Assessment of Sumicsid's TCB18 response

Prepared for
TenneT TSO B.V.

14 May 2021

Strictly confidential

www.oxera.com

oxera
compelling economics

Contents

Figures and tables

Oxera Consulting LLP is a limited liability partnership registered in England no. OC392464, registered office: Park Central, 40/41 Park End Street, Oxford OX1 1JD, UK; in Belgium, no. 0651 990 151, branch office: Avenue Louise 81, 1050 Brussels, Belgium; and in Italy, REA no. RM - 1530473, branch office: Via delle Quattro Fontane 15, 00184 Rome, Italy. Oxera Consulting (France) LLP, a French branch, registered office: 60 Avenue Charles de Gaulle, CS 60016, 92573 Neuilly-sur-Seine, France and registered in Nanterre, RCS no. 844 900 407 00025. Oxera Consulting (Netherlands) LLP, a Dutch branch, registered office: Strawinskylaan 3051, 1077 ZX Amsterdam, The Netherlands and registered in Amsterdam, KvK no. 72446218. Oxera Consulting GmbH is registered in Germany, no. HRB 148781 B (Local Court of Charlottenburg), registered office: Rahel-Hirsch-Straße 10, Berlin 10557, Germany.

Executive Summary

The pan-European Transmission Cost Benchmarking Project (TCB18) was carried out by the Council of European Energy Regulators through its consultant, Sumicsid. Sumicsid used data envelopment analysis (DEA) to estimate the relative efficiency of the participating TSOs.

From a study commissioned by all the participating TSOs, we concluded that TCB18 suffers from a **number of significant flaws**, some of which are **fundamental**. Sumicsid's response to our report has **not** alleviated these concerns. Instead, Sumicsid's response provides further evidence to support the concerns and conclusions of our original study, including that:

- the TCB18 process and the published material are opaque, such that an independent assessor of the study could not clearly follow what Sumicsid had done, thus casting doubts on the benchmarking exercise;
- there was significant uncertainty in the data used by Sumicsid—the raw dataset contained several material errors, the uncertainty of which was compounded by numerous adjustments to the data;
- the model development procedure was arbitrary and not supported by robust conceptual or empirical evidence;
- the final TCB18 model was not robustly validated through appropriate sense checks.

At different times, Sumicsid now directly contradicts its statements from the TCB18 report, arbitrarily shifting modelling criteria and rationale. Moreover, Sumicsid misrepresents our arguments and those of academic articles in order to justify its position. Concerningly, Sumicsid appears to resist any reasonable margin of error at any stage and does not fully engage with the pragmatic solutions presented in our report. Therefore, we maintain that the flaws covered in our main report and associated work demonstrate that the estimated efficiency scores and suggested cost savings for the TSOs **are not robust** and **cannot be used in their current form for regulatory, operational or valuation purposes**.

The pan-European Transmission Cost Benchmarking Project (TCB18)[1] was carried out by the Council of European Energy Regulators (CEER) through its consultant, Sumicsid. Sumicsid used cost and asset data provided by the 17 participating electricity Transmission System Operators (TSOs), in addition to environmental and input price data from external sources, to estimate the relative efficiency of TSOs through data envelopment analysis (DEA).

A consortium of all the TSOs that participated in TCB18 commissioned Oxera to validate and review the results from TCB18. The overall conclusion of our assessment ('our main report'),[2] which was published in April 2020, was that the TSOs' estimated efficiency scores from the study cannot be used for regulatory, operational or valuation purposes in their current form. Our main report concluded with 15 recommendations regarding the procedural and

[1] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July.
[2] Oxera (2020), 'A critical assessment of TCB18 electricity', April.

methodological issues in TCB18 to improve the robustness of such international benchmarking exercises.

In addition to our main report, which provided a general assessment that was relevant to all TSOs, several participating TSOs, including Statnett SF,[3] SHE-Transmission,[4] and TenneT TSO B.V. ('TenneT'),[5] commissioned Oxera to assess their individual position in TCB18. These projects examined specific issues with TCB18 that had a material impact on the various TSOs.

The authors of TCB18 (Professors Per Agrell and Peter Bogetoft for Sumicsid) has responded to our main report in October 2020 ('the TCB18 response').[6] TenneT has commissioned Oxera to review and assess the arguments brought forward by Sumicsid in its TCB18 response, and to determine whether they alter any of the conclusions of our main report or the assessment of TenneT's individual efficiency.

Sumicsid's TCB18 response broadly follows the structure of our main report. Therefore, we continue to group our concerns with TCB18 and the TCB18 response into four themes, as follows.

- **Transparency and engagement process**—the TCB18 response **directly contradicts** the TCB18 report and its associated appendices repeatedly. For instance, Sumicsid changes the stated purpose of the second-stage analysis,[7] and its justification for the returns-to-scale assumption.[8] Instead of improving third parties' understanding of the TCB18 project, the TCB18 response compounds the confusion through inconsistent and selective arguments.
- **Data collection and validation**—one of the conclusions of our main report is that the data is measured with significant uncertainty. **This remains valid and is supported by the TCB18 response**. Even while acknowledging the existence of several data errors in the TCB18 dataset, such as an error that understated the weighted lines variable by 15% on average, Sumicsid maintains that the dataset is of high quality and that no adjustment for statistical noise is required.[9] Similarly, while acknowledging that there is considerable uncertainty when making (or not making) adjustments to the raw TSO data to capture operational features, Sumicsid still does not consider that this uncertainty needs to be accounted for when estimating efficiency scores. As such, there is a **clear contradiction in Sumicsid's position in acknowledging errors and uncertainty on the one hand but being unwilling to make robust allowances for these** in the final outputs on the other.
- **Model development—**Sumicsid shifts its rationale and methodology on model development substantially between the TCB18 report and its TCB18 response such that **large parts of the original documentation are no longer relevant**. For instance, the regression results presented in the individual reports are now uninformative or irrelevant.[10] Indeed, it is no longer clear what (or if) model development took place in TCB18, nor is it clear how Sumicsid selected and constructed the cost drivers included in

[3] Oxera (2019), 'Impact of Price Level adjustment in TSO Benchmarking', December.
[4] Oxera (2019), 'Scottish Hydro Electric Transmission's cost assessment', December.
[5] Oxera (2020), 'Analysis of TenneT's estimated efficiency under TCB18', August.
[6] Sumicsid (2020), 'Response to the Oxera Report on TCB18 ETSO', October—hereafter 'Sumicsid response (2020)'.
[7] See section 2 and Table 1.
[8] See section 5.1.
[9] See section 3.1.
[10] For example, see Sumicsid response (2020), para. 3.70.

the final model. The lack of clear documentation of the methodology in TCB18 highlighted in our main report[11] is affirmed by Sumicsid's shifting stance on the process and approach followed in determining the model specification. This shifting stance on several fundamental aspects of TCB18 casts doubt on the entire benchmarking exercise, as the authors themselves appear to be unclear or inconsistent as to what process was followed.

- **Model assumptions and output validation**—Sumicsid has not justified the assumptions that it has made in its final model. **The additional evidence presented in the TCB18 response is either unconvincing or in contradiction to Sumicsid's earlier position**. For example, Sumicsid now cites the academic literature in support of its returns-to-scale assumption, yet most of the papers cited by Sumicsid directly contradict Sumicsid's approach.[12] Similarly, Sumicsid cites an academic paper stating that negative frontier shift is possible, yet the paper itself estimates positive and at times sometimes statistically insignificant frontier shift.

These four themes are discussed in more detail below.

The TCB18 response directly contradicts the TCB18 main report and its associated appendices repeatedly. Instead of improving third parties' understanding of the TCB18 project, the TCB18 response compounds the confusion and exacerbates our earlier concerns.

Our main report presented several methods of achieving transparency while maintaining confidentiality (such as through the use of anonymised code or data rooms),[13] yet Sumicsid does not engage with these possibilities in its TCB18 response. Sumicsid mentions the issue of transparency **only once** in the TCB18 response. Its stated rationale for omitting documentation on model development was 'data confidentiality'.[14]

In addition, the TCB18 response directly contradicts the main outputs of TCB18 and its associated appendices on several occasions. The contradictions not only add to the lack of transparency and confusion for an independent assessor of the study, but also cast doubt on the entire benchmarking exercise, as the authors themselves do not appear to be clear or consistent about what the process was. Some of the contradictions are documented in the table below.

[11] For example, see Oxera (2020), 'A critical assessment of TCB18 electricity', April, p. 8.
[12] See section 5.1.
[13] Oxera (2020), 'A critical assessment of TCB18 electricity', April, pp. 81 and 83–4.
[14] Sumicsid response (2020), para. 3.65

Table 1 Material inconsistencies and contradictions between the TCB18 main report and the TCB18 response

Issue	Sumicsid's position in TCB18	Sumicsid's position in the TCB18 response	Material inconsistency or direct contradictions
Criteria for cost driver selection	'[…] [W]e have in this project focused on the following generic criteria: Exogeneity – Output and structural parameters should ideally be exogenous, i.e. outside the influence of the TSOs. Completeness – The output and structural parameters should ideally cover the tasks of the TSOs under consideration as completely as reasonable. Operability – The parameters used must be clearly defined and they should be measurable or quantifiable. Non-Redundancy – The parameters should be reduced to the essential aspects, thus avoiding duplication and effects of statistical multi-collinearity.'[15]	'[R]egulatory benchmarking primarily should be based on a specification that is **deterministic, complete, feasible and verifiable**'[16] [emphasis added]	Sumicsid's previous position describes relatively standard criteria for variable selection in efficiency benchmarking. The TCB18 response retains two criteria: completeness and operability (the latter renamed as 'feasibility'). The two other criteria—exogeneity and non-redundancy—are no longer covered, but replaced by the new criteria: being 'deterministic' and 'verifiable'. In fact, Sumicsid now states that non-redundancy is not an issue.[17] While neither of the new criteria, 'deterministic' or 'verifiable', are defined in the TCB18 response, their use by Sumicsid appears to be in conflict with the previous criteria. For instance, it is not clear how a cost driver can be 'deterministic' from the TSOs' perspective while also being exogenous. Exogenous factors always contain an element of randomness and are not entirely deterministic.
Aggregation of NormGrid	'The calibration of the [NormGrid] asset weight **systems is made through linear regression** towards the **Capex and Opex data obtained in the project**'[18] [emphasis added].	'The NormGrid is **not to be estimated from the TCB18 data**, but from engineering expertise'[19] [emphasis added].	Sumicsid previously stated that the calibration of NormGrid uses linear regression and expenditure data from TCB18.[20] It now states that no estimation on the data took place.[21] A calibration towards CAPEX and OPEX data through linear regression **is** an estimation of components of NormGrid on TCB18 data. We have here a direct contradiction from Sumicsid.
Support for the RTS assumption	'For all possible model specifications, we have also tested which of the returns to scale assumptions in the DEA model **fit data the best** […]. We have done so **using F-tests based on a goodness-of-fit measure** as explained in the Method chapter. The general finding is that the IRS assumption […] is the best assumption to invoke. This is **supported also by parametric analyses for a logarithmic model**, where the coefficients sum to less than one for the selected parameters'[22] [emphasis added].	'The assumption for NDRS is based on **techno-economic arguments** linked to the abundant **evidence from other studies**, previous transmission studies, concession areas and the model specification (NormGrid covers all relevant assets)'[23] [emphasis added].	The TCB18 main report made no mention of academic references or precedent supporting the RTS assumption. Yet in the TCB18 response this is now said to be the key input into the decision. The empirical evidence on the current model is still not present. Note that the new statement is also not backed up with robust evidence.

Issue	Sumicsid's position in TCB18	Sumicsid's position in the TCB18 response	Material inconsistency or direct contradictions
Purpose of the second stage analysis	'In order to **investigate whether some potentially relevant variables have been omitted** in the final model specification, a so-called second stage analysis has been performed.'[24] 'Extensive second-stage analyses shall be undertaken to **see if any of the non-included variables should be included**.'[25]	'The post-run second-stage process is intended to detect potential bias in the scores, **not the inclusion of specific parameters**.'[26]	This is a direct contradiction. The purpose of the analysis is stated on multiple occasions without ambiguity as testing for the *inclusion* of specific parameters. The TCB18 response now explicitly states that this was **not** the purpose.

[15] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, para. 4.83.
[16] Sumicsid response (2020), para. 3.81.
[17] Ibid., para. 3.70.
[18] Sumicsid (2019), 'Norm Grid Development – Technical report', p. 2.
[19] Sumicsid response (2020), para. 3.86.
[20] Sumicsid (2019), 'Norm Grid Development – Technical report', p. 2.
[21] Sumicsid response (2020), para. 3.86.
[22] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July. 5.11.
[23] Sumicsid response (2020), para. 3.102.
[24] Sumicsid (2019), 'Project TCB18: Individual Benchmarking Report Fingrid – 131', p. 35.
[25] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July. 4.09.
[26] Sumicsid response (2020), para. 3.127.

As well as these direct contradictions, we also discovered several *misrepresentations* of both our own study and the wider academic literature when reviewing Sumicsid's TCB18 response. For example, Sumicsid uses Yatchew (2000) to justify its returns-to-scale assumption, claiming that Yatchew (2000) finds increasing returns to scale for smaller utilities.[27] However, the full quote from the abstract of the paper is:[28]

> The data reveal substantial evidence of increasing returns to scale with minimum efficient scale being achieved by firms with about 20,000 customers. **Larger firms exhibit constant or decreasing returns**. [emphasis added]

That is, Yatchew (2000) **does not support** Sumicsid's stated non-decreasing returns-to-scale assumption. Similar observations hold for most of the papers that Sumicsid quotes to support its returns-to-scale assumption.

Similarly, our main report argued that performing model development in levels (as opposed to logarithms) could be inappropriate—logarithmic models are used more widely in regulatory and academic contexts because they are better able to account for heteroskedasticity and lead to coefficient estimates that are readily interpretable.[29] In its TCB18 response, Sumicsid states:

> Contrary to the claim of Oxera (2020), the practice of non-parametric benchmarking using levels is very common for regulation.[30]

Oxera **never stated** that it was uncommon to estimate the *non-parametric model in levels*. Indeed, all of the DEA models presented in our main report were estimated using data in levels. Oxera suggested the logarithmic functional form *for model development*, which *is* the most common form and has been applied by Sumicsid in previous studies.[31] Sumicsid instrumentalises this misinterpretation to dismiss valid criticism of its cost driver analysis. Had Sumicsid followed a more robust approach to model development, a different set of cost drivers could have been chosen, which may have materially affected the TSOs' estimated efficiencies.

Misrepresentations like these are common in Sumicsid's TCB18 response, and we address them throughout this report.

Despite acknowledging the existence of several data errors in the TCB18 dataset, no adjustment for statistical noise is made. Similarly, Sumicsid has now acknowledged that there is considerable uncertainty in adjustments to the raw data, yet this uncertainty is not accounted for when estimating efficiency scores.

Data errors

Real-world data is inherently noisy. Due to a variety of factors such as measurement error, misreporting and miscommunication, the data entering a dataset can contain material errors. Moreover, errors can be further introduced by the consultant when processing the data.[32] TCB18 appeared to employ a standard data gathering and processing methodology in principle, and this was

[27] Sumicsid response (2020), para. 3.102.
[28] Yatchew (2000), 'SCALE ECONOMIES IN ELECTRICITY DISTRIBUTION: A SEMIPARAMETRIC ANALYSIS', p. 188.
[29] Oxera (2020), 'A critical assessment of TCB18 electricity', April, section 4.1.
[30] Sumicsid response (2020), para. 3.59.
[31] Sumicsid and Swiss Economics (2016), 'Benchmarking European Gas Transmission System Operators FINAL REPORT', June, p. 45.
[32] These are discussed in the subsequent sections.

reiterated by Sumicsid in its TCB18 response. However, in our main report, we document several significant errors in the final TCB18 data that biased the estimated efficiency scores.[33]

In TCB18, Sumicsid did not consider the impact of data errors on the TSOs' estimated efficiencies. In our main report, we noted that this was a material omission in the TCB18 study, because DEA (as applied by Sumicsid, and by its own admission)[34] is highly sensitive to data errors. We proposed a Monte Carlo simulation to determine the impact of these errors and other noise in the data on the estimated results. Specifically, we simulated a +/- 10% error margin to all cost and cost driver data, noting that the data issues that were identified in the main report were materially higher than this margin.

Sumicsid has now acknowledged the existence of several of these data errors. For example, it has acknowledged a material data processing error, which **affected all** weighted lines data for 2013–16,[35] resulting in the variable being understated by 15% on average[36] across the entire sample. Another error that has been acknowledged meant that one TSO was mistakenly not assigned any 'weighted lines'. Despite its reversal of views, Sumicsid maintains that the TCB18 data is of a 'very high quality'.[37] Sumicsid also suggests that Oxera's proposed component of noise is excessive, ignoring the fact that the error that it has belatedly identified on weighted lines is significantly larger in magnitude than our proposed margin.[38] We presented several other data errors where the magnitude was up to 30%,[39] again significantly higher than our proposed margin.[40]

In light of multiple statements from TSOs highlighting the difficulty of providing the data required,[41] several known (and some publicly acknowledged) data errors in the final TCB18 data,[42] and the publication of the additional data processing error in the TCB18 data by Sumicsid,[43] the random component that we considered in our main report remains appropriate and conservative. In any case, once these known errors are corrected, the entire methodology (i.e. data processing, model development, efficiency analysis and output validation) needs to be repeated and the results shared with the participants.

Defining the input variable

Sumicsid uses a one-input total expenditure (TOTEX) model in TCB18, thereby assuming a strict one-to-one trade-off between *benchmarked operating expenditure* (OPEX) and *benchmarked capital expenditure* (CAPEX). In our

[33] Oxera (2020), 'A critical assessment of TCB18 electricity', April, section 3.1.
[34] Sumicsid response (2020), para. 3.13.
[35] Ibid., para. 3.50.
[36] We do not have access to Sumicsid's original (erroneous) data, so we cannot speculate as to the magnitude of the error for any specific TSO. Nonetheless, an error of 15% is already above the 10% margin assumed in our Monte Carlo analysis.
[37] Sumicsid response (2020), para. 3.12.
[38] Ibid., para. 3.11.
[39] For example, an error discovered in a TSO's CAPEX overstates its TOTEX by 32%. See Oxera (2020), 'A critical assessment of TCB18 electricity', April, p. 35.
[40] The ACM (without reference to supporting evidence) claims that all data errors raised have been addressed (see ACM (2021), 'Methodebesluit Transporttaken TenneT 2022-2026', April, para. 332). However, we would expect the correction of such a material issue to result in a correction to the TCB18 report as it is published, including relevant discussion on the impact of the error on the cost driver analysis, outlier procedure, estimation of efficiencies, and validation of outputs. To our knowledge, a corrigendum to the TCB18 report has not been issued, and nor have the affected TSOs been notified. Regardless of whether such corrections have taken place, the underlying point is that the data errors that have been discovered are indicative of further undiscovered data errors, as the TSO and NRA level of engagement with TCB18 varied greatly.
[41] For example, see SHE-Transmission (2018), 'CEER financial reporting notes', October.
[42] Sumicsid response (2020), para. 3.09.
[43] Ibid., para. 3.49.

main report, we highlighted that this is unnecessarily restrictive and possibly inappropriate as the OPEX/CAPEX input mix may not be:

- **controllable**—TSOs cannot choose freely between OPEX and CAPEX solutions, for instance, due to national laws determining ownership of the grid. Cost reduction targets based on a reallocation of OPEX and CAPEX are not achievable in this context;
- **equivalent**—it is unlikely that one unit of benchmarked OPEX leads to the same outcome as one unit of benchmarked CAPEX, due to the different operating rules and data normalisations applied.[44]

Sumicsid's TCB18 response does not engage with these principled arguments. Instead, Sumicsid justifies its choice by referring to regulatory precedent of TOTEX regimes.[45] However, as highlighted in our main report, there exist several options for TOTEX benchmarking that do not rely on the strict one-to-one trade-off between OPEX and CAPEX imposed by Sumicsid, which are also employed in regulatory benchmarking, for instance, in the UK energy sector.[46]

Adjusting for differences in input prices

Sumicsid only adjusts 5.9% of the cost base for differences in input prices in TCB18.[47] This risks conflating uncontrollable price differences with inefficiency. In its TCB18 response, Sumicsid recognises that the correction for input prices can materially impact the modelling results, and that the question of *how* input prices are accounted for cannot be definitively answered as no perfect price index exists.[48] However, Sumicsid fails to acknowledge that *not* correcting for input prices amounts to an extreme assumption that does not resolve the issue of uncertainty. Moreover, Sumicsid made no attempt to model (or even acknowledge) this uncertainty in TCB18.

Sumicsid is also incorrect when it states that other studies did not correct for price levels. In fact, one of the studies referenced in the TCB18 response does adjust *all* costs with a price level index, as also indicated in our main report.[49]

Nonetheless, we welcome that Sumicsid appears to acknowledge some of the issues with its correction for input prices in the TCB18 model, as evidenced by the fact that it on multiple occasions refers the ultimate responsibility for price level adjustments to the national regulatory authorities (NRAs).[50]

Sumicsid shifts its rationale and methodology on model development substantially between the TCB18 report and its TCB18 response, such that large parts of the original documentation are no longer relevant.

Cost driver analysis

Operating, maintaining and enhancing the capabilities of an electricity transmission network is an extremely complex operation, and finding a set of

[44] Oxera (2020), 'A critical assessment of TCB18 electricity', April, section 3.2.
[45] Sumicsid response (2020), para. 3.15.
[46] Oxera (2020), 'A critical assessment of TCB18 electricity', April, p. 36.
[47] Ibid., April, section 3.3.
[48] Sumicsid (2020), 'Response to the Oxera Report on TCB18 ETSO', para. 3.33.
[49] Jamasb, T., Newbery, D. and Pollitt, M. (2007), 'International Benchmarking and Regulation of European Gas Transmission Utilities', p. 27.
[50] For example, see Sumicsid (2020), 'Response to the Oxera Report on TCB18 ETSO', paras 3.28 and 3.31.

cost drivers that can completely describe the functions of a TSO is a difficult task. Therefore, a robust model development process must be in place to ensure that the results of the empirical analysis are robust.

The TCB18 report stated that a combination of econometric techniques and operational insight was used to develop the TCB18 benchmarking model. Specifically, the report stated that the steps taken in model development are as follows:[51]

> 1. Definition of parameter candidates. In a first step we established **a list of parameter candidates** which **may have an impact** on the costs of TSOs.[…]
> 2. Statistical analysis of parameter candidates. […] The main advantage of statistical analysis is that it allows us to e**xplore a large number of candidate parameters** and to **evaluate how they individually and in combination allow us to explain as much as possible of the cost variation**.
> 3. Plausibility checks of final parameters. […][emphasis added]

In our main report, we argued that Sumicsid's stated cost driver analysis procedure is flawed, because it had relied on inappropriate statistical models and tests, and had provided insufficient evidence that its final model was materially better than alternatives. In addition, the model outputs were not properly validated. In response, Sumicsid has changed its characterisation of the model-development procedure that it had followed in TCB18. Sumicsid now states that:[52]

> Whereas the reasoning in Oxera (2020) seems to suggest a wide range of different factors of unknown influence to be investigated, the **structure of a good cost model can be derived from some simple principles**, proven both from engineering practice, reference network analysis and transmission system benchmarking.

That is, Sumicsid no longer appears to consider statistical analysis to be a critical component in the model development procedure. We consider that this change in argumentation between the publication of the main report and the TCB18 response is symptomatic of wider transparency and opacity concerns—either Sumicsid has never considered statistical analysis to be important in model development (in which case its final report and workshop slides are misleading) or Sumicsid has somehow changed its preferred approach between the TCB18 report and its response (in which case the model developed under TCB18 is unlikely to be robust according to Sumicsid's current model development criteria). As a result of this inconsistency, much of the material presented by Sumicsid in the TCB18 outputs is no longer relevant.

Despite Sumicsid's change in stance, it still challenged the conceptual and empirical critiques outlined in Oxera's main report. Therefore, we cover Sumicsid's response to these critiques in this report.

Sumicsid has made several factual errors in its response and has misrepresented Oxera's main report. For example, Oxera's main report challenged Sumicsid's decision to suppress the intercept when estimating its model as it was inconsistent with Sumicsid's returns to scale assumption.[53]

---

[51] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators—Main report', July, para. 4.85.
[52] Sumicsid response (2020), para. 2.13.
[53] Supressing the intercept (i.e. setting the intercept to zero) implies that a TSO's costs are zero when all of the cost drivers are zero. In other words, it assumes that there are no 'fixed costs' to production and (as the

Sumicsid responded that 'Contrary to the claim in Oxera, the model specification reported at Workshop 4 […] show [sic] regression results including intercept for all steps of the development'.[54] Sumicsid's reference to the models in Workshop 4 is irrelevant to this discussion—the individual reports for TSOs, which were published alongside the main TCB18 report, *clearly show that a regression without an intercept was used*, so our criticism remains valid.

In addition, there are conceptual errors in Sumicsid's TCB18 response. For instance, it states that:[55]

> [T]he models developed for validation of the minimal necessary size using Lasso regression were presented at Workshop 3, but these were not essential to the further developments of the model.

We note that Lasso cannot be used to make a *general* observation about optimum model size. Rather, Lasso can be used to narrow down a *specific* set of cost drivers to a smaller subset of the *specific* set of drivers *on the specific dataset* based on a specific criterion. Thus, Lasso cannot be used to assess whether the final TCB18 model is of 'optimal size', as both the cost drivers (e.g. weighted lines was not in the set of possible cost drivers in Workshop 3) and the dataset itself have changed substantially since Workshop 3 (e.g. through the correction of data errors, section 3.1). If Sumicsid has repeated the Lasso regression on the new dataset (and as noted earlier, the entire methodology), it would be helpful to share this analysis with the project participants to assess whether it supports the TCB18 model.

Sensitivity to the sample selected

In our main report, we argued that Sumicsid had not undertaken sufficient validation of its model specification using appropriate statistical tests or alternative methods. For example, we find that the estimated model is highly sensitive to the inclusion of specific TSOs, indicating that a few unusual TSOs are unduly skewing the model specification.[56] Moreover, the specification and estimated efficiency scores were not stable with respect to the time period of analysis.

Sumicsid claims in its TCB18 response that this robustness check is irrelevant as the TCB18 sample is 'complete' (i.e. not a subset of a larger population), and that the model aims to establish an efficiency estimate for 2017 only.[57] According to Sumicsid, this avoids the need to test the sensitivity of the final model to the sample selected. Sumicsid's characterisation of the TCB18 sample as 'complete' is **misleading and incorrect** for the following reasons.

- The set of participants is only a subset of a larger population (all European TSOs, or all TSOs including non-European TSOs).
- 2017 is one year of many in which the TSOs have operated.

In fact, the Dutch authority for consumers and markets (ACM, a member of the project steering group) is proposing to use the efficiency information from 2017 to inform cost reduction targets in 2022–26—more than half a decade after the year on which efficiency scores are estimated. On the assumption that the

cost drivers are modelled linearly) there are constant returns to scale. This is inconsistent with Sumicsid's non-decreasing returns to scale assumption that it applied in the DEA model.

[54] Sumicsid response (2020), para. 3.57.
[55] Sumicsid response (2020), para. 3.57.
[56] Oxera (2020), 'A critical assessment of TCB18 electricity', April, section 4.2.
[57] Sumicsid response (2020), para. 3.54.

TCB18 model is only a valid tool for estimating a TSO's efficiency in 2017, and that it is reasonable for a TSO's efficiency to vary significantly from year-to-year as was stated by Sumicsid,[58] the TCB18 model cannot be used as evidence for TenneT's (or any other TSOs') current or future efficiency levels. Therefore, under Sumicsid's own assertions, a NRA cannot reasonably use TCB18 to set regulated revenues for a TSO, as ACM is suggesting.

Thus, sensitivities to the sample selected is a critical robustness check to validate the conclusions from the study before they are applied in a regulatory context. Especially considering that the data contains material errors (see section 3.1), the model development approach is flawed (see section 4), DEA is deterministic and sensitive to errors and no validation of results has been undertaken (e.g. using SFA or alternative models).

Selecting candidate cost drivers

In TCB18, Sumicsid stated that the asset-based measures it used as cost drivers are highly correlated with TSO cost drivers, such as network capacity and routing complexity. However, as highlighted in our main report, these statements remain unsubstantiated and valid alternatives were not considered. More generally, there are advantages and disadvantages to using asset-based variables as cost drivers,[59] and there is a need to ensure that the results from any one model is corroborated and cross-checked against alternative specifications. In the case of TCB18, alternative specifications could include 'pure outputs' (such as load by grid level), as opposed to the asset-based measures of capacity.

In the TCB18 response, Sumicsid does not address the general concern with focusing on asset-based cost drivers in benchmarking or the lack of validation on the data. Instead, the TCB18 response appears to alter the criteria for cost driver selection (as highlighted in Table 1).

As noted, the cost drivers previously had to be complete, non-redundant, feasible and exogenous.[60] Sumicsid's TCB18 response now states that it will limit the set of cost drivers to include only 'deterministic' variables.[61] While this term is not defined anywhere in the TCB18 outputs, it appears to imply that it precludes any non-asset based measures. Under this definition, many drivers of TSOs' expenditure, such as faults, network availability and peak load *are* stochastic in nature. That is, some drivers of TSOs' expenditure will be driven, at least in part, by volatile events. Therefore, when these drivers are excluded based on this newly introduced criterion, there is a material risk that random variation in costs (that is, variation that is not explained by the deterministic drivers included in the model) is conflated with inefficiency.

Aggregation of NormGrid

TCB18 used constructed variables such as Normalised Grid (NormGrid), which reflect an aggregation of a number of classes of assets using weights that are themselves estimated with a degree of uncertainty. Our main report noted that in this context, it may be more appropriate to consider each asset class as a

[58] Sumicsid response (2020), para. 3.71.
[59] In our main report, we noted that the use of asset-based measures may skew the model towards companies that deploy asset-based solutions. See Oxera (2020), 'A critical assessment of TCB18 electricity', April, section 4.3.
[60] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, para. 4.83.
[61] Sumicsid response (2020), para. 3.81.

separate cost driver and to allow the DEA model[62] to determine the correct weights on each asset class or otherwise include a margin for error.[63]

In the TCB18 response, Sumicsid focuses its critique on hypothesised 'absurd dual weights' in the model that can control for the main asset types separately.[64] Sumicsid's proposal that dual weights must be carefully examined for operational intuitiveness is welcome, and we note that this is precisely one of the recommendations made in our main report that Sumicsid failed to consider in TCB18.[65] Sumicsid should have carefully examined whether the dual weights in its model were appropriate, which it had not.

In this respect, we note that the model presented in our main report demonstrated how the efficiency scores change if TSOs are benchmarked to TSOs that deploy a similar mix of assets. The fact that doing so had a material impact on TSOs' efficiency scores indicated that TSOs were benchmarked against TSOs that deployed materially different assets. For instance, a TSO may be required to use underground cables instead of overhead lines in densely populated areas. When it is now compared to a TSO mainly deploying overhead lines, this solution is not available and the proposed cost reduction target can be infeasible. Importantly, the alternative model presented in our main report was not intended to provide a fully developed standalone benchmarking model.

The TCB18 response also dismisses all suggestions for a proposed margin of error in the aggregation of NormGrid. The TCB18 response criticises the regressions presented in our main report that are used to derive the error margin, as Sumicsid supposedly used an engineering assessment to determine the NormGrid aggregation weights. As noted in Table 1, this statement **directly contradicts** the appendices to the TCB18 main report where Sumicsid stated that it used linear regression to derive the NormGrid aggregation weights.[66] **These statements are irreconcilable and obfuscate the approach followed by Sumicsid**. We do not consider that any party other than Sumicsid could reasonably discern how the NormGrid weights were derived from the TCB18 outputs.

Nonetheless, the need for a margin of error is independent of the exact method used to derive the weights. Measures of uncertainty are built into statistical methods of setting the NormGrid aggregation weights (e.g. the confidence intervals on coefficient estimates), as demonstrated in our main report. However, expert judgement is not without uncertainty either—there is bound to be some disagreement in any panel of experts regarding the exact cost weight to attach to certain assets, regardless of whether the cost weights are informed by any quantitative or qualitative assessment. Indeed, we would expect that an individual expert would provide a feasible range rather than a precise point estimate in all cases. The need to account for this uncertainty is not addressed in Sumicsid's TCB18 response.

Environmental adjustments

Sumicsid's environmental adjustment to NormGrid in TCB18 is not supported by statistical, economic or operational evidence. We could find no external references in Sumicsid's outputs to support the weights it had used, nor was any robust statistical or operational evidence presented. Indeed, we found that

[62] This can include weight restrictions if required.
[63] Oxera (2020), 'A critical assessment of TCB18 electricity', April, section 4.4.
[64] Sumicsid response (2020), para. 3.88.
[65] Oxera (2020), 'A critical assessment of TCB18 electricity', April, section 5.3.
[66] Sumicsid (2019), 'Norm Grid Development – Technical report', p. 2.

the complexity factor weights were *counterintuitive*, as TSOs that operate in more complex regions (as defined by Sumicsid's complexity factor) have *lower* costs per NormGrid on average.

Sumicsid now reaffirms that, in a deviation from the project plan,[67] the weights were derived from engineering assessment (which was not presented or even cited in the TCB18 response) and not verified using econometric techniques.[68] As a result of this unseen assessment, Sumicsid does not engage with the operational and statistical arguments brought forward.

Sumicsid has not justified the assumptions that it has made in its final model. The additional evidence presented in the TCB18 response is unconvincing or contradicts Sumicsid's earlier position.

Returns to scale assumption

In our main report, we found no conclusive statistical evidence to support Sumicsid's returns-to-scale assumptions (non-decreasing returns to scale, NDRS, also called increasing returns to scale (IRS) by Sumicsid), and alternative returns-to-scale assumptions (including variable returns to scale, VRS) were generally preferred.[69] This is despite the fact that we conducted the same tests that Sumicsid outlined in its TCB18 report.[70] In its TCB18 response, Sumicsid reasserts that the NDRS assumption is supported by econometric tests, yet it does not report the results of these tests, nor does it attempt to explain the contradictory results that we found when using the same dataset and benchmarking model.

Instead of providing empirical arguments, Sumicsid focuses on the academic literature in its TCB18 response; it states that the use of NDRS is supported by the literature, quoting *seven* academic studies in which returns to scale have been estimated in the electricity industry.[71] After reviewing these studies in detail, we found that *only two of the seven* (Dismukes et al. (1998) and Filippini et al. (2004)) can be used to support an NDRS assumption, with the majority providing some support for a VRS assumption. **Therefore, Sumicsid's characterisation of these studies as providing 'abundant evidence'[72] for NDRS is manifestly incorrect**.

In its TCB18 response, Sumicsid further argues that 'no valid reason can be found to suggest that larger operators would be unable to organize their services and assets as efficiently as smaller units'.[73] However, for these efficiency gains to be realised, either the large TSO needs to be reduced to the most productive scale size *for its mix of outputs*, or split up into several TSOs, each one enjoying the most productive scale size. It is not clear whether either of these approaches are feasible in practice. Moreover, even if Sumicsid's assertion that TSOs cannot operate under decreasing returns to scale (DRS) was correct, Sumicsid must address the question of why large TSOs tend to be estimated to be inefficient (which could be driving the statistical tests to conclude that a VRS technology is appropriate). One feasible hypothesis is that the TCB18 model omits material drivers of expenditure that happen to be

[67] Sumicsid (2019), 'Norm Grid Development – Technical report', p. 4.
[68] Sumicsid response (2020), para. 3.92.
[69] Oxera (2020), 'A critical assessment of TCB18 electricity', April, section 5.1.
[70] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, para. 5.11.
[71] Sumicsid response (2020), para. 3.102.
[72] Ibid., para. 3.103.
[73] Ibid., para. 3.106.

correlated with scale in the sample. However, through its non-engagement on the topic, Sumicsid has implicitly assumed that the largest TSOs in its sample happen to be inefficient by chance, without justification.

Sumicsid asserts that 'the default approach in DEA should be to impose the minimally necessary assumptions', and that this 'would correspond to variable returns to scale'.[74] Given that the academic literature is (at best) mixed on the appropriate returns-to-scale assumption and that the statistical evidence does not support an NDRS assumption, a VRS assumption would be the most appropriate in TCB18 on Sumicsid's own rationale.

Outlier analysis

In TCB18, Sumicsid largely relied on the German regulatory ordinance (ARegV) for outlier detection.[75] As we discussed in our main report, the outlier procedure set out in the ARegV is neither legally binding nor sufficient in an international benchmarking context. Indeed, the scientific flaws of the ARegV's outlier procedure are well known.[76]

The TCB18 response does not engage with the merits of the different outlier tests discussed in our main report.[77] Instead, Sumicsid refers to the fact that its outlier procedure is still in use in Germany. This statement is irrelevant in the context of TCB18. The flaws of the ARegV's outlier procedure, such as the inability to detect masked outliers,[78] the bias towards non-detection,[79] the application of untested assumptions,[80] and the counterproductive nature of the tests in volatile samples,[81] still apply regardless.

In principle, we welcome that Sumicsid now states that visual inspection was used as an additional tool used for outlier analysis in TCB18 to supplement this flawed procedure.[82] However, we note that visual inspection is, by nature, not a rigorous process and only helps in identifying relatively obvious anomalies. The usefulness of this procedure could be supported if the visuals were presented in the TCB18 outputs, but they were not. Without specifying the nature of this comparison and the steps taken to identify why this TSO was unusual, it is impossible to verify whether it was appropriate.

DEA weights

Our main report highlighted that Sumicsid did not examine whether its final model's outputs were consistent with economic and operational expectations. For example, Sumicsid stated that NormGrid is the primary driver of expenditure, yet an analysis of DEA weights reveals that most TSOs' efficiency scores are not primarily driven by NormGrid.[83] In its TCB18 response, Sumicsid now claims that this statement was based on least absolute

[74] Sumicsid response (2020), para. 3.101.
[75] Oxera (2020), 'A critical assessment of TCB18 electricity', April, section 5.2.
[76] For example, see discussion in Kumbhakar, S., Parthasarathy, S. and Thanassoulis, E. (2018), 'Validity of Bundesnetzagentur's dominance test for outlier analysis under Data Envelopment Analysis', August; Deuchert, E. and Parthasarathy, S. (2018–19), five-part series of articles on the German energy regulator's benchmarking framework covering efficiency methods (DEA and SFA), functional form assumptions, cost driver analysis, outlier analysis, and model validation, in *ew–Magazin für die Energiewirtschaft*.
[77] Sumicsid response (2020), para. 3.115.
[78] E.g. Deuchert, E. and Parthasarathy, S. (2019), 'Sind die Verfahren zur Ausreißererkennung belastbar?', *ew–Magazin für die Energiewirtschaft*.
[79] Kumbhakar, S., Parthasarathy, S. and Thanassoulis, E. (2019), 'Rejoinder to Sumicsid's response to the Oxera report on Bundesnetzagentur's dominance test in DEA', May.
[80] Kumbhakar, S., Parthasarathy, S. and Thanassoulis, E. (2018), 'Validity of Bundesnetzagentur's dominance test for outlier analysis under Data Envelopment Analysis', August.
[81] Oxera (2020), 'A critical assessment of TCB18 electricity', April, p. 68.
[82] Sumicsid response (2020), para. 3.116.
[83] Oxera (2020), 'A critical assessment of TCB18 electricity', April, section 5.3.

shrinkage and selection operator (Lasso) and ordinary least squares (OLS) regressions.[84] However, this directly contradicts its statement in the main report that explicitly referred to NormGrid being 'the strongest candidate in the **frontier** models'.[85] We note that neither Lasso nor OLS regression is a frontier model according to Sumicsid's own definitions.

Importantly, Sumicsid has failed to grasp the deeper conceptual issue that the DEA model must be validated against operational expectations. Despite briefly discussing the use of dual weights to identify potential outliers in its TCB18 report,[86] and its apparent support for examining the operational relevance of the dual weights in its TCB18 response,[87] Sumicsid presents no evidence that it has examined the dual weights in its TCB18 model.

Second stage analysis

In the TCB18 report, Sumicsid stated that it used second-stage analysis (regression analysis where the dependent variable is the estimated efficiency score) to detect potential cost drivers that were omitted from its final model. Specifically, it stated:[88]

> [T]o **investigate whether some potentially relevant variables have been omitted** in the final model specification, a so-called second stage analysis has been performed. [emphasis added]

Our main report argued that there is **no theoretical basis** to support this validation approach to identify omitted outputs.[89] Moreover, we **demonstrated that Sumicsid's second-stage analysis fails to validate its own model specification**.[90] That is, Sumicsid's second-stage analysis cannot detect its identified cost drivers as omitted (i.e. relevant).

In its TCB18 response, Sumicsid now states:[91]

> The critique against the post-run (second-stage) analysis is not well posed since **the purpose is not related to model-specific variable selection**.

This directly contradicts the purpose of second-stage analysis stated in Sumicsid's previous statement. If Sumicsid had identified a cost driver as omitted in the second-stage analysis, one could reasonably assume that it would attempt to alleviate the bias (e.g. by controlling for this factor in the model), rather than simply concluding that the model is biased. Note that Sumicsid does not provide any further detail regarding how second-stage analysis can be used in the model development *or* model validation processes.

As noted above, Sumicsid has argued that NormGrid is the strongest cost driver. Therefore, one would expect that omitting NormGrid from the model specification would cause a systematic bias in efficiency scores—the same systematic bias that Sumicsid now states the second-stage analysis is designed to detect. However, we have already demonstrated that Sumicsid's second stage analysis **fails to detect NormGrid** as a relevant omitted output.

[84] Sumicsid response (2020), para. 3.120.
[85] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July. 5.01.
[86] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, p. 30.
[87] Sumicsid response (2020), para. 3.88.
[88] See Sumicsid (2019), 'Project TCB18 Individual Benchmarking Report Fingrid – 131', July, chapter 5.
[89] For example, see discussion in Kumbhakar, S., Parthasarathy, S. and Thanassoulis, E. (2018), 'Validity of Bundesnetzagentur's cost driver analysis and second-stage analysis in its efficiency benchmarking approach', February.
[90] Oxera (2020), 'A critical assessment of TCB18 electricity', April, section 5.4.
[91] Sumicsid response (2020), para. 3.126.

Therefore, Sumicsid's revised position on the issue remains conceptually and empirically incorrect.

Stochastic Frontier Analysis (SFA)

In TCB18, Sumicsid did not cross-check the results of its analysis using well-established alternative methods such as stochastic frontier analysis (SFA), despite having panel data available.[92] SFA applied to Sumicsid's model and dataset suggests that there is no statistically significant inefficiency among the TSOs.[93] Sumicsid now offers a limited explanation for the non-consideration of SFA, with its arguments falling into one of two categories.

- The use of an SFA model is not important when the data and model specification is deterministic.[94]
- The sample is too small to allow for a robust estimation of SFA models.[95]

The former argument relies on the assumption that the TCB18 data is without error, which is demonstrably wrong according to Sumicsid's own admission.[96] The various adjustments that are made (or not made) to the data create additional layers of uncertainty (see sections 3.3 and 3.4). Moreover as covered in section 4.3, a TSOs' expenditure in a given year is in part driven by 'Random' events as well as omitted factors. In this context, it is *incorrect* to state that the TCB18 model and data are deterministic.

The latter argument relies on a misrepresentation of a previous Oxera report that stated:[97]

> SFA models with fewer than 30–40 observations can fail to produce results.

We note that the TCB18 dataset contains *more than twice* the number of observations that are defined as a 'lower bound' in the quote above (the TCB18 dataset contains 81 observations, while the quote suggests difficulties with fewer than 30–40 observations). Sumicsid's argument regarding sample size rests on its failure to consider the time-series dimension of the dataset properly for assessing the TSOs' efficiency levels, even though it relied on this structure for model development.[98] Nowhere in its main report or in its TCB18 response does it provide a credible reason for this inconsistency, which results in an unnecessary loss of useful information.

Moreover, while SFA *can* fail to produce results with fewer than 30–40 observations, this does not mean that it *will* fail. With the exception of extremes,[99] the appropriate sample size will always be an empirical issue that cannot be tied to a fixed number. We demonstrated in our main report that SFA models *can* be estimated on the TCB18 dataset, and the analysis supports our high-level conclusions from other methods considered in our review.

[92] A panel dataset contains data over time across TSOs and thus contains more information than a single year of data.
[93] Oxera (2020), 'A critical assessment of TCB18 electricity', April, section 5.5.
[94] Sumicsid response (2020), para. 3.136.
[95] Ibid., para. 3.136.
[96] Oxera (2020), 'A critical assessment of TCB18 electricity', April, section 3.1.
[97] Oxera (2012), 'How can the NMa assess the efficiency of GTS?', June, p. 2.0.
[98] Note that Sumicsid has subsequently changed its view on the use of statistical analysis in model development in its TCB18 response.
[99] For example, an SFA model cannot be estimated if the number of observations is less than the number of parameters to be estimated.

Frontier shift

The TCB18 data and model suggest that there has been a material regress in the efficient frontier in the period 2013–17.[100] In our main report, we noted that a frontier regress is inconsistent with the estimates usually applied to TSOs' expenditure as part of incentive regulation regimes that tend to show a positive frontier shift (i.e. frontier progress). Moreover, the unusual result provided further evidence that Sumicsid's TCB18 model is mis-specified. That is, if the model cannot capture *changes in efficient costs* over time, then it is unlikely that the model can capture *differences in efficient costs* between TSOs.[101]

Sumicsid asserts that 'neither productivity progress, nor regress per se should be directly interpreted as linked to the model structure'.[102] This statement is meaningless as it is, or at the very least requires further explanation. The frontier shift calculation involves calculating a TSO's performance relative to an estimated frontier—how that frontier is defined (e.g. which cost drivers are included in the model and the modelling assumptions involved) will have an impact on what the estimated frontier shift is.

Sumicsid quotes two results that it states support its observation of a frontier regress.[103] We find that one of its references actually estimates a positive frontier shift. Llorca et al. (2016) states:[104]

> The coefficient on the time trend is negative (nonetheless it is not significant in some of the models), which indicates that costs decrease over time, i.e. there is positive technical change.

It is not clear whether Sumicsid has robustly interrogated its sources when responding to our main report. Nonetheless, the existence of one or two isolated academic papers does not validate Sumicsid's approach, and this constitutes another example of Sumicsid's failure to engage fully with the arguments presented in our report.

[100] Sumicsid (2020), 'PROJECT CEER-TCB18 Dynamic efficiency and productivity changes for electricity transmission system operators MAIN REPORT', April.
[101] Oxera (2020), 'A critical assessment of TCB18 electricity', April, section 5.6.
[102] Sumicsid response (2020), para. 3.139.
[103] Ibid., para. 3.140.
[104] Llorca, M., Orea, L. and Pollitt, M.G. (2016), 'Efficiency and environmental factors in the US electricity transmission industry', *Energy economics*, **55**. pp. 234–46.

1 Introduction

The pan-European Transmission Cost Benchmarking Project (TCB18)[105] was carried out by the Council of European Energy Regulators (CEER) through its consultant, Sumicsid. Sumicsid used cost and asset data provided by the 17 participating electricity Transmission System Operators (TSOs), in addition to environmental and input price data from external sources, to estimate the relative efficiency of TSOs through data envelopment analysis (DEA). The TCB18 study concluded that the efficiency of participating TSOs ranges between 66% and 100%, with a mean value of 89%, indicating a total annual savings potential of €713m for the sample of the 17 participating TSOs.

A consortium of all of the TSOs that participated in TCB18 commissioned Oxera to validate and review the results from TCB18, and to recommend robust solutions to any issues that emerge. As part of this study, we reviewed outputs produced and shared with the TSOs by Sumicsid from TCB18 and had access to the complete underlying dataset that was used.

The overall assessment of this exercise ('our main report'),[106] which was published in April 2020, was that the estimated efficiency scores cannot be used for regulatory, operational or valuation purposes in their current form. To remedy the issues identified, we provided a total of 15 recommendations on the study's procedural and methodological issues. We also suggested that it could be helpful to consider debriefs involving all parties to help future studies.

The authors of TCB18 (Professors Per Agrell and Peter Bogetoft for Sumicsid) responded to this assessment in October of 2020 ('the TCB18 response').[107]

We also analysed the impact of several issues on TenneT's individual efficiency. We refer to this assessment where appropriate. TenneT has now commissioned Oxera to review and assess the arguments brought forward by Sumicsid in its reply to determine whether it alters any of the conclusions of our main report or the assessment of TenneT's individual efficiency.

[105] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July.
[106] Oxera (2020), 'A critical assessment of TCB18 electricity', April.
[107] Sumicsid (2020), 'Response to the Oxera Report on TCB18 ETSO', October—cited in footnotes as 'Sumicsid response (2020)'.

2 Transparency

The level of transparency in the TCB18 exercise was a key issue in our main report. Sumicsid’s outputs do not contain the necessary information for third parties to clearly follow its analysis, validate its analysis or validate its sources without considerable effort.

We understand that the TCB18 data is confidential and could thus not be published in full. However, public availability of data is only one aspect of transparency; there are several additional aspects that should have been considered to ensure sufficient transparency, such as the publication of analysis files or a clear description of processes followed. If those aspects are properly executed, the data would not need to be published to ensure third parties could have a sufficient understanding of the analysis.[108]

Specifically, in our main report, we observe that the approach chosen and analysis undertaken must be presented in a way that enables easy comprehension and validation. In addition to a clear and comprehensive description of the work, this could take the form of a publication of anonymised analysis files (e.g. scatterplots for visual inspection of outliers) and modelling codes. We have provided an example of such a code as part of our main report.[109] This is considered practice in regulatory applications, even when the data itself is deemed confidential.[110] For instance, the German regulator, Bundesnetzagentur, published the modelling code for the sectoral productivity factor on its website. It remains published, even though the data was removed from the website following a court order.[111]

The final report and its appendices lack a clear and comprehensive description of the work and analysis undertaken. Oftentimes, the statements that Sumicsid makes to support its strong assumptions are not supported by any statistical, operational or literary evidence. Even where Sumicsid states that its assumptions are supported by, for example, statistical evidence, the evidence itself was not shown (e.g. in the form of regression tables).

To illustrate this deficiency we provide an example from TCB18: Sumicsid’s ‘returns to scale’ assumption. Returns to scale (RTS) relates to how changes in inputs (i.e. TOTEX) are linked to changes in outputs (e.g. NormGrid) for efficient companies.

- The returns to scale assumption is one of the key inputs into a DEA model and is mentioned in one slide of the first workshop.[112] Sumicsid gives no indication what RTS assumption will be applied in TCB18 in this workshop, nor did Sumicsid discuss how the appropriate RTS assumption will be determined.
- Despite being a key assumption in the DEA, the RTS assumption is not mentioned again in the material of the subsequent workshops 2–5. No

[108] Note that publishing data and modelling codes can still be helpful in any study even if the written documentation is fully transparent. For example, it is unlikely that specific data errors could be identified by third parties (as Oxera did in its main report) if the data is not shared.
[109] Oxera (2020), ‘A critical assessment of TCB18 electricity’, April.
[110] Ofwat (2019), ‘Final determinations models’, available at: https://www.ofwat.gov.uk/final-determinations-models/ (last accessed 13 April 2021); NVE (2018), ‘IRiR - a script for calculating Revenue Caps for Norwegian Electricity DSOs’, available at: https://github.com/NVE/IRiR (last accessed 13 April 2021).
[111] Bundesnetzagentur (2018), ‘Festlegung des generellen sektoralen Produktivitaetsfaktors fuer Elektrizitaetsversorgungsnetze (3. Regulierungsperiode)’, available at: https://www.bundesnetzagentur.de/DE/Service-Funktionen/Beschlusskammern/BK04/BK4_76_Prodfakt/BK4_Prodfakt_Strom_basepage.html (last accessed 13 April 2021).
[112] Sumicsid (2018), ‘CEER-TCB18 project Methodological Approach’, January, slide 16.

project participant could have been aware of the RTS chosen ahead of the publication of the final report.

- The final report then includes the following statement (which is also an **exact copy of the text used in e2gas**,[113] a predecessor study to TCB18) and TCB18 gas:[114,115]

> For **all possible model specifications**, we have also tested which of the returns to scale assumptions in the DEA model fit data the best […]. We have done so **using F-tests** based on a goodness-of-fit measure as explained in the Method chapter. The general finding is that the IRS assumption […] is the best assumption to invoke. This is **supported also by parametric analyses for a logarithmic model**, where the coefficients sum to less than one for the selected parameters. [emphasis added]

- In e2gas, the results of a logarithmic regression were presented alongside this statement.[116] However, this was not the case in TCB18.[117]
- The direct predecessor study to TCB18 electricity, e3grid, also uses results from a log-linear regression model (which were presented) to support the returns to scale assumption.[118]

From the published material, one would be led to believe that the RTS assumption was made based on strong statistical evidence, despite the fact that the evidence is not presented. An independent assessor of the study cannot be expected to simply trust Sumicsid's statements—Sumicsid must support its choices with evidence rather than rely on statements alone.

Indeed, in our main report, we **could not find any statistical evidence supporting the RTS assumption** invoked by Sumicsid using the exact same tests described. In contrast, we found some evidence supporting a variable returns to scale assumption.[119] Since our analysis was conducted on the same dataset that Sumicsid used in TCB18, it is unclear to us how Sumicsid could have reached a different conclusion.

In the TCB18 response, Sumicsid now states that:[120]

> The assumption for NDRS is based on techno-economic arguments linked to the **abundant evidence from other studies**, previous transmission studies, concession areas and the model specification (NormGrid covers all relevant assets). [emphasis added]

This statement indicates that statistical considerations were **not** the primary reason behind the NDRS assumption. Rather, the decision was based on a supposedly established consensus in the literature**.** We disagree that such a consensus exists, even in the literature quoted by Sumicsid. Indeed, many of

[113] Sumicsid and Swiss Economics (2015), 'PROJECT E2GAS Benchmarking European Gas Transmission System Operators – Final report', section 5.32.
[114] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators', July, section 5.12.
[115] The tests mentioned are an F-Test based on the estimated efficiencies in the sample (equivalent to the dominance test of the Bundesnetzagentur outlined in section 5.2 of our main report Oxera (2020), 'A critical assessment of TCB18 electricity', April) and a test of the sum of coefficients in a logarithmic model.
[116] Sumicsid and Swiss Economics (2015), 'PROJECT E2GAS Benchmarking European Gas Transmission System Operators – Final report', Figure 5-3.
[117] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators', July, section 5.12.
[118] Consentec, Frontier Economics and Sumicsid (2013), 'E3GRID2012 – European TSO Benchmarking Study', p. 78.
[119] Oxera (2020), 'A critical assessment of TCB18 electricity', April, section 5.1.3.
[120] Sumicsid response (2020), October, para. 3.107.

the studies quoted by Sumicsid do not find evidence for NDRS over VRS.[121] For example, Sumicsid uses Yatchew (2000) to justify its RTS assumption, claiming that Yatchew (2000) finds increasing returns to scale for smaller utilities. However, the full quote from the abstract of the paper is:[122]

> The data reveal substantial evidence of increasing returns to scale with minimum efficient scale being achieved by firms with about 20,000 customers. **Larger firms exhibit constant or decreasing returns**. [emphasis added]

That is, Yatchew (2000) does not support Sumicsid's RTS assumption. Similar disparities between Sumicsid's positioning of academic articles and their content can be found for several other papers quoted (see section 5.1.2).

Importantly, the statistical evidence that was stated as underpinning the decision to apply an NRDS assumption to the TCB18 data appears to have actually been secondary to theoretical considerations (which themselves appear inconsistent with the literature quoted). It is unclear how a project participant, let alone a third party not directly involved in the project (such as a valuations agency), could have discerned this from the TCB18 report.

The type of transparency outlined above relates to the ability for TSOs and third parties to follow exactly what Sumicsid has done, such that alternatives or improvements to the analysis can be proposed and the limitations of the analysis can be understood. However, in addition to this type of transparency, the reports resulting from Sumicsid's study do not contain key outputs of the benchmarking model that would be fundamentally useful for TSOs and NRAs.

In particular, information on peer companies[123] and the importance of different cost drivers to the TSO (virtual weights) are missing entirely. These outputs are essential in helping the TSOs (and the NRAs) to understand what drives the (in)efficiency of the TSO in the model. For example, a TSO may be interested in understanding whether it is being compared to a TSO with a strong distribution component or a pure point-to-point transport TSO. TSOs can utilise these outputs and learn from the benchmarking exercise and thus improve their efficiency by comparing themselves to peers. Nevertheless, these outputs are not given to the TSOs in TCB18.

There are many other instances in which Sumicsid's level of transparency was insufficient, including with respect to:

- the purpose of the second stage analysis (section 5.4);
- the process of model development in TCB18 (section 4);
- how the NormGrid weights were derived (section 4.4);
- the evidence underpinning the environmental adjustments (section 4.5).

We will highlight these and other misleading statements made by Sumicsid throughout this report. These statements not only add to the general lack of transparency in TCB18, but also cast doubt on the validity of results from the benchmarking exercise, as the authors contradict themselves or are unclear on material points, severely limiting the robustness of the process that was followed in the cost driver analysis.

[121] See section 5.1.
[122] Yatchew (2000), 'SCALE ECONOMIES IN ELECTRICITY DISTRIBUTION: A SEMIPARAMETRIC ANALYSIS', p. 188.
[123] Note that this would not necessarily entail naming the efficient TSOs (which could breach confidentiality), but rather describing them qualitatively and quantitatively.

3 TCB18 data collection and construction

3.1 Data errors

In TCB18 no provision was made for data errors: TCB18 simply states that 'although no approach will be fully safe' the datasets are of a 'good quality'.[124]

Our assessment demonstrated that the TCB18 data, which was used to develop the benchmarking model and estimate efficiencies, contains significant errors which are not accounted for.[125] As the benchmarking method used, DEA, is unable to directly account for noise, we proposed a Monte Carlo simulation for noise to account for it. Specifically, we recommended adding a noise component of +/-10% to all data points and re-estimating the model numerous times to derive a distribution of efficiency scores for each TSO.

In the TCB18 response, Sumicsid repeats the data collection process from the main report and emphasises the role of the NRAs. Nonetheless, the TCB18 response confirms the existence of several data errors in the final data used in TCB18. These include:

- that the data for the weighted lines variable was inaccurate for all years except 2017 due to a calculation error made by Sumicsid;[126]
- that a TSO incorrectly reported no angular towers;[127]
- that one TSO overreported circuit ends[128] (in fact we can confirm through correspondence with the participating TSOs that this issue extended to at least one other TSO).

The magnitude of the first error is summarised in Table 3.1 below. On average the error led to the TSOs' weighted lines value being understated by 15%. Although we cannot assess the magnitude of the error for specific TSOs, note that the magnitude does vary at different points in the distribution: Sumicsid's data overstated the value of weighted lines by c. 4% for the first quartile, but understated the value by 41% for the third quartile.

Table 3.1 Impact of the calculation error in weighted lines

	Mean	First quartile	Median	Third quartile
Data used by Sumicsid in TCB18	1,772	678	1,286	1,752
Corrected data	2,040	651	1,378	2,475
Difference (%)	15%	-4%	7%	41%

Source: Oxera analysis based on Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, Table 5-2.

However, the TCB18 response claims that while 'minor individual errors' were not spotted, 'the final data material is of very high quality',[129] and that the noise component simulated in our main report to account for data errors (+/-10%) is thus excessively large. Given that we have evidence of existing data errors in

[124] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, sections 3.06–3.07.
[125] Oxera (2020), 'A critical assessment of TCB18 electricity', April, section 3.1.
[126] Sumicsid response (2020), para. 3.50.
[127] Ibid., para. 3.09.
[128] Ibid.
[129] Ibid., para. 3.12.

excess of our applied margin, we disagree with Sumicsid's statements regarding our proposed error margin.

While Sumicsid may be able to correct for the specific data errors that we highlighted in our main report, the errors that we identified should rather be seen as indicative of wider issues that cannot be fixed by simply adjusting the data for individual TSOs. Importantly, the impact of the errors on the entire process—i.e. cost driver analysis, outlier procedure, efficiency estimation and validation of outputs—should be investigated and shared with the participants.

In this regard, Sumicsid acknowledges that the ability of DEA to account for noise is limited.[130]

3.1.1 Material omissions in Sumicsid's response

The TCB18 response does not fully engage with the core argument in our main report. Real-world data is inherently noisy and the potential sources of error are manifold, as outlined in our assessment of TCB18, and data uncertainty and errors exist in nearly all applications of regulatory benchmarking, even when data collection has been standardised, publicised and is ongoing for decades.[131]

The fact that TCB18 required the collection of a largely novel dataset (especially for TSOs that did not participate in previous international benchmarking exercises) stretching back 45 years that included information not readily available to many TSOs may help to explain the scale and prevalence of data errors (and other data uncertainty). For instance, one TSO stated in correspondence with Sumicsid that:[132]

> Our accounting system **does not hold asset counts** nor is there any further breakdown of our project spend into the Asset cost split as required by CEER and will take a **significant amount of re-work** / investigation with **a lot of judgement** being used. [emphasis added]

In the face of such explicit statements that judgement played a major role in determining the reporting of TCB18 data, Sumicsid's assertion that the data is 'deterministic' and 'not stochastic' is clearly incorrect—judgements and assumptions have been made by TSOs and Sumicsid in the data collection and construction phase, and this will add to the uncertainty in the benchmarked costs and cost drivers.

Indeed, there appears to be significant uncertainty in the *raw data* that TSOs submitted to Sumicsid, without even considering the additional uncertainty caused by Sumicsid's explicit assumptions when some data was unavailable. For instance, as one TSO was unable to produce a CAPEX stream preceding 2013, the previous values of assets were calculated from the regulated asset value and distributed to asset categories based on 2013–17 spending (e.g. assuming that the split of Lines to Cables in those five years was representative for the 50 preceding years).[133] At least eight other TSOs,

[130] Sumicsid response (2020), para. 3.13.
[131] For example, the use of 'average pumping head' (a measure of pumping capacity) in econometric models in the English and Welsh water sector was debated extensively in the PR19 CMA appeals. Ofwat, the regulator, had decided not to include the variable in its econometric models due to uncertainty regarding its calculation. Note that, unlike Sumicsid's application of DEA, Ofwat's econometric approach takes statistical uncertainty into account to some extent. See CMA (2020), 'Anglian Water Services Limited, Bristol Water plc, Northumbrian Water Limited and Yorkshire Water Services Limited price determinations: Provisional findings', September, pp. 112–5.
[132] SHE-Transmission (2018),'CEER financial reporting notes', October.
[133] Sumicsid (2019), 'Correction of initial AV - 115', March.

including three peers and two super-efficient outliers, were affected by similar issues relating to the initial asset value.[134]

On top of this, Sumicsid's cost normalisations (e.g. the correction for input prices, the environmental adjustment to NormGrid) introduce an additional layer of uncertainty. These adjustments to the data are discussed elsewhere in this report.

The stated assumption that the data is deterministic is also inconsistent with statements made by Sumicsid elsewhere in the report. For instance, Sumicsid observes that the level of spending varies year by year.[135] The volatility of expenditure can be explained by several factors, some of which are stochastic. For example, adverse weather or changes in input prices are largely stochastic variables (from the perspective of the TSO), and these could have a material impact on expenditure in any given year. Moreover, expenditure can be driven by deterministic factors that are unaccounted for in the model, such as the costs associated with an ageing network.[136] The stochastic factors will have had random realisations in 2017, affecting estimated efficiencies and necessitating a margin of error.

We thus maintain that it is likely that the substantial and numerous data errors and inconsistencies discovered in TCB18 data and discussed in our assessment of TCB18 are only indicative of wider issues. In addition, Sumicsid appears not to have taken any steps to correct the three examples mentioned in section 3.1.1 of our main report—namely the following.

- One TSO mistakenly enters some of its towers as 'steel' as opposed to 'wood'.
- One TSO submits CAPEX data that has already been adjusted for inflation, but is subsequently adjusted again by Sumicsid (therefore resulting in a double-counting of the inflationary impact).
- Some TSOs aggregate their towers in a way that seems to indicate that they have no angular towers, which was not spotted by Sumicsid.

Moreover, to our knowledge, Sumicsid has not issued a corrigendum to the TCB18 report that addresses the data errors that were identified and their impact on the entire exercise.

3.1.2 Conceptual errors in Sumicsid's response

Sumicsid correctly acknowledges that its chosen method (DEA) is less able to account for statistical noise. Given the prevalence of noise in the data and models (as highlighted above and throughout our main report), this noise must be accounted for through additional modelling.

In our main report, we offered pragmatic and straightforward approaches to account for the statistical noise in the data, which have regulatory precedent. These include the use of alternative estimation approaches that can better account for statistical noise such as SFA, which Sumicsid itself had used as a validation exercise in E2GAS,[137] as well as the use of Monte Carlo analysis to simulate the impact of data uncertainty on the DEA model itself. Note that the

[134] This count is based on the amount of TSOs with a significant (>20%) opening balance or an investment stream going back less than ten years.
[135] Sumicsid response (2020), para. 3.71.
[136] In TenneT's case, these factors may explain why maintenance expenditure varies around a clear trend.
[137] Sumicsid and Swiss Economics (2015), 'PROJECT E2GAS Benchmarking European Gas Transmission System Operators – Final report'.

latter also forms part of the approaches to testing the representativeness of the DEA model which was recommended by Economic Insight Ltd in a report for the ACM.[138]

Sumicsid states that the 'impact of data errors should be proportionate to the magnitude and frequency of errors expected in the sample'[139] when conducting Monte Carlo analysis. We agree with this statement, which is precisely why we assumed a 10% error in our report.[140] Indeed, in the presence of *several known data errors of over 25%*, a 10% error margin would be conservative. These known data errors include the following.

- The error discovered in a TSO's CAPEX overstates its TOTEX by 32%, resulting in its efficiency being understated by more than 25 percentage points and affecting the shape of the efficient frontier (as it would be a peer with corrected data).
- The value for weighted lines is overstated by 27% for one TSO (contrary to Sumicsid's assertion that issues in the tower data are not material), resulting in a 11.4% change in estimated efficiency for the affected TSO.
- One TSO incorrectly received a zero value for the weighted lines variable.
- The weighted lines variable was calculated erroneously for 80% of the panel (2013–16). On average the value was understated by 15%, affecting all dynamic results and the model development.

The first two of these errors were noted to the relevant NRAs and formed part of regulatory submissions.[141] We understand that the relevant regulators subsequently decided to not apply an efficiency challenge based on TCB18.[142]

It is unclear why Sumicsid has stated that the 10% error margin is 'exaggerated' in this context.[143] As is evident from the examples above, and in direct contradiction to Sumicsid's repeated assertions that any data errors are 'minor' and 'individual', data errors are substantial and continue to have a material influence on estimated efficiencies and the shape of the efficient frontier for several operators.

Sumicsid has acknowledged the presence of multiple data errors in the dataset on which it performed its final analysis.[144] However, it has implied that these data errors are not significant on the context of the TCB18 as the affected TSOs are either inefficient or outliers, meaning that they have no impact on the estimated efficient frontier. In this respect, we note the following.

- The data errors would have had an impact on the affected TSOs' estimated efficiency scores and the efficiency scores of other TSOs. Had the affected

[138] Economic Insights (2017), 'Topics in efficiency benchmarking of energy networks: Choosing the model and explaining the results', December, p. iv.
[139] Sumicsid response (2020), para. 3.13.
[140] Note that Sumicsid states that our report is unclear on whether we apply a 20% range or a 10% range in our report. (see Sumicsid (2020), 'Response to the Oxera Report on TCB18 ETSO', para. 3.11). We refer Sumicsid to section 3.1.2 of our report, where we explained that the assumed error was plus or minus 10% (i.e. a range of 20%) by way of example. See Oxera (2020), 'A critical assessment of TCB18 electricity', April, p. 33.
[141] Data errors for any one TSO are likely to affect the estimated efficiency scores of many other TSOs. This is most obvious if the TSO with data errors is a peer, as it directly affects the shape of the efficiency frontier. However, even data errors for non-peer TSOs will have an impact on the model development and outlier analysis and thereby on the overall results.
[142] For example, see Ofgem (2021), 'RIIO-2 Final Determinations Electricity Transmission System Annex (REVISED)', February.
[143] Sumicsid response (2020), para. 3.13.
[144] Towers, historic, para. 3.05.

TSOs' respective NRAs used the results from TCB18 to set regulated revenues, the impact of these errors (and Sumicsid's failure to account for them) would have had material implications for the TSOs and their consumers.

- The existence of data errors for any TSO will have an impact on Sumicsid's outlier analysis, irrespective of whether they are on the frontier in the final estimation.
- If Sumicsid has used any statistical analysis to support or develop its model, data errors would have an impact on the final model specification that is developed.[145] This could have a material impact on several TSOs' efficiency scores.

Sumicsid's assertions regarding the high-quality nature of its dataset are value judgements. Indeed, the implications in its outputs are that the data is measured without uncertainty or error,[146] but this *cannot* be true given that many errors have been identified. Sumicsid should make clear what it means by 'high-quality'. Is it high quality relative to previous international benchmarking studies that Sumicsid has conducted, relative to national benchmarking applications, or relative to some metric?

Note that, in addition to any uncertainty in the raw data, the modelling choices that Sumicsid makes subsequently, such as the adjustment for input prices, add an additional layer of uncertainty. This is covered in sections 3.2, 3.3 and 3.4.

3.2 Defining the input variable

Sumicsid uses TOTEX as the input variable in TCB18. TOTEX is defined as the sum of OPEX and CAPEX, which are calculated using different normalisations. This was summarised in Table 3.1 of our main report, which is reproduced as Table 3.2 below.[147]

Table 3.2 Sumicsid's cost normalisation approach

Issue	OPEX	CAPEX
Time period over which it is assessed	Annual (i.e. only 2017 OPEX is considered)	A sum of annuitised investments from 1973
PLI adjustment	Adjustment for labour prices using the PLI for civil engineering works	No PLI adjustment
Inflation adjustment	N/A1	Adjusted for inflation in overall goods

Note: [1] Sumicsid's final model is estimated using OPEX from the year 2017 only. As such, the inflation adjustment is not applicable.

Source: Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', sections 4.14–4.16.

[145] Note that Sumicsid's outputs do not make it clear how (or if) statistical analysis fed into the TCB18 project, and many of Sumicsid's statements on the topic appear to be contradictory (these are discussed elsewhere in the report).
[146] If any data is measured with any level of uncertainty, then the data cannot be 'deterministic', as Sumicsid often asserts.
[147] Sumicsid appears to be confused by the lack of an inflation adjustment for OPEX in this table. As this table is concerned with the data entering Sumicsid's benchmarking model, which only covers 2017, no such inflation adjustment was made for OPEX.

In our main report, we highlighted that this aggregation relies on certain assumptions to provide a robust estimate of efficiency and the correct incentives. Namely they must be:

- **equivalent**—at the margin, €1 of CAPEX should have the same worth as €1 of OPEX in terms of supporting output levels. That is, there must be a one-to-one trade-off between OPEX and CAPEX;
- **controllable**—the ratio of OPEX to CAPEX needs to be within the control of management.

We demonstrated that, due to the different normalisations involved and the different solutions available to the different TSOs in an international setting, the use of a single input TOTEX model may not give achievable estimates of potential savings. We thus suggested several alternatives to a single-input TOTEX model including the use of a two-input model (with or without weight restrictions) to ensure a TSO is only compared to TSOs with similar OPEX to CAPEX ratios and activity-level benchmark models.

Sumicsid makes three observations on the use of TOTEX in its TCB18 response.

- Sumicsid concedes that different normalisations in OPEX and CAPEX can influence efficiency scores. It cites the impact from its own sensitivity analysis on labour costs in CAPEX to suggest that this was covered.
- Sumicsid dismisses our suggested alternative models and cites regulatory regimes that employ one-input TOTEX benchmarking as evidence that TOTEX benchmarking is the preferred solution.
- Sumicsid presents the example of AST leasing its grid as an argument *in favour* of TOTEX modelling, while Oxera's main report argued that it was a reason to explore alternatives.

3.2.3 Misrepresentation of Oxera's main report

In its TCB18 response, Sumicsid implies that Oxera suggested benchmarking OPEX only.[148] This is not true. We offered three possible alternatives to the one-input TOTEX model used by Sumicsid, all three of which would ultimately use an aggregate TOTEX cost base as inputs.

- **A two-input DEA model**, as presented in our main report, uses both OPEX and CAPEX. The substantial difference between this and a TOTEX model is that this would ensure that a TSO is compared only to TSOs with similar OPEX and CAPEX ratios. This would directly account for the heterogeneity between the cost categories.
- **A two-input DEA model with restricted weights**: Sumicsid appears particularly dismissive of this suggestion calling it 'without any merit', 'arbitrary' and 'redundant'. However, this simply constitutes a pragmatic compromise between a one-input TOTEX model proposed by Sumicsid and a two-input OPEX–CAPEX model favoured by our main report. This would recognise that the trade-offs between OPEX and CAPEX may be skewed in different directions for different TSOs by the normalisation methods used and the different operating environments they are facing, while also ensuring that dual weights are within an operationally intuitive range. Note that the one-input TOTEX model is simply the extreme case of this

[148] Sumicsid response (2020), para. 3.15.

suggestion where the possible range of dual weights and thus the margin for error is reduced to zero.

- **Multiple activity-level models** that can be aggregated to a TOTEX basis: this goes beyond the two previous suggestions in that it suggests entirely new model development along reasonable delineations of activity such as maintenance, grid planning and new CAPEX projects. Modelling at the disaggregate level allows for more and more specific cost drivers to be included in the models. Of course, comparability of spend at the disaggregated level would be ensured as well.

While all of these suggestions relax the strict assumptions on the 1:1 relationship between OPEX and CAPEX assumed by Sumicsid's one-input model, all suggestions benchmark at the TOTEX level. Thus, Sumicsid's suggestion that Oxera is proposing an OPEX model is void.

Note that, as discussed in our main report, any or all of these can also be used as a cross-check of the one-input TOTEX model used by Sumicsid. Sumicsid does not reply to this suggestion at all.

3.2.4 Material omissions

Sumicsid does not address the main issue—that the trade-off between OPEX and CAPEX may not be controllable to the TSO and the estimated efficiencies and savings may thus not be achievable at all. In international benchmarking, unlike in national benchmarking, this issue is further exacerbated by the different legal environments that TSOs operate in, an issue unique to international benchmarking and not covered by the regulatory precedent quoted by Sumicsid.

Sumicsid also does not sufficiently address the issue of equivalence due to differences in normalisation. The sensitivity analysis that it presents focuses exclusively on the adjustment for price levels, which we cover in detail in section 3.3 of the main report and section 3.3 in this reply.[149] However, the normalisation of CAPEX and OPEX also covers two additional distinct steps, as outlined in our main report.

- CAPEX represents an annuitised stream of investments while OPEX is based on one year only.
- The CAPEX stream is adjusted for inflation before being summed up (OPEX in 2017 does not need to be adjusted for inflation).

Differences in normalisation and their effect on estimated efficiencies are still not sufficiently addressed by Sumicsid. Issues around the normalisation of CAPEX that were pointed out in our main report are well known and not unique to TCB18.[150] Uncertainties around CAPEX normalisation may be addressed by calculating CAPEX using several approaches and setting efficiency targets using a combination of the various approaches (e.g. the maximum efficiency score from either approach). Specifically, several regulators including the German and Austrian regulator referenced by Sumicsid calculate CAPEX in an annuity approach (as used by Sumicsid with a flat CAPEX impact of historical investments over time) and in a depreciated historic cost approach (in which the CAPEX impact of investments declines over time until fully depreciated).

[149] Note that the suggested impact of the sensitivity is substantial, implying an uncertainty range of up to 9%.
[150] Oxera, 'A critical assessment of TCB18 electricity', April, p. 36.

In addition to the calculation approach, a comprehensive sensitivity analysis would cover interest rate, lifetimes of assets, and price level adjustments both individually and in combination. The limited sensitivity analysis presented by Sumicsid (which nevertheless shows a significant impact on estimated efficiencies) is thus incomplete.

3.3 Adjusting for differences in input prices

In TCB18, Sumicsid adjusted approximately 5.9% of TOTEX for differences in input prices, using nominal exchange rates to convert the rest of the expenditure. In our main report we criticised this choice as insufficient. It does not capture all of the material differences in price levels between countries. In fact, the TCB18 study assumes that maintenance services and all investment goods can be procured for the same price in Norway as they can be in Germany, Slovenia and the UK. Sumicsid also did not provide sufficient evidence to validate its hypotheses.[151]

In response, Sumicsid agrees that a correction for local input prices is the right approach in principle if price differences are 'exogenous and well-identified'.[152] It also states that:

> The correction of input prices by general price indexes is not harmless. Without observing the origin and controllability of the expenditure, it may lead to undue protection of inefficient procurement in high-cost areas and to unfair penalties for procurement in low-cost areas.

At several points in its response, Sumicsid also acknowledges that the correction made may be insufficient. However, it states that it would be the task of the NRA to 'take into consideration specific suboptimal conditions that explain its occurrence'.[153]

3.3.5 Factual errors and misrepresentations of Oxera's first report

In the TCB18 response, Sumicsid claims that there is no prior international benchmarking exercise in energy that uses a price level correction for the entire cost base.[154] First, we note that the restriction to benchmarks 'in energy' is arbitrary as it excludes other industries, such as rail, where the existence of national monopolies also necessitate international benchmarking. International benchmarking in rail, such as ORR's benchmarking of network rail, commonly adjusted the entire cost base for differences in price levels.

However, even within the narrow focus of the energy sector, one of the papers quoted by Sumicsid (Jamasb et al. (2007))[155] explicitly *does* use a correction for the price level differences between countries. It states:[156]

> [W]e use Purchasing Power Parities (PPPs), which by construction equalize average domestic price levels.

The only referenced works that do not account for differences in price level, then, are those prepared by Sumicsid itself.

It is not clear whether Sumicsid has considered the relationship between PPP and a PLI. Sumicsid states that the correction used by the ORR in its

---

[151] Oxera, 'A critical assessment of TCB18 electricity', April, section 3.3.
[152] Sumicsid response (2020), para. 3.26.
[153] Ibid., para. 3.31.
[154] Ibid., para. 3.34.
[155] Jamasb, T; Newbery, D; Pollitt, M (2007), 'International Benchmarking and Regulation of European Gas Transmission Utilities'.
[156] Ibid., p. 27.

benchmarking of Network Rail (transforming international, nominal, data to GBP using PPP exchange rates and then deflated using the UK inflation index) is different from the one applied here (deflating using national inflation indices, converting to EUR using exchange rates in 2017 and applying a PLI). This method is however equivalent in its theoretical approach to the one proposed in our assessment of TCB18, since PPP is defined as the product of a PLI with the exchange rate.[157] Jamasb et al., ORR and our main report all propose adjustments for price level, inflation *and* exchange rate. As these three operations are performed with multiplication, the order in which they are performed is irrelevant.[158]

The exclusion of an adjustment for PLI is thus limited to Sumicsid's works alone.

3.3.6 Conceptual errors in Sumicsid's response

Sumicsid correctly identifies the theoretical need for correction in input prices in its reply, stating that:[159]

> Methodologically, the correction for local (potentially operator-specific) input prices is the correct approach when the said prices are exogenous and well-identified.

Sumicsid also correctly identifies one of the empirical challenges in international benchmarking: that no perfect index for the price differences between TSOs exists and differences in input prices are thus not easily isolated from other factors and corrected for.

We already emphasised the uncertain nature of price level adjustments in our main report.[160] As a direct consequence of this uncertainty, our main report stressed the need (i) to perform sensitivities around the choices for input price corrections made and (ii) to include margins of error for the estimated efficiency scores.[161]

Despite this, Sumicsid fails to acknowledge that its choice to ignore input price differences for the vast majority of costs does not solve the issue of which exact price level is appropriate. Not correcting for input prices is simply one assumption of many possible assumptions regarding this issue, and is likely to conflate uncontrollable price differences with inefficiency.

We also note that Sumicsid's position has no empirical support. While Sumicsid appears to suggest that any item that can, in theory, be procured internationally (Sumicsid quotes invoicing, repairs, communication, equipment and installation as mobile goods for instance) has a near-uniform price level, we note that price level differences demonstrably exist even in closely linked economies and for relatively mobile goods. As can be seen in Table 3.3, even those products mentioned by Sumicsid as mobile such as machinery, and even (theoretically) perfectly mobile goods such as software, exhibit differences in prices of more than 20% across the countries included in the TCB18 sample.

[157] Eurostat (2019), 'Statistics explained - Glossary: Price level index (PLI)', available at: https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Price_level_index_%28PLI%29 (last accessed 21 January 2021).
[158] Note that, while the operations are conceptually equivalent, the empirical outcome may depend on the order. This is because the PLI and inflation rate are often measured using different baskets of goods or otherwise track the price of goods imperfectly.
[159] Sumicsid response (2020), para. 3.26.
[160] Oxera (2020), 'A critical assessment of TCB18', April, p. 41.
[161] Ibid.

Table 3.3 Price level differences of mobile goods in the TCB18 sample

Basket of goods	Machinery and equipment	Fabricated metal products and equipment	Software	Total goods	Capital goods
Minimum	93.8	93.4	96.4	80.3	75.6
Maximum	123.8	121.4	118.5	147.1	148.9
Standard deviation	9.1	8.9	6.2	17.7	21.8

Source: Oxera analysis of EUROSTAT data. The sample contains the 15 countries included in TCB18.

Price level differences thus have strong empirical and theoretical backing.[162] Because OPEX as well as CAPEX include activities and materials that differ in price across countries, the PLIs chosen in our central scenario (PLI for overall GDP in OPEX and for Civil engineering projects in CAPEX)[163] were judged to represent the best available and reasonable approximation of actual price level differences. Nevertheless, due to the uncertainties surrounding the precise adjustment the need to perform sensitivities around the choices for input price corrections made and including margins of error for the estimated efficiency scores. The materiality of these margins of error can be seen in Figure 3.3 of our main report. In fact Sumicsid's own analysis shows that the estimated efficiency of one TSO, Statnett, is highly sensitive to the adjustment made and could lie anywhere between 92.9% and 68.3%.[164] Oxera analysis on behalf of Statnett concluded that Statnett's estimated efficiency would become 86.2%, based on the price level adjustment alone.[165]

We however welcome that Sumicsid appears to have some awareness of the inadequacy of its approach, as evidenced by the fact that it repeatedly refers responsibility for price level adjustments to the NRAs.[166]

3.4 Indirect cost allocation mechanism

Like most large businesses, TSOs have overarching support functions (described as 'Indirect Support' by Sumicsid), such as finance, IT support and human resources. Some of these support functions may be directly relevant to specific activities, but others may be sufficiently general that they cannot be allocated to any one activity.

In TCB18, Sumicsid's allocation rule uses the proportion of all costs except depreciation and energy relative to these costs across activities. This includes direct manpower cost, landowner compensation, taxes and levies, insurance, control centre expenses, research and development, office rent, and other expenses. Notably, it thus includes *all taxes and levies* irrespective of whether they are likely to increase indirect expenditure.[167]

Our main report pointed out that this includes material items for some TSOs that bias results. In particular the allocation rule has a significant impact for one

[162] For studies of price levels in closely linked economies, see, for example, Berka, M. and Devereux, M.B. (2010), 'What determines European real exchange rates?', National Bureau of Economic Research; and Engel, C. and Rogers, J.H. (1996), 'How wide is the border?', *The American Economic Review*, **86**:5, pp. 1112–25.
[163] Sumicsid response (2020).
[164] Sumicsid (2020), 'Sensitivity Analysis for TCB18 Results for Statnett', May, Tables 4 and 16.
[165] Oxera (2019), 'Impact of Price Level adjustment in TSO Benchmarking', December.
[166] Ibid., paras 3.28 and 3.31.
[167] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, section 4.97.

TSO, where the inclusion of a large levy in the allocation rule results in only 0.8% of indirect costs being allocated to benchmarked scope activities.

Sumicsid acknowledges that some elements of the allocation key 'could' be excluded but states that the allocation of indirect costs based on taxes is appropriate as it is 'considered as relevant and economically beneficial to the core activity'. It also states that the impact was minimal.

3.4.7 Conceptual errors in Sumicsid's response

The allocation of expenditure to benchmarked activities is a conceptually important issue. While the empirical impact of our correction is relatively small as the affected TSO was not on the frontier in Sumicsid's model, the affected TSO could have been on the frontier if the model or sample was different. Indeed, the ACM has already made some corrections to the TCB18 model relating to the environmental adjustment, and the affected TSO *is* on the frontier in this model. As such, the TSOs for which the affected TSO is a peer will have biased efficiency scores.[168]

Sumicsid states that taxes (and other non-benchmarked activities) were included as drivers of indirect costs as they were 'considered as relevant and economically beneficial to the core activity'. We find this reasoning to be unconvincing. While taxes, or rather the administration of tax and levy claims, is a relevant activity, it is unclear how this corresponds to the scale of tax and levy payments as the driver of indirect costs. Especially in the case of some abnormally large tax payments observed in TCB18, it is unlikely that this payment corresponded to a substantial effort in overarching corporate functions.

The fortunate coincidence that the TSO whose allocation key was clearly and materially biased was not a peer in the TCB18 base model (despite an unjustified reduction in costs by more than 14%) does not absolve the conceptual flaws.

In any future iteration it may be the case, as it was in the ACM's application, that the allocation of indirect costs has a material impact on efficiencies. Therefore, it is essential that in future iterations, unlike what was the case in TCB18, the allocation rule is sufficiently justified, tested and excludes abnormal items that are unlikely to drive indirect expenditure.

[168] ACM's method decision addressed this issue in its earlier proposals, increasing TenneT's estimated efficiency by 7 percentage points.

4 Model development

4.1 Cost driver analysis

Operating, maintaining and enhancing an electricity transmission network is an extremely complex operation. Finding a set of cost drivers that can completely describe the functions of a TSO is therefore a difficult task, and a robust model-development process must be in place to ensure that the results from the empirical analysis are robust.

The model-development process should take into account the operational and economic rationale for including specific cost drivers, as well as their statistical validity. In performing the analysis, the assumptions of any statistical model should be justified and, where possible, empirically tested. The empirical testing of cost drivers using statistical analysis is an essential step in cost driver analysis. It offers a simple and straightforward method for assessing a large number of cost drivers objectively. The assessment is verifiable, easy to understand and interpret, and impartial. It challenges and verifies the essentially subjective assessment of operational factors by experts and insiders. In addition, statistical analysis explicitly considers the available data and can thus determine whether the hypothesised relationship is economically and statistically significant for the given sample. In considering the data, it can also point towards other issues with the data and model (e.g. data errors) when the results of the statistical analysis fail to align with intuition.

Based on our review of published TCB18 materials and the associated workshop slides, we reasoned in our main report that Sumicsid used three statistical estimators to examine the relationship between costs and outputs and Lasso regression to inform model size. We also noted that it was not clear from the outputs which cost drivers had been considered or which other model specifications were considered. Indeed, we highlighted in several instances that the model development process was not clearly explained—we could find no consistent rationale for the procedure in the outputs presented in the final report or its appendices.

Sumicsid now claims in its response report, without quoting any sources or references, that:[169]

> [T]he structure of a good cost model can be derived from some simple principles, proven both from engineering practice, reference network analysis and transmission system benchmarking.

This suggests that the cost driver analysis in TCB18 was limited in scope, and that our critique of the methods that Sumicsid *stated* it had used in the TCB18 exercise may not have been aimed at the core of the issue—namely, that no coherent and comprehensive cost driver analysis was performed. Sumicsid nevertheless responds to several of our specific critiques.

Sumicsid also admits that the calculation of the 'weighted lines' variable contained an error.[170] This calls into question all estimates presented by Sumicsid on the full TCB18 sample as they were performed on demonstrably erroneous data.

[169] Sumicsid response (2020), para. 2.13.
[170] Ibid., para. 3.50.

4.1.1 Non-transparency and inconsistency of Sumicsid's approach

The critique in our main report focused on econometric techniques. This was based on Sumicsid's previous description of its model development approach. In the TCB18 report, Sumicsid states that:[171]

> The process of parameter selection combines engineering and statistical analysis. We have in this project used the following steps:
>
> 1. Definition of parameter candidates. In a first step we established a list of **parameter candidates** which may have an impact on the costs of TSOs.[…]
> 2. Statistical analysis of parameter candidates. […] The main advantage of statistical analysis is that it allows us to explore a large number of candidate parameters and to **evaluate how they individually and in combination allow us to explain as much as possible of the cost variation**.
> 3. Plausibility checks of final parameters. […] [emphasis added]

These statements were also consistent with the material presented as part of the workshops, which emphasised the use of statistical analysis.

Sumicsid now states that no comprehensive model development has taken place on the TCB18 sample, as the cost function is 'well established'.[172] The 'overall techno-economic analysis' that Sumicsid claims to have undertaken is an opaque process and unscientific—the exact process followed, modelling assumptions/alternative options, motivations unpinning particular modelling choices have to be fully documented and subject to external scrutiny. However, it is clear from Sumicsid's description of the process that, if a relevant cost driver did not conform to their 'techno-economic' expectations of 'main sources of costs', it is in fact highly unlikely that it would have even be considered.

For instance, we know that, conceptually, there are many drivers of expenditure that are not properly accounted for in Sumicsid's model (e.g. service quality, asset age, density). We also presented evidence that certain drivers, such as population density and price levels, are insufficiently and/or inappropriately accounted for (both in the main report and follow-on work for several TSOs).[173] Sumicsid's flawed method of cost driver analysis likely resulted in the omission of relevant cost drivers. This could have been avoided if sufficient econometric analysis of cost drivers had been considered.

Sumicsid also defends its lack of transparency by claiming that the final report needs to be 'Readable for a layman'.[174] This seems to unveil a fundamental misunderstanding of TCB18's primary audience. We consider that the primary readers of the study will be TSOs, regulators, academics and others involved in the transmission industry, not 'laymen'. Moreover, it is clear that omitting key information from the report does not serve the goal of 'Readability'. Rather, it makes it impossible for laypeople and experts alike (without access to the data time and resources that we had) to make informed judgements or decisions based on the study.

If Sumicsid considers that its study is too technically complex even for the specialised audience it reaches, then it should outline these technical concepts (e.g. in boxes or annexes) and explain *why* it has followed its approach, rather than leave it up to the reader to have faith in the judgement made.

[171] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators—Main report', July, para. 4.85.
[172] Sumicsid response (2020), para. 2.13.
[173] E.g. Oxera (2020), 'Analysis of TenneT's estimated efficiency under TCB18', August.
[174] Sumicsid response (2020), para. 3.52.

Indeed, if Sumicsid were concerned that results could confuse laypeople, technical descriptions could have been included in a published appendix, not unpublished slides (which are unavailable to non-participants, are outdated as Sumicsid uses erroneous data, and are incomplete as Sumicsid does not show how the final model was selected).

4.1.2 Factual errors and misrepresentations in Sumicsid's response

Suppressing the intercept

In our main report, we highlighted the inconsistency in Sumicsid's returns to scale assumption. Specifically, we stated:[175]

> Sumicsid has not justified why it has forced the intercept in the regression to be equal to zero. **Supressing the intercept assumes that there are no 'fixed costs'** to the operations the transmission network, which in turn assumes that there are constant returns to scale. This is **inconsistent with the non-decreasing returns-to-scale assumption** applied by Sumicsid in its estimation of efficiency scores in the DEA model.

In response, Sumicsid states:[176]

> Contrary to the claim in Oxera, the model specification reported at Workshop 4 (and documented in the slides) show regression results including intercept for all steps of the development.

We note that our main report made no reference to the suppression of intercepts in the model specifications presented in Workshop 4. Indeed, as we were discussing the returns to scale assumption applied in Sumicsid's *final* model, we expected it to be clear that we were discussing the final model in the paragraph above. For clarity, Sumicsid *does* restrict the intercept to equal zero in its final report. **Error! Reference source not found.** below re-presents Table 4.1 from Sumicsid's individual benchmarking reports.

Table 4.1 Sumicsid's final model

	Sumicsid's model
yNG_yArea	0.302***
yTransformers_power	4,196.088***
yLines.share_steel_angle_mesum	16,770.490***
Observations	81
R-squared	0.981
Adj. R-squared	0.980
Residual Std. Error	59,571,597.000 (df=78)

Source: Sumicsid (2019), 'Project TCB18 Individual Benchmarking Report Fingrid – 131', August, Table 3.1.

We can infer that no constant was estimated (i.e. that the intercept was restricted to zero) from two aspects of the table.

- First, if the constant *was* estimated by Sumicsid, it would be common practice to present the value of that estimate in the regression table. If the constant was estimated but not presented in the table (e.g. because Sumicsid considers that it is not a parameter of interest), this should have been clarified in the text of the individual reports.

[175] Oxera (2020), 'A critical assessment of TCB18 electricity', April, p. 49.
[176] Sumicsid response (2020), para. 3.57.

- Second, the presented degrees of freedom (df) are 78, while the presented sample size is 81. This indicates that three parameters were estimated.[177] However, as we can see from the table that three coefficients on cost drivers were estimated, we can infer that no other coefficients (such as the intercept) were estimated.

Importantly, the slides for Workshop 4 do not contain the final TCB18 model, so it is not clear how Sumicsid derived its final model, nor why (and at what stage) it opted to remove the intercept. Note that Sumicsid does not dispute Oxera's conceptual statement that suppressing the intercept in a levels regression is equivalent to assuming constant returns to scale. This is inconsistent with Sumicsid's stated assumption that the sample exhibits increasing returns to scale. This inconsistency if it were correctly identified and addressed could lead to different cost drivers being selected in the cost driver analysis, thus materially affecting the estimated efficiencies.

Modelling in levels

In our main report, we argued that Sumicsid's use of a linear functional form in the model development could be inappropriate. We noted that it is common practice to model in logarithms because it alleviates concerns regarding heteroskedasticity (that can bias statistical inference) and makes the estimated coefficients more easily interpretable. Indeed, given that the statistical analysis indicated that *neither* levels nor logarithms were appropriate, we considered that Sumicsid should have followed precedent and logic to model in logarithms.

In response, Sumicsid states:[178]

> Contrary to the claim of Oxera (2020), the practice of non-parametric benchmarking using levels is very common for regulation. The use of level data lets the DEA formulation determine the exact shape of the production frontier and its response to a change in an input or output at a given point, rather than using the interpolation from a general technical assumption that may not hold at parts of the frontier.

This is not a relevant statement. At no point in Oxera's main report did we argue that DEA models should be estimated using data in logarithms, or any other non-linear transformed data. Our argument was referring to the functional form *assumed in the model development phase.* In this respect, we note that Sumicsid itself used logarithmic models in previous international benchmarking studies. For example, in E2GAS, Sumicsid used logarithmic models to estimate SFA models as a cross-check to its DEA analysis.[179] Indeed, in e3grid2012, Sumicsid's econometric analysis (in logarithms) was used *directly* to inform weight restrictions in its DEA model.[180]

Alternatives to the logarithmic functional form do exist. As highlighted by Sumicsid, some regulators have used normalised linear (i.e. unit cost) models to set efficiency targets.[181] However, it is uncommon to focus on the use of a non-transformed linear functional form in model development, as Sumicsid has done in TCB18. This may be part of the reason why several conceptually important cost drivers such as asset age and population density are not

[177] The difference between the number of observations (81) and the degrees of freedom (78) is the number of parameters estimated—here, 81 - 78 = 3.
[178] Sumicsid response (2020), para. 3.59.
[179] Sumicsid and Swiss Economics (2016), 'Benchmarking European Gas Transmission System Operators FINAL REPORT', June, p. 45.
[180] Consentec, Frontier Economics and Sumicsid (2013),' E3GRID2012 – European TSO Benchmarking Study', July, p. 92.
[181] For example, the UK water regulator, Ofwat, uses unit cost models for residential retail activities. Ofwat (2019), 'PR19 slow track draft determinations: Securing cost efficiency technical appendix'.

considered sufficiently by Sumicsid. As shown in Oxera's work for TenneT, in a logarithmic functional form, these drivers are estimated to be significant drivers of cost.[182]

Sumicsid also alludes to the linear models performing 'better' than the logarithmic models.[183] It is unclear what Sumicsid means by 'better' in this context—it may refer to the average cost regressions with and without the intercept, which may be referring to Table 4.3 of our main report. However, we cannot see how Sumicsid has used the analysis in that table to support the use of levels. In both the linear and logarithmic specifications, the coefficient on one cost driver is negative (in violation of operational expectations) and/or statistically significant. Moreover, statistical tests presented in Table 4.4 on the same page indicate that *neither* logarithmic nor linear models are appropriate.

4.1.3 Conceptual errors in Sumicsid's response

Sumicsid appears to have misunderstood the purpose of Lasso regression. It states:[184]

> the models developed for validation of the minimal necessary size using Lasso regression were presented at Workshop 3, but these were not essential to the further developments of the model

Lasso regression is an algorithmic process for model selection. Starting from a model with a large number of possible cost drivers, the drivers are removed if they do not contribute sufficiently to the model fit. The details of the method are explained in Box 4.1.

[182] Oxera (2020), 'Analysis of TenneT's estimated efficiency under TCB18', August, pp. 23 and 35.
[183] Sumicsid response (2020), para. 3.59.
[184] Sumicsid response (2020), para. 3.52.

Box 4.1 An overview of Lasso regression

Lasso regression is a type of linear regression aimed at reducing model size (i.e. the number of cost drivers). Lasso regression introduces a penalty term for non-zero parameter estimates, which causes cost drivers with small or statistically insignificant coefficients to be set to zero (i.e. excluded from the model). The regression standard model is set up as follows.

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_n x_{ni} + \varepsilon_i$$

where

- y_i is the TOTEX of TSO *i*;
- β_0 is the intercept, which represents the value of *y* when all cost drivers are zero;
- β_j is the cost impact of cost driver *j*;
- x_{ji} is the value of cost driver *j* for TSO *i*;
- ε_i is the random error (residual) for TSO *i*.

OLS estimates the parameters by minimising the sum of squared residuals across TSOs, as follows.

$$\min \sum \varepsilon_i^2$$

Lasso regression extends the usual OLS minimisation problem to include a penalty term for non-zero values of β_j.

$$\min \sum \varepsilon_i^2 + \lambda \sum |\beta_j|$$

where:

- λ is the degree to which the model is penalised for non-zero values of β_j;
- $|\beta_j|$ is the absolute value of the coefficient on cost driver *j*.

A larger penalty term leads to more cost drivers being excluded from the model, relative to OLS regression. The size of the penalty term is typically chosen through a cross-validation procedure (e.g. minimising the mean squared error in a validation dataset). If the penalty term is set to zero, the Lasso regression is equivalent to OLS and no cost drivers are excluded from the model.

Source: Oxera.

Lasso cannot be used to make a *general* observation about optimum model size. Rather, Lasso can be used to narrow down a *specific* set of cost drivers to a smaller subset of the *specific* set of drivers *on the specific dataset* based on a specific criterion.

Lasso is specific to the cost drivers, data and functional form used. Once a model has been selected with Lasso, no parameter can be replaced without re-estimating the Lasso regression (if Lasso is to form part of the evidence). Moreover, Lasso is sensitive to the data used, any changes to the dataset (such as the correction to the weighted lines data) will have an impact on the optimum model size and the parameters selected. However, Sumicsid replaced several cost drivers after performing Lasso, without re-examining the results.

For instance, according to the documentation of Workshop 4, the Lasso selected a model including an output for infrastructure density (yShare.motorways) and the 'minimal model' does not include NormGrid,[185] yet the subsequent models presented do not include infrastructure density in any form. Conversely, the final model includes NormGrid, a variable that was not selected by the Lasso. Note that we do not consider Lasso sufficient as a

[185] Sumicsid (2019), 'CEER-TCB18 project - Model Specification – ELEC, April, slides 61–2.

standalone approach in model development either. If the results from Lasso regression are to be used, however, *all* of the outputs of the approach should be considered in the model development, rather than focusing on a limited subset of supposedly relevant outputs (as Sumicsid has done). The application of Lasso in the current case does not constitute evidence on model size.

Note that the Lasso results presented do not support the restriction to three output parameters in the specific case either. The results presented as part of Workshop 4 show that the decision criterion in fact supports anywhere from three to six parameters.

Sumicsid arbitrarily chooses the lower end of this range and thus risks introducing bias in the estimates as relevant parameters are not included.

The use of OLS-type estimators in model development

Sumicsid argues that the use of the OLS and 'Robust' OLS (ROLS) estimators is justified since the alternative estimation approach (SFA) requires (i) large samples; and (ii) parametric assumptions on the distribution of inefficiency. We have already addressed the 'large sample' issue in our main report, and we elaborate on both issues in section 5.5 of this document. Sumicsid further argues that:[186]

> [T]he model specification for an average cost model should be robust without a frontier model, the opposite would be highly counter intuitive.

In principle, this statement is partially correct. The estimated coefficients in an OLS model will be unbiased, regardless of the shape of the distribution of the residual. However, Sumicsid does not acknowledge that almost all of the statistical inference that it used[187] in the model development phase relies on the residuals being normally distributed, which is not the case if inefficiency is present in the sample. It is inconsistent to assume no inefficiency in that data in the model development process and in a subsequent stage attempt to extract inefficiency in the data. This inconsistency will cause errors in the model development phase and can lead to inappropriate cost drivers being selected. This may be part of the reason why several conceptually important cost drivers such as asset age and population density are not considered sufficiently by Sumicsid. In an SFA model for instance, these drivers come out to be significant drivers of cost.[188]

Sumicsid also argues that an OLS-type estimator is appropriate as Oxera's analysis could find no statistically significant inefficiency (hence, the residuals *are* normally distributed).[189] The observation that no statistically significant inefficiency is present in the model does indicate that the statistical inference can be informative in an OLS context.[190] However, as the statement indicates, it also means that the data and model are consistent with there being no inefficiency in the sample, yet this interpretation is not discussed in Sumicsid's outputs. Moreover, had Sumicsid used SFA to inform its model development procedure, it might have developed a different model that *could* distinguish between statistical noise and inefficiency.

Note that Sumicsid leaves another concern with the ROLS estimator unaddressed. In section 4.1.3 of our main report, we point out that the ROLS

[186] Sumicsid response (2020), para. 3.47.
[187] Given Sumicsid's response to our main report, it is not clear what or whether statistical inference informed the development of the model.
[188] Oxera (2020), 'Analysis of TenneT's estimated efficiency under TCB18', August, pp. 23 and 35.
[189] Oxera (2020), 'A critical assessment of TCB18 electricity', April, section 5.5.
[190] Assuming that other statistical criteria are met.

estimator artificially inflates conventional measures of fit, such as the R-squared, which biases the model fit results displayed by Sumicsid. Thus, the results of a high R-squared presented by Sumicsid in Table 5-3 of the main report is not relevant, as it is driven by the estimator used. This result can therefore not be used to support the current model specification.

Alternative functional form

Sumicsid states that it sees 'no reason' for exploring more complex functional forms as the level specification is a 'good fit' already.[191] However, it does not respond to the specific point raised in our main report—that cost drivers that may fit the data poorly in a linear model may explain the data well in more flexible models. These cost drivers with non-linear relationships with expenditure may have been chosen as cost drivers for the benchmarking model if cost driver analysis was performed in a more appropriate functional form.

Sumicsid dismisses any suggestion of investigating alternative functional forms as it suggests the model could not be validated on the cross section (i.e. the 17 observations in 2017), which was used in DEA. Sumicsid gives no reason why this would be required in the first instance—Sumicsid has used the full panel dataset (81 observations) for its own econometric validation. The need for validation on the cross-section is thus an arbitrary reason presented with no supporting evidence and invalid, even by Sumicsid's own standards.

Considering alternative functional forms such as squared terms in the cost driver analysis may lead to the selection of other cost drivers that in turn will directly affects estimated efficiencies. For example, Ofwat (the UK water regulator) controls for population density and population density squared as cost drivers in its econometric model.[192] However, density is statistically insignificant when included in the model without the squared term, and therefore may not have been selected if functional form was ex-ante restricted.

4.1.4 Material omissions in Sumicsid's response

In our main report we also highlighted that, while the outputs from TCB18 do not enable us to robustly replicate and review the model-development process, we pointed out several issues that lead us to the conclusion that the process was flawed. This was partly based on three issues that were not addressed by Sumicsid's reply:

- the results published in final report (and other associated outputs);
- the unjustified restriction of the model to three outputs;
- the limited consultation with the TSO group on these.

The concerns regarding the lack of engagement with TSOs, unjustified restriction to three outputs, and limited results published remain unaddressed.

4.2 Sensitivity to the sample selected

Sumicsid does not present any analysis in its final outputs that demonstrates how its model reacts to minor changes in the data used in the modelling.

---

[191] Note that, as pointed out in section 4.1.3 of our main report this 'good fit' is partially driven by the use of the ROLS estimator and the suppression of the intercept.
[192] Ofwat (2019), 'PR19 final determinations: Securing cost efficiency technical appendix', December, Table A2.1.

However, it removes one TSO from the sample before estimating efficiency scores because:[193]

> The analyses of the raw data as well as the analysis of a series of model specifications, i.e. models with alternative costs drivers, suggest that one of the 17 TSOs almost always is an extreme outlier.

In our main report, we conducted a robustness check on the econometric results by removing one TSO from the sample at a time (similar to the step taken by Sumicsid). We found that the relationship between the costs and cost drivers in Sumicsid's model is highly dependent on the exact sample of TSOs used and unreliable.[194]

Similarly, we assessed the change in coefficients and estimated efficiencies if the base year of analysis was modified. We found that while the econometric results are more stable over time (than they are when a single TSO's data is excluded), the estimated efficiency scores varied significantly depending on the base year of analysis chosen.

Sumicsid now claims that this robustness check is irrelevant as the TCB18 sample is 'complete' and not drawn from a larger distribution. Sumicsid states:[195]

> TCB18 is the best practice cost function for exactly the TSOs included for the reference year, nothing else.

In addition, Sumicsid claims that this does not impact its view of relevant cost drivers as (according to Sumicsid) cost drivers were selected based on a set of known cost drivers with 'immediate causality', and regression results do not influence this choice. Sumicsid also refers to collinearity between cost drivers as the likely explanation why its hypothesised relationship does not hold in a regression model.

Sumicsid states that volatility of estimated efficiency through time is caused by substantial changes in efficiency year on year. [196]

4.2.5 Non-transparency and inconsistency of Sumicsid's approach

We note that Sumicsid's own statement on the volatility of companies' efficiency year on year precludes the use of TCB18 to set regulatory revenues in any year other than 2017. **Sumicsid explicitly states the following three assumptions in the TCB18 response**.

- The TCB18 model provides a robust estimate of TSOs' efficiency in 2017.[197]
- Actual efficiency can vary significantly, even in short periods of time.[198]
- The TCB18 model is the best practice cost function for *only* the 16 TSOs in the sample in 2017.[199]

Under these assumptions, **it is highly unlikely that a TSO's efficiency in 2020 (or later) corresponds in any way to that estimated by Sumicsid in**

[193] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, p. 35.
[194] Oxera (2020), 'A critical assessment of TCB18 electricity', April, pp. 52–3.
[195] Sumicsid response (2020), para. 5.54.
[196] See Sumicsid response (2020), p. 16.
[197] This is implied by Sumicsid's closing statements in its main report. See Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, section 7.
[198] See Sumicsid response (2020), October, p. 16.
[199] Ibid., p. 14.

TCB18. Therefore, following the logic of Sumicsid's own statements, TCB18 cannot be used to set regulatory revenues for any other period than 2017. Note that we strongly dispute all of the assumptions listed above.

Despite the obvious interest of regulators in setting regulatory revenues based on the international benchmarking (as they did previously), Sumicsid did not discuss these limitations of the TCB18 study in its report or its appendices.[200] Indeed, Sumicsid stated that:[201]

> PSG [the project steering group, consisting of Sumicsid, ACM, BNetzA (German NRA), CNMC (Spanish NRA), NVE (Norwegian NRA) and PUC (Latvian NRA)] cannot think of any reason why these could not be trusted.

Moreover, Sumicsid's statements regarding the unimportance of statistical econometric analyses in developing a robust benchmarking model[202] are inconsistent with the TCB18 material. For example, Sumicsid mentions the following step in its cost driver analysis in its main report:[203]

> Statistical analysis of parameter candidates. […] The main advantage of statistical analysis is that it allows us to explore a large number of candidate parameters and to **evaluate how they individually and in combination allow us to explain as much as possible of the cost variation**. [emphasis added]

We have covered this in detail in section 4.1.1 above.

4.2.6 Conceptual errors in Sumicsid's response

Necessity of robustness to sample

Sumicsid makes several references to TCB18 not being a stochastic exercise, and states that 'the set of TSOs and years are not random from an infinite pool, the study is composed exactly by the participants'.[204] **This is misleading**—the set of participants **is** only a subset of a larger population (all European TSOs, or all TSOs including non-European TSOs), and 2017 is one of many years in which the TSOs have operated. In fact, ACM (a member of the project steering group) is proposing to use data from 2017 to set revenues for 2022–26, almost a full decade after TCB18. While the sample might not be considered 'Random' in a statistical sense, it is certainly a subset of a larger population. There is nothing about this subset that makes it unique, other than the decision to participate in TCB18. In this context, the validation of the analysis on a slightly different subset of TSOs is a **critical robustness test**, especially considering that the data contains material errors (see section 3.1), the model development approach is flawed (see section 4), DEA is deterministic and sensitive to errors and no validation of results has been undertaken (e.g. using SFA or alternative models).

Note that Sumicsid itself alters the sample—for instance, we know that one TSO was removed from the sample because it was deemed a 'persistent outlier'. It seems entirely inconsistent to argue that (i) the sample is complete and cannot be altered and (ii) to then alter the sample anyway. The leave-one-out validation exercise carried out in our main report is a common validation

[200] See Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, section 7.
[201] See Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, section 7.04.
[202] For example, see Sumicsid response (2020), October, para. 3.69.
[203] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators—Main report', July, para. 4.85.
[204] Sumicsid response (2020), October, p. 16.

exercise in econometric and DEA (e.g. super-efficiency assessment that is used in outlier analysis) contexts.

The econometric results are in violation of operational intuition

Sumicsid has built a model with the intention of explaining differences in efficient costs across TSOs, and it stated in the TCB18 report that the cost driver selection was informed by statistical evidence from robust OLS (among others).[205] The fact that this model is highly sensitive to which TSOs are included in the sample implies that the cost drivers do not robustly explain differences in expenditure.

Sumicsid now attributes all issues with the sign, stability and significance of the coefficients to multicollinearity.

While a negative estimated coefficient in an econometric model may be driven, to some extent, by multicollinearity, a negative coefficient could indicate that the characteristic that a cost driver is intended to capture is already captured by the other cost drivers included in the model (at least on average). In this instance, alternative, more appropriate cost drivers could be omitted and the specification must be re-examined.

We note that coefficients on cost drivers being of the right sign in an econometric analysis is generally seen as one of the basic requirements in cost driver analysis, including by Sumicsid in the documentation to TCB18[206] and many other assignments.[207]

Sumicsid acknowledges the issues with cost driver stability but refers to their desire for 'completeness' in cost drivers as the reason for including highly collinear drivers. According to Sumicsid, completeness refers to the following:[208]

> **Completeness** – The output and structural parameters should ideally cover the tasks of the TSOs under consideration as completely as reasonable.

We agree—in principle—that a benchmarking model should be 'complete' as defined above. However, this stated desire is in direct contradiction to Sumicsid's position on the 'minimal necessary size', presented in Workshop three and referred to in the final report, which arbitrarily restricts the number of outputs to three, whereas up to six could have been chosen following Sumicsid's own evidence.[209]

This 'Rule', along with the exclusion of several material cost drivers in the model, such as service quality, asset age and density seems inconsistent with the stated desire for 'completeness'.

Controlling for these drivers as opposed to focusing on those that are highly collinear with the already included drivers would make better use of the (according to Sumicsid) limited model size. Note that we do not agree with Sumicsid's assessment that a strictly limited model size exists (see section

[205] See Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, p. 32.
[206] Sumicsid (2019), 'CEER-TCB18 project Model Specification Model Results', April, slides 42–5; and Sumicsid (2018), 'CEER-TCB18 project Methodological Approach', January, slide 26.
[207] See, for example, Sumicsid and Swiss Economics (2019), 'Effizienzvergleich Verteilernetzbetreiber Strom der dritten Regulierungsperiode', April, p. 59; and Sumicsid and Swiss Economics (2015), 'Project E2gas - Benchmarking European Gas Transmission System Operators', June, p. 3.
[208] See Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, p. 28.
[209] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, p. 35.

4.1.3) as the objective of a cost driver analysis should rather be the identification of well-specified model(s) that is(are) not under- or over-fitted.

While Sumicsid refers to their 'completeness' criteria, it fails to mention another one of its criteria: non-redundancy. Sumicsid defined this as follows:[210]

> **Non-Redundancy** – The parameters should be reduced to the essential aspects, thus avoiding duplication and effects of statistical multi-collinearity and interdependencies that would affect the clear interpretation of results.

Clearly, by Sumicsid's own admission, its output parameters are affected by exactly the statistical multicollinearity and cannot be interpreted clearly in their current form. While we appreciate a desire for completeness, Sumicsid's lack of concern for non-redundancy in this case is inconsistent with the responses that project participants received to their various cost adjustment claims. Here, Sumicsid concluded that diverse factors such as icing, population density, topography and soil conditions were 'captured by correlations to other parameters'[211] without a reference to how their exclusion would affect the 'completeness' of the model.[212]

4.3 Selecting candidate cost drivers

Sumicsid did not consider output parameters and instead restricted itself to using asset-based variables as cost drivers in TCB18. Sumicsid discussed outputs such as energy delivered and peak load in its main report,[213] but it does not present analysis to justify excluding these outputs in the model-development process.

In our main report, we criticised the ex-ante focus on asset-based cost drivers as favouring certain solutions over others. We considered that if asset-based measures are to be used in benchmarking, these must be validated through comparisons to pure outputs.

We stated our concern that close substitutes of the used cost drivers, such as substituting transformer power with circuit end power, lead to substantially different estimated efficiencies.

4.3.7 Non-transparency and inconsistency of Sumicsid's approach

Sumicsid opportunistically switches its criteria for cost driver selection. In the TCB18 main report, the criteria were listed as:[214]

- exogenous;
- complete;
- operable (feasible);
- non-redundant.

In its reply to our main report, Sumicsid now states that the specification should be:[215]

---

[210] Ibid.
[211] Ibid., p.13.
[212] Ibid., p. 28.
[213] Ibid., section 4.81.
[214] See Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, p. 35.
[215] Sumicsid response (2020), p. 18.

- deterministic;
- complete;
- feasible;
- verifiable.

Two criteria (feasibility and completeness) are unchanged, but the inclusion of the 'deterministic' criterion in particular appears opportunistic. While it is not clear what Sumicsid considers 'deterministic' data (as the term is not clearly defined anywhere in the TCB18 response), the term seems to have been used as a contrast to 'stochastic' data that appears to be defined by Sumicsid as variables fluctuating over time depending on external (outside of the control of a TSO) factors.[216]

This new restriction (which was not discussed or even mentioned in the TCB18 report or its appendices) appears to exclude a whole class of exogenous drivers of TSOs' costs such as peak load, energy not supplied, service quality measures such as supply interruptions and many others.

The arbitrariness of this criterion is reinforced by the fact that real-world data, including that used by Sumicsid, is inherently noisy and subject to data issues from misreporting, measurement error or data manipulation errors. At the scale and complexity of a project such as TCB18, no 'deterministic', noise-free data is likely to exist. This is covered in more detail in section 3.1.

In addition, this criterion completely ignores the input side of the benchmarking equation. The input (TOTEX) is subject to stochastic year-on-year fluctuations caused (at least in part) by the 'stochastic' cost drivers. By ignoring the 'stochastic' cost drivers, Sumicsid conflates the entirety of this random fluctuation with inefficiency.

This general observation can also be applied to many of the specific examples of possible cost drivers mentioned in our main report. For instance, Sumicsid claims that including energy not supplied as a driver may penalise operators in a rural setting (without considering that operating environment could be controlled for), but *not* including energy not supplied conversely disadvantages those TSOs that have to spend significant resources maintaining their line network to make it robust against failures and favours those that do not.

4.3.8 Material omissions in Sumicsid's response

Sumicsid does not address the general concern with asset-based cost drivers in benchmarking, namely that they artificially skew the benchmark towards asset based solutions. In our main report, we offered an illustrative example of TSOs reacting to a more volatile energy supply with two possible solutions—one that involves investment in assets to increase network capacity, and one that involves investment in software to manage energy supply and demand. The former solution will materially increase the size of the asset base and the latter will not.

A TSO choosing the latter solution will therefore be disadvantaged in a model that controls for NormGrid, regardless of the 'true' efficiency of the investment.

Sumicsid does not address this point in its response.

[216] For example, see Sumicsid response (2020), paras 2.17, 2.23 and 3.97.

4.3.9 Conceptual errors in Sumicsid's response

The idea of using circuit end power as an alternative proxy for capacity stems from the issue of the TSO–DSO distinction being inconsistent across the sample.[217] This induces a bias against certain TSOs, which is covered in detail in our individual assessment of TenneT's efficiency.[218] Circuit end power was tested as one of several alternative proxies as it is more consistently associated with grid assets than Sumicsid's transformer power variable, and does not suffer from the same issues regarding ownership structure as transformers.

Importantly, circuit end power was presented as an example of how using a close substitute for the cost drivers included in Sumicsid's model has a material impact on TSOs' efficiency scores. Sumicsid does not address the conceptual argument that alternative asset-based proxies exist and can lead to different outcomes for TSOs.

Note that Sumicsid now raises the comparability of the circuit ends data. However, this variable uses the same data that Sumicsid considered to be appropriate for a calculation of NormGrid (which also uses both power and voltage of circuit ends as inputs). If there are errors or uncertainties in the circuit end data, then this should have been acknowledged and addressed in the TCB18 report.

4.4 Aggregation of NormGrid

NormGrid is defined as the weighted sum of different asset classes. Sumicsid chooses the weights on assets based on an average cost estimate of each asset.[219] Specifically Sumicsid states that it used linear regression to inform the aggregation weights.[220]

> The calibration of the [NormGrid] asset weight systems is made **through linear regression** towards the **Capex and Opex data obtained in the project**. [emphasis added]

As different TSOs deploy assets in different proportions to each other, the choice of aggregation weights impacts estimated efficiency. Thus, we stated in our main report that: [221]

- the weights used in aggregation must be robustly justified;
- the uncertainty within the estimation of the weights must be accounted for.

We presented a regression estimate of NormGrid weights to calculate appropriate standard errors for the weights and efficiency scores.[222]

Now, Sumicsid states that:[223]

> The NormGrid is not to be estimated from the TCB18 data, but from engineering expertise.

[217] Sumicsid states that it is unclear whether the circuit end power was corrected for ownership of substations. We can confirm that it was corrected for this.
[218] We also present statistical evidence in the form of results of Sumicsid's model estimated using SFA, replacing Transformer power with circuit end power. We find that the coefficient on circuit end power is positive and significant. Oxera (2020), 'Analysis of TenneT's estimated efficiency under TCB18', August.
[219] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, section 4.72, p. 25.
[220] Sumicsid (2019), 'Norm Grid Development – Technical report', p. 2.
[221] Oxera (2020), 'A critical assessment of TCB18 electricity', April, p. 58.
[222] Ibid., pp. 58–9.
[223] Sumicsid response (2020), para. 3.86.

It justifies this by citing collinearity and bias from inefficient operators as possible issues in a regression.[224]

In our main report, we also suggested that it may be more appropriate to not combine asset classes and include them as separate outputs in the DEA model so that the weights are chosen endogenously.[225] Sumicsid is dismissive of this approach proposed in its TCB18 response, stating that 'aggregation is inevitable' and that 'absurd dual weights' would force weight restrictions that would mean 'a return to the original expert assessment'.[226] Sumicsid go as far as saying:[227]

> An output specification using open dual weights has no merit.

4.4.10 Material omissions in Sumicsid's response

Sumicsid does not respond to either of the two conceptual points raised in our main report. The lack of documentation for the process is not addressed at all and Sumicsid never acknowledges the need for a margin of error. Instead, it responds to the specific method used in our main report.

4.4.11 Factual errors and misrepresentations

In order to investigate the justification for Sumicsid's aggregation weights, we chose to estimate a linear regression. This was informed a statement from Sumicsid's main report in which it stated that:[228]

> The calibration of the asset weight systems is made through linear regression towards the Capex and Opex data obtained in the project.

Based on this, we concluded, as any reasonable reader would have, that Sumicsid estimated a linear regression to derive the relative asset weights before aggregating. Sumicsid now directly contradicts this statement and claims that 'engineering expertise' was used instead.[229] It claims that estimating a regression to derive the weights is inappropriate, citing collinearity and inefficiency in the sample as the reasons for this. This directly contradicts its own approach from the original report.

While this contradiction is emblematic of an opaque process, the issues identified in our main report are relevant regardless of whether the weighted were derived through regression analysis, engineering assessments, or any other method. The weights require a **robustly justified** estimation that in turn requires a reasonable margin of error.

The approach suggested in our main report was simply the one that was consistent with the approach taken by Sumicsid (linear regression). Measures of uncertainty are built into statistical methods of weight-derivation (e.g. the confidence intervals on coefficient estimates), as demonstrated in our main report. However, expert judgement is not without uncertainty—there will always be some disagreement in any panel of experts regarding the exact cost weight to attach to certain assets, especially if these cost weights are not informed by any quantitative analysis. Indeed, we would expect even an individual expert to be able to provide a feasible range.

[224] Ibid., para. 3.84.
[225] Oxera (2020), 'A critical assessment of TCB18 electricity', April, p. 59.
[226] Sumicsid response (2020), para. 3.88.
[227] Ibid., para. 3.89.
[228] Sumicsid (2019), 'Norm Grid Development – Technical report', p. 2.
[229] Sumicsid response (2020), para. 3.86.

Sumicsid does not dispute the necessity of a margin of error in its reply and in fact did present a basic version of this same sensitivity (covering only lines or transformers holding all else constant) in the main report itself.[230] However, Sumicsid misrepresents Oxera's position on the materiality of this issue. Sumicsid states that:[231]

> As Oxera (2020) notes, the sensitivity with respect to the relative weights between and within asset classes is very low.

While NormGrid is not an important driver of costs for many TSOs (see section 5.3 of our main report) and the estimated efficiencies of those TSOs are not very sensitive to this change, the impact on several TSO's estimated efficiency is substantial. As our main report notes:[232]

> the impact is highly material for one TSO, where the confidence interval on the NormGrid weights leads to a range of efficiency estimates from 65% to 94%.

As presented in Figure 4.5 of our main report, the issue is not limited to this operator either. Depending on the exact weight used, four additional TSO's efficiency estimate changes by at least eight percentage points. We do not consider these variances 'low' in the context of setting regulatory revenues and maintain the need for an appropriate margin of error.

4.4.12 Conceptual errors in Sumicsid's response

The TCB18 response devotes a paragraph to the issue of aggregation in benchmarking in general—we never disputed that aggregation of some form is necessary in benchmarking (one could argue that the very concept of aggregating a set of inputs and outputs into an efficiency score itself constitutes aggregation). In fact, both approaches proposed in our main report (regression based weights and keeping asset classes as separate outputs in the DEA) are alternative forms of aggregation. They are, however, ones that require less stringent assumptions and recognise that trade-offs can be flexible.

We note that the model presented in our main report that controlled for the four largest components of NormGrid was presented to demonstrate how the efficiency scores change if TSOs are not benchmarked against TSOs that deploy a substantially different mix of outputs, rather than providing a fully developed standalone benchmarking model. The fact that doing so had a material impact on TSOs' efficiency scores indicated that TSOs were benchmarked against TSOs that deployed materially different assets. Therefore, the solutions available to the 'efficient' TSOs (as determined by the TCB18 model) may not be appropriate for the 'inefficient' TSOs. For instance, a TSO may be required to use underground cables instead of overhead lines in densely populated areas. When it is now compared to a TSO mainly deploying overhead lines this solution is not available and the cost reduction target based on a NormGrid approach is thus not feasible.

The TCB18 response to this suggestion focuses on the hypothesised wide range of dual weights in this model. Sumicsid goes as far as saying:[233]

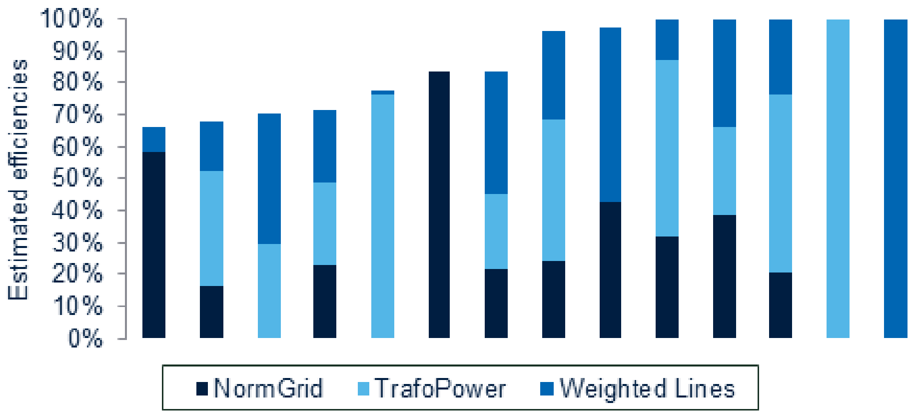
> An output specification using open dual weights has no merit.

---

[230] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, pp. 39–40.
[231] Sumicsid response (2020), para. 3.85.
[232] Oxera (2020), 'A critical assessment of TCB18 electricity', April, p. 59.
[233] Sumicsid response (2020), para. 3.89.

For Sumicsid's model (as well as the model controlling for different asset types separately) the dual weights are also not restricted, meaning that they can take any positive value (regardless of other dual weights). They are not even discussed in the main report, the associated appendices or workshop slides. In fact, as can be seen in Figure 5.4 of our main report, reproduced below as Figure 4.1, Sumicsid's model exhibits a similarly wide range of dual weights as well. For instance, four TSOs do not place any weight on NormGrid, according to Sumicsid the 'primary' cost driver of TSOs, at all.[234] Similar observations hold for the other two cost drivers as well.

Figure 4.1 Weights of outputs in Sumicsid's DEA



Note: This figure does not include the weights for outliers that were removed from the final sample. The weights for fully efficient units are not unique.

Source: Oxera analysis.

The concern that Sumicsid shows for this is in direct contrast to the lack of concern for the same matter it displays in the assessment of its own chosen model. It is unclear why Sumicsid considers it 'without merit' to have an output specification with unrestricted dual weights in one case, but not the other, as they are very similar conceptually.

Sumicsid also once again ignores any possible compromise between aggregation and flexibility. Sumicsid states:[235]

> Obviously, leaving the asset classes as separate outputs leads to absurd dual weights in DEA, a problem that in reality would have to be dealt with using e.g. weight restrictions – i.e. a return to the original expert assessment of the relative weighting.

In contrast to what is stated in the TCB18 reply, a restriction of dual weights would not automatically lead to fixed weights as in Sumicsid's approach. Rather, it could result in a reasonable range of dual weights being admissible, resulting in a margin for error in aggregation. This margin of error in the DEA could be informed by our understanding of the margin of error in the estimation of the relative aggregation weights. In effect, this would result in a similar solution to the one proposed in section 4.4.3 of our main report, with each TSO choosing the optimal dual weighting for it (i.e. the upper end of the confidence interval).

[234] Oxera (2020), 'A critical assessment of TCB18 electricity', April, p. 71.
[235] Sumicsid response (2020), para. 3.88.

4.5 Adjusting for environmental factors

Sumicsid agrees that environmental complexity is a key cost driver for electricity TSOs.[236] In TCB18, Sumicsid used factors ranging from 1 (e.g. grassland) to 3.5 (Infrastructure) and multiplied those by the percentage of land used by this feature in an operators operating country (or operating region for the sub-national TSOs National Grid, Scottish Power, SHE-T and TenneT Germany) to arrive at a single 'complexity weight', which was then multiplied by NormGrid.

In our main report, we criticised this approach on several counts.

- We were unable to find a source or convincing narrative supporting the environmental weights used by Sumicsid in any of its outputs.
- There is a lack of correlation between the environmental adjustment and unit costs.
- The biggest driver of most TSOs' environmental adjustment is the proportion of their service area that is covered in forests (as forests cover a large percentage of area), and more intuitive drivers of expenditure, such as urbanity, were given inadequate weight (as they cover a more concentrated area).
- The environmental factor adjustment to NormGrid does not take the distribution of assets into account.
- The impact of taking into account environmental factors is negligible, contrary to reasonable expectation.

Our main report suggested that a more appropriate way of correcting for environmental factors would use the available data to estimate the cost impact of environmental conditions on the available data or to treat them as exogenous drivers of expenditure in a DEA model.

In response to our concern regarding the non-intuitive impact of environmental conditions on the benchmark, Sumicsid states that the weights used were based on 'independent expert assessments, as used in any techno-economic study'.[237] Sumicsid also states that the lack of correlation between unit costs and the environmental complexity weight is explained by 'other outputs and the level of efficiency' of the TSOs.[238]

Sumicsid refers to sensitivity analysis, which was not presented, showing an unquantified 'Relatively low' sensitivity of efficiencies to the environmental adjustment.[239]

4.5.13 Factual errors and misrepresentations

Sumicsid alleges that Oxera would indiscriminately regress environmental factors on costs in order to completely absorb any variability in costs. Specifically, Sumicsid states:[240]

> Another approach, **advocated by Oxera**, would be to let the environmental factors absorb any variability in cost at the average cost function stage. For

[236] Ibid., pp. 6–7.
[237] Sumicsid response (2020), para. 3.94.
[238] Ibid., para. 3.93.
[239] In a subsequent report, Sumicsid quantifies the possible change in the base model at 5.8%. Sumicsid (2021), 'Response to the Oxera Report on TenneT TSO in TCB18-ETSO', April, p. 19.
[240] Ibid., para. 2.28.

> example, if particularly inefficient operators have more railroad per surface area, a regression would suggest that the cost should be attributed to the railroads rather than to inefficiency.

This assertion is false on several accounts. First and most obviously, Oxera never suggested an indiscriminate approach in which non-intuitive variables are added to a regression without reference to operational intuition. Specifically, our main report states:[241]

> The model-development process should take into account both the **operational and economic rationale** for including specific cost drivers, as well as their **statistical validity**. [emphasis added]

If the statistical analysis demonstrates that the relationships in the data are completely counter to what is expected from operational expectations, then this should be documented and explored, rather than ignored (as Sumicsid has done).

Second, the source of our critique on environmental adjustments in particular builds on operational rationale with the correction factor for forested areas appearing especially excessive. At no point within our main report or any associated work do we advocate the use of or even present results of a regression without considering the operational rationale.

4.5.14 Material omissions in Sumicsid's response

Sumicsid only explicitly addresses one of the four points made in our main report, claiming that the lack of correlation between unit costs and the environmental complexity weight is explained by 'other outputs and the level of efficiency' of the TSOs.

We note that the regression result that demonstrates a negative correlation between unit costs and environmental complexity as defined by Sumicsid was designed as a basic cross check, not as the basis of an adjustment. However as the simple cross check was unsuccessful, we explored the construction of the adjustment and offered two possible explanations for the result:

- that the biggest driver of most TSOs' environmental adjustment is the proportion of their service area covered in forests and more intuitive drivers of expenditure, such as urbanity (which even according to Sumicsid's interim results[242] and Sumicsid in e3grid are relevant drivers of expenditure) were given little weight;
- that the environmental factor adjustment to NormGrid does not take the distribution of assets into account.

Sumicsid does not address these points in its reply at all. Sumicsid also does not address our concern at the minor impact environmental adjustments have in TCB18, which are likely a result of the correction approach.[243]

4.5.15 Non-transparency and inconsistency of Sumicsid's approach

Sumicsid did not provide any relevant documentation of the 'independent expert assessment' on environmental cost factors in electricity transmission that was supposedly undertaken in any of the outputs reviewed by Oxera. In fact, despite an earlier section promising that the appendix would cover the

[241] Oxera (2020), 'A critical assessment of TCB18 electricity', April, p. 45.
[242] Sumicsid (2018), 'CEER-TCB18 project Model Specification ELEC', February, slide 67.
[243] Oxera (2020), 'A critical assessment of TCB18 electricity', April, p. 62.

'candidate variables and hypothesised cost impact',[244] the appendices to the TCB18 main report simply note that:[245]

> [T]he relevant factors are a subset of those already described for gas transmission pipeline construction.

Sumicsid fails to specify the subset of factors that were considered 'Relevant', what the cost weight is, or how it was derived.

The same document also suggested that an econometric assessment of environmental factors would be undertaken. In the appendices to the main report, Sumicsid still states that participants would receive the following:[246]

> Estimation results from the econometric team for the candidate variables individually, as well as the retained factors with their numerical estimates and possible intervals of uncertainty.

In its TCB18 response, Sumicsid now also confirms that no calibration was undertaken:[247]

> [T]here was no direct calibration of the TCB18-data towards the environmental complexity factors.

This appears to be a deviation from the project plan that was not elaborated on in the TCB18 report or its appendices.

So while the initial plan appears to have been a two-pronged approach, covering both the bottom-up engineering assessment and top-down econometric assessment, the former approach was not documented at all while the latter appears to have been abandoned.

Estimating cost impacts of environmental factors without any element of validation that would ensure that the derived weights are relevant and appropriate to the sample used and take relevant context into account is an ambitious task. It is not only an uncertain exercise, but risks that oversights and disconnects between the different teams are not discovered and remedied. This appears to have been an issue in TCB18, as the weight placed on the area covered in forests appears excessive when compared to more intuitive drivers of expenditure, such as urbanity and mountainous areas, as outlined in our main report.[248] This may have, for instance, occurred because the engineering assessment is concerned with the cost impact of an individual asset's location on this individual asset's cost, not overall adjustment of an aggregate cost base.[249]

Yet Sumicsid applied these factors to a country-wide land use cover, not linked to the location of assets within that country. Thus, even if the cost impact of the individual assets was estimated accurately the overall adjustment applied by Sumicsid is biased towards categories such as forests as they can cover vast, unserved areas.

We are conscious that mapping the grid of each TSO is not a straightforward task—however, if Sumicsid wanted to apply these factors directly to the assets

[244] Sumicsid (2019), 'Norm Grid Development', February, p. 3.
[245] Ibid., p. 40.
[246] Ibid., p. 4.
[247] Sumicsid response (2020), para. 3.93.
[248] Oxera (2020), 'A critical assessment of TCB18 electricity', April, p. 61.
[249] For instance, the appendix on environmental factors in gas states that safety regulations and thus the price of a pipeline depend on the 'urbanisation degree and population density in the immediate vicinity of the narrow corridor within which the pipeline is constructed.' Sumicsid (2019), 'Norm Grid Development', February, p. 22.

themselves, this degree of granularity is required for an accurate assessment. In the absence of such detailed data, the use of assumptions may alleviate the issue to a degree. For instance, it is likely that infrastructure such as power lines are concentrated close to population centres, necessitating a higher weight on urban areas. Such assumptions (much like the weights themselves) will never be fully accurate, but could be validated through top-down analysis of the kind presented in our main report.

Nevertheless, as Sumicsid did not provide the relevant materials, it is not possible to reconstruct where exactly in the development of the weights Sumicsid made an error.

As outlined in our main report, there are several proven and pragmatic alternatives to this restrictive approach taken by Sumicsid. For instance, Sumicsid could have included density or mountainous areas as exogenous drivers of expenditure (which can be explicitly accounted for) in the DEA model.[250]

[250] Oxera (2020), 'A critical assessment of TCB18 electricity', April, p. 62.

5 Model assumptions

5.1 Returns to scale

'Returns to scale' relates to how changes in outputs (e.g. NormGrid) are linked to associated changes in inputs (e.g. TOTEX) for efficient companies. In applications of DEA, typical assumptions include constant returns to scale (CRS) or variable returns to scale (VRS).[251] Sumicsid used a non-decreasing returns to scale assumption (NDRS) that is a mixture of CRS and VRS assumptions. In the final report for TCB18, Sumicsid argued that it had used the following statistical analysis to support this assumption.

- **A Banker F-test for returns to scale on the DEA efficiencies.** In this test, the efficiencies estimated under CRS are compared to efficiencies under alternative RTS assumptions. The test statistic is similar to that used in the dominance test (outlined in Box 2.3 of our main report), and compared to an F-distribution. If the test statistic is statistically significant, this is taken as evidence for the alternative (non-CRS) assumption.
- **The sum of coefficients in a log-linear regression.** If the sum of coefficients is less than one, adding 1% to every output increases costs by less than 1%, indicating increasing (or non-decreasing) returns to scale.[252]

However, unlike in the predecessor studies to TCB18,[253] Sumicsid did not present the results of these statistical tests in the main report, and we could find no evidence for such an assumption in our replication of TCB18. Indeed, we found that the statistical evidence was inconclusive, but supported either a CRS, VRS or a DRS returns to scale assumption.[254] Importantly, not a single statistical test that we conducted supported Sumicsid's returns to scale assumption. In response to Oxera's main report, Sumicsid appears to have repositioned its justification for the NDRS assumption. Sumicsid does not explicitly dispute Oxera's findings that the statistical tests that Sumicsid stated it had conducted in its final report do not support its returns to scale assumption. Instead, Sumicsid uses three academic arguments.

1. VRS is rejected by all published studies relating to estimating production functions in the energy industry.
2. There are specific regulatory arguments for applying an NDRS assumption, as small TSOs cannot readily increase the scale of their operations if they are bounded by regulatory borders.
3. There is no techno-economic reason for large TSOs to be less productive than a combination of smaller TSOs.[255]

[251] Under CRS, it is assumed that, when outputs rise or fall by a certain amount (say 5%), efficient companies would be expected to increase or decrease their cost by that same amount (5%). In contrast, under VRS, when outputs rise or fall by a certain amount (say 5% again), efficient costs can rise or fall by a percentage greater than, less than, or equal to 5%, depending on whether decreasing, increasing or constant returns to scale are expected to prevail.
[252] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, p. 35.
[253] E.g. Sumicsid and Swiss Economics (2015), 'PROJECT E2GAS Benchmarking European Gas Transmission System Operators – Final report', Figure 5-3; Consentec, Frontier Economics and Sumicsid (2013), 'E3GRID2012 – European TSO Benchmarking Study', p. 78.
[254] Oxera (2020), 'A critical assessment of TCB18', April, section 5.1.3.
[255] Sumicsid response (2020), para. 3.107.

5.1.1 Material omissions in Sumicsid's response

Sumicsid states that its returns to scale assumption is based on 'econometric tests, observations from the empirical distribution of efficiency and techno-economic considerations from the model specification'.[256] However, Sumicsid does not present any of the econometric tests, or the observations from the empirical distribution of efficiency, to support its argument. In the face of the statistical analysis presented in Oxera's main report, we consider that it is essential that Sumicsid presents its own analysis so that the parties can investigate the causes of the contradictory findings.

5.1.2 Factual errors and misrepresentations

Sumicsid presents a list of seven academic studies that, it states, supports the presence of NDRS technology in the electricity sector. However, of the seven studies, only two actually supported Sumicsid's claim, as shown below.

Nerlow (1961)

Sumicsid states that this paper presented evidence of increasing and constant returns to scale in electricity. However, this is a narrow, incomplete and misleading interpretation of Nerlow (1961). Indeed, the paper found IRS at small plant sizes, but the returns to scale gradually reduced until it was DRS for the largest firms.[257] This is consistent with a VRS assumption, not the NDRS assumption applied by Sumicsid.

Christensen and Greene (1976)

Sumicsid states that this study shows that most operators operate under IRS or CRS. This is a factual statement by Sumicsid—the *majority* of the firms examined in the Christensen and Greene (1976) paper did indeed operate under IRS or CRS. However, the study also showed that one firm in the sample operated under statistically significant DRS,[258] thus supporting the use of a VRS assumption. Note that the study is an assessment of electricity generation, rather than transmission. One would expect the input/output correspondence to differ across generation and transmission, meaning that the exact point of the optimal scale size will likely differ.

Sumicsid's characterisation of this study as supporting the use of NDRS is inappropriate.

Yatchew (2000)

Sumicsid states that Yatchew (2000) finds increasing returns to scale for smaller utilities. Again, Sumicsid is correct that the smaller utilities in the sample exhibit IRS, but Sumicsid has omitted the observation in Yatchew (2000) that large firms exhibit DRS:[259]

> The data reveal substantial evidence of increasing returns to scale with minimum efficient scale being achieved by firms with about 20,000 customers. **Larger firms exhibit constant or decreasing returns**. [emphasis added]

Moreover, Yatchew (2000) compares its analysis to similar studies that have also estimated the existence of VRS technology in the electricity sector,

[256] Ibid., para. 3.100.
[257] Nerlow (1961), 'Returns to Scale in Electricity Supply', Table 5 and Figures 2–3.
[258] Christensen and Greene (1976), 'Economies of scale in U.S. Electric Power Generation ', Table 10 and Figure 3.
[259] Yatchew (2000), 'SCALE ECONOMIES IN ELECTRICITY DISTRIBUTION: A SEMIPARAMETRIC ANALYSIS', p. 188.

indicating that the author believes a VRS assumption is in line with the academic literature. For Sumicsid to characterise this paper as supporting an NDRS assumption is clearly erroneous.

Filippini et al. (2004)

This is one of the two references cited by Sumicsid that provides any support for the application of an NDRS assumption. However, we note that the study imposes strict assumptions on the relationship between cost and outputs such that VRS *could never* be identified. Indeed, the only tests that can be conducted in the model estimated in Filippini et al. (2004) is between IRS, CRS and DRS. Filippini et al. (2004) acknowledges this limitation in its study:[260]

> The translog cost function offers an appropriate functional form for answering questions about economies of scale. However, the small data set available for this study does not allow the use of this functional form, which is characterised by the high number of coefficients to be estimated.

Therefore, the use of this study is of limited practical use in the context of Sumicsid's TCB18 response that rejected the use of VRS.

Filippini and Wild (2001)

Sumicsid states that Filippini and Wild (2001) finds IRS in Swiss electricity distribution. This is a misleading statement, as Filippini and Wild (2001) finds that there is IRS *for the average* electricity distribution company. Indeed, Filippini and Wild (2001) estimates a positive and statistically significant coefficient on the squared output term,[261] indicating that the largest firms operate under DRS technology.[262] That is, Filippini and Wild (2001) shows that the Swiss electricity distribution sector is characterised by VRS technology.

Dismukes et al. (1998)

Dismukes et al. (1998) supports Sumicsid's position on the use of NDRS. It finds increasing returns to scale in a translog model covering the US transmission industry in 1986–91. However, this research is dated and pre-dates the formation of stand-alone TSOs and may thus be less relevant to the current case.[263]

Llorca et al. (2016)

Sumicsid states that Llorca et al. (2016) finds IRS technology in the US transmission sector. Again, this is highly misleading—Llorca et al. (2016) estimates IRS for approximately two-thirds of the sample, and DRS for the remaining third.[264] This would support the use of a VRS technology.

Overview of academic literature

We expect that there *are* academic studies in the public domain that would support the use of an NDRS assumption, thereby supporting Sumicsid's modelling assumption. However, it is telling that the majority of the studies cited by Sumicsid support the use of a VRS assumption, contrary to Sumicsid's assertions on the topic. Nevertheless, Sumicsid should have validated its

[260] Filippini et al. (2004), 'EFFICIENCY AND REGULATION OF THE SLOVENIAN ELECTRICITY DISTRIBUTION COMPANIES', p. 12.
[261] Filippini and Wild (2001), 'Regional Differences in Electricity Distribution Costs and their Consequences for Yardstick Regulation of Access Prices', Table 2.
[262] Ibid., Figure 1.
[263] Dismukes et al. (1998), 'Capacity and economies of scale in electric power transmission', p. 1.
[264] Llorca et a. (2016), 'EFFICIENCY AND ENVIRONMENTAL FACTORS IN THE US ELECTRICITY TRANSMISSION INDUSTRY', February, Figure 6.

modelling assumptions on the data at hand—if there is a strong ex ante justification for imposing the NDRS assumption (which is not the case here, as demonstrated above) then the supposedly counterintuitive statistical analysis on the data should be investigated and explained, rather than simply ignored (as Sumicsid has done).

5.1.3 Conceptual errors in Sumicsid's response

Sumicsid's final argument for the use of an NDRS assumption rests on its 'techno-economic' evidence. That is, Sumicsid asserts that 'no valid reason can be found to suggest that larger operators would be unable to organise their services and assets as efficiently as smaller units'.[265] The policy recommendations from such an approach is, if a TSO is too large, then it should be separated into smaller operating networks (or management units).

For these productivity gains to be made, either the large TSO needs to be reduced to the most productive scale size *for its mix of outputs*, or split up into several TSOs, each one enjoying the most productive scale size. Even ignoring the substantial reorganisation costs involved in this, neither approach is feasible in practice. Moreover, breaking up the large TSO cannot guarantee that any one of the resulting smaller TSOs will necessarily have the same mix of outputs (terrain, NormGrid, connections, etc.) as the large TSO but at the most productive scale size. In fact, it is possible that any feasible subdivision of the TSO will itself operate at a non-optimal scale.

Moreover, if Sumicsid insists on using an NDRS assumption, it must address the question of why large TSOs tend to be estimated to be inefficient (which causes the statistical tests outlined above to conclude that a VRS technology is appropriate). One hypothesis (implicitly assumed by Sumicsid) is that larger TSOs happen to be inefficient by chance in this sample, and that a larger or different sample would show a different relationship between size and productivity. However, this cannot be assumed without evidence, as Sumicsid has done. An alternative hypothesis could be that Sumicsid has omitted key drivers of expenditure that happen to be correlated with scale in the sample, causing a bias in the estimated efficiency scores of large TSOs. For example, several of the largest TSOs in both gas and electricity happen to operate in relatively densely populated countries (NGET and TenneT in electricity; NGGT and GTS in gas). Additionally, it may be the case that TSOs that have accumulated a number of assets over time (thereby being 'large') now operate a relatively old grid. As Sumicsid's model does not control for asset age, this could bias the analysis against large TSOs.[266]

In any case, if the observed relationship between cost and output is not consistent with Sumicsid's operational expectations (that are themselves not justified in its report), Sumicsid should have investigated and documented the causes of such inconsistencies instead of ignoring the problem (as it has done in this case).

Note that we agree with Sumicsid's observation that VRS should be the default approach in DEA:[267]

> The default approach in DEA should be to impose only the minimally necessary assumptions, to avoid a priori influence on the data and the efficiency

[265] Sumicsid response (2020), para. 3.106.
[266] In previous analysis, we found that omitting the impact of asset age on costs led to an underestimation of TenneT's efficiency. See Oxera (2020), 'Analysis of TenneT's estimated efficiency under TCB18', August, section 3.
[267] Sumicsid response (2020), para. 3.101.

assessment. In this case, it would correspond to variable returns to scale (VRS), meaning that there could be both increasing and decreasing returns to scale in the data set.

As demonstrated throughout this section, Sumicsid's justifications for applying an NDRS assumption are inherently flawed. Neither statistical evidence, nor the academic literature that Sumicsid cites supports the NDRS assumption. Instead, Sumicsid relies on unsubstantiated claims of 'techno-economic analysis' to support deviating from VRS.

If the regulator has an explicit objective for companies to operate at the optimal scale size (e.g. through mergers and demergers), then it may be appropriate to assume a CRS technology in the benchmarking model, even if the 'true' technology is VRS.[268] For example, the Norwegian energy regulatory framework ensures that the distribution system operators (DSOs) are incentivised to undertake structural reforms and are compensated if their efficiency position in the benchmarking model worsens due to any such reform. This policy perspective, however, is not applicable in the TCB18 assessment, which is a cross-country comparison. The scale of a TSO, particularly those that cover their entire country, is not within management control (as was also noted by Economic Insights Ltd in a report for the ACM).[269,270] Thus, if the true technology is VRS,[271] there is no valid reason to assume any other technology.

5.2 Outlier analysis

The DEA model, as applied by Sumicsid, determines the frontier based on those TSOs deemed to be efficient without making any allowance for errors. Thus, the estimation of an efficient frontier is greatly influenced by the existence of outliers, especially in deterministic methods of benchmarking such as DEA.

Sumicsid follows the same outlier procedure as the Bundesnetzagentur, which is an approach that was developed to fall in line with the legal requirements of the ARegV. In addition to the formal outlier tests outlined in the ARegV, Sumicsid removes one TSO from the sample ex ante.

Our main report pointed out several flaws in the outlier procedure used by Sumicsid, including the following.

- The dominance test requires several parametric assumptions (which Sumicsid did not test), and relies on large sample properties (which Sumicsid appears to be unconcerned with, in contrast to their issues with SFA).
- The dominance test is biased towards non-rejection (against detecting outliers).

[268] For example, this is the case in German DSO regulation where BNetzA employs a CRS technology assumption with the objective of encouraging structural reforms to achieve efficient scale. We would note that there are conceptual and empirical inconsistencies in the application of this policy decision of Bundesnetzagentur by its consultants.
[269] Economic Insights (2017), 'Topics in efficiency benchmarking of energy networks: Choosing the model and explaining the results', December p. 11
[270] For example, in addition to TSOs that may be unable to increase their scale due to national boundaries (or regulatory imposed boundaries), it may not be feasible for a TSO to reduce its size if regulatory restrictions are in place.
[271] VRS may be present in the model if (i) there are genuine diseconomies of scale in the organisation of TSOs or (ii) Sumicsid's model is mis-specified and omits important drivers of cost for the large TSOs in the sample. For a discussion on the latter argument, see Banker (2016), 'The Returns to Scale Assumption in Incentive Regulation', March.

- The super efficiency test is unlikely to identify outliers in a volatile sample.
- Both tests are vulnerable to masked outliers (units that are close to each other, but far from the rest of the sample).

5.2.4 Material omissions in Sumicsid's response

Sumicsid does not engage with the merits of the different outlier tests proposed in our main report.[272] Instead, it refers to the fact that its outlier procedure is still used in Germany. We make the following observations.

- There has been no official decision on the merits of the German dominance test relative to the bootstrap test yet.
- The decision to apply the super-efficiency test once and not sequentially (decreasing its vulnerability to outliers) or with a fixed threshold (increasing its likelihood to detect outliers in a volatile sample) is mandated by law in German regulatory benchmarking and was thus not part of that appeal to begin with. TCB18 is not bound by this legal requirement.
- The situation in Germany is largely irrelevant for international benchmarking exercises, such as TCB18. The flaws of the ARegV's outlier procedure, such as the inability to detect masked outliers, still exist regardless.

Sumicsid now also refers to various other outlier procedures it supposedly used to determine outliers (visual inspection, cooks distance). While we agree that visual inspection can form a part of the outlier analysis, these were not documented in the final TCB18 report and we thus cannot evaluate whether they were performed rigorously and consistently.

5.3 Model validation—DEA weights

When developing a cost assessment model, the outputs of that model need to be validated to ensure that it is consistent with operational expectations. In a DEA context, this could involve the assessment of the weights on the input and output factors, and the peers and their weights for the inefficient TSOs. If it is believed ex ante that some outputs are stronger drivers of cost than others, an examination of the DEA weights could support or contradict this and lead to further model development. Moreover, as alluded to in Sumicsid's TCB18 report,[273] an examination of DEA weights could be used to identify unusual TSOs.

In our main report, we questioned Sumicsid's lack of validation through examination of DEA results. Without an evaluation of DEA weights, many statements made in the report—including that concerning the importance of NormGrid—are unverifiable. For instance, we found that NormGrid was not the primary driver of efficiency in Sumicsid's model, despite Sumicsid's assertions that it was the 'strongest' parameter.

In addition, we questioned the lack of peer analysis in TCB18. The average scaling factor is 4.1, indicating that the average inefficient TSO is compared to TSOs approximately four times smaller than itself. Indeed, the highest scaling factor is 12.15. For this TSO, the TSOs to which it is being compared are so different in scale and complexity that many of the solutions implemented by the

[272] Sumicsid response (2020), para. 3.115.
[273] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, p. 30.

peer may not be feasible. In any case, the outputs from DEA should be validated to be useful for learning or any other purpose.

Sumicsid does not directly respond to our examination of DEA weights or comment on their notable absence in TCB18. Instead, Sumicsid claims that the specific statement that NormGrid was the strongest parameter was based on 'initial Lasso and OLS regressions for a single variable'.[274]

The TCB18 response does not comment on the lack of peer analysis in TCB18, instead stating that the result of an average scaling factor of 4.1 is 'not surprising', as scaling factors have a lower bound of one due to the NDRS assumption.

Sumicsid also replies to our analysis on the second-stage analysis of its own model in its reply to section 5.3. We address this response in section 5.4, in line with the structure of our main report.

5.3.5 Factual errors and misrepresentations in Sumicsid's response

In Sumicsid's final report for TCB18, it states:[275]

> In general, the strongest candidate in the **frontier models** is the normalized grid. [emphasis added]

However, our main report demonstrated that NormGrid was not the most important driver in the DEA model for most TSOs. Indeed, four TSOs' efficiency scores are not determined by NormGrid at all, and a further eight TSOs do not have NormGrid as the main driver of their estimated efficiency scores.[276] In response, Sumicsid has confirmed that it did not examine the DEA weights to inform statements regarding the importance of cost drivers. It states:[277]

> The statement that NormGrid is the strongest parameter is related to the **initial Lasso and OLS regressions** for a single variable, where invariably NormGrid comes out as the most informative.

We note that this statement is contradictory to the main report. Neither Lasso nor OLS regressions are frontier models. Indeed, as Sumicsid has rejected the use of alternative frontier models (e.g. SFA), the *only* frontier model that Sumicsid can be referring to in its final report is the DEA model. Therefore, the only evidence that can feasibly support the 'importance' of a cost driver is an examination of the DEA weights, as we had done in the main report.[278]

Of course, one could argue that NormGrid is the most important driver of costs from a 'techno-economic' perspective. However, it would then be essential to interrogate the data to examine *why* the importance of NormGrid is not borne out by the data. The fact that the model outputs are inconsistent with operational intuition should be seen as a cause to investigate further, and not simply dismissed as Sumicsid has done.

---

274 Sumicsid response (2020), para. 3.120.
275 Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, para. 5.01.
276 See Oxera (2020), 'A critical assessment of TCB18', April, Figure 5.4.
277 Sumicsid response (2020), para. 3.120.
278 If one believes that second-stage analysis has merit, then one could also test the relative strength of cost drivers through second-stage analysis. However, as demonstrated in our main report and this response, second-stage analysis does not have merit in this context. In any case, NormGrid is not identified as a material driver using this approach.

5.3.6 Material omissions in Sumicsid's response

Sumicsid correctly observes that the weights on peers (i.e. 'lambdas') must be greater than one in a DEA model if NDRS is assumed. However, Sumicsid does not comment on the magnitude of these lambdas. The average scaling factor is 4.1, indicating that the average inefficient TSO is compared to TSOs approximately four times smaller than itself. Indeed, the highest scaling factor is 12.15. For this TSO, the TSOs to which it is being compared are so different in scale and complexity that many of the solutions implemented by the peer may not be feasible. The observation in our main report that these lambdas are 'unusually' large is supported by the statistical tests regarding the most appropriate returns to scale assumption.

Moreover, Sumicsid has not commented on the merits of 'peer analysis' to examine the comparability of inefficient TSOs to their peers. It is possible in DEA to perform a precise one-to-one assessment of the homogeneity of the units. This assessment could cover several factors, such as the regulatory environment, ownership structure and quality of service. While such an approach is time-intensive and becomes less feasible as the dataset gets larger, Sumicsid should have performed such an analysis before concluding that the DEA results were robust and that the estimated efficiency targets are feasible.

5.4 Model validation—identification of omitted cost drivers

If relevant cost drivers are omitted from the model specification, the resulting efficiency scores will be biased for particular TSOs. The relevance of potential cost drivers is commonly assessed with the cost driver analysis in the model development phase of a benchmarking study and validated further through extensive sensitivity analysis.

Sumicsid conducted second-stage regressions to test for the exclusion of relevant cost drivers in TCB18. The main report states:[279]

> Extensive second-stage analyses shall be undertaken to **see if any of the non-included variables should be included**.

In addition, in the individual reports, it stated:[280]

> In order to in**vestigate whether some potentially relevant variables have been omitted** in the final model specification, a so-called second stage analysis has been performed.

Oxera argued in its main report that second-stage analysis, as conducted by Sumicsid, suffers from fundamental conceptual flaws. Indeed, we are not aware of any academic literature supporting the use of second-stage regressions to assess the relevance of omitted outputs in a DEA model. Moreover, we demonstrated by way of example that the second-stage analysis cannot detect whether relevant cost drivers have been omitted from the model. To show this, we estimated efficiency scores in a model controlling for two of the three cost drivers used in the final model, and we use Sumicsid's second-stage approach to test whether Sumicsid's third cost driver is deemed to be 'omitted' or not. We found that Sumicsid's approach only identifies transformer power, and not NormGrid, weighted lines or the environmental adjustment, as a relevant omitted variable.

[279] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, para. 4.09.
[280] Sumicsid (2019), 'Project TCB18: Individual Benchmarking Report Fingrid – 131', p. 35.

In response, Sumicsid states:[281]

> The critique against the post-run (second-stage) analysis is not well posed since the purpose is **not related to model-specific variable selection**. The post-run analysis aims at investigating and validating the potential presence of systematic bias for operators with specific conditions with respect to the efficiency score used in the regulation. The list of parameters includes also elements already in the model, for information about potential impact and not as an omitted variable.

Sumicsid also states that Oxera's second-stage analysis is not a reason to select or to deselect a specific parameter and that the 'techno-economic significance' of a parameter must be taken into account.

5.4.7 Factual errors and misrepresentations in Sumicsid's response

Sumicsid argues that 'the post-run second-stage process is intended to detect potential bias in the scores, not the inclusion of specific parameters.'[282] This contradicts Sumicsid's statements in the main report. It also contradicts statements made in the individual reports and e3grid2012 (and there will be other instances in which Sumicsid has been explicit about the use of second-stage analysis), such as:

> In order to **investigate whether some potentially relevant variables have been omitted** in the final model specification, a so-called second stage analysis has been performed. The idea of the second stage analysis is to investigate if some of the remaining variation in performance can be explained by any of the unused cost drivers.[283]

> We have further undertaken so-called second stage analysis. **The purpose of a second stage analysis is to ensure that we have appropriately specified the best model using the available data**. We do so by testing if any excluded variables should potentially have been included. In a second stage analysis, the efficiency scores are regressed against an excluded variable to determine whether it has a significant impact on efficiency scores. If the variable were to significantly explain the efficiency scores, this could be an indication that the respective variable should have been included in the base model. Therefore, second stage regression analysis provides a valuable control of the model specification. [emphasis added] [284]

The main report implies that, if Sumicsid had identified a statistically significant omitted variable from the second stage, it would have included (or, at least, investigated the inclusion of) the omitted variable in the model. **However, Sumicsid's response contradicts this**. The response also begs the question: what would Sumicsid have done if a variable were identified as omitted in the second-stage analysis (i.e. a systematic effect were identified)? The process is not explicitly discussed in the TCB18 documents, and Sumicsid's response appears to indicate that it would not impact the final model. So the second stage validation is used as a *false basis* by Sumicsid to conclude that its model is well specified. We conceptually and empirically showed that it is a baseless test that should be abandoned altogether.

Sumicsid also does not appear to have understood the purpose of our analysis demonstrating that NormGrid would not have been detected as an omitted cost driver using the second-stage analysis. Our argument was that Sumicsid claimed that NormGrid is the most relevant cost driver, but even this cost driver

[281] Sumicsid response (2020), para. 3.127.
[282] Sumicsid response (2020), para. 3.127.
[283] TCB18 individual report, section 5.
[284] e3grid2012, pp. 11–2.

could not be detected by Sumicsid's second-stage analysis. The point of the analysis was to demonstrate that the second-stage analysis is not fit for purpose, not to argue that NormGrid should be excluded from the model.

Specifically, we stated:[285]

> Note that we do not present these results as arguments for excluding NormGrid and weighted lines from the model. Rather, these results **illustrate that the second-stage validation used by Sumicsid is not able to identify relevant omitted cost drivers**. [emphasis added]

That is, we agree with Sumicsid's comments regarding the use of second-stage analysis in model development (i.e. that it cannot be used), but go further by stating that it has no value in a model validation exercised. We also agree that operational considerations are important when selecting and modelling cost drivers, as clearly stated in our main report.[286]

5.5 Model validation—SFA

DEA is only one method of assessing the efficiency of TSOs. It suffers from being a deterministic method of efficiency assessment (as applied by Sumicsid) and the results are contingent on assumptions imposed on the model. In this respect, it treats the data 'as given' and makes no allowance for uncertainty in the variables. Therefore, it is common for regulators to use alternative benchmarking methods, either to directly inform the efficiency target or as a cross-check on the results from DEA. SFA is particularly relevant when we suspect a lot of uncertainty in the data, as in the current case. It is therefore important that the results from the TCB18 model are supported by SFA models.

Sumicsid has not used SFA or any other benchmarking method to validate its model in TCB18 (unlike in previous benchmarking exercises, such as e2gas),[287] nor has it used SFA to assess the robustness of the TSOs' efficiency scores.

In our main report, as is common in real-world applications, we tested different assumptions on the distribution of inefficiency, as outlined below.

- **Pooled SFA.** Each observation (a TSO in a specific year) is treated as independent. This means that a TSO's inefficiency in one year is uncorrelated with its inefficiency in another.
- **Time-invariant SFA.** The observations for a TSO are assumed to be correlated and a TSO's inefficiency is assumed to be fixed throughout the modelling period. While this might be an inappropriate assumption for long panels covering multiple years, it might be a reasonable approximation for a relatively short panel (e.g. five years).
- **Time-varying SFA.** The observations for a TSO are assumed to be correlated and TSOs' inefficiency are assumed to follow a common, deterministic trend.

Under all three modelling assumptions, we found that the estimated inefficiency in the sample is statistically insignificant; that is, the data used by Sumicsid is consistent with there being no statistically significant inefficiency

[285] Oxera (2020), 'A critical assessment of TCB18 electricity', April, p. 73.
[286] Ibid., p. 45.
[287] Sumicsid and Swiss Economics (2015), 'PROJECT E2GAS Benchmarking European Gas Transmission System Operators – Final report'.

among the TSOs, with the deviations due to statistical noise being potentially identified as inefficiency by the DEA model.

Sumicsid disagreed with our analysis regarding SFA, with its arguments falling into one of two categories.

- The use of a stochastic model is not important when the TCB18 data and model specification are deterministic.
- The sample is too small to allow for a robust estimation of SFA models.

5.5.8 Factual errors and misrepresentations

Sumicsid alludes to a hypothesised inconsistency between our main report and previous work for ACM with regards to the uses and limitations of SFA. Sumicsid presents the following quote from an Oxera report.[288]

> SFA models with fewer than 30–40 observations can fail to produce results

We have four points of contention with the argument posed by Sumicsid. First, and most obvious, we have over double the number of observations that we defined as the 'lower limit' (the panel models presented in our report used 81 observations). Even if there was an extensive outlier procedure that reduced the sample size, it is highly unlikely that the number of observations would fall below 40.

Second, Oxera (2012) states that the model *can* fail with fewer than 30–40 observations, not that the model *will* fail. An SFA model can be estimated with significantly fewer observations than that available to Sumicsid. If the model converges and the results are intuitive from an economic, operational and statistical perspective, then there is no reason to disregard evidence from SFA models, however 'small' the sample is.

Third, in the same report, we argue that 'as SFA is stochastic, where noise is prevalent in the data it is likely to produce more robust results than DEA'.[289] As we highlighted throughout our main report, data errors are prevalent throughout the TCB18 dataset. Indeed, we note that Sumicsid has already acknowledged and corrected one form of data error (the data for weighted lines) *subsequent* to the publication of the final report. Moreover, the various adjustments that are made (or not made) to the data, such as the adjustment for regional wages, will create an additional layer of uncertainty that is not accounted for in Sumicsid's deterministic approach at all and will lead to biased estimates of inefficiency. SFA is much less sensitive to data errors than DEA.

Fourth, Sumicsid has presented no evidence that its sample is sufficiently large for DEA to produce robust estimates of (in)efficiency. As noted above, all models are estimated with an associated 'error' or 'uncertainty', including DEA. These uncertainties will likely be larger when the sample size is small, as indicated by the SFA models presented in Oxera (2020). Relatedly, Sumicsid refers to the 'convention of 3(#inputs+#outputs)'[290] to support its restriction of the model to account for only three outputs. **To our knowledge, there is no robust scientific exposition of this derivation**. We would also observe that a key characteristic of a robust benchmarking model, especially in a regulatory context, is not whether a model obeys a certain 'convention' or has a particular number of efficient units, but rather whether the model allows for a robust like-

[288] Oxera (2012), 'How can the NMa assess the efficiency of GTS?', June, p. 20.
[289] Ibid., p. 19.
[290] Sumicsid response (2020), section 5.09.

for-like (i.e. homogenous) comparison that can be demonstrated from an economic, statistical and operational perspective. That benchmarking must be undertaken on a homogenous basis is in fact dictated in the law in many jurisdictions.

For the reasons outlined above, we consider that Sumicsid's representation of Oxera (2012) is both inaccurate and irrelevant in the context of TCB18.

5.5.9 Conceptual errors in Sumicsid's response

Sumicsid advances the following claim:[291]

> An application [of SFA] to a time-series or a pooled dataset would introduce a number of a priori assumptions that are neither empirically verifiable, nor relevant to solve the question of the incumbent cost efficiency in the reference year.

As noted above, we presented results using several different distributional assumptions regarding the inefficiency term, and all led to the same conclusion that all of the deviation between TSOs' predicted and observed costs can be attributed to statistical noise. A more detailed investigation into the most appropriate distributional assumption might have been warranted if the models produced contradictory results. However, the models failed to detect any statistically significant inefficiency in all cases. **Note that Sumicsid's statements regarding the empirical verifiability of different distributional assumptions is misleading and incorrect**. It is possible to conduct statistical tests to determine which distributional assumption is more appropriate, which can be corroborated using sector-specific knowledge. Also, it is possible to test whether the data should be pooled or if a panel model is more appropriate.

It is unclear how Sumicsid can conclude that panel data is not relevant for assessing the cost efficiency of a TSO in the reference year. **This is wrong and misleading**, especially when one can test the assumptions underlying the panel models. Panel data increases the size of the dataset and allows a more robust estimation of the parameters in the model (both the coefficients on the cost function and the shape of the inefficiency distribution). It is simple to estimate a TSO's inefficiency in **any** reference year using a model that is estimated using panel data. If Sumicsid believes that efficiency varies significantly year-by-year then it could, for instance, have run a pooled SFA and taken the 2017 efficiencies as their estimate.

Furthermore, note that Sumicsid used panel data in its model development and used the simplest estimator, pooled robust OLS.[292] That is, it did not use any model that took into account the panel nature of the dataset, as we have done in our analysis. Similarly, we understand that pooled data was used for the Lasso regression that, according to Sumicsid, confirmed the appropriateness of NormGrid and informed the optimal number of cost drivers in the model.

Sumicsid also states:[293]

> Oxera (2020) claims that SFA should be relevant for cross-validation, but fails to find any significant gamma term (individual inefficiency), thus the exercise confirms our claim [that SFA should not be used in this context].

[291] Sumicsid response (2020), section 3.131.
[292] We gleaned this from the fact that Sumicsid always lists the number of observations as 81 (16 operators over five years plus an additional operator in 2017 only). For example, see Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, Tables 5-2 and 5-3.
[293] Sumicsid response (2020), section 3.133.

In this respect, we note that many of the TSOs operate in regulatory regimes where there are strong incentives for cost reduction. Indeed, some TSOs have been in such regimes for decades. If the regime is successful in encouraging efficiency improvements, there is a reasonable possibility that the TSOs are operating efficiently (or nearly efficiently), as in competitive markets in the neoclassical model. A primary advantage of SFA is that it is possible to test the hypothesis of whether TSOs are operating efficiently, whereas, for all practical applications, a DEA model is likely to confound noise for inefficiency even if there is no real inefficiency in the sample.

Sumicsid's assumption that there is no noise or uncertainty in the efficiency scores estimated from its DEA model is particularly concerning. In any model, there is always an 'error' or 'uncertainty'; the difference between DEA and SFA is that, in SFA, the measure of uncertainty (such as standard deviation or statistical significance) is immediately visible whereas, in DEA, further analysis is required. **Just because Sumicsid's naïve DEA cannot produce them, does not mean that uncertainty does not exist**. For example, the 'error' or 'uncertainty' can be estimated using bootstrapping methods or modelled using Monte Carlo analysis, neither of which were presented by Sumicsid in TCB18.

Furthermore, we note that Sumicsid has been inconsistent in its treatment of statistical analysis. On the one hand, Sumicsid has made conclusive statements based on statistical inference on smaller datasets than what we used to estimate the SFA models—Sumicsid's outlier test (i.e. a statistical test) is estimated using 16 observations and the second-stage analysis uses 17 observations, yet Sumicsid concludes that there are no dominant units or omitted variables based on the statistical significance of these tests. On the other hand, Sumicsid appears to have rejected or ignored statistical tests when it comes to the returns to scale assumption and the presence of inefficiency in the sample.

5.6 Model validation—frontier shift

Frontier shift relates to the ability of the most efficient operators in an industry to improve productivity. In a DEA context, frontier shift can be estimated by assessing the evolution of the efficient frontier over time. Alternatively, in an SFA context, frontier shift can be estimated by including time variables (e.g. a time trend or time dummies) in the model specification. In Oxera's main report, we argued that the large and negative estimated frontier shift of 4% p.a. (i.e. a frontier *regress* of 4% p.a.) was potentially unintuitive when compared to the frontier shift estimated in regulatory applications. In turn, this could be indicative of model mis-specification rather than genuine deteriorations in productivity. For example, we noted that Sumicsid's model did not control for factors that may change over time (such as asset health) and the position of a firm in the investment cycle. Moreover, we argued that the inability of the model to explain costs *over time* raises additional concerns regarding the ability of the model to account for differences in efficient expenditure *between TSOs*.[294]

Subsequent to the publication of our main report, we have reviewed the dynamic efficiency analysis report that Sumicsid published after its final report for TCB18. Sumicsid corrected for the error in angular towers for this analysis (see section 3.1), leading to an estimated frontier regress of c. 1.7% p.a. In this report, Sumicsid stated that the frontier shift is estimated on real expenditure

[294] Oxera (2020), 'A critical assessment of TCB18 electricity', April, section 5.6.

data (i.e. adjusted for inflation), meaning that means that the negative frontier shift should be interpreted as a frontier regress relative to the wider economy.

In the TCB18 response, Sumicsid argued that:[295]

> Neither productivity progress, nor regress per se should be directly interpreted as linked to the model structure.

Moreover, Sumicsid presents the results of two studies (Llorca et al. (2016) and AER (2019)) that it states support the existence of a negative frontier shift in energy transmission.

5.6.10 Factual errors and misrepresentations

Sumicsid's statement that Llorca et al. (2016) finds frontier regress in the USA is incorrect. Llorca et al. (2016) estimates an SFA model that controls for environmental factors as a determinants of inefficiency and finds that efficient costs are *decreasing* over time, counter to Sumicsid's assertion. Llorca et al. (2016) states:[296]

> The coefficient on the time trend is negative (nonetheless it is not significant in some of the models), which indicates that costs decrease over time, i.e. there is positive technical change.

Indeed, we consider that the results of Llorca et al. (2016) support our argument in the main report that an estimated technical regress could be a sign of model mis-specification. Llorca et al. (2016) controls for weather conditions that, in principle, are expected to vary through time (e.g. due to random faults caused by weather events or storms that need to be fixed through OPEX) and across regions (e.g. due to climate affecting the deterioration of assets). The fact that Sumicsid's model does not account for weather conditions could be one of the many reasons for the volatility of the estimated efficiency scores and the estimated frontier shift.

It is not clear whether Sumicsid has robustly interrogated its sources when responding to our main report. Nonetheless, the existence of one or two isolated academic papers does not validate Sumicsid's approach.

5.6.11 Conceptual errors in Sumicsid's response

Sumicsid's argument that the estimated frontier regress is unrelated to the model structure is clearly incorrect. The frontier shift calculation involves calculating a TSO's performance relative to a frontier. Therefore, how that frontier is defined will have an impact on what the estimated frontier shift would be. Had Sumicsid used an alternative set of cost drivers, the estimated frontier shift would be different. The same would be true if Sumicsid had used different cost normalisation (e.g. deflated using a different price index), different aggregation weights for NormGrid, or different outlier analyses.[297] Indeed, nearly every modelling decision made throughout TCB18 will have an impact on the estimated frontier shift.

[295] Sumicsid response (2020), para. 3.139.
[296] Llorca, M., Orea, L., and Pollitt, M. (2016), 'Efficiency and environmental factors in the US electricity transmission industry', *Energy Economics*, **55**, pp. 234–46.
[297] We note that Sumicsid's outlier analysis in the dynamic report is not clear. It states that it has performed the same outlier procedure as in the static analysis (see Sumicsid (2020), 'PROJECT CEER-TCB18 Dynamic efficiency and productivity changes for electricity transmission system operators MAIN REPORT', April, para. 3.04). However, it also states that the frontier shift was estimated on a sample of 16 TSOs (see Sumicsid (2020), 'PROJECT CEER-TCB18 Dynamic efficiency and productivity changes for electricity transmission system operators MAIN REPORT', April, Table 3.7), which implies that Sumicsid did not perform the outlier analysis.

It is true that it is *possible* for productivity to regress in the TSO sector. However, given that European TSOs generally operate under regulatory regimes that encourage cost reduction, the nearly universal regress in productivity warrants further investigation. Indeed, the estimated regress could be a symptom of omitted cost drivers or inappropriate cost normalisations, as we have documented throughout our interactions on TCB18.

6 Conclusion

In our main report, we voiced significant concerns over the use of the TCB18 study. We conclude that the study suffers from a **number of significant flaws**, some of which are **fundamental**. The TCB18 response has not served to alleviate these concerns.

Instead, the TCB18 response and Sumicsid's statements provides further evidence for the following.

- The TCB18 process and the published material are opaque, such that an independent assessor of the study could not accurately follow what Sumicsid had done. Indeed, Sumicsid contradicted itself on several occasions.
- The TCB18 dataset contained material errors and is measured with a significant degree of uncertainty.
- The model development procedure was arbitrary and not supported by empirical evidence.
- The final TCB18 model was not robustly validated.

On numerous occasions Sumicsid directly contradicts its statements from the main report, arbitrarily shifting modelling criteria and rationale, and misrepresents our arguments and those of academic articles.

Concerningly, Sumicsid appears to resist any reasonable margin for error at any stage and does not appear to engage with the pragmatic solutions presented in our main report. Therefore, we maintain that the flaws covered in our main report and associated work mean that the estimated efficiency scores and suggested cost savings are not robust and **cannot be used in their current form for regulatory, operational or valuations purposes**.