A critical assessment of TCB18 electricity

Prepared for
the European electricity TSOs

30 April 2020

Final

www.oxera.com

Contents

Figures and tables

Short Management Summary

The pan-European Transmission Cost Benchmarking Project[1] (TCB18) 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 66–100%, 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 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.

Key messages

An overarching issue with TCB18 is that Sumicsid's outputs do not contain necessary information for third parties to clearly follow its analysis, validate its analysis or its sources without considerable effort. As such, the level of transparency exhibited by Sumicsid falls short of what would be considered regulatory good practice.

In addition, we identified some significant issues specific to each stage of Sumicsid's benchmarking analysis, which are summarised below under three general themes.

1. Sumicsid's data collection and construction process do not enable a sufficiently harmonised dataset to undertake robust cost benchmarking.

Sumicsid states that it carried out a rigorous data collection and validation exercise, involving a number of iterations with independent auditors, TSOs and national regulatory authorities (NRAs).[2] Nevertheless, TSOs have informed us of several data errors in the final dataset on which Sumicsid's benchmarking was performed. Sumicsid did not robustly consider the impact of such errors on the estimated efficiencies, especially as DEA, as applied by Sumicsid, is highly sensitive to data errors. It is also widely recognised that 'real' data is noisy, hence such robustness checks are necessary even where the data is supposedly free of errors (which is not the case in TCB18). We undertook extensive Monte Carlo simulations to estimate the impact of data uncertainty, which concluded that both the classification of TSOs as efficient or otherwise, as well as the level of the estimated efficiencies, are sensitive to small errors in the data. For example, four other TSOs that are currently not identified as peers become peers.

Furthermore, Sumicsid's decision to model total expenditure (TOTEX) as a single input assumes a strict, one-to-one trade-off between operating expenditure (OPEX) and capital expenditure (CAPEX). This is inappropriate

[1] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July.
[2] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, section 3.2.

and unnecessary for several reasons, and could inappropriately conflate heterogeneity with inefficiency. Accounting for this type of heterogeneity changes the efficiency classification as well as increasing the efficiency scores of some TSOs by up to 17 percentage points.

Sumicsid has not adequately adjusted for differences in input prices across TSOs. Specifically, Sumicsid currently adjusts only for manpower (i.e. labour) OPEX, which translates to approximately 5.9% of TOTEX (on average across TSOs) being normalised for price-level differences. No adjustment is made to CAPEX or other cost items within OPEX. We find that the estimated efficiency of some TSOs can change by up to 40 percentage points if price levels are better accounted for.

2. Sumicsid's approach to model development appears arbitrarily restrictive and inconsistent with the scientific literature.

Sumicsid has not undertaken sufficient validation of its model specification (i.e. the relationship between TOTEX and the cost drivers identified) using statistical tests or other methods. For example, we find that the estimated model is highly sensitive to the inclusion or exclusion of specific TSOs, indicating that a few unusual TSOs are driving the model specification. Similarly, Sumicsid does not present any compelling statistical analysis to support key assumptions in the model development process. Moreover, Sumicsid has also not effectively used all the information it has at its disposal, and has, for example, without justification, focused on a single year's data without cross-checking the impact of this.

Sumicsid states that the asset-based measures it uses as cost drivers are highly correlated with TSO cost drivers, such as network capacity and routing complexity, but these statements are unsubstantiated in its report and alternatives were not considered. Moreover, where we used alternative asset-based outputs to capture similar operating characteristics, this has a significant impact of up to 39 percentage points on the TSOs' estimated efficiencies, emphasising the uncertainty surrounding the chosen proxies. The lack of alternative model specifications involving outputs (rather than asset measures) is a significant omission in the TCB18 study.

Sumicsid considers NormGrid to be 'the strongest candidate in the frontier models'.[3] Constructed variables such as NormGrid reflect an aggregation of a number of classes of assets using weights that are themselves estimated with a degree of uncertainty. In this context, it may be more appropriate to consider each asset class as a separate cost driver and to allow the DEA model[4] to determine the correct weights on each asset class. We find that replacing the outputs in Sumicsid's model with the main components of NormGrid as separate outputs has a material impact of up to 29 percentage points on the estimated efficiencies of individual TSOs.

Sumicsid's environmental adjustment to NormGrid 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 the complexity factor weights were counterintuitive, as TSOs that operate in

[3] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, p. 32.
[4] This can include weight restrictions if required.

more complex regions (as defined by Sumicsid's complexity factor) have *lower* costs per NormGrid on average.

3. Sumicsid has not justified the assumptions that it has made in its model, and its approach to model validation is incapable of detecting flaws or omissions in its model.

Sumicsid makes several statements regarding statistical tests that it has undertaken to support its modelling assumptions, yet it does not present the empirical evidence in its final outputs. For example, we find no conclusive statistical evidence to support Sumicsid's returns-to-scale assumption (non-decreasing returns to scale). Alternative returns-to-scale assumptions lead to improvements in efficiency of up to 30 percentage points for some TSOs.

Sumicsid has relied on the German regulatory ordinance (ARegV) for outlier detection. However, the outlier procedure set out in the ARegV is neither legally binding nor sufficient in an international benchmarking context. In addition, the scientific flaws of the ARegV's outlier procedure are well known.[5] Where alternative and scientifically appropriate outlier tests are considered, we find that some TSOs' estimated efficiencies are underestimated by up to 17%.

As part of its validation procedure, Sumicsid uses regression analysis involving the estimated efficiency scores from DEA and potentially omitted cost drivers. However, there is no theoretical basis to support this validation approach to identify omitted outputs.[6] Moreover, we show that Sumicsid's own model will not be supported by its validation approach. Thus, Sumicsid has not demonstrated that no relevant variables were omitted from its sole model.

Sumicsid has not examined whether the DEA outputs are consistent with economic and operational expectations. For example, Sumicsid states that NormGrid is the primary driver of expenditure, yet most TSOs' efficiency scores are not primarily driven by NormGrid. Furthermore, Sumicsid has not examined whether the peers for the inefficient TSOs, and how they are scaled, are appropriate. Indeed, we find that some TSOs are being compared against peers that are up to 12 times smaller.

Importantly, Sumicsid has not cross-checked the results of its analysis using well-established alternative methods such as stochastic frontier analysis (SFA), despite having a panel of data available.[7] SFA applied to Sumicsid's model and dataset suggests that there is no statistically significant inefficiency among the TSOs. The SFA model not finding statistically significant inefficiency is not a reason to use DEA; rather, it suggests that caution is warranted against interpreting any estimated inefficiency in the DEA as actual inefficiency rather than statistical noise, and/or that the model specification should be re-examined.

Finally, dynamic efficiency analysis casts further doubt on the validity of Sumicsid's model and dataset. For example, DEA indicates a frontier *regress*

[5] 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, *ew–Magazin für die Energiewirtschaft*.
[6] 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.
[7] A panel dataset contains data over time across TSOs and thus contains more information than a single year of data.

of 4% p.a.[8] Moreover, the frontier shift using SFA is estimated over a wide confidence interval and is statistically indifferent from zero (consistent with the conclusion of the individual inefficiency estimates). Such a volatile, large and negative frontier shift result is indication that Sumicsid's model cannot capture changes in costs over time. 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.

Conclusion of our review of TCB18

International benchmarking **can be a powerful tool** for companies and regulators to assess the efficiency of network operators. This is especially true in the context of the electricity transmission industry, where the sector is often characterised by national monopolies, thus making national benchmarking challenging. In this sense, we welcome projects such as TCB18, which have attempted to develop a framework for periodic assessment of TSOs.

Nevertheless, the TCB18 study itself suffers from a **number of significant flaws**, some of which are **fundamental**. These flaws mean that the estimated efficiency scores and suggested cost savings are not robust and **thus cannot be used in their current form for regulatory, operational or valuations purposes.**

Some of these weaknesses, such as consistency in reporting guidelines, are partly driven by the lack of maturity in the international benchmarking process, and we expect this to improve with time. However, Sumicsid's concluding remarks are concerning, as they are not consistent with the significant issues and areas for future work identified through our comprehensive review. For example:

> Regulatory benchmarking has reached a certain maturity through this process and model development, signaling both procedural and numerical robustness […]
>
> […] future work can be directed towards further refinement of the activity scope and the interpretation of the results, rather than on the model development.

By incorporating the recommendations presented in this report, we consider that CEER will be better able to develop a process and methodology for international cost benchmarking that are informative and fit for purpose. In this regard, it can also be helpful to consider debriefs involving all the parties on process and methodology to help future studies.

[8] Sumicsid published the results of the dynamic efficiency analysis after the finalisation of this report. See Sumicsid (2020), 'Dynamic efficiency and productivity changes for electricity transmission system operators', April.

Long-form Executive Summary

Background

The pan-European Transmission Cost Benchmarking Project[9] (TCB18) was carried out by the Council of European Energy Regulators (CEER) through its consultant, Sumicsid. It is a follow-up to previous studies, such as ECOM+[10], e3grid2012[11] and e3grid2008.[12]

The TCB18 study involved an international comparison of 17 electricity TSOs based in 15 European countries. Sumicsid used cost and asset data provided by TSOs, in addition to environmental and price-level data from external sources, to assess the relative efficiency of TSOs. As in previous benchmarking exercises, Sumicsid used data envelopment analysis (DEA) to estimate the efficiency of the European electricity TSOs.

A consortium of all the European TSOs that participated in TCB18[13] commissioned Oxera to validate and review the results from TCB18, and to recommend robust solutions to any issues that emerge.

We understand that the results from TCB18 could be used by some national regulatory authorities (NRAs) as evidence to set regulatory revenues for the TSOs concerned. Hence, it is essential that the limitations of Sumicsid's analysis are fully understood.

Assessment

As part of this study, we have reviewed a number of outputs, including:

- the final report and appendices as published by Sumicsid, which are available on CEER's website;[14, 15]
- the TSO-specific outputs that detail individual TSOs' data and performance;[16]
- workshop slides that Sumicsid shared with the TSOs through the course of the TCB18 project.[17]

We also received the final underlying dataset from the TSOs as used by Sumicsid in its analysis. We have used this dataset to validate Sumicsid's work. It should be noted that Sumicsid's outputs do not contain the necessary

[9] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July.
[10] Sumicsid (2006), 'ECOM+ Results 2005 FINAL REPORT', June
[11] Sumicsid, Frontier Economics, Consentec (2013), 'E3GRID2012 – European TSO Benchmarking Study A REPORT FOR EUROPEAN REGULATORS', July
[12] Sumicsid (2009), 'International Benchmarking of Electricity Transmission System Operators e 3GRID PROJECT – FINAL REPORT', September
[13] A full list of the participating TSOs can be found in Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, Table 2-2.
[14] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July.
[15] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators appendix', July.
[16] These are generally not publicly available. However, Fingrid has published its report online and we will reference this report where appropriate. See Sumicsid (2019), 'Project TCB18 Individual Benchmarking Report Fingrid – 131', July, found here https://energiavirasto.fi/documents/11120570/12862527/tcb18_indrep_final_elec_131_FI+%28Fingrid+Oyj%29.pdf/cda330f6-ea39-a345-3e8b-cf4c1170522a/tcb18_indrep_final_elec_131_FI+%28Fingrid+Oyj%29.pdf, last accessed 31 January 2020.
[17] See Sumicsid (2019), 'Model Specification Model Results', April; Sumicsid (2018), 'Validation of NormGrid and Preliminary Environmental Results', November; Sumicsid (2019), 'CEER-TCB18 project Model Specification ELEC V1.3', February.

information for third parties to clearly follow its analysis, validate its analysis or validate its sources without considerable effort. In our replication, small deviations in some of the associated outputs presented by Sumicsid exist due to the lack of information provided by Sumicsid in its various outputs. The level of transparency exhibited by Sumicsid in the project falls short of what would be considered good practice.

Nevertheless, our replication was close enough for us to identify issues and conclude on the quality of the benchmarking. In fact, we have identified a number of significant issues with Sumicsid's analysis; these can be summarised under three themes, as follows.

1. Sumicsid's data collection and construction process do not enable a sufficiently harmonised dataset to undertake robust cost benchmarking.

Sumicsid states that it carried out a rigorous data-collection exercise involving a number of iterations with independent auditors, TSOs and national regulatory authorities (NRAs).[18] In theory, its procedure should produce a relatively robust dataset for benchmarking purposes. However, despite such a lengthy, iterative process, we have identified several issues with the dataset that Sumicsid used in its analysis. Furthermore, the adjustments that Sumicsid makes to the data are insufficient and not adequately justified.

i. Sumicsid has not adequately ensured that the final dataset is free from significant data errors and inconsistencies

TSOs have informed us of several data errors in the final dataset—for example, miscommunication regarding the reporting guidelines, leading to misreporting of data, and measurement error. For example, some TSOs aggregated their data for towers in a way that indicated that they have no angular towers (thus understating the weighted lines variable by 100%). This would clearly underestimate the level of output for these TSOs, and bias the resulting efficiency scores.

As part of our assessment, we had to take the data collated and processed by Sumicsid largely as given and could only make specific changes for particular TSOs (i.e. we were not able to tackle systematic or pervasive errors). However, we illustrate the impact of data errors and data uncertainty on the estimated efficiency scores of each TSO through Monte Carlo simulations, which have been considered by regulators in international and national benchmarking exercises.[19] The analysis indicates that most TSOs' efficiency scores are highly sensitive to possible errors in the data. For example, four TSOs that are estimated to be inefficient in Sumicsid's analysis have a 100% efficiency score in at least 5% of the simulations.[20] This demonstrates that, based on data uncertainty alone (i.e. ignoring all of the modelling flaws in Sumicsid's analysis), Sumicsid's analysis is not able to robustly identify

[18] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, section 3.2.
[19] Specifically, we assume that the actual value of a variable is the measured value plus an error. This error is assumed to be uniformly distributed and +/- 10% of the observed value of a variable (the level of error, while conservative, is informed by the scale of errors noted by the participating TSOs and can also be informed by the standard error of the cost drivers from the regression model). We re-estimate Sumicsid's model 1,000 times with a different error each time, and this creates the distribution of inefficiency scores.
[20] In this context, we focus on a right tailed test where the estimated efficiencies from the simulations are sorted in ascending order. We focus on the 95th percentile of this ordered sequence of scores to see if a TSO deemed inefficient under Sumicsid's analysis is estimated to be 100% efficiency because of data errors in at least 5% of the simulations.

efficient and inefficient TSOs, nor robustly estimate the level of inefficiency present in the TSOs.

ii. Sumicsid's choice of input variable does not appropriately capture the trade-off between different types of expenditure

Sumicsid models expenditure on a total expenditure (TOTEX) basis, where TOTEX is the sum of operating expenditure (OPEX) and capital expenditure (CAPEX). This implicitly assumes that there is a one-to-one trade-off between OPEX and CAPEX. However, OPEX and CAPEX are calculated differently (and are not strictly a measure of a TSO's TOTEX in a year) and subject to different normalisations[21] that may limit the extent to which the two types of expenditure are comparable. If TSO's have different ratios of OPEX to CAPEX dictated by national regulatory and legislative frameworks and operational characteristics, TOTEX modelling, as considered by Sumicsid, could inappropriately conflate TSO heterogeneity with inefficiency.

There are several methods to account for this heterogeneity, none of which has been properly examined by Sumicsid. For example, OPEX and CAPEX can be kept as separate inputs in the DEA model; this ensures that TSOs are only benchmarked against peers with similar OPEX to CAPEX ratios, and mitigates the risk that a TSO is benchmarked against a peer with a very different cost structure. Alternative approaches include developing separate models (econometrically or through DEA) for OPEX and CAPEX, while recognising the trade-offs between the two and without imposing unnecessary assumptions.

Accounting for the heterogeneity in the expenditure categories by modelling OPEX and CAPEX as two distinct inputs leads to two previously inefficient TSOs becoming peers, and swings in estimated efficiency as large as 17 percentage points for some TSOs.

iii. Sumicsid has not sufficiently accounted for differences in input prices across TSOs

Sumicsid adjusts only for manpower (i.e. labour) OPEX by an index of civil engineering price levels to account for differences in input prices across TSOs. This translates to approximately 5.9% of TOTEX (on average across TSOs) being normalised for price-level differences. No adjustment is made to CAPEX or other cost items within OPEX. This approach raises a number of issues, each of which can significantly affect TSOs' estimated efficiency.

- The civil engineering price-level index (PLI) contains prices for non-labour inputs (such as raw materials like metals, plastics and concrete). Its application to labour costs is therefore insufficiently substantiated.
- Sumicsid has limited the scope of the adjustment to a specific cost line within OPEX. In reality, a significant proportion of CAPEX is driven by labour or labour-related costs.
- The differences in non-labour input prices, such as raw materials (which would impact both OPEX and CAPEX), are not accounted for at all.

In our view, based on discussions with the TSOs, one option would be to adjust all OPEX with the price level index (PLI) for overall GDP and to adjust all

[21] For example, OPEX is calculated on an annual basis and is adjusted for differences in labour input prices. CAPEX, on the other hand, is calculated as the sum of annuities in specific investments, and no adjustment is made for differences in input prices.

CAPEX with the PLI for civil engineering, as has been considered in other international benchmarking applications.

Most of the TSOs' efficiency scores are highly sensitive to the method of indexation (such as the choice of PLI and the proportion of expenditure adjusted), and any adjustment (or lack thereof) requires careful consideration and a robust justification.

The impact of price levels on estimated efficiencies can be as large as 40 percentage points, with the estimated efficiency of one TSO increasing by 16 percentage points.

iv. Sumicsid's allocation of indirect costs to assessed OPEX is arbitrary, and evidence supporting its allocation rule is not presented in the report

Sumicsid allocated indirect costs (e.g. human resources expenditure, IT support) to activities considered within the scope of benchmarking based on unsubstantiated allocation rules. Specifically, Sumicsid allocates indirect expenditure to activities based on the percentage of OPEX (minus energy costs and depreciation) in that activity. Large, uncontrollable cost items that are unrelated to indirect expenditure (such as taxes and levies) can have a significant impact on the amount of expenditure allocated to in-scope activities.

We recommend amending the allocation rule to exclude all costs that are considered outside of the scope of benchmarking. This mitigates the risk that indirect expenditure is arbitrarily allocated to activities based on cost items that are unrelated to indirect expenditure. While the impact of this adjustment is material for only one TSO in Sumicsid's model, the allocation of indirect expenditure is an important conceptual issue; it can have a material impact in alternative model specifications and methods that Sumicsid has overlooked. Consideration of the allocation rule would clearly be important for future iterations of the benchmarking study.

2. Sumicsid's approach to model development appears arbitrarily restrictive and inconsistent with the scientific literature.

A robust model-development process is necessary to ensure that the results from an empirical investigation are robust. This process should take into account operational and economic rationale for including or excluding specific cost drivers and should be supported by statistical analysis and operational evidence. Sumicsid's model-development process is not clearly presented in any of its outputs, nor does it consistently follow scientific best practice.

i. Sumicsid's cost driver analysis is not transparent and is based on assumptions that have not been validated in the current context

Sumicsid uses a combination of OLS regression (with and without outliers) and 'robust OLS' (ROLS) regression to validate the relationship between costs and cost drivers (asset-based measures, in this case). Sumicsid does not present analysis behind its model-development process in its final reports,[22] but it does present the coefficients of an ROLS regression on its final model in the TSO-specific outputs.[23]

[22] Some alternative models are presented in workshops throughout the TCB18 study. However, the final model was not justified in these workshops.
[23] Sumicsid (2019), 'Project TCB18 Individual Benchmarking Report Fingrid – 131', July, Table 3.1.

Given this lack of transparency, we cannot validate the econometric analysis that Sumicsid presents in its report. Nonetheless, we have identified several flaws that undermine the robustness of Sumicsid's analysis.

- It is well established that the assumptions required for OLS estimators to provide valid statistical inference are not met if inefficiency (i.e. systematic deviation) is present in the dataset. In the present case, Sumicsid has concluded that about 10% of inefficiency is present (i.e. systematic deviations are present) on average across the TSOs.[24] Therefore, any comment regarding the statistical significance of coefficients in the first stage regression model is not conclusive.
- The estimated coefficients in the final model are highly sensitive to the exclusion of certain TSOs, indicating that a few unusual TSOs may be driving the observed relationship between costs and cost drivers.[25] Sumicsid does not appear to consider such essential robustness checks.
- The estimated efficiency scores in the sample are highly sensitive to the year in which efficiency is assessed. This could indicate that the model is not able to explain changes in expenditure over time (the dynamic efficiency estimates on Sumicsid's model are also highly volatile). For example, the impact of the investment cycle may not be fully captured by Sumicsid's data and modelling adjustments, and the estimated efficiency of TSOs may therefore be influenced by their relative position in the investment cycle.
- Sumicsid's functional form assumption (i.e. a linear relationship between costs and cost drivers) dictating the model specification is not substantiated by statistical analysis. For example, a non-linear relationship between costs and cost drivers can result in alternative cost drivers being identified as relevant. Sumicsid has not presented sufficient evidence to validate its assumptions, nor has it considered the impact of alternative assumptions.
- Sumicsid does not present sufficient detail of the analysis it has used to develop its final model in the main report or appendices. Furthermore, Sumicsid has not shared modelling codes with the TCB18 participants (which could have been anonymous, avoiding any confidentiality issues). As such, the exact process that Sumicsid used to develop its models is not open to allow for third parties to understand the process followed or the results. In this regard, a sample of our modelling code that replicates parts of Sumicsid's analysis is available in Appendix A1 for reference.

As a result of these modelling flaws, it is unlikely that Sumicsid's model development procedure has led to an appropriate final model from which unbiased efficiency scores could be estimated.

ii. Sumicsid has restricted itself to the use of asset-based measures of output, without providing empirical justifications for doing so

All of the 'outputs' used in Sumicsid's final model (NormGrid, transformer power and weighted lines) are measures of assets (i.e. inputs). Indeed, Sumicsid does not present any sensitivity where outputs (such as variability of network supply and total energy transmitted) are used. Using asset-based measures instead of outputs could bias the efficiency assessment in favour of TSOs that choose asset-based solutions. It is also unusual to solely focus on models where the cost drivers are asset-based measures, as the interpretation

[24] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, Table 5-4.
[25] For this robustness assessment, we are using 'robust OLS', the same method as used by Sumicsid.

of the regression outputs from such an input–input model would be unclear (as cost (i.e. TOTEX) is equated with other measures of cost (i.e. assets), resulting in a tautological relationship).

Sumicsid states that the asset-based measures are highly correlated with outputs, such as network capacity and routing complexity. However, Sumicsid has not substantiated these statements with statistical analysis or operational evidence and alternative outputs were not considered. For example, we observe that using alternative, asset-based measures to capture the same operating characteristics has a significant impact of up to 39 percentage points on the TSOs' estimated efficiency scores; this emphasises the uncertainty surrounding these proxies. The lack of alternative model specifications involving outputs is a significant omission in the TCB18 study.

iii. The weights attached to each asset class when aggregating to a NormGrid measure are not robustly validated

Sumicsid considers NormGrid to be 'the strongest candidate in the frontier models'.[26] Constructed variables such as NormGrid carry an inherent risk of favouring some TSOs at the expense of others; this is because such variables reflect an aggregation of a number of classes of assets using weights that are themselves estimated with a degree of uncertainty. Depending on the weights used and the mix of assets of each type that a TSO uses in reality, a TSO may be favoured or disadvantaged relative to other TSOs that differ substantially on the mix of assets. The weights used in the aggregation of NormGrid should therefore be robustly justified.

Sumicsid states that it used linear regression to derive the appropriate OPEX and CAPEX weights on NormGrid.[27] The results of this regression analysis are not presented in any of Sumicsid's outputs. Furthermore, we are unable to validate the weights that Sumicsid used when we conduct similar regression analysis—that is, the coefficients of the regression analysis did not match the weights used by Sumicsid. Changing the weights used to aggregate asset classes to those implied by our regression analysis has a positive impact of up to 8 percentage points on one of the TSOs in Sumicsid's model.

Deriving weights based on regression analysis relies on parametric assumptions that are inconsistent with the non-parametric nature of DEA. In this context, it may be more appropriate to consider each asset class as a separate output and to allow the DEA model (if required with weight restrictions) to determine the correct weights on each asset class.

We find that replacing the outputs in Sumicsid's model with the four largest components of NormGrid has a material impact of up to 29 percentage points on the estimated efficiency of individual TSOs.

iv. The environmental adjustment to NormGrid is not supported by statistical evidence

To account for exogenous environmental factors, Sumicsid adjusts the NormGrid measure with an 'environmental complexity factor'. This complexity factor is based on the land-use characteristics of the area served by the TSOs (for example, the proportion of service area that is urban). Sumicsid presents the weights it uses to adjust NormGrid in one of the workshops (the W5

[26] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, p. 32.
[27] Sumicsid (2019), 'Norm Grid Development', February, p. 2.

workshop),[28] but it does not provide evidence to support the values of the weights.

Our analysis of the environmental adjustment indicates a negative relationship between unit costs (defined as unadjusted TOTEX per NormGrid) and the environmental complexity factor—that is, TSOs that operate in more complex environmental conditions (as defined by the environmental complexity factor) have a lower unit cost. This counterintuitive relationship may be partially explained by the way in which the environmental complexity factor is constructed. In particular, the percentage of service area covered by forests is the land-use characteristic that has the biggest impact on the overall complexity factor for most TSOs. Also, factors that may be more operationally intuitive drivers of expenditure (such as urbanity or mountainous areas) have a low impact on the overall complexity factor.

3. Sumicsid has not justified the assumptions that it has made in its model, and its approach to model validation is incapable of detecting flaws or omissions in its model.

For the results from any benchmarking model to be considered reliable, the assumptions of the model must be justified and the model itself must be robustly validated. Sumicsid makes several claims regarding the statistical tests that it has undertaken to support its modelling assumptions, yet it does not present the empirical evidence in its final outputs.

i. Sumicsid's returns-to-scale assumption is not supported by statistical evidence

One of the key assumptions in DEA relates to the specification of the returns-to-scale assumption. Sumicsid has assumed a 'non-decreasing returns to scale' (NDRS) technology when estimating TSOs' efficiency scores, and it states that this is supported by statistical evidence. However, this statistical evidence is not reported in the outputs.

In our replication of Sumicsid's tests, we do not find conclusive evidence supporting the NDRS assumption in the final model. Analysis of the estimated efficiency scores in the DEA model indicates that a 'variable returns to scale' (VRS) assumption may fit the data more appropriately.

If the model is estimated using the variable returns-to-scale assumption, four additional TSOs become peers, increasing their estimated efficiency by up to 30 percentage points, while two TSOs that are peers in Sumicsid's analysis become inefficient.

ii. Sumicsid's outlier procedure is not justified in its report and is scientifically inadequate

Sumicsid has relied on German regulatory precedent to detect outliers. Specifically, Sumicsid has performed dominance and super-efficiency tests based on the Bundesnetzagentur's approach to outlier detection, as outlined in the Incentive Regulation Ordinance (ARegV). The decision to follow the outlier procedure specified in the ARegV is not justified, nor is the outlier procedure likely to be sufficient in an international benchmarking context.

We recommend the following amendments to Sumicsid's outlier procedure.

[28] Sumicsid (2019), 'Model Specification Model Results', April, slide 55.

- **Dominance test.** Following the recommendations of Kumbhakar, Parthasarathy and Thanassoulis (2018) [29] in their expert opinion on Sumicsid's dominance test, we apply a 'bootstrap-based test' for dominant TSOs. Sumicsid's dominance test has no theoretical foundation in the context of outlier analysis. The bootstrap-based test provides a robust foundation as it is a non-parametric test consistent with DEA and can better take the specific context (i.e. outlier analysis) into account. In the current case, the bootstrap test identifies one additional outlier, improving average efficiency across the sample by 2 percentage points.

- **Super-efficiency test.** Consistent with the recommendations in Deuchert and Parthasarathy (2019)[30] and Thanassoulis (1999),[31] we apply the super-efficiency test iteratively until no more super-efficient outliers are identified. This modification increases the number of detected outliers by three and results in a more homogenous sample on which DEA can be performed. In the current case the iterative application of the super-efficiency test identifies three additional outliers, improving average efficiency across the sample by 3 percentage points.

There are further issues with these tests that are not addressed with these amendments. We also note that outlier procedure is not a replacement for robust data collection, validation and model-development process.

iii. Sumicsid's second-stage analysis is incapable of detecting omitted cost drivers and does not support the final model

In order to test whether relevant drivers of expenditure have been omitted from the final model specification, Sumicsid uses regression analysis involving the estimated efficiency scores and the omitted cost drivers.

As noted in Kumbhakar, Parthasarathy and Thanassoulis (2017),[32] 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. In addition, the use of second-stage analysis requires assumptions that need to be justified, and Sumicsid has not presented such justification in its output.

We demonstrate that Sumicsid's second-stage analysis is unable to validate its own model specification. To show this, we estimate 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 'omitted' or not. We find that Sumicsid's approach only identifies transformer power, and not NormGrid, weighted lines or the environmental adjustment, as a relevant omitted variable. Thus Sumicsid has not demonstrated that there are no relevant variables omitted in its sole model.

[29] Given the non-applicability of the ARegV in the current context, as noted in Kumbhakar, Parthasarathy and Thanassoulis (2019),[29] the test can be easily amended to improve on its discriminatory power. See Kumbhakar, S., Parthasarathy, S. and Thanassoulis, E. (2018), 'Validity of Bundesnetzagentur's dominance test for outlier analysis under Data Envelopment Analysis', August.
[30] 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, *ew–Magazin für die Energiewirtschaft*.
[31] Thanassoulis, E, (1999) 'Setting Achievement Targets for School Children', *Education Economics*, **7**:2, pp. 101–19.
[32] 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.

iv. Sumicsid has not examined whether the DEA outputs are consistent with operational expectations

Sumicsid has not validated the outputs from its DEA modelling. For example, DEA weights can be used to assess the importance of each cost driver in determining a TSO's efficiency. Sumicsid states that such weights can be used to identify potential data errors,[33] but it does not present any analysis of DEA weights in its final outputs. Sumicsid has not ensured that the DEA outputs are operationally intuitive and valid for the TSOs. This is necessary to show that that method is appropriate for the dataset and model.

Furthermore, Sumicsid makes statements regarding the importance of each cost driver that are not supported by empirical evidence. For example, Sumicsid states that environmentally adjusted NormGrid is the primary driver of expenditure in its DEA model, yet four TSOs' efficiency scores are not determined by NormGrid at all (i.e. NormGrid has zero weight) and a further six TSOs do not have NormGrid as the main driver of efficiency (i.e. NormGrid has less than 33% weight). If the ex ante expectation is that adjusted NormGrid is the primary driver of costs, then this analysis of DEA weights is concerning and could indicate data errors or model mis-specification.

Moreover, each inefficient unit will have its corresponding set of efficient peers scaled up or down to provide an efficient benchmark. Sumicsid has not presented any discussion of whether the identified peers and their weights as estimated by its model are appropriate. Indeed, we find evidence that inefficient TSOs are being benchmarked against peers that are up to 12 times smaller, and thus are not necessarily comparable. There are other instances of such unusual scaling factors that have not been validated by Sumicsid.

v. Sumicsid has not cross-checked the results of its analysis using alternative methods

The results from DEA are contingent on certain assumptions imposed on the model (and clearly on the underlying dataset) that have not been sufficiently justified by Sumicsid. Moreover, Sumicsid's application of DEA is deterministic and unable to account for data errors or modelling uncertainty. Because of this, alternative benchmarking techniques such as stochastic frontier analysis (SFA)[34] should be used as a cross-check to the DEA results, alongside operational evidence.

In its main report, Sumicsid acknowledges SFA as a valid tool for efficiency benchmarking, but states that the sample size was too small to allow for 'a full scale application of SFA as a main instrument'.[35] However, the appropriate sample size is an empirical question, applies to all empirical methods, and cannot be decided ex ante—there is no fixed rule as to how many observations a model needs for the analysis to be robust. SFA has been performed on

[33] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, p. 30.
[34] SFA is an econometric approach to benchmarking regulated companies. For a more detailed discussion on SFA, see Kumbhakar, S. and Knox Lovell, C.A. (2000), *Stochastic Frontier Analysis*, Cambridge University Press, Kumbhakar, S.C, Wang, H-J and Horncastle, A. P. (2015), *A Practitioner's Guide to Stochastic Frontier Analysis Using STATA*, Cambridge University Press, and 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, *ew–Magazin für die Energiewirtschaft*.
[35] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, p. 29.

smaller samples in regulatory context[36] and Sumicsid had access to a panel dataset.[37] Equally, Sumicsid should have validated the outputs from DEA (e.g. peers and weights) to show that the method was appropriate for the dataset, but it has not.

As a general related observation, we note that despite deriving its model on a panel dataset (i.e. data over time across TSOs), Sumicsid has not effectively used all the information that it has at its disposal. Instead, it has—without justification—focused on a single year's data for estimating the TSOs' efficiency levels.

We have managed to apply SFA on Sumicsid's model on a cross-sectional basis, as well as on a panel basis. In our estimation of SFA models, we find 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 among the TSOs, with the deviations due to statistical noise being potentially identified as inefficiency by the DEA model. The SFA model not finding statistically significant inefficiency is not a reason to use DEA; rather, it suggests caution is warranted against interpreting any estimated inefficiency in the DEA as actual inefficiency rather than statistical noise, and/or that the model specification should be re-examined.

vi. Dynamic efficiency (frontier shift) analysis casts additional doubt on the validity of the data and model

Dynamic efficiency relates to the ability of the most efficient companies in an industry to improve productivity. Despite discussing dynamic efficiency results in one workshop,[38] Sumicsid has not presented any relevant analysis in the final outputs. On Sumicsid's final model, DEA indicates that there has been a frontier *regress* of 4% p.a.[39] That is, efficient costs have been increasing at a rate of 4% p.a. over the period of assessment (i.e. 2013–17). When applying SFA on the same model, we estimate a similar rate of frontier shift, although it is statistically indifferent from zero (consistent with the conclusion of the individual inefficiency estimates).

Such a large and negative frontier shift result is unusual when compared with what is commonly estimated in regulatory settings, and this could indicate that Sumicsid's model cannot capture changes in costs over time—for example, relevant cost drivers that control for the position of a TSO in the investment cycle (such as asset health) are missing.

This is concerning; if the model cannot capture *changes in efficient costs over time*, then it is likely that the model cannot capture differences in *efficient costs between TSOs*.

The volatility in expenditure that is not captured by changes in the cost drivers is further evidence that the data is measured with a high level of uncertainty. For this reason, the resulting efficiency scores are likely to be estimated with a large degree of uncertainty; while DEA, as applied by Sumicsid, does not provide information on uncertainty surrounding the estimated efficiencies, this

[36] E.g. the Office for Rail and Road (ORR) performed SFA on a sample of 50 observations for its determination of the efficiency of Network Rail as part of the PR18 price control. Office of Rail and Road (2018), 'PR18 Econometric top-down benchmarking of Network Rail A report', July
[37] A cross sectional dataset contains one observation per unit (i.e. TSO) for one year. A panel dataset contains data over time across TSOs
[38] See Sumicsid (2019), 'Model Specification Model Results', April, slide 81.
[39] Sumicsid has now published the results of the dynamic efficiency analysis after the finalisation of this report. See Sumicsid (2020), 'Dynamic efficiency and productivity changes for electricity transmission system operators', April.

uncertainty must be accounted for in some way (for example, through Monte Carlo simulations or SFA) if the subsequent scores are to inform regulatory or operational or valuations applications.

Summary of our assessment

The table below summarises the impact of the individual adjustments we suggest under the three themes described above. Our recommendations are principles-driven and consistent with the scientific literature, and they can therefore result in a decrease or increase in the estimated efficiency of a TSO in a particular sensitivity.

We have not presented the combined results, in which all our recommendations are jointly implemented, as we have identified numerous fundamental issues with Sumicsid's analysis. Further, given the unreliable nature of the data and the overarching conclusion from SFA that the data is too noisy, we have not developed alternative model specifications with outputs as cost drivers.

The TCB18 study suffers from a number of fundamental weaknesses that mean that the estimated efficiency scores cannot be used for regulatory, operational or valuations purposes in their current form. We have outlined a number of recommendations to be considered at each step of Sumicsid's analysis. These require significant additional work to ensure that the outcomes from the TCB18 are robust and fit for purpose.

Summary of our assessment

Issues in the TCB18 study		Specific empirical impact	Wider modelling implications	Report section
Data collection and construction	Data errors	Monte Carlo simulation indicates that four inefficient TSOs become efficient at the 90% significance level. The widest confidence interval is 48 percentage points.	Data errors will have an impact at all stages of the benchmarking, including model development.	3.1
	Defining the input variable	In a two-input model, one efficient TSO becomes inefficient and two inefficient TSOs become efficient. Alternative options exist and need to be explored.	The modelling implies the relationship between OPEX and CAPEX assumed by Sumicsid is restrictive and that alternative models must be explored.	3.2
	Adjusting for input prices	The impact of our preferred PLI adjustment on TSOs' efficiency is in the range of -34–16 percentage points.	Most TSOs' efficiencies are highly sensitive to the method of adjustment, highlighting the importance of the adjustment and the need for sensitivity analysis.	3.3
	Indirect cost allocation	One TSO reduces its efficiency by 18 percentage points.	The allocation of expenditure is a conceptually important issue and can have material impact if other adjustments are made to Sumicsid's analysis and in future iterations.	3.4
Model development	Cost driver analysis	Our validation of Sumicsid's econometric analysis does not support its final model. Its assumed functional form and estimation approaches are not conclusively supported by empirical evidence and cast doubt on the statistical validity of the model.	Sumicsid's approach invalidates any of its statements regarding statistical significance.	4.1
	Sample sensitivity	The coefficients of the econometric model are highly sensitive to the exclusion of some TSOs. Many TSOs' efficiencies are highly sensitive to the year in which they are assessed.	If a model is sensitive to the inclusion of some TSOs, it may be poor at explaining industry-wide costs. The variability in a TSO's relative efficiency from year to year indicates yearly effects and investment cycles are not appropriately captured.	4.2
	The use of asset-based outputs	Using an alternative capacity measure increases one TSO's efficiency by 15 percentage points and reduces another's by 39 percentage points.	Asset-based models favour TSOs that have deployed expensive assets for the same contextual features and may therefore bias the efficiency estimates in favour of costly TSOs. It is also unusual to only consider input–input models.	4.3
	Aggregation of NormGrid	Modelling components of NormGrid separately increases efficiencies of some TSOs by 29 percentage points and significantly changes the distribution of efficiency.	If NormGrid is to be used, the weights must be robustly justified. This will impact final efficiency scores and model development.	4.4
	Adjusting for environmental factors	The environmental adjustment is not positively correlated with unit costs.	The environmental adjustment has not been validated. Models with environmental characteristics as exogenous drivers of expenditure and other alternatives should be explored.	4.5

Application and validation	Returns-to-scale assumption	We find no conclusive statistical evidence supporting the NDRS assumption. Under VRS, the classification and estimates of the TSOs' efficiency changes significantly.	The RTS assumption should be consistent with other areas of the analysis and supported by statistical and operational evidence.	5.1
	Outlier analysis	The bootstrap-based dominance test identifies two additional outliers and the iterative super-efficiency test identifies three additional outliers.	Sumicsid's outlier procedure is insufficient and flawed. Any outlier analysis is not a replacement for a robust data collection and model-development process.	5.2
	DEA outputs	NormGrid is not the primary driver of costs for most TSOs. Peers and their weights are unusual in some cases and non-validated.	The DEA weights suggest that NormGrid is not the primary cost driver for most TSOs. Scaling factors on peers suggest that TSOs are being benchmarked to peers that are not comparable.	5.3
	Identification of omitted cost drivers	In a two-output model, we are not able to detect NormGrid or weighted lines as relevant omitted variables.	Sumicsid's second-stage analysis for model validation has no theoretical foundation and is not able to detect relevant omitted cost drivers. Relevant cost drivers must be tested in the model-development and validation phases.	5.4
	SFA	The SFA models (applied on a cross-sectional and panel basis) do not detect any statistically significant inefficiency.	This is further evidence that the data is 'noisy' and the estimated efficiency gaps as identified by DEA do not represent genuine differences in efficiency.	5.5
	Dynamic efficiency	The estimated dynamic efficiency is -4% p.a. using the DEA model. This is supported by SFA models. Moreover, the frontier shift is statistically indifferent from zero, consistent with the conclusion of the individual inefficiency estimates.	The dynamic efficiency analysis indicates that the model cannot capture changes in costs over time and provides further evidence that the analysis is contaminated by statistical noise.	5.6

Source: Oxera analysis.

Recommendations for further development of TCB18 and in future iterations

Some of the weaknesses in Sumicsid's model, such as consistency in reporting guidelines, could be partly driven by the lack of maturity in international benchmarking processes and may improve with time. However, the analysis presented in the TCB18 study also requires significant improvements. In this regard, we consider that it will be helpful to have debriefs involving all the parties on process (for example, in terms of data processing and validation) and methodology to help future studies.

Our recommendations include the following.

- Provide a clear conceptual (and, where possible, empirical) justification for any assumptions that feed into each stage of the benchmarking process.
- Relatedly, provide detailed description in the outputs and publish modelling codes to aid transparency (similar to the output presented with this report).
- Establish an iterative data-collection procedure (including validation and cross-checking exercises) to ensure that data is reported correctly and consistently across TSOs and validate the reported data.
- Use statistical analysis, such as Monte Carlo simulations, to evaluate the impact of any potential data errors (especially if using deterministic methods for efficiency estimation). This could then be used for deriving confidence intervals around the estimated efficiency scores. Alternative methods, such as SFA, could also inform the extent of the uncertainty adjustment applied in the simulations, apart from operational evidence/expert judgement.
- Robustly capture the impact of all input price differences on expenditure to avoid conflating efficiency and this exogenous factor.
- Perform a scientifically valid model-development process that includes consultations with the TSOs throughout and: (i) is based on realistic modelling assumptions; (ii) tests the significance of alternative model specifications; (iii) tests the sensitivity of the analysis to small changes in the sample: and (iv) avoids the restriction of cost drivers to asset-based outputs.
- Relatedly, the analysis should not be too sensitive to the year in which efficiency is assessed. If the estimated efficiency of TSOs fluctuates significantly from year to year, the causes of this (e.g. investment cycles) must be explored.
- If asset-based measures are used, these must be validated through comparisons to outputs.
- Provide robust statistical evidence to support modelling assumptions.
- Develop a robust outlier-detection procedure that is consistent with the academic literature and appropriate in an international benchmarking context.
- Analyse the outputs of a DEA model, such as cost driver weights, peers and their weights, to ensure that they are consistent with economic and operational intuition.
- Avoid relying on second-stage validation to detect omitted cost drivers. In a DEA context, the impact of omitted cost drivers should be assessed by testing the sensitivity of the results to alternative model specifications.

- Cross-check the analysis with alternative benchmarking methods, such as SFA, to validate whether the estimated efficiency scores can be attributed to genuine differences in efficiency, or data uncertainty, or the choice of benchmarking method.
- Make effective use of panel data and estimate dynamic efficiency on data and model to validate the results, as this can also help to identify flaws with the model that are not evident from cross-sectional analysis.

1 Introduction

The transmission cost benchmarking project (TCB18), a study of the cost performance of European transmission service operators (TSOs), covering gas[40] and electricity,[41] was commissioned by the Council of European Energy Regulators (CEER) and performed by its consultancy, Sumicsid. A consortium of European electricity TSOs asked Oxera to perform a shadow benchmarking exercise of the TCB18 project.

In electricity, TCB18 covered 17 European TSOs. Oxera has obtained the full sample of data for all electricity TSOs that participated in TCB18 through the shadow study. Using this dataset, we were able to validate, critique, and improve upon Sumicsid's analysis.

Through the extensive analysis in this report, we highlight several significant flaws with the study. We offer recommendations on how the results from the study should be interpreted and on the additional research required before the results could be used to determine revenue cap. We also offer recommendations on how the analysis could be improved in future editions of the study.

This report is structured as follows:

- section 2 provides a brief factual summary of the TCB18 study;
- section 3 critically examines Sumicsid's data-collection and data-construction exercises;
- section 4 assesses Sumicsid's approach to model development;
- section 5 reviews Sumicsid's application and validation of its final model;
- section 6 concludes.

In carrying out this shadow benchmarking exercise, we have drawn on Sumicsid's publicly available final report, the associated appendices,[42] the TSO-specific reports shared with the TSOs,[43] and the slides from the workshops undertaken by Sumicsid as part of the TCB18 study.

[40] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for gas transmission system operators', July.
[41] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July.
[42] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators appendix', July.
[43] For example, see Sumicsid (2019), 'Project TCB18 Individual Benchmarking Report Fingrid – 131', July.

2 Summary of Sumicsid's TCB18 approach

This section gives an overview of the TCB18 study to provide context for the issues outlined in sections 3–5.

An efficiency benchmarking exercise can broadly be divided into three phases.

1. **Data collection and construction.** Here, data from TSOs is collected, audited and screened for errors (such as misreporting or anomalies). After one is confident that the data is (to a reasonable degree) without errors, it needs to be normalised for differences in reporting (e.g. accounting guidelines can vary, expenditure needs to be converted into a single currency), regulatory frameworks and operational characteristics to ensure that the cost base is comparable across TSOs.

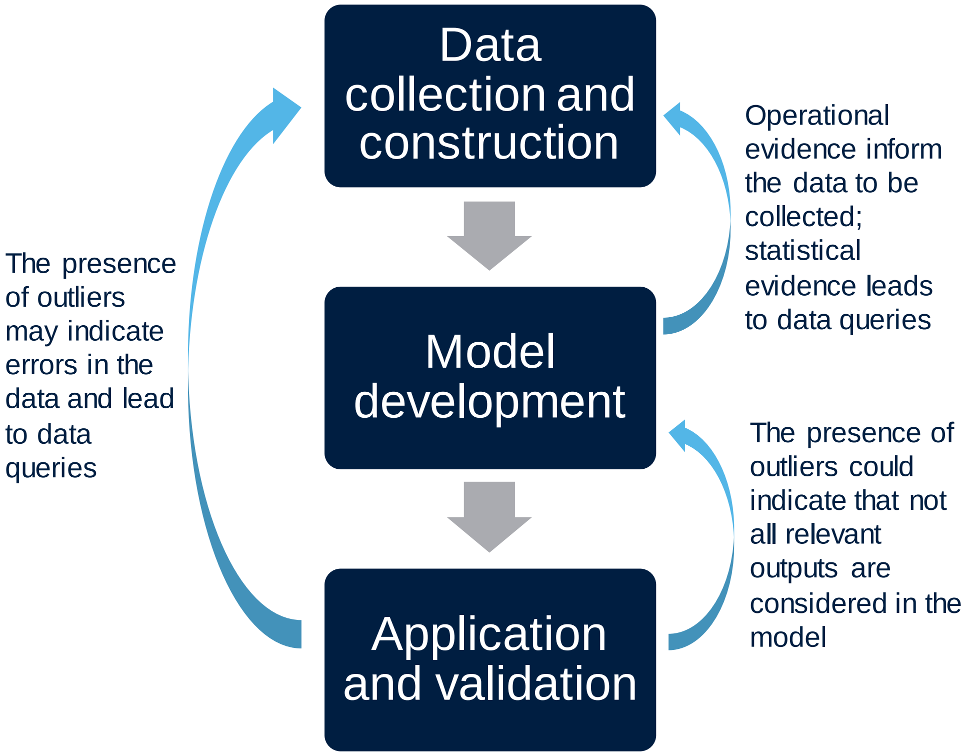
2. **Model development.** Given the data, the model for benchmarking is derived based on a combination of scientific procedure and expert judgement. This concerns the definition of costs (e.g. TOTEX), cost drivers (e.g. network length, environmental factors), the approach to the treatment of outliers given the chosen model, and the choice of benchmarking model, such as data envelopment analysis (DEA)[44] and stochastic frontier analysis (SFA),[45] and motivating the assumptions underpinning each step.

3. **Application and validation.** Once the model specification(s) and method are selected based on best available data, the model needs to be robustly estimated. This involves, among other things, validating the results from the model, as well as undertaking robust sensitivities to ensure that the results are not driven by specific assumptions made by the modeller.

In reality, this is not a sequential procedure but a highly iterative one, with feedback occurring between the various steps. Even the results from the final application of the benchmarking model may highlight additional data and modelling queries that necessitate further analysis. The process is illustrated in Figure 2.1 below.

[44] DEA is a mathematical non-parametric approach that is widely used when benchmarking regulated companies. For a more detailed discussion on DEA, see Thanassoulis, E. (2001), *Introduction to the Theory and Application of Data Envelopment Analysis: A Foundation Text with Integrated Software*, Springer.
[45] SFA is an econometric approach to benchmarking regulated companies. For a more detailed discussion on SFA, see Kumbhakar, S.C, Wang, H-J and Horncastle, A. P. (2015), *A Practitioner's Guide to Stochastic Frontier Analysis Using STATA*, Cambridge University Press.

Figure 2.1 Benchmarking process



Source: Oxera.

We summarise each step taken by Sumicsid below.

2.1 Sumicsid's approach to data collection and construction

In order to benchmark the cost performance of the European TSOs, it is essential to construct a homogenised dataset on the cost and outputs of the participating TSOs to enable a like-for-like comparison. That is, cost and outputs must be reported consistently (for example, in terms of allocating costs to specific activities), and the activities performed by the TSOs must be broadly similar.

Sumicsid states that it followed a six-stage approach to data collection and validation.[46]

1. **Asset system and audited financial statements.** This involved the collection of asset system data and audited financial statements of TSOs.
2. **Clear guides/templates.** CEER, Sumicsid and the TSOs worked interactively to establish reporting definitions to translate the data from the first stage into something that could be used for international benchmarking.
3. **Interaction (e.g. workshop).** There was interaction between the TSOs, NRAs and Sumicsid at all stages in the data-collection process to ensure the correct interpretation of reporting guidelines.
4. **National validation.** The national regulatory authorities (NRAs) validate the data to ensure the data is 'complete, consistent, correct and plausible'.

[46] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, section 3.2.

5. **Cross-validation.** Sumicsid performed an additional cross-validation exercise to identify and correct for any remaining misinterpretations of reporting guidelines.

6. **Data analysis.** Related to the cross-validation process, Sumicsid continued to review the data during the modelling process. The modelling process may detect errors or misreporting that were not identified in the previous steps.

As part of its final data checks, Sumicsid states that all TSOs participating in the TCB18 study received ‘a dump of asset and financial files’[47] that they could review for missing or incorrect data. At this stage in the process, there were ‘a few final corrections’ for ‘many’ TSOs.[48]

For the TCB18 study, Sumicsid also collected data from external sources relating to inflation rates,[49] price-level differences[50] and environmental factors.[51]

Activities assessed in the benchmarking study

The management of an electricity transmission network is an extremely complex operation. Sumicsid categorises each activity undertaken by TSOs as follows.[52]

- **Transport (T).** This entails the core service of TSOs in transporting electricity from generators to connection points (such as a downstream network, another TSO, or an end-client).
- **Grid maintenance (M).** Grid maintenance involves the repair of grid assets (both preventative and reactive) and the replacement of degraded equipment.
- **Grid planning (P).** This involves the analysis, planning and drafting of network expansion.
- **Indirect support (I).** Indirect expenditure includes administrative support functions (such as human resources and IT) that cannot be allocated to a specific activity. This includes central management costs.
- **System operations (S).** This involves assessing the real-time energy balance, failure detection and analysis, and maintaining technical quality.
- **Market facilitation (X).** Market facilitation involves the facilitation or management of electricity marketplaces.
- **Offshore transport (TO).** This entails the transport of electricity through offshore assets .
- **Other activities (O).** TSOs may perform activities that cannot be classified into the above seven activities.

[47] Sumicsid (2019), ‘Pan-European cost-efficiency benchmark for electricity transmission system operators main report’, July, p. 8.
[48] Sumicsid (2019), ‘Pan-European cost-efficiency benchmark for electricity transmission system operators main report’, July, p. 8.
[49] Sumicsid (2019), ‘Pan-European cost-efficiency benchmark for electricity transmission system operators main report’, July, Table 4.3.
[50] Sumicsid (2019), ‘Pan-European cost-efficiency benchmark for electricity transmission system operators main report’, July, Table 4.1.
[51] Sumicsid (2019), ‘Pan-European cost-efficiency benchmark for electricity transmission system operators main report’, July, Table 3.1.
[52] For more detail regarding how Sumicsid has defined these activities, see Sumicsid (2019), ‘Pan-European cost-efficiency benchmark for electricity transmission system operators main report’, July, sections 4.4–4.12.

Sumicsid considers T, M, P and a subset of I to be within the scope of the benchmarking study, and allocated costs and outputs on this basis. It claims that this was driven by ‘analysis of common factors in cost reporting, the variability and homogeneity of the data and the separability of the activity’.[53]

Cost construction

Sumicsid assesses cost performance of the TSOs on a total expenditure (TOTEX) basis, where TOTEX is the sum of operating expenditure (OPEX) and capital expenditure (CAPEX). The OPEX is calculated on an annual basis in real terms (i.e. taking into account general price inflation) with no smoothing across years. Cost items within OPEX that are deemed uncontrollable or out-of-scope[54] are excluded from the analysis. Sumicsid applies an adjustment to labour costs using the price level index (PLI) for civil engineering works to account for differences in wage rates across the TSOs. Finally, the OPEX for TSOs operating outside of the euro area is converted to euros using the average exchange rate in 2017 (for all years).

CAPEX is typically ‘lumpy’; a TSO may construct a large segment of the network in a particular year, and this will be registered as a large increase in investment. To avoid the efficiency estimates being driven by a TSO’s position in the investment cycle (or its ‘age'), regulators often consider methods of ‘smoothing’ CAPEX.

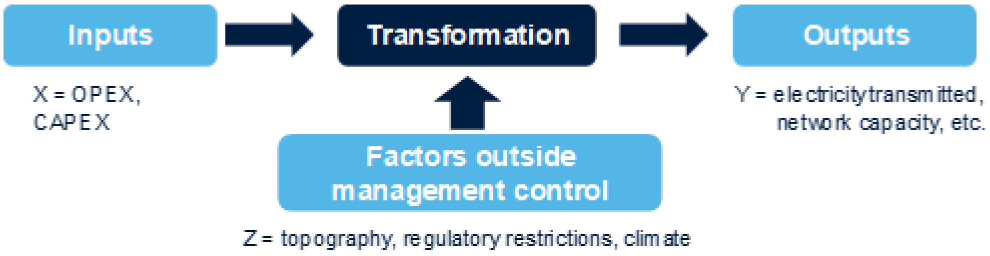
In this case, Sumicsid has taken the annuity approach to CAPEX measurement, whereby the cost of the investment in a particular asset is spread over the asset’s lifetime. Specifically, Sumicsid uses investment stream data from 1973, assumed certain ‘techno-economic lifetimes’ of each asset, and assumed a real interest rate of 3% to calculate the annuity on each asset.[55] Assessed CAPEX in a particular year is the sum of the annuities for assets within the scope of benchmarked activities.

2.2 Sumicsid’s approach to model development

Cost benchmarking requires the comparison of homogenous units. However, TSOs operate at different scales, produce different mixes of outputs, and have different regulatory requirements and operating conditions (such as differences in topography and demography). In order to account for such severe heterogeneity, it is essential that the drivers of (efficient) expenditure are sufficiently understood and controlled for. In this context, we assess the process shown in Figure 2.2 below.

[53] Sumicsid (2019), ‘Pan-European cost-efficiency benchmark for electricity transmission system operators main report’, July, p. 18.
[54] These include energy costs, landowner compensation, right-of-way and easement fees, taxes and levies, depreciation, research and development and the rent of the main office building
[55] $A = CAPEX * \frac{r}{(1-(1+r)^{-T}}$, where A is the annuity, r is the assumed interest rate and T is the assumed lifetime of the asset. CAPEX is the expenditure for the asset. This formula splits CAPEX into T constant payments of A. For example an investment of 1,000,000€ in an asset with a lifetime of 50 years at an interest rate of 3% would result in an annuity of $A = 1{,}000{,}000€ * \frac{0.03}{1-(1+0.03)^{-50}}$ =38,865.49€ per year for 50 years.

Figure 2.2 Production process



Source: Oxera.

Relevant cost drivers (outputs or environmental factors) are included in the benchmarking model in order to homogenise the characteristics of different TSOs. Sumicsid states that in an ideal setting, the cost drivers should be:

- exogenous—i.e. outside of management control;
- complete—i.e. accounting for all operating characteristics;
- operable—i.e. clearly defined and measurable;
- non-redundant—i.e. the set should be as small as possible to avoid unnecessary duplication.[56]

While Sumicsid does not discuss this in the published report, it notes in some workshop slides that it has used econometric analysis to test the relevance of candidate cost drivers. In particular, Sumicsid states it has used 'robust OLS' (ROLS) to estimate and validate the relationship between costs and cost drivers in the main report.[57] ROLS is an extension of the OLS estimator, which attaches less weight to observations that are further from the regression line[58] (i.e. observations that fit the assumed model less well). Based on Sumicsid's final report, it appears to have particularly focused on the model-fit of such cost drivers (i.e. the extent to which variations in the cost drivers in the model can explain variations in costs) when determining the appropriate outputs to use in its benchmarking model.[59]

The final outputs used in Sumicsid's benchmarking model are the following.

- Environmentally adjusted Normalised Grid ('NormGrid')—a measure of the assets used to deliver outputs, adjusted for the land-use characteristics in the TSO's service area. This output is discussed in more detail below.
- Transformer power—the sum of power across all transformers owned by TSOs.

[56] See Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, p. 28.
[57] See Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, Table 5.3.
[58] Regression analysis is a statistical method that separates out the impact of different factors in explaining movements in the key variable of interest. This variable of interest is known as the 'dependent variable' (TOTEX in this context). The regression identifies how the different cost drivers contribute to explaining the observed values of the dependent variable, and whether the cost drivers are statistically significant. Simple regression, with one dependent variable and one cost driver, can be visualised graphically as a 'line of best fit'. This visualisation becomes more difficult as more cost drivers are added to the model.
[59] For example, Sumicsid presents only the adjusted R^2 for three models in a table in its report (see Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, Table 5.3). In addition, the workshop slides tended to focus on the adjusted R^2 for the validation of NormGrid (see Sumicsid (2018), 'Validation of NormGrid and Preliminary Environmental Results', November, slides 6–11).

- Weighted lines—the total length of lines multiplied by the share of angular towers and share of steel towers.

NormGrid construction

NormGrid can be considered as an aggregated measure of the in-scope assets deployed by TSOs. Sumicsid describes it as 'a cost-norm for the construction costs for the standard assets'.[60] Specifically, NormGrid is calculated as the weighted sum of in-scope assets according to the equations below.

$$NormGrid = NormGrid_{OPEX} + NormGrid_{CAPEX}$$

$$NormGrid_{OPEX} = \sum_{t}\sum_{a}(N_{at} * w_a)$$

$$NormGrid_{CAPEX} = \sum_{t}\sum_{a} n_{at} * v_a * \alpha(r, T_a)$$

Where:

- N_{at} is the number of assets of type a, acquired at time t;
- w_a is the OPEX weight for assets of type a;
- n_{at} is the number of assets of type a, acquired at time t and in prime age;
- v_a is the CAPEX weight for assets of type a;
- α is the annuity function, which is determined by:
  - the real interest rate, r;
  - the 'techno-economic life' of the asset, T_a.

The weight on each asset is supposed to account for the characteristics of the asset within an asset class. For example, it is typically more costly to construct and operate a high-voltage line than a low-voltage line. Sumicsid outlines the formula for each asset weight in an appendix to its main report,[61] but the formulae are not clearly supported by evidence in the report or its appendices.

Environmental adjustment

Sumicsid states:

> Environmental conditions influence the investment cost for, in particular overhead lines and underground cables, to a lesser extent the costs for transformers and other assets. […] Operating costs, including maintenance costs, are affected by a some additional factors by virtue of the location and configuration of the assets[62]

Sumicsid considered factors that could account for these differences in operating characteristics, including land use type, topography, vegetation type, soil humidity, subsurface features, extreme temperatures, and salinity. In its final report, Sumicsid states that 'extensive statistical tests revealed

[60] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, p. 25.
[61] Sumicsid (2019), 'Norm Grid Development', July, section 4.
[62] Sumicsid (2019), 'Norm Grid Development', July, section 5.

correlations and interaction between several of the factors', and that 'the most important factor for electricity was land use'.[63]

Sumicsid adjusted the NormGrid variable to account for land use type by multiplying it by an overall environmental factor. This factor is calculated by multiplying the share of a TSO's service area covered in a certain feature with a complexity factor. We illustrate the calculation of this factor for the average of the 17 TSOs' operating areas in Table 2.1 below.

Table 2.1 Calculation of the environmental adjustment

Feature	Factor (A)	Share of area covered (B)	Resulting adjustment (A*B)
Urban areas	1.5	3.8%	0.06
Infrastructure	3.5	0.2%	0.01
Agricultural, cultivated	1	24.7%	0.25
Forest	1.55	30.0%	0.47
Grass, meadows	1	8.9%	0.09
Shrubs, bushlands	1.1	5.4%	0.06
Without use	1	3.1%	0.03
Lakes, rivers, ponds	1.2	4.3%	0.05
Other	1	19.5%	0.20
Sum		100%	1.20

Source: Oxera analysis, based on Sumicsid data.

2.3 Sumicsid's approach to efficiency estimation and model validation

Sumicsid uses DEA[64] to estimate the relative efficiency of the TSOs. DEA uses linear programming models to estimate the minimum level of TOTEX needed for a TSO to meet its outputs (i.e. the cost drivers considered in the model) given the cost and output data for all the TSOs. The relatively efficient TSOs would be those for which no better performing TSO ('peer') could be identified in the dataset, and these TSOs would form the efficient frontier. DEA is further explained in Box 2.1.

[63] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, p. 33.
[64] Thanassoulis, E. (2001), *Introduction to the Theory and Application of Data Envelopment Analysis: A foundation text with integrated software*, Kluwer Academic Publishers.

Box 2.1 An overview of DEA

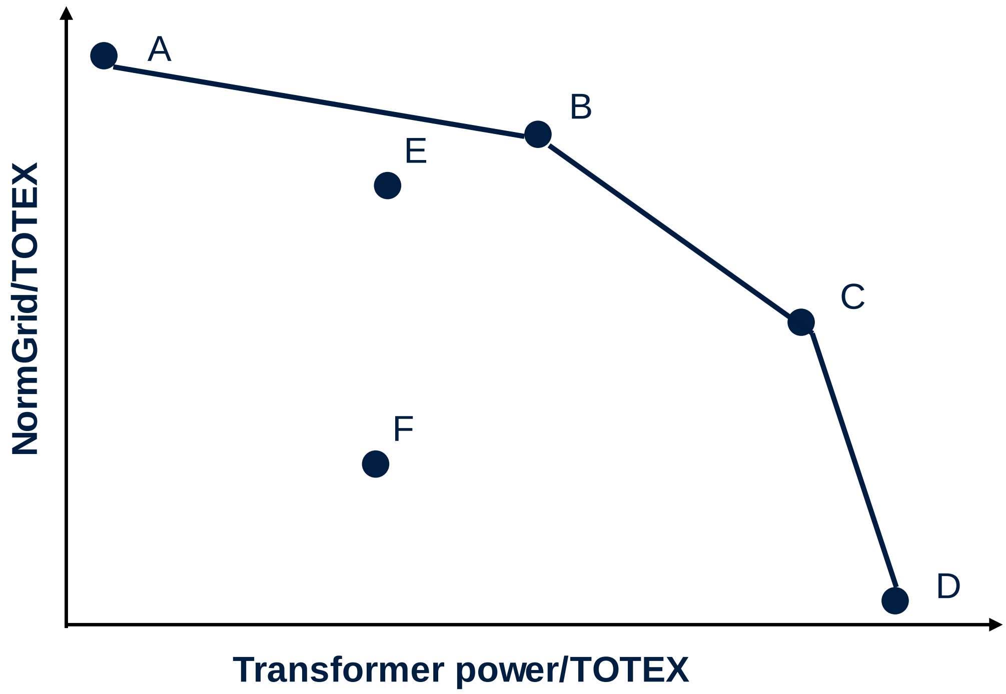
DEA has been used by a number of regulators and academics to assess the efficiency of companies and public service providers (such as in health and education). In its simplest form, DEA is an intuitive and transparent method for estimating the efficient frontier.

Suppose that all companies are identical in their production volume—for example, all firms serve the same number of customers who live in a comparable area. All one needs to do is order firms by their cost, and the one with the lowest cost sets the benchmark (it also has the lowest unit cost) or the efficient frontier.

However, the reality is far more complicated. The TSO dataset is characterised by severe heterogeneity, and a simple cost-by-cost comparison is not feasible. Multiple outputs are involved, and they differ across the TSOs. In this context, for an individual TSO, DEA tries to identify a peer TSO from existing (efficient) ones by combining them in certain proportions by means of linear programming. The peer TSO produces the same or more output than the TSO in question, but at lower cost (i.e. a lower level of TOTEX).

Figure 2.3 presents a stylised example of DEA in which NormGrid and transformer power are the two outputs, TOTEX is the input, and the technology exhibits constant returns to scale. TSOs A, B, C and D are identified as efficient as no other TSO in the sample produces more of any one output without producing less of another output for a unit of TOTEX. TSO E is estimated to be inefficient, as TSO B can produce more of both outputs for the same unit of TOTEX. Similarly, TSO F is estimated to be inefficient, as TSOs B and C can produce more outputs for the same unit of TOTEX. DEA can be extended to account for any number of inputs or outputs.

Figure 2.3 Stylised example of DEA



Note: It is assumed that interpolation between real TSOs leads to a feasible point for any TSO to operate. Therefore each point on the graph enveloped by the frontier ABCD and the frontier itself represents a real or virtual TSO. ABCD is the efficient frontier in that no real or virtual TSO can exceed one of its outputs without attaining less on the other output for a unit of TOTEX.

Source: Oxera.

Sumicsid states that according to convention, the number of input variables plus the number of output variables in a model should be less than one third of the number of observations.[65] In this case, the number of inputs is 1, the number of outputs is 3, and the number of observations is 17. Sumicsid states that as (1+3)=4 and 17/3~5.7, the model is therefore sufficiently discriminatory.

Returns to scale

One of the key steps in the application of DEA relates to the specification of the returns-to-scale assumption. This is explained in Box 2.2.

Sumicsid assumes non-decreasing returns to scale (NDRS)[66] in the model. NDRS can be seen as a variant of variable returns to scale (VRS), whereby TSOs may have increasing returns to scale when scale size is small (i.e. an increase (or decrease) in cost results in a more-than-proportionate increase (or decrease) in outputs), but operating under constant returns to scale (CRS) beyond a certain scale (in other words, no portion of the efficient frontier exhibits decreasing returns to scale). Sumicsid claims that this returns-to-scale assumption is supported by statistical evidence, but it does not present evidence to support this claim in its final outputs.

[65] Sumicsid (2019), ‘Pan-European cost-efficiency benchmark for gas transmission system operators’, July, p. 35.
[66] Otherwise known as ‘increasing returns to scale’ (‘IRS’).

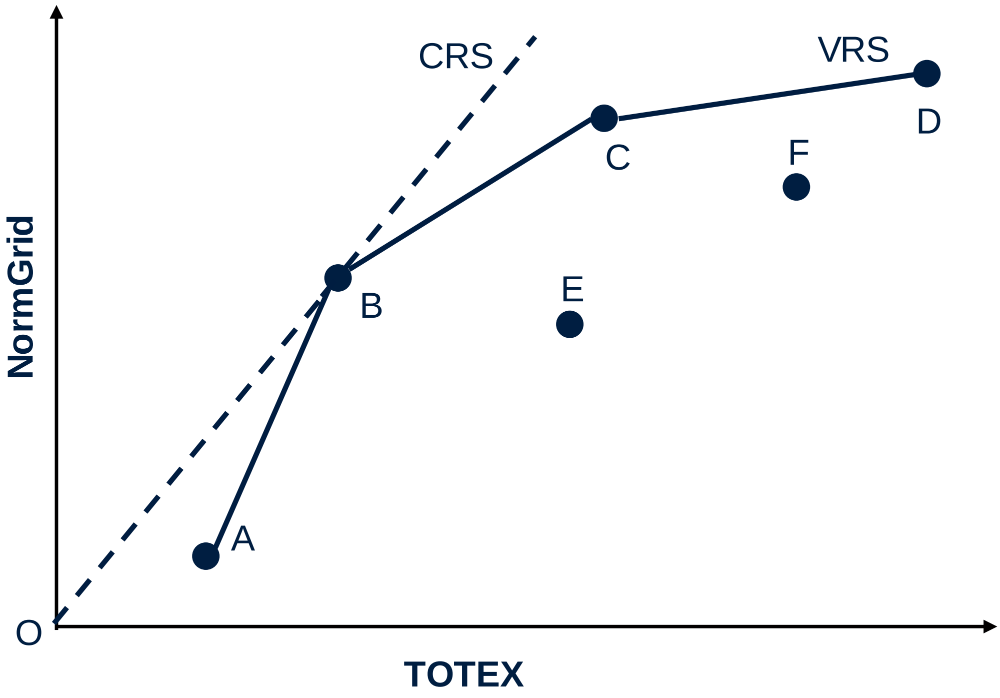
Box 2.2 Returns to scale

'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, the typical assumptions are constant returns to scale (CRS) or variable returns to scale (VRS).

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 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.

Figure 2.4 shows an illustrative example of different returns-to-scale assumptions in a single-input, single-output context. The line passing through OB represents the CRS efficient frontier, while the line ABCD represents the VRS frontier. Under VRS, the line AB exhibits increasing returns to scale in that for a unit rise in TOTEX (input) we have a more than proportional rise in NormGrid (output). In contrast, on BC and CD we have decreasing returns to scale in that for a unit rise in TOTEX we have a less than proportional rise in NormGrid. The key practical implication is that B exhibiting maximum NormGrid per unit TOTEX cannot be scaled under VRS to provide benchmarks as it can along OB under CRS. VRS is therefore generally a less demanding assumption than CRS where benchmark performance is concerned.

Figure 2.4 Illustration of returns to scale



Source: Oxera.

Outlier analysis

The particular form of DEA that Sumicsid has considered in its analysis is deterministic. Such an approach takes no account of statistical noise (e.g. random data errors) in estimating the efficient frontier or individual efficiency scores. To mitigate the impact of potential outliers, Sumicsid follows the same outlier procedure as applied by the Bundesnetzagentur and outlined in the

German Incentive Ordinance (‘ARegV’). Specifically, Sumicsid performs a dominance and a super-efficiency test (explained in Box 2.3) to detect and remove outliers.

Box 2.3 The Bundesnetzagentur’s outlier procedure

The Bundesnetzagentur is required to follow the methodology outlined in the ARegV to detect and remove outlier observations. In its application of DEA, the Bundesnetzagentur (i.e. or, specifically, the consultancy it has engaged) must remove the dominant and super-efficient outliers; these are defined below.

Dominance test

The aim of the dominance test is to identify companies that exert a substantial effect on the efficiencies of many other companies. The test, as outlined in the ARegV, compares the mean efficiency of all companies, including the potential outlier, with the mean efficiency calculated after excluding the potential outlier. If the efficiencies computed with and without the potential outlier are statistically different from each other at the 95% confidence level, the company is deemed dominant and removed from the sample.

By construction, removing one efficient company will increase the efficiencies of all companies to which it is a peer. To determine whether the difference in efficiencies with and without the company are statistically significant, the following test statistic is computed.

$$\frac{\sum_{k\in K\setminus i}(E(k;K\setminus i)-1)^2}{\sum_{k\in K\setminus i}(E(k;K)-1)^2}$$

Where:

- K is the total number of units (in this case, 16 TSOs);
- $E(k;K\setminus i)$ is the efficiency of TSO K estimated from the sample **excluding** the potential outlier, i;
- $E(k;K\setminus i)$ is the efficiency of TSO K estimated from the sample **including** the potential outlier, i.

This test statistic is assumed to follow an F distribution, and the company in question is removed if the value of the test statistic has a less than 5% chance of being randomly observed in the sample (i.e. a 95% confidence level).

Super-efficiency

The super-efficiency test aims to identify companies that are significantly more efficient than the rest of the sample. As defined in the ARegV, a company is considered super-efficient if its efficiency score when assessed relative to the rest of the companies (i.e. without it) exceeds the third-quartile efficiency value by more than 1.5 times the interquartile (i.e. between the third and first quartiles) range of efficiency values.

Furthermore, Sumicsid removed one TSO *before* the implementation of the outlier procedure. Sumicsid states that this TSO was ‘almost always an extreme outlier’ in its model-development phase,[67] but the empirical evidence for this was not provided in Sumicsid’s final outputs.

[67] Sumicsid (2019), ‘Pan-European cost-efficiency benchmark for electricity transmission system operators main report’, July, p. 35.

Model validation

After estimating the model, Sumicsid performs a certain ‘robustness analysis’[68] to test the sensitivity of the results to changes in the following specific modelling assumptions: interest rate, NormGrid calibration between OPEX and CAPEX, NormGrid weight for lines vs other assets, and salary corrections for capitalised labour in investments.[69] Sumicsid presents the impact of the changes on the average efficiency score in the sample. The impact on individual TSOs is not presented in the main report.

In the TSO-specific outputs, Sumicsid carries out second-stage analysis to test whether any relevant cost drivers have been omitted in the first-stage model used for efficiency estimation. Our understanding from the TSO-specific outputs is that in the second-stage analysis, Sumicsid regresses the estimated efficiency scores from the DEA model against potentially omitted drivers of expenditure (such as NormGrid weighted with slope factors) one at a time, according to the equation below.

$$Efficiency_i = \beta_0 + \beta_1 Y_i + \varepsilon_i$$

Where:

- $Efficiency_i$ is the estimated efficiency of TSO i;
- Y_i is the value of the omitted output of TSO i;
- ε_i is a random error component.

Sumicsid states that if the estimated coefficient $\hat{\beta}_1$ is statistically insignificant, the output in question, Y, is ‘already considered in the model and do[es] not merit specific post-run corrections’.[70]

It is not clear from the individual report whether the regression is estimated using OLS, ROLS or some other estimator, but Sumicsid states in the main report that:

> […] second-stage analyses are typically done using graphical inspection, non-parametric Kruskal-Wallis tests for ordinal differences and truncated Tobit regressions for cardinal variables.[71]

Sumicsid further states that such second-stage analysis of this sort is ‘routinely done’ to identify omitted cost drivers,[72] but provides no evidence to support this statement.

Having summarised Sumicsid’s approach, in the next three sections we critically review Sumicsid’s approach to data collection and construction, model development, and application and validation.

[68] Sumicsid (2019), ‘Pan-European cost-efficiency benchmark for electricity transmission system operators main report’, July, section 5.5.
[69] Sumicsid (2019), ‘Pan-European cost-efficiency benchmark for electricity transmission system operators main report’, July, p. 38.
[70] Sumicsid (2019), ‘Project TCB18 Individual Benchmarking Report Fingrid – 131’, July, p. 35.
[71] Sumicsid (2019), ‘Pan-European cost-efficiency benchmark for electricity transmission system operators main report’, July, paragraph 4.09.
[72] Sumicsid (2019), ‘Project TCB18 Individual Benchmarking Report Fingrid – 131’, July, p. 35.

3 TCB18 data collection and construction

This section outlines the issues that we have identified regarding Sumicsid's approach to the data collection, cleaning, validation, and normalisation procedures in TCB18. The following issues are covered.

- The dataset used in TCB18 contains significant data errors that were not corrected by Sumicsid (section 3.1).
- TSOs differ significantly in their cost structure. We consider the risk that this impacts on the estimated efficiency (section 3.2).
- Most of the cost data was not normalised for price-level differences (section 3.3).
- The allocation of indirect costs is influenced by uncontrollable costs. This risks conflating uncontrollable factors with inefficiency (section 3.4).

3.1 Data errors

3.1.1 Description of the issue

Any empirical analysis relies on the accuracy of the data being used, and real data tends to be noisy. These errors can, for example, result from issues such as:

- misreporting—for example, one TSO mistakenly enters some of its towers as 'steel' as opposed to 'wood';
- miscommunication—for example, 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);
- measurement errors—for example, some TSOs aggregate their towers in a way that seems to indicate that they have no angular towers, which was not spotted by Sumicsid.

All of the above scenarios occurred in TCB18 and are present in the final dataset on which the published results were estimated. These are empirical issues and their effect on the dataset needs to be explored.

While some methods are better able to account for data inaccuracies than others, DEA as applied by Sumicsid in TCB18 does not account for them at all. DEA (as applied by Sumicsid) is particularly sensitive to data errors because it takes no account of statistical noise (which includes certain types of data errors that are random and symmetrical). Efficiency estimates exclusively rely on the identified benchmark (i.e. relatively efficient) TSOs. Therefore, any TSO identified as a benchmark and placed on the efficient frontier because of a data error will affect the results for other TSOs for which this particular TSO acts as a benchmark. The reverse is also true, in that a TSO failing to be identified as a benchmark through a data error could also affect the position and shape of the efficient frontier.[73]

Therefore, a rigorous data-screening process is required before any empirical assessment of efficiency analysis. This is especially true in a regulatory

[73] Placing a TSO by error on the efficient frontier would adversely affect the efficiencies of TSOs for which it is a peer. Similarly, failing to place a TSO by error on the efficient frontier would benefit the efficiencies of TSOs for which it might have been a peer. As data errors can be random, in a comparative assessment we cannot know which TSOs benefited and which suffered in their comparative efficiency through TSOs being erroneously placed on the frontier or, conversely, through failing to feature on the frontier.

context, where data errors can have significant financial consequences for the regulated entities, which can result in a wider adverse impact for consumers.

3.1.2 Sumicsid's approach to data errors

As outlined in section 2.1, Sumicsid states that its data collection and validation procedure consisted of six stages.[74] Sumicsid states that 'although no approach will be fully safe' the datasets are of 'good quality'.[75]

3.1.3 Critique and proposed solution

While Sumicsid's description of the overall procedure to data collection that it has followed appears relatively consistent with good practice, in the shadow study, several TSOs noted that their data used by Sumicsid is not accurate. The examples in section 3.1.1 illustrate how all three potential data issues resulted in errors in the TCB18 data. Some TSOs flagged data inaccuracies to Sumicsid over the course of the study, but these were not corrected for in Sumicsid's final analysis. While we can correct for data errors that were pointed out to us, this is indicative of wider issues that cannot be fixed by adjusting the data for individual TSOs.

Sumicsid must therefore ensure that the patterns observed in the data—across TSOs as well as over time, as well as with respect to the chosen models and efficiency estimation methods—are consistent with operational expectations (determined through engineering and economic analysis). Where data errors exist, an appropriate solution is to have a rigorous data-cleaning process and a comprehensive sensitivity analysis. This should involve:

- extensive data checks to ensure that the relationships in the data are in line with operational expectations for every TSO;
- iterative consultations with the TSOs to ensure that the submitted data is correct and reported in a consistent fashion.

Even after these steps have been taken, processes must be in place to either correct or remove any irregular observations remaining in the dataset.

It is beyond the scope of this report to pursue any of these process-related solutions relating to data reporting and screening, which should have been considered comprehensively before undertaking any analysis. As part of the shadow study, we had to take the data collated and processed by Sumicsid as given and could only make specific changes for particular companies (and we were therefore unable to tackle any systematic or pervasive errors).

Thus, to account for data errors, we have estimated an appropriate margin of error relative to the TSOs' efficiency scores. While the data errors are extensive, in this report, we suggest the use of a Monte Carlo simulation to account for the data errors at least to some extent. The Monte Carlo simulation adds a random component to all inputs and outputs,[76] and the magnitude of this random component can be informed by our knowledge on the prevalence of data errors. Such simulations have been used, for example, by the Office of Rail and Road (ORR) in the UK in its PR13 benchmarking of Network Rail

[74] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, section 3.2.
[75] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, section 3.06–3.07.
[76] For example, if a TSO reported a TOTEX of €10m then we would draw a number between €9m and €11m from a uniform distribution. We do the same thing for all other outputs and for all TSOs. We then estimate the efficiencies. We repeat this process 1,000 times to arrive at a distribution of estimated efficiency scores.

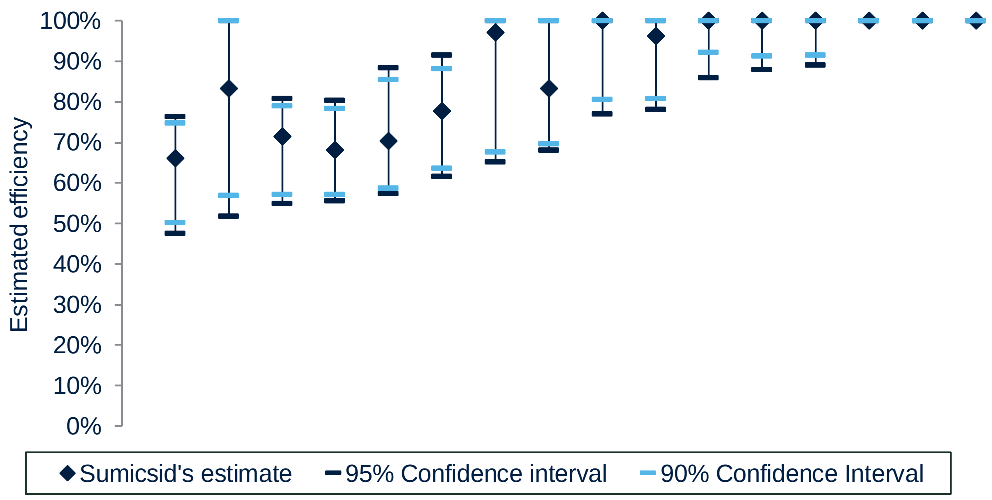
against international rail infrastructure managers, in which a 2.5% random component was applied to the data.[77]

In the present case, we chose to apply a 10% random component. While this is a larger random adjustment than that applied by ORR, it is significantly smaller than many of the errors we found in the data. For example, an error discovered in a TSO's CAPEX overstates its TOTEX by 32%. In another case, the value for weighted lines is overstated by 27%. In this context, a 10% random component is relatively conservative.

The margins of error for the TSOs' efficiency scores based on this method are presented in Figure 3.1.[78] We note the following.

- While Sumicsid estimates seven TSOs as efficient, eleven TSOs have 100% efficiency within the margin of error. That is, given the data uncertainty, we cannot be certain (at the 90% significance level) that these four additional TSOs are *not* operating efficiently.
- Some TSOs are particularly sensitive to the inclusion of random noise. For example, the fifth TSO (reading from left to right in Figure 3.1) has an estimated efficiency score of 70% in Sumicsid's analysis. However, the 90% confidence interval lies in the range of 59–88%.[79]

Figure 3.1 Confidence intervals based on the Monte Carlo simulation



Note: Based on 1,000 draws of a random component of 10% in TOTEX, Adj. NormGrid, transformer power and weighted lines.

Source: Oxera analysis.

If the efficiency scores from the TCB18 study are to be used to inform cost allowances, it is essential that the uncertainty inherent in the data and modelling is taken into account. To that end, a regulator needs to be confident that the proposed savings are feasible.

Note that the Monte Carlo simulations we have considered focus only on random and symmetrical statistical noise. As such, the estimated efficiency

[77] Office of Rail Regulation (2013), 'PR13 Efficiency Benchmarking of Network Rail using LICB', August, p. 55.
[78] We note that while Sumicsid identifies three outliers with the outlier tests it describes in Sumicsid (2019), we are only able to identify two. This leads to the estimated efficiencies in our replication being slightly lower than in Sumicsid's output. We know which TSO was identified as the additional outlier, if we do so too we match the efficiencies exactly.
[79] This interpretation assumes that the model itself is fundamentally correct (e.g. all relevant outputs are considered) and the only 'flaw' in the modelling is a random data error.

scores from TCB18 will lie within the estimated confidence intervals by construction. Other sections of this report address some of the systematic issues that bias the results for some TSOs.

3.2 Defining the input variable

3.2.1 Description of the issue

Sumicsid uses TOTEX as the input in their models, where its measure of TOTEX is constructed by summing OPEX and CAPEX. For this approach to provide the correct incentives and a robust estimate of managerial efficiency, it must be assumed that OPEX and CAPEX are as follows.

- 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.

If the above conditions do not hold, then it may be more appropriate to explicitly account for the trade-off between OPEX and CAPEX in the benchmarking model.

3.2.2 Sumicsid's approach to defining the input variable

Sumicsid uses a TOTEX model to assess the efficiency of TSOs. It states that such a model provides incentives for TSOs to balance OPEX and CAPEX solutions,[80] but the equivalence and controllability conditions are not discussed in its report.

3.2.3 Critique and proposed solution

Sumicsid has made separate adjustments and normalisations to OPEX and CAPEX, as shown in Table 3.1.

Table 3.1 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 annuitized investments from 1973
PLI adjustment	Adjustment for labour prices using the PLI for civil engineering works	No PLI adjustment
Inflation adjustment	N/A[1]	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.

Operationally, it is unlikely that a euro of CAPEX should have the same impact as a euro of OPEX. Moreover, the separate treatment of OPEX and CAPEX in the cost normalisation process casts doubt on the equivalence between the two. A manager could re-allocate €1 from OPEX to CAPEX and the resulting *normalised* TOTEX will be different. Similarly, two TSOs that have equivalent levels of TOTEX may have different level of *normalised* TOTEX due to differences in cost reporting. In this context, it is inappropriate to impose a one-

[80] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, p. 19.

to-one trade-off between OPEX and CAPEX by treating TOTEX as a single input.

Furthermore, the significant variance in CAPEX shares across TSOs may be indicative of fundamentally different operating models or reporting differences. Different operating models could be driven, for example, by maturity of the networks, legislative and regulatory obligations, or operational characteristics. For example, the Latvian TSO, AST, does not own the electricity transmission network assets, as mandated by the Latvian electricity market law.[81] A TSO that does not own the electricity transmission assets it uses will almost certainly be less CAPEX-intensive than TSOs that own these assets, since leasing is treated as OPEX. This illustrates that the CAPEX–OPEX mix may be a matter of regulatory policy and not completely within the control of the TSOs.

The treatment of OPEX and CAPEX is especially concerning in the current context, where the share of CAPEX in TOTEX varies significantly across TSOs. In 2017, CAPEX as a percentage of TOTEX varies from 23% to 90%, with most TSOs having a share of CAPEX within TOTEX significantly above 50%. The CAPEX/TOTEX ratio can be even more extreme in other years. For example, Sumicsid's data suggests that some TSOs had a share of CAPEX within TOTEX of 11% in 2013.

To ensure that a TSO is compared only to TSOs with similar OPEX and CAPEX ratios, one option would be to consider OPEX and CAPEX as separate inputs in the DEA model.[82] This would directly account for the heterogeneity between the cost categories.

Figure 3.2 illustrates the effect of this change. The results show:

- one TSO assessed to be efficient in Sumicsid's model becomes inefficient in the two-input model;
- two TSOs assessed to be inefficient in Sumicsid's model become efficient;
- the seventh TSO (reading from left to right) has one of the highest shares of CAPEX in TOTEX (84%) and is compared to a TSO for whom CAPEX is a mere 25% of TOTEX in Sumicsid's TOTEX model. When the heterogeneity is taken into account in a two-input model, the former TSO becomes 100% efficient.

[81] Augstsprieguma tikls (2017), 'Financial statements 2017', p. 62.
[82] Multi-input models can also be estimated in an SFA context. See Kumbhakar, S.C, Wang, H-J and Horncastle, A. P. (2015), *A Practitioner's Guide to Stochastic Frontier Analysis Using STATA*, Cambridge University Press, chapter 6.

Figure 3.2 Impact of moving to a two-input model

Source: Oxera analysis.

An alternative approach to a two-input model could be to construct TOTEX as a weighted average of OPEX and CAPEX (where the weights are based on expert judgement), recognising that some TSOs are able to make trade-offs, but maintaining flexibility regarding the exact relationship.[83] It is also possible to model activities at a disaggregate level before aggregating to a TOTEX efficiency. This is one of the approaches used by UK regulators.[84] Such models could also serve as a cross-check for models developed on a TOTEX basis as they can capture drivers of specific types of costs more robustly.

3.3 Adjusting for differences in input prices

3.3.1 Description of the issue

Price-level differences can persist even in closely linked economies and for relatively mobile goods.[85, 86] TSOs are likely to choose their input mixes in the way that optimises the impact of price-level differences. For example, a TSO facing high labour costs may choose to invest more in CAPEX, thereby reducing its maintenance costs. However, it is not possible for a TSO to fully mitigate the impact of higher input prices across all inputs. That is, input prices are exogenous (i.e. not within management control).

[83] The weighted average representation of TOTEX would be: TOTEX = w*CAPEX +(1-w)*OPEX = OPEX +w*(CAPEX-OPEX). The weight w can be estimated through a regression of OPEX on CAPEX. For additional flexibility, a squared or cubed term can be included. Typically there is a confidence interval around the weight and it can be introduced in the two-input DEA model in the form of an additional constraint specifying the relationship between the two.
[84] Ofwat (2019), 'PR19 final determinations: Securing cost efficiency technical appendix', December; Ofgem (2019), 'Consultation - RIIO-2 tools for cost assessment', June.
[85] 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.
[86] For studies on the effect of a single currency, see for example Engel, C. and Rogers, J.H. (2004), 'European product market integration after the euro', *Economic Policy*, **19**:39, pp. 348–84; and Eurostat (2019), 'GDP per capita, consumption per capita and price-level indices', https://ec.europa.eu/Eurostat/statistics-explained/index.php?title=GDP_per_capita,_consumption_per_capita_and_price_level_indices#Relative_volumes_of_GDP_per_capita, accessed 26 November 2019.

Controlling for the impact of input prices on expenditure is an important step in normalising costs over different jurisdictions, and is often done in national and international benchmarking studies. [87]

3.3.2 Sumicsid's approach to adjusting for input prices

Sumicsid acknowledges that price-level differences exist for some input factors, namely labour. It further concedes that input prices are exogenous:

> In order to make the operating costs comparable between countries a correction for differences in national salary cost levels has been applied. Otherwise TSOs would be held responsible for cost effects, e.g. high wage level, which is not controllable by them.[88]

In particular, Sumicsid applies a correction to direct manpower costs within OPEX using the price-level index (PLI) for civil engineering from Eurostat. The index 'includes construction not classified under buildings, for example railways and bridges'.[89, 90]

Price-level differences for other inputs are not discussed in the outputs.

3.3.3 Critique and proposed solution

Sumicsid correctly recognises the need to correct for price-level differences across TSOs. However, we have identified two broad issues with Sumicsid's approach.

- Adjusting labour costs by the PLI for civil engineering does not take into account the fact that labour costs may vary by more than overall civil engineering price levels, which includes factors of production that may be more mobile across borders within the European Economic Area (EEA) such as raw materials (e.g. construction materials such as metals, plastics and concrete).
- Sumicsid does not apply a correction to any cost item other than direct manpower cost. This means that Sumicsid only accounts for input price differentials for a very small proportion of the cost base (approximately 5.9% of TOTEX, on average across TSOs) without sufficient justification. As such, Sumicsid assumes there are no input price differences across the participating TSOs for all the other components of OPEX, such as purchase of external maintenance, personnel leasing, consultancies, office supplies and control centre costs, as well as all components of CAPEX including setup costs.

The adjustment made by Sumicsid is therefore insufficient as 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 does not provide sufficient evidence to validate its hypotheses, nor does it consider that:

[87] For example, see Office for Rail and Road (2013), 'PR13 Efficiency Benchmarking of Network Rail using LICB', August, pp.12–14, August.
[88] Sumicsid (2019), Pan-European cost-efficiency benchmark for electricity transmission system operators main report, section 4.46, p. 20, July
[89] Eurostat (2018), Glossary: Civil engineering work, available at: https://ec.europa.eu/Eurostat/statistics-explained/index.php/Glossary:Civil_engineering_work, accessed 8 December 2019.
[90] Eurostat (2019), Purchasing power parities (PPPs), price-level indices and real expenditures for ESA 2010 aggregates, available at: https://appsso.Eurostat.ec.europa.eu/nui/show.do?dataset=prc_ppp_ind&lang=en, accessed 29 November 2019

- a significant proportion of CAPEX is labour or labour-related costs (e.g. installation costs) and not covered by the adjustment to gross labour costs in OPEX;
- differences in the price of more mobile factors of production, such as raw materials, do exist across the EEA due to, for example, transportation costs, as evident in differences in the investment goods, total goods and other price indices;[91]
- regulations such as local procurement rules sometimes restrict the scope for arbitrage, and therefore limit the ability of the market to homogenise input prices across jurisdictions;
- much of the (44-year) CAPEX investment stream used to calculate TOTEX was incurred decades ago, before the close integration of the EU and EEA was complete. Intuitively, the TSOs operating in the former Soviet sphere of influence were likely not part of the same market as western European TSOs prior to 1990, due to stringent trade restrictions.[92] It is therefore unlikely that they faced the same input prices.

The approach chosen by Sumicsid thus does not normalise the cost sufficiently and risks conflating the uncontrollable price-level differences faced by companies with managerial inefficiency.

In national and international benchmarking exercises, it is common to adjust all expenditure to account for regional and international differences in prices. In the UK, for example, the Office of Rail and Road (ORR) adjusted all cost data (100% of OPEX and CAPEX) using the PLI for GDP adjustment in its international benchmarking of Network Rail's efficiency for its 'PR13' price review.[93] Differences in input prices are sometimes considered in national benchmarking exercises.[94]

Although it is clear that input price adjustments are required across a material proportion of the cost base, if not the entire cost base, the precise method of correcting for price differences requires careful consideration. In particular, one needs to consider:

- the basket of goods represented by the PLI (e.g. consumer prices, construction prices);
- the base year of the PLI (e.g. as the published values of price levels can vary across years, should we express prices in price levels for 2017, 2016 or an average over the modelled period?);
- the percentage of the cost base subject to the PLI adjustment (e.g. are there cost components whose prices do not vary across countries or TSOs?);

[91] Eurostat (2019), Purchasing power parities (PPPs), price-level indices and real expenditures for ESA 2010 aggregates, available at: https://appsso.Eurostat.ec.europa.eu/nui/show.do?dataset=prc_ppp_ind&lang=en, accessed 29 November 2019.
[92] Trade within the Soviet sphere of influence was conducted under a different system of exchange. See Broadman, H. G. (ed.). (2006), *From disintegration to reintegration: Eastern Europe and the former Soviet Union in international trade*, The World Bank, Box 1.1, p. 52.
[93] Office for Rail and Road (2013), 'PR13 Efficiency Benchmarking of Network Rail using LICB', August, pp.12–14, August; (2008), 'Periodic review 2008 Determination of Network Rail's outputs and funding for 2009-14', October, p. 122. For the latest benchmarking exercise, which is part of PR18, the ORR did not perform an international benchmarking exercise.
[94] For example, in its RIIO-ED1 price control, the Office of Gas and Electricity Markets (Ofgem) applied a correction for regional labour costs within the UK to its cost base. See Ofgem (2014), 'RIIO-ED1 final determinations for the slow-track electricity distribution companies Business plan expenditure assessment', 28 November, p. 41.

- relatedly, the type of cost that is subject to the PLI adjustment (e.g. OPEX or all costs?);

The estimated efficiency scores for individual TSOs are likely to be highly sensitive to these choices, illustrating the need for sensitivities around the choices. We have estimated the efficiency score of each TSO under the following price-level sensitivities.

- **Choice of index.** We considered indices relating to overall GDP, construction, civil engineering works, total goods, capital goods and total services to adjust expenditure.
- **Time of adjustment.** We considered adjusting expenditure using the PLI for 2017, an average based on the last five years of data, an average based on the last 10 years of data, and an average based on all of the PLI data that was available.
- **Proportion of adjustment.** Sumicsid makes no adjustment for CAPEX. However, as discussed above, a 100% adjustment has regulatory precedent. We considered adjusting 60%, 80% and 100% of CAPEX for difference in price levels. For OPEX, Sumicsid adjusts labour costs, which is approximately 5.9% of TOTEX; however, as discussed above, a 100% adjustment has regulatory precedent. As such, we considered adjusting 60%, 80% and 100% of OPEX for difference in price levels.

As a central scenario, we propose to adjust 100% of OPEX with the PLI for overall GDP and 100% of CAPEX with the PLI for civil engineering. Furthermore, we have chosen the five-year average of 2013–17 as the base year for the PLI adjustment. These are justified below.

- As OPEX contains a wide array of goods and services bought by TSOs (e.g. grid maintenance, grid planning and business support activities), the PLI for overall GDP is a reasonable compromise between the higher price-level differences in services and the lower price-level differences in goods.
- For CAPEX, the basket of goods is different. This cost category contains mainly investment projects and thus fewer (but still a considerable amount of) services. The ESA category that matches these expenses the closest is ‘civil engineering works’. This is defined by Eurostat as ‘a construction not classified under buildings, for example railways, roads, bridges, highways, airport runways and dams’.[95] An alternative would be to use the ‘construction’ category, but this will contain residential construction as well, which is likely less relevant for the specialised works required by TSOs.[96]
- Taking an average over the time period is a compromise that takes into account the volatility of PLIs and the fact that contracts and investments are usually multi-year commitments whose price is not set in the year they are recorded.

the Office of Gas and Electricity Markets (Ofgem) applied a correction for regional labour costs within the UK to its cost base. See Ofgem (2014), ‘RIIO-ED1 final determinations for the slow-track electricity distribution companies Business plan expenditure assessment’, 28 November, p. 41.

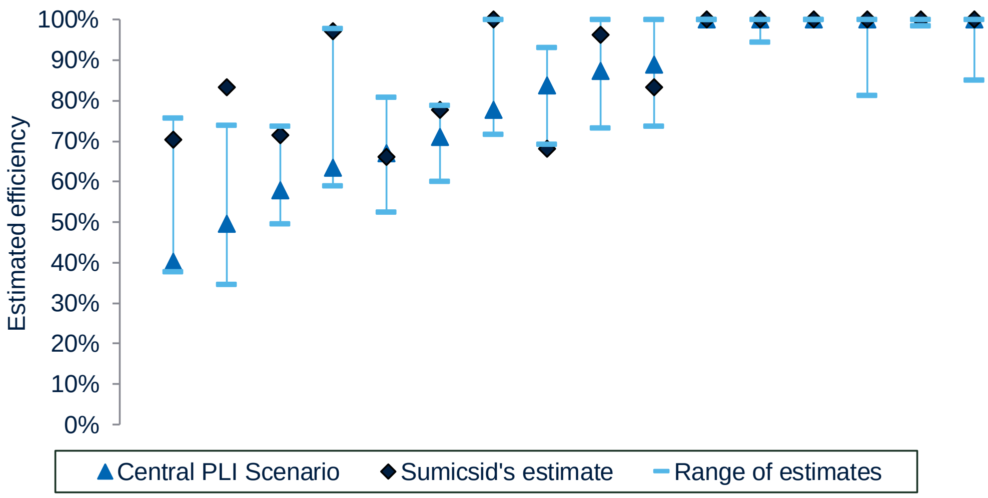
[95] Eurostat (2018), ‘Glossary: Civil engineering work’, available at: https://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:Civil_engineering_work, accessed 5 February 2020.

[96] The range and standard deviation of the PLI for construction is roughly 20% wider than the corresponding values for civil engineering.

Figure 3.3 shows the estimated efficiency scores for each TSO in the sensitivities we have considered. Adjusting the method of accounting for input prices has a significant impact on most TSOs' efficiency scores. For example:

- Two TSOs that are assessed to be inefficient by Sumicsid become efficient in at least one PLI adjustment.
- Furthermore, five of the seven TSOs estimated to be efficient by Sumicsid become inefficient in at least one of the sensitivities. Indeed, one TSO that is estimated to be 100% efficient in Sumicsid's analysis is estimated to be 78% efficient in our central PLI scenario. Thus, Sumicsid's peers may well be not on the best-practice frontier.

Figure 3.3 Impact of PLI adjustments



Source: Oxera analysis.

Clearly, the method of accounting for price-level differences is not a trivial decision and must be robustly justified. In addition, some sensitivity analysis should be undertaken. Sumicsid has not sufficiently motivated its choices in its final outputs and has therefore potentially conflated the estimated efficiency scores with price-level differences. Our analysis demonstrates that the impact of the price-level adjustments can be material for many of the TSOs.

3.4 Indirect cost allocation mechanism

3.4.1 Description of the issue

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.

As many TSOs perform activities that are beyond the scope of the TCB18 benchmarking project, some of the costs incurred in 'Indirect Support' will be driven by activities that are not assessed. For example, a TSO that undertakes a significant amount of system operations and market facilitation (activities that are outside of the scope of benchmarking) may have a larger expenditure on IT support than other TSOs that do not undertake such activities.

Thus, to avoid a TSO's efficiency score being driven by activities deemed outside the scope of the study, indirect costs need to be allocated to activities within the scope of benchmarking. Ideally, the allocation rule should consider how much of the indirect costs are driven by each activity (i.e. where the indirect expenditure is incurred).

3.4.2 Sumicsid's approach to allocating indirect expenditure

Sumicsid's allocation rule uses the proportion of all costs except depreciation and energy relative to these costs across activities.[97]

$$Indirect_{TPM} = \frac{OPEX_{TPM} - Depreciation_{TPM} - Energy_{TPM}}{OPEX_{total} - Depreciation_{total} - Energy_{total}} * Indirect_{total}$$

Where the subscripts TPM and $total$ refer to in-scope (i.e. transport, planning and maintenance) and TOTEX, respectively. Note that this allocation rule includes cost items such as taxes and research and development, which are not part of the efficiency benchmarking.

Sumicsid states that it has tested 'several allocation methods' for indirect expenditure.[98] However, sensitivities regarding the allocation rule were not presented in any of its reports.

3.4.3 Critique and proposed solution

Sumicsid's allocation rule is never justified in its main report. In some cases, the allocation of indirect expenditure is driven by large cost items that are unrelated to where indirect costs are incurred. For example, one TSO has a large tax payment in an activity that is considered to be out-of-scope. Because of this, a large percentage of indirect expenditure is allocated to this out-of-scope activity, even though the tax payment itself should not be a material driver of indirect expenditure.

As a solution to this issue, we consider that it would be more robust to only use in-scope costs to allocate indirect OPEX. This would avoid the risk that indirect costs are allocated to activities based on cost items that are unrelated to indirect expenditure. As a sensitivity, we have also considered the possibility of excluding or including all indirect costs in the benchmarking. The assumptions behind such allocations are explained below.

- Excluding all indirect costs—this assumes that either (i) there is no trade-off between indirect costs and other expenditure items such that they can be excluded without tainting the analysis of other areas of expenditure; (ii) indirect costs are efficiently incurred by TSOs; or (iii) indirect costs are incurred at the same level of efficiency as other expenditure items.
- Including all indirect costs—this assumes that indirect costs are a 'fixed cost' that do not vary significantly across TSOs that perform different activities. For example, a TSO that only undertakes in-scope activities may spend the same amount on central management (e.g. CEO compensation) as a TSO that undertakes more activities. This is equivalent to assuming there are significant economies of scope with respect to indirect expenditure.

We are aware that these sensitivities suffer from limitations and do not consider the controllability of indirect expenditure or the trade-offs across the activities that TSOs undertake. These are intended to serve as a reference

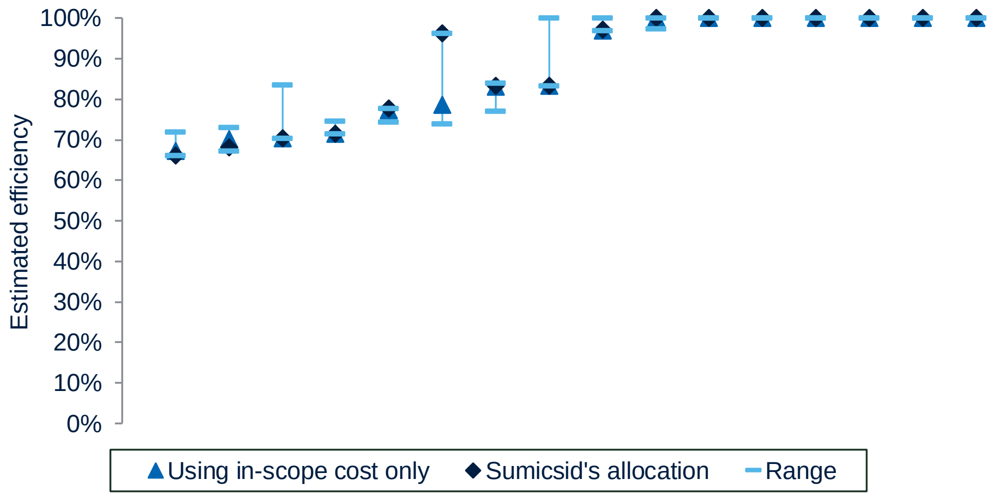
[97] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, section 4.97.
[98] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, section 4.97.

points as to the maximum possible impact of sensitivities to indirect costs on overall efficiency.

Figure 3.4 shows the impact of different allocation rules on TSOs' estimated efficiencies. We note that:

- most TSOs' estimated efficiency scores are insensitive to the allocation of indirect expenditure;
- one TSO estimated to be efficient in the TCB18 study becomes inefficient under one allocation rule;
- the allocation rule has a significant impact for one TSO, the eighth from the left, where the inclusion of a large levy results in only 0.8% of indirect costs being allocated to benchmarked scope activities, despite in-scope activities constituting 25% of benchmarked OPEX.

Figure 3.4 Impact of different indirect cost allocation rules



Source: Oxera analysis.

In TCB18, the allocation of indirect expenditure has a small impact on the efficiency of most TSOs. However, the allocation of expenditure to benchmarked activities is a conceptually important issue and may have a more material impact on a different sample. It is therefore essential that the allocation rules are robustly justified.

4 TCB18 model development

In this section, we critically review Sumicsid's approach to model development. As Sumicsid did not share modelling codes with the TSOs, our review is limited to what is described in the final outputs. We note the following.

- The cost driver analysis that Sumicsid states it has followed suffers from empirical and theoretical flaws (section 4.1).
- Sumicsid does not appear to have tested whether its final model is sensitive to changes in the sample (section 4.2).
- Sumicsid's decision to restrict the set of outputs to asset-based data is inconsistent with academic and regulatory literature and is not justified by empirical evidence or conceptual reasoning (section 4.3).
- The analysis used to derive OPEX and CAPEX weights in the construction of NormGrid is not presented in the final outputs. Furthermore, our replication of Sumicsid's stated approach of deriving NormGrid weights does not support the weights Sumicsid used (section 4.4).
- Sumicsid's adjustment of NormGrid for environmental factors is not justified in its final outputs and is not supported by empirical evidence (section 4.5).

4.1 Model development—cost driver analysis

4.1.1 Description of the issue

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 both 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.

4.1.2 Sumicsid's approach to cost driver analysis

Sumicsid used three statistical estimators to examine the relationship between costs and outputs:

- Ordinary Least Squares (OLS) regression;
- OLS regression excluding outliers as defined by the cooks distance metric;
- Robust OLS (ROLS) regression (an estimator where observations far from the regression line are given less weight).[99]

In its workshop slides, Sumicsid states that it uses Lasso regression[100] to justify its use of NormGrid and detect and validate alternative output variables, but this analysis is not presented in the final outputs.

[99] Sumicsid (2019), 'CEER-TCB18 project Model Specification ELEC V1.3', February, slide 31.
[100] 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). A larger penalty term leads to more cost drivers being excluded from the model and the size of the penalty term is within the control of the practitioner. If the penalty term is set to zero, the Lasso regression is equivalent to OLS and no cost drivers are excluded from the model.

4.1.3 Critique and proposed solution

It is not clear from the TCB18 report exactly which cost drivers were considered in the project and how the final set of cost drivers were derived. There is insufficient detail in the report for us to robustly replicate and review the model-development process. However, given (i) the results published in final report (and other associated outputs); (ii) the unjustified restriction of the model to three outputs; (iii) the inappropriate functional form assumed in the modelling; and (iv) the limited consultation with the TSO group on these; we consider that Sumicsid's model-development process was flawed.

Alternative models are presented in the workshop slides, but the slides do not contain the final model used in the main report, so it is unclear how this final model was selected. Furthermore, the statistical assessment of the final model is limited to two tables in the final outputs: one table in the main report shows the adjusted R-squared for the model (and how the adjusted R-squared changes as outputs are added sequentially) and one table in the TSO-specific report shows the estimated coefficients in the final model.

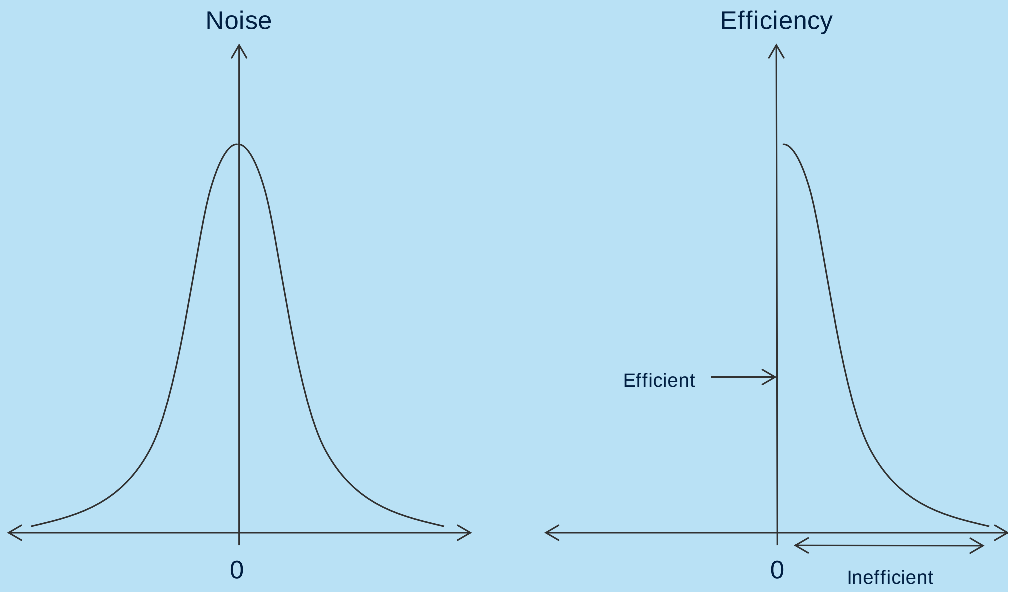
In general, the adjusted R-squared can provide useful information as to how well variations in cost drivers can explain variations in average costs. However, the adjusted R-squared is not the only parameter of interest in cost driver analysis—it is also essential that the estimated relationships between costs and cost drivers are in line with operational intuition and statistically significant. Furthermore, the value of adjusted R-squared in a regression of costs against NormGrid is not informative. NormGrid is a representation of the average costs required to build and operate a particular asset base and will therefore be highly correlated with observed costs by construction.

The use of ROLS

Sumicsid's use of ROLS is inappropriate, especially when only the results from this estimation technique are presented. This estimator explicitly gives less to observations that are further from the regression line and the adjusted R^2 is therefore inflated. More fundamentally, both the OLS and ROLS used by Sumicsid for cost driver analysis assume that the error term is symmetrical and normally distributed. However, if Sumicsid expects that some TSOs are operating inefficiently, the error term will be skewed, as shown in Box 4.1. Any statistical inference that Sumicsid may have performed in selecting cost drivers is therefore inconclusive.

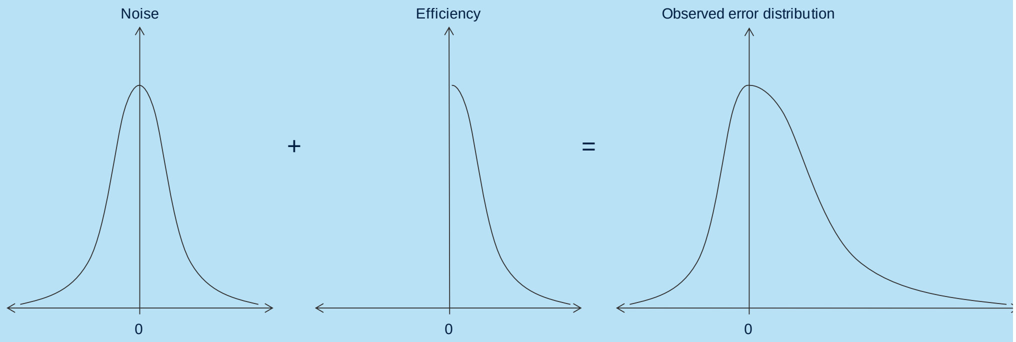
Box 4.1 Statistical inference in the presence of inefficiency

If inefficiency is present in the sample, the residual term of an OLS regression represents two effects. The first effect is pure, normally distributed statistical noise. Statistical noise can have a positive or negative impact on TOTEX, but the mean and median residual will equal zero. The second effect is the inefficiency effect. A TSO's inefficiency will equal zero if it is fully efficient and will be greater than zero if it is inefficient. By definition, a firm cannot be more than fully efficient and the distribution of inefficiency in therefore one-sided, reflecting the higher costs of an inefficient firm relative to a fully efficient firm. A comparison of the distribution of the statistical noise and inefficiency components of the error term is shown in the figure below.



Source: Oxera.

The total error term therefore represents a combination (i.e. sum) of these two components and will be asymmetric, as shown in the figure below.



Source: Oxera.

Our replication of Sumicsid's analysis

We have been unable to exactly reconstruct all of the data used by Sumicsid. In particular, the historical (i.e. pre-2017) data we have received for weighted lines does not match that used by Sumicsid.[101] As such, any replication of

[101] Sumicsid's data here implies an unreasonable growth in this variable of 9.2% p.a. whereas we arrive at a more realistic 3.3% p.a..

analysis that uses the full dataset is expected to be different to what Sumicsid presents in its final outputs.

Table 4.1 shows our replication of Sumicsid's full model. Given the uncertainties in the data and ambiguities in the report, we have conducted a number of sensitivities regarding the sample size and time period of analysis (discussed below). However, we are unable to validate the analysis presented in Sumicsid's final outputs.

- **Full sample.** Here, we estimate the econometric model using all 81 observations,[102] as Sumicsid stated it used in the individual reports.[103] The estimated coefficient on NormGrid is negative and statistically insignificant, which is operationally unintuitive.
- **2017 only.** As we have identified an inconsistency between the dataset we have access to and that which Sumicsid states it has used for years prior to 2017, we have estimated the model using data from 2017 only, where the two datasets are consistent.[104] The coefficients in this model are aligned with our replication on the full sample, i.e. they remain unintuitive.
- **One TSO removed.** Sumicsid states in its main report that one TSO is 'permanently removed from the reference set' as it is almost always an extreme outlier.[105] We therefore estimate the model with this TSO removed from the sample. The coefficient on NormGrid becomes positive and statistically significant and is more consistent with the results published in individual reports. However, the coefficient on weighted lines becomes negative and insignificant, which is inconsistent with operational intuition.

Table 4.1 ROLS regression results

	Sumicsid	Full sample	2017 only	One TSO removed	One TSO removed, 2017 only
Adj. NormGrid	0.302***	-0.1	-0.13	0.46***	-0.02
Transformer Power	4196***	3839***	4776***	2839***	4502***
Weighted Lines	16770***	66920***	57482***	-2605	49673***
Adj. R-squared	0.98	0.99	0.99	0.99	0.99
Observations	81	81	17	76	16

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

Our inability to exactly replicate Sumicsid's analysis points to several potential flaws in both the TCB18 procedure and the final model itself, including:

- the final report is not sufficiently detailed for third parties to follow, replicate and critically assess the analysis undertaken by Sumicsid;
- Sumicsid's non-publication of modelling codes from the final set of outputs compounds the issue highlighted above;

[102] Note that while there are 17 TSOs in the 2017 sample, the sample for 2013–2016 contains only 16 TSOs, hence the full sample is an unbalanced panel of 81 observations.
[103] See Sumicsid (2019), 'Project TCB18 Individual Benchmarking Report Fingrid – 131', July, Table 3.1.
[104] We have matched the 2017 used in our analysis with the data presented in the individual reports. Furthermore, we are able to replicate Sumicsid's DEA model results that are estimated using the 2017 data only. For this reason, we are confident that we have access to the complete data for the year 2017.
[105] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, p. 30.

- the model is sensitive to changes in the historical data (specifically, the data for weighted lines).
- the chosen model is highly sensitive to the inclusion or exclusion of particular TSOs, indicating that the model is not a good predictor of industry-wide costs;
- the model is highly sensitive to the inclusion or exclusion of particular years, indicating that the relationship between cost and cost drivers may be different in different time periods.

The latter two points are discussed in more detail in section 4.2.

The choice of functional form

Sumicsid made three choices regarding the functional form of its cost diver analysis that are not robustly validated:

- the intercept is set to zero, instead of being estimated by the model;
- the model is estimated in levels rather than logarithms (logs);
- there is no test for the presence of more flexible relationships in the data (e.g. quadratic relationships or interactions).

A satisfying analysis regarding the points on functional form may include the following:

- **Visual inspection**—this can, for example, include scatterplots of costs against cost drivers to form an initial understanding of the heterogeneity and possible non-linear form of the relationship;
- **Statistical analysis**—this initial understanding can then be confirmed, supplemented or refuted by specialized statistical tests.

Supressing the intercept

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.

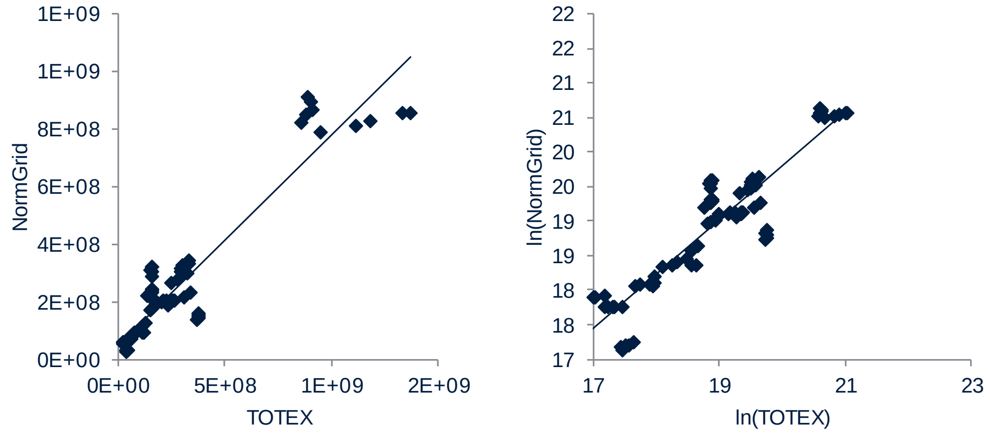
Modelling in levels

The choice to model in levels rather than transforming the data with logarithms is not supported with visual or statistical evidence in the final outputs. A justification of this decision is required as it represents a departure from the common practice in regulatory benchmarking to model in logarithms.[106] Modelling in logs generally alleviates heteroskedasticity[107] and makes coefficients more easily interpretable.[108]

[106] For example Ofwat, Ofgem and the Bundesnetzagentur model in logs. For example. see Ofwat(2019), ‘PR19 final determinations: Securing cost efficiency technical appendix‘, December.
[107] Heteroskedasticity occurs when the standard error of the residual is correlated with the cost drivers in the model. If heteroskedasticity is present, then the standard errors on the estimated coefficients are biased. In such cases, statistical inference is not possible.
[108] When modelling in logs, we can say that an additional 1% increase in outputs would be associated with a b% increase in inputs. By contrast, this interpretation is scale-dependent in levels.

An example of heteroskedasticity issues in TCB18 can be seen in Figure 4.1. A visual inspection of TOTEX against NormGrid reveals that two TSOs are significantly larger than the rest of the sample and appear to be further from the regression line. This may lead to biased results, as the two large TSOs have a large influence on the shape of the regression line. It is evident from the figure that this issue is less pronounced if the data is transformed using logarithms.

Figure 4.1 NormGrid–TOTEX comparison



Source: Oxera analysis.

The results in Table 4.2 show that statistical evidence also does not support the choice to model in levels. A statistical PE-test of the logarithmic model against the model in levels is inconclusive.[109] In fact the test indicates that neither form is sufficient. In light of this, a prudent approach would be to model in logarithms due to the concerns regarding heteroskedasticity and precedent raised above.

Table 4.2 p-values of the PE test

	Intercept set to zero (Sumicsid's approach)	Intercept not set to zero
Linear Model	0.003	0.006
Logarithmic Model	1.332e-11	3.728e-09

Source: Oxera analysis.

Table 4.3 shows Sumicsid's model when it is re-estimated on log-transformed data. Several key relationships between costs and cost drivers have changed. In logarithms, the coefficient on NormGrid is positive and statistically significant. However, the coefficient on weighted lines becomes negative and statistically insignificant. This suggests that if Sumicsid had modelled in logs, an alternative set of cost drivers may have been selected.

[109] The PE-test considers whether the difference in fitted values between the models is an explanatory factor. It is proposed in: J. MacKinnon, H. White, R. Davidson (1983). Tests for Model Specification in the Presence of Alternative Hypotheses: Some Further Results. *Journal of Econometrics*, **21**, pp. 53–70.

Table 4.3 Regression results in logs

	Intercept set to zero		Intercept not set to zero	
	Levels	Logs	Levels	Logs
Intercept			-7557411.36	3.12**
Adj. NormGrid	-0.1	0.74***	-0.05	0.50***
Transformer Power	3838.68***	0.47***	3871.44***	0.63***
Weighted lines	66920.36***	-0.01*	62399.85***	-0.01*
Adj. R-squared	0.98	0.99	0.97	0.88
Observations	81	81	81	81

Source: Oxera analysis.

Mis-specification

An initial visual inspection of the scatterplots (Figure 4.1) does not reveal a clear nonlinear relationship. However, as there are many possible relationships and interactions to consider, visual inspection is insufficient to be conclusive. The RESET test is a statistical test designed to detect the presence of non-linear relationships variables.[110] When performed on Sumicsid's model, the RESET test does not indicate significant mis-specification when the model is estimated in levels. However, there is evidence of mis-specification if the model is estimated in logarithms. Where mis-specification is detected, a more flexible functional form—for example, one that includes some interactions and squared terms—may be needed. This is essential in the selection of cost drivers; cost drivers that may fit the data poorly in a linear model may explain the data well in more flexible models and would have thus been chosen as cost drivers for the benchmarking model if cost driver analysis was performed in a more appropriate functional form.

Table 4.4 p-values of the RESET test

	Intercept set to zero (Sumicsid's approach)	Intercept not set to zero
Linear model	0.36	0.47
Logarithmic model	0.04	0.04

Source: Oxera analysis.

Overall, there is no conclusive evidence on the functional form. Therefore, there is also no evidence to support Sumicsid's departure from common practice. The choice of functional form impacts cost driver analysis and may thus change the composition of the benchmarking model.

Further to our concerns regarding Sumicsid's choice of estimator and functional form, a robust procedure should also include the following.

- **Sensitivity analysis.** The final model should include outputs that can explain differences in industry-wide expenditure. As such, it should not be

[110] The RESET test considers if powers of the residuals can explain the dependent variable. It is proposed in: Ramsey. J. B. (1969), 'Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis'. *Journal of the Royal Statistical Society, Series B* **31**, pp. 350–71.

sensitive to the inclusion or exclusion of specific TSOs or specific time periods of analysis.

- **Valid estimation approaches.** The cost driver analysis should include alternative estimation procedures that can explicitly account for the skewness of the residual terms, such as SFA.
- **Transparency.** At the very least, the process of model selection should be well-documented in the final outputs. If modelling codes are also shared, this could allow third parties to follow exactly what has been done and identify any errors or inconsistencies in the modelling process.[111]

4.2 Model development—sensitivity to the sample selected

4.2.1 Description of the issue

The final model should include outputs that can explain differences in industry-wide expenditure. As such, it should not be sensitive to the inclusion or exclusion of specific TSOs or specific time periods of analysis.

4.2.2 Sumicsid's approach

Sumicsid does not present any analysis in its final outputs that demonstrates that its model is insensitive to the data used in the modelling. However, it removes one TSO from the sample before estimating efficiency scores because:

> 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.[112]

It is unclear how exactly this TSO was identified as an 'extreme outlier'.

4.2.3 Critique and proposed solution

Based on the fact that one TSO is permanently removed from the reference set, it seems that Sumicsid conducted some analysis regarding the influence of single TSOs on the estimated model. Sumicsid states that this TSO was an 'extreme outlier' in 'a series of alternative model specifications'.

However, it is unclear what tests were actually performed and how exactly this one TSO was identified as an outlier. It is not clear if this TSO was removed on the basis of econometric or DEA results.

For our assessment of individual TSO's influence on the econometric model, we exclude one TSO from the dataset at a time and estimate the econometric model on the reduced dataset, containing the observations for the remaining 16 TSOs. Thus we obtain 17 estimates for each parameter. This provides us with the range of estimates that is supported by the reduced datasets given in Table 4.5. If the econometric model was robust, we would expect the relationship between costs and cost drivers to be stable. In the current case the range of supported coefficients is wide and includes zero for two of Sumicsid's cost drivers. This means that on average, an increase in the cost driver is not associated with an increase in costs. Thus the relationship

[111] For example, Ofwat made an error in its modelling of a specific wastewater programme in its draft determination. Upon release of the Excel analysis, the error was spotted by water companies and their consultants. The error was then corrected for the final determination. See Ofwat (2019), 'PR19 final determinations: Securing cost efficiency technical appendix', December, p. 97.
[112] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, p. 35.

between the costs and cost drivers in Sumicsid's model is highly dependent on the exact sample of TSOs used and unreliable.

Table 4.5 Sensitivity of 'robust regression' results to exclusion of one TSO

	Minimum	Median	Maximum
Intercept	-13713113	-9717442	2641704
Adj. NormGrid	-0.14	-0.09	0.54
Transformer Power	2629.86	4103.23	4535.93
Weighted Lines	-3498.12	64817.61	71005.72
Observations	76[1]	76[1]	76[1]

Note: [1] For one TSO only one year of data is available; if that TSO is removed, we would have 80 observations instead of 76.

Source: Oxera analysis.

The impact of a single TSO on DEA efficiencies is assessed through the dominance test. We provide our assessment of Sumicsid's dominance test in section 5.2.

2017 was likely chosen as the base year for the analysis as it was the latest year with data available when the TCB18 process began. However, all the other years of data in the TCB18 dataset are similarly valid for benchmarking. Sumicsid should have tested if the results hold in different base years as well.

We can assess the impact of the base year by estimating the econometric model on one year of data at a time. This way we can verify if the proposed relationship between costs and cost drivers is in fact stable over time. As can be seen in Table 4.6 below, the relationship is more stable over time than it is to the exclusion of some TSOs. The coefficient on NormGrid switches from positive to negative when using the data from 2013, and the coefficient on weighted lines is statistically insignificant.

Table 4.6 Sensitivity of 'robust regression' results to year of data used

	2013 only	2014 only	2015 only	2016 only	2017 only
Intercept	-1893375	-9836324	-11567559	-8315200	-5676202
Adj. NormGrid	0.3	-0.08	-0.1	-0.09	-0.11
Transformer Power	2580***	3628***	3858***	4486***	4748***
Weighted lines	16246	70372***	70865***	60619***	56749***
Observations	16	16	16	16	17

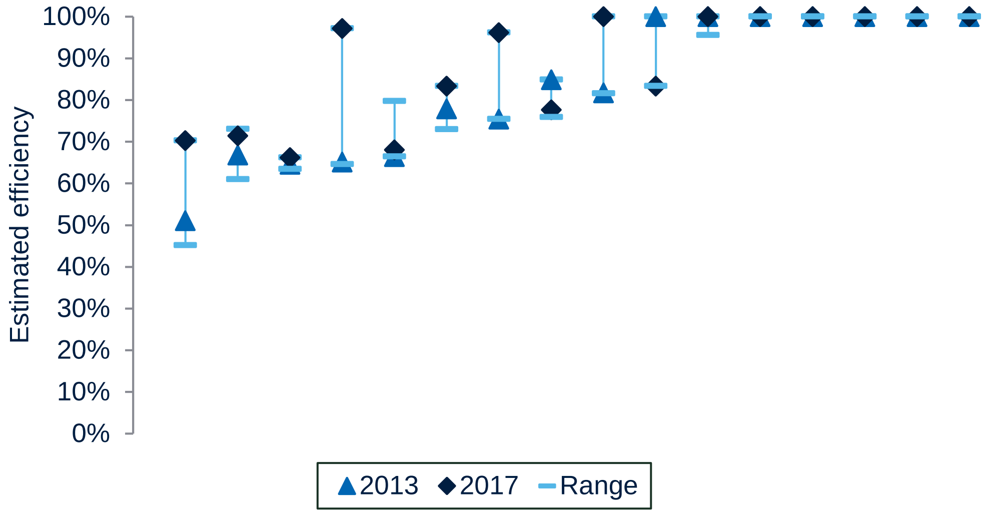
Source: Oxera analysis.

Despite the coefficients being relatively consistent over time, the DEA results for some TSOs are still highly sensitive to the base year of the analysis, as is evident from Figure 4.2. In particular:

- one TSO is estimated to be fully efficient from 2013–15, but is then estimated to be inefficient in 2016–17;
- one TSO that is a peer in Sumicsid's analysis is estimated to be 82% efficient prior to 2016;

- several more TSOs are significantly more efficient in 2017 than in earlier years. One TSO improves its score by 31 percentage points from 2013–17.

Figure 4.2 Sensitivity of efficiencies to base year



Source: Oxera analysis.

Such large changes in efficiency should not occur if the model is robust. However, they may occur because:

- relevant cost drivers may be missing, e.g. declining asset health may lead to higher expenditure while asset characteristics (and thus Sumicsid's cost drivers) remain unchanged;
- the costs do not react to changes in Sumicsid's cost drivers;
- there may be data issues, e.g. the expenditures and assets are not recorded in the same year;
- Sumicsid's cost normalisation may have understated input price inflation, thus some large recent investment projects may have been inefficient, according to Sumicsid's benchmarking model.

It is especially important to robustly justify the occurrence of large swings and how this swing in estimated efficiency also coincided with a change in the management efficiency of the TSOs. This was not considered by Sumicsid, and the causes of the swings remain unexplained.

4.3 Model development—selecting candidate cost drivers

4.3.1 Description of the Issue

In a regulatory setting, it is considered best practice for a regulator to define a set of cost activities, suggest variables to capture each activity and invite participating companies to critique the variables and to suggest alternatives. For example, in advance of its determinations, Ofwat consulted the water industry on the use of econometric models to assess expenditure.[113] In response to feedback from the industry, it amended its modelling approach in later stages of the price review.[114] As part of its upcoming RIIO-2 price control the Office of Gas and Electricity Markets (Ofgem) has asked companies to

[113] Ofwat (2018), 'Cost assessment for PR19: a consultation on econometric cost modelling', March.
[114] Ofwat (2019), 'Supplementary technical appendix: Econometric approach', January.

express their views on the appropriate outputs for each category.[115] This process of iterative consultations limits the risk that significant cost drivers are insufficiently considered. It also gives regulators an opportunity to respond to critique from the industry and, where appropriate, justify their choices with empirical evidence.

In consultation with the TSOs participating in this shadow benchmark, we determined that key drivers of expenditure in the electricity transmission industry include energy transmitted, network length, peak demand, load density, energy not supplied, network availability, connected volumes, variability of energy flows, asset health, and the amount of power supplied by renewable sources.

4.3.2 Sumicsid's approach to selecting candidate cost drivers

Sumicsid does not consider the output parameters mentioned above and has instead restricted itself to using asset-based variables as cost drivers.

Sumicsid discusses outputs such as energy delivered and peak load in its main report,[116] but it does not present analysis to justify excluding these outputs in the model-development process.

4.3.3 Critique and proposed solution

One issue with an asset-based model is that it can create perverse incentives—for example, it could lead to TSOs 'gaming' the benchmarking model by, for example, installing unnecessary assets. Sumicsid acknowledges this in its main report,[117] and we agree with Sumicsid's view that this specific type of gaming seems unlikely in the current context.

However, the use of asset measures such as NormGrid in benchmarking models may still be problematic. For example, a TSO facing increasingly volatile energy supply due to increased generation from renewable sources of energy may consider two solutions with similar levels of investment: 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. Put more broadly, a company that deploys a more expensive asset base than another—when both face the same contextual and demand levels—would appear more efficient, when in fact it is less efficient, even if both have exactly the same TOTEX level. Thus, while assets can be used in comparing the relative efficiency of TSOs, it is necessary to cross-check these models with alternatives that control for TSOs' outputs.

NormGrid is a particularly unusual driver, even excluding the general arguments relating to the use of asset-based outputs outlined above. NormGrid is itself a measure of average costs, so the interpretation of a regression output of TOTEX (costs) against NormGrid (a measure of costs) is unclear.

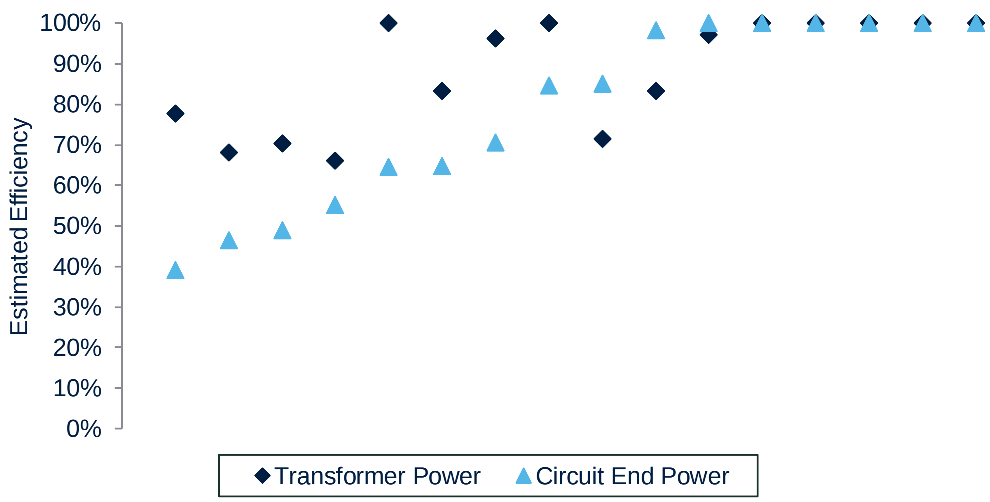
As for the other outputs, Sumicsid states that transformer power and weighted lines are highly correlated with output factors such as capacity and routing

[115] I.e. Ofgem (2019), 'RIIO-2 tools for cost assessment', June
[116] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, section 4.81.
[117] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, section 4.81.

complexity, respectively.[118] These statements should be substantiated with statistical evidence or operational evidence in the final report or its appendices; however, they are not. It is concerning that using close substitutes for these outputs can significantly change the estimated efficiency of some TSOs. For example, Figure 4.3 below shows TSOs' estimated efficiencies when transformer power is replaced with the power of circuit ends. We note that:

- two TSOs that are identified as efficient in Sumicsid's model become inefficient with the alternative capacity cost driver;
- one TSO's efficiency score reduces by 39 percentage points with the alternative measure;
- one TSO increases its efficiency by 15 percentage points.

Figure 4.3 Efficiencies using the power of circuit ends



Source: Oxera analysis.

Developing a cost function (i.e. costs as a function of outputs and input prices), for benchmarking the TSOs, as consistent with the academic literature, is not feasible for this shadow benchmarking study. The dataset constructed by Sumicsid does not contain output variables that would allow us to perform a complete model-development process. Collecting, validating and processing data on additional cost drivers is beyond the scope of the shadow benchmarking exercise.

Nevertheless, it is a significant omission that the cost drivers in Sumicsid's final model were not validated through a comparison to output parameters. For example, the correlation between asset capacity (either transformer or power of circuit ends) and measures of output (such as an appropriate measure of peak demand) could provide useful evidence in supporting the use of one capacity measure over another, even if Sumicsid does not directly include outputs in the model specification.

[118] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, section 5.05.

4.4 Aggregation of NormGrid

4.4.1 Description of the issue

The TCB18 benchmarking exercise uses NormGrid as the primary output variable. As NormGrid represents a weighted sum of different asset classes, it is necessary that the weights on each asset are justified. This is particularly important if TSOs deploy assets in very different ratios to each other. If the weights are inappropriate, a derived variable such as NormGrid may mistake heterogeneity in operating characteristics with inefficiency.

4.4.2 Sumicsid's approach to aggregating asset classes

Sumicsid chooses the weights on assets based on an average cost estimate of each asset.[119] In the appendices to its main report, Sumicsid states:

> The calibration of the asset weight systems is made through linear regression towards the Capex and Opex data obtained in the project. This step scales the relative NormGrid metric towards average practice (not best practice) such that the relevant cost measures are attributed to the size proxy.

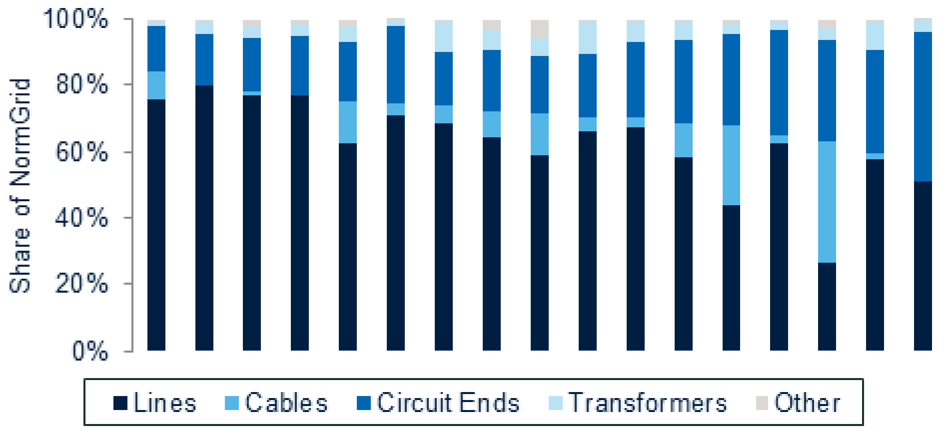
The results of this regression are not part of the final report or its appendices and there appears to have been a separate treatment of OPEX and CAPEX weights. Specifically:

- CAPEX weights appear to have been undertaken by asset category, as the groups have different weights;
- OPEX weights seem to have been estimated on an overall basis rather than by asset category.[120]

4.4.3 Critique and proposed solution

As illustrated in Figure 4.4, the distribution of assets by category across TSOs is heterogenous. For example, the share of Lines in NormGrid varies from 80% to 27% across the sample. Because of this, the weight attached to an asset class may change the relative positions of the TSOs, thus impacting estimated efficiencies.

Figure 4.4 Breakdown of NormGrid by asset categories



Source: Oxera analysis.

[119] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, section 4.72, p. 25.
[120] All assets have the same NormGrid OPEX weight.

We recognise that if NormGrid is to be used as one cost driver, some method of aggregation is required. However, the derivation of weights must be clearly described and robustly justified in the final outputs.

Table 4.7 shows the estimated coefficients from a regression of the four predominant asset classes in NormGrid on TOTEX, alongside the weights used by Sumicsid.[121] As Sumicsid states that they used the results of a regression on OPEX and CAPEX, we also regress asset classes' NormGrid OPEX and CAPEX components on OPEX and CAPEX as a sensitivity. DEA results are generally consistent with our TOTEX approach, however some coefficients are insignificant in the CAPEX regression.

The CAPEX weight Sumicsid attaches to lines and cables is similar to that implied by a regression of the assets on TOTEX. However, the weight on transformers and (to a lesser extent) circuit ends are less similar. Furthermore, the weights implied by the OPEX and CAPEX regressions (bottom two rows) are very different to the OPEX and CAPEX weights Sumicsid uses. As these regressions are not presented in Sumicsid's outputs, we are not able to identify the causes for such large discrepancies.

In addition, Sumicsid's weights do not account for the uncertainty with which the coefficients are estimated. As can be seen in Table 4.7, the standard errors on the estimated coefficients (particularly on cables, circuit ends and transformer power) are large.

Table 4.7 NormGrid aggregation weights

	Lines	Cables	Circuit Ends	Transformer
Sumicsid's weight (CAPEX)	15.79	16.99	17.83	18.91
Sumicsid's weight (OPEX)	437.08	437.08	437.08	437.08
Regression coefficient (TOTEX)	15.57*** (2.28)	17.4** (8.44)	13.52** (6.08)	292*** (35.5)
Regression coefficient (OPEX)	1789.6*** (111.0)	19142.9*** (4515.8)	217.07*** (65.1)	6595.5*** (1034.1)
Regression coefficient (CAPEX)	7.1*** (2.16)	11.2 (7.91)	9.1 (5.85)	340.7*** (33.5)

Note: Standard errors in parentheses, Stars indicate statistical significance at the 1% (***), 5% (**) and 10% (*) level.

Source: Oxera analysis.

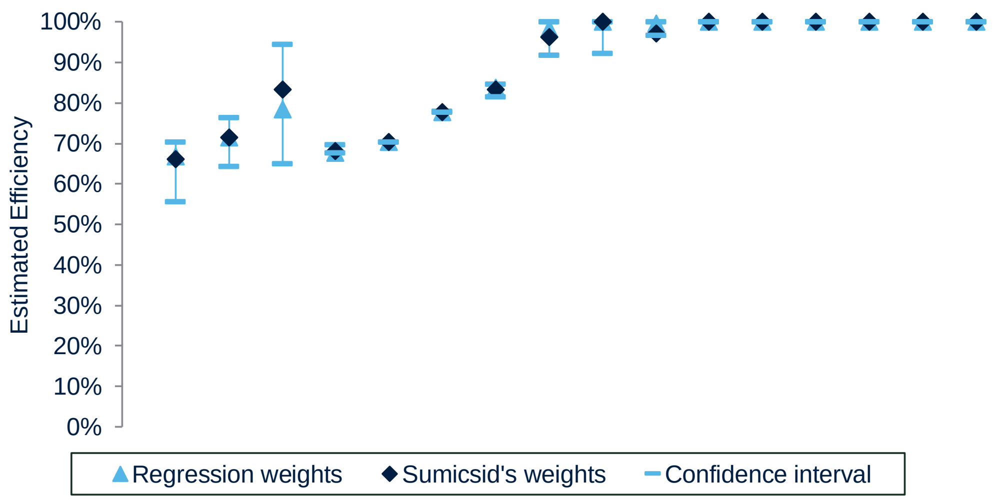
Figure 4.5 shows the estimated efficiency scores when the NormGrid weights are derived from the TOTEX regression in Table 4.7. We also present the range of estimated efficiency scores implied by the confidence interval on the estimated coefficients. The figure shows that:

- the impact of changing the NormGrid weights is small for most TSOs;[122]

[121] We find a negative and significant intercept in this regression. This is not an issue, since the focus of the investigation is the relative weight of the asset classes.
[122] This may be driven by the lack of importance attached to NormGrid in the DEA model. This is discussed further in section 5.3.

- the average efficiency score across all TSOs does not change (to one decimal place);
- 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%.

Figure 4.5 Impact of using regression-based NormGrid weights



Note: Confidence intervals for the DEA efficiencies are calculated by calculating the efficiency score for by adding or subtracting the standard error times two from the estimated regression weight.

Source: Oxera analysis.

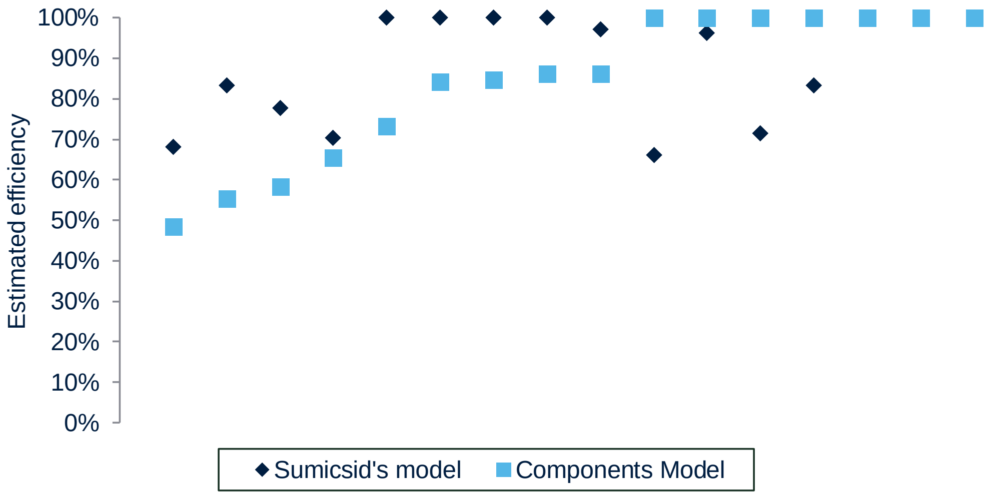
The analysis presented above still relies on parametric assumptions to derive the weight on each asset. However, Haney and Pollitt (2012)[123] argued that the use of the cost weights for the aggregation of the physical assets contradicted the principle of DEA, which chooses input and output weights in such a way as to give the firm the highest efficiency score possible.

Figure 4.6 shows the estimated efficiency scores in a model that includes the four largest components of NormGrid (lines, cables, circuit ends and transformers) as outputs and no other outputs. Controlling for each asset class separately has a relatively large impact on the estimated efficiency of some TSOs. In particular:

- three previously inefficient TSOs become efficient in this model—indeed, one TSO increases its estimated efficiency score by 29 percentage points;
- two TSOs that were previously identified as efficient are now classified as inefficient.

[123] Brophy Haney, A and Pollitt, M. G. (2013), 'International benchmarking of electricity transmission by regulators: A contrast between theory and practice?', *Energy Policy*, **62**, November, pp. 267–81.

Figure 4.6 NormGrid component model



Source: Oxera analysis.

4.5 Adjusting for environmental factors

4.5.1 Description of Issue

As a TSO's main task consists of transporting energy across a country, the features of that country's environment, such as land use, climate and topography, can be a significant driver of their costs. Costs associated with environmental factors can take the form of access costs, increased maintenance costs and increased costs of routing lines around obstacles.

Although TSOs should act in such a manner as to mitigate the impact of these environmental costs (e.g. through grid planning), these environmental characteristics are not controllable by management and thus need to be accounted for in the benchmarking process, either through the model directly or through post-modelling adjustments.

4.5.2 Sumicsid's approach to accounting for environmental factors

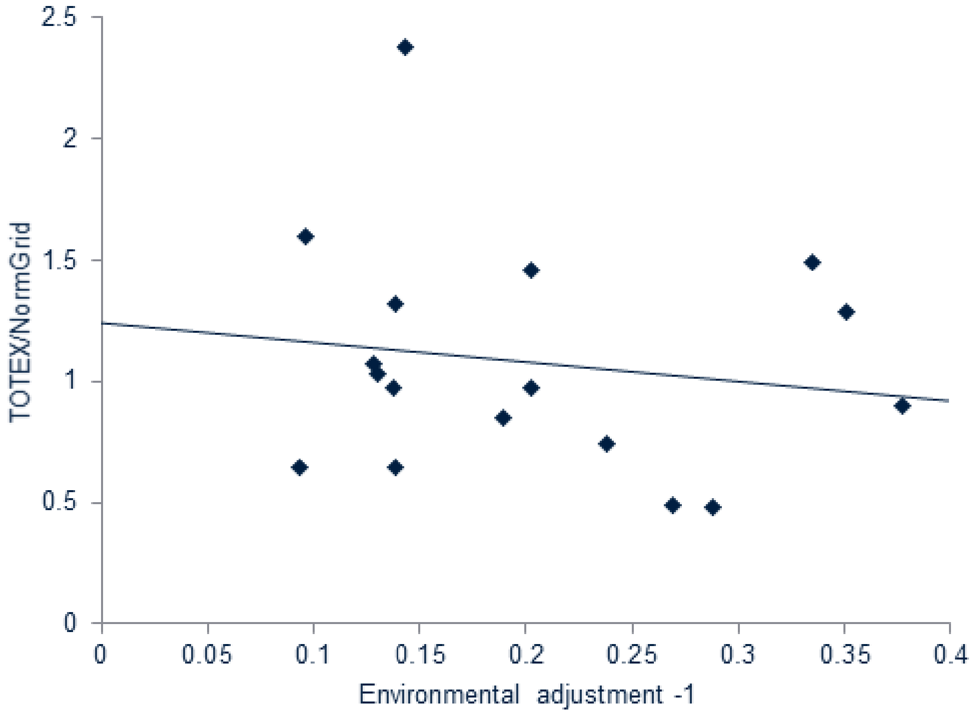
Sumicsid controls for differences in operating environments by multiplying NormGrid with a single environmental factor based on land use. The weights used to construct the environmental adjustment are not discussed or presented in the final report or its appendices, but are presented in one of the workshop.[124]

4.5.3 Critique and proposed solution

We have been unable to find the source of the environmental weights used by Sumicsid in any of its outputs.[125] Furthermore, the environmental adjustment cannot be justified with econometric techniques, such as a unit cost regression. As illustrated in Figure 4.7, TSOs with higher unit costs based on unadjusted NormGrid receive slightly lower environmental adjustments on average. That is, TSOs that operate in areas that should be more costly to operate in (based on land use factors) have lower unit costs than TSOs that operate in areas that should be less costly.

[124] Sumicsid (2019), 'Model Specification Model Results', April, slide 55.
[125] Indeed, Sumicsid presents a number of alternative environmental weights in its appendix, none of which are used in the final report. See Sumicsid (2019), 'Norm Grid Development', Table 2-5.

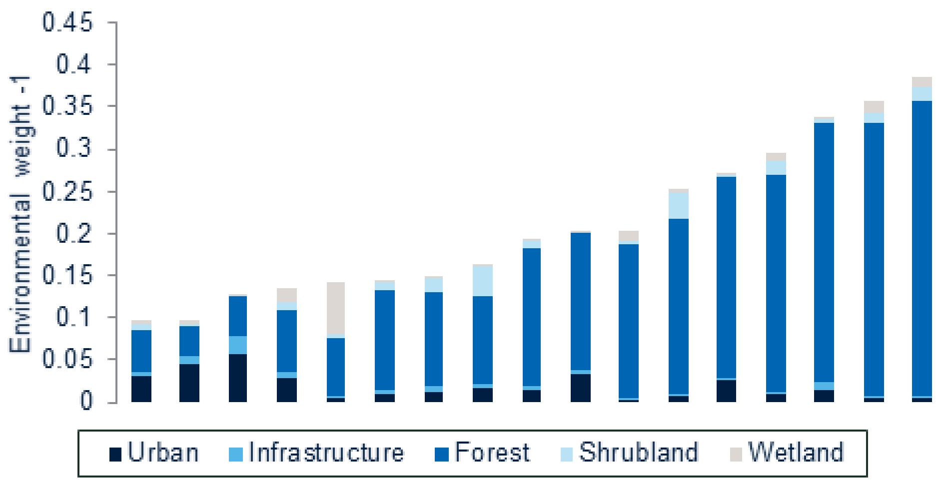
Figure 4.7 Relationship between Environmental adjustment and TOTEX per NormGrid



Source: Oxera analysis.

The lack of correlation between the environmental adjustment and unit costs could be driven by the way in which the environmental adjustment is derived. As shown in Figure 4.8, the biggest driver of most TSOs' environmental adjustment is the proportion of their service area covered in forests. Factors that may be more operationally intuitive drivers of expenditure, such as urbanity, mountainous area and the proportion of surface area covered in infrastructure, have a relatively small impact on the overall environmental adjustment.

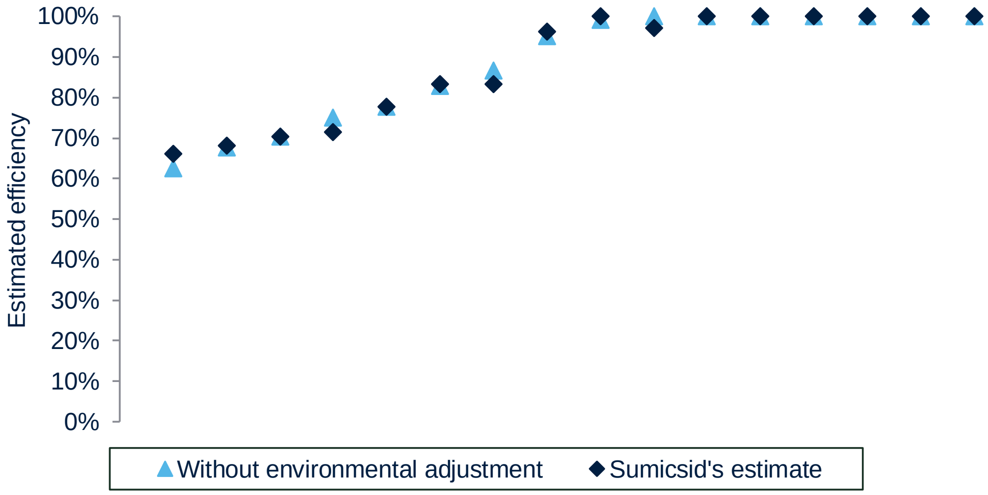
Figure 4.8 Impact of different land use categories on NormGrid weight



Source: Oxera analysis.

As shown in Figure 4.9, the environmental adjustment also has a negligible effect on estimated efficiencies. If one assumes that environmental factors are a key driver of expenditure, then one would expect an environmental adjustment to have a larger impact on the estimated efficiency scores of TSOs,[126] especially for those TSOs that operate in either extremely simple or extremely complex environments.

Figure 4.9 Impact of Sumicsid's environmental adjustment



Source: Oxera analysis.

Moreover, a satisfactory environmental factor adjustment to NormGrid should take the distribution of assets relative to the population served in the service area into account. Serving a large but relatively unpopulated area of forest does not introduce the same complexities as serving a densely populated area with significant demand for electricity. In the absence of such data, one approach to better capture the actual impact of environmental factors would be to expand the regression including the components of NormGrid suggested in section 4.4.3 to also include environmental factors. Alternatively, exogenous drivers of expenditure can be explicitly accounted for in the DEA model.

[126] Part of this is driven by the fact that the weight on NormGrid is small for most TSOs (see section 5.3).

5 TCB18 application and validation

In this section, we critically assess Sumicsid's application of DEA and its model-validation procedure. In particular, we identify the following issues.

- Sumicsid does not present an empirical justification of its returns-to-scale assumption, and we find contradictory evidence in our replication (section 5.1).
- The outlier analysis conducted by Sumicsid does not follow scientific best practice and is biased towards the non-detection of outliers (section 5.2).
- Sumicsid does not ensure that its modelling results are consistent with operational intuition and economic expectation (section 5.3).
- Second-stage analysis, as conducted by Sumicsid to validate the model, is not supported by the academic literature and is unable to detect the existence of omitted cost drivers (section 5.4).
- Sumicsid's conclusions in terms of the TSOs' efficiency gaps are not validated by alternative modelling techniques, such as SFA (section 5.6).

5.1 Returns-to-scale assumption

5.1.1 Description of the Issue

'Returns to scale' relates to how changes in inputs (i.e. TOTEX) are linked to changes in outputs (e.g. NormGrid) for efficient companies.

The choice could be a matter of policy in a national benchmarking exercise. For example, if a regulator wishes to encourage firms to move to a more productive scale size (e.g. through mergers with other firms), it could use a CRS assumption. 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.[127] As such, a VRS assumption may be more appropriate. In any case, the returns-to-scale assumption—as with any other assumption made in the modelling—should be empirically validated.

5.1.2 Sumicsid's approach

Sumicsid uses a non-decreasing returns to scale (NDRS) assumption. Sumicsid states that this is supported by the following statistical evidence.

- **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,[128] and compared to an F-distribution. If the test statistic is statistically significant, this is taken as evidence for the alternative assumption.

[127] 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.
[128] See Box 2.3 for more details.

- **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 returns to scale.[129]

However, Sumicsid does not present the results of these tests in its final outputs.

5.1.3 Critique and proposed Solution

We have performed the Banker F-test and the test on the sum of coefficients noted by Sumicsid, but we do not encounter the same results as Sumicsid.

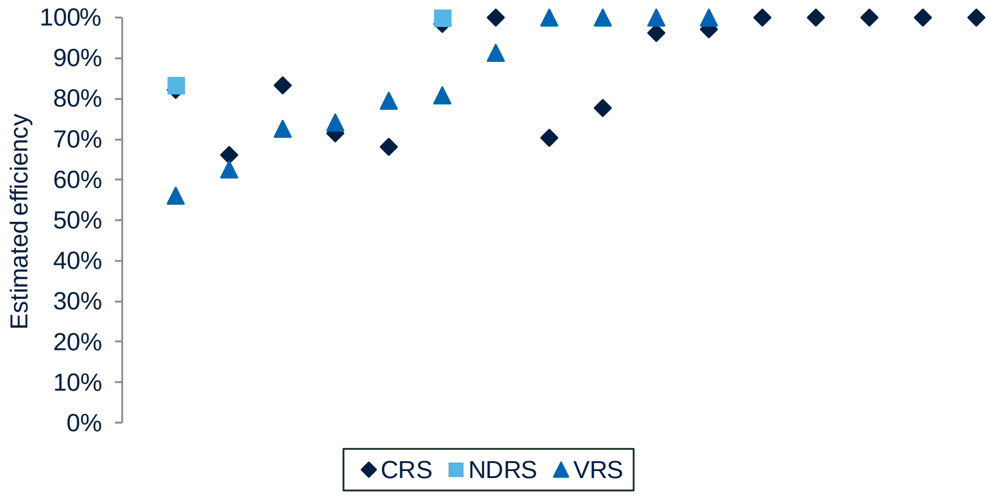
- Our application of the Banker F-test suggests that the NDRS assumption is overly restrictive. Instead, the test indicates that the VRS assumption is more appropriate.
- The sum of coefficients in logs is larger than one, but the result is not significantly different from one. This indicates that a 1% increase in cost drivers leads to an increase in costs of more than 1%. This suggests decreasing returns to scale or CRS (given the statistical insignificance of the result).[130]
- We conducted an additional parametric test by examining the coefficient of the intercept in a regression in levels. This test comes to the same conclusion as the test of the coefficients in logs, weakly supporting decreasing returns to scale.

Given the sample size, the evidence is inconclusive, and results must be combined with operational reasoning. The statistical evidence supporting the NDRS assumption is not as clear as Sumicsid states. If the test statistics and procedures used to support the statement were part of the final report and detailed in a modelling code, we may have identified reasons for our diverging findings. As it stands, however, we cannot assess the merits of Sumicsid's claims.

Figure 5.1 below shows TSOs' estimated efficiency scores under CRS, NDRS and VRS assumptions. The efficiencies of TSOs do not significantly differ between NDRS and CRS assumptions. However, under the VRS assumption, four TSOs that are assessed to be inefficient under NDRS are assessed to be efficient. Furthermore, two TSOs that are assessed to be efficient under NDRS are inefficient under VRS.

[129] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, p. 35.
[130] These simple econometric tests cannot distinguish between VRS and DRS. This could be resolved with a 'translog' model, which includes not only the cost drivers, but also their interactions and squared terms. The size of this model, with at least nine independent variables on a dataset of 17 TSOs, is not practical.

Figure 5.1 Returns to scale sensitivity



Note: NDRS and VRS efficiency scores are only displayed when they differ from the CRS efficiency score.

Source: Oxera analysis.

The returns-to-scale assumption has a significant impact on the shape of the efficient frontier and the efficiency scores of TSOs. For this reason, any decision made must be robustly justified by both conceptual arguments and empirical evidence, neither of which have been presented by Sumicsid.

5.2 Outlier analysis

5.2.1 Description of the Issue

The determination of achievable and robust efficiency targets is highly dependent on the procedure used to detect outliers and, more broadly, heterogeneity in the sample (i.e. the uncontrollable structural, operational, environmental characteristics) that is not accounted for in the model specification.

The DEA model, as applied by Sumicsid, determines the frontier based on those TSOs deemed to be efficient without making any allowance for errors. For the TSOs to be sufficiently comparable to conduct robust DEA, data should be measured without errors, all relevant exogenous differences should be considered as cost drivers, and any excluded heterogeneity in the model specification should be addressed through an effective outlier procedure. Thus the estimation of an efficient frontier is greatly influenced by the existence of outliers, especially in deterministic methods of benchmarking such as DEA.

In the context of this study, where the TSOs being assessed operate in very different operating environments and data errors are particularly prevalent (see section 3.1), outliers are likely to be present in the sample.

5.2.2 Sumicsid’s approach

Sumicsid follows the same outlier procedure as the Bundesnetzagentur. As such, the approach 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 as it states this TSO is ‘almost always [...] an extreme outlier’ in the various model specifications it tested.

Other outlier tests are discussed in the main report,[131] but it is not clear if or how these have been used in the final analysis.

5.2.3 Critique and proposed solution

As with other parts of its analysis, Sumicsid does not provide any rationale for focus on the outlier procedure specified in the ARegV. The purpose of the ARegV is to enable national comparisons among German regulated entities. It is neither legally binding nor sufficient in an international benchmarking context.

Setting aside the insufficiency of the ARegV in TCB18, the Bundesnetzagentur’s outlier procedure has been challenged in the past as it is inconsistent with the academic literature and fails to identify all outliers. Furthermore, Sumicsid mentions alternative approaches to outlier detection in the main report (such as an examination of DEA weights and the econometric method)[132] but does not explain how or if these methods were used at any stage in the TCB18 study.

Specific issues relating to the dominance test and super-efficiency are outlined below.

Dominance test

In an expert opinion on Sumicsid’s dominance test in the benchmarking of German TSOs using DEA, Kumbhakar, Parthasarathy and Thanassoulis (2018)[133] concluded that the dominance test is neither legally consistent with the ARegV nor based on any theoretical foundation. The legal consistency is less of an issue in the TCB18 study as there are no binding legal requirements for this international benchmarking exercise. More importantly, however, the dominance test is not supported in the academic literature and suffers from a number of flaws.

- Although not relevant to TCB18, Sumicsid’s dominance test in itself addresses only the second requirement of the ARegV (i.e. that the impact of removing the potential outlier must be statistically significant), as the test statistic is indifferent to the number of units affected by an outlier (required by the first requirement).
- The dominance test requires the TSO efficiencies to follow a half-normal distribution, which is inconsistent with the non-parametric nature of the DEA method.
- Banker (1993) notes that his tests are asymptotically valid under the maintained assumptions, which means that they are appropriate only for large samples.[134]
- Efficiencies of the same units are being compared in the numerator and denominator of the test statistic. This invalidates the Banker tests (1993, 1996), which require independent samples of efficiencies to be compared.

[131] Sumicsid (2019), ‘Pan-European cost-efficiency benchmark for electricity transmission system operators main report’, July, p. 30.
[132] Sumicsid (2019), ‘Pan-European cost-efficiency benchmark for electricity transmission system operators main report’, July, p. 30.
[133] Kumbhakar, S., Parthasarathy, S. and Thanassoulis, E. (2018), ‘Validity of Bundesnetzagentur’s dominance test for outlier analysis under Data Envelopment Analysis’, August.
[134] Banker, R. D. (1993), ‘Maximum Likelihood, Consistency and Data Envelopment Analysis: A Statistical Foundation’, *Management Science* **39**:10, pp. 1265–73; (1996), ‘Hypothesis tests using data envelopment analysis’, *The Journal of Productivity Analysis*, **7**, pp. 139–59.

In summary, the dominance test has no theoretical basis and cannot be used to identify dominant units, regardless of the sample size.

In response to Kumbhakar, Parthasarathy and Thanassoulis (2018), Sumicsid (2019) has, without exception, agreed with the four limitations highlighted in our expert report and summarised above. Sumicsid (2019)[135] specifically notes that:

> In summary, we consider the four objections towards the F-distribution raised by the Oxera note as valid.[136]

Nevertheless, Sumicsid (2019) argues that the test 'makes the most use of available information' and that it 'is a cautious test'.

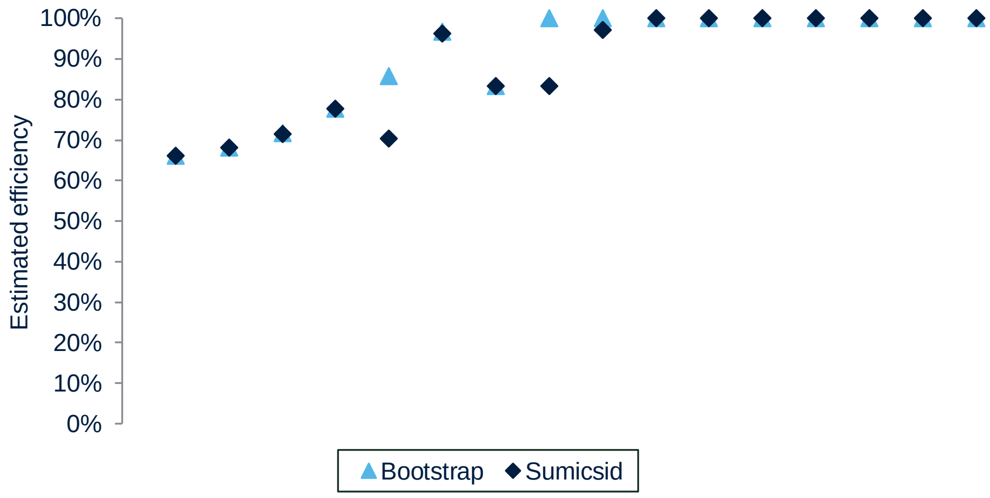
In response to Sumicsid (2019), Kumbhakar, Parthasarathy and Thanassoulis (2019)[137] showed not only that the test lacks a theoretical foundation, but also that it is biased towards non-rejection (i.e. non-identification of potential outliers). The claim made in Sumicsid (2019) that its test is cautious does not stand up to scrutiny which we also evidenced empirically in Kumbhakar, Parthasarathy and Thanassoulis (2019).

There are better alternatives for the dominance test, but they require further theoretical development. The most promising candidate is the bootstrap test proposed in Kumbhakar, Parthasarathy and Thanassoulis (2018). The bootstrap-based test is a non-parametric test consistent with the non-parametric nature of DEA, requires fewer assumptions than other non-parametric options explored in Kumbhakar, Parthasarathy and Thanassoulis (2018), and explicitly takes into account the pairing structure of efficiencies both with and without a potential outlier.[138]

Figure 5.2 shows the estimated efficiency scores when Sumicsid's dominance test is replaced with our recommended bootstrap-based test. The test identifies two additional dominant outliers[139] and increases the number of efficient TSOs by two.

[135] Sumicsid (2019), 'Outliers in DEA based regulatory benchmarking response to the Oxera report', October.
[136] Sumicsid (2019), 'Outliers in DEA based regulatory benchmarking response to the Oxera report', October, p. 16.
[137] 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.
[138] Kumbhakar, S., Parthasarathy, S. and Thanassoulis, E. (2018), 'Validity of Bundesnetzagentur's dominance test for outlier analysis under Data Envelopment Analysis', August.
[139] In our replication of Sumicsid's analysis, we did not identify any dominant outliers.

Figure 5.2 Impact of using a bootstrapped dominance test



Source: Oxera analysis.

Super-efficiency test

Although the super-efficiency test as applied by Sumicsid has a theoretical foundation, it is insufficient to detect all abnormally super-efficient units.

- **Outliers are unlikely to be identified in volatile samples.** The critical efficiency value, above which a TSO is identified as an outlier, is directly proportional to the inter-quartile range (IQR) of the estimated efficiency scores. Therefore, if the results from a DEA model are highly volatile, this could increase the IQR on average, which would imply that the critical efficiency value would be higher for a unit to be identified as an outlier. Such a high-volatility sample could be a result of unidentified heterogeneity or an imperfect model.

- **Sumicsid's analysis is vulnerable to masked outliers.** For example, there could be two network operators that are similar in characteristics and far removed from the rest of the sample, but one is masked (covered or hidden) by another. This is referred to as a 'masked' outlier. In such a case, a mechanistic application of the super-efficiency test could fail to identify either one as an outlier. Note that the same issue can also occur with respect to the dominance test.

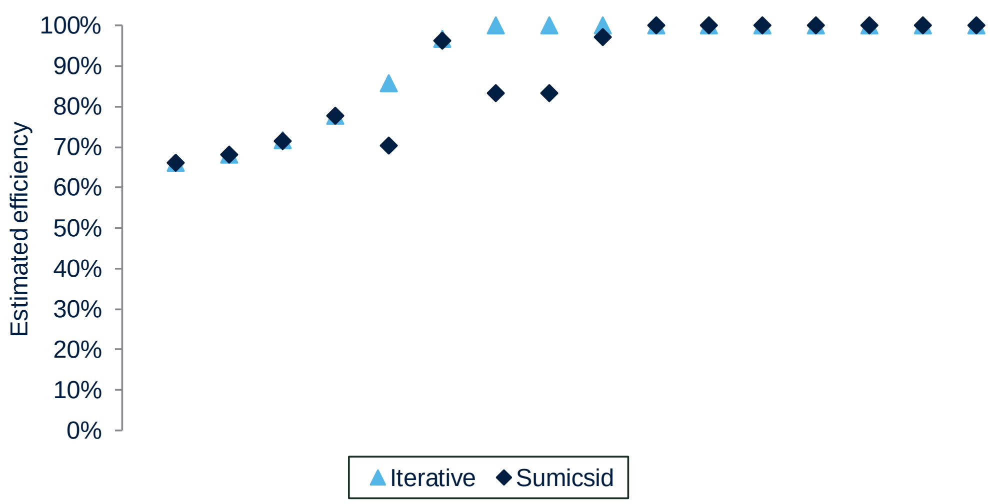
Only a sequential application of the super-efficiency outlier tests that allow for the possibility of masking could reveal two or more 'joint' outliers. Deuchert and Parthasarathy (2019)[140] and Thanassoulis (1999)[141] suggest that a sequential exploration of outliers should be pursued. After each step, conspicuous companies based on the absolute super-efficiency threshold are excluded and the remaining companies are investigated for additional abnormalities.

A sequential application of super-efficiency outlier tests would result in the identification of more outliers, resulting in a more homogeneous sample for efficiency estimation.

[140] 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, ew–Magazin für die Energiewirtschaft.
[141] Thanassoulis, E, (1999) 'Setting Achievement Targets for School Children', *Education Economics*, **7**:2, pp. 101–19.

Figure 5.3 shows the distribution of estimated efficiency scores when the super-efficiency test is applied iteratively (here we have not adjusted for the super-efficiency threshold issue highlighted above). In this case, three more outliers are identified and the three more TSOs are estimated to be fully efficient.

Figure 5.3 Impact of using an iterative super-efficiency test



Source: Oxera analysis.

It is important to note that all outlier procedures rely on assumptions and is contingent on the model specification. Furthermore, most tests require somewhat arbitrary thresholds to determine whether a particular observation is an outlier. For this reason, it is essential that the data collection and construction process is robust to identify obvious data errors. Furthermore, the final model used to estimate efficiency scores must appropriately capture differences in operation characteristics across TSOs. In other words, even a robust and academically valid outlier procedure is not a replacement for proper data processing and model development.

5.3 Model validation—DEA weights

5.3.1 Description of the issue

In an econometric setting, it is relatively straightforward to assess whether a particular cost driver has an operationally intuitive impact on efficient expenditure—the sign, magnitude and statistical significance of the estimated coefficients can determine whether the estimated relationship between cost and cost drivers in a benchmarking model is appropriate. Similar validation is required in DEA, where the assessment is of the weights on the input and output factors, and the peers and their weights for the inefficient TSOs. This is an important step to show that the method (i.e. DEA) is appropriate for the dataset (as would be expected of any other method).

DEA weights can be used to assess whether a model is operationally intuitive. 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. Indeed, in past iterations of the TCB18 study, Sumicsid imposed restrictions on the weights to ensure that the relative importance of particular cost drivers remain operationally intuitive for all TSOs.[142]

Alongside aiding model development, the weights can be used to identify unusual TSOs. If a TSO's efficiency score is driven largely by one less relevant cost driver, or if the weights imply an unintuitive relationship between the cost drivers, then the TSO may be an outlier.

5.3.2 Sumicsid's approach

Sumicsid discusses the use of DEA weights in its main report. For example, it states that the 'marginal substitution ratios can reveal whether an observation is likely to contain errors',[143] where the marginal substitution ratios are calculated as the ratios of the DEA weights. Furthermore, Sumicsid briefly mentions the use of weight restrictions and how they are not necessary in the current context.[144]

However, the DEA weights are not discussed anywhere in the final outputs (including the TSO-specific outputs), so it is unclear exactly how the weights (or in fact, any of the DEA outputs) have been used in the model-development and validation processes.

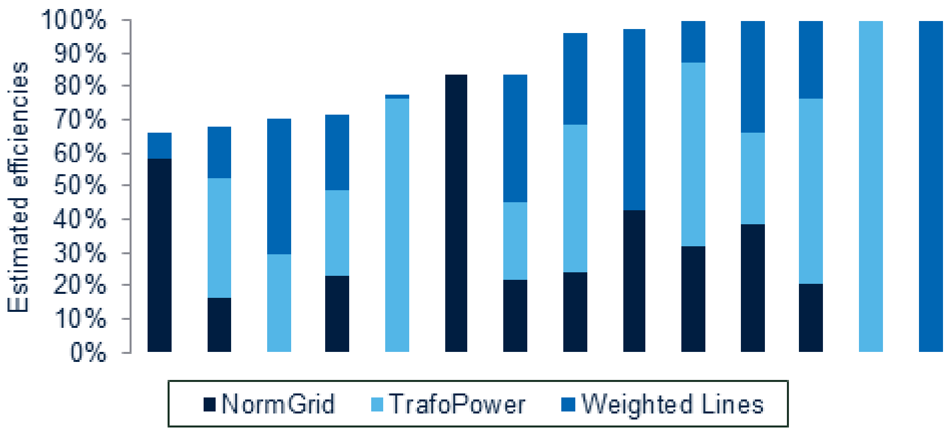
5.3.3 Critique and proposed solution

The apparent lack of validation through examination of DEA results is concerning. In particular, some of Sumicsid's statements that are unsubstantiated in the main report could be confirmed or contradicted by an examination of DEA weights. For example, Sumicsid's statement that adjusted NormGrid is the primary cost driver, followed by weighted lines,[145] could be empirically supported by an examination of the DEA weights.

Figure 5.4 shows the importance of each cost driver in determining the efficiency of the TSOs. Interestingly, 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. This is counterevidence to Sumicsid's statement that NormGrid is the strongest cost driver.[146]

[142] For example, see Frontier Economics, Sumicsid and Consentec (2013), 'E3GRID2012 – European TSO Benchmarking Study, A report for European Regulators', July.
[143] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, p. 30.
[144] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, p. 46.
[145] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, pp. 32–34.
[146] It is well known that multiple optimal solutions (i.e. more than one set of DEA weights) exist for the efficient units (i.e. peers). However, Sumicsid has not demonstrated that NormGrid is always the main efficiency driver by exploring alternative optimal DEA weights sets for firms where the initial finding is otherwise.

Figure 5.4 Weights of outputs in Sumicsid's DEA



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

Source: Oxera analysis.

Without an evaluation of DEA weights, many statements made in the report—including the one about the importance of NormGrid's—are unverifiable. From the results presented in Sumicsid's report, we do not know if NormGrid is actually instrumental in determining efficiency. In fact, an examination of DEA weights reveals that NormGrid is not the primary driver of efficiency for most TSOs. Additionally, removing NormGrid does not have the largest effect on overall efficiency.

Furthermore, other outputs from DEA modelling, such as peer analysis, appear to have been ignored in Sumicsid's final reports. In DEA, each inefficient TSO will have a corresponding set of efficient peers. It is therefore possible 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.

An alternative to the assessment of homogeneity is to examine the weights on peers (known as 'scaling factors' or 'lambdas'). The scaling factors indicate how much an efficient TSO has to be scaled up or down to assess the inefficient TSO to which it is a peer. The scaling factors for the inefficient TSOs are summarised in Table 5.1 below.

Table 5.1 Sum of scaling factors for inefficient TSOs

	Average	**Minimum**	**Maximum**
Scaling factor	4.10	1.00	12.15

Source: Oxera analysis.

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. In any case, the outputs from DEA should be validated to be useful for learning or any other purpose.

One efficient TSO is a peer to six of the nine inefficient TSOs. For one inefficient TSO, this peer is scaled up by a factor of 9.8. For another inefficient

TSO, the peer is scaled up by a factor of 5. These scaling factors are unusual and should have been validated.

5.4 Model validation—identification of omitted cost drivers

5.4.1 Description of the issue

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.

5.4.2 Sumicsid's approach

Sumicsid uses second-stage regressions to test for the exclusion of relevant cost drivers. It is not clear from the individual report whether the regression is estimated using OLS, ROLS or some other estimator, but Sumicsid states in the main report that:

> second-stage analyses are typically done using graphical inspection, non-parametric Kruskal-Wallis tests for ordinal differences and truncated Tobit regressions for cardinal variables.[147]

Sumicsid further states that such second-stage analysis of this sort is 'routinely done' to identify omitted cost drivers.[148] This analysis is presented in the individual report.

5.4.3 Critique and proposed solution

The second-stage analysis forms the core of Sumicsid's model validation in the individual reports.[149] However we are not aware of any literature that specifically focuses on using DEA, followed by regression, to justify a set of input–output variables.[150] Second-stage regressions are sometimes used to adjust efficiencies from a first stage DEA model;[151] however, this serves a different purpose to that stated by Sumicsid. Furthermore, the second-stage analysis used in the literature is valid if the following conditions hold.

- The location of the efficient frontier based on the first-stage analysis is **not** affected by the variables used in the second-stage analysis.[152]
- The first- and second-stage variables should be independent of each other, although they can be correlated within first- or second-stage variables.

---

[147] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, paragraph 4.09.
[148] Sumicsid (2019), 'Project TCB18 Individual Benchmarking Report Fingrid – 131', July, p. 35.
[149] For example, see Sumicsid (2019), 'Project TCB18 Individual Benchmarking Report Fingrid – 131', July, chapter 5.
[150] The Bundesnetzagentur has used such analysis in the past to identify omitted cost drivers. but such use is not supported by academic literature and was strongly challenged by the industry.
[151] For example, see the Norwegian Water Resources and Energy Directorate, 'Guidelines for revenue cap calculation in R', section 4.2.
[152] See Simar, L. and Wilson, P.W. (2007), 'Estimation and inference in two-stage, semi-parametric models of production processes', *Journal of Econometrics*, **136**:1, pp. 31–64; (2011), 'Two-stage DEA: caveat emptor', *Journal of Productivity Analysis*, **36**:205. Banker and Natarajan (2008) adopt a different variant of separability, as noted below. See Banker, R.D. and Natarajan, R. (2008), 'Evaluating contextual variables affecting productivity using data envelopment analysis', *Operations Research*, **56**:1, pp. 48–58.

- Estimates of efficiency from the first stage should first be adjusted for the serial correlation bias, because the dataset in the first stage is finite. The bias correction can be estimated using a bootstrapping procedure.[153]

As the first-stage DEA model does not offer information on statistical significance of the outputs and its second-stage analysis for further model validation is erroneous, there is a need to provide additional robustness checks to see what happens to the DEA efficiency scores if some of the omitted variables are used alongside others in the first stage. This could potentially identify TSOs that are particularly susceptible to certain cost drivers, even if those cost drivers are not significant at the industry level.

We use an empirical example to illustrate the invalidity of Sumicsid's approach. We first estimated the efficiencies of TSOs with two of Sumicsid's cost drivers, excluding one. We then tested whether the omitted cost driver would be identified as such using Sumicsid's second-stage analysis. Sumicsid makes a number of claims with respect to its final model, including the importance of each cost driver in explaining costs.[154] One would therefore expect all cost drivers to be statistically significant at the second stage—especially adjusted NormGrid, given that Sumicsid states that this is the main cost driver.

Table 5.2 below shows that for two out of three cost drivers, the omitted cost driver was statistically insignificant and would not be identified as an omitted variable in Sumicsid's second-stage analysis. For example, if Sumicsid had started with a two-output model controlling for weighted lines and transformer power, it would not have identified NormGrid as an omitted output and would have concluded that the two-output model is validated. Furthermore, the 'next strongest' cost driver, weighted lines, would also not have been identified as an omitted driver. Indeed, the only driver that would be identified is transformer power.

Table 5.2 Results of the second-stage validation

Omitted cost driver	Coefficient on omitted cost driver[1]	P-value[2]	Relevant (omitted) variable?[3]
Adj. NormGrid	-1.1E-10	0.48	No
Transformer Power	-2E-06	0.01	Yes
Weighted Lines	-2.6E-05	0.19	No
Environment	0.06014	0.88	No

Note: [1] Coefficient of omitted cost driver in a regression on efficiency scores determined with a two-output model omitting this driver. [2] Probability of the coefficient occurring, given that the 'real' coefficient is zero. [3] Significantly different coefficient from zero at the 5% level.

Source: Oxera analysis.

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. It cannot be argued, therefore, on the basis of this procedure, that no relevant cost drivers were omitted, nor that the first-stage cost model is validated.

[153] See Simar, L. and Wilson, P.W. (2007), 'Estimation and inference in two-stage, semi-parametric models of production processes', *Journal of Econometrics*, **136**:1, pp. 31–64; (2011), 'Two-stage DEA: caveat emptor', *Journal of Productivity Analysis*, **36**:205.
[154] Sumicsid (2019), 'Pan-European cost-efficiency benchmark for electricity transmission system operators main report', July, pp. 32–34.

5.5 Model validation—SFA

5.5.1 Description of the issue

DEA is one method of assessing the efficiency of TSOs. Although it has certain advantages over other methods (for example, it does not impose strict assumptions on the relationship between inputs and outputs), it suffers from being a deterministic method of efficiency assessment (as considered by Sumicsid) and the results are contingent on assumptions imposed on the model (e.g. returns to scale) which have not been sufficiently motivated by Sumicsid. In this respect, it treats the data ‘as given’ and makes no allowance for uncertainty in the variables.

It is therefore common for regulators to use alternative benchmarking methods, either to directly inform the efficiency target or as a cross-check to the results from DEA. For example, the Bundesnetzagentur uses four models to estimate a DSO’s relative efficiency, two of which are estimated via DEA and two of which are estimated via SFA. A DSO’s efficiency score is the best of the four, with a lower bound of 60%.

SFA is particularly relevant when we suspect a lot of uncertainty in the data. As highlighted in section 2, there is significant uncertainty in the data used by Sumicsid (this includes issues with its data processing and data adjustments), and it is therefore important that the results from DEA are supported by SFA models.

5.5.2 Sumicsid’s approach

Sumicsid has not used SFA or any other benchmarking method to validate its model, nor has it used SFA to assess the robustness of the TSOs’ efficiency scores.

Sumicsid states:

> In a study of European electricity TSOs, the number of observations is too small for a full-scale application of SFA as main instrument. We have therefore used DEA as our base estimation approach, in line with regulatory best practice and earlier studies such as E2GAS and E3GRID.[155]

Furthermore, Sumicsid has stated at previous workshops that there are convergence issues[156] when attempting to estimate the SFA models on the current dataset.

5.5.3 Critique and proposed solution

All models are imperfect representations of reality, and any one model could overestimate or underestimate an individual TSO’s efficiency. As such, it is important to consider valid alternatives, both in terms of cost driver selection and estimation technique.

While most empirical investigations (including both DEA and SFA) perform better on larger samples, there is no fixed rule as to how many observations a model needs. Indeed, SFA and other econometric methods have been used on smaller sample sizes than that available to Sumicsid.[157] As such, sample size

[155] Sumicsid (2019), ‘Pan-European cost-efficiency benchmark for electricity transmission system operators main report’, July, p. 29.
[156] If a model fails to converge, it means that the iterative procedure used to estimate the models results in an endless loop. In this sense, the model cannot be estimated.
[157] For example, the ORR used estimated SFA models with 14 infrastructure managers (although the time series component was longer). The ORR also performed SFA on a sample of 50 observations for its

in and of itself is not a valid justification for ignoring SFA, and the appropriateness of a method has to be determined empirically on the data and model used.

As a general related observation, we note that despite deriving its model on a panel dataset (i.e. data over time across TSOs), Sumicsid has not effectively used all the information it has at its disposal. Instead, it has focused on a single year's data without justification. Also, Sumicsid should have validated the outputs from DEA (e.g. peers and weights) to show that the method was appropriate for the dataset, but it has not.

The non-consideration of SFA is also in conflict with the lack of consideration given to sample size in Sumicsid's current approach—for example, with respect to its second stage analysis and the dominance test, which also rely on large samples (among other conditions) to be appropriate. Sumicsid made no attempt to justify the size of its sample for this analysis and has not validated the DEA output to show that it is valid for the dataset, yet it has relied on sample size alone to justify ignoring SFA models from the evidence base.

The lack of convergence that Sumicsid has discussed (even though no evidence in the form of, say, modelling code was made available) may cast doubt on the appropriateness of the model specification that it has selected. Lack of convergence typically has three interpretations:

- the model is mis-specified, which could mean that key drivers have been omitted and/or the functional form is incorrect;
- there is too much statistical noise in the data for the residual to be decomposed into noise and inefficiency;
- there is no inefficiency in the sample.

Overall, the non-convergence of SFA models should not be used as an argument against using SFA. When SFA models are estimated on our dataset, we do not encounter convergence issues. For example, Table 5.3 shows the SFA models when estimated on a cross section (i.e. year by year).

It is possible to test for the absence of inefficiency using a likelihood ratio (LR) test. If inefficiency is not present in the sample, the SFA model reduces to a standard OLS model with normally (symmetrically) distributed errors. As can be seen in Table 5.3, the estimated efficiency scores are statistically insignificant in all years, suggesting that much of the estimated efficiency gap is due to statistical noise rather than inefficiency.

determination of the efficiency on Network Rail as part of the PR18 price control. See The Office of Rail and Road (2013), 'PR13 Efficiency Benchmarkings of Network Rail using LICB', August, p. 6; (2018), 'PR18 Econometric top-down benchmarking of Network Rail A report', July, p. 43.

Table 5.3 Cross-sectional SFA

	2013 data	2014 data	2015 data	2016 data	2017 data
NormGrid	0.631***	0.609**	0.591**	0.507**	0.472**
Transformer Power	0.465**	0.548**	0.516**	0.560***	0.559***
Weighted Lines	0.0478	0.0272	0.002	0.0011	0.0115
Constant	1.398	1.073	2.061	3.556	4.169
Number of observations	16	16	16	16	17
Does the LR test detect statistically significant[1] inefficiency?	No	No	No	No	No

Note: All models presented in this table assume that inefficiency follows a half-normal distribution and are estimated in logarithms. The models do not materially change if alternative distributional assumptions are made (such as exponential or truncated normal) but the models are more difficult to estimate and sometimes do not converge. [1] Statistical significance is determined at the 5% significance level.

Source: Oxera analysis.

The results in Table 5.3 should be interpreted with caution due to the size of the sample on which the model is estimated. Samples of 16 or 17 observations are generally considered small to allow for robust statistical inference, but regulatory benchmarking applications exist on similarly sized datasets.

The models converge, but the estimated inefficiency is statistically insignificant in all of the models presented in the table.

Table 5.4 shows similar results estimated on the full sample (i.e. all 81 observations included in one model). Specifically, we present results for:

- pooled SFA models, where each observation is assumed to be independent;
- time-invariant SFA, where a TSO's inefficiency is assumed to be constant in the analysis period;
- time-varying SFA, where parameters describing the distribution of inefficiency are allowed to vary through time, according to a linear trend.

The models converge, but the estimated inefficiency is statistically insignificant in all of the models presented in the table.

Table 5.4 Panel SFA

SFA estimator	Does the LR test detect statistically significant[1] inefficiency?	Intuitive interpretation of coefficients?[2]
Pooled SFA	No	Yes
Time-invariant SFA	No	Yes
Time-varying SFA	No	Yes

Note: All models presented in this table assume that inefficiency follows a half-normal distribution and are estimated in logarithms. A linear trend is added to the cost driver specification to account for changes in expenditure through time. [1] Statistical significance is determined at the 5% significance level. [2] In this context, 'intuitive interpretation' means that the estimated coefficient on each output variable is positive.

Source: Oxera analysis.

The SFA model not finding statistically significant inefficiency is not a reason to use DEA; rather, it suggests caution is warranted against interpreting any estimated inefficiency in the DEA as actual inefficiency rather than statistical noise, and/or that the models specification should be reconsidered.

5.6 Frontier shift

5.6.1 Description of the issue

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. The assessment of frontier shift is a critical aspect of regulatory benchmarking, as the frontier shift productivity improvements (or deteriorations) can be achieved by all companies in an industry.

5.6.2 Sumicsid's approach

Sumicsid does not discuss frontier shift in its main report or associated appendices. We understand that frontier shift was discussed at one workshop, but the results are not presented in the slides.[158, 159]

5.6.3 Critique and proposed solutions

Because Sumicsid does not discuss frontier shift analysis, we unable to assess the validity of the methods it used or the robustness of the final outputs. However, we have estimated frontier shift using the same input and output variables that Sumicsid has used in its analysis of relative efficiencies. Consistent with scientific best practice, we use a CRS assumption when estimating frontier shift.[160] The frontier shift results are shown in Figure 5.5.

The analysis indicates that the frontier has been *regressing* at a rate of 4% p.a. on average. This suggests, all else equal, that efficient costs are increasing in the analysis period. Such a large and negative frontier shift result differs greatly to what is often applied in regulatory contexts[161] and could be indicative of model mis-specification rather than genuine deteriorations in productivity. In particular, as Sumicsid's cost drivers do not explain changes in expenditure well over time, relevant cost drivers that explain changes in costs over time (such as asset health) may be missing, and the position of a TSO in the investment cycle (as they have not been sufficiently normalised) may also impact the estimated results.

The inability of the DEA model to account for changes in efficient expenditure over time also raises additional concerns regarding the ability of the model to

---

158 See Sumicsid (2019), 'Model Specification Model Results', April, slide 81.

159 Sumicsid published the results of the dynamic efficiency analysis after the finalisation of this report. See Sumicsid (2020), 'Dynamic efficiency and productivity changes for electricity transmission system operators', April.

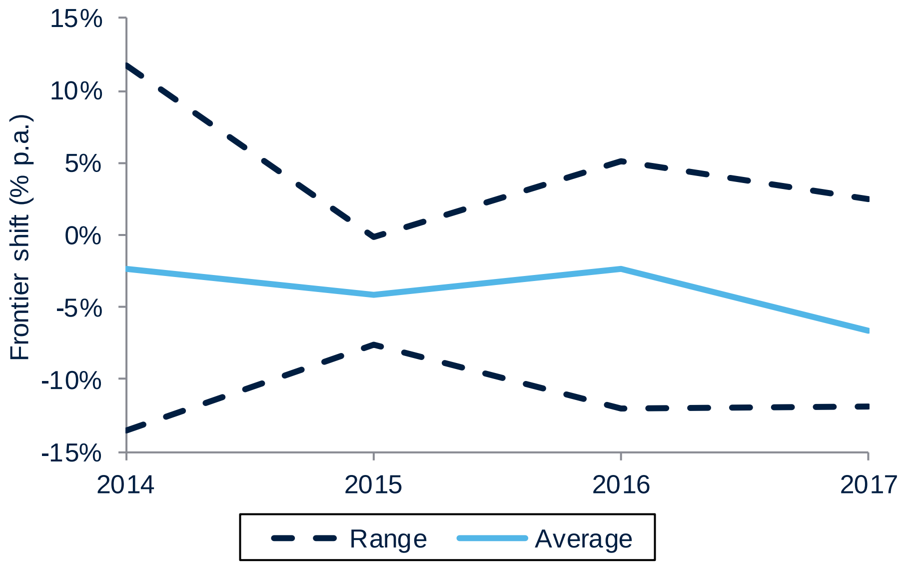
160 See, for example, Thanassoulis, E. (2001), *Introduction to the Theory and Application of Data Envelopment Analysis: A foundation text with integrated software*, Kluwer Academic Publishers, pp. 177–178, and Färe, R., Grosskopf, S. and Margaritis, D. (2008), 'Efficiency and productivity: Malmquist and more', *The measurement of productive efficiency and productivity growth*, **5**, pp. 522–622.

161 For example, Ofwat has applied a 1.1% p.a. frontier shift challenge to water companies in the PR19 final determination. See Ofwat (2019), 'PR19 final determinations Securing cost efficiency technical appendix', December, p. 168. The Bundesnetzagentur applied a frontier shift target (known as Xgen) of 0.9% p.a. in electricity distribution. See Bundesnetzagentur (2018), 'BK4-18-056 Beschlusskammer 4', November. Ofgem applied a -0.3% ongoing efficiency target (frontier shift net of input price pressure) to gas distribution networks and a -0.7–0.1% p.a. ongoing efficiency target to transmission and system operators. See Ofgem (2012), 'RIIO-T1/GD1: Real price effects and ongoing efficiency appendix', December, Table 3.1.

account for differences in efficient expenditure across TSOs. For example, if the model is unable to capture the general trend of increasing regulatory burden over time, then it is unlikely that the model can capture differences in regulatory burden between TSOs. In this case, a TSO's estimated efficiency score will be driven by differences in regulatory burden as well as (or rather than) genuine differences in efficiency.

Similar arguments could be made in relation to other factors that are not captured by the model, such as changes in input prices, weather conditions, asset health, economic environment, objections to new assets from local residents, and quality of service.

Figure 5.5 Frontier shift—DEA Malmquist



Note: A positive number indicates an improvement in productivity.

Source: Oxera analysis.

The analysis of DEA models is supported by results from SFA. Table 5.5 shows the SFA models that we have used to determine the rate of frontier shift in the sample. The estimated coefficient on the time trend suggests that efficient costs are increasing at a rate of 2.3–4.7% p.a., which is consistent with the 4% p.a. increase in efficient costs estimated by the DEA model. However, the estimated frontier shift is statistically insignificant in some specifications, indicating that the data is also consistent with there being no frontier shift at all. This is further evidence that the dataset is 'noisy' and Sumicsid's focus on deterministic approaches that do not allow for statistical inference is misleading.

Table 5.5 Frontier shift—SFA

	Pooled SFA	Panel SFA (time-invariant)	Panel SFA (time-varying)
NormGrid	0.551***	0.929***	0.891***
Transformer Power	0.520***	0.147	0.12
Weighted lines	0.0146	0.065	0.115
Time trend	0.0464	0.0284***	0.0229
Constant	2.688*	-1.593	-0.933
Implied frontier shift (% p.a.)	**-4.7%**	**-2.9%**	**-2.3%**

Note: All models presented in this table assume that inefficiency follows a half-normal distribution and are estimated in logarithms.

Source: Oxera analysis.

The inability of the models to capture changes in expenditure over time should have been identified in the cost driver analysis phase of the benchmarking study (for example, through statistical tests to determine whether the relationship between costs and cost drivers stays constant through time).

The relative efficiency scores and frontier shift estimates in the TCB18 study are fundamentally related. It cannot be that one piece of evidence from TCB18 is deemed more robust than another, given that both parameters are determined by the same data, model and methodology. Therefore, if the relative efficiency scores are to be used by NRAs to set cost allowances for TSOs, then the frontier shift results must also be considered and used in validating the outputs from the study. Sumicsid's omission of frontier shift analysis from its final outputs is particularly concerning, as it does not allow NRAs to use complete information when setting cost allowances for TSOs.

6 Conclusion

International benchmarking can be a powerful tool for companies and regulators to assess the efficiency of TSOs. This is especially true in the context of the electricity transmission industry, where the sector is often characterised by national monopolies, thus making national benchmarking challenging. In this sense, we welcome projects such as the TCB18 study and its predecessors, which have attempted to develop a framework for the regular assessment of TSOs.

Nevertheless, the TCB18 study itself suffers from a number of weaknesses that mean the estimated efficiency scores cannot be interpreted as 'true' differences in efficiency in their current form. Some of these weaknesses, such as consistency in reporting guidelines, are partly driven by the lack of maturity in the international benchmarking process. We expect that some of these issues could be resolved with time and in future iterations of the study, as TSOs and NRAs become more familiar with the process. However, Sumicsid's concluding remarks in its main report are concerning, as they are not consistent with the significant issues and areas for future work identified through our review. For example:

> Regulatory benchmarking has reached a certain maturity through this process and model development, signaling both procedural and numerical robustness. Drawing on the work, the definitions and data standards as well as the model, CEER can readily plan for a repeated regular benchmarking at a considerably lower cost in time and resources, to the benefit of all involved. Although the current model brings improvements in particular in environmental factors, the inflation and salary corrections and the NormGrid definitions, the relative symmetry with the earlier model from E3GRID can be seen as a confirmation of the type of parameters and approaches chosen, leading to stable and predictable results. In this manner, the **future work can be directed towards further refinement of the activity scope and the interpretation of the results, rather than on the model development**.
>
> [emphasis added]

We have identified several areas of the benchmarking project that require significant future work.

- **Data collection and construction.** The dataset on which the cost model was derived contains multiple data errors and inconsistencies that reduce the robustness of any analysis. Furthermore, the adjustments for price-level differences and operating environment that Sumicsid has made to the data are insufficient to capture the heterogeneity across TSOs.
- **Model development.** Sumicsid's sole reliance on asset data as cost drivers rather than outputs, its limited statistical analysis and non-validation of outputs and modelling assumptions, and the sensitivity of its final model to small changes in the modelling assumptions indicate that Sumicsid's final model is not validated.
- **Application and validation.** The assumptions that fed into Sumicsid's application of DEA are not well-justified. Its outlier procedure is not consistent with academic practice and not justified for the current context. Furthermore, Sumicsid's approach to validating its model through second stage analysis is neither theoretically valid nor sufficient,
- **Transparency.** Sumicsid's level of transparency falls significantly short of what would be considered good practice.

Clearly, these issues cannot be corrected through simple refinements to the current model. We provide a number of recommendations in this report to improve the analysis for future iterations of the study, including the following.

- Provide a clear conceptual (and, where possible, empirical) justification for any assumptions that feed into each stage of the benchmarking process.
- Relatedly, provide detailed description in the outputs and publish modelling codes (which can be anonymised) to aid in transparency.
- Establish an iterative data-collection procedure that ensures data is reported correctly and consistently across TSOs and validate these.
- Use statistical analysis, such as Monte Carlo simulations, to evaluate the impact of any potential data errors. This could then be used to adjust the estimated efficiency scores for setting cost allowances. Alternative evidence, such as SFA modelling, could also inform the extent of the adjustment.
- Robustly capture the impact of all input price differences on expenditure to avoid conflating efficiency and this exogenous factor.
- Perform a scientifically valid model-development process that: (i) is based on realistic modelling assumptions; (ii) tests the significance of alternative model specification; (iii) tests the sensitivity of the analysis to small changes in the sample: and (iv) avoids the arbitrary restriction of cost drivers to asset-based outputs.
- Relatedly, the analysis should not be too sensitive to the year in which efficiency is assessed. If the estimated efficiency of TSOs fluctuates significantly from year to year, the causes of this must be explored.
- If asset-based outputs are used, these must be validated through comparisons to pure outputs.
- Provide statistical evidence to support its modelling assumptions. In particular, its returns-to-scale assumptions must be justified.
- Develop a robust outlier-detection procedure based on academic and scientific best practice. This need not include exact tests recommended in this study (i.e. the bootstrap based dominance test and the iterative super-efficiency test); however, any assumptions that feed into the outlier tests should be clearly explained and supported.
- Analyse the outputs of a DEA model, such as cost driver weights, peers and lambdas, to ensure they are consistent with operational intuition.
- Avoid relying on second-stage validation to detect omitted cost drivers. In a DEA context, the impact of omitted cost drivers should be assessed by testing the sensitivity of the results to the inclusion of alternative cost drivers.
- Cross-check the analysis with alternative benchmarking methods, such as SFA, to validate whether the estimated efficiency scores can be attributed to genuine differences in efficiency or data uncertainty.
- Estimate frontier shift. Not only is this an essential parameter in setting cost allowances, but it can also help to identify flaws with the model that are not evident from cross-sectional analysis.

By incorporating the recommendations presented in this report, CEER and Sumicsid (or any future consultant) will be better able to develop a robust model (or set of models) for cost benchmarking. In this regard, it can also be helpful to consider debriefs involving all the parties on process and methodology to help future studies.

A1 Sample R script

A1.1 Description

The script below performs Sumicsid's second stage analysis on its model. It is written in the programme R,[162] and uses data on cost and the cost drivers as calculated by Sumicsid, as well as the estimated efficiencies of TSOs calculated on a reduced model, leaving one cost driver out at a time. (See Table 5.2 of the report and accompanying text for a description of our review on this issue.)

This script shows, by way of an example, that it is possible to publish modelling codes to enable better transparency regarding the approach followed by the consultant, while preserving the confidentiality of the TSO data.

[162] R is free statistical programming software and is available at https://www.r-project.org/.

A1.2 Sample R script

```
#This Script uses the efficiency scores calculated using Sumicsid's procedure

#The file imported as 'Effscores' here contains the estimated efficiencies for
16 TSOs
#In line with Sumicsid we have excluded one TSO ex ante
#The efficiencies were calculated using only two of Sumicsid's three cost
drivers.
#Specifically the columns contain the following:
# company - names of the TSOs
# noNG    - Estimated efficiencies with Transformer Power and Weighted Lines as
#           cost drivers (excluding NormGrid)
# noTP    - Estimated efficiencies with Adjusted NormGrid and Weighted Lines as
#           cost drivers (excluding Transformer Power)
# noWL    - Estimated efficiencies with Transformer Power and Adjusted NormGrid
#           as cost drivers (excluding Weighted Lines)
# noEnv   - Estimated efficiencies with NormGrid, Transformer Power and Weighted
#           Lines as cost drivers (excluding the environmental adjustment)

Effscores<-read.csv("secondstagevalidation.csv")

#To perform the second stage validation we also import the cost driver data
Data<-read.csv("collated.csv")

#As DEA efficiencies are estimated on 2017 data the second stage uses only 2017
data
Data<-Data[Data$year==2017,]

#For the analysis we merge the two datasets
Data<-merge(Data,Effscores,by="company")

#We run the second stage analysis. The cost driver which was
#ommitted in the estimation of efficiencies is regressed on the estimated
efficiencies.
noNG<-lm(noNG~yNG_yArea,data=Data)
noTP<-lm(noTP~TrafoPower,data=Data)
noWL<-lm(noWL~yLines_steel_angle,data=Data)
noEnv<-lm(noEnv~weight,data=Data)

#Looking at the result of the second stage analysis for NormGrid
summary(noNG)$coefficients

#Output (Table 5.2 in the report)
#               Estimate      Std. Error     t value     Pr(>|t|)
# (Intercept)  8.628687e-01 6.163569e-02 13.9994980 1.264021e-09
# yNG_yArea   -1.084348e-10 1.505139e-10 -0.7204307 4.831162e-01

#Looking at the result of the second stage analysis for Transformer Power
summary(noTP)$coefficients

#Output (Table 5.2 in the report)
#               Estimate      Std. Error     t value     Pr(>|t|)
# (Intercept)  9.339792e-01 4.683268e-02 19.942895 1.117836e-11
# TrafoPower  -1.959299e-06 6.227405e-07 -3.146252 7.144662e-03

#Looking at the result of the second stage analysis for Weighted Lines
summary(noWL)$coefficients

#Output (Table 5.2 in the report)
#                        Estimate      Std. Error     t value     Pr(>|t|)
# (Intercept)          8.407058e-01 6.114351e-02 13.749713 1.600502e-09
# yLines_steel_angle -2.586587e-05 1.872571e-05 -1.381302 1.888356e-01

#Looking at the result of the second stage analysis for the environmental
adjustment
summary(noEnv)$coefficients

#Output (Table 5.2 in the report)
#               Estimate   Std. Error  t value     Pr(>|t|)
#(Intercept) 0.81346405  0.4792434 1.6973924 0.1117308
#weight      0.06014338  0.3984780 0.1509328 0.8821819

#For two out of three cost drivers, Normgrid and Weighted Lines, the omitted
cost driver was statistically insigificant (p>0.1) and would not be identified
as an omitted variable in Sumicsid’s second-stage analysis.
```

www.oxera.com

compelling economics