Introduction

1.1 This Annex presents a more detailed explanation of our cost assessment methodology in a step-by-step guide format, with a focus on our approach to regional factors and econometric modelling. It provides further detail on the analysis we have carried out and set out in the RIIO-2 Final Determinations – Gas Distribution Sector decision. Our analysis has taken into account relevant considerations, including stakeholders' responses to our Draft Determinations. Details on the rationale underlying our methodological choices can be found in the GD Annex and the company annexes.

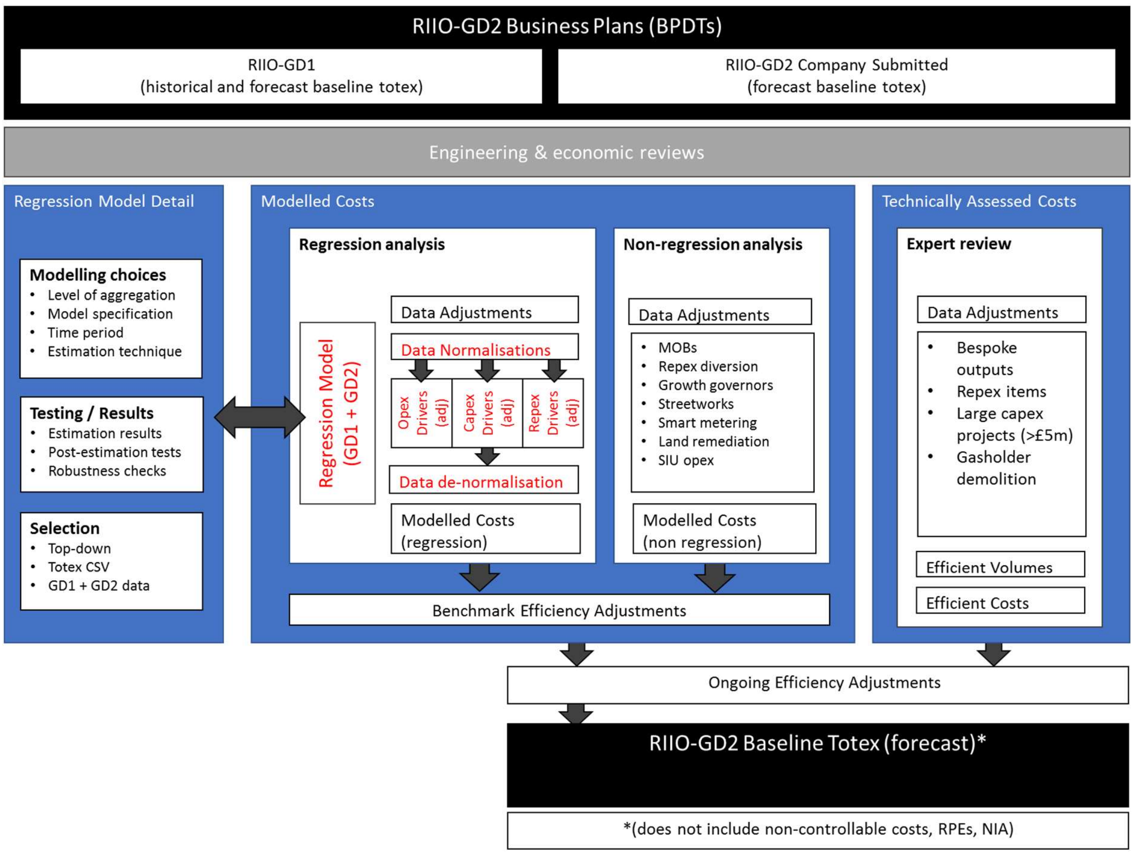
1.2 The first section presents an overview of the process we have undertaken to set the baselines in our FD decision. The second section details the normalisations and adjustments we have made on the submitted costs. Our econometric modelling and efficiency assessment are also set out in the following sections.

Overview of the totex benchmarking process

1.3 We have used regression, non-regression and technical assessment to determine baseline totex for RIIO-GD2. The overall process is summarised in chapter three of the GD Annex.

1.4 Figure 1 below provides a more detailed overview of our econometric modelling approach and of how it fits in the overall process.

Figure 1 Econometric modelling in the overall cost assessment process



Submitted and normalised data

1.5 The data we used for benchmarking was submitted by the GDNs in the RIIO-GD2 Business Plan Data Templates (2013-14 to 2025-26) resubmitted in September 2020 further to DDs. We adopted total controllable expenditure (totex) as our measure of total costs, which we defined as:

Controllable Totex = controllable opex + capex + repex.

1.6 As in RIIO-GD1, we used a seven-year rolling average to smooth capex because of related sporadic expenditure in some of the GDN cost activities, particularly LTS and other capex. To do this, we used capex data prior to RIIO-GD1 (back to 2007-08).

1.7 Before carrying out our regression analysis, we made data normalisations and adjustments which we have explained in the GD Annex. These adjustments include[1]:

- embedded ongoing efficiency adjustments
- exclusions
- volume related adjustments
- reclassifications
- non-regression and technical assessment
- regional factors.

A full list of data normalisations and adjustments for each GDN is provided in the company annexes.

1.8 **Embedded ongoing efficiency adjustments** were made to submitted costs based on information received by the companies via supplementary questions. We stripped out ongoing efficiency from submitted costs in order to improve comparability between GDNs, resulting in an improved benchmarking analysis.

1.9 **Exclusions** were made to historical costs which were previously classified as controllable costs but are now classified as non-controllable costs (eg Xoserve, PPF Levy costs). Moreover, we excluded capex relating to historical large projects (above £5m), in order to align with our approach for forecast large projects, and maintain a consistent dataset over the 13-year period. We also excluded pass-through items and costs we have proposed to be subject to an uncertainty mechanism.

1.10 **Volume related adjustments** were made to specific cost activities that did not satisfy a needs case, such as asset management repex programmes which did not meet our CBA payback criteria. Different to our Draft Determinations, we made upwards adjustments to repairs costs and workload to account for disallowed repex workload. We also made upward adjustments to some GDNs' costs (and corresponding

[1] Our reference to Business Plans as a source of information includes information from the plans, Business Plan Data Templates (BPDTs) submitted in December 2019 and September 2020, as well as any corrections/clarifications provided though supplementary questions (SQs).

workloads) where we found the baseline volume assumption to be less than a "P50 forecast", for example for Cadent's connections and mains reinforcement expenditure. Moreover, to enhance comparability with forecast data, we adjusted historical emergency costs upwards to account for loss of meterwork.

1.11 **Reclassifications** were made where we considered that a GDN reported certain cost activities incorrectly or differently to the majority of GDNs.

1.12 We removed certain forecast costs for **separate assessment**. We distinguished between costs separated for a detailed technical assessment and those suitable for a modelled non-regression assessment.[2]

- Technical assessment – around 6% of forecast totex. This includes large Capex projects, Gasholder demolitions, Physical security expenditure and the majority of Bespoke outputs (which were considered unsuitable for modelling). Our efficient view of these costs was added to modelled costs, and was not subject to a further efficiency catch-up challenge. Different from Draft Determinations and following stakeholders' feedback, at Final Determinations we included IT&T capex into the totex figure for regression analysis.
- Non-regression assessment – around 8% of forecast totex. This includes MOBs, Streetworks, Repex Diversions, Streetworks, Smart Metering, Land remediation, SIU opex and Growth Governors. At Final Determinations, we have adopted our Draft Determinations position and removed these cost activities and modelled the costs, then added these costs to overall modelled costs prior to applying the efficiency benchmark.

1.13 In order to ensure comparability between GDNs, we applied regional labour, urbanity and sparsity adjustments to submitted totex. A detailed explanation of how the related indices were computed following stakeholders responses to our Draft Determinations can be found in Appendices A and B. As detailed in the company annexes, we also made adjustments for company specific factors.

1.14 Finally, we retrieved the capex data we used to calculate the seven-year rolling average from GDPCR Regulatory Reporting Packs (RRPs) over the period 2007-08 to 2012-13.

[2] Technical and non-regression assessment are detailed in chapter three of the GD Annex.

We normalised this data to reflect the adjustments made to the data for the period 2013-14 to 2025-26 (eg regional factors and exclusions).

1.15 We used the costs we derived at the end of this process (ie normalised and adjusted costs) for our regression analysis. We assessed costs on a gross basis (ie including customer contributions) and then adjusted to net costs after modelling. A GDN commented that costs should be assessed on a net basis, because net costs are what consumers need to fund. Nonetheless, we have adopted our Draft Determinations position and assessed costs on a gross basis, as we consider that the level of efficiency is better assessed on the overall costs incurred by networks, independently of how these costs are funded. Although different from RIIO-GD1, the approach we adopted is in line with RIIO-ED1.

Econometric analysis

1.16 This section describes our econometric approach to cost assessment for Final Determinations. It mainly reports our methodological decisions as compared to RIIO-GD1 and Draft Determinations. For details on stakeholders' feedback and rationale underlying these decisions, see chapter three of the GD Annex and company annexes.

A single totex model

1.17 In RIIO-GD1, the econometric analysis was performed at two different levels of aggregation, top-down and bottom-up, which were then combined using equal weight. At our Draft Determinations fro RIIO-2, we proposed a single top-down regression model for our econometric analysis.

1.18 As set out in our Final Determinations GD Annex, we have implemented our Draft Determinations position and retained the use of a single top-down model. We consider that using a single top-down model better accounts for cost complementarities, trade-offs and potential reporting inconsistencies between GDNs than alternative approaches.

Selected sample

1.19 The RIIO-GD1 econometric models were estimated on four years of historical data (2008-

09 to 2011-12) and on two years of forecast data (2013-14 to 2014-15). This allowed us to take into account both historical and expected relative performance of GDNs.

1.20 At Draft Determinations, we proposed using RIIO-GD1 and RIIO-GD2 data (2013-14 to 2025-26) in order to increase the size of the sample used in the regression analysis and to ensure that changes in technology and scope for future efficiency gains are explicitly taken into account. We have implemented the Draft Determinations position for Final Determinations and used RIIO-GD1 (historical and forecast) and RIIO-GD2 forecast data for our econometric analysis.

Model specification and estimation

1.21 As in RIIO-GD1 and Draft Determinations, we employed a Cobb-Douglas functional form. This is a standard approach used in cost assessment literature as it allows for economies of scale to be captured.

1.22 Specifically, our totex model takes the following form:

$$ln(totex_{it}) = \beta_0 + \beta_1 \, ln(totex\ CSV_{it}) + \beta_2 t1 + \beta_3 t2 + \epsilon_{it},$$

Where β_0 is a constant term, β_1 is the coefficient associated with the cost driver (totex CSV) and ε_{it} is the error term representing the component of costs not explained by the cost driver (ie noise, measurement errors and inefficiency) for GDN *i* at time *t*.[3]

1.23 To account for unobserved time effects, this specification also includes a linear trend for historical data (t1) and another one for forecasts (t2). This allows us to capture changes in real expenditure due to frontier shift and potentially other exogenous factors such as changes in service quality. Following a GDN's response to our Draft Determinations, we slightly modified the specification of the time trend variables. The historical time trend t1 takes value 1 in 2014 and increases by one after each year until 2026, while the forecast time trend t2 takes value 0 between 2014 and 2020, value 1 in 2021 and increases by 1 afterwards.

[3] It is worth noting that the logarithmic transformation of the variables results in two advantages. First, the corresponding estimated coefficients can be interpreted as cost elasticities. Second, the transformed variables follow more closely a normal distribution, thus better reflecting the assumptions underlying our estimation approach.

1.24 As in RIIO-GD1 and Draft Determinations, the model we have decided to use in Final Determinations was estimated via Ordinary Least Squares (OLS), which produced an average relationship between totex and the cost driver, under the assumption that the data points are independent. We estimated the model using clustered robust standard errors to account for the fact that, in reality, data points relating to the same GDN are correlated and thus not fully independent (ie to address potential heteroskedasticity) and to increase accuracy when assessing statistical significance.

1.25 Since the number of comparators (eight GDNs) remained unchanged, we have decided not to adopt more data-intensive estimators as our main approach. Nonetheless, as shown in para 1.63-1.66, we checked the robustness of the totex model by estimating it via both Random Effects (RE) and Stochastic Frontier Analysis (SFA).

Totex Composite Scale Driver (CSV)

1.26 In presence of limited sample size, the inclusion of a relatively high number of drivers in the model specification is normally not considered appropriate from an econometric perspective. However, missing out relevant drivers of costs might limit the explanatory power of the model itself. A way to conveniently address this issue is to use a composite scale variable (CSV).

1.27 A composite scale variable (CSV) is a weighted average of different drivers. This ensures a parsimonious model (ie a single driver) while incorporating as much information as possible.

1.28 As in RIIO-GD1 and Draft Determinations, for Final Determinations we have decided to use a CSV in our totex model. Specifically, we used the same driver as in RIIO- GD1, totex CSV, with the same individual components. These components include a mix of scale and workload drivers, reflecting the drivers of the disaggregated cost activities included in our definition of totex.

1.29 The individual components in the CSV correspond to the drivers used in the bottom-up regression models in RIIO-GD1: emergency CSV, maintenance MEAV, total external condition reports, repex synthetic costs, mains reinforcement synthetic costs, connections synthetic costs and MEAV. By using the drivers from the disaggregated models we have retained the information that we used in the bottom-up analysis, while allowing the model to solve the trade-offs between the expenditure on different activities.

1.30 In terms of weights assigned to the individual components we have decided in Final Determinations tofollow the RIIO-GD1 and Draft Determinations approach, where these weights were based on industry spend proportions (ie ratios of controllable, normalised and adjusted costs) for the disaggregated cost activities to which the drivers apply, with the residual weight assigned to the scale driver Modern Equivalent Asset Value (MEAV). This approach was able to take into account the relative importance of each cost driver based on knowledge of GDNs' costs. The resulting totex CSV is as follows:

$$totex\ CSV = (emer\ CSV)^{\delta 1} * (maint\ MEAV)^{\delta 2} * (tot\ ext\ cond\ rep)^{\delta 3} * (repex\ syn)^{\delta 4} * (reinf\ syn)^{\delta 5} * (conn\ syn)^{\delta 6} * (MEAV)^{\delta 7},$$

where δ_i (*i=1, …, 7*) are the weights corresponding to the individual components (with δ_7=1- δ_1-…- δ_6). The values of these weights are showed in Table 1 below.

Table 1 Cost activities, totex CSV components and weights

Cost activities	Totex CSV component	Weight
Emergency	Emergency CSV[1]	0.05
Maintenance	Maintenance MEAV	0.08
Repairs	Total external condition report	0.05
Repex	Repex synthetic cost	0.38
Mains reinforcement	Mains reinforcement synthetic cost	0.01
Connections	Connections synthetic cost	0.06
Work Management, Business Support, Other Direct Activities, Training and Apprentices, Other Capex	MEAV	0.37

[1] Composite scale variable including customer numbers (0.80) and total external condition reports (0.20).

1.31 In order to account for both fixed and variable elements of emergency costs, the driver for GDNs' emergency activity is a CSV comprising customer numbers (0.80) and total external condition reports (0.20). Customer numbers capture the fixed element of these costs, while total external condition reports are assumed to drive the variable element. The latter are also assumed to be the main driver of repair activities (ie costs for site attendance, excavation, repair of leaking mains and road reinstatement).

1.32 The selected driver for maintenance activities is maintenance MEAV, a subset of MEAV only including above ground assets (ie those primarily requiring both routine and non-routine maintenance). Different to RIIO-GD1 and Draft Determinations, we have decided in Final Determinations to include embedded gas entry points (EGEPs) into maintenance MEAV. We have not included risers to avoid double counting, as Multi Occupancy Buildings were separately assessed via non-regression analysis.

1.33 Finally, we describe below our approach to updating the other components of the totex CSV: MEAV and synthetic cost drivers.

Updating MEAV

1.34 MEAV is the current replacement value of an asset. The sum of MEAVs corresponding to a GDN's assets provides a proxy for the GDN's scale of operation. In RIIO-GD1, the assets included in MEAV calculations were: mains, governors, Local Transmission Systems (LTS), storage, National Transmission System (NTS) offtakes, Pressure Regulating Stations (PRS) and services (proxied by customers number).

1.35 Differently from other scale variables such as network length and customer numbers, in RIIO-GD1 and at Draft Determinations MEAV was deemed to better reflect the complexity of the network, and thus was the preferred scale driver in previous price controls. We took the same approach for RIIO-GD2 and employed MEAV as the main scale driver at Draft Determinations. We have implemented this approach for Final Determinations.

1.36 Nonetheless, in response to stakeholders' feedback at Cost Assessment Working Groups and Draft Determinations, for Final Determinations we have updated the MEAV driver with respect to assets' composition and, partially, replacement values of these assets. In terms of assets' composition, we note that in RIIO-GD1 both EGEPs and risers were excluded from MEAV asset base. In order to ensure MEAV better reflects the scale of GDNs' operation and following stakeholders' comments at CAWGs and in response to our Draft Determinations, we included both asset types in the asset base for RIIO-GD2. While the inclusion of EGEPs did not lead to substantial changes to MEAV values, the inclusion of risers had a relevant impact in MEAV calculations, especially for London and Southern networks.

1.37 In terms of asset replacement values, we used the same values as in RIIO-GD1 (converted into 2018-19 prices) for governors, LTS, storage, NTS offtakes, PRS and services. At Draft Determinations we proposed to adjust the replacement values of governors, NTS and PRS to account for differences in throughput between GDNs. For Final Determinations we updated the GDN-specific replacement values to account for differences in both throughput and number of assets, using 2019 as the reference year for the adjustments.

1.38 For gas mains, to address the difference in diameter bands reporting between RIIO-GD1 and RIIO-GD2 Business Plan Data Templates (BPDTs), we used linear interpolation as follows:

- We assigned the unit costs to set RIIO-GD1 allowances to the mid-point of historical band
- We interpolated the unit costs based on the difference between the midpoint of the old bands and the new bands.

1.39 As for EGEPs, based on the engineering view, at Draft Determinations we consulted on the replacement value of EGEPs to be 25% of that for a PRS, with no differentiation for pressure level (above or below 7bar). An EGEP is much simpler than a PRS as there isn't any pressure reduction equipment and therefore the pressure level it is connected to has little impact on its configuration. Following consideration of responses to Draft Determinations, including a GDN's response to, at Final Determinations we have revised downwards our estimate of the replacement value for an EGEP.

1.40 Finally, for risers, at Final Determinations we have updated both replacement values and number of assets using information submitted by the GDNs following a coordination effort to ensure reporting consistency.

Updating synthetic cost drivers

1.41 In RIIO-GD1, synthetic unit costs were used to calculate cost drivers in the repex, connections and reinforcement regressions. For each type of mains replacement activity (defined by material and/or diameter)4 and related services interventions, as well as for each type of mains reinforcement and connections activity, a fixed synthetic unit cost was calculated for all GDNs.

1.42 These synthetic unit costs consisted of average industry costs determined using historical data. These were then multiplied by the GDN specific workloads for each activity and summed to arrive at a single synthetic cost driver (defined in £m) for each distribution

[4] Mains replacement activities included capitalised replacement. We did not treat capitalised replacement separately because the nature of the activity is the same as mains replacement, the difference being only in reporting.

network, which was used in the regression analysis. At SSMC and RIIO-2 Tools for cost assessment consultations, we consulted on the RIIO-GD1 being appropriate for RIIO-2.

1.43 In response to these consultations, the majority of stakeholders suggested updating the synthetic unit costs. We engaged CEPA to develop an assessment framework and methodology for the update. CEPA proposed a process which considered both quantitative and qualitative criteria for the calculation of appropriate synthetic unit costs covering both RIIO-GD1 and RIIO-GD2 periods. A detailed explanation of the analytical framework and rationale for each of the proposed criteria can be found in the Draft Determinations Annex "GD2: Synthetic Unit Costs Update Annex".

1.44 At Draft Determinations we slightly modified the process proposed by CEPA. Specifically, we applied the following quantitative criteria before computing industry average unit costs:

1. **Minimum number of observations:** data must be provided for a minimum of two historical reporting years and for a minimum of two GDNs

2. **Outlier test:** unit costs must be within 100% of the industry average unit cost over the same period

3. **Maximum unit cost variability between GDNs:** to check whether individual GDN unit costs are within 40% of the industry average over the same period

4. **Maximum unit cost variability over time:** to check whether unit costs calculated in each year are within 40% of the average unit cost over the considered period.

1.45 The first two criteria were assigned a higher level of importance than the last two, implying that failure of criteria three and four didn't necessarily result in discarding a unit cost automatically. In line with CEPA's suggested framework, at Draft Determinations we also applied five qualitative criteria to complement the quantitative assessment. Specifically, we looked at data quality and comparability, routineness of work and materiality before the quantitative assessment, and considered potential drivers that cause differences in unit costs between GDNs and/or over time after the unit costs were computed. For Final Determinations, we have followed the same approach.

1.46 In line with RIIO-GD1, at Draft Determinations we proposed to update synthetic unit costs based on historical information. To increase the number of observations, we initially considered data prior to RIIO-GD1 as well. However, given the lack of a sufficient level of disaggregation in the data, we only used RIIO-GD1 historical data (ie 2013-14 to 2018-19). For Final Determinations, following stakeholders' feedback and further analysis, we have decided to use both RIIO-GD1 and RIIO-GD2 data for the calculation of all synthetic unit costs. This allowed us to derive synthetic unit costs that better reflected the engineering logic compared to those computed at Draft Determinations.

1.47 As a starting point, for all repex and capex (mains reinforcement and connections) activities, at Draft Determinations we considered the lower level of disaggregation to which apply the proposed criteria. First, we ensured a sufficient number of observations (criterion 1) and removed outlier observations (criterion 2). Then, if the calculated synthetic unit cost failed to meet the other selection criteria, we first re-iterated the procedure at a higher level of aggregation (ie summing up similar cost activities). If the criteria were still not met, we computed the synthetic unit cost for the activity by applying a scaling factor to the closest activity for which it was possible to compute unit costs within this framework. The scaling factor was based on the assumption that the percentage difference between unit costs of different activities was the same as between the synthetic unit costs used in RIIO-GD1. We have implemented an analogous approach for Final Determinations.

1.48 Our RIIO-GD2 approach did not split out differences in replacement techniques and innovative processes (eg CISBOT) due to a lack of consistent information at a more granular level. However, we did account for differences in regional wages and productivity by applying the same updated indices used to normalise submitted costs.

1.49 The updated synthetic unit costs were then used to calculate repex and capex synthetic cost drivers for the regression analysis as the sum of the products of synthetic unit costs and workloads of the corresponding activities. The complete list of cost categories for which we have calculated a synthetic unit at Final Determinations is provided in Appendix C.

1.50 For repex, the workloads associated with each of the following activities are included within the synthetic cost driver: Tier 1 iron mains, Tier 2A iron mains, Tier 2B iron mains, Tier 3 iron mains, steel mains <=2", steel mains >2", iron mains >30m from a building, other policy & condition mains, services associated with all of the aforementioned mains replacement activities, services not associated with mains replacement.

1.51 Other changes from the RIIO-GD1 calculation of synthetic repex are the exclusion of non-rechargeable diversions (separately assessed) and the inclusion of services not associated with mains replacement. We have included services not associated with mains replacement within the totex regression, so as to capture any interplay with GDNs' opex activities.

1.52 Moving to capex, the synthetic cost driver for mains reinforcement distinguishes between mains below and above 180mm. In terms of synthetic unit costs, no distinction was made between general and specific reinforcement. The Final Determinations approach is in line with the approach taken at RIIO-GD1.

1.53 Finally, the synthetic cost driver for connections accounted for mains and services workload, distinguishing between mains below and above 180mm diameter. However, differently from RIIO-GD1, at Final Determinations we have aggregated new housing, existing housing and non-domestic, implying that the two types of connections were assumed to exhibit similar unit costs. Moreover, we included connections related to the Fuel Poor Network Extension Scheme (FPNES), which in RIIO-GD1 were assessed separately.

Results and post-estimation tests

1.54 For Final Determinations, we adopted the Draft Determinations position and used OLS with clustered robust standard errors to estimate the coefficients of our totex model on the sample covering the period 2013-14 to 2025-26. Table 2 below shows the regression results. Specifically, column *OLS1* in the table reports the results of the totex model we have proposed, while columns *OLS2* and *OLS3* are alternative specifications that we estimated as a robustness check.

Table 2 Totex regression results

Ln_totex	OLS1	OLS2	OLS3
Ln_totex_csv	0.786*** (0.000)	0.785*** (0.050)	0.786*** (0.000)
Ln_totex_csv_sq			-0.018 (0.046)
t1	-0.003 (0.533)		-0.003 (0.005)
t2	0.004 (0.513)		0.004 (0.006)
Year2015		-0.009 (0.013)	
Year2016		-0.020 (0.012)	
Year2017		-0.025 (0.017)	
Year2018		-0.053 (0.018)	
Year2019		-0.038 (0.023)	
Year2020		-0.021 (0.029)	
Year2021		-0.017 (0.036)	
Year2022		-0.019 (0.026)	
Year2023		-0.022 (0.028)	
Year2024		-0.029 (0.028)	
Year2025		-0.034 (0.029)	
Year2026		-0.033 (0.027)	
Constant	-0.059 (0.867)	-0.040 (0.036)	5.506 (0.038)
Adj R^2	0.918	0.911	0.917
Obs.	104	104	104
Standard errors in parentheses. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.			

1.55 Model fit of our model (*OLS1*) is good, as the adjusted R^2 is 0.918 (higher compared to the Draft Determinations model, which showed an adjusted R^2 of 0.865). The estimated coefficient of the totex CSV is 0.786, implying that, everything else equal, a 1 percent increase in the totex CSV would result in a 0.786% increase in totex. The estimated coefficients of the two time trends are not statistically significant.

1.56 We obtained similar results to *OLS1* when we replaced the two time trends with year dummies (column *OLS2*) and when we estimated the model considering different time periods.[5]

1.57 In line with CEPA and our academic advisor recommendations[6], we also performed the following post-estimation tests to check the stastistical robustness of our Final Determinations model:

- **Normality:** to test whether residuals follow a normal distribution as per standard OLS assumptions. It is worth noting that the failure of this test does not affect the properties of the OLS estimator and is only a problem if the sample size is very small.
- **Heteroskedasticity:** to test whether residuals have a constant variance (ie are homoskedastic). However, the presence of heteroscedasticity would only affect standard errors, as the OLS estimates would still be unbiased.
- **Pooling:** to test whether the coefficients of the OLS model are significantly different from the true coefficients of the same model run on each individual cross-section of the data. The failure of this test indicates that panel data analysis might not be appropriate.
- **RESET:** to test whether there are any omitted non-linearities in the model. If this test fails, it might be appropriate to test a different model specification (eg inclusion of a quadratic term in case of univariate regression or a translog specification).

1.58 Table 3 provides a summary of each of the above tests on the selected model (*OLS1*).

---

[5] The estimated coefficient of totex CSV was 0.788 on historical data (2013-14 to 2019-20), 0.792 on RIIO-GD1 data (2013-14 to 2020-21) and 0.774 on RIIO-GD2 data (2021-22 to 2025-26). The corresponding R^2 (0.899, 0.904 and 0.934, respectively) were also similar to that of our main model.

[6] See Annexes to RIIO-2 Tools for Cost Assessment Consultation available here.

Table 3: Model test result summary

Test	Result	Note/Action
Normality	Passed (Chi-square 0.2617)	We cannot reject the hypothesis that residuals follow a normal distribution. No action taken.
Heteroskedasticity	Passed (p-value 0.5061)	Residuals are homoscedastic.
Pooling	Passed (p-value 1)	Panel data analysis seems appropriate. See robustness checks below.
RESET	Failed (p-value 0.0000)	Tested alternative model specification (*OLS3*, Table 2)

1.59 As can be seen from Table 3, the selected totex model passed both normality and heteroskedasticity tests. However, the pooling test indicated that panel data analysis might be more appropriate than OLS. It's worth noting that these results should be taken with caution, as the cross-sectional regressions (one for each year) on which the test is based only consider eight data points. Nonetheless, we address this issue below, where we show the results of additional robustness checks that explicitly account for the panel nature of the data.

1.60 Moreover, as for the totex model proposed at Draft Determinations, our Final Determinations' totex model failed the RESET test, suggesting the presence of omitted non-linearities. To address this issue, we estimated a model that included a quadratic term for the totex CSV. Column *OLS3* in Table 2 shows the results of this alternative model specification.[7]

[7] We normalised the driver variables with respect to the sample mean to avoid difficulties in coefficients interpretation due to their different magnitude.

1.61 The signs of the coefficients are reasonable from an economic perspective (positive for the logarithm of totex CSV and negative for its square), indicating a U-shaped relationship between totex and totex CSV (ie at first totex increase with the driver, then they decrease). However, coefficients are not all statistically significant and model fit does not improve substantially compared to our main model. Moreover, model *OLS3* also fails the RESET test. We obtained similar results to *OLS3* when we estimated a translog functional form to check for additional non-linearities in the model. Thus, we didn't have strong reasons to discard the selected model *OLS1* based on the RESET test results.

1.62 Finally, in order to explore the stability of the model, we estimated the same model by removing individual years or GDNs. The removal of any year from the sample size resulted in substantially unchanged regression estimates in terms of both magnitude of the estimated coefficients (between 0.782 and 0.792) and model fit. However, as expected the estimated coefficient of totex CSV exhibited some variation when individual GDNs were excluded from the sample, although within an acceptable range (between 0.743 when Scotland was excluded and 0.821 when London was excluded).

Additional robustness checks

1.63 We performed additional checks to ensure robustness of the totex model. This was primarily done by comparing the results obtained via OLS estimation with those from different estimation techniques (RE and SFA).

Random Effects (RE) and Stochastic Frontier Analysis (SFA) models

1.64 The selected totex model was estimated via OLS with clustered robust standard errors to account for the fact that the observations in the sample are not fully independent but clustered by GDN. Nonetheless, the pooling test indicated that panel data analysis might be appropriate, and the Breusch-Pagan test result suggested using a Random Effects (RE) estimator instead of OLS.

1.65 We further investigated the robustness of our totex model by testing alternative Stochastic Frontier Analysis (SFA) models with the support of our academic advisor Prof. Andrew Smith. SFA models explicitly consider the separation between inefficiency and statistical noise. In line with preliminary analysis detailed in the Annex "Note for Ofgem on Alternative Methodologies: Some Preliminary Analysis", the following SFA models were tested:

- Pooled: it doesn't account for the panel nature of the data. The inefficiency term varies over time, but in an unstructured way
- Battese and Coelli (1988): time-invarant inefficiency (*BC88*)
- Battese and Coelli (1992): time-varying inefficiency (*BC92*).

Table 4 Estimation results

Ln_totex	OLS1	RE	SFA Pooled	BC88	BC92
Ln_totex_csv	0.786*** (0.046)	0.709*** (0.064)	0.788*** (0.042)	0.699*** (0.055)	0.730*** (0.119)
t1	-0.003 (0.005)	-0.003 (0.005)	0.000 (0.009)	-0.003 (0.005)	0.001 (0.006)
t2	0.004 (0.05)	0.003 (0.006)	0.001 (0.010)	0.003 (0.006)	0.002 (0.005)
Constant	-0.059 (0.341)	0.486 (0.452)	-0.172 (0.272)	0.457 (0.376)	0.201 (0.884)
Adj R^2	0.918				
Log-likelihood	134.84	180.40	136.08	181.92	183.67
Obs.	104	104	104	104	104
Standard errors in parentheses. *, ** and *** indicate statistical significance at the 10%, 5% and 1% level, respectively.					

1.66 As shown in Table 4, all of the estimated models exhibit very similar results, indicating that our Final Determinations totex model is robust to different estimation techniques. Specifically, models *OLS1* and pooled SFA produce similar coefficients, while RE results are similar to the SFA models that explicitly account for the panel structure (*BC88* and *BC92*).

Interestingly, time variation in *BC92* model is not significant, which might explain the similarity with *BC88* results. As expected, efficient costs from SFA models are on average lower than with OLS and RE, which would result in tougher allowances. However, given our data limitations, we prefer not to rely on models that are more data intensive and are based on discretionary distributional assumptions for the error term. Thus, we implemented our Draft Determinations position and used the OLS model for Final Determinations, also in light of its higher degree of transparency. Further detail of the decision can be found in the GD Annex.

Determining modelled costs

Modelled totex

1.67 In line with RIIO-GD1 and Draft Determinations, we used the following formula to compute modelled costs from our regression analysis for each GDN *i* and each year *t*:

$$Modelled\ totex_{it} = \alpha * \exp\left(\beta_1 + \beta_2 \ln\left(tote\quad CSV_{it}\right) + \beta_3 t1 + \beta_4 t2\right)$$

Where α is an alpha correction factor and the coefficients are those estimated from the selected model (*OLS1*). At Final Determinations, and as a change from Draft Determinations in response to consideration of DD responses, we used the unsmoothed totex CSV variables instead of the smoothed ones to compute modelled costs. This ensures better comparability between modelled and submitted costs.

1.68 Indeed, as we used a logarithmic transformation of the data for our totex regression, the exponential transformation into costs would tend to underestimate modelled costs. To resolve this, we followed the RIIO-GD1 approach and multiplied modelled costs with an estimate of the expected value of residuals (ie the above mentioned alpha correction factor). The alpha correction factor corresponds to the estimated coefficient from a linear regression of normalised adjusted totex on those predicted from the selected model without a constant. The computed alpha factor was 1.002 (equal for all GDNs due to homoscedasticity), implying that the adjustment to totex due to the logarithmic transformation was minimal.

1.69 We computed modelled costs using both unadjusted and adjusted cost drivers so to derive the effect of workload adjustments. However, as explained in the following section, only adjusted modelled costs were then used to compute the efficiency score for each GDN.

Efficiency assessment

Calculating efficiency scores and choosing the efficiency benchmark

1.70 For each GDN, we calculated a totex efficiency score for the RIIO-GD2 period as the ratio between submitted normalised adjusted costs and adjusted modelled costs:

$$Efficiency\ score = \frac{Submitted\ (normalised\ adjusted) costs}{Adjusted\ modelled\ costs}$$

1.71 For Final Determinations we used the 85th percentile efficiency score (0.96) from the GDNs' efficiency scores as benchmark totex (see Table 6), which will be applied in the last two years of RIIO-GD2. We selected the 85th percentile score rather than the frontier to acknowledge that part of the difference in costs across GDNs related to factors other than GDNs' relative efficiency (ie measurement errors and statistical noise). As described in more detail in the GD Annex, further to consideration of responses to Draft Determinations, at Final Determinations we have applied a glide path to the 85th percentile starting from the 75th percentile in the first year of RIIO-GD2.

1.72 It is worth noting that, compared to Draft Determinations, the range of efficiency scores at Final Determinations has decreased substantially (from 0.28 to 0.19). This confirms that the methodological changes made following stakeholders' responses and additional analysis has resulted in an increased robustness of our approach.

OFG1161

Table 6: GDNs' efficiency scores (RIIO-GD2 period)

GDN	Final Determinations	Draft Determinations
EoE	0.97	1.10
Lon	1.10	1.17
NW	1.02	1.04
WM	0.98	1.04
NGN	0.91	0.89
Sc	0.97	0.95
So	1.04	0.98
WWU	0.96	1.00

Applying the benchmark efficiency score

1.73 To determine the benchmark efficiency score, we took each GDN's modelled costs and added back our modelled view of the separately assessed costs. We also added back the pre-modelling adjustments made for regional factors and workload adjustments. We then converted the modelled gross costs to modelled net costs (ie net of customer contributions), based on the ratio of submitted gross costs to submitted net costs in each year.

1.74 We took the modelled costs for each GDN post reversal of adjustments, and multiplied these by the benchmark efficiency score to determine adjusted modelled costs post efficiency challenge ("efficient modelled costs", which exclude ongoing efficiency challenge). This provided efficient modelled costs at the totex level for each GDN in each year of RIIO-GD2.

Disaggregating efficient modelled totex and applying ongoing efficiency assumptions

1.75 The first step we took to disaggregate allowance was to calculate an implied adjustment factor for each GDN by dividing each GDN's efficient modelled costs by the submitted modelled costs (post exclusions and reclassifications).

1.76 We then multiplied the submitted modelled costs (post exclusions, which include volume reductions, and reclassifications) for each disaggregated cost activity by the implied adjustment factor to give totex for adjusted bottom-up modelled cost, post efficiency challenge, including ongoing efficiency and excluding frontier shift and RPEs. This totex number was then disaggregated into different workload activities, using weights derived from company submitted net costs after exclusions and reclassifications. Our approach at Final Determinations ensures that the catch up efficiency challenge is applied evenly to totex, and the disaggregated cost activities reflect the exclusions and reclassifications previously made.

1.77 The next step in the calculation was to apply our frontier shift and ongoing efficiency assumptions to efficient modelled costs and costs assessed via technical assessment to determine overall baseline totex allowances for each GDN. Further details of our decision on ongoing efficiency can be found in the GD Annex.

1.78 For activities associated with specific mechanisms (ie PCDs or volume drivers), we undertook a further level of disaggregation to identify a specific activity-level allowance, consistent with the level required for each mechanism. For activites with a clearly defined cost allowance derived through technical assessment, we have used this cost to set the final allowance. For all other activities included within the totex model, we used weights derived from net submitted costs adjusted for exclusions and reclassifications to allocate allowance to specific activities. For Final Determinations, we have only defined activity allowances for those activities that have a specific mechanism associated with them. For the remaining share of allowances, we have separated these into NARM and non-NARM associated expenditure.

Methodology for calculating unit costs

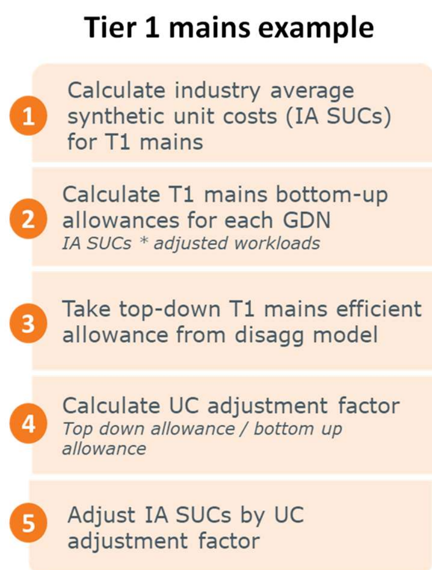
1.79 We need to set unit costs for certain activities that have specific mechanisms associated with them in RIIO-GD2. We have used two methods to set unit cost allowances for specific activities in RIIO-GD2:

- Bottom-up assessment of defined activities
- Derivation from top-down allowances

1.80 Where we have assessed specific activities on a bottom-up basis (eg gas holder demolition), we have used the efficient unit costs from this assessment to set unit costs at Final Determinations.

1.81 For activities where efficient costs are set through the totex model, we have derived unit costs based on allocating the top-down allowance across different activities within a specific cateogry (ie decommissioning different diameter bands for Tier 1 mains). The methodology we have used is outlined in Figure 2.

Figure 2: RIIO-GD2 methodology for calculating unit costs for cost activities with top-down derived allowances



1.82 This was a five step process, as outlined in Figure 2. Step 1: calculate an individual or set of industry average synthetic unit cost(s) (IA SUCs) for the activity. We based these on company submitted data and applied the same outlier tests when calculating synthetic costs as explained in [X.X]. For Tier 2A, we decided at Final Determinations to use synthetic unit costs from RIIO-GD1, as the submitted unit costs for RIIO-GD2 were highly variable, owing to very low workloads, and did not result in logical outturn synthetic costs.

1.83 Step 2: Using the industry average IA SUCs, we calculated a bottom-up allowance for each activity, multiply each IA SUC with the relevant submitted workloads, adjusted for exclusions and reclassifications. The sum of individual activities within a category (ie decommissioning of different mains diameter band within Tier 1) is equal to the full BU allowance for a specific activity (ie total Tier 1 mains).

1.84 Step 3: We calculated the top-down efficient allowance for the activity (ie Tier 1 mains). This was derived from our totex model, based on the disaggregation methodology described in the section above. This allowance is inclusive of benchmarking efficiency challenges and ongoing efficiency, but exclusive of RPEs.

1.85 Step 4: For each GDN we calculated a unit cost adjustment factor by dividing the top-down efficient allowance by the BU allowance for a specific activity (ie Tier 1 mains). This gives a scalar factor which can be applied to scale unit cost.

1.86 Step 5: The GDN-specific scalar factor is multiplied by the IA SUCs for each workload activity. The result is a set of GDN-specific unit costs for each workload activity (ie Tier 1 diameters bands), which are consistent with the top-down efficient allowance for the specific activity (ie total Tier 1 mains). This ensures that unit costs specified within certain mechanisms, such as PCDs, are consistent with top-down efficient totex allowances.

1.87 At Final Determinations we have applied this methodology consistently across mechanisms where we've calculated efficient unit costs from top-down allowances, with the exception of Tier 2A unit costs for SGN. SGN's submitted comparatively low costs and volumes for Tier 2A in both its Scotland and Southern networks. This resulted in very high unit costs for some categories, significantly above the industry average. While we considered these forecast costs as justified at Final Determinations, following review of the engineering information provided, we did not think that they were representative of average unit costs expected from this type of work and may have resulted in perverse incentives for SGN to over-deliver work through the Tier 2A volume driver. Therefore, we have decided at Final Determinations to set SGN's Tier 2A unit costs for Scotland and Southern in line with the industry average[8]. This effectively means that we have assigned a smaller weighting when attributing the share of totex assigned to the Tier 2A allowance for Scotland and Southern, but does not result in a reduction in totex.

[8] With an upwards regional adjustment for Southern.

Appendix A - Methodology for calculating regional labour indices

1.88 We consider that the wage differential between London, the South-East, and the rest of Great Britain still appears to be wide enough to warrant an adjustment in our benchmarking. In line with RIIO-GD1, in our Draft Determinations we proposed to use regional labour indices to make pre-modelling cost adjustments. We have decided to use this approach in our Final Determinations.

1.89 In our Draft Determinations we estimated labour indices using updated BPDT information on the GDNs' Full-Time Equivalents (FTEs) by employment category, ASHE data on regional wages, and ONS population data. We largely followed the same seven-step process used in RIIO-GD1, but with some changes.

1.90 Following Draft Determinations responses and bilaterals with GDNs and undertaking our own analysis, we have further revised our approach to calculating labour indices in our Final Determinations. Table 7 summarises the changes in our approach between RIIO-GD1, RIIO-GD2 Draft Determinations and Final Determinations.

Table 7 Calculating regional labour indices, RIIO-GD1 and RIIO-GD2

Step	RIIO-GD1	RIIO-GD2 Final Determinations
1. Calculate occupational weights	GDNs split their direct and contract labour across 3-digit Standard Occupational Classification (SOC) codes. For each SOC code, we averaged the GDNs' spend relative to total labour spend to obtain an industry average.	No change to our DD position. We calculate industry average occupational weights based on FTEs rather than labour spend. We calculate industry average occupational weights at the 2-digit SOC code level.

2. Calculate regional wage indices	For each administrative region of the UK and occupational category, we calculated the region's mean annual wages relative to the UK mean wage. Then, we averaged these relative wages across occupational categories, using the weights calculated in Step 1, to obtain regional wage indices. This was based on 3-digit SOC wage data from the Annual Survey of Hourly Earnings (ASHE) published by ONS.	First, for each region, we average regional wages across occupational categories using the weights calculated in Step 1. We also do this with UK wages. Then, we divide the regional average wage by the UK average wage to obtain regional wage indices. This is a change to our DD position and better reflects wage differentials across occupational categories. In line with our DD position, we calculate average wages and wage indices at the 2-digit SOC level to reduce uncertainty and missing data in the ASHE wage estimates. We also continue to use gross hourly mean wages (including overtime) rather than annual wages, as these are more robust to regional differences in the number of hours worked.

3. Calculate the wage index for 'Elsewhere'	We calculate the wage index for 'Elsewhere' as the average of the regional wage indices calculated at Step 2 (excluding the London and South-East regions), weighted according to the regions' population.	No change to our DD position. Same approach as RIIO-GD1, except Northern Ireland is excluded from the Elsewhere index as it is not served by any GDN. We also rescale indices so that the Elsewhere index equals 1, meaning that only GDNs operating in London and the South-East will have an adjustment applied, making it easier to detect adjustments.
4. Estimate GDNs' work across the London, South-East and Elsewhere regions	We assumed that GDNs' work was distributed across these three areas in the same proportion as the area's share of the GDN's total population. Two GDNs, London and Southern, have the majority of their operations in London and the South-East, and East of England has a small share of its population in London. All other GDNs operate exclusively in the Elsewhere region.	No change to our DD position. Same approach as RIIO-GD1.

5. Estimate work that should be done locally in the London and South- East regions	To reflect the fact that some work does not need to be carried out locally and can be done in lower cost regions, we adjusted the amount of work done by the GDNs in the London and South-East regions (Step 4) by applying an average percentage of local work across all activities. We assumed that only 40% of Work Management needed to be done locally, whereas the remaining activities were 100% local.	In line with our DD position, we do not apply an average local work percentage across all activities as part of the calculation of the indices. Instead, we apply a specific local work percentage to each cost activity when making the labour adjustments. This makes it unnecessary to calculate separate direct and contract labour indices. We assume that 44% of Work Management and 85% of Training & Apprentices occur locally, whereas the remaining activities are 100% local.
6. Calculate the GDNs' labour indices	For each GDN, the labour index was the average of the regional wage indices for London (Step 2), South-East (Step 2), and Elsewhere (Step 3), weighted by the amount of work that the GDN needs to carry out in each region (Step 5).	No change to our DD position. The labour index is the average of the regional wage indices, weighted by the region's share of the GDN's population.
7. Standardise the labour indices	Lastly, we divided each GDN's labour index by the indices' average and used these standardised indices to make labour cost adjustments for each cost activity.	No change to our DD position. We do not standardise labour indices to avoid losing the benefit of scaling in Step 3.

1.91 In addition to the changes reported in Table 7, our RIIO-GD2 approach at both Draft and Final Determinations differs in the way that the indices are rolled forward to cover years in which historical data is not available. In RIIO-GD1, we calculated the labour indices for 2008-09, 2009-10 and 2010-11, then applied the 2010-11 indices to later years. For our RIIO-GD2 Final Determinations, we have calculated the indices between 2014-15 and 2018-19 and set the indices for later years equal to the average of this period. This approach makes use of the latest information available[9], while continuing to ensure robustness to year-to-year variations in the historical indices.

Proportion of expenditure related to labour

1.92 Calculating the proportion of expenditure that is related to labour and therefore subject to labour adjustments is not necessary to calculate the labour indices, but is required to determine the size of each GDN's labour adjustments (as well as other regional adjustments that apply to labour costs, ie for sparsity and urbanity).

1.93 In RIIO-GD1 and Draft Determinations, we calculated industry average labour ratios based on GDNs' actual expenditure, then adjusted them based on labour and sparsity indices.

1.94 In our RIIO-GD2 Final Determinations we apply industry average labour ratios to all GDNs for each cost activity, but calculate these after adjusting the GDNs' expenditure for labour, sparsity, and urbanity regional factors. This avoids distortions to labour ratios due to the impact of regional factors on GDNs labour spend. Using notional weights also ensures that we do not reward a potentially inefficient company.

Calculating occupational weights and regional wage indices

1.95 The SOC is a common classification of occupational information for the UK. It is a hierarchical structure that categorises jobs in four increasing levels of detail: 1-digit SOC codes indicate nine broad occupational categories which are further broken down into 25 2-digit groups, 90 3-digit groups, and 369 4-digit units.[10]

---

[9] 2020 ONS ASHE data was not published in time for it to be incorporated into the Final Determinations.
[10] Data from the Annual Survey of Hours and Earnings (ASHE) 2014-2019, Table 15.5a. Available here.

1.96 As lower-digit (shorter) groups are aggregates of higher-digit (longer) groups, the decision of which level to adopt presents a trade-off between robustness and granularity. Lower-digit wage estimates refer to more broadly defined occupational categories which may encompass more jobs than those strictly relevant to the GDNs but are based on larger samples and are more reliable than higher-digit estimates.

1.97 GDNs reported FTEs by SOC code at a 3-digit level in their Business Plans, however there were some inconsistencies in reporting across GDNs. For example, some GDNs did not report historical data, and some GDNs classified a large number of FTEs under different SOC codes with similar names. We therefore asked the GDNs to resubmit this data on a consistent basis and clarify any differences in reporting prior to our Draft Determinations.

1.98 At Final Determinations, we have decided to implement the position set out in our Draft Determinations, to use 2-digit SOC codes in our calculation of regional labour indices. This is in line with our approach in RIIO-ED1 and appears to have a stronger statistical basis than using 3-digit SOC codes. Using 2-digit codes also reduces the occurrence of missing data from the ASHE wage estimates.

1.99 Cadent noted in its Business Plan that hourly wages better represent the price of labour compared to annual wages because they are not affected by people in some regions working more hours than in other regions. We agree and have used mean hourly wages to calculate the regional wage indices. This approach is in line with our RIIO-ED1 decision.

1.100 In our Final Determinations, in line with our proposed position at Draft Determinations, we have decided to use industry average occupational weights based on FTEs as a starting point for the calculation of regional wage indices. However, different to Draft Determinations, we average regional wages across occupational categories before calculating the ratio between regional and UK mean wages. This approach ensures that we also take into account the amount paid for different job types.

Regions requiring a labour adjustment

1.101 In RIIO-GD1 we made a labour adjustment for three regions: London, South East, and Elsewhere (ie the rest of Great Britain). As in RIIO-GD1, we have decided on a three-region adjustment in our RIIO-GD2 Final Determinations. Although not as high as London, the South East wages are still systematically higher than the other regions (excluding London) and the national mean.

1.102 At Final Determinations we have also decided to implement our Draft Determinations position and rescale the wage indices so that the Elsewhere index equals 1. This means that only Cadent's London and East of England networks and the SGN Southern network will have an adjustment applied.

Estimating work that should be done locally in the London and South-East regions

1.103 In RIIO-GD1 we assumed that, for most cost activities, all work needed to be done locally. For Work Management (opex), we assumed that 40% of work was done locally. We calculated the overall proportion of work needing to be done locally as the average percentage of local work across various activities, weighted by the activity's proportion of the GDN's spend. This calculation varied between direct and contracted labour, as these had a different mix of the various activities. Therefore, we calculated separate labour indices for direct and contracted labour.

1.104 We note that using an average percentage is reasonable when assessing totex, as all the costs are summed together. But when assessing a specific cost activity, eg repex, which is estimated as being 100% done locally, it is not appropriate to use a labour index that has been calculated to reflect the fact that another activity, ie work management, is only partly done locally.

1.105 To address this inconsistency, we adopted the Draft Determinations position and calculated a single labour index for each GDN and apply this to each activity's labour proportion. In the calculation of the labour adjustment to Work Management, we applied a correction to the labour index to reflect the fact that Work Management is only partly done locally. This approach also made separating direct and contracted labour unnecessary. In addition to Work Management, at Final Determinations we also apply a correction to the labour index for Training and Apprentices, which we assume is only partly done locally.

1.106 In our Draft Determinations, we proposed to update the local work proportion of Work Management to 44%, based on Cadent's submission. Cadent noted that this was calculated as approximately 66% for all GDNs over the period from 2013-14 to 2018-19, reduced by one third to reflect Operations Management costs that are centrally incurred (with the proportion based on actual data for 2018-19).[11] We did not receive any other information to suggest a different local work proportion for Work Management. We have implemented the Draft Determinations proposed posiiton in our Final Determinations.

1.107 For Training and Apprentices, at Final Determinations we have decided to assume a local work proportion of 85%, based on Cadent's Draft Determinations consultation response.

Table 8: RIIO-GD2 regional labour indices (2020-2026)

GDN	Indices
EoE	1.01
Lon	1.18
NW	1.00
WM	1.00
NGN	1.00
Sc	1.00
So	1.10
WWU	1.00

[11] Cadent Business Plan, Appendix 9.21.

Appendix B - Methodology for calculating sparsity indices

1.108 In RIIO-GD1 we implemented a sparsity adjustment to compensate for the productive time lost during the additional time spent on travelling in a sparse area when attending Emergency and Repairs. At Draft Determinations, we consulted on maintaining the same adjustment. For Final Determinations, we have decided to implement the Draft Determinations position and maintain the sparsity adjustment.

1.109 We estimated population density (people per km2) in each Local Authority (LA) in Great Britain using ONS Open Geography Portal data on the LAs' land area and ONS population estimates. We also calculated industry-level density by dividing Great Britain's total population by total area.

1.110 In our Draft Determinations, we proposed to eliminate from the analysis the LAs that were identified in RIIO-GD1 as having no gas network coverage. For our Final Determinations, we have reviewed the list of excluded LAs in light of new information we received from the GDNs in response to Draft Determinations and in light of our own independent analysis. At Final Determinations we only exclude the sparsest LAs, as well as some islands that GDNs have confirmed have no gas network coverage. The list of LAs excluded from the analysis is presented in Table 9.

Table 9: List of excluded LAs.

Local Authority	GDN	People per km^2 (2019)	Included in RIIO-GD1 sparsity calculations	Included in RIIO-GD2 FD sparsity calculations
Na h-Eileanan Siar	Sc	9	Yes	No
Highland	Sc	9	No	No
Argyll and Bute	Sc	12	No	No
Shetland Islands	Sc	16	No	No
Orkney Islands	Sc	23	No	No
Dumfries and Galloway	Sc	23	No	Yes
Scottish Borders	Sc	24	No	Yes
Powys	WWU	26	No	Yes
Perth and Kinross	Sc	29	No	Yes
Isles of Scilly	WWU	136	Yes	No

1.111 As at Draft Determinations, for Final Determinations we calculated the population served by each GDN in each LA using the same assumed split of LAs between GDNs as in RIIO-GD1. This is the same split that was used to allocate GDNs' work to regions in the calculation of labour indices. Most LAs are included entirely in one GDN's service area, but there are a number of cases where we split a LA's population between two GDNs.

1.112 We classified LAs with a population density lower than the industry density as sparse and calculated their sparsity indices as the ratio between the LA's density and the industry density. We normalised these indices by converting them into deviations from 1.

1.113 We calculated GDNs' unstandardised sparsity indices as the average of LAs' sparsity indices, weighted by the LA's proportion of the GDN's total population. Our approach at Final Determinations is the same as at Draft Determinations, and differs from RIIO-GD1 in the way the indices are rolled forward to cover years in which historical data is not available. In RIIO-GD1, we calculated the sparsity indices for one year only, based on 2010 population data, then applied the same indices to later years. In our Final Determinations, we have calculated the indices using historical data until 2018-19 and set indices for later years equal to the 2014-15 – 2018-19 average.

1.114 At Final Determinations, we have implemented the Draft Determinations approach and calculated the ratio between the GDNs' unstandardised indices and the unstandardised index for WWU. We then multiply these ratios by the percentage sparsity adjustment applied to WWU's Emergency and Repair labour costs (-13%). This means that each GDN receives a sparsity adjustment that reflects its sparsity relative to WWU's. All the sparsity adjustments will be negative (ie a reduction to modelled costs) aside from the Cadent London network, which does not receive an adjustment as it is considered to have no sparse areas in its network. For consistency with the other regional factors, we converted the percentage adjustments into standardised sparsity indices, which are presented in Table 10 below.

Table 10: RIIO-GD2 sparsity indices (2020-2026)

GDN	Indices
EoE	1.08
Lon	1.00
NW	1.01
WM	1.06
NGN	1.10
Sc	1.15
So	1.04
WWU	1.15

Table 11: RIIO-GD2 regional factors, by cost activity

Cost activity	Regional labour	Sparsity	Urbanity reinstatement	Urbanity productivity
Work Management	Yes	No	No	No
Emergency	Yes	Yes	Yes	Yes
Repairs	Yes	Yes	Yes	No
Maintenance	Yes	No	Yes	No
Other Direct Activities	Yes	No	Yes	No
Business Support	No	No	No	No
Training and Apprentices	Yes	No	No	No
LTS Pipelines, Storage & Entry	Yes	No	No	No
Connections	Yes	No	No	Yes
Reinforcement	Yes	No	No	Yes
Governors	Yes	No	No	No
Transport & Plant	No	No	No	No
Other Capex	Yes	No	No	No
Repex	Yes	No	No[12]	Yes

[12] We apply a separate reinstatement adjustment to repex costs. See chapter three of the GD Annex, as well as Cadent and SGN Annexes for more details.

Appendix C – List of repex and capex categories for synthetic unit costs calculation

1.115 Table 12 and 13 list the repex and capex categories for which we derived synthetic unit costs.

Table 12 Repex synthetic unit costs

Repex asset	Diameter/Type	Scaling factor?
Mains		
Tier 1 iron	<75mm	No
Tier 1 iron	75mm to 125mm	No
Tier 1 iron	125mm to 180mm	No
Tier 1 iron	180mm to 250mm	No
Tier 1 iron	250mm to 355mm	No
Tier 1 iron	355mm to 500mm	No
Tier 1 iron	500mm to 630mm	Yes
Tier 1 iron	>630mm	Yes
Tier 1 Steel	<75mm	No
Tier 1 Steel	75mm to 125mm	No
Tier 1 Steel	125mm to 180mm	No
Tier 1 Steel	180mm to 250mm	Yes
Tier 1 Steel	250mm to 355mm	Yes
Tier 1 Steel	355mm to 500mm	Yes
Tier 1 Steel	500mm to 630mm	Yes
Tier 1 Steel	>630mm	Yes
Tier 2	<75mm	No
Tier 2	75mm to 125mm	No
Tier 2	125mm to 180mm	No
Tier 2	180mm to 250mm	No
Tier 2	250mm to 355mm	No
Tier 2	355mm to 500mm	No
Tier 2	500mm to 630mm	No
Tier 2	>630mm	Yes
Tier 3	<75mm	No
Tier 3	75mm to 125mm	No
Tier 3	125mm to 180mm	No

Tier 3	180mm to 250mm	No
Tier 3	250mm to 355mm	No
Tier 3	355mm to 500mm	No
Tier 3	500mm to 630mm	Yes
Tier 3	>630mm	Yes
Iron >30m	<75mm	No
Iron >30m	75mm to 125mm	No
Iron >30m	125mm to 180mm	No
Iron >30m	180mm to 250mm	No
Iron >30m	250mm to 355mm	No
Iron >30m	355mm to 500mm	No
Iron >30m	500mm to 630mm	No
Iron >30m	>630mm	Yes
Steel>2”	<75mm	No
Steel>2”	75mm to 125mm	No
Steel>2”	125mm to 180mm	No
Steel>2”	180mm to 250mm	No
Steel>2”	250mm to 355mm	No
Steel>2”	355mm to 500mm	No
Steel>2”	500mm to 630mm	No
Steel>2”	>630mm	Yes
Other Policy and Condition	<75mm	No
Other Policy and Condition	75mm to 125mm	No
Other Policy and Condition	125mm to 180mm	No
Other Policy and Condition	180mm to 250mm	No
Other Policy and Condition	250mm to 355mm	Yes
Other Policy and Condition	355mm to 500mm	No
Other Policy and Condition	500mm to 630mm	No
Other Policy and Condition	>630mm	Yes
Services		
Relay associated with mains replacement	Domestic	No
Relay associated with mains replacement	Non-domestic	Yes
Transfer associated with mains replacement	Domestic	No
Transfer associated with mains replacement	Non-domestic	No
Relay not associated with mains replacement	After escape	No
Relay not associated with mains replacement	Other	No

Table 13 Capex synthetic unit costs

Asset type	Diameter	Scaling factor?
Mains reinforcement		
General and Specific	<180mm	No
General and Specific	>180mm	No
Connections		
New and Existing Housing, Domestic and Non-Domestic Mains	<180mm	No
New and Existing Housing, Domestic and Non-Domestic Mains	>180mm	No
Services	-	No
FPNES Mains	<180mm	No
FPNES Mains	>180mm	Yes
FPNES Services	-	No