TCB18 – PROJ 37801

Response to the Oxera Report on TCB18 ETSO

Per J. AGRELL
Peter BOGETOFT

2020-10-30 ver V1.0

Disclaimer

This is a specific report by the authors, professors Per AGRELL and Peter BOGETOFT for the SUMICSID GROUP.

The methods, assumptions, conclusions and recommendations in this report represent only the viewpoint of the authors that assume all responsibility for any errors of fact or logic.

Title: Response to the Oxera Report on TCB18 ETSO

Final report, open. Project: TCB18/37801
Release date: 2020-10-30

Sumicsid Group
Rue Maurice Lietart 56
B-1150 Brussels, BELGIUM
www.sumicsid.com

Table of contents

1. Overview

1.1 Context

1.01 This note is a comment to the Oxera report "A critical assessment of TCB18 electricity", called Oxera (2020) below, released 30/04/2020 on behalf of the TCB18 TSOs participating in the electricity benchmarking.

1.02 The format of the note is brief as most documentation is provided in the following documents, released during the project:

1) Sumicsid (2019) *Norm Grid Development, Technical* Report V1.3, 2019-02-27.
2) Sumicsid and CEER (2019a) *Pan-European cost-efficiency benchmark for electricity transmission system operators*, Main report V1.2, 2019-07-12.
3) Sumicsid and CEER (2019b) *Project TCB18 Individual Benchmarking Report*, V1.0, 2019-07-25. (Released by several TSO).
4) Sumicsid and CEER (2020) *Dynamic efficiency and productivity changes for electricity transmission system operators*, Main report V1.2, 2020-09-11.

1.03 The outline of the response restates the main arguments of Oxera (2020) in an orange shaded paragraph. In some cases, the original statements have been summarized and reformulated without intention of changing the contents and bearing of the argument.

1.04 The response provides an open discussion in a normal paragraph, concluding in a shaded grey paragraph as to our assessment of the impact of the argument on the viability of the TCB18 benchmarking results.

1.2 Outline

1.05 In Chapter 2 we recall the principles of the study and the methodological choices made in it, as well as the differences in focus in Oxera (2020) and the TCB18 project.

1.06 In Chapter 3 we respond in more detail on the main critique raised by chapter in Oxera (2020).

2. TCB18 Model Conception

2.1 Scope and purpose of the Oxera work

2.01 Oxera (2020) is to be seen as a compilation of separate sensitivity analyses, applied to each element in a benchmarking study such as TCB18. The separate comments illustrate, at best, the range and frequency of impact for changes to assumptions, parameters and data in the study. However, since the development of an alternative model (or process) are not in scope of Oxera (2020), the sections cannot be compiled to a common assessment of the model quality of TCB18. E.g., in the critique of the model specification to be composed of asset-based output parameters, no comment is made to the choice of a deterministic model (DEA) and its consequences. Elsewhere, the method choice is criticized for being an unconstrained DEA, rather than a parametric method (OLS or SFA) although this would assume equally strong assumptions of the distribution of errors or inefficiency. Thus, to clarify the fundamental choices in the TCB18 study, we here revisit the data processing, the method choice and the model specification.

2.2 Data processing and validation

2.02 Data quality is primordial for benchmarking and particular attention has been given to the design of an optimal data collection and validation system.

2.03 The principles for the data collection are to ensure full understanding of the data protocol by all project participants. In TCB18 this was implemented by separate releases of the data specifications and guides in December 2017 with several rounds of reviews and two project workshops, leading to a final release in March 2018. Specific templates in Excel were developed and also revised. The project participants had ample of time and opportunity to ask questions about the data definitions, both at the interactive workshop and on open and closed areas of the project platform. Choices of principal nature, such as the activity decomposition and the scope of the benchmarking were discussed and decided jointly with the NRAs in the CEER project steering group (PSG). It can therefore be asserted that the data protocol is well known by the project participants.

2.04 The obligation to comply with any data collection procedure for a TSO is ultimately defined and enforced by the corresponding NRA. It was therefore an important principle to pass the data collection and primary data validation through the NRA, thereby inciting commitment and awareness of the TSO operations and concerns. All data exchanges, both submissions, requests for clarifications and releases of processed data, passed over the NRA to ensure full compliance.

2.05 The primary data validation was performed by the NRAs using a specific data auditing protocol, requiring the NRA to explicitly endorse the quality of the data at submission.

2.06 The role of the consultants in the data validation was to assist in the cross validation, since some TSOs did not allow other NRAs to access their data. The consultants preformed data validation of both technical and economic data in addition to the checks performed by the NRAs. The results for the data validations, frequently resulting in questions and comments, were uploaded to the project platform. Thus, each party in the project brings a specific skill to the task, improving overall quality, independency and consistency.

2.3 Choice of method

2.07 The choice of the benchmarking method (DEA) already hints at the type of model and functional form that will be privileged: the most intuitive and natural deterministic form that explains the current data and that is consistent with existing knowledge about electricity transmission cost causality.

Functional form

2.08 A strong advantage with DEA is the absence of a priori assumptions on the functional form. The piece-wise linear non-parametric structure of DEA can be shown to be the minimal covering hull for the data set, merely using convexity (in standard models) as assumption. The properties of this functional form are strong and intuitively attractive in that they allow for different relative costs between different outputs and for different levels of scale (under e.g. NDRS).

2.09 Other functional forms are possible and frequently used in scientific studies where the aggregate or sector effects are sought, not the individual efficiency scores. Sumicsid and its collaborations have used and published such models, e.g. in Agrell and Brea (2017), for translog models for electricity distribution. However, the alternative functional forms violate the condition of a priori assumptions and/or introduce additional parameters to estimate, which may be infeasible with given data material. It is therefore important to see these more specific functional forms as answers to a series of questions of fit, only introducing a more complex form if a simpler fails to deliver satisfactory performance.

2.10 Oxera (2020) makes a number of statements regarding the loglinear form, without showing that it has a better fit than the chosen level form. The relevant question for a loglinear form is high heteroskedasticity and regular nonlinearity in the residual. The loglinear solved these two problems for many applications, but other solutions also exist (e.g. normed linear form, Agrell-Bogetoft ((2006). It is important to note that the seemingly attractive form brings a significant conceptual problem in the production space. As we are estimating a cost function, the production possibility set covered by the function is not convex. In other words, the iso-cost curves (which are straight lines in the log-log space) are such that a linear combination of two points on this curve is outside the production possibility set. Possibly due to this conceptual difference, the results have fairly low correlation with DEA efficiency scores.

2.11 The conceptual problems related to the loglinear cost function are illustrated in Figure 1 below. The graph on the left presents an imaginary data set (purple points) and a straight (blue) line that corresponds to an iso-cost curve that is produced by fitting a linear model in this type of data. The right hand side presents the same iso-cost curve in the original scale and shows how a DEA frontier fitted in the imaginary data set would look like (purple line). Note that the shape of the loglinear iso-curve (blue) is inverted to the curve formed by the evidence (red line).

2.12 The principle to start with the simplest functional form and test for fit and nonlinearity in the residual using graphs and statistical test is equivalent to Occam's razor: between two models fitting the data equally, the simpler is preferred. In the case of TCB18, the linear level model fits data better than alternative models, it has a straightforward interpretation and a monotonous response (increasing one unit of output is not transformed using any postulated efficiency). Oxera (2020) is not refuting that the level formulation has superior performance, meaning that their critique is principally one of documentation for the performed tests.

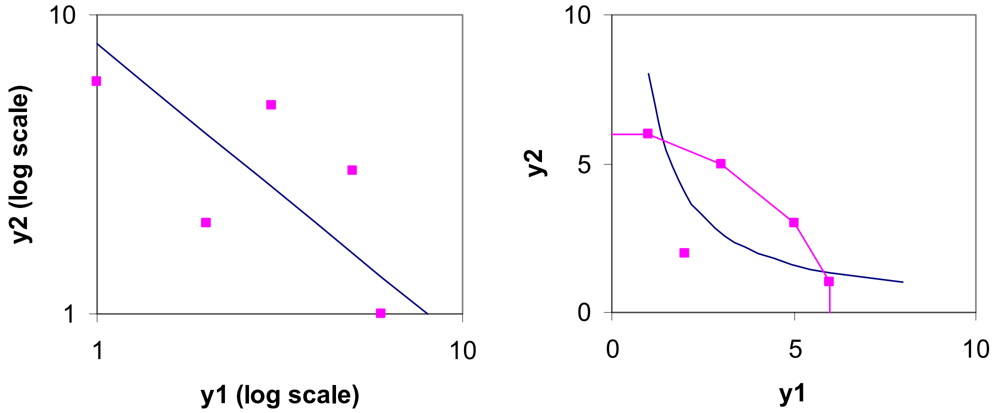


Figure 1 Production frontier (output) under DEA (purple) and loglinear (blue) specification.

2.4 Model specification

2.13 Cost function modelling in electricity transmission is not a new science, it is well established both in the power engineering and production economic literature. Whereas the reasoning in Oxera (2020) seems to suggest a wide range of different factors of unknown influence to be investigated, the structure of a good cost model can be derived from some simple principles, proven both from engineering practice, reference network analysis and transmission system benchmarking.

2.14 An electricity transmission system operator is a techno-economic system with four main sources of cost:

1) Transport work (direct variable cost for the transport of energy)
2) Capacity provision (fixed and variable cost for the capability to deliver instantaneous power to the system users)
3) Grid provision (fixed and variable cost for the connection of a grid user to the main grid at a given spatial location).
4) Customer service (variable costs for the administration of grid users, prevention of faults, training and information).

2.15 As will be discussed below, the inclusion or exclusion of the categories above are consequences of the choice of model in the study.

Transport work

2.16 In the majority of academic published papers on transmission systems, the output for transport work is a measure of delivered energy (GWh), readily available as public information. The output is intuitively attractive, it seems natural to relate the output to the predominant tariff-basis for most TSOs.

2.17 However, whereas most studies are made in US or for countries with vertically integrated utilities, the situation in the unbundled EU energy market is different. Excluding the cost of the commodity (this pertains to the retailer and/or the generator), the residual cost impact of transport work on the unbundled transmission system operator is related to the losses in the network, proportional to the volume transferred, the voltage and type of connection (line or cable) used. There are several problems to incorporate this factor in the benchmarking used to induce general managerial best practice. First, the volume is a stochastic variable, it depends on external factors (weather, business cycle, social and medical incidents) that may be unpredictable and unequally distributed among the operators. Second, the losses also depend on the timing of use (day, night, temperature), largely exogenous or stochastic. Third, the cost of covering the losses depend on the prices for the commodity at the location and time of contracting, the types of contracts allowed by law and regulation and the aversion of risk acceptable in energy trading for the operator. As already investigated in ECOM+ (2005), these conditions vary across operators for valid reasons and the overall cost for losses (<5% of transported volume) is relatively low (<2% of Totex). For these reasons, transported energy is not included in TCB18 as a relevant cost driver, although frequently present in academic studies.

Grid provision

2.18 The grid provision dimension in infrastructure is fairly evident, just as a road operator offers connections to a set of locations and cities, the transmission system operator connects physical locations with a functioning and effective power system. Without a dimension capturing spatial dimension of the grid provision, the model must integrate some proxy such as service area and/or length of roads, both highly imperfect in explaining the grid structure. The most common metrics here are the total line or route length of network, alternatively the circuit/route length of overhead lines and that of cables. Since all TSOs operate multiple voltage levels and single, double or triple lines, the question poses how to treat the different assets. In simplified studies, no differentiation is made among different assets and 1 km of 420kV overhead line is equivalent to 1 km of 130kV single line. However, many regulators (NO, DK, ES, UK) have hired consultants to establish relative weights to create a more cost-relevant proxy for the grid. The idea is not new, an early example for European transmission is Montero et al. (2001) that developed a weighted 'normalized line measure' across voltages to explain cost differences.

2.19 In short, the line component of NormGrid is exactly this parameter and the cited studies from the different countries have been used by the power engineers in TCB18 to create a good measure with the relevant features.

2.20 The logic for other grid assets follows that for lines: a more complex task requires compensating equipment, serial compensation, control centers et c. These assets are also assigned relative weights in the same manner, reflecting their costs for investment and operation.

Capacity provision

2.21 The second most important cost factor in transmission is related to the primarily fixed costs invested in transformers to efficiently transmit electricity over large distances and deliver to underlying grids. The factor has two dimensions, the assets and their effective use. Since they are interrelated, there is normally a choice to make among the two depending on the focus of the study.

2.22 A metric related to the assets, i.e. the transformer capacity, is easily obtained and measured in MVA. Combined with the grid provision metrics above, it already explains a large part of the total cost if used in a cost function.

2.23 An alternative exogenous output is to use peakload (MW), usually across the service area. As the peakload is bounded by the transformer capacity, the two are intimately related. However, peakload is a stochastic utilization metric that is influenced by demographic changes, weather, business cycles etc. Inclusions of this parameter would require the application of a stochastic model, but foremost it would load the risk of underutilization on the TSOs. The latter is not without risk in a regulatory regime as the cost of undercapacity is not well defined or even negative in some systems. Thus it is not in the interest of CEER to penalize the (temporary) utilization of the power system when the future is highly uncertain and the more serious challenge is investment incentives.

2.24 When including the assets in the grid proxy (NormGrid) the construction of the metric mimics the Totex in that older assets disappear from the capex as they are fully depreciated. However, just as peakload is the physical measurement of the maximum use, irrespective on which assets are used, the total transformer power (MVA) above is a metric that complements the NormGrid component, particularly in the operating cost dimension.

2.25 For this reason, the capacity provision is included in two ways, both as part of the grid proxy NormGrid, and its effective capacity as a separate output. Note that an equivalent double inclusion (peakload in MW and substation capacity in MVA) is made in e.g. Llorca et al. (2016) for a stochastic frontier analysis model.

Service provision

2.26 The service provision is often measured as the number of connection points or customers when dealing with distribution system operators. However, for electricity TSOs the number of customers is usually very limited and most countries have delegated invoicing to lower levels. Further, given the very high explanatory power for models including the grid and capacity provision, the service dimension is usually considered included in the two former categories.

Environmental factors

2.27 Remains to control for relevant environmental factors that for a given level of grid and capacity output create higher total expenditure (investments and operating cost). As already developed and documented in E3GRID (2012), the routing complexity can be captured through the information related to the tower structure. As a TSO is facing challenges in urban sprawl, infrastructure or topology, it cannot construct straight lines using cable-stayed suspension towers in wood, although this would give the lowest cost per circuit length. We here developed a multiplicative output parameter for the circuit length weighted with the share of angular towers and the share of steel towers. The parameter successfully separates the sample in TSOs with more complex routing from more transport-oriented TSOs with long straight lines with lower investment and maintenance costs.

2.28 An additional level of environmental correction was achieved by independently letting the power engineers derive and list the complexity factors increasing costs from a technical perspective. Note the methodological difference between doing this step prior to the specification of the model compared to a 'data mining' approach where various factors with unknown effects are used indiscriminately as regressors in a cost function. The risk with the latter approach is to find a set of factors that may fit a particular data set by second-order or spurious correlation, but without any techno-economic rationale. Another approach, advocated by Oxera, would be to let the environmental factors absorb any variability in cost at the average cost function stage. For example, if particularly inefficient operators have more railroad per surface area, a regression would suggest that the cost should be attributed to the railroads rather than to inefficiency. The staged method in TCB18 avoids this problem since the magnitudes of the impact is estimated a priori without inefficiency.

3. Oxera critique

3.1 Data collection and validation [Oxera, ch 3]

3.01 Oxera (2020, Chapter 3) makes the argument that the TCB18 data collection and construction process does not enable a sufficiently harmonized dataset to undertake robust cost benchmarking.

Data errors and validation [Oxera, section 3.1]

3.02 Data errors coming from (i) misreporting, (ii) miscommunication and (iii) measurement errors occur in the sample.

3.03 The methods used in TCB18 do not account for error.

3.04 As a development of earlier projects (ECOM+, e3GRID), TCB18 has reinforced the data collection and the data validation in several aspects.

3.05 The project has defined data collection standards, written guides and templates, for consultation with all project participants. The data collected by the TSO passed through several rounds of NRA validation before submission to the consultant, performing a cross validation of both economic and technical data.

3.06 The data validation involved several rounds of written requests for clarification and control of asset reporting, outputs, investment and cost data. Each TSO received also data sheets with the processed data in each step, including also intermediate data. The following 12 reporting stages assured that each participant (TSO-NRA) had all relevant information about the data reported, processed and used in the benchmarking.

1) Data cross validation by consultant Q3/Q4-2018
2) Data revalidation by NRA Q3/Q4-2018
3) Release I Financial data Q3-2018
4) Release I Assets Q3-2018
5) Release II Assets Q1-2019
6) Release II Financial data Q1-2019
7) Release III Assets Q1-2019
8) Release III Financial data Q1-2019
9) Release IV Individual report Q3-2019 (54 pages)
10) Release V Rundata: normgrid Q3-2019
11) Release V Rundata: totex/opex/capex Q3-2019
12) Release V Rundata: outputs Q3-2019

3.07 A number of data errors were detected and corrected in the intensive data processing above. Indeed, for each round NRA, TSO and consultants collaborated to review the processed data and the validation protocols to explain any deviations.

3.08 Oxera (2020) mentions three errors as examples of data errors and alludes to "extensive" errors and "some TSOs [that] flagged data inaccuracies to Sumicsid over the course of the study, but [without obtaining] correction in Sumicsid's final analysis" without giving any details or substantiation.

3.09 After the final run, we obtained information from two TSOs for various deviations:

1) A TSO reported no angular towers in the data. It received a clarification request for towers in Step 1 of Art. 3.06, the TSO updated the asset reporting (including towers) 08/03/2019, received the subsequent releases without any comment prior to the post-run session in September 2019. The TSO was classified as inefficient and has therefore no impact on the frontier.

2) A TSO likely overreported circuit ends which was spotted in the cross validation of Step 1 of Art. 3.06. The matter was investigated by the NRA and the project steering group. The operator fell out as outlier from the DEA model and therefore the reporting has no impact on the frontier.

3.10 Oxera (2020) mentions also some misreporting (material choice for some towers, prior inflation correction of some investment data). In spite of detailed reporting guidelines, specific reporting templates with controls and four steps of data releases pre-run, these TSOs did not signal the errors. Note, however, that the data material includes 417,719 towers. It is therefore highly unlikely that this error would impact the final score for any operator.

3.11 To illustrate the effects of errors, Oxera (2020) uses a Monte-Carlo simulation where each input and output is subject to a 10% uniform noise term, unclear whether this actually gives a range of 20% in estimates (-10% to +10%) or whether a 5% range has been used. (The reference to ORR (2013) is a 2.5% error, i.e. 5% range.). Clearly, this simulation is exaggerated and cannot be justified with the claims for particular data errors. As will be discussed, the model specification for a deterministic DEA model is based on deterministic and verifiable data, not stochastic variables. The existence of 2-3 minor individual errors in a large dataset cannot be extrapolated in this manner.

3.12 TCB18 has designed and operated a well-functioning data collection and data validation process, based on well prepared data standard protocols, interactive workshops, multiple rounds of validation reports, multiple rounds of data releases before and after the runs, as well as individual and common reports. The existence of some minor individual errors slipping through TSO, NRA and consultants' eyes does not change the conclusion that the final data material is of very high quality.

3.13 A sensitivity analysis for the impact of data errors should be proportionate to the magnitude and frequency of errors expected in the sample. A generalized Monte-Carlo simulation with an exaggerated standard deviation simply illustrates the capacity of the benchmarking method to absorb pure noise, which of course is lower for a deterministic method. However, this ignores the variable selection and data validation procedures for deterministic variable.

Choice of TOTEX [Oxera, section 3.2]

3.14 TCB18 uses TOTEX, this is only acceptable if OPEX and CAPEX are equivalent at the margin and the ratio is controllable.

3.15 The choice of choosing Total expenditure (TOTEX) as the dependent variable or input in the benchmarking is the correct choice both theoretically and practically. TOTEX is used in the electricity regulation in a number of European countries, such as Austria, Germany, Lithuania, Netherlands, Norway, Portugal, and Sweden, in benchmarking, TFP-estimations and as basis for the revenue-cap calculations, see CEER (2017). The opinion that TOTEX is a sound basis is also shared by other stakeholders, customers and operators as reported in CEER (2018). Contrary to the argumentation in Oxera (2020), benchmarking limited to e.g. OPEX would be extremely sensitive to the exact ratio (OPEX/CAPEX) that Oxera (2020) considers as partially non-controllable. By changing from leasing to direct investment, a TSO could show radical improvements in partial OPEX efficiency, but potentially without any positive impact on overall efficiency. Oxera (2020) provides an example for a TSO leasing its grid, we agree and provide it as an example of the appropriateness of the method.

3.16 Oxera (2020) illustrates this aspect with a simulation using OPEX and CAPEX as separate inputs in a DEA model. Naturally, this implies widely different valuations of CAPEX and OPEX, depending on the dual weights. Oxera (2020) passingly alludes to this problem by suggesting a convex combination of the cost elements based on "expert judgement". We find the overall approach without any merit and the suggestion arbitrary (or even redundant, since DEA sets the TSO-optimal weights).

3.17 The only substantive argument in this section concerns the cost normalization differences for elements that could potentially appear in either OPEX or CAPEX. Contrary to what is stated in Oxera (2020, Table 3.1), we confirm that the time period and the inflation indeed are adjusted year by year when a pooled model or analysis is made. However, the personnel cost in OPEX is normalized using the PLICI index, whereas the CAPEX is only inflation-adjusted. The reason for this is the lack of verifiable information concerning the labor element in the investments, origin and composition. To explore the sensitivity with respect to this factor, a sensitivity analysis is included in Sumicsid and CEER (2019a, art 5.26). The labor part is assumed to be between 0% and 25% of the overall investment amount. The relative difference is shown to be minimal (<1%) for the mean score and individually in the range between (-9% to +3%). Additional analyses in Agrell and Bogetoft (2020) confirm this result, but here also including a change of index to the general LCIS index.

3.18 The critique against the choice to TOTEX lacks substance, both in theory and practice. The simulation presented lacks relevance and is inconsistent with the premises of the stated argument since it makes OPEX and CAPEX fully independent inputs, which is non-sensical.

3.19 TOTEX is the only robust input for regulatory benchmarking, since it makes the financial and operational solutions irrelevant. TCB18 has fully explored the sensitivity with respect to labor-cost corrections in CAPEX, finally not made for the general run due to lack of verifiable data.

3.20 The partial efficiencies on OPEX at a given level of CAPEX and for CAPEX at a given level of OPEX, are presented and made available to all project participants as part of Sumicsid and CEER (2019b).

Indexation of OPEX and CAPEX [Oxera, section 3.3]

3.21 Oxera (2020) considers that the price-level differences are incorrectly adjusted for. PLICI does not consider other production factors beyond civil engineering.

3.22 There is no correction for price-levels besides direct manpower cost.

3.23 TCB18 assumes open markets for all services and goods.

3.24 Objective differences may exist due to transport costs across Europe. Investments are governed by local regulation Investments over time have had different conditions

3.25 The choice of index for input-price adjustments has a methodological and an empirical side.

3.26 Methodologically, the correction for local (potentially operator-specific) input prices is the correct approach when the said prices are exogenous and well-identified. An operator required to buy land for its assets in a specific location cannot be responsible for the overall expenditure since the location is forced by the nature of the service. In the same manner, the permanent staff of a transmission system operator must be recruited and hired in accordance with national employment conditions. On the other hand, services such as invoicing, repairs, or communication could potentially be subcontracted or outsourced to service providers in the same or in neighboring states, employing part or all of the labor force under other conditions. Likewise, whereas the land and legal cost of right-of-way are intrinsically local, the value of the equipment itself and its installation are less bound to the national price-level. Frequently, transmission system operators are the only eligible buyers of certain equipment and services in their respective countries, which means that they hardly can rely upon local suppliers to provide for their needs. An erroneous correction of input prices, such as assuming that a TSO in a low-labor cost area can also acquire e.g. transformer cores less expensively than in a high-labor cost area, will artificially skew the benchmarked OPEX negatively for the operator, irrespective of the observed cost.

3.27 Empirically, the question at hand is whether the basis for the input price correction can be well identified or even exists. Ideally, we would desire an exogenous index for the price development for all services required by a TSO and that for each country. Naturally, such index cannot be produced due to endogeneity in most countries and also the task variation across TSOs and over time. The second alternative is then to find well-defined exogenous indexes for the services for which correction is desired. Provided that such indexes exist over sufficient time and for all countries involved in the benchmarking, the operation also requires verifiable data separated over all such indexed services. E.g., an index for civil engineering involves a certain share of administrative IT-services, for which an alternative index exists, as well as construction, maintenance, auditing, et c. In reality, the choice is better guided towards a robust and well-defined basis and the closest widely-available index.

3.28 Oxera (2020) explicitly mentions potential differences in the investment conditions between different areas of Europe (East vs West, North vs South). To the extent that these differences concern prices for input and labor, the general discussion on the adequacy of general indexes apply (just as for individual countries). The claim that there would be intrinsic or systemic differences beyond the prices has not been substantiated and must be left to the individual NRA as one potential explanatory factor among others.

3.29 In TCB18 the choice has been made to adjust for the local salary differences using the civil engineering index PLICI from EUROSTAT. The index is exogenous, available for all countries and defined for staff-intensive services without much outsourcing within the TSOs.

3.30 Oxera (2020) argues that TCB18 should adjust for all services and for investment goods. We will analyze these suggestions in turns.

Full service-price adjustments

3.31 Oxera (2020) claims that the adjustment of labor cost is insufficient and that not enough evidence is provided to validate the hypotheses behind the methodological choice. It is not clear what type of evidence Oxera would consider necessary or relevant in this case. Assuming that TSO X is shown to buy some services more expensively than TSO Y, is this evidence of varying input prices or inefficient procurement? Detailed evidence of outsourced services in other sectors, would this be representative? The argument is tautological and ignores the purpose of the benchmarking – to provide a stable platform for best-practice performance. Ad hoc adjustments to particular conditions, legacy systems and traditions would invalidate the status of the best-practice peer, as its status might heavily depend on ad hoc assumptions of past or current conditions. In the case of market changes, opening and improved procurement, the benchmarking would no longer converge to the optimal best-practice cost, but to an arbitrary state trying to explain the past. The benchmarking in itself is not the regulatory ruling, it is the NRA using the information for reviewing the performance of the TSO that would take into consideration specific suboptimal conditions that explain its occurrence.

3.32 Thus, in TCB18 and as before, we turned the question around and invited the TSOs in the operator-specific data collection in TCB18 to provide evidence of operator- or country-specific regulations or conditions that would be lasting, material and exogenous.

3.33 The correction of input prices by general price indexes is not harmless. Without observing the origin and controllability of the expenditure, it may lead to undue protection of inefficient procurement in high-cost areas and to unfair penalties for procurement in low-cost areas. The technical and economic experts in the team have observed throughout several benchmarking projects examples of services and goods procured internationally for transmission services. It has been judged more stringent in this project to refrain from assumptions regarding the nature of outsourcing (e.g. labor contents).

Adjustments of investment costs

3.34 Oxera (2020) argues for a 100% adjustment of the investment cost using PLICI. It is claimed that this corresponds to regulatory practice, citing the PR13, ORR (2013) study and one specific example for national differences in salary in OFGEM RIIO-ED1. In the case of ORR (2013), the application is actually different: the international data is transformed to GBP (the reference currency) using PPP for a five-year horizon using a TOTEX measure in nominal value, then using the UK inflation conversion. Notwithstanding some sensitivity analyses in several countries (Norway, Germany), we note no utilization of this drastic correction in any prior international benchmarking in energy, such as ECOM+ (2003, 2005), Jamasb et al. (2007), e3GRID (2009, 2012), e2GAS (2013). The claim thus stays with Oxera alone. As above, the approach is economically dubious: a major part of the investment cost in energy infrastructure is composed of materials (steel, copper) and components manufactured by a few global suppliers. Local adjustments would assume that the entire basket of overall investments would be correlated to the labor-intensive civil engineering part, usually corresponding to about 25% of the total investment. Oxera (2020) concludes from their simulation (p.42) that the impact of price-level adjustments can be material, but the simulation illustrates a radically different market for investment goods without any empirical support.

3.35 In TCB18, the impact of the choice of index is illustrated in Agrell-Bogetoft(2020) and the impact of labor-cost differences in the final report (cf. 3.17 above).

3.36 The base year for the adjustment is always 2017, no other alternative is shown.

3.37 The impact of a change of reference year to 2013 for the exchange rate is investigated in a separate sensitivity analysis in Agrell and Bogetoft (2020), showing a minimal impact on mean efficiency (0.5%) but some adjustments for countries with non-EUR currencies. Averaging on an index (as suggested by Oxera) that is to be applied to the same period is not a recommended action, it distorts the correction and may blur efficient and inefficient action since differences are smoothed.

3.38 The choice of index as well as the basis for correction have been extensively discussed during the TCB18 workshops, the choice is guided by access to reliable exogenous data and forward-looking principles for the establishment of best-practice cost, avoiding ad hoc or legacy considerations. The project included a structured approach to address specific circumstances that materially deviated from the assumptions.

Allocation of indirect costs [Oxera, section 3.4]

3.39 Cost allocation for indirect costs is not justified (p.43)

3.40 The causation and treatment of support cost (indirect cost in TCB18) for TSOs is a recurrent question. In some past international benchmarking, all indirect costs have been included (e3GRID, 2013), in others various allocation keys have been used to create a fair comparison.

3.41 In TCB18, calculations were made as suggested in Oxera (2020, p.43) based on no indirect costs, full allocation, individual keys and common allocation keys. The results of these various options were analyzed.

3.42 The basis did include a few elements such as taxes and R&D costs that could be excluded. We confirm the simulation results by Oxera (2020) showing minimal differences (cf. Agrell-Bogetoft (2020)), making the issue minor.

3.43 The allocation policy for indirect cost is primarily a choice of principle, the impact of on the efficiency results is minimal. The project steering group decided to use a partial allocation based also on non-benchmarked activities, as these were considered as relevant and economically beneficial to the core activity.

3.2 Model development [Oxera, ch 4]

3.44 Oxera (2020) generally states that the TCB18 model development appears arbitrarily restrictive and inconsistent with the scientific literature.

Cost driver analysis [Oxera, section 4.1]

3.45 Process is flawed based on (i) results in reports, (ii) restriction to three outputs, (iii) inappropriate functional form.

3.46 OLS and ROLS are used on data with one-sided noise, biased assessment.

3.47 The question of model specification using a regression-based average cost function versus a frontier cost model is a classical question in benchmarking. As Oxera states, the residual can be expected to be skewed if there is substantial inefficiency in the sample. The alternative would therefore be to use a parametric frontier model, such as SFA, to assist in the model specification phase. Applying a parametric frontier model, as will be shown, requires a number of non-trivial technical assumptions (distribution of the error term, structure of inefficiency term, et c.) that in themselves affect the outcome, but also a sufficiently large dataset to perform the assessment. In practice, and this includes all subsequent analyses in Oxera (2020) as well, the model specification for an average cost model should be robust without a frontier model, the opposite would be highly counter intuitive. In Oxera (2020), an application of SFA application without significant separation of noise and inefficiency (p. 74-77) actually serves to confirm the average cost model – the parametric model cannot prove the individual term, effectively turning into a standard OLS on the log-variant of the model.

3.48 Oxera could not reconstruct the line data (p. 47) leading to problems in replicating the dynamic analysis, but calculations for 2017 can be replicated.

3.49 Intrigued by this result, we have undertaken an in depth review of all the data and codes for the panel data 2013-2016. This control did reveal a vector error in the dynamic data for the parameter angular-tower weighted lines, yLines.share_steel_angle_mesum, not adjusting for the difference in the number of TSOs for one year. We have corrected this error and the OLS estimates of Oxera then correspond to ours. A corrected revision of the dynamic report has been produced (revision 1.3).

3.50 Oxera is correct in detecting an error for the pre-2017 line data. The error concerns the parameter yLines.share_steel_angle_mesum for 2013-2016 and has been corrected. The correction has no impact on the static DEA results for 2017.

3.51 The model is not sufficiently documented.

3.52 Oxera (2020) argues at several instances that the model(s) are not sufficiently documented. However, given the rich documentation and the necessity to keep the final reports somewhat readable for a layman, some specific information for the model specification was primarily presented and discussed with the project participants during the workshops. E.g., the models developed for validation of the minimal necessary size using Lasso regression were presented at Workshop 3, but these were not essential to the further developments of the model.

3.53 The model specification is sensitive to changes in historical data, to inclusion/exclusion of TSOs, and to inclusion/exclusion or specific years.

3.54 A number of simulations can be made by removing individual or groups of units as well as time periods. This approach is valid and sound when the data material constitutes a random sample drawn from some distribution. The exact constitution of the sample should then not impact the estimates, since the observations are equally likely to occur in the drawing. However, this idea is not valid and relevant if the data constitutes the complete sample, which is the case in TCB18. Here, removing a TSO or a year will have a direct impact on the interpretation of the results. TCB18 is the best practice cost function for exactly the TSOs included for the reference year, nothing else. Robust regression work by temporarily removing (weigh down) large units off-center as not to influence the estimates for the median operator. This type of exclusion is based on the statistical properties of the observation (leverage, i.e. disproportional impact on the estimation of the regression coefficients due to size) and not simply existence. An intuitive comparison can be made with an international championship: removing athletes from top countries will have a major impact on the result, transforming the sense of the event. The opposite is a statistical survey of the athletic capacity of high-school students: here the individual student is merely a random pick and the result should not depend on the drawing.

3.55 The presence of potential outliers in the data is anticipated in the data validation, model development tools and outlier filter for the calculation of the final scores. The outcomes in terms of number of outliers in each step are documented, the individual TSOs are also informed about their classification in their individual reports.

Functional form

3.56 Model has intercept at zero, assuming no fixed costs, but still non-decreasing returns to scale (NDRS) is assumed in TCB18.

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

3.58 Model estimated in levels and not logs is not regulatory practice, it is worse for heteroskedasticity and worse for interpreting coefficients. Heteroskedasticity (F4.1) may lead to biased results for large TSOs. However, no test shows misspecification of the model as estimated in levels, but for logarithms.

3.59 Our results, as those of Oxera (2020) confirm that the performance of the model in level is superior to a logarithmic transformation. The results were clear for different specifications with and without intercept for the average-cost function. The performance of the standard level-formation was good, which was expected given the inclusion of the NormGrid parameter that captures the grid assets. Contrary to the claim of Oxera (2020), the practice of non-parametric benchmarking using levels is very common for regulation. The use of level data lets the DEA formulation determine the exact shape of the production frontier and its response to a change in an input or output at a given point, rather than using the interpolation from a general technical assumption that may not hold at parts of the frontier.

3.60 No flexible forms are tested (quadratics, interactions).

3.61 In particular in regulatory benchmarking, the steps in the model specification should be guided by the minimum influence from technical a priori assumptions by analysts. The very choice of the non-parametric and fully flexible DEA method to let the data define the production frontier and its features is a proof of this. Thus, given the good fit for a natural model both for average cost and as a non-parametric frontier in DEA, we see no reason to develop an entirely different parametric model with a more complicated functional form. First, the intended use is in DEA on a cross-section, which means that the model can likely not be validated on the same sub data sets. Second, using Occam's razor, a more complex formulation should only be chosen if a simpler form shows some limitations that need to be overcome.

3.62 Oxera tries to fit a log-model and discovers a mis-specification. They agree that the linear model is not mis-specified, but claim that it should have been "proven" by some "evidence". Logs are effective against heteroskedasticity, but this is the rationale for using ROLS. Occam's razor should apply for using more complicated forms.

3.63 Oxera (2020) is not correct regarding regulatory practice; at least Austria, Belgium, Germany, Netherlands, Norway, and Sweden use level-models for their regulatory benchmarking. These models are monotonous, do not rely on a priori functional forms and have good isoquants, which are the most frequent arguments for their use.

3.64 Tools to be used in model development are not (sufficiently) documented, like visual inspection and statistical analysis.

3.65 The constraints imposed upon the documentation were to assure that no element in the open report, table or graph, should enable an external party to reconstruct the data or to identify the participating units and their respective scores. As a consequence, part of the model development using graphs and visual detection (as in returns to scale below) could not be documented in the final report but the method used is the same as forwarded by Oxera (2020).

Sensitivity to sample selected [Oxera, section 4.2]

3.66 Oxera suggests running bootstrapping removing one TSO at a time, presenting min-median-max, two parameters have zero in the range.

3.67 Oxera finds that running ROLS for different years changes the estimation, also differences in efficiency among years for the same operator. The choice of base year has a large importance.

3.68 Large changes in efficiency should not occur, it can be a sign of missing cost drivers, data errors, or understated price inflation.

3.69 Contrary to average-cost functions, non-parametric DEA estimations are by construction dependent upon the subset of best-practice performances. For a large data set, the sensitivity with respect to "layers" or "peels" can be explored using bootstrapping or similar techniques. For a smaller data set such as a cross section with 16 TSOs, it is evident that the removal of units will lead to changes in both the parameter estimates and the efficiency scores. However, the model specification in regulatory benchmarking for energy infrastructure is not a data mining exercise driven by a unknown and hypothesized relations, it is foremost based on the construction and compilation of a set of cost-drivers that have immediate causality on the relevant cost metric.

3.70 As discussed elsewhere, the interpretation of regression coefficients for collinear parameters, such as NormGrid and the angular-tower-lines, cannot be separate. We did not use the average-cost estimation to gauge the performance of the units and a marginal interpretation of a negative (or positive) coefficient is erroneous. Contrary to the exposition in Oxera (2020), a non-parametric DEA model may perfectly well contain collinear parameters that have contradictory signs in the regression. This comes from the desire of *completeness* (Agrell and Brea, 2017, p. 347) for the benchmarking specification, i.e. the set of outputs should fully contain the activities causing the input consumption in the model. Thus, unless two different outputs are perfectly correlated, the exclusion of any output in a non-parametric cannot be made on purely statistical arguments. The science and art of combining such outputs to a tractable set corresponding to the available data are the essence of regulatory infrastructure benchmarking.

3.71 The hypothesis alluded to by Oxera that efficiency is a stationary (time-invariant) feature only linked to the operator has no support from the data or any other argument. A mere observation of the annual accounts for the organizations in a diverse sector like energy transmission would reveal that there are variations across time and across operators.

3.72 As discussed above, the set of TSOs and years are not random from an infinite pool, the study is composed exactly by the participants. The fact that the regression estimates change when operators are removed is natural and expected, in particular for the small reference set. The choice of parameters is not only based on a particular subset of data but an overall techno-economic analysis, including inferences to other and earlier models of electricity transmission systems. Frontier estimates in DEA depend on the best-practice observations, which are by definition a subset of the overall dataset. It is therefore also expected and sound that the estimates change when peers are removed from the reference set, just as the relative ranking of athletes depend on number and quality of the top competitors at an event.

Output choice [Oxera, section 4.3]

3.73 Key drivers should be 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.

3.74 TCB18 only uses asset-based variables as cost-drivers.

3.75 NormGrid is creating disincentives for non-asset solutions.

3.76 Transformers and lines have unknown correlation with other capacity and routing complexity parameters, no justification is made for their choice.

3.77 Significant omission that there is no comparison of output parameters. The correlations between asset capacity and output (peak) are missing.

3.78 The Oxera (2020) simulation replacing transformer capacity with circuit end power is not substantiated by any techno-economic or even statistical argument, the circuit ends do not correspond to the same capacity function as the transformers and are not used in any benchmarking to our knowledge. Transformer power is uniformly understood and used, e.g. by ENTSO-E and for peakload comparisons, whereas the technical configuration of circuit ends depends on country and the layout of the substation. It is also not clear whether the simulation by Oxera (2020) corrects for jointly owned installations.

3.79 As discussed above, the choice of a deterministic DEA model in a static perspective is consistent with the commitment of the NRAs not to expose the TSOs to volume risk. Oxera (2020) mentions a number of utilization metrics (energy delivered, peakload) as well as stochastic quality indicators (network availability, energy not supplied, asset health) and some parameters related to renewable power (amount of RES power). The first category of parameters is certainly interesting from a dynamic perspective, but inconsistent for use in a deterministic static application of DEA. An operator exposed to milder climate, specific business cycles or exogenous events related to assets would appear as overcapitalized and inefficient. Likewise, a TSO anticipating grid growth by asset construction not fully utilized (to avoid blackouts) would be ranked as inefficient compared to capacity-lagging TSOs that may be hampering e.g. the energy transition. A similar argument applies to stochastic quality parameters, to be evaluated in a longer time series and with particular attention to the type and cause of the delivery interruptions. E.g. including a single year of ENS without controlling for the operating environment would naturally penalize operators with long lines in remote rural areas, although the optimal policy may be different. A specific survey was made in TCB18 for the collection of quality indicators and it was decided by the PSG not to include such measures in the model for 2017. Finally, the RES impact at transmission level has been investigated in other projects (e.g. Sumicsid, 2006) revealing that aggregate measures for the input to underlying levels are not adequate cost drivers for transmission. In the current project, the existing data were not detailed enough to analyze the particular impact of renewables on TSOs cost.

3.80 Thus, considering that the TCB18 is a deterministic exercise and not a stochastic average cost function, no specific data collection was made for the utilization metrics (that often relate to the national energy system and not uniquely to the TSO). Thus, it is natural that the TCB18 model is focused at asset-driven parameters and not parameters related to the market size, utilization or reliability of the services.

3.81 As discussed in Chapter 2, the selection of output parameters for TCB18 follows a principle of covering the major service dimensions (transport, grid and capacity provision and customer service) based on a sequential analysis that was explained during workshops and partially in the final report. It is not an open data-mining exercise to find any parameter, as regulatory benchmarking primarily should be based on a specification that is deterministic, complete, feasible and verifiable. In the particular case of TSO benchmarking, CEER did not wish to pursue utilization-based metrics as they penalize early investments and also demographic/economic demand changes for irreversible investments.

NormGrid construction [Oxera, section 4.4]

3.82 Derivation of NormGrid is unclear, the shares of different assets among the TSOs vary, weights may have an impact on efficiency.

3.83 Comparing OLS weights with the calibration constants, the differences are large.

3.84 Oxera (2020) compares a regression estimate for the various asset categories and derives some differences. However, given that the dependent variable Totex in addition is reflecting the other output dimensions and operating efficiency, a direct estimation of the regression coefficients from the cost data will overfit the function to the NormGrid function. From collinearity across categories, it could also lead to incorrect relative weights among NormGrid-components. The very idea of having an independent relative grid proxy estimate is to avoid endogeneity, making it possible to test the overall fit of the grid proxy.

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

3.86 The observation that the TSOs have different shares of the asset categories is correct and normal. As noted elsewhere in Oxera (2020) the structure of the networks vary somewhat as a function of the TSO-DSO interfaces. The NormGrid is not to be estimated from the TCB18 data, but from engineering expertise (cf. Sumicsid and CEER (2019b), Appendix F). The sensitivity to changes in the relative coefficients is very low.

New model: four outputs (lines, cables, circuit ends, transformers) [Oxera, section 4.4.3]

3.87 Haney and Pollitt (2012) argue that aggregation in DEA contradicts the principle of benchmarking. Oxera (2020) show a simulation with four asset classes as only output, leading to different results.

3.88 Aggregation is inevitable in benchmarking, in particular when there are qualitative differences between different outputs or their parts. Even in the simulation, using the asset class "lines" as an output requires some rule to combine different dimensions of lines, voltages and ages into a single scalar. In Jamasb and Pollitt (2001a), the authors use "network length" as one of the outputs for a DEA model. The output is calculated as the unweighted sum of all network lines and cables, irrespective of double/single line, voltage and location. Of course, equal weighting is a special case of aggregation, assuming that 1 km of 110 kV single line is equally worth to 1 km of 420 kV double line. Obviously, leaving the asset classes as separate outputs leads to absurd dual weights in DEA, a problem that in reality would have to be dealt with using e.g. weight restrictions – i.e. a return to the original expert assessment of the relative weighting.

3.89 An output specification using open dual weights has no merit, aggregation is inevitable in any asset provision benchmarking.

Environmental factors [Oxera, section 4.5]

3.90 The environmental factors are not documented in the final report.

3.91 No correlation between unit cost (UC) and environmental adjustment.

3.92 The two-stage process for estimating the environmental impact is new in TCB18. Without resorting to the TSO data, the engineering teams derived exogenous cost drivers and their relative impact for discussion with the participants. Separately, detailed GIS data were gathered in six relevant environmental categories: landuse, slope, subsoil and top soil texture, gravel, and soil wetness. These aspects where then matched with the engineers' estimates for the exact definitions used in the GIS databased (as compared to the proprietary data from the engineers). The resulting complexity factors were combined with the grid proxy using a multiplicative model.

3.93 The empirical result for unit cost compared to environmental adjustments is partial: the unit cost depends also on other outputs and the level of efficiency. To avoid endogeneity in the estimation in which the environmental factors would serve as coefficients covering for various other cost effects, there was no direct calibration of the TCB18-data towards the environmental complexity factors.

3.94 The environmental complexity factors are independent expert assessments, as used in any techno-economic study. They are designed not to cover other effects from a multi-output and efficiency perspective. Detailed sensitivity analysis in Agrell and Bogetoft (2020) shows a relatively low sensitivity to the choice of environmental factors or their parameters, also confirmed by the analysis in Oxera (2020).

3.3 Model assumptions [Oxera, ch 5]

3.95 Main Oxera point: Lack of justifications for model assumptions. Returns to scale assumption and outlier detection are not scientifically based.

Returns to scale [Oxera, section 5.1]

3.96 Banker F-test and sum of coefficients in log-linear regression are not presented in the report.

3.97 Banker test: Oxera finds variable returns to scale (VRS), not non-decreasing returns to scale (NDRS). Sum of coefficients is less than 1, but not significantly (constant returns to scale; CRS)

3.98 Testing intercept in levels: weak support for decreasing returns to scale (DRS).

3.99 Oxera (2020) reports one unit as efficient under CRS and inefficient under VRS (should be impossible).

3.100 The returns to scale assumption in TCB18 is based on econometric tests, observations from the empirical distribution of efficiency and techno-economic considerations from the model specification.

3.101 The default approach in DEA should be to impose only the minimally necessary assumptions, to avoid a priori influence on the data and the efficiency assessment. In this case, it would correspond to variable returns to scale (VRS), meaning that there could be both increasing and decreasing returns to scale in the data set.

3.102 TCB18 is not the only or first study to deal with electricity power systems and returns to scale. Already in Nerlow (1961) evidence was presented for increasing and constant returns to scale in electricity, Christensen and Greene (1976) confirmed these results using a series of parametric formulations, essentially confirming an L-shaped cost function with increasing returns followed by constant returns for most operators. Yatchew (2000) investigate Canadian electricity distributors and finds increasing returns to scale for smaller utilities. Filippini et al. (2004) show increasing returns to scale in electricity distribution in Slovenia using an SFA model. Filippini and Wild (2001) also find increasing returns to scale in distribution for Swiss electricity distribution using three different parametric models. Dismukes et al. (1998) find strong evidence for increasing returns to scale in electricity transmission in USA using extensive tests on a translog cost function. Llorca et al. (2016) also finds increasing returns to scale for US transmission operators.

3.103 Besides the abundant evidence from other studies for increasing (or to be more exact NDRS), there are also specific regulatory arguments for this choice. Experience and data also from other energy network benchmarking highlight the observation that the concessions for national transmission system operators cannot readily be extended outside of their borders, for technical, regulatory and economic reasons. It may be infeasible to gauge a small operator against a larger with the argument that full scale-efficiency should be obtained. A striking example here is the inclusion of the TSO in Malta in e3GRID, limited even by the surface available for expansion. A strong argument is therefore voiced in favor of increasing (non-decreasing) returns to scale, NDRS.

3.104 Oxera (2020) presents certain results from Banker tests, as well as the coefficients from the SFA and logarithmic average cost functions indicating (non-significant) signs of decreasing returns to scale. However, being non-significant, the only conclusion from these tests is that CRS cannot be rejected.

3.105 First, a simple look at the distribution of efficiency in TCB18 electricity provides concrete evidence for the assumption. A plot shows observations of high or full efficiency across all sizes, from the smallest to the largest operators. Even under NDRS (the score is almost identical for CRS, as also seen in Oxera (2020)), we can observe inefficient and efficient observations across different sizes. It is true that there are only two very large units in the sample, but also among these units there is considerable differences in efficiency.

3.106 VRS as assumption is rejected by all published studies and the previous regulatory benchmarking, but also simple techno-economic evidence: no valid reason can be found to suggest that larger operators would be unable to organize their services and assets as efficiently as smaller units. Oxera (2020) provide no other argument for why this assumption would be valid beyond purely small sample size effects.

3.107 The assumption for NDRS is based on techno-economic arguments linked to the abundant evidence from other studies, previous transmission studies, concession areas and the model specification (NormGrid covers all relevant assets). One cannot use SFA results with non-significant coefficients as arguments for e.g. VRS, with the limited sample size this confounds the efficiency effect (that Oxera does not validate) with the general cost function shape.

Outlier analysis [Oxera, section 5.2]

3.108 A technical section largely based on Oxera work for the industry against BNetzA and the incentive regulation in Germany.

3.109 TCB18 uses outlier detection as prescribed in German law (ARegV), but Oxera has supported industry appeals against ARegV and BNetzA in that respect.

3.110 Scores are not half-normal, but DEA is non-parametric. The Banker test is of value only for large samples.

3.111 Efficiencies of the same unit in both numerator and denominator is not consistent with Banker independent samples.

3.112 Sumicsid has agreed on the objections against the F-test in other legal processes.

3.113 Super efficiency should be iterative (Thanassoulis, 1999)

3.114 Oxera (2020) is partially repeating arguments from an appeal that they were involved in on behalf of operators against BNetzA concerning the implementation of the outlier detection in accordance with ARegV. Oxera argues against dominance, super-efficiency tests and in favor of bootstrapping and/or sequential application of outlier filtering.

3.115 We will not repeat the lengthy arguments in the appeal against BNetzA here (cf. Agrell and Bogetoft, 2019), it suffices to clarify that contrary to the allusions in Oxera (2020), the appeal was rejected and the outlier detection procedure in Germany practiced by BNetzA for both electricity and gas networks at all levels remain the most advanced and best practice in regulatory benchmarking. To our knowledge and not contradicted by Oxera, no NRA uses the suggested bootstrapping procedure.

3.116 Outlier detection is a vast area of academic discussion, see also Agrell and Niknazar (2014) with various models advanced for detecting outliers, defined in various ways. As noted in CEER-Sumicsid (2019), outlier detection for the type of sample used in TCB18 is not merely a mechanical application of the criteria in ARegV, it also includes econometric reviews such as Cook's distance and foremost studies of how individual units appear in different graphs for unit costs and certain partial measures. This holistic approach, combined with the data cross validation, warrants for the greatest possible protection of the replicability of the efficient frontier. The exclusion of one TSO is based on a combination of econometric and techno-economic observations.

3.117 The outlier detection in TCB18 follows best practice for regulatory benchmarking, well beyond studies cited in Oxera (2020) such as the ORP (2013) or others in which mainly ad hoc inspection is used. Outlier detection in a small data set is always a careful multi-tool balance, not a mechanical application.

Model validation process [Oxera, section 5.3]

3.118 Oxera claims generally that the model validation process in TCB18 is incapable of detecting model flaws or omissions. The main discussion here focuses on the DEA weights.

3.119 Dual weights for NormGrid do not provide evidence of NormGrid as the strongest cost driver.

3.120 DEA is based on a multi-dimensional output space where the surfaces are supported by an individual subset of the output parameters. Unless perfectly aligned, the operators will select different dual weights as to put forward their particularity in the comparison. The statement that NormGrid is the strongest parameter is related to the initial Lasso and OLS regressions for a single variable, where invariably NormGrid comes out as the most informative.

3.121 The sequence of added parameters may impact a model specification choice based on OLS significance. In section 5.3 Oxera (2020) shows that NormGrid would not have been selected if two others were already in the model, only transformers.

3.122 The selection of variables in regulatory models is not a datamining exercise where any parameter can replace any other. The techno-economic significance must be taken into account when considering a parameter candidate, not just the predictive ability from a purely statistical viewpoint. E.g., the early Lasso-models tested in the project sometimes had correlated parameters of secondary economic importance (e.g. length of high-ways) that marginally dominated more techno-economically relevant parameters in terms of explanatory power. However, clearly this finding in itself is not a reason to select or to deselect a specific parameter.

3.123 Lambdas are high, mean = 4.1, max =12.

3.124 The multipliers in DEA for a non-decreasing returns to scale formulation are necessarily higher than 1. As discussed elsewhere, it is more realistic to imagine that a large organization should be able to match the collective output of a series of smaller operators than the opposite, i.e. a lambda of 0.25 would mean that a TSO is directly compared to a peer that is four times larger. Restricting the lambdas beyond the returns to scale is not a structured approach and the finding of Oxera (2020) is not surprising.

Omitted cost drivers [Oxera, section 5.4]

3.125 Second-stage analysis not correct, the second-stage parameters are not independent from the first-stage parameters, no correction for serial correlation.

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

3.127 The model specification process includes a structured approach for covering the services of a TSO; grid provision, capacity provision, customer service. An additional stage included systematic incorporation of environmental conditions. The post-run second-stage process is intended to detect potential bias in the scores, not the inclusion of specific parameters.

Use of SFA [Oxera, section 5.5]

3.128 No use of SFA in TCB18.

3.129 Oxera uses SFA on 16 TSO but finds no significant inefficiency, claimed to be proof of a flaw in the model specification.

3.130 NormGrid and Transformers are significant in SFA, lines are not.

3.131 As discussed before, the application of a stochastic frontier model is not relevant in this project, operating a relatively small cross-section of 16 operators. An application to a time-series or a pooled dataset would introduce a number of a priori assumptions that are neither empirically verifiable, nor relevant to solve the question of the incumbent cost efficiency in the reference year.

3.132 Sumicsid is using SFA for cross-validation or direct assessment in projects with a larger number of operators included. However, in TCB18 the limited dataset, the static focus, the DEA method and the particular attention on transmission systems have prompted a purely deterministic model specification. Thus, the cross-validation with a stochastic method applied to a time series cannot be expected to return any directly useful information.

3.133 Oxera (2020) claims that SFA should be relevant for cross-validation, but fails to find any significant gamma term (individual inefficiency), thus the exercise confirms our claim.

3.134 In another project, Oxera (2012b) recognizes that in discussing the data requirements for SFA in regulatory benchmarking concludes that "*SFA models with fewer than 30-40 observations can fail to produce results*" (p.20). The findings in Oxera (2020) are consistent with their initial observation.

3.135 In passing, we notice that Oxera (2020) seems to have some difficulties in the modelling. Although various estimations for time-series and pooled formulations can be found, the relevant cross-sections perform poorly for the stated reasons. For this reason, SFA (along with more advanced flexible forms) is not an adequate cross-validation tool for the efficiencies in TCB18.

3.136 DEA has been chosen by CEER for the TCB18 benchmarking on cross-sectional data for its absence of a priori assumptions on the production function and the structure of the potential inefficiency in the sector. The model specification for TCB18 is based on verifiable deterministic output parameters and a structurally comparable TOTEX. The fact that the data sample is too small for a regular SFA application is irrelevant for TCB18 and the validity of its results.

Frontier shift [Oxera, section 5.6]

3.137 Regress in dynamics means that the model is wrong, it misses changes in regulatory burden over time.

3.138 Regress means that the data is noisy and stochastic.

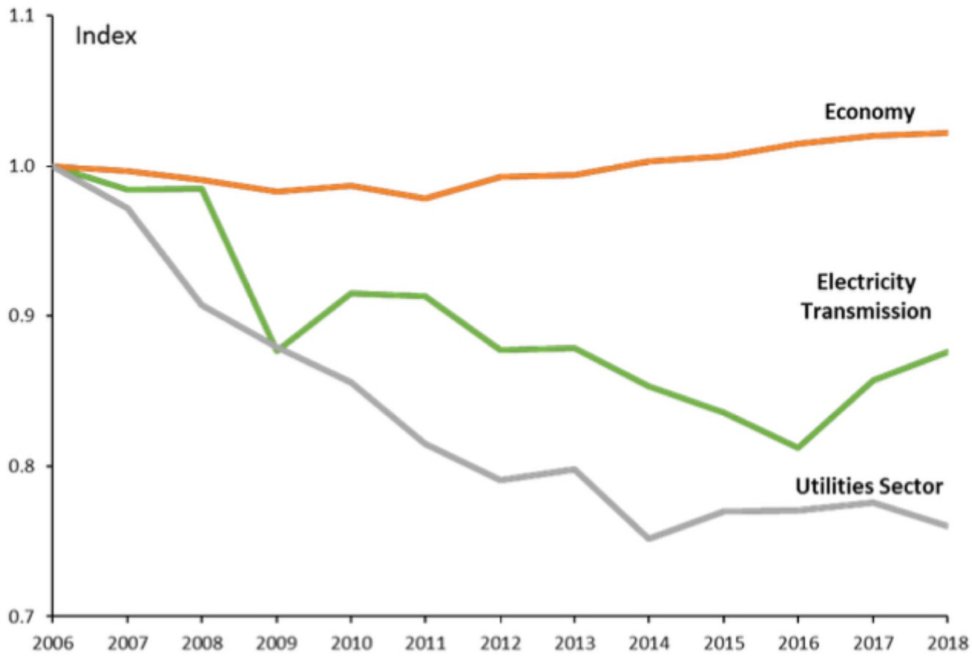
3.139 The initial jump in the Malmquist estimates in Sumicsid-CEER (2020) is a result of the data error for angular-tower lines. When corrected for this, the dynamic productivity estimates are fairly stable as seen in . The results indicate still a productivity regress, but a series of potential reasons for this are discussed in the report. Neither productivity progress, nor regress per se should be directly interpreted as linked to the model structure. The dynamic results are resulting from the relative performance of the firms over time, the data set includes also firms that were classified as outliers in the static calculations.

	Malmquist	Efficiency Change	Technical Change	Number of DMUs
2013 - 2014	0.984	0.975	1.010	16
2014 - 2015	0.964	1.012	0.954	16
2015 - 2016	0.952	0.984	0.969	16
2016 - 2017	0.976	0.978	0.997	16
Averages	0.969	0.987	0.983	

Table 3-1 Dynamic results, Malmquist, TCB18/ELEC, 2013-2017.

3.140 Oxera (2020) voices concerns of model misspecification due to the dynamic regress results, advancing references to how NRAs implement non-negative frontier shifts. However, the application of frontier shifts in general X-factors has no causality on the past productivity progress relative to the general economy. Regress in electricity transmission is not rare, e.g. Llorca et al. (2016) find regress in USA during the period 2001-2009 using an SFA application correcting for environmental effects. Using a TFP approach, AER (2019) reports continuous regression in electricity transmission from about 2009 to 2018, see also Figure 2. The reasons for regress are probably multiple and beyond this note.

Figure 3.1 Electricity transmission industry, utilities sector, and economy productivity indices, 2006–2018



Source: Economic Insights; Australian Bureau of Statistics.

Note: The productivity of the Australian economy and the utility industry is from the ABS indices within 5260.0.55.002 Estimates of Industry Multifactor Productivity, Australia, Table 1: Gross value added based multifactor productivity indexes (a). We have rebased the ABS indices to one in 2006.

Figure 2 Total factor productivity for industry, electricity transmission and utilities, 2006-2018. AER (2019).

3.141 The interpretation of the dynamic scores in Oxera (2020) is speculative and could as well explain any decrease in performance. Other published sources using other methods document equivalent results.

3.142 The process for operator specific conditions in TCB18 was established to let operators present specific static or dynamic cost-increasing conditions that could affect performance. Since these are already addressed in the current dataset, the interpretation of Oxera (2020) cannot be substantiated.

3.4 Oxera summary [Oxera, ch 6]

3.143 The final summary in Oxera (2020, Chapter 6) is more moderate than the subsections. Many of the points in the summary are valid to consider in benchmarking, which is done in TCB18.

3.144 Provide a clear conceptual (and, where possible, empirical) justification for any assumptions that feed into each stage of the benchmarking process.

3.145 Relatedly, provide detailed description in the outputs and publish modelling codes (which can be anonymised) to aid in transparency.

3.146 Establish an iterative data-collection procedure that ensures data is reported correctly and consistently across TSOs and validate these.

3.147 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.

3.148 Robustly capture the impact of all input price differences on expenditure to avoid conflating efficiency and this exogenous factor.

3.149 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.

3.150 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.

3.151 If asset-based outputs are used, these must be validated through comparisons to pure outputs.

3.152 Provide statistical evidence to support its modelling assumptions. In particular, its returns to scale assumptions must be justified.

3.153 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.

3.154 Analyze the outputs of a DEA model, such as cost driver weights, peers and lambdas, to ensure they are consistent with operational intuition.

3.155 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.

3.156 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.

3.157 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.

3.158 The general summary in Oxera (2020) contains a number of elements that Sumicsid can agree on as important for a regulatory benchmarking.

3.159 Overall, Oxera (2020) provides a good sensitivity analysis for certain elements of the study. While e.g. the sensitivity to the units or years included is less relevant in a DEA study, other suggestions illustrate the inevitable tradeoffs in conducting an international infrastructure benchmarking, such as the overhead allocation basis and the choice of reference year or currency.

3.160 CEER-Sumicsid (2019) is one piece of documentation among a large mass of documentation for the project. Although considerable effort has gone into designing informative feedback prior to and after workshops, as well as prior to runs and calculations, this process can always be improved and Oxera (2020) provides some suggestions how to do so. Notwithstanding, in TCB18 the overall amount and level of detail in the information for project process reporting and data validation are the highest observed for an international benchmarking to our knowledge.

3.161 TCB18 is an ambitious study of TSO performance using a CEER-developed data collection and data validation scheme and an interactive model development process with several path-breaking innovations (GIS-environmental factors, capital harmonization methods), ample of time for TSO-NRA consultations of guidelines, data templates and processed data, as well as safeguards for specific conditions. Oxera (2020) discusses ways to improve the process documentation, in particular in the model specification and data validation phase. These points are valid and could lead to an improved process, in the interest of all project participants.

References

AER (2019) *Annual benchmarking report: Electricity transmission network service providers,* November 2019. Australian Energy Regulator.

Agrell, P. J. and P. Bogetoft (2007) *Development of benchmarking models for German electricity and gas distribution*, Final report 2007-01-01, Bundesnetzagentur (BNetzA), SUMICSID AB.

https://www.bundesnetzagentur.de/SharedDocs/Downloads/DE/Sachgebiete/Energie/Unternehmen_Institutionen/Netzentgelte/Gas/Effizienzvergleich%20Verteilernetzbetreiber/GA_Sumicsid_Effizienzvergleich.pdf?__blob=publicationFile&v=3

Agrell, P. J. and P. Bogetoft (2009) *International Benchmarking of Electricity Transmission System Operators*. Project e^{3}GRID. SUMICSID. http://e3grid.sumicsid.com/pub/2009-03-09_e3grid_final_report_open_main.pdf

Agrell, P. J. and P. Bogetoft (2014), International Benchmarking of Electricity Transmission System Operators *Proceedings of European Energy Market Conference EEM14, IEEE Proceedings*, pp. 1-5 doi: 10.1109/EEM.2014.6861311

Agrell, P. J. and P. Bogetoft (2017) *Reponse to Prof. Kumbhakar and Prof. Thanassoulis – Validity of Bundesnetazagentur's cost driver analysis andsecond-stage analysis in its efficiency benchmarking approach*. Report to the Court, 2017-05-15. SUMICSID.

Agrell, P. J. and P. Bogetoft (2019) *Outliers in DEA-based regulatory benchmarking*. Response to Oxera, V1.3, 2019-02-03. SUMICSID.

Agrell, P. J. and P. Bogetoft (2020) *Sensitivity Analysis for TCB18 Results for Statnett*. Final report for NVE, 2020-05-04. SUMICSID. Confidential.

Agrell, P. J., and Brea-Solís, H. (2017). Capturing heterogeneity in electricity distribution operations: A critical review of latent class modelling. *Energy Policy*, 104, 361-372.

Agrell, P. J., P. Bogetoft, C. Riechmann, A. Rodgarkia-Dara, C. Zimmer (2013) *e3GRID 2012 – European TSO Benchmarking Study*. Project e3GRID 2012. SUMICSID, Consentec and Frontier Economics. Report for CEER. https://www.researchgate.net/publication/265592571_E3GRID2012_European_TSO_Benchmarking_Study

Agrell, P.J., P. Bogetoft, and J. Tind (2005), DEA and Dynamic Yardstick Competition in Scandinavian Electricity Distribution, *Journal of Productivity Analysis*, 23, 173–201.

Agrell, P. J., and Niknazar, P. (2014). Structural and behavioral robustness in applied best-practice regulation. *Socio-Economic Planning Sciences*, 48(1), 89-103.

Akkemik, K. A. (2009). Cost function estimates, scale economies and technological progress in the Turkish electricity generation sector. *Energy Policy*, 37(1), 204-213.

Banker, R.D. (1996), Hypothesis Test Using Data Envelopment Analysis, *Journal of Productivity Analysis*, 7, pp. 139-159.

Battese, G.E. and Coelli, T.J. (1992), Frontier Production Functions and the Efficiencies of Indian Farms Using Panel Data from ICRISAT's Village Level Studies, 5(1), 327-348.

Bogetoft, P. (1997), DEA-Based Yardstick Competition: The Optimality of Best Practice Regulation, *Annals of Operations Research*, 73, 277-298.

Bundesnetzagentur (2006), *Bericht der Bundesnetzagentur nach § 112a EnWG zur Einführung der Anreizregulierung nach § 21a EnWG*. 01.07.2006.

Cameron, L. (2001). Transmission investment: Obstacles to a market approach. *The Electricity Journal*, 14(2), 25-38.

CEER (2017) *CEER Report on Investment Conditions in European Countries*. C16-IRB-29-03.

CEER (2018) *Incentives Schemes for Regulating Distribution System Operators, including for innovation*. CEER Conclusions Paper C17-DS-37-05.

Filippini, M., & Wild, J. (2001). Regional differences in electricity distribution costs and their consequences for yardstick regulation of access prices. *Energy Economics*, 23(4), 477-488.

Edvardsen, D. F., & Førsund, F. R. (2003). International benchmarking of electricity distribution utilities. *Resource and Energy Economics*, 25(4), 353-371.

Haney, A. B., & Pollitt, M. G. (2013). International benchmarking of electricity transmission by regulators: A contrast between theory and practice?. *Energy Policy*, 62, 267-281.

Jamasb, T., Newbery, D., Pollitt, M., & Triebs, T. (2007) *International benchmarking and regulation of European gas transmission utilities*. Report prepared for the Council of European Energy Regulators (CEER). http://portal.e-control.at/portal/page/portal/EER_HOME/EER_PUBLICATIONS/CEER_PAPERS/Gas/2006/ERGEG_Cost_benchmark_final_tcm7-107140.pdf

Jamasb, T. and Pollitt, M. (2000), *Benchmarking and Regulation of Electric Distribution and Transmission Utilities: Lessons from International Experience. Workshop on International Benchmarking of Electric Utilities*, Robinson College, Cambridge, December.

Jamasb, T. and Pollitt, M. (2001a), *International Benchmarking and Regulation of European Electricity Distribution Utilities*. Final report for CEER, May 2001.

Jamasb, T. and Pollitt, M. (2001b), *International Benchmarking and Yardstick Regulation: An Application to European Electricity Utilities*. DAE Working Paper 0115, Dept of Applied Economics, University of Cambridge.

Jamasb, T., Pollitt, M., & Triebs, T. (2008) Productivity and efficiency of US gas transmission companies: A European regulatory perspective. *Energy Policy*, 36(9), 3398-3412.

Kittelsen, S. A. C. (1993), *Stepwise DEA; Choosing Variables for Measuring Technical Efficiency in Norwegian Electricity Distribution*, SNF-arbeidsnotat nr. A 55/93, SNF, Oslo.

Kumbhakar, S. C., & Hjalmarsson, L. (1998). Relative performance of public and private ownership under yardstick competition: electricity retail distribution. *European Economic Review*, 42(1), 97-122.

Llorca, M., Orea, L., & Pollitt, M. G. (2016). Efficiency and environmental factors in the US electricity transmission industry. *Energy Economics*, 55, 234-246.

Montero, F. L. P., Pérez-Arriaga, I. J., & Odériz, F. J. R. (2001). Benchmark of electricity transmission tariffs in the countries of the internal electricity market of the European Union. *Utilities Policy*, 10(1), 47-56.

Nerlove, M. (1961). Returns to scale in electricity supply. In *Measurement in Economics – Studies in Mathematical Economics and Econometrics in Memory of Yeshuda Grunfeld*, C.F. Christ (Ed.), Stanford University Press.

Sumicsid (2019) *Norm Grid Development*, Technical Report V1.3, 2019-02-27.

Sumicsid and CEER (2019a) *Pan-European cost-efficiency benchmark for electricity transmission system operators*, Main report V1.2, 2019-07-12.

Sumicsid and CEER (2019b) *Project TCB18 Individual Benchmarking Report* V1.0, 2019-07-25. (Released by several TSO).

Sumicsid and CEER (2020) *Dynamic efficiency and productivity changes for electricity transmission system operators*, Main report V1.2, 2020-09-11.

von Geymueller, P. (2009). Static versus dynamic DEA in electricity regulation: the case of US transmission system operators. *Central European Journal of Operations Research*, 17(4), 397.

Von Hirschhausen, C., Cullmann, A., & Kappeler, A. (2006). Efficiency analysis of German electricity distribution utilities–non-parametric and parametric tests. *Applied Economics*, 38(21), 2553-2566.

Yatchew, A. (2000). Scale economies in electricity distribution: A semiparametric analysis. *Journal of Applied Econometrics*, 15(2), 187-210.

SUMICSID SPRL (BELGIUM)
Rue Maurice Lietart 56
B-1150 Bruxelles, BELGIUM
Tel: +32 26 75 15 23
Fax: +32 27 06 53 58
belgium @ sumicsid.com

SUMICSID APS (DENMARK)
Fru Ingesvej 19
DK-4180 Sorø, DENMARK
Tel +45 57 83 15 18
Fax: +32 27 06 53 58
denmark @ sumicsid.com

http://www.sumicsid.com