA analysis of Ofgem data from December business plans.

Regression models: Phase 1

As shown in Table 11, for the regression models for the transmission companies, we considered multiple variations of MEAV and CSV (as constructed for BSCs for the transmission companies in Section 6.2.1). Specifically, the regression models considered were:

- Univariate models with MEAV (Model 1) or CSV (Model 3) as the only cost driver
- The addition of a time trend to the above respective models (Models 2 and 4) to control for real expenditure or RPE changes over time relative to cost drivers
- The inclusion of a GT dummy variable (Models 5 to 7), to control for inherent differences between the ET and GT sectors
    - On a related note, we also considered dropping NGGT (TO) from the sample to run an ET-only regression (Model 8).

The coefficients for MEAV and the CSV were consistently significantly positive across our model specifications and the models also exhibit reasonably strong statistical fits (R-squared ranging from 0.68 to 0.77). The inclusion of a time trend (Models 2, 4, and 6) did not have a significant impact on any model.[61]

[61] Year dummy variables were eschewed due to the small sample size in order to maintain degrees of freedom.

Table 11 Transmission BSC regression models

	MEAV (Model 1)	MEAV + time trend (Model 2)	CSV (Model 3)	CSV + time trend (Model 4)
Sectors	GT and ET	GT and ET	GT and ET	GT and ET
MEAV coefficient	0.729***	0.722***		
CSV coefficient			0.754***	0.748***
GT dummy coefficient				
Time trend		0.063		0.061
Constant	-3.726***	-3.883***	3.127***	2.913***
Observations	24	24	24	24
Adjusted R-squared	0.681	0.684	0.761	0.766
	MEAV + GT dummy (Model 5)	**MEAV + GT dummy + time trend (Model 6)**	**CSV + GT dummy (Model 7)**	**MEAV – ET only (Model 8)**
Sectors	GT and ET	GT and ET	GT and ET	ET
MEAV coefficient	0.825***			0.825***
CSV coefficient		0.793***	0.800***	
GT dummy coefficient	-0.504***	-0.308***	-0.314***	
Time trend		0.059		
Constant	-4.495***	3.000***	3.210***	-4.490***
Observations	24	24	24	18
Adjusted R-squared	0.725	0.779	0.774	0.730

Notes: *** p<0.01, ** p<0.05, * p<0.1. All regressions were run with POLS.

However, despite the apparently strong statistical fit, inconsistencies arose when converting these models to efficiency scores, with modelled costs significantly differing from RIIO-1 actuals and / or RIIO-2 forecasts for each network. While this may simply indicate inefficiency, general concerns remained about the comparability of GT and the ETs, as well as within the ETs given the difference in scale between NGET and the Scottish ETs (although a regression approach should help control for differences of scale). We considered a simple MEAV ratio benchmark to better understand the underlying dynamics (as we did for GD BSCs in Section 6.2.3 and T CAIs in Section 6.3), but the resulting benchmark gave implausible results, e.g. efficiency scores well above 2 for SPT and below 0.7 for both NGGT and SHET.

Our proposed solution to this apparent inconsistency was a within-model adjustment: including a GT dummy variable. Given general concerns about comparability between sectors, we considered it a prudent modelling adjustment to explicitly acknowledge this in the regression equation. By applying a GT dummy variable, we allowed the constant in the regression equation to differ between GT and ET. We also considered an ET-only model that dropped NGGT (TO) from the sample (Model 8), but the econometric results were unchanged and the modelled cost results were immaterially different for the ETs; we also have reservations about reducing the sample size, and including a GT dummy variable already effectively separates ET and GT.

As seen in Models 5-7 of Table 11, the GT dummy variable is statistically significant, which supports its inclusion and it also contributes to materially improved Adjusted R-squared values. We compare the resulting efficiency scores between the original CSV-only regression and the CSV regression that includes a GT dummy variable in Table 12. With the exception of SHET, the inclusion of the GT dummy significantly lowers the dispersion of the efficiency scores.

Table 12 Implied efficiency scores from the T regression models

Network	CSV regression (Model 3)		CSV + GT dummy regression (Model 7)	
	RIIO-1 actuals	RIIO-2 forecast	RIIO-1 actuals	RIIO-2 forecast
NGGT (TO)	0.82	0.86	1.01	1.05
NGET	1.32	1.22	1.14	1.05
SHET	0.61	0.82	0.59	0.78
SPT	1.60	1.60	1.56	1.53
Benchmark:	Average model fit		Average model fit	

Source: ECA analysis.

SPT remains an inefficient outlier in Model 7. We considered both the inclusion of IT&T costs and the exclusion of insurance costs as sensitivity checks as potential explanations for this below. One interpretation could be that SPT, as the smallest network, is judged as inefficient due to the presence of fixed costs, but we note that our use of a regression model, with a significantly positive intercept term, should partly address the fixed costs issue. In contrast, SPT's efficiency score is greater than 2 for both RIIO-1 and RIIO-2 when using a simple ratio benchmark.

Regression models: Phase 2

With the CSV + GT dummy variable regression (Model 7) being our regression model of choice, and having rejected a ratio benchmark alternative, we turned to **Phase 2** of our regression selection process, subjecting the model to a series of diagnostic tests (as described in Table 6). The results are summarised in Table 13.

Table 13 BSC Transmission model diagnostic tests

	Model 7: POLS CSV cost driver with GT dummy	Importance	Test Pass/Fail
Phase 1 results			
CSV coefficient	0.800***	High	Pass
GT dummy coefficient	-0.314***	High	Pass
Constant	3.210***	N/A[1]	N/A
Adjusted R-squared	0.774	High	Pass
F-test	40.433***	High	Pass
Hausman test	FE	Medium	Fail[2]
Phase 2 results			
Coefficient(s) robust to robust standard errors?	Yes	Low	Pass
VIFs are below 10	Yes	Medium	Pass
Breusch-Pagan test	Null of no panel effects is rejected	Low	Fail
RESET test	Null of no omitted non-linearities is rejected	Low	Fail
Jarque-Bera test	Null of normal residuals not rejected	Low	Pass

Notes: *** p<0.01, ** p<0.05, * p<0.1. [1]The constant being statistically significant is not directly relevant to the confidence in the model, but it does indirectly suggest that the model accounts for fixed costs to some extent. [2]Test failure is relative to using a RE model.

We address each of the Phase 2 tests in turn:[62]

- **Multicollinearity**: excessive correlation between dependent variables reduces the precision of the coefficient estimates. A measure of this distortion is the VIF (variance inflation factor). If the VIF is below 10 we do not consider multicollinearity to be a concern. The CSV + GT dummy model passes this test.

- **Hausman test**: this tests for endogeneity caused by the correlation of explanatory variables with the individual company effects. The test does this by comparing the RE estimates with the FE estimates. A rejection of the null hypothesis implies RE (and therefore OLS) estimates are biased and inconsistent. However, as discussed in section 6.2.2, there are doubts about the FE estimator's precision in small samples and its appropriateness in a regulatory setting. For this reason, despite the Hausman test rejecting the null hypothesis, we opted to retain the OLS estimator. However, the result of this test suggests greater weight be placed on the use of the FE estimator as a sensitivity test, which is considered in our sensitivity checks below.

- **Breusch-Pagan test**: used to test for presence of panel effects. The null hypothesis was rejected, implying the presence of panel effects. Although this result is sometimes used to support use of the RE estimator, which accounts

[62] Note that we did not test for heteroscedasticity because we used robust standard errors by default, which correct for heteroscedasticity.

for panel effects, this should be balanced by the OLS estimator's superior small sample properties, which is applicable in this case. We again considered this insufficient reason to reject use of the OLS estimator. However, the result of this test suggests greater weight be placed on the use of the RE estimator as a sensitivity test, which is considered below.

- **RESET test**: used to help identify misspecification by testing for omitted non-linearities. The null hypothesis was rejected, indicating omitted non-linearities may be present. Rejection of the null hypothesis is sometimes taken as evidence that additional terms or alternative specification should be used to address non-linearities. However, this must be balanced against the need for a simple and transparent model, with regressors backed by economic and technical logic, particularly in a small sample setting. Likewise, as discussed in Section 5.2.2, the RESET test does not usually provide sufficient evidence for rejecting a model outright.

- **Jarque-Bera test**: used to test normality of the model's residuals. The assumption of normal residuals is relevant as a test of statistical significance in small samples. The null hypothesis is not rejected, suggesting residuals are normal.

Having put the model through Phase 2 tests, while the model may not have ideal properties with respect to all the panel-based diagnostic tests, given our RE and FE sensitivity tests below, we were not compelled to reject the model over other proposed alternatives. The model satisfied our key requirements as outlined in Section 6.

Sensitivity checks

The sensitivity checks we conducted for the transmission BSCs were:

- **Random and Fixed Effects**: In recognition of the panel structure of the dataset, which was confirmed by the Breusch-Pagan test for panel effects and the Hausman test rejecting the null hypothesis for our model of choice, we ran both Random and Fixed Effects specifications on the CSV + GT dummy regression model (the FE model making the inclusion of the GT dummy variable redundant). The coefficient on the CSV went above 1 under these specifications, which counterintuitively suggests diseconomies-of-scale, and the R-squared dropped significantly. Perhaps due to the small sample size, we therefore did not consider the RE or FE specifications to be reliable. We therefore considered a POLS specification to be preferable due to its reliable small sample properties and its relative simplicity and transparency.

- **Forecast data**: As set out in Section 4.1, we preferred to use only historical data for our baseline regressions. However, we still considered the use of forecast data as a model sense-check. The CSV + GT dummy model results were immaterially different when using both historical and forecast data in the sample. [REDACTED]. Hence, the modelled results would be robust to Ofgem deciding to include forecast data across the price control.

- **ET-only regression**: Given general concerns about the comparability of GT and ET, as a sense-check on the GT dummy variable model, we checked the results if only the ET networks were included in a MEAV-only regression (Model 8 in Table 11). However, the same dispersion of results between NGET, SHET, and SPT were present in the ET-only model. The coefficient on MEAV was essentially unchanged and the statistical fit was similar to the GT + ET model. Furthermore, small sample size concerns are even greater if NGGT is excluded from the sample.

- **Including IT&T / excluding insurance costs**: Sense checks were conducted that considered including IT&T and / or excluding insurance costs. [REDACTED]

T-only BSC model of choice

The high dispersion of implied efficiency scores gave pause about the regression approach, despite the models having apparently strong statistical fits. An attempt to simplify the model to a ratio benchmark only exacerbated the apparent issue.

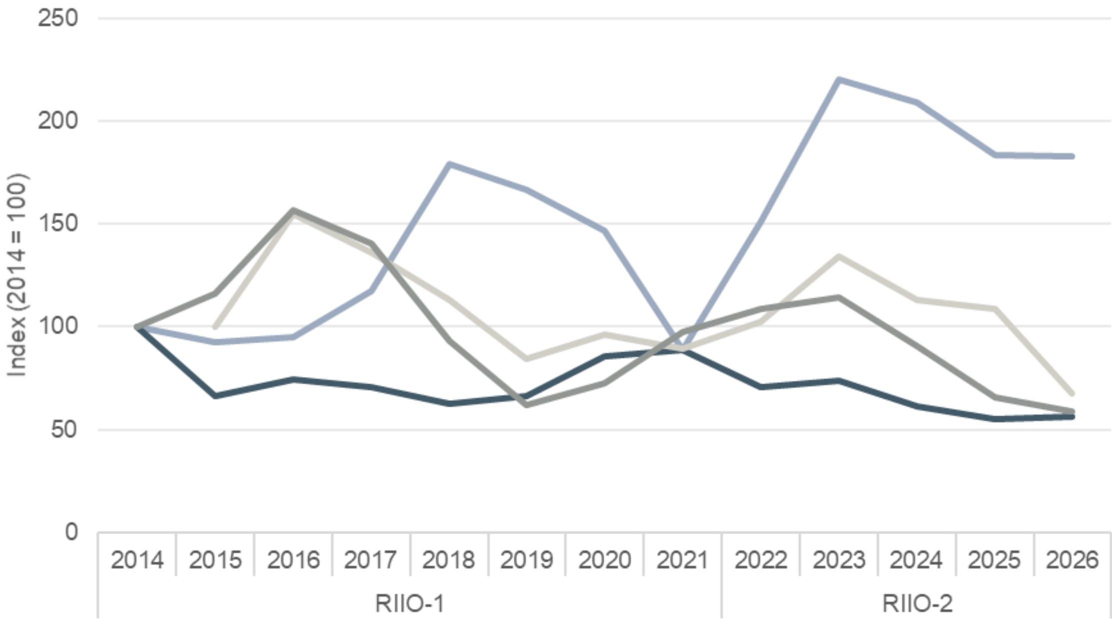
We therefore considered a number of within-model adjustments for the regression. Given general concerns about the comparability of the ET and GT sectors for benchmarking, we chose to explicitly model for this, including a GT dummy variable, which proved to be statistically significant. This modelling adjustment materially improved the statistical fit of the model and lowered the dispersion of the model's efficiency scores, giving us more confidence in the model's results. Hence, **a pooled CSV regression model that includes a dummy variable for the GT sector is our model of choice for BSC Ts.**

The modelled cost result for **SPT** requires further scrutiny. However, our sensitivity checks, including other estimators, the use of forecast data, and the inclusion of IT&T / the exclusion of insurance costs, are consistent in finding SPT's RIIO-2 submission to be inefficient relative to the model benchmarks.

6.3 CAI modelling

We now consider CAIs, which can generally be expected to be driven by workloads. Looking at Total Capex (Figure 14), as an indicator of workload, its relative 'lumpiness' is apparent given the large-scale programs of the Ts. Hence, given the year-to-year variability of CAIs, Total Capex may need to be combined with a 'steadier' scale variable, such as MEAV, in a multivariate regression to help explain both the scale and variance of CAIs.

Figure 14 Total Capex trends for transmission companies, 2014 - 2026



Source: ECA analysis of Ofgem data from December business plans. The index has been set to 2015 for SHET for scale as Total Capex in 2014 was a significant downward outlier for SHET. Note that 2014 SHET data was still included in all benchmarking analysis.

It is notable that in NGGT's December 2019 RIIO-2 submission, NGGT (TO)'s Total Capex is forecast to significantly increase in RIIO-2, with an average annual of £191m across RIIO-1 rising to £294m for RIIO-2. This is to some extent in line with its CAI trends, which are projected to rise from an annual average of £40m in RIIO-1 to £54m in RIIO-2 (Section 3.2).

Regression models: Phase 1

We first considered univariate regressions on MEAV (Model 1) and Total Capex (Model 2) of CAIs for ET and GT (Table 14). In terms of model fit, the Total Capex univariate regressions perform much better than MEAV (R-squared of 0.73 vs 0.34). **We therefore considered Total Capex to be our primary cost driver**. The inclusion of a time trend was considered (Model 3), but it was found to be statistically insignificant.[63]

Given the apparent 'lumpiness' of Total Capex noted above, and the MEAV coefficient still being significant on its own, we considered including MEAV as part of a multivariate regression (Model 4). **The Total Capex + MEAV regression (Model 4) had robust cost driver coefficients and the *adjusted* R-squared improved** from 0.73 to 0.79.[64] Despite the

[63] A time trend could not be reliably included in the multivariate regression due to degrees of freedom constraints, but we consider the result in Model 3 sufficient to conclude that time trends are not a significant factor for CAIs.
[64] Note that an *adjusted* R-squared only increases if the inclusion of the new term improves the model fit more than would be suggested by chance. Otherwise, R-squared values automatically - and possibly spuriously - increase when extra explanatory variables are added to a model. We considered this to be an important adjustment given that the small sample size limits the degrees of freedom, hence additional variables need to be well-justified.

small sample size, this motivated including both cost drivers in a multivariate regression on statistical grounds, in addition to the intuitive reasoning that Total Capex and MEAV should together reflect both the workload and scale effects that drive CAIs.

Table 14 CAI regression models

	MEAV (Model 1)	Total Capex (Model 2)	Total Capex + time trend (Model 3)	Total Capex + MEAV (Model 4)
Sector	ET and GT	ET and GT	ET and GT	ET and GT
Estimator	POLS	POLS	POLS	POLS
MEAV coefficient	0.424**			0.198**
Total Capex coefficient		0.860***	0.862***	0.735***
Time trend			-0.014	
Constant	0.100	-0.978**	-0.939*	-2.093***
Observations	24	24	24	24
Adjusted R-squared	0.340	0.728	0.716	0.786

*** p<0.01, ** p<0.05, * p<0.1

However, when using Model 4 to produce modelled costs, there was a notable spread in the resulting efficiency scores (see Table 15 below), particularly for NGET and SHET, and SPT for RIIO-1. NGET was found to be highly inefficient (especially for RIIO-2), while the efficiency score for SHET was particularly low, i.e. efficient, for RIIO-1. SPT's RIIO-2 submission showed improvement, but it was assessed as notably inefficient in RIIO-1. **Despite Model 4's apparently strong statistical fit, these implied efficiency scores need to be explored in detail**. We therefore considered the results from different estimators, as well as the use of a simple ratio benchmark.

Table 15 Efficiency scores based on Model 4 (Total Capex + MEAV)

Network	RIIO-1	RIIO-2
NGGT (TO)	0.93	0.91
NGET	1.20	1.42
SHET	0.65	0.87
SPT	1.33	1.13

Source: ECA analysis

We first considered a Fixed Effects specification, which allows for a different constant in the regression equation for each network. FE models can better capture any unobserved heterogeneity in companies' CAIs, although, as discussed in Section 5.2.1, there are concerns about their appropriateness in a small-sample regulatory setting. As shown in Table 16, the coefficient on MEAV turned insignificant (and, unintuitively, negative) when running an FE specification on the Total Capex + MEAV model (Model 6), so we instead report the efficiency scores from an FE specification on the *univariate* Total Capex model (Model 7)[65] in our comparative table below (Table 17).

[65] A Hausman test rejected the null that the RE model is unbiased and consistent for the univariate Total Capex model, which suggests that an FE model should be used to account for panel effects.

We next considered a Random Effects specification (Model 5), which a Hausman test found to be unbiased and consistent compared to an FE model for the Total Capex + MEAV multivariate regression (see the Phase 2 diagnostic tests in Table 18 below).

The regression results for Models 5, 6, and 7 are summarised in the table below, with Model 4's results again reported for comparison.

Table 16 CAI regression sensitivity checks

	Total Capex + MEAV POLS (Model 4)	Total Capex + MEAV RE (Model 5)	Total Capex + MEAV FE (Model 6)	Total Capex FE (Model 7)
Sector	ET and GT	ET and GT	ET and GT	ET and GT
Estimator	POLS	RE	FE	FE
MEAV coefficient	0.198**	0.044	-0.247	
Total Capex coefficient	0.735***	0.415***	0.381***	0.395***
Time trend				
Constant	-2.093***	1.212		
Observations	24	24	24	24
Adjusted R-squared	0.786	0.531	0.595	0.488

*** p<0.01, ** p<0.05, * p<0.1

Finally, we compared the efficiency scores for all transmission companies to a simple Capex ratio benchmark (which we label as 'Model 8').[66] Table 17 compares the efficiency scores across the original univariate Total Capex regression (both the POLS (Model 2) and FE (Model 7) specifications), the POLS (Model 4) and RE (Model 5) specifications of the multivariate Total Capex + MEAV regression, and the Total Capex ratio benchmark (Model 8).

Table 17 Efficiency scores by CAI model

Network	Total Capex POLS (Model 2)		Total Capex FE (Model 7)		Total Capex + MEAV POLS (Model 4)		Total Capex + MEAV RE (Model 5)		Total Capex ratio benchmark (Model 8)	
	RIIO-1	RIIO-2	RIIO-1	RIIO-2	RIIO-1	RIIO-2	RIIO-1	RIIO-2	RIIO-1	RIIO-2
NGGT (TO)	1.15	1.08	1.01	1.15	0.93	0.91	0.87	0.98	1.02	0.90
NGET	1.32	1.61	1.00	1.14	1.20	1.42	1.97	2.25	0.93	1.16
SHET	0.57	0.82	0.99	1.41	0.65	0.87	0.61	0.85	0.46	0.66
SPT	1.07	0.99	1.01	0.83	1.33	1.13	0.98	0.80	0.91	0.87

Source: ECA analysis. Regression models were evaluated against the average model fit. The Total Capex ratio benchmark was set at the 2019 median ratio. The numbering of the models follows from Table 14 and Table 16.

We make the following observations:

In contrast, in our Phase 2 diagnostic tests on the MEAV + Total Capex model, the Hausman test retains the null that the RE model is unbiased and consistent.

[66] We also considered a MEAV ratio benchmark, but it produced implausible modelled cost results, e.g. an efficiency score of over 3 for SPT in RIIO-1.

- The efficiency scores under Model 7 (Total Capex FE) is narrower, suggesting that an FE model better captures unobserved cost dynamics, but the model also suggests an entirely opposite conclusion for SHET compared to the Total Capex + MEAV POLS model (Model 4): that its RIIO-2 submission is highly *inefficient* (efficiency score of 1.41 compared to 0.87).

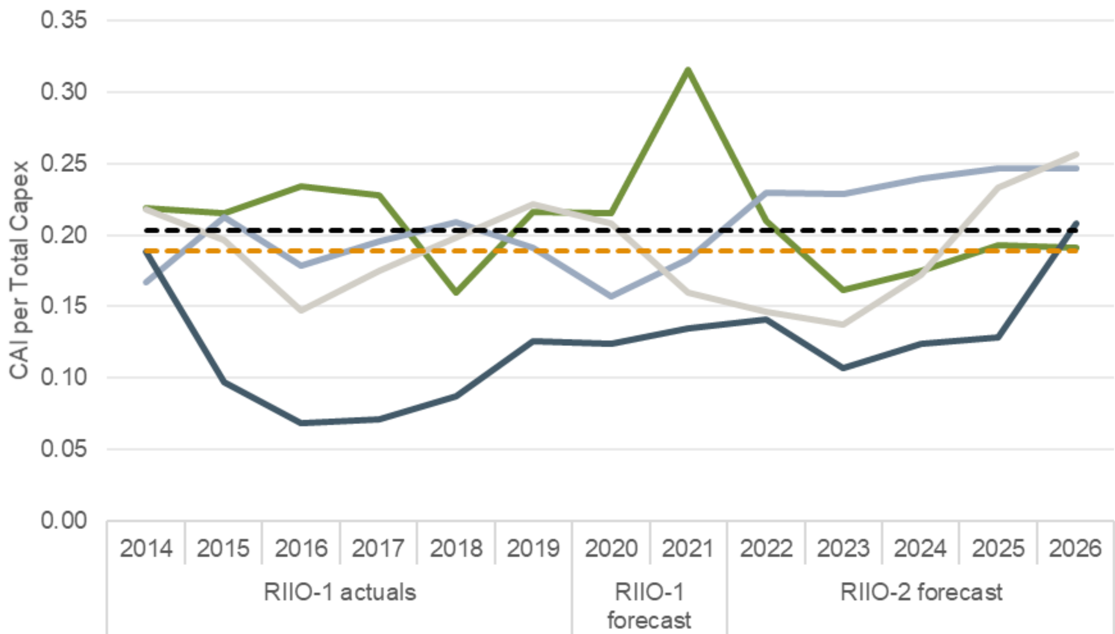
- Model 5 (Total Capex + MEAV RE) has the same broad conclusions about networks' efficiency with the Total Capex + MEAV POLS model, although the efficiency scores for NGET (1.97 in RIIO-1 and 2.25 in RIIO-2) cannot be considered plausible. We therefore prefer the POLS model, which has more reliable small sample properties.

- As when running a POLS regression on Total Capex (Model 2) or Total Capex + MEAV (Model 4), the Total Capex ratio benchmark concluded that SHET has been at the efficiency frontier for both RIIO-1 and RIIO-2. Also, NGET's apparent inefficiency for RIIO-2 remains, but it is less pronounced.

As a visual aid, we chart the Total Capex ratio benchmark [REDACTED]

[REDACTED]

Figure 15 T CAI-Total Capex ratio trends for RIIO-1 and RIIO-2



Source: ECA analysis.

The FE specification does raise questions as to why SHET's RIIO-2 submission appears to be inefficient relative to its RIIO-1 actuals, but considering our misgivings about using FE in a small sample regulatory setting, we do not consider it to be an appropriate model on its own.

The ratio benchmark has similar conclusions to the POLS regression models, although the efficiency score results for SHET, particularly the score of 0.46 for RIIO-1, do not appear to be plausible. A ratio benchmark is simple and transparent, which can be an attractive property, but in this case, its simplicity likely prevents it from robustly capturing the dynamics of CAIs.

We therefore concluded that a **Total Capex + MEAV POLS regression is the best available option**, so we proceeded to our Phase 2 diagnostic tests for this model. **However, Ofgem will need to consider the efficiency scores in Table 17 in more detail, as the results do raise questions about the robustness of the model for setting final costs**. We discuss the results on a network-by-network basis in our conclusion to this section.

Regression models: Phase 2

With the MEAV + Total Capex POLS regression (Model 5) being our model of choice, we turned to Phase 2 of our regression selection process. The results of the diagnostic tests are presented in Table 18.

Table 18 CAI model diagnostic tests

	Model 4: Total Capex + MEAV POLS	Importance	Test Pass/Fail
Phase 1 results			
MEAV coefficient	0.198**	High	Pass
Total Capex coefficient	0.735***	High	Pass
Constant	-2.09***	N/A[1]	N/A
Adjusted R-squared	0.786	High	Pass
F-test	89.18***	High	Pass
Hausman test	RE	Medium	Pass
Phase 2 results			
Coefficient(s) robust to robust standard errors?	Yes	Low	Pass
VIFs are below 10	Yes	Medium	Pass
Breusch-Pagan test	Null of no panel effects is rejected	Low	Fail
RESET test	Null of no omitted non-linearities is rejected	Low	Fail
Jarque-Bera test	Null of normal residuals not rejected	Low	Pass

*** p<0.01, ** p<0.05, * p<0.1. [1]The constant being statistically significant is not directly relevant to the confidence in the model, but the negative coefficient does raise concerns about the model's indirect treatment of fixed costs.

Our preferred model passed all but two of the diagnostic tests. For brief summaries of the diagnostic tests that passed, see the discussion of the transmission BSC regression model's Phase 2 results in Section 6.2.4. In relation to the two failed tests, we note the following:

- **Panel effects**: the failure of the **Breusch-Pagan test** suggests that panel effects should be taken into account in the model, while the **Hausman test** concludes that a Random Effects specification is not biased or inconsistent compared to a Fixed Effects specification. However, as mentioned in Phase 1 above, the Random Effects specification results in implausible efficiency scores for NGET, and we consider a POLS model to be more reliable in a small sample setting.

- **RESET test**: used to help identify misspecification by testing for omitted non-linearities. The null hypothesis was rejected, indicating omitted non-linearities may be present. Rejection of the null hypothesis is sometimes taken as evidence that additional terms or alternative specification should be used to address non-linearities. However, this must be balanced against the need for a simple and transparent model, with regressors backed by economic and technical logic, particularly in a small-sample setting. Given we are already employing a multivariate regression, the scope for adding more terms or non-linearities is limited. Likewise, as discussed in Section 5.2.2, the RESET test does not usually provide sufficient evidence for rejecting a model outright.

The Phase 2 diagnostic tests suggest the model may not sufficiently account for panel effects, which does raise **some caution about relying exclusively on Model 4 to *set* allowances**, but given our doubts about the applicability of FE models and the implausible NGET efficiency scores resulting from the RE model, we are not compelled to reject the POLS model.

Sensitivity checks

We next considered a series of sensitivity checks that may help explain the results[67]:

- **Forecast data**: As set out in Section 4.1, we have preferred to use only historical data for our baseline regressions. However, we still considered the use of forecast data as a model sense-check and to consider whether a 'structural break' may have occurred between RIIO-1 actuals and RIIO-2 submissions. In Table 19, we can see that the econometric outputs are broadly consistent, while the statistical fit notably improves. The resulting efficiency scores in Table 20 draw generally consistent conclusions about the networks. NGET does see an efficiency score improvement when forecast data is included, improving from 1.42 to 1.20 for RIIO-2 when only forecast data is used, but the takeaway remains that NGET does appear to be inefficient regardless of the chosen sample period. We therefore conclude that the modelled results would be robust to the inclusion of forecast data.

Table 19 Econometric outputs for the CAI Total Capex + MEAV model by sample period

	Historical data only	Historical + forecast data	Forecast data only
MEAV coefficient	0.198**	0.229***	0.264***
Total Capex coefficient	0.735***	0.733***	0.716***

[67] Note we have already discussed Random and Fixed Effects sensitivity checks in the regression model selection discussion above.

Constant	-2.093***	-2.334***	-2.542***
Observations	24	52	28
Adjusted R-squared	0.786	0.849	0.898

*** p<0.01, ** p<0.05, * p<0.1.

Table 20 Total Capex + MEAV CAI regression model efficiency scores by sample period

Network	Historical data only		Historical + Forecast data		Forecast data only	
	RIIO-1 actuals	RIIO-2 forecast	RIIO-1 actuals	RIIO-2 forecast	RIIO-1 actuals	RIIO-2 forecast
NGGT (TO)	0.93	0.91	0.88	0.87	0.83	0.82
NGET	1.20	1.42	1.12	1.33	1.07	1.27
SHET	0.65	0.87	0.64	0.85	0.64	0.85
SPT	1.33	1.13	1.34	1.13	1.37	1.14

Source: ECA analysis. All networks are evaluated against the average model fit.

- **Inclusion of an NGET dummy variable**: In seeking to understand why the implied efficiency scores for NGET appear to be an outlier in the regressions, we considered a specification where a dummy variable was only included for NGET (similar to our use of a GT dummy variable for BSC Ts); the argument being that NGET has unobserved characteristics that cannot be explained by a pooled model. However, such a specification was considered inappropriate because:
  - The inclusion of the NGET dummy variable caused the coefficients for both MEAV and Total Capex to become insignificant, which seems unfounded, especially for Total Capex given that its significance in every other specification suggests it should be a robust (and intuitive) cost driver for CAIs.
  - A dummy variable is a crude instrument, particularly in small samples, and its usage in this case is not as justifiable as when we applied a GT dummy variable to our BSC T model, where inherent unobserved differences may exist between the GT and ET sectors. A similar argument could be made with respect to the *scale* of NGET's operations compared to the other ETs, which may entail more inherent costs, but the inclusion of MEAV in the multivariate regression should already control for this.
  - We also note that the FE specification, which effectively applies a dummy variable for *all* networks, also concludes that NGET's RIIO-2 submission is inefficient.
- **Inclusion of IT&T costs**: We ran the same models with IT&T included in the aggregate CAIs. The implied efficiency scores were not materially different from the model without IT&T. This would suggest that excluding IT&T CAIs may not be necessary for any final top-down assessment.

CAI model of choice

Despite our regression model of choice (Total Capex + MEAV POLS regression) having both robust cost drivers and a strong statistical fit, we had concerns about the dispersion of the implied efficiency scores, particularly for NGET and SHET.

An FE specification, which captures the effect of unobserved company heterogeneity, reduced the dispersion of the efficiency scores, but we were concerned that the FE model had an entirely opposite efficiency conclusion for SHET, apparently penalising SHET for its past outperformance. We also had misgivings about applying FE models in a small sample regulatory context (as discussed in Section 5.2.1).

We also considered a simple ratio benchmark, which helped intuitively explain the contradictory results for SHET, as well as giving some context to NGET's inefficient score. However, the ratio benchmark yielded implausible results for SHET: it models SHET's RIIO-1 actuals as being 55% lower than they 'should' have been for RIIO-1 and their RIIO-2 submission to be 50% too low. It should also be noted that given the Total Capex ratio is only based on 2019 actuals, it utilises less of the sample and the impact of any changes to the input data will be more acute than a regression model, which will have multiple fluctuating parameters (we noted how this was less of a deficiency for the GD BSC ratio model due to concerns about the comparability of RIIO-1 historical data). A Total Capex ratio on its own does not appear able to sufficiently explain CAI dynamics.

Given these models have opposite conclusions about SHET (and, to a lesser extent, for NGGT (TO)) and they are in broad 'agreement' about NGET's inefficiency and SPT's efficiency, we considered whether it may be appropriate to 'mix' the FE and Total Capex ratio models. Mixing models has been used in previous price controls in recognition of the imperfection of any one model. For example, Ofgem's RIIO-ED1 assessment of IT&T BSCs was based on a 50 / 50 split of expert review and benchmarking. In its advice to Northern Ireland's Utility Regulator (UR) for its RP6 price control, CEPA advised UR to consider 'triangulating' a mix of models, having identified trade-offs to multiple approaches.[68]

We do not think such an approach is justified in this case. CEPA's advice for UR cited above is in reference to a case where they deemed different models, using alternative cost drivers, to *all* perform well. Different cost drivers may capture different elements of modelled costs or there may be clear trade-offs in how one model structurally measures efficiency versus another.

In this case, a Total Capex ratio may capture that SHET's CAIs appear to be efficient relative to the sector median ratio, while an FE model, in capturing unobserved heterogeneity with a company-specific intercept, may capture that SHET's efficiency in RIIO-2 relative to RIIO-1 may have declined. However, these details are specific to the resulting assessments of SHET rather than explicitly addressing the differences in the approaches (and potential defects) of each model.

We therefore considered the MEAV + Total Capex POLS model to be the best available option. The model has generally attractive statistical properties, both cost drivers are

[68] CEPA, 'RP6 Efficiency Advice', The Northern Ireland Utility Regulator (UR), Final Report, March 2017, p. 39.

robust and have economic / technical logic, and sensitivity checks generally arrive at the same conclusions about networks' efficiency.

The resulting efficiency scores (as well as networks' CAI and cost driver submissions) do require further scrutiny by Ofgem, outside of this modelling process, to understand whether an efficiency challenge is appropriate. We summarise and discuss the implications of the model's efficiency assessments for each network in turn.

NGET

The MEAV + Total Capex regression model converts to average annual efficiency scores of 1.20 in RIIO-1 and 1.42 in RIIO-2. Hence, NGET is found to be inefficient in RIIO-1 and increasingly inefficient for RIIO-2. The score of 1.42 for RIIO-2 is a significant challenge to NGET's efficiency, but this finding is consistent whether applying an RE or FE specification, a univariate Total Capex regression or ratio model, or including forecast data. The inclusion of MEAV in the regression should also help the model control for NGET's underlying scale compared to the other Ts (and its inclusion does lower NGET's RIIO-2 efficiency score from 1.61 to 1.42).

We therefore concluded that the MEAV + Total Capex model can be used as a *basis* for challenging NGET given that this model, and other sensitivity checks, are unanimous in judging NGET's submission as inefficient.

Further technical review may be needed of NGET's MEAV and / or Total Capex numbers, in conjunction with its CAI submission. Our understanding is that Ofgem is reviewing these inputs. Given the small sample size, the regression result could change significantly if the inputs change. NGET's CAI actuals / submissions could be further scrutinised on an item-by-item basis, but as noted in Section 5.1.4, some inconsistency in CAI reporting between RIIO-1 and RIIO-2 precluded a disaggregated assessment.

SHET

The models consistently assessed SHET's CAIs as efficient (except for the FE specification), although there is concern that the high efficiency score results for RIIO-1, ranging from 0.57 to 0.64, are unrealistic. This is less of a concern for the RIIO-2 assessments, where the modelled results range from 0.82 to 0.87.

However, the key question for Ofgem to investigate is why SHET's RIIO-2 submission appears to be less efficient than its RIIO-1 actuals, as its per unit CAI costs are on an upward trend. This is the conclusion of every model specification and is stressed by an FE model, even if SHET's CAIs do still appear to be efficient relative to the other networks.

SPT

There is a contradiction between the MEAV + Total Capex POLS model, which finds SPT to be inefficient (efficiency scores of 1.33 and 1.13 in RIIO-1 and RIIO-2, respectively), and the univariate Total Capex regression (except for RIIO-1) and Total Capex ratio models that deem SPT to be efficient. However, given MEAV appears to be a robust and reasonable cost driver to include in the model, and the Total Capex ratio is a crude / simplistic model, this contradiction is not necessarily an issue.

The negative effect of including MEAV for SPT's assessment may not be surprising given the findings of the BSC models, where using MEAV as the key cost driver also found SPT to be considered inefficient relative to the other networks. **Hence, SPT's MEAV submission relative to the other networks may need some additional scrutiny to confirm the results**.

The presence of fixed costs could be an issue for SPT as a small, standalone transmission network (setting aside its parent company, Iberdrola). We argued in Section 4.5 that regression models should be able to better address the fixed costs / overheads issue compared to a ratio model due to the presence of an intercept. However, in this case the univariate Total Capex regression deems SPT to be less efficient than the Total Capex ratio model (the MEAV + Total Capex multivariate model cannot be compared to the Total Capex ratio model).

This is a concern about the model, but in addition to regression models controlling for fixed costs, we also note that:

- The regression models do acknowledge the apparent efficiency improvement in SPT's RIIO-2 submission.
- Including MEAV in the multivariate regression should also partially control for scale effects. Further scrutiny of SPT's MEAV submission may clarify this issue.
- We do not think that the presence of 'overhead' costs for CAIs is as apparent an issue compared to BSCs.

NGGT (TO)

The dispersion of efficiency score results for NGGT (TO) is more muted across the models. The assessment 'switching' from inefficient to efficient between the Total Capex and Total Capex + MEAV regression models is notable, but the resulting scores are not materially different from '1' and MEAV does appear to be an intuitive and robust cost driver to include, so a change in results when adding a significant variable to a regression is not unexpected.

7 Conclusion

Informed by our pre-modelling considerations (Section 4) and our discussion of different potential cost drivers and model specifications (Section 5), our model selection process has led us to a set of distinct preferred models for GD BSCs, for transmission BSCs, and transmission CAIs, respectively. The split in approaches for GD and transmission BSCs reflects the apparent unreliability of a cross-sector approach for BSCs as we sought to use the available regulatory toolkit to apply an approach that best fits the data for each sector. For CAIs (transmission only), we identified a preferred regression model but given the dispersion of implied efficiency scores under different model specifications, we suggest that further scrutiny of the cost submissions is undertaken outside the modelling process.

We compare our resulting modelling framework with the approaches taken to BSCs and CAIs in RIIO-1 in Table 21 overleaf (the RIIO-1 approaches are summarised in more detail in Annex A1). A cursory examination may suggest inconsistencies with RIIO-1, but our overall approach has taken a similar 'toolkit' approach to reach our recommendations, seeking to apply those methodologies with regulatory precedent that are most appropriate for assessing BSCs and CAIs for RIIO-2. In addition to the data shown in Table 21, we have conducted a range of sensitivity checks on our models – for example, we tested whether and confirmed that the BSC model for transmission appears to be robust to the inclusion of forecast data.

We have not considered ex-post adjustments in this analysis, but we highlight that Ofgem has often used such adjustments to reconcile model results that do not appear to be reasonable. Our approach has concentrated on pre-modelling and within modelling adjustments in order to best fit the data and set a reasonable efficiency benchmark, highlighting the trade-offs and intuition of each modelling approach.

We highlight how our modelling approach for CAIs (transmission only) differs from that of RIIO-T1, as we took an aggregated top-down approach compared to relying on a bottom-up combination of trend analysis and expert review for RIIO-T1. Our approach was more in-line with the regression approach for CAIs in RIIO-ED1, but the robustness of a regression-only approach may be limited by a smaller sample size. Hence, we recommend that some of the results of the CAI model need further scrutiny before they are used to set allowances.

Table 21 Comparison of RIIO-1 final assessment approaches to ECA proposed approach for RIIO-2

	RIIO-GD1	RIIO-T1	RIIO-ED1	ECA recommendations for RIIO-2
Business Support Costs				
Gross or Net costs	Gross	Gross	Gross	Gross
Group or Network level	Group	Group	Group	**GD:** Group **T:** Network
Top-down or Bottom-up	Top-down	Top-down	Top-down	Top-down
Cost driver(s)	CSV	CSV	MEAV	**GD:** MEAV **T**: CSV
Chosen method	Ratio analysis	Ratio analysis	Ratio analysis	**GD:** Ratio analysis **T**: Regression
Sample period	Historical	Historical	Historical + Forecast	Historical
IT&T	Included in top-down model	Included in top-down model	Separately assessed by ratio analysis and expert review	Excluded
External benchmark used?	Yes	Yes	No	No
Applied benchmark	Upper quartile	Upper quartile	Median	**GD**: Median **T**: Average model fit
Ex-post adjustments?	Yes	Yes	No	N/A[1]
Closely Associated Indirects				
Group or Network	-	Network	Network	Network
Top-down or Bottom-up	-	Bottom-up	Top-down	Top-down
Cost driver(s)	-	N/A	MEAV + Asset Additions	MEAV + Total Capex
Chosen method	-	Trend analysis / expert review	Regression	Regression
Sample period	-	N/A	Forecast	Historical
IT&T	-	N/A	Not included in CAIs	Excluded from baseline models
External benchmark used?	-	No	No	No
Applied benchmark	-	N/A	Average model fit	Average model fit
Ex-post adjustments?	-	N/A	Yes	N/A

Source: ECA. [1]Ex-post adjustments were outside the scope of this report. [2]We were not able to identify a satisfactory top-down model for assessing CAIs for the Ts.

ANNEXES

A1 [REDACTED]

A2 RIIO-1 assessment approaches

For point of reference, we briefly summarise Ofgem's approach for assessing BSCs and CAIs across RIIO-GD1, RIIO-T1, and RIIO-ED1.

A2.1 BSC assessment for RIIO-GD1 and RIIO-T1

In brief, Ofgem's approach for assessing BSCs for the slow-track RIIO-GD1 and RIIO-T1 price controls comprised of:

- **Assessing BSCs on a gross basis**, i.e. "*Where a company has allocated a proportion of its business support costs to direct opex, capex, or repex, or to non-network businesses then these are allocated back to the submitted net costs as pre-benchmark normalisations.*"[69] As networks may have different cost allocation policies, this was intended to ensure valid comparisons between network companies.

- **Assessing BSCs at the group ownership level**, rather than network BSCs. For example, the BSCs of National Grid's electricity and gas transmission, and gas distribution networks were assessed together.

- **Assessing BSCs in aggregate** (i.e. pooled), except for insurance which was separately assessed.[70] Ofgem changed its approach from initial proposals (of a disaggregated (bottom-up) approach) to address concerns over 'cherry picking' and be more consistent with other activity assessments.

- **Applying normalisations to the baseline**. In addition to the net to gross normalisation, Ofgem applied other pre-benchmark normalisations where a network company had activity in 2010-11 that would not be continued in RIIO-1 or additional activities in RIIO-1 that were not in the baseline (subject to the company providing sufficient justification).

- **Deriving a composite cost driver.** The cost driver was a composite of cost drivers (e.g. IT end-users, revenue, employees, expenditure) for individual BSC activities weighted by the cost of each activity.

- **Using external comparators**. At final proposals, Ofgem used external comparators, rather than network costs to derive baseline allowances, except for CEO and group management where a composite of network and external

[69] Ofgem, RIIO-GD1: Initial proposals – Supporting document – Cost efficiency, Appendix 6, July 2012, 1.12.

[70] Ofgem stated that insurance costs were excluded from benchmarking "*given differences in risk appetite and appropriate levels of coverage between companies and sectors*", Appendix 6, para 1.36, *ibid*.

benchmarks was used. This represented a shift from initial proposals, where Ofgem used the lower of the network and external benchmark at a disaggregated level. In aggregate, the difference between the network and external benchmarks was negligible (although there were significant differences for some individual activities).

- **Using an upper quartile, ratio benchmark**. Ofgem applied the upper quartile benchmark (a ratio of total cost to a composite cost driver) to the relevant groups' composite cost driver to derive total efficient business support costs. Although Ofgem applied the external upper quartile (except for CEO and group management), there was little difference between this and that of the networks.

- **Applying ex-post adjustments.** Ofgem applied ex-post adjustments to benchmarked costs. These included where a company identified and justified exceptional costs over RIIO-1 and where they submitted robust evidence of cost efficiency (at an activity level). Ofgem also made additions (at final proposals) for:

  - An increase in NG's baseline to reflect operational growth in NGET TO

  - An increase to reflect additional costs network companies face relative to the external comparators

  - An increase in SO costs, recognising the IT intensive nature of the SO role, based on an assessment by consultants.

A2.2 BSC assessment RIIO-ED1

In the electricity distribution (ED) slow track price controls, applying from 1 April 2015 to 31 March 2023, Ofgem adopted a slightly different approach to assessing BSCs than in RIIO-GD1 and RIIO-T1. In brief, the approach comprised:

- **Assessing BSCs at the group level.** Ofgem stated this addresses the problem of differences in allocation methodologies across ownership groups as well as accounting, to some degree, for the sharing of costs across Distribution Network Operators (DNOs) within a group.

- **Assessing BSCs in aggregate** (finance and regulation including insurance, HR and non-operational training, property management, and CEO and group management), except for IT & telecoms.[71]

- **Using data from 2010/11 to 2022/23.**

- **Using MEAV as the cost driver**. Other drivers (including some used for fast track determinations), such as employee numbers, "*were rejected for three key*

[71] IT & telecoms allowances were based 50% on a benchmarking exercise and 50% on expert review (which was in combination with a review of operational IT&T and non-operational capex IT&T).

reasons: the lack of economic rationale; their endogenous nature; and significant changes were made to them by DNOs between the fast-track and the slow-track submissions, reducing our confidence in them.”[72]

- **Using DNOs’ own BSCs** to set the benchmark, rather than external comparators.
- **Applying a median ratio benchmark** to determine allowances.

A2.3 CAI assessment in RIIO-T1

Across both electricity transmission (ET) and gas transmission (GT) slow-track price controls, Ofgem assessed CAIs at a disaggregated level. Ofgem made use of its own analysis as well as recommendations from technical consultants. The approaches adopted in setting allowances were trend analysis and expert review. Ofgem did not use regression analysis or other forms of benchmarking in setting allowances.

In the case of GT, NGGT forecast a reduction in average annual CAI operating costs for RIIO-1, compared to the previous price control period. Ofgem set allowances based on NGGT’s forecasts, but subject to an assumed efficiency of 1.5%

In the case of ET, NGET forecast a slight increase in average annual CAI operating costs for RIIO-1, compared to the previous price control period. Ofgem set allowances for some CAIs based on NGET’s forecasts, some on the basis of its consultant’s recommendations,[73] and some on its own analysis.

A2.4 CAI assessment in RIIO-ED1

In the RIIO-ED1 slow track assessments of CAI, Ofgem used regression analysis across eight sub-categories (aggregated) and ratio analysis for the remaining three.[74] Some of the key elements of the regression analysis were:

- **MEAV and asset additions were the cost drivers of choice**, as proxies for scale and workload.
- **Only forecast CAI data were regressed** as forecast ED1 costs were much lower than the previous price control period.
- **Regressions were conducted at the network level**, rather than at the ownership group level. Group level regression result in fewer observations than network level regressions and did not give plausible results. Also, Ofgem

[72] Ofgem, RIIO-ED1: Final determinations for slow-track electricity distribution companies Business plan expenditure assessment, November 2014, 10.55.
[73] The consultant’s approach involved taking a 2010/11 baseline, making adjustments about changes on the network and applying a 2.25% efficiency factor.
[74] The three areas not subject to regression analysis were wayleaves, vehicles and transport, and operational training.

did not believe there was sufficient evidence of shared costs across the CAI sub-categories subject to regression.

- **Ofgem aggregated eight sub-categories of CAI** to address boundary issues in reporting of CAI costs.

- **Ex-post adjustments** were made for three networks for which the regression results were considered to be harsh.

- Ofgem also tried **grouping CAI categories based on common cost drivers**.[75] This proposal was a move away from the previous price control, which had used direct expenditure as a driver, and which could create perverse incentives, e.g. by rewarding inefficient companies.

[75] In the strategy consultation Ofgem proposed two groups: (A) those that almost entirely support the delivery of direct activities and would 'flex' with the volume of work; and (B) those that have a substantially 'fixed' cost. This approach was intended to allow for cost drivers that are as closely aligned to the activity as possible.