ACER-PAEC

AEC Method and Process

FINAL REPORT D04

Per J. AGRELL
Axel GAUTIER
Andreas STRITT
Urs TRINKNER

FINAL

2025-12-11 ver V1.0

Disclaimer

This is a technical report prepared by SUMICSID GROUP and Swiss Economics on behalf of ACER and in collaboration with ACER and national regulatory authorities (NRAs), for guidance in completing the ACER Cost Efficiency Comparison (AEC), starting in 2024 and to be published for the first time by 5 August 2027. However, this report reflects the views of SUMICSID GROUP and Swiss Economics and does not necessarily represent the views of ACER. ACER may adapt, modify, or deviate from the methodology, recommendations, or processes proposed in this report.

The report takes into account the proposals presented at a public workshop in Brussels on 9-10 July 2025, which was complemented by a public consultation between 17 June and 17 July 2025 and a review by an independent group of experts commissioned by ACER. The final revised document has been prepared solely by SUMICSID GROUP and Swiss Economics, who assume full responsibility for any errors of fact or logic. ACER is not liable for any consequence resulting from the use of information contained in this report.

Title: AEC Final Report D04

Final report, open. Project: PAEC
Version: V1.0
Release date: 2025-12-11

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

Table of Contents

Glossary

Term	Definition	Art (§) Section (S) Chapter (C)
ACER	European Union Agency for the Cooperation of Energy Regulators	
AE	Allocative Efficiency (metric)	S 7.7
AEC	ACER Efficiency Comparison	
CAPEX	Capital expenditure (investment cost annuities)	§ 6.04
CE	Cost Efficiency (Totex, metric)	S 7.4
CE(2)	Decomposed Cost Efficiency (Opex and Capex, metric)	S 7.5
ED	Execution-focused Dynamic model	S 8.9
ES	Execution-focused Static (single-year) model	S 8.9
CEA	Cross Efficiency Analysis (method)	§ 8.02
D02–D03	Reports (deliverables) in the project PAEC	§ 1.04–1.09
DEA	Data Envelopment Analysis (method)	§ 8.37
Disc	Discrete variable (e.g., material choices, …)	Appendix E
DSO	Distribution System Operator	§ 6.24
DSOE	Distribution System Operator (Electricity)	§ 8.122
DSOG	Distribution System Operator (Gas)	§ 8.47
E2GAS	CEER gas TSOG benchmarking (2016)	§ 3.06
EC / TC / M	Efficiency Change / Technological Change / Malmquist index	S 7.8
EG	Expert Review of AEC, Report 15/10/2025	§ 0
FERC	U.S. Federal Energy Regulatory Commission	Appendix F
FTE	Full-Time Equivalent staff (measure)	Appendix E
GIS	Geographic Information System (data for asset locations)	Appendix E
H-gas / L-gas	High-calorific / Low-calorific natural gas	Appendix D
HP / MP / LP	High / Medium / Low pressure classes	Appendix D
NCE	National Currency Equivalent (e.g. EUR converted values)	Appendix E
NCU	National Currency Unit (local currency values)	Appendix E
NRA	National Regulatory Authority	§ 1.01
OPEX	Operating expenditure (costs for operation, maintenance,)	§ 6.03
PAEC	Project for ACER Efficiency Comparison (Phase I)	
PCG / PCN	Piece count gross / net (unit counts, un-/weighted)	Appendix D
Q(t)	Quality incentive term in revenue cap (optional)	§ 6.12
RAB	Regulatory Asset Base – book value of regulated assets	§ 9.04
RDEA	Robust DEA (method)	§ 8.02
RTOG	Regional Gas Transmission Operator	§ 11.08

SA	Sensitivity analysis	Table 8-24
SE	Scale Efficiency (scale metric)	S 7.6
SFA	Stochastic Frontier Analysis (method)	§ 8.31
StonED	Stochastic Semi-Nonparametric Envelopment of Data (method)	§ 8.42
T0–T3	Data tiers by disclosure: T0 public → T3 confidential	Appendix D
T01-T03	Work packages and discussion notes in project PAEC	§ 1.04–1.09
TE	Technical Efficiency (metric)	S 7.3
TCB21 / TCB18	CEER benchmarking projects in 2018 and 2022	§ 3.06
Tm	Length of efficiency catch-up period (years)	§ 6.14
TMP	Scope: transport, maintenance, planning	§ 11.07
TOTEX	Total expenditure = OPEX + CAPEX	§ 6.02
TSOG / TSOG	Transmission System Operator (Gas)	§ 1.01
UC	Unit cost (measure, e.g cost/volume)	S 7.2
UD	Utilisation-focused Dynamic model	S 8.9
UoM	Unit of measurement	Appendix D
WACC	Weighted Average Cost of Capital	§ 6.04
WS1 / WS2	PAEC workshops 1 and 2	§ 3.04
Xi	Individual efficiency improvement term	§ 6.12
Xgen	General efficiency improvement requirement	§ 6.12
Z(t)	By-pass factor in revenue cap	§ 6.12

Statistical technical terms

Term	Definition	Context / Reference
Inference metrics		
Std. error	Estimated standard deviation of a coefficientś sampling distribution	Regression output
z-value	Ratio of estimate to its standard error (large-sample normal test)	Inference (z-test)
t-value	Ratio of estimate to its standard error (small-sample t-test)	Inference (t-test)
$\Pr(>\lvert t \rvert)$	Probability that the coefficient represents random noise (p-value)	Significance testing
Model fit and diagnostics		
R^2	Proportion of total variance in dependent variable explained by model	Goodness-of-fit
Adj. R^2	R^2 adjusted for the number of predictors (penalizes overfitting)	Model comparison
Residual std. error	Standard deviation of residuals; unexplained variability	Model diagnostics
F-statistic	Overall test of joint significance of all coefficients	Model significance
Estimation context		
OLS	Ordinary Least Squares; minimizes sum of squared residuals	Estimation method
SFA	Stochastic Frontier Analysis	Estimation method
(OLS) σ	Estimated residual standard deviation (noise variance)	Model precision
(SFA) γ	Share of total variance due to inefficiency component	Stochastic Frontier Analysis
Error and performance measures		
MD	Mean Deviation; average absolute difference between observed and predicted values	Accuracy measure
MAD	Mean Absolute Deviation; mean of absolute residuals	Accuracy measure
MSE	Mean Squared Error; average of squared residuals (model accuracy)	Accuracy measure
MRC	Mean Rank Correlation; mean Pearson correlation of scores	Efficiency
Rank()	Ordering of units or coefficients by efficiency, score, or magnitude	Ranking / evaluation
Estimation methods		
MoM	Method of Moments; matches sample and theoretical moments	Estimation approach
ML	Maximum Likelihood; maximizes the likelihood of observed data	Estimation approach
PL	Pseudo-Likelihood (or Partial Likelihood); approximate likelihood estimation	Simplified ML variant

1. Summary

1.1 Background

1.01 Regulation (EU) 2024/1789 establishes the requirement for ACER to publish a study comparing the efficiency of natural gas transmission system operators (TSOG) which should be published by 5 August 2027 and every four years thereafter.

1.02 The ACER efficiency comparison (AEC) is a benchmark for gas transmission system operations in the European Union. The AEC will assess the relative cost efficiency of all TSOG active in the EU.

1.2 Report

1.03 ACER has commissioned Sumicsid and Swiss Economics (thereafter Sumicsid) to organize an interactive planning and method development for the AEC in collaboration with ACER and national regulatory authorities (NRAs). This report reflects the views of SUMICSID GROUP and Swiss Economics and does not necessarily represent the views of ACER. The document additionally considers input from stakeholders collected in several dedicated exchanges.

1.04 The report provides the recommendation made by Sumicsid for the methodology and process of the AEC, which ACER can use to carry out phases II and III, as well as the iterations following the first publication. ACER may adapt, modify, or deviate from the methodology, recommendations, or processes proposed in this report.

1.3 Process

1.05 ACER has structured the work leading to the first publication of the AEC in three different phases:

1.06 *Phase I* establishes the objectives of the comparison in the context of decarbonisation and to define the methodology and the data approach that will be used in the comparison.

1.07 *Phase II* is intended to collect and validate the TSOG data that will be used in the comparison. This second part of the process will focus on finalizing the data request to be sent to transmission system operators for gas (TSOG), collecting the data and validating the TSOG submissions. The final objective of this phase is to provide data with the level of quality necessary to apply the methods established in phase I.

1.08 *Phase III* is intended to complete the modelling work based on the methodology defined in phase I and the data collected in phase II, performing the calculations, validating and reporting the results.

1.4 Objectives

1.09 The AEC will be used in the context of increasing complexity for TSOG during the energy transition, characterized by declining gas volumes, the integration of renewable gases, and preparations for hydrogen transport. The report assesses the methods that can be used in this context to measure the relative efficiency across TSOG as well as the adaptation to a context of decreasing network utilisation and potential removal of assets from the regulatory asset base for decommissioning or repurposing.

1.10 The AEC establishes a harmonised EU-wide database to measure TSOG performance. The proposed metrics are intended to be robust, relevant and comparable across Member States and should be consistent over time. The process for the calculation of the AEC is based on extensive transparency requirements to allow for an auditable review aimed at strengthening the regulatory enforceability of the result when used in frontier-based incentive regulation for TSOG.

1.11 The metrics to be delivered, the features of the models to calculate these metrics, the data used, the controls for heterogeneity in operating conditions and history, and the means applied to assure robust estimates, are derived from the NRA needs identified (section 6.4). in the short and long run with a view to safeguard both cost efficiency and effectiveness in the energy transition. These requirements support Article 17 of Regulation (EU) 2024/1789, according to with the costs incurred by TSOG *correspond to those of an efficient and structurally comparable network operator*.

1.5 Metrics, methods and models

Metrics

1.12 Cost and performance comparisons can be used to calculate a range of metrics, such as technical efficiency (TE), scale efficiency (SE), cost efficiency (CE), partial efficiencies or performance indicators (PI), allocative efficiency (AE) as well as dynamic efficiency (DE).

1.13 Given the requirements put by ACER for the regulatory use by NRA to implement cost-efficiency in short and long-run perspectives, the main metrics to be calculated are cost efficiency (CE) and dynamic efficiency (DE). Cost efficiency CE (in %) expresses the ratio of minimal cost to actual cost for transmission system services under comparable conditions. A cost efficiency score of 95% indicates that 5% of the total expenditure is in excess of what an efficient operator pays for the same services.

1.14 Dynamic efficiency DE measures concisely how the efficiency of operators evolve over time, both by emulating the best (catching-up inefficiency gaps) and by innovating to offer more value for the same cost (frontier-shift). Together, the two effects express the collective productivity development (in total cost) for the gas transmission sector.

1.15 In addition, CE is proposed to be accompanied with a decomposed measure, split into the changes in operating expenditure and capital expenditure, CE(2) to fully understand and decompose the sources of efficiency. Finally, scale efficiency SE measuring the share of possible inefficiency resulting from a suboptimal scale of operations, is a complementary metric to CE, which in itself is corrected for scale effects.

1.16 Metrics that do not integrate the cost consequence (TE, AE) or that are incomplete (PI) are deemphasized.

Methods

1.17 The report reviews the existing families of efficiency analysis methods that are or could be used in regulatory cost comparisons, including an extensive literature review of published work on gas network efficiency analysis (transmission and distribution), noting methods, data and metrics.

1.18 Considering the features required from the methods in terms of variable characteristics (deterministic or uncertain), the size of the reference set (single year or panel data), the importance of endogenous identification of peers and the identification of outliers, as well as the capability of integrating advanced dynamic analyses for the structure of inefficiency. The report (Chapter 8) concludes with the recommendation to use the two methods Data Envelopment Analysis (DEA) and Stochastic Frontier Analysis (SFA).

1.19 DEA (see §8.37 ff) is the most well-established efficiency measurement method in academic and applied studies, including regulatory benchmarking (cf. Section 8.5) since (i) it is does not require any technical assumptions of the functional form or type of inefficiency, (ii) it handles multiple outputs and inputs naturally in the model formulation, necessary for transmission models, (iii) it provides individual performance targets based on identifiable peer units that can be challenged, (iv) the calculations are fully transparent, replicable without any technical configuration, also for smaller and medium-size data sets, (v) accepted in regulation precedents to accommodate structural heterogeneity through its formulation comparing to most relevant peers rather than average performance. The weaknesses of DEA are the dependency on the assumption that all observations above best-practice cost are due to inefficiency, the sensitivity to outliers and data errors or variables that are subject to random noise, as well as the lack of statistical indicators for the uncertainty in its results (see §8.41**)**.

1.20 SFA (see §8.31 ff) is the strongest parametric efficiency method with frequent use also in regulation (cf. Section 8.5). The method has several important strength for AEC: (i) it separates inefficiency from random noise, whether it is due to data errors or random output variables, (ii) it gives full statistical inferences for test of hypotheses, confidence intervals and significance, (iii) it is less sensitive to outliers and misreporting, (iv) it is powerful in dynamic and time-series analyses, allowing checks for specific effects over time, specific shocks or common trends, and (v) its allows integrated modelling and testing of environmental and contextual variables, whereas DEA needs these to be established beforehand. The weaknesses of SFA (see §8.36) are the restrictive advanced technical assumptions prior to a run, sensitivity to the model specification (such as omitted variables) and the data requirements forcing relatively large datasets to enable effective analyses (see §8.61 and Section 8.8).

1.21 Observing that the weaknesses of DEA correspond to the strengths of SFA and vice versa, it is proposed to use a combination of the two methods for AEC (Section 8.13). Noting that frontier analysis is essentially constructing a cost function for transmission operations, there is no methodological reason to restrict it to a specific method: any coherent model can be calculated with the two methods if the data material allows. Combining frontier methods is considered best-practice in regulatory analyses (§8.107) for both methodological and validation reasons (see §8.106). It is also recommended by the Expert Review (ER) in their assessment that even suggested presenting jointly formed ("best-of") scores such as done in the regulation in Austria and Germany, among others (ER, p. 16-17).

1.22 As mentioned, the two complementary methods DEA and SFA are the most widely used in regulation and academic studies of productivity and efficiency for the sector (cf. Section 8.5). Given the importance of the task and the dataset, it is not seen as relevant to use methods (KPIs, COLS and ideal engineering networks) with methodological limitations (see §8.58 ff), neither is it necessary to force uniformity in choosing a single methodology if the model assumptions are differentiated. The

Expert Review (ER) supported this assessment and made additional suggestions for model extensions using these two methods, such as multi-year assessments of efficiency (ER, p. 15) and the use of latent class methods for complementary outlier detection (ER, p. 17, Agrell and Brea-Solis, 2017).

Models

1.23 The analysis of the current gas transmission challenges (Chapter 5) and the regulatory needs resulting from them (Chapter 6) translate into a series of requirements on the model design for the AEC. First, the NRAs need robust and economically sound models to assess the total cost performance without risk of contradictions (§6.27), capable of addressing increasing heterogeneity in operations and assetbase (§6.29). In a medium-term perspective to inform revenue allowances with a given assetbase, the assessment should be flexible and consistent also for different data and horizons (§6.29). These requirements suggest the use of an *execution-oriented* model based on the outputs provided by the asset provision of the TSOG in a year, without considering the past or future use of the network as outputs (§6.28).

1.24 However, the radical changes of the transport tasks, the energy transition and the ongoing conversion of existing assets also expect the NRA to have the AEC informing them about the dynamic development of grid structures, capacity and utilisation to meet the long-term challenges (§6.30). For these needs, a *utilisation-oriented* model is proposed to address network development as a function on network needs, market evolution and random risks in demand and supply (§6.31).

1.25 The AEC is therefore based on a dual design perspective: *an execution model* , aimed at measuring the cost efficiency of a given network, at given times and over time, and a *utilisation model*, aiming at measuring the overall cost efficiency of the network usage and investments over time.

1.26 Combining the time dimension (single-year, time-series) with the two risk perspectives results in three models being proposed for AEC, illustrated in Figure 1-1 below.

1.27 A *static execution model* (ES) takes uses the TSOG outputs to calculate efficiency for each individual year over a period. This model excludes exogenous events, and will provide efficiency factors that can be used in NRA decisions to set the revenue of the TSOG, either in the form of Total expenditure (Totex) or separately for capital expenditure (Capex) and operating expenditure (Opex).

1.28 The Expert Review also supported the proposal to combine the methods for the core model ES (execution-model with single-year data) to provide methodological cross-validation and hedge against potential uncertainty in the data (ER, p. 16)

1.29 A *dynamic execution model* (ED) based on the yearly results of the execution static model. These results will determine the catch up and frontier shift, showing how TSOG evolve over time, but still without exposing the operator to the utilisation risk. This model will show the efficiency changes of the most efficient TSOG, which set the efficiency frontier, and of the TSOG that are less efficient. These results are useful in setting the common productivity factors in revenue caps, as well as for monitoring individual efficiency improvements across time relative to TSOG performance and to best practice, respectively.

1.30 Finally, a *dynamic utilisation model* (UD) is presented, extending the variables from models ES and ED to include also utilisation variables (transported volume, maximum capacity used, etc). As mentioned above, the utilisation focus (Section

6.2, §6.07 ff) is necessary for long-run analysis by NRA (Section 6.4), desired by users and shippers (Appendix B, Section 2.5), and supported by the Expert Review (ER, p. 6). Utilisation rates cannot be included directly in ex-post assessment of efficiency (as in the ES) as this induces risks for the TSOG. Unless the cost for unserved demand is included in models or regulation, the evaluation would be biased against early investments. In addition, since grid investments are lasting and costly to remove, a point-wise evaluation of utilisation may also give the wrong conclusions for past investments that may have been optimal at the time of investment. On the other hand, limiting the regulatory cost review only to execution focus without demand factors would potentially give incentives for overinvestment or delayed restructuring. Likewise, if no assessment is ever made of the TSOG forecasting accuracy, the entire planning system is left outside of the regulation with potential impact on investment choices. The transport-capacity driven dynamic UD model complements the ES-ED model in these aspects, allowing for more nuanced review of adaptive capacity of the TSOG facing the gas sector challenges in Chapter 5.

1.31 Note that the dynamic utilisation model is different from dynamic *efficiency* (DE, Section 7.8) since it addresses changes in input and output ("effectiveness") as opposed to minimum cost for a given output ("efficiency") as in the model ED, tracking how the efficiency developing. The use of dynamic models is well established in gas transmission studies and SFA is the most common method used (Section 8.5). The Expert Review insisted on maintaining the panel-data approach to exploit the maximum information about the units (ER, p. 15) as also discussed in the AEC reference data analysis (Section 8.8). The applications are important to analyse the decreases in gas volumes (Section 9.4) and changes in direction. The variable choices for the UD model, including exogenous variables, is discussed in Section 13.1.

1.32 Structural comparability in the models is achieved through harmonised definitions of assets, costs, and outputs, with environmental and contextual adjustments explicitly integrated.

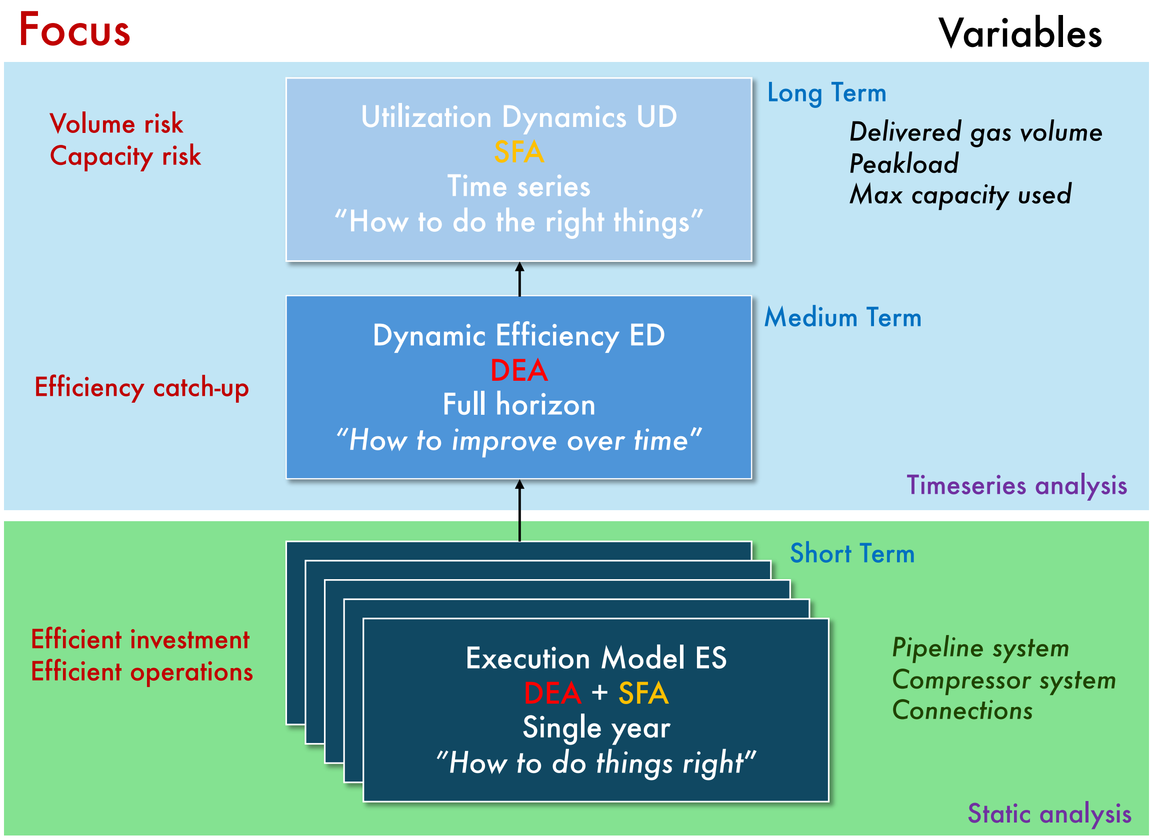


Figure 1-1 Core models ES, ED, and UD

Data and time horizon

1.33 Efficiency can be assessed repeatedly using a snapshot at a given time, using different data and assumptions. Given the dual needs identified for the NRA, both for a robust reference year assessment for e.g. a revenue cap and for a dynamic assessment of productivity gains, the report (8.8) proposes the full use of the time-series data for the estimation of models that can the extracted for one year (ES) or used as dynamic assessments (ED and UD). This approach is consistent with the strengths and weakness of the methods as well, since the method SFA performs best when exposed to larger dynamic datasets, whereas DEA dominates for a single-year analysis.

Sensivity analyses, outliers and peers

1.34 The AEC, as a cost comparison, essentially provides a minimal cost function. As such, the impact of atypical observations (outliers) and of minimal cost performance (peers) is paramount to the quality and precision of the resulting cost estimates. In this regard, the report reviews established outlier detection methods in regulatory efficiency analysis (section 8.14) to conclude that that an endogenous three-stage procedure (dominance analysis, superefficiency filter and peeling analysis) is to be recommended for adequate stability of the frontier.

1.35 Further, a thorough peer analysis (section 8.15) is explained and strongly recommended to ensure reliable results. The impact of parameter settings (such as interest rates, inflation indexes, etc) is proposed to be assessed using a set of sensitivity analyses where the relevant parameters are changed and the impact of individual and mean scores are documented and reported (section 8.16).

1.36 In addition to the sensitivity analyses, the report also recommends a few specific runs with different policy parameters in use (e.g. neutralisation of legacy investments, section 10.5, or national cost of capital, section 10.3). These runs are not limited to the mere impact on scores but fully reported and analyzed as complementary assessments.

Comparability

1.37 Each TSOG operates in a specific environment, exposed to different market conditions, operating in a regulatory environment that is country-specific and within a service area that is not directly comparable. It is essential for the AEC to consider these specificities and to propose solutions to make a proper comparison between the different TSOGs. If the benchmark fails to take that into account the resulting efficiency measures will be biased and cannot be used for improving efficiency.

1.38 The report presents several measures to ensure the comparability of TSOGs. These include corrections of operating expenditure (OPEX) capital expenditure (CAPEX) and outputs:

1.39 Raw accounting data for operating expenditure is incomparable since it contains heterogenous activities, non-controllable national costs, varying overhead costs and local labour costs. OPEX is therefore adjusted to include only in-scope functions through activity analysis with predefined functions for any TSOG. Non-controllable costs such cost of energy for losses and compressors, taxes, fees, and right-of-way fees are eliminated from in-scope OPEX. Support costs (overhead, management, indirect functional support) are fully reported, but reallocated to the in-scope OPEX using a standardized allocation key for all TSOG. The local staff costs for direct and indirect functions are adjusted to European standard cost levels using a labour cost index. Finally, the OPEX across time is inflation adjusted using a relevant inflation index for all operators.

1.40 Book-value extracted residual values cannot be used for benchmarking since the initial investments are depreciated over different periods, may exclude direct subsidies or contributions from third parties, and potentially include non-standard elements such as financial costs during construction, land investments or assets for third parties (joint assets). Moreover, diverse financial costs across Europe cannot be directly compared since they may be nominal or real, include or not the same elements, involve different components in the calculation of allowed capital cost and be applied to asset bases that are different for identically aged and dimensioned assets. For these reasons, CAPEX is recalculated bottoms-up from gross investments by using standardized techno-economic lifetimes and standardized capital costs in real terms, excluding fiscal depreciation or local financial costs, limiting the asset base to in-scope asset categories only with correction for jointly owned assets, asset ages, upgrades-lifetime prolongations, and repurposing. The investment stream is indexed over time using relevant material price and inflation indexes, applied consistently for all operators

1.41 The TSOG outputs in terms of grid, service and capacity provision are measured through objective, verifiable and technical variables that are calculated from the underlying reports, such as aggregations of grid-pipeline sections, compression equipment, connection points, regulators, metering stations and other installations. For the utilisation-focused model, the outputs also include transported volumes, maximum contractual or real transport volume, and derived variables. To achieve full comparability of the rendered services, the outputs are corrected with environmental complexity factors that apply to specific output factors, e.g. soil conditions to pipeline assets. The individual environmental corrections are based on the GIS-data for grid locations and environmental spatial databases for landuse cover, slope, soil, humidity, wetness and weather. In addition, the output data related to grid assets are corrected for joint operations, age and repurposing.

1.6 Robustness

1.42 The AEC includes safeguards to strengthen the robustness of the methodology, which includes:

Measures to support the use in regulatory decisions:

1. The choice of method, data set, environmental corrections and cross validation correspond to international scientific best practice for regulatory benchmarking, see section 8.18.

2. The use of proprietary or fully confidential data render comparisons fragile and difficult to validate. AEC is proposed based on a largely transparent data structure with a specific protection only for confidential and security-sensitive data through a three-tier information structure: open information (T1), commercially sensitive information (T2) and spatial-technical information (T3). This approach is designed to both cater legitimate needs for data disclosure as well as for guaranteeing high-precision estimates using good cost and spatial information from TSOGs.

3. Data quality is proposed ensured by a multi-level data validation architecture including independent data validation by NRAs, consultants, cross validation by peer TSOG and NRAs, as well as third-party auditors.

4. Transparent application of methods applied to control for structural, economic and environmental heterogeneity is proposed prior to, and during the calculations of results.

Measures to ensure the relevance of the results:

5. The metrics and models are aligned with EU and national frameworks for the energy transition, such as decommissioning, repurposing and asset rehabilitations both for use in periodic revenue allowances and for qualitative monitoring of progress in terms of decarbonisation, repurposing and graceful dismantling of existing grid systems.

6. AEC is proposed to provide specific runs for variations of national capital costs (section 10.3) and decommissioning premium (section 9.2), as well as neutralisation of legacy investments prior to EU-regulation (section 10.5), improving the range of potential application of AEC results in national regulation.

7. Model specification of two types; executive and utilisation focus, to inform about different dimensions of efficiency (medium-term vs long-term), at a given time and over a period.

1.7 Data quality

1.43 Data collection is based on common templates covering assets, financial data, and indicators, subject to multiple validation layers: TSOG self-checks, NRA reviews, independent audits, and cross-validation across all TSOG data.

1.44 Sumicsid proposes a four-year cycle, with annual updates to maintain accuracy and comparability. At the same time, it should be noted that ACER will decide on the frequency of the AEC once the workload associated with the data request and the modelling is visible.

1.45 Transparency across the data is ensured through published definitions, open access to non-sensitive data, and step-by-step audit trails enabling reproducibility.

1.8 Transparency and publication requirements

1.46 The report identifies transparency as a key requirement for the AEC. The earlier benchmarks performed by CEER were carried out on the basis of voluntary participation, which ultimately limited the disclosure the results and underlying data used for the modelling. The report proposes a two-fold approach to transparency to ensure that commercially sensitive data, or data with other sensitivity markings, is protected, while the rest of the information can be subject to transparency requirements. The report recommends disclosing the underlaying data to be used for the AEC modelling. At the same time, the report recommends that ACER carefully assesses the sensitivity of the data to be published, checking this status with NRAs and TSOGs.

2. Introduction and Terms of Reference

2.1 Legal mandate of ACER

2.01 Article 19(2) of Regulation (EU) 2024/1789 establishes the requirement for ACER to carry out an efficiency comparison for natural gas transmission system operators' (TSOG) costs. By 5 August 2027 and every four years thereafter, ACER should publish a study comparing the efficiency of TSOGs' costs, subject to the protection of data considered by ACER to be commercially sensitive. The AEC is a benchmark that will assess the relative cost efficiency for TSOG in the EU.

2.2 Process of ACER for AEC

2.02 ACER has structured the work leading to the first publication of the AEC in three different phases:

1) *Phase I* establishes the objectives of the comparison in the context of decarbonisation and to define the methodology and the data approach that will be used in the comparison. Phase I is hereafter referred to as PAEC.
2) *Phase II* is intended to collect and validate the TSOG data that will be used in the comparison. The data request will be sent to TSOG and the data will be validated by NRAs and ACER supported by a consultant. The final objective of this phase is to provide data with the level of quality necessary to apply the methods established in phase I.
3) *Phase III* is intended to complete the modelling work based on the methodology defined in phase I and the data collected in phase II.

2.3 Terms of reference

2.03 For Phase I, ACER has commissioned this study from Sumicsid and Swiss Economics (thereafter Sumicsid).

2.04 The terms of reference for this study are established in the direct service contract (ACER/NEG/GHR/22/2024) according to which the Sumicsid shall assist the Agency in the design of a TSOG efficiency comparison in the context of decarbonisation for all the TSOG active in the Union.

2.05 Sumicsid is responsible for elaborating, in coordination with ACER and with NRAs, a report on the AEC. The Report should cover, first, the approach and objectives for the AEC in the context of decarbonisation; second, methodology for carrying out the AEC. As part of the approach and objectives for the AEC, the report should cover the current and future challenges of TSOG, the trends on gas TSOG revenue regulation, the role of an efficiency comparison promoting efficiency in natural gas transmission networks and the approach and objectives of the AEC.

2.06 As part of the design of the methodology for carrying out the AEC, the report should cover the calculation methods to perform an efficiency comparison for TSOG, the parameters to be used as an input to the methodology, the data request and the process.

2.07 Sumicsid should further organise two workshops, one with NRAs and a second one with all stakeholders.

3. Project setup and structure of the report

This Chapter explains the project setup and provides the outline of the report.

3.01 Phase I of the project (PAEC) was divided in two parts (T02 and T03), each composed of task notes, resulting in two reports, D02 for Objectives and Criteria and D03 for Method, Data and Process, respectively. The two parts converged for a public discussion at workshop 2 (WS2) and were subject to a public consultation as well as an expert review.

3.02 The four task notes in Part T02 are devoted to identifying the objectives and criteria for the AEC, (see Figure 3-1 below). After discussion and analysis with the stakeholders, the notes in T02 were merged to a synthesis document D02.

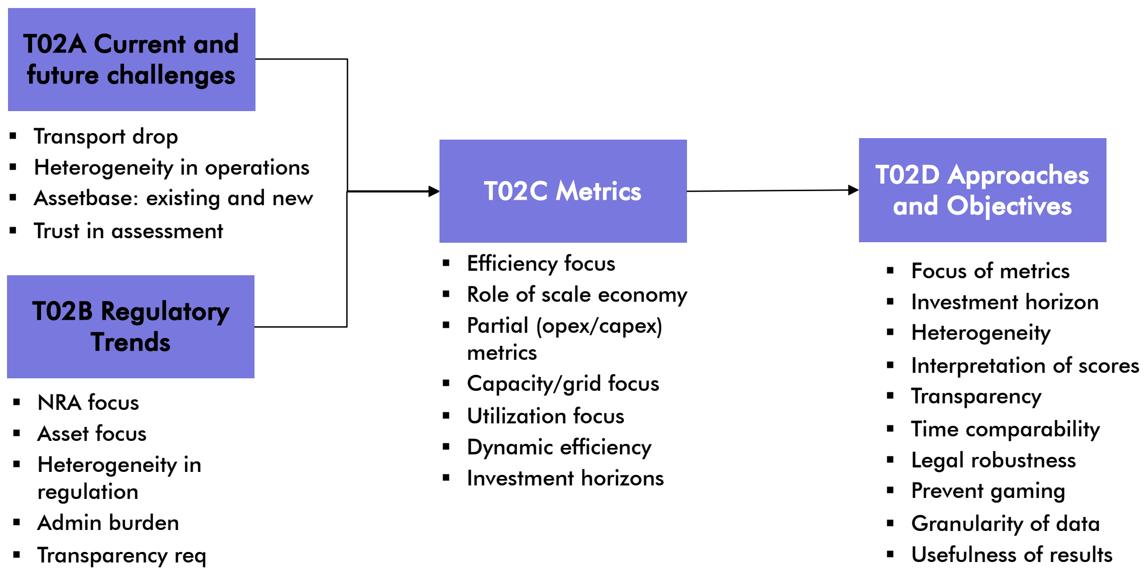


Figure 3-1 Overview of subproject T02 task notes.

3.03 The second part (T03) is devoted to determining methods for assessing efficiency, to outline necessary variables and parameters, to define requirements for collecting and verifying data, and to define a possible process for the AEC, based on the prior part T02, as illustrated in Figure 3-2 below. After consolidation and discussions at workshop 1, the task notes T03 were summarized in a synthesis document D03.

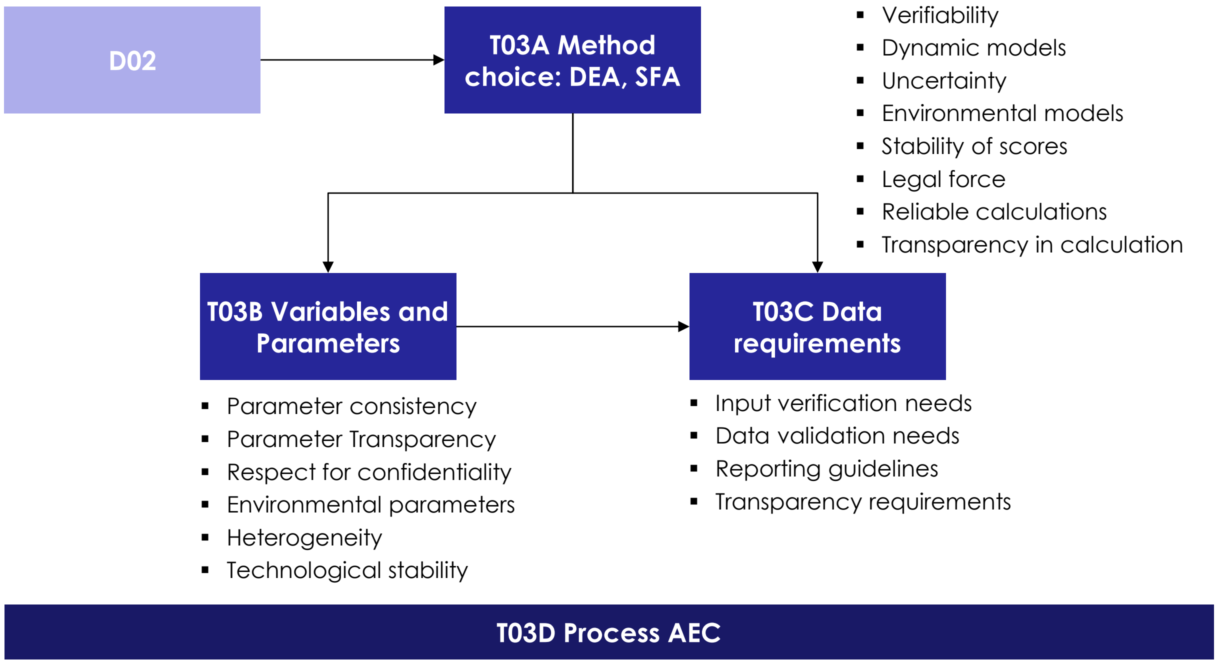


Figure 3-2 Overview of subproject T03 task notes.

3.04 At workshop 2 (WS2) in Brussels on July 9-10, 2025, the project proposals and the two synthesis documents D02 and D03 were presented and discussed. In addition, a public consultation was carried out based on a set of questions (Appendix A). A summary of the public consultation responses is provided in Appendix B.

3.05 The entire proposal was reviewed by an independent expert review after workshop 2. The summary of the Expert Review findings on the original document that is published as a separate document, called ER in the text. A summary of the main suggestions by the Expert Review and how they are integrated in the current proposal is found in Appendix C. The group was composed by the following academic experts:

1) Prof. Dr-Ing. Joachim MÜLLER-KIRCHENBAUER, Technical University Berlin
2) Prof. Dr. Mette BJØRNDAL, Norwegian School of Economics and Management,
3) Prof. Dr. Tooraj JAMASB, Copenhagen Business School
4) Prof. Dr. Anne NEUMANN, Norwegian University of Science and Technology
5) Prof. Dr. Luis OREA, University of Oviedo

3.06 This synthesis is based on the following documents:

1) CEER: Future role of gas (FROG) study (2018).
2) DNV study for CEER, DNV (2018)
3) DNV study for ACER, DNV (2022)
4) NRA Survey conducted in the PAEC project in March 2025
5) Information from the European Commission and European Scientific Board on Climate Change studies, Enerdata (2024)
6) Findings of previous projects e.g., 3.06, Sumicsid-CEER (2024) and EFG4 Swiss Economics, Sumicsid, IAEW (2023)
7) Model specifications of other benchmarks, e.g. EFG4 Swiss Economics, Sumicsid, IAEW (2023),

8) Reference network models developed by SE, IAEW and DNV,
9) CEER-Sumicsid (2023) Note on Environmental Modelling (Appendix B),
10) Scientific literature (a list of references is provided in the appendix)
11) The AEC Project reports D02 and D03, published before the AEC workshop 2 (WS2), each drawing on their preceding notes T02 and T03, respectively

3.07 This final report document (D04) draws also on written input from the public consultation of documents D02 and D03 following Workshop 2 (WS2), workshop interactions, and Expert Review feedback.

Outline

3.08 The rest of the report is organised in 12 chapters followed by an Appendix.

1) Chapter 4 explains ACER's foreseen process to complete the first AEC.
2) Chapter 5 reviews the current and future trends in the gas sector.
3) Chapter 6 treats the current regulatory methods used in Europe for gas TSOG.
4) Chapter 7 analyzes existing efficiency metrics and proposes indicators for the AEC.
5) Chapter 8 summarizes efficiency methods and AEC models
6) Chapter 9 looks at solutions for problems linked to the energy transition.
7) Chapter 10 summarizes the harmonizing of capital costs and expenditures.
8) Chapter 11 outlines how structural comparability can be achieved.
9) Chapter 12 discusses the robustness measures for the AEC.
10) Chapter 13 explains the different variables and parameters for the AEC models.
11) Chapter 14 discusses the data collection and validation.
12) Chapter 15 proposes a process of AEC with actions, timing and frequency.

3.09 The Appendix contains complementary information

1) Public consultation questions
2) Summary of public consultation responses
3) Response to the Expert Review
4) Data structure and confidentiality levels
5) New data collection items

4. Process

4.1 Phase I: Objectives, methodology and process

4.01 Article 19 of Regulation (EU) 2024/1789 requires that Acer publish the AEC by 5 August 2027 and every four years thereafter. ACER has designed the process for the publication of the first results in three stages, the first of which is completed with the publication of this report.

4.02 Phase I establishes the objectives of the comparison in the context of decarbonisation, defines the proposed methodology and the data approach that can be used in the comparison.

4.03 From Chapter 5 onwards, this report documents the results of this first phase.

4.04 The work completed during Phase I provides the objectives, methods, data approach and process to be used in the next two phases. The suggested process is summarised in Figure 4-1.

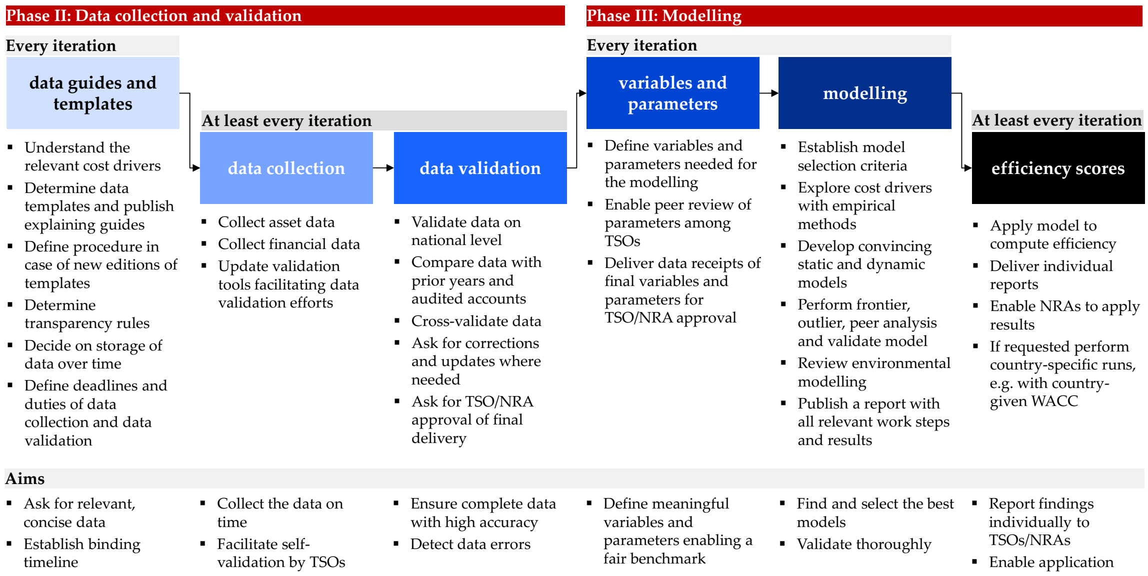


Figure 4-1: Overview of the proposed process for phases II and III.

4.2 Phase II: Data collection and validation

4.05 The objective of phase II is to collect and validate the TSOG data that will be used for the calculation of the efficiency results. The data request will be based on the data collection carried out as part of TCB18 (Sumicsid and CEER, 2019) and TCB21 (Sumicsid and CEER, 2024).

4.06 In these projects CEER together with NRAs developed detailed data request templates with data definitions and accompanying guides. In TCB21, in addition data validation tools were provided to facilitate national data validation and inform on how the data will be processed by CEER's consultants. The process of TCB21 is illustrated in Figure 14-3.

4.07 While phase II will be designed by ACER, this report proposes several steps to carry out the work.

4.08 First, the data request should be reviewed to include additional data related to labour costs in outsourced services (section 11.4), labour costs in capex (see §12.29) and costs related to the adaptation of asset to transport hydrogen (section 9.3). These additions are further summarised in Appendix E.

4.09 Second, the data request should be discussed with NRAs and TSOG to validate the approach, requested data points, timeline, interactions and IT tools used.

4.10 Third, the data request is sent to TSOG, who should fill in the relevant templates. This step should ensure sufficient time for TSOG and might require assistance from ACER.

4.11 Fourth, once the data has been received the validation process can start. The report recommends several approaches to complete this work, which include the validation of the TSOG submission by NRAs, a third-party audit on the data submission and the cross-validation of all the data. Sumicsid further recommends the partial disclosure of the data request to enable a final level of cross-validation where TSOG and market parties can intervene.

4.12 The experience of TCB18 and TCB21 has shown that binding deadlines for deliveries, corrections and approvals are of crucial importance to ensure a streamlined process. The process should assume multiple iterations (in rare cases up to 10 iterations). Furthermore, the quality of the validation is of utmost importance, as errors in the data could potentially impact the efficiency frontier. During TCB21, some TSOG questioned the quality of the validation across all data submissions. The proposed data collection and validation process in this report aims at avoiding the identified risks and shortcomings from past benchmarking projects.

4.13 Article 19 of Regulation (EU) 2024/1789 further requires that data confidentiality is taken into account. To this aim, the report proposes a two-fold approach to transparency that should be considered during phase II. Data that is considered commercially sensitive, or that falls under some other sensitive category, should remain protected by ACER across all steps of the AEC. Data that is not considered to fall under these categories can potentially be made public. ACER should assess the forms in which this information can be published, during phase II, for the the validation of the TSOG data submissions.

4.14 An important part of the data request is the locational data of TSOG assets, which is discussed under section 13.6. The report proposes the use of detailed GIS data to be used for the AEC. This is further supported by the Expert Review (summary in Appendix C). At the same point, the report acknowledges that a detailed level of granularity could raise security concerns. For this purpose, the report proposes alternatives to GIS data, which include NUTS2 and NUTS3 data. In TCB21 public data from OpenInfrastructureMap (OIM) was used for pipeline locations, if existing and complete. In cases where the OIM data was incomplete, NRA or TSOG data with different granularity was collected for processing with a safety width of up to 10 km around the asset locations.

4.3 Phase III: Modelling

4.15 The last phase of the AEC aims at completing the modelling work on the basis of the AEC Method and Process using the data validated during this phase.

4.16 The modelling includes the following main steps:

1. Definition and approval of variables / potential cost drivers: After alignment with NRAs and TSOGs , a list of variables and definitions is published. Based on the validated data from phase II, all variables are computed and delivered individually to TSOG in the form of data receipts for formal approval. Subsequently, parameters in T1 that can be published according to the transparency rules are published for peer-review among TSOGs.
2. The modelling phase consists of an empirical cost driver analysis, model development and model specification with the aim of finding a valid static and dynamic frontier. The frontiers are validated, including in-depth analysis of environmental factors.
3. Sumicsid recommends ACER to publish a common report including all work steps, considerations and relevant results

4.17 Subsequently, the resulting frontier models can be applied in each year with little effort, allowing the determination of TS-specific efficiency scores. To this end, standardized TSOG-specific reports are provided to the TSOG and respective NRAs. Upon request of NRAs, country-specific runs with a country-given WACC can be carried out as well as specific dynamic analyses.

5. Current and future trends in gas markets

In this Chapter, we discuss the major changes in the gas markets and their consequences for the TSOG. Understanding these changes is critical as the benchmarking exercise requires taking them into account. The Chapter further proposes solutions for guaranteeing the comparability of the TSOG that are unequally affected by these changes. The options to assure comparability and align with the energy transition objectives that are further discussed in Chapter 9.

5.1 National development plans and decreasing gas volumes

5.01 The energy transition has led and will continue to lead to significant changes in natural gas markets and transportation requirements. With the EU's Third Energy Package national development plans (NDPs) were introduced, also aiming to interconnect existing infrastructures and provide more diversified sources, sometimes also leading to changes of market zones and gas quality. The EU has the ambition to be carbon neutral by 2050, an economy with net-zero greenhouse gas emissions. Consistently, all the Member States have national energy and climate plans that organize the phasing out of natural gas from fossil sources. Overall, *demand for natural gas is expected to decline steadily*, especially in sectors/regions where electrification and renewable energy solutions are viable. Since the decrease of consumption does not necessarily lead to a decrease of peak demand, opportunities for decommissioning may be limited.

5.02 Decreasing gas volume expose the TSOG to the utility *death spiral* (cf. Felder & Athawale, 2014, Costello & Hemphill, 2014). Investments being made for long periods (up to 50 years) and assets are not easily redeployable, implying that network costs are dominated by fixed costs. A decrease in the volume of gas consumption then implies that these essentially fixed costs will be allocated to a lower volume of consumption or a lower number of consumers, implying an increase in the regulated tariff, which could further accelerate the decrease in demand.

5.03 Decreasing consumption and decreasing the number of clients may also trigger the choice between extending or replacing assets. TSOG must decide whether gas network assets at the end of their regulatory lives are replaced by pipeline infrastructure of similar or smaller size (reinvestment) or whether the operation of these assets can be extended. According to DNV (2022), between 2010 and 2022, replacement investments became more relevant compared to expansion investments, with the share of replacements of total investments increasing from 24% to 48% (DNV). This is particularly important since the depreciation of regulated asset vary widely among member states and NRAs, between 27 and 70 years for pipelines and 13 to 60 years for compressors, DNV(2022).

5.04 The gas infrastructure providers (TSOG and DSOG) in natural gas directly impacted by the decreasing gas volumes but have no active role in changing the demand.

5.2 Change in the gas sourcing

Biogas, sector coupling (P2G, G2P)

5.05 New entry points for biomethane (biogas) injection and power-to-gas (P2G) are also important elements of the energy transition. The REPowerEU plan targets to scale up biomethane production to 35 bcm (billion cubic meters) by 2030, replacing about 20% of current natural gas imports. This will require new injection points for biomethane production, as well as pipelines to connect to these injection points. These new biomethane entry points may also change flows in relevant parts of the transmission networks. To scale up biomethane production, these connections are also often subsidized.

5.06 About P2G, the use of electrolysers to produce hydrogen from green electricity will require new entry points with similar implications as the biomethane entry points, in particular for power plants with connection to the natural gas network. The developments also lead to new exit points such as Gas and steam turbines / Combined Cycle Gas Turbine Plants (CCGTs) and combined heat and power units (CHPs).

Russian invasion-related challenges

5.07 The invasion of Ukraine by Russia in 2022 was accompanied by an abrupt decrease in Russian gas supply. To reestablish supply security in the medium term, large investments were made in liquefied natural gas (LNG) terminals, leasing contracts for floating storage and regasification units (FSRUs) and infrastructure to connect these new port terminals to the network (cf. Agrell et al, 2025). This has led to a rapid expansion of LNG capacity.

5.08 Another trend accelerated by the Russian invasion of Ukraine is the relevance and usage of reverse flow capabilities. While the bi-directional capacities of interconnection points between EU member were already a requirement in the Gas SoS Regulation (EU Regulation 2017/1938)[1], the reduction of Russian gas supplies made the reverse flow capability a critical component for redistributing imported LNG and gas from alternative sources, leading to different flows in existing pipelines. These trends also allowed the diversification of supply sources and increased LNG imports from alternative countries (e.g., the United States).

5.3 Hydrogen

Hydrogen-ready pipelines

5.09 There is still substantial uncertainty about future supply and demand for both gas and hydrogen. Hydrogen-ready pipelines transport both natural gas and hydrogen or blends of the two, enabling a gradual transition to a hydrogen economy. Currently, new pipelines are built hydrogen ready (e.g. in Greece or connection to new LNG terminals in Germany), and in the coming years, an increasing share of H2-ready assets is expected. Depending on national and European hydrogen network plans (TYNDP) there are regional differences in the hydrogen readiness of pipelines.

5.10 Capex may increase when pipelines are hydrogen-ready, but it is not clear whether there is an effect on Opex. In terms of outputs, hydrogen-ready pipelines are

[1] With the possibility for member states to apply for derogations.

expected to increase in length and volume, while the corresponding values for total pipelines remain mostly unchanged.

5.11 Hydrogen-ready pipelines are efficient when repurposing is planned in the future. TYNDP can be used to check whether this is the case. Parameters to differentiate between conventional and hydrogen-ready pipelines should be available for benchmarking, necessitating a refinement of the asset data collection. Possible non-investment costs for the preparation, technical studies or setup of H2 facilities undertaken by TSOG should be reported separately with a consistent treatment of subsidies/public contributions.

Repurposing of existing pipelines to H2

5.12 The hydrogen pipeline infrastructure that will be deployed in the next decades will partially use repurposed gas pipelines. The costs for repurposed gas transmission pipelines are expected to be between 60 to 90% lower compared to building new hydrogen pipelines, see ACER (2021), Cerniauskas et al. (2020) or Monsma et al. (2024). The possibility that a natural gas pipeline can be repurposed depends strongly on its geographical location and whether a hydrogen corridor or pipeline is planned in that region.

5.13 The activities of TSOG to repurpose assets are linked to different costs: costs to assess technical feasibility and need of an adaptation, additional cost to ensure the assets are hydrogen-ready, costs associated with the separation of assets and organisation and cost related to the actual transfer to the H2 TSOG, as the European framework foresees a horizontal unbundling between Gas and H2 TSOs. The repurposing may increase Opex of the TSOG in the short term but decrease it in the long term. Capex decreases after repurposing. The natural gas outputs decrease. Hydrogen outputs of the H2 TSOG increase, in particular asset-related ones (e.g. pipeline length and pipeline volume).

5.14 The scope of the benchmarking will cover TSOG only, hence H2-repurposed pipes should not be included. The benchmarking should ensure that natural gas networks that shift their assets to hydrogen networks have neither an advantage nor a disadvantage to repurpose unused pipelines. For that reason, we propose to apply a uniform transfer value to H2 operators. This point is discussed further in section 9.3 below.

Decommissioning of existing gas pipelines

5.15 Some pipelines cannot be repurposed due to technical or economic constraints and will be decommissioned. The decommissioning occurs especially in regions where gas demand is rapidly declining, e.g., those reliant on Russian gas pipeline imports. The decommissioned pipelines will be physically removed or left in the ground. The TSOG will have additional costs for the decommissioning and dismantling of pipelines. Article 19 of EU Regulation 2024/1789 proposes a modification of the depreciation profile of transmission assets to prevent a situation where the cost recovery of transmission system operators through tariffs threatens the affordability of natural gas.

5.16 Before decommissioning, changes in depreciation profiles lead to higher Capex, while Opex is likely to decrease as the pipeline is less or no longer maintained. With decommissioning, Capex decreases for fully depreciated and stranded assets while the outputs (e.g. pipe length or pipe volume) decrease as well.

5.17 To compare the TSOGs , having a uniform treatment for decommissioned assets is important. This implies that the benchmarking by default standardizes the values of

Totex, Capex, Opex, and output. We explain this in greater detail in section 9.2 below.

5.18 However, to provide incentives for phase-out of underused assets, we also propose a specific run with a higher value, see article 9.08 below.

6. Current regulation for TSOG

In this Chapter, the current regulatory prerogatives and methods for regulating European TSOGare discussed. This Chapter provides thereby the requirements from the NRAs on the features and outcomes of the AEC that are designed in the rest of the report.

6.1 Overview

6.01 The recovery of TSOG investment costs is a two-step procedure. In a first step, the total allowed revenue for the regulated firm is determined and, in a second step, the tariff methodology is designed to recover the allowed revenue from users of the infrastructure.

6.02 In a nutshell, to determine the total allowed revenue, regulators either consider total expenditure (TOTEX) or distinguish between operating expenditure (OPEX) and capital expenditure (CAPEX).

$$\text{TOTEX} = \text{OPEX} + \text{CAPEX}$$

6.03 Operating expenditure OPEX is further divided into controllable and non-controllable OPEX (or controllable and non-controllable TOTEX). The idea is to isolate expenditures over which TSOG exert some control from those who are not controllable. In practice, the distinction between fully controllable, partially controllable and non-controllable costs varies from one country to another. Furthermore, certain countries distinguish between permanently and temporary non-controllable cost components in Opex. We state the formula (the components are discussed further below)

$$\text{OPEX} = \text{OPEX-C} + \text{OPEX-NC}$$

where OPEX-C are controllable (in-scope) costs of goods and services and OPEX-NC are non-controllable or out-of-scope costs of goods and services.

6.04 For capital expenditure CAPEX, the classical regulation consists in remunerating the regulated asset base (RAB) at a given rate, the weighted average cost of capital (WACC). For non-controllable OPEX, a common solution is a cost-plus regulation. The formula for standardized CAPEX in the AEC

$$\text{CAPEX} = \mathit{annuity}(r,T)\ \text{CAP-REAL}$$

Where CAP is a vector of the investment values in real value, r is a real interest rate, T is the vector of techno-economic lifetimes[2] for asset categories in CAP-REAL, and *annuity* is the annuity function. CAPEX is a real annuity of capital expenditure for the financing and use of the investments CAP over their effective lifespan T.

6.05 For controllable TOTEX/OPEX, there is an incentive regulation where the regulator fixes a cap, eventually completed by a common (X) and potentially a firm-specific efficiency factor (X_i). The NRA survey shows that many regulators use efficiency

[2] Note that the techno-economic lifetime is standardized per asset, it may be different from the fiscal national depreciation periods that are applied for accounting reasons.

factors.[3] These factors can be determined by an appropriate benchmarking exercise. For such a benchmarking exercise, it is important to properly define the scope of the benchmark and the perimeter of activities and the associated costs and outputs that are included in the benchmark (see section 6.5). The efficiency factor(s) resulting from the benchmark should only be applied to the costs that are effectively benchmarked.

6.2 Use of efficiency factors in regulation

6.06 Efficiency can relate to the construction and utilisation of specific assets (*utilisation focus*) or the planning and execution of grid services with a given infrastructure (*execution focus*).

Utilisation focus

6.07 In a vertically separated sector, the utilisation of the infrastructure depends on the matching of supply and demand (load), but subject to exogenous factors and prices, resulting in a downwards uncontrollable indicator. Used ex post, utilisation-focused regulation (such as based on share of capacity invested used or similar) could reward and penalize TSOG subject to demand increases and decreases, respectively. However, as the unobserved demand (full request for volume and capacity) normally is not contractually binding for the lifetime of the specific assets (no binding long-term take-or-pay contracts from all clients), the main effect is a punishment for temporal or permanent drops in transport demand. Consequently, the one-sided risk borne by the TSOG will have to be integrated in some way in the expected return or the optimal TSOG policy will be to lag in capacity building.

Execution focus

6.08 For these reasons, most regulatory benchmarking in Europe is *execution-based*, meaning that efficiency primarily relates to the cost of procurement, construction, operation and maintenance of grid assets put at the users' disposal. A TSOG that makes poor preparations, costly and ineffective tendering and inadequate maintenance interventions will show up as having an execution-based inefficiency compared to a best-practice TSOG. For many TSOG, the primary sources of inefficiency are operational management and procurement, leading to lower staff productivity, high outsourcing costs, as well as high equipment, installation, and sitting costs due to choices of kits, standards, timing, suppliers and locations.

6.09 However, the execution focus does not depend on the network planning, the asset intensity, dimensions or delays in finishing projects (planning efficiency). Nor does it consider the (varying real) utilisation of the assets, in part of general (utilisation focus).

6.10 The execution focus is a necessary but not sufficient element to achieve full effectiveness in operations (see below). Given the uncertainty of demand, infrastructure development and transport flows, it is nevertheless the sound and safe approach to monitor TSOG investments and operations, without adding risk.

Integration in revenue caps

6.11 The efficiency factors can be used both in low-powered (cost-recovery) and high-powered (incentive) regulation. The use in ex-post cost recovery regulation (cost-

[3] List of countries in NRA survey: AT, BE, BG, CZ, DE, DK, ES, FI, GR, HR, IE, IT, LV, PL, PT, SI, SK.

plus) is obvious but rare, we therefore focus on incentive regulation using a revenue-cap regime with the standard parametrisation for inflation correction, quality adjustments and by-pass of non-controllable costs. Ex-ante regulation could also be normative, for instance using investment unit cost values obtained from engineering studies or benchmarking, without concerning past investment or operating expenditures. The incentive effects of such regulation should be carefully reviewed.

6.12 Assume that we have revenue cap window of years 1,2,..,T years, and we define the baseyear as year 0. A very simplified formula for application (without individual efficiency requirements, that is a common X, also called Xgen) is:

$$Revenue(t) = Revenue(0)*CPI(t) * (1- X)^t + Z(t) + Q(t)$$

6.13 where Revenue(0) is the allowed revenue the base year 0, CPI(t) is an inflation adjustment for year *t*, X (or Xgen) is the technical change or productivity gain per year observed, Z(t) are by-pass costs not in the revenue formula (losses, new investments, non-controllable costs) and Q(t) is a possible quality adjustment factor (bonus-malus for accidents, supply interruptions etc). Note that Xgen is the same for all operators, it is the common productivity requirement.

6.14 When benchmarking is used for individual targets, an individual term for each TSOG is calculated, Xi, using e.g. the relationship:

$$Xi = (1- CE(i))/Tm$$

Where CE(i) is the cost efficiency of TSOG i and Tm is the catch-up period decided. It might not be possible for an operator to catch-up and adopt best practices immediately. For that reason, regulators may spread the effort over a longer time period Tm and allow the (inefficient) operator to catch-up progressively.

6.15 Retaining the earlier X-factor that applies for all TSOGs , adding the individual Xi assuming that the Xi factor should not apply to costs that are out of the scope of the benchmark (which are then included in the term Z(t)), the formula becomes:

$$Revenue(t) = Revenue(0)*CPI(t) * (1- X-Xi)^t + Z(t) + Q(t)$$

6.16 We will discuss hereafter the different measures of efficiency. In a nutshell, with an appropriate method and metric (dynamic efficiency, see section 4.7), it is possible to distinguish general technical progress, what we call *frontier shift* (TC) from pure *efficiency changes* (EC), that is laggers who are catching up (see section 7.8 below). The frontier shift can be used for the common factor X and the catching-up for the individual specific factor Xi, measures that NRA could include or not in their gas regulation.

6.3 Role and use of efficiency factors in the energy transition

6.17 The AEC can assist NRA to support and incite TSOG in the energy transition in multiple ways.

General efficiency requirements

6.18 The execution focus of AEC provides the NRA with information to monitor and incentivize procurement, investments, staffing, operations and maintenance for existing assets. This dimension is intrinsic to frontier-based benchmarking and promotes best-practice and penalizes low-productivity staffing, gold-plating and slack in procurement, excessive outsourcing, and the inclusion of inflated asset values to increase revenues and profits. Through adoption of best practice, identified in AEC for structurally comparable operators, the operators may also optimize the degree of inhouse/outsourcing and level of internal delegation and scale.

Incentives for asset extensions

6.19 Conventional benchmarking will create disincentives for significant asset rehabilitations and life-time extensions. In AEC, these measures are seen as highly relevant to avoid reinvestments for a shorter period. Consequently, the Capex values are adjusted as to incentivize these life-extending investments as long as they are economically rational and technically feasible

Prevent perverse incentives for repurposing

6.20 Without analysis, there are perverse incentives for TSOG to sell or transfer pipeline assets with inefficiently high investment values to hydrogen operators under setup. To avoid nominal values in transfer, an operator might initiate asset reconfigurations, land preparation or soft planning actions that are capitalized with the incumbent asset value. This increases capex-efficiency for the TSOG and hides gold-plating by asset transfer. In AEC, these transfers are reported separately and monitored for the original value and compared to mean unit-cost values.

Incentives for decommissioning

6.21 By default, the AEC model is neutral to decommissioning unused assets at RAB-values. However, in the specific run R4 (see Table 8-23 below), NRAs may create incentives for decommissioning by allowing for flat replacement values in the capital stock[4], meaning that decommissioning of incumbent efficient investments will also be promoted. The detailed process for decommissioned assets is discussed in section 9.2 below.

Utilisation incentives

6.22 A utilisation-based model can be used to monitor the development of the asset utilisation for a TSOG over time, controlling for random factors in volume and capacity load. This may assist NRAs in the dialogue with TSOG to guide decommissioning, repurposing and/or new investments during the transition period until 2050. This type of model can also be used for forecasting of future loads and asset uses, which in turn may improve the effectiveness of TSOG.

Heterogeneity

6.23 With the radical changes of the natural gas sector brought by the energy transition and the supply security crisis, the differences of operation conditions among TSOG are ever more important as not all countries have the same demand and transport patterns. NRA will need to carefully assess and evaluate the individual preconditions

---

[4] Decommissioning an asset with an inefficient investment level increases the score in a benchmarking. Here we address the decommissioning of efficient investments that are underused. By deducting a higher value than the original investment from the Capex base the score may increase also for efficient assets. The padded value can e.g. include the net present value of future avoided capital and maintenance cost.

for their TSOG when comparing to best-practice options of investments and operations. This requires considering spatial, geological, hydrological, economic and structural conditions in their area. Since some of these individual heterogeneity factors vary over time, the methods will need to be able to address uncertainty in input data.

Sector transparency and accountability

6.24 The AEC promotes transparency and accountability. The AEC provides objective and harmonised cost efficiency comparisons, ensuring that TSOG are accountable as other network operators (DSO, TSOE, RTO) for equity. Stakeholders (governments, investors, consumers) can use AEC data to track progress in gas infrastructure decarbonisation.

Dynamic improvement incentives

6.25 The use of **dynamic efficiency (see section 7.8 below)** allows for the decomposition of efficiency changes into general **technical progress** and **efficiency catch-up**. This has the following implications:

1) Incentives can be provided to inefficient TSOG that improve their relative position over time, even if not fully efficient. This trajectory may be optimal if the restructuring costs are higher than the benefits from optimal operations. TSOG that demonstrate consistent efficiency improvements (catch-up) are therefore assessed as dynamically improving, although they may be statically inefficient. The proposed methods in AEC allow moreover the identification and isolation of potential causes for the initial problems, e.g. legacy investments or distorted capital costs.
2) For the sector, it is important to gauge the overall development of the efficient peers leading the best practice. If necessary, these operators may require higher returns to share the benefits. Here we refer to the 'superefficient' operators that contribute to frontier shift (TC) by implementing new technology and operational regimes earlier than others. NRA may consider sharing the benefits of this overperformance as e.g. done in some national regulation (DE, NO) using sharing schemes for superefficiency or frontier shifts.

6.4 NRA needs for AEC information

6.26 The NRA needs can be summarized in two points:

6.27 First, to perform their recurrent medium-term tasks under EU Regulation with respect to setting target and allowed revenue, NRAs need the AEC to inform about the Totex efficiency of the current assets and operations of their TSOG, correcting for structural differences, heterogeneity and uncertainty. As the upcoming period will be economically strained for TSOG, shippers and users alike, it is important that the metrics and methods chosen are unanimously accepted as techno-economically sound and relevant in their assumptions and interpretations.

6.28 To deliver this, we recommend an *execution-oriented* model that is focused on the current asset structure and capable of identifying best practice in the short and medium run, also for a smaller set of comparable TSOG like in a given reference year prior to a regulatory period.

6.29 NRAs will also need to provide practical guidance and incentives in asset restructuring and repositioning under the national regulatory regimes they operate, if necessary, based on annual or short-term data. This means that the execution-oriented model should guarantee production of consistent estimates for key metrics for varying data sets and horizons. To reinforce the validation of findings from the models and metrics, NRA and stakeholders are in favor of combining models (see section 8.13 below).

6.30 Second, NRAs need to monitor, guide and incentivize the path of TSOG towards the objectives in the energy transition to achieve the energy supply security and decarbonisation goals. For this objective, the application of execution-oriented models needs to be accompanied by a dynamic assessment of changes to the asset base and the scale of operations for TSOG in Europe. This implies studying the development of cost and asset stocks for changing demand over time, as well as learning and adopting best-practice method for operations of existing assets. The grid users in the stakeholder consultation as well as the Expert Review stressed the importance of a dynamic evaluation of grid use.

6.31 To support NRAs in this dimension, we recommend a *dynamic utilisation-focused* model that captures the transformation of European TSOG and the effectiveness of the system to meet energy supply demands over time. This model should be capable of addressing uncertainty, dynamic events and of investigating the deeper causes and limits to cost efficient operations and investments.

7. Efficiency metrics

In this Chapter, we review metrics that could be calculated from AEC, the needs from NRA and TSOG, and propose a set of deliverables from the AEC to fulfill these needs in an informative manner.

7.1 Efficiency and effectiveness

7.01 Efficiency in network operations refers to how well resources (e.g., labour, pipelines, and compressors) are used to deliver grid services. TSOGs combine multiple inputs to deliver outputs. Inputs can be measured in units (staff, km of pipelines, etc.) or in monetary units. In the latter case, it is usual to consider two inputs OPEX and CAPEX or a single input TOTEX. The benchmark compares how TSOGs use inputs to deliver outputs. Efficiency measures are usually expressed as an efficient score in between 0 and 1 (or 100%), 1 being the highest possible value. In order to understand and interpret the (relative or absolute) scores from the efficiency comparison, it is best practice to decompose the results into several different dimensions or metrics:

1) Partial efficiency, unit cost UC – *cost per asset unit*
2) Technical efficiency, TE – *whether resources are minimized at given output*
3) Cost efficiency, CE – *whether costs are minimized at given output*
4) Scale efficiency, SE – *whether scale of operations is optimal*
5) Allocative efficiency, AE – *whether the right mix of resources is used*
6) Dynamic efficiency: Technological Change (TC) and Efficiency Change (EC) – *how efficiency changes over time*

7.02 Depending on estimation method used some of the measures may not be readily calculated, but most established frontier methods can produce all metrics. For all measures, except for dynamic efficiency, the metric can be computed with a cross-section, that is a unique year of observations. For dynamic efficiency and to measure how efficiency changes over time, it is necessary to have observations spanning over several years. Hereafter, we briefly present the six different metrics.

7.2 Unit cost, UC

7.03 An intuitive approach is to calculate a ratio of (some) cost to (some) output, e.g., cost per delivered nm3 or investment cost per km of pipeline. These measures are called *partial efficiency* measures or simply *unit costs.* The advantage of UC metrics is its simplicity. The drawback is that the partial focus renders interpretation difficult.

7.04 Unit cost measures are only reliable if the cost element is standardized and only applies to the output selected (e.g., maintenance costs for compressors without overhead cost related to compressors). Unit cost metrics are partial, green field, not adjusted for scale and do not explain total expenditure. Thus, they cannot be directly applied in regulatory mechanisms for total expenditure such as revenue-caps or incentive-based WACC plans.

How can it inform regulation?

7.05 Unit cost measures, if correctly calculated and interpreted, may provide useful information for NRA to monitor investment plans, values for new investments in hydrogen-ready equipment, repurposing or asset transfers from TSOG to other operators. Unit cost can also be used qualitatively to identify specific best practice and quality differences among TSOG. Given that unit cost metrics only concern direct investment values without considering environmental factors, associated investments or other system features, the idea of using minimal values to define some type of best practice is not methodologically correct.

7.3 Technical efficiency, TE

7.06 In addressing multiple inputs and outputs, but without involving prices, we obtain *technical efficiency TE* as the ratio of the actual resources to the minimal volume of resources for the same level of output. To illustrate, suppose that there are two inputs TSOG staff (FTE) and capital invested in transmission assets and one output. In Figure 7-1 we represent the input/output ratios for several operators (US interstate gas transmission system operators, 1987, FERC, data from Granderson and Linvill, 1999)

7.07 *The ideal in productivity would be to consume as few resources per output as possible. In Figure 7-1 this corresponds to the direction of the origin: lowest capital intensity and lowest staff intensity (or highest staff productivity). Consider TSOG D with the lowest staff intensity and third lowest capital intensity. No TSOG has lower staff intensity; no TSOG in the sample can match the staff productivity of TSOG D. Same applies for TSOG F but for grid assets. The TSOGs D and F, are therefore called technically efficient units (100%) since they operate with minimal resources for their output and profile. The efficient units are called peers, and their combination is called the (efficient) frontier. The frontier can be estimated with an econometric method; this point will be discussed below.*

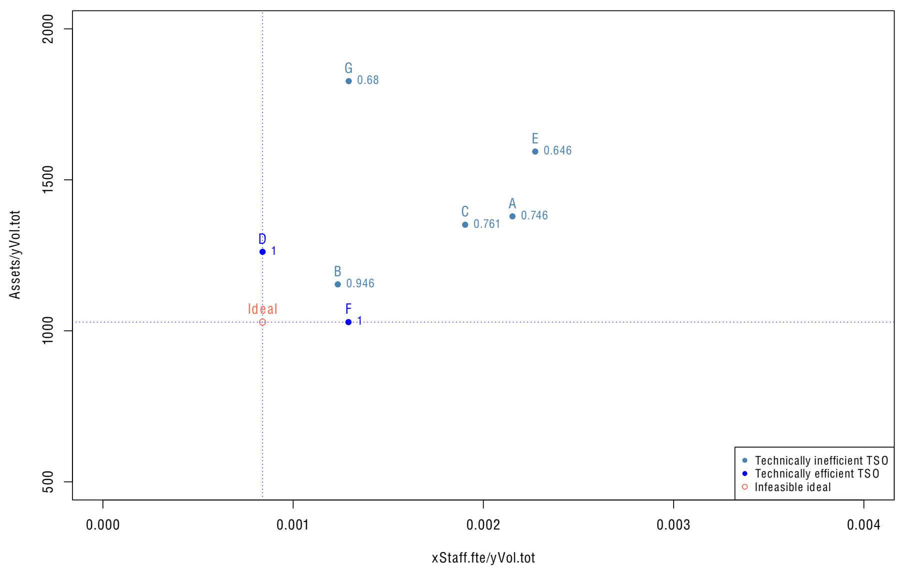


Figure 7-1. Technical efficiency TE (two inputs, per volume delivered), US-TSOG, 1987.

7.08 *All the other TSOGs than D and F in Figure 7-1 are technically inefficient (TE < 100%) since we can obtain their output at lower resource use by comparing with one or several of the efficient peers. E.g., TSOG B is using about 5.4% more resources than optimal, that is it has a technical efficiency of 94.6%. TSOG G and E are far from the frontier with technical efficiencies that are 68% and 65%,* *respectively. The technical efficiency holds for any input prices, meaning that whatever the prices and ratios of prices between capital and labour, the partial productivity for D and F, respectively, holds. Thus, as will be shown below, it may not be economically (cost) rational to be the most labour productive if that comes at a high price in substitution to capital (replacing labour), but the TE metric stays stable.*

How can it inform regulation?

7.09 Technical efficiency measures can be used for good technical information about the number and dimensions of assets, physical measures of operating services (hours, FTE) as well as physical measures for outputs and services. If possible, environmental conditions should be corrected for, as well as state (age) of equipment if relevant.

7.10 Technical efficiency TE means optimality for some input prices, that may not exist. Cost efficiency CE on the other hand, requires optimality with respect to the actual prices at the time of comparison. Thus, TE is an upper bound for the efficiency of a TSOG, the cost efficiency can never by higher than the technical efficiency. If the input prices are changing rapidly or have done so in the past, TE could be a relevant measure. However, for aggregate inputs as Opex and Capex, the input prices are limited to the cost of capital vs cost of labour and services, which can be addressed with a two-input model.

Limitations of use

7.11 There are two limitations to TE. First, high dimensional TE-models rapidly lose the capability of distinguishing among efficient operations. Moreover, TE models allow for extreme trade-offs between inputs that may be unreasonable. Second, the decomposition of TOTEX into TE for gas transmission would require data on technical assets (which indeed are available), as well as labour input and e.g. energy in physical terms, which only partially exist. The relevance for TOTEX is therefore indirect and TE is not directly applicable for AEC.

7.4 Cost efficiency, CE

7.12 *Cost efficiency* (CE) in gas transmission operations measures how well a TSOG is minimizing costs while delivering gas services. It looks at whether the TSOG is using the lowest possible cost mix of resources to achieve its services. Instead of measuring inputs in physical units (as in TE), inputs are measured in monetary units.

7.13 ***High cost efficiency:*** *A TSOG optimizes spending on operations (opex) and investments (capex), reliably transporting gas at lowest possible cost.*

7.14 ***Low cost efficiency:*** *A TSOG might have overpaid staff, expensive external services, and/or poor procurement strategies, leading to higher investment cost (capex) and operating expenditure than necessary for the same level of service.*

7.15 **Key factor – the right input mix:** Even if a TSOG operates efficiently in technical terms (e.g., no wasted resources), it can still be cost-inefficient if it relies too much on expensive inputs. For being cost efficient, a TSOG should use the right input mix.

TOTEX cost efficiency (CE)

7.16 The primary model for most regulatory benchmarking is a single-input model with total expenditure (TOTEX) as dependent variable.

7.17 We start with the example above while limiting the model to a single output for simplicity: volume transported. Thus, we can calculate the cost efficiency for the sample by their ratio to the peer F in Table 7-1 below. We note that TSOG F has the lowest Totex per output (133 USD/(1000)Mcf).

Table 7-1 Example Cost efficiency CE (US TSOG, 1987)

TSO	A	B	C	D	E	F	G
Totex	70953591	59307514	212926194	148387173	83480115	153155352	89512920
Totex/Mcf	219	173	222	177	269	133	315
CE	0.606	0.766	0.598	0.748	0.494	1	0.422

7.18 The interpretation of the CE-score is straightforward: if TSOG D were fully cost-efficient it would have a cost per output of 133, that is 1-0.748 = 25.2% lower than now. The actual Totex (real) for TSOG D is 148,387,133 USD (Table 7-1), meaning that the savings potential is 148,387,133*0.252 = 37,393,558 USD.

7.19 Now, consider the frontiers in Figure 7-2 below. The frontier for F is the least steep, meaning that the cost increases less per volume transported. Thus, compared to F, all TSOGs are cost inefficient, the distance to the red frontier indicates the magnitude of inefficiency. However, as we will discuss below under Scale efficiency, TSOG F is the largest in the sample and smaller TSOG may not be able to increase their volume due to restrictions of geography and concession areas (in the case of AEC). Note in this case that the total cost falls for the TSOGs G, E, A and B although the volume transported increases. This naturally explains by this group is dominated by TSOG B, delivering more volume at a lower absolute cost. On the other hand, TSOGs D, C and F are larger (in Totex) than B and can therefore not dominate TSOG B due to scale differences. This reasoning is behind the attention paid to the appropriate returns to scale in Section 7.6 below.

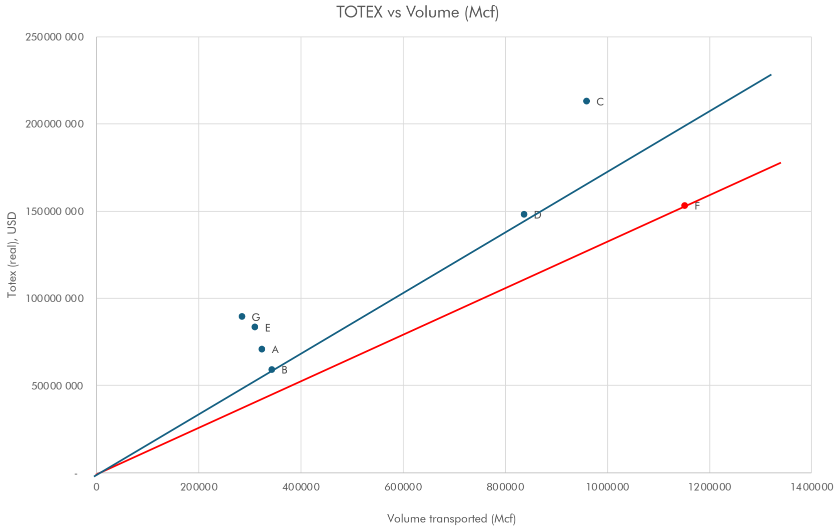


Figure 7-2. Cost efficiency (Totex vs volume), CE(1), TSOG-US, 1987.

7.20 *In TCB21, the mean cost efficiency for TSOG after outlier removal and environmental corrections was estimated to 84.3% - 86.3% (for 2020). In TCB18, the corresponding score was 79.1% (for 2017).*

7.5 Decomposed cost efficiency, CE(2)

7.21 We noted that for Totex, TSOG F was the most efficient and its example could inspire other TSOGs as an actual peer realizing an output per EURO that is higher than any other TSOG (in the sample). However, the question remains by what means TSOG F has achieved this standing. Is it due to better investments (Capex), higher staff productivity (Opex) or an optimal combination of both? For a TSOG with a limited investment budget, the answer to this question may be very important to gauge whether the results could be emulated or not. For this reason, to explore the sources of efficiency (or inefficiency), we decide to decompose CE into the two components Capex and Opex.

7.22 We denote this decomposed, two-input model, by CE(2) to differentiate it from the TOTEX single-input model (cf. Figure 7-2 above).

7.23 In Figure 7-3 we illustrate the same TSOG as above but now in terms of Opex (that is the sum of staff, service and fuel cost) and Capex (annuity, in real terms, adjusted for inflation and lifetime) per output (here simplified for illustration in the graph to a single output: gas volume transported in Mcf, yVol.tot). In the Figure, the Totex (sum of Opex and Capex) is marked above each point. The dotted lines show mean Opex and Capex levels (adjusted for inflation to the reference year).

7.24 The decomposed cost efficiency, compared to the Totex cost efficiency, allows us to better understand the sources of efficiency in the sector. E.g., as seen in Figure 7-3, TSOG D has the lowest staff intensity and consequently the lowest Opex per output, but more expensive assets than the peer. Likewise, TSOG F is efficient uniquely due to a high utilisation factor for its assets, i.e., its size: the TSOGs A, B, D are more OPEX-efficient than F. Thus, F is not the optimal source for learning Opex efficiency, but perhaps the best example of procurement and construction. Further data analysis might reveal the underlying sources for the productivity differences of E and G, although the exercise here is not realistic in a single-output case without any environmental factors. Recall, however, the caveats we made before concerning the scale of operations (i.e. investments): it would not be relevant (or fair) to compare the investment cost per volume for a smaller unit. Finally, note that the comparison here for visibility is simplified to a single output (volume), in a real model this is not the case.

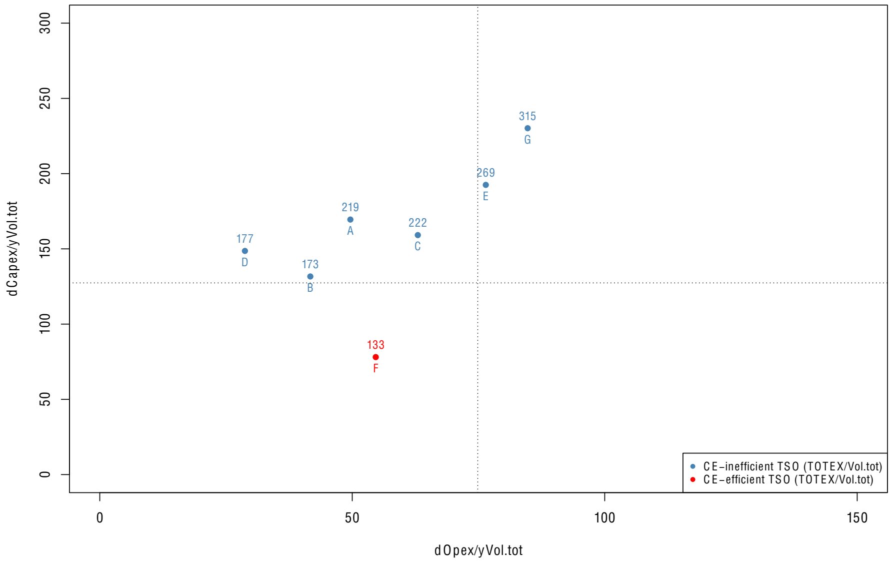


Figure 7-3 Cost efficiency (decomposed), (Capex and Opex), CE(2), TSOG-US, 1987

7.25 Cost efficiency requires reliable cost information for Opex but not necessarily split in physical amounts (hours), to be standardized for currency, taxes and fees, inflation, labour cost differences, quality differences, overhead allocation and out-of-scope activities. For Capex, the information should include gross investment costs and years by type of asset, to be standardized for currency, inflation, taxes and fees, quality differences, techno-economic lifetimes. The capex is annuitized with a standardized real interest rate. These issues will be discussed further in Section 6.

7.26 Aggregate CE is the primary and most relevant metric for incentive regulation as it directly links to value (output) per total expenditure. CE gives an estimation of the budget saving that can be achieved by adopting best practices. It can be used as Xi-factors in revenue cap formulas. The extraction of inefficiency could be split over a period that corresponds to the catch-up speed of the TSOGs. The decomposition in two components, Opex and Capex, as in CE(2) complements the information from CE for the NRA interested in the sources to inefficiency that can have different time horizons for adaptation or constraints in terms of negative scale changes.

7.6 Scale efficiency, SE

7.27 Scale efficiency (SE) in gas transmission operations refers to how well a TSOG is utilizing its size to minimize costs. It measures whether the operator is working at an optimal scale—neither too small nor too large—to achieve the lowest possible cost per unit of gas transported. Recall the sources for capital efficiency in Figure

7-1 Figure 7-2 above, where the largest unit sets the frontier for all other due to utilisation factors.

7.28 Under **increasing returns to scale (IRS), or non-decreasing returns to scale (NDRS),** a TSOG with a smaller network may have higher per-unit costs because it cannot spread fixed costs over a large enough volume of transported gas. Conversely, under **decreasing returns to scale (DRS),** a TSOG that has grown beyond an optimal size may experience coordination inefficiencies, higher maintenance costs, and diminishing cost advantages. At the optimal scale **(most productive scale size – MPSS)**, the TSOG operates at a size where it fully utilizes its infrastructure and workforce, achieving the lowest cost per unit of gas transported. Under **Constant returns to scale (CRS),** there is no inefficiency controlled for by size.

7.29 Return to scale is a characteristic of the *production technology*. Following the literature, past results and experiences in benchmarking studies for electricity and gas transmission, the expected empirical outcome of the efficiency analysis is either CRS or IRS/NDRS.[5] Scale efficiency is calculated in the same models as for TE and CE, no other information is necessary. Normally, SE is validated in studies to determine scale issues.

7.30 Now, revisit the earlier example using the assumption of increasing (IRS or NDRS) rather than constant returns to scale. The cost efficiency results are listed in Table 7-2 below along with the difference in CE due to scale (CE-IRS – CE-CRS). As seen, it is now revealed that we have two peers in the group: B and F. The smaller group (A,B,E,G) forfeits about 22-24% due to scale differences. The larger inefficient TSOGs (C, D) do not have the same problem, the difference is marginal compared to CRS.

Table 7-2 Cost efficiency under constant (CRS) and increasing (IRS) returns to scale.

TSO	A	B	C	D	E	F	G
CE-CRS	0.61	0.77	0.60	0.75	0.49	1.00	0.42
CE-IRS	0.84	1.00	0.61	0.79	0.71	1.00	0.66
Scale losses	0.23	0.23	0.02	0.04	0.22	-	0.24

How can it inform regulation?

7.31 As illustrated in the example above, SE is important in regulation as it separates the sources of inefficiency with respect to managerial (technical) and scale, which may not be controllable. A larger TSOG may benefit from better prices in procurement due to bigger and more frequent investments, a larger overhead may include specialists in reengineering and ICT that could perform more optimisation, fixed costs for indivisible assets (metering, regulators) may be split over a larger grid size, etc. It also informs about potential gains (to be shared) through structural changes (expansion, mergers or split-ups), challenging transfers of concession areas. E.g., smaller TSOGs could gain in scale by mergers to pool procurement and construction tendering together for similar project, same for unique indivisible resources. Likewise, a too large TSOG with diseconomies of scale due to internal bureaucracy may decentralize certain services, like maintenance, to sub-entities like regions or functions, emulating the optimal scale of operations. Finally, notice that SE alone

[5] Previous benchmarks have confirmed IRS/NDRS or CRS as returns to scale for gas transmission. In TCB18 and TCB21, IRS/NDRS was used to protect smaller TSOG from targets imposed by larger target TSOG, thereby forcing expansion or merger.

cannot be used in revenue regulation, it is a complement to get the correct adjustments for TE and CE.

7.32 *In TCB18 and TCB21, the scale assumption for TSOG was non-decreasing returns to scale (NDRS), reflecting large scale heterogeneity among the participants.*

7.33 For AEC we propose that scale efficiency be investigated using flexible assumptions, CRS, IRS-NDRS and VRS, in multiple models and for both parametric and non-parametric methods. The returns to scale finally used in AEC will then follow from the empirical results obtained.

7.7 Allocative efficiency, AE

7.34 In gas transmission, allocative efficiency (AE) means that the operator uses the *optimal mix* of inputs (e.g., labour, compressors, pipelines, fuel) given their prices, i.e., costs are minimized for the achieved output level.

7.35 Even if technical efficiency (TE) is perfect (no waste), allocative inefficiency arises when resources are used in the wrong proportions relative to their economic costs e.g., using excessive compressor fuel instead of cheaper pipeline capacity to deliver the same gas volumes.

7.36 *Allocative efficiency*, AE, is the ratio of the cost with no waste to the optimal cost, a number between 0% and 100%. In Figure 7-4 below we graph the AE and CE for the five TSOG sampled.

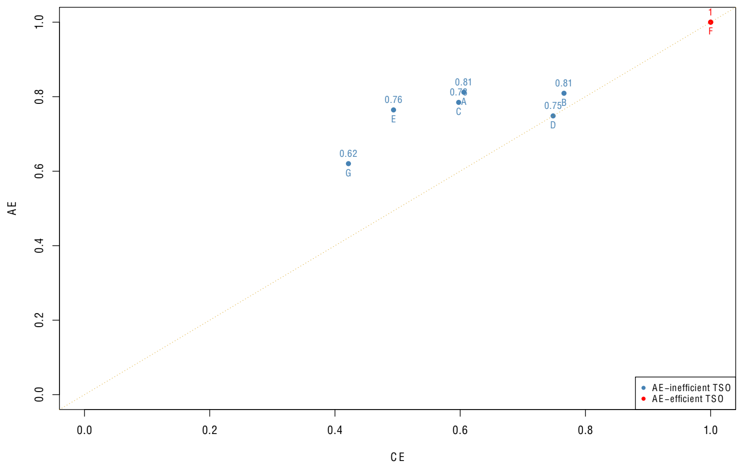


Figure 7-4 Cost efficiency CE vs Allocative efficiency AE, US TSOG.

7.37 To be fully cost-efficient, the unit also needs to be full allocatively efficient, AE=1, in addition to being fully technically and scale efficient. We then get the full expression for cost efficiency:

$$CE(y) = TE\ SE\ AE$$

$$(cost\ efficiency) = (no\ waste)(optimal\ scale)(optimal\ mix)$$

7.38 Thus, we note that a TSOG may be allocatively efficient (AE=1) and still cost inefficient (CE<1) if it suffers from waste of resources (TE<1) and/or operates at the wrong scale (SE<1).

7.39 For (A,B,E,G) in Figure 7-4 the lower AE is primarily due to scale effects: the smaller TSOGs cannot split capital costs over the same volume. However, in Figure 7-4, a different situation is shown for TSOG D. This TSOG is technically efficient (see Figure 7-1) and large (see Figure 7-2). However, it has only 71% in allocative efficiency in Figure 7-4 since it suffers from the wrong input proportions (too high capex intensity compared to opex), but no waste and almost no scale losses (Table 7-2).

How can it inform regulation?

7.40 AE is useful to separate the effect of input mix changes that could take longer time (Capex…) from Opex changes, thereby analyzing cost efficiency improvement options. Furthermore, it can estimate the sensitivity to standardized vs local prices. However, allocative efficiency is only one necessary condition for cost efficiency, permitting an analysis of the input or output profile of the TSOGs, but it cannot be used in separation to determine total expenditure. For this (see formula above), we also need the technical efficiency TE and the scale efficiency SE results. In practice, AE is used to review similarities of TSOGs in terms of profile, a screening metric to help in the interpretation of scores. In particular, since the scale effects in gas transmission likely are permanent and noncontrollable, the usefulness of AE in separation is limited for regulation.

7.41 We do not recommend to use AE as a primary metric for AEC since it incorporates the scale effects that are likely to be substantial in the sector.

7.8 Dynamic efficiency, DE

7.42 Dynamic efficiency (DE) measures how well a TSOG improves over time. Unlike static efficiency, which looks at performance in a single year, dynamic efficiency tracks whether a TSOG is keeping up with industry advancements and reducing inefficiencies over multiple years. The data requirement differs for computing dynamic efficiency as several years of data are needed to compute the evolution of efficiency over time.

7.43 TSOGs continuously invest and try to improve their practices. The use of dynamic efficiency allows to separate technological progress from catch-up. Efficient operators will push the frontier and achieve performance that were previously non-reachable. Other operators will catch-up, adopt best practices and improve their efficiency. In a world where technologies are continuously changing, it is important to identify in a dynamic model, the impact of the technological progress on the frontier and how the relative efficiency of the TSOGs evolve over time.

7.44 There are **two key components** of dynamic efficiency, TC and EC.

Technological Change TC (Frontier Shift)

7.45 A technological change or frontier shift (TC) happens when the best-performing TSOGs peers push the efficiency frontier forward by adopting new technologies, optimizing processes, or improving infrastructure. Even top performers need to keep innovating to stay on the frontier.

7.46 *Example: the common productivity requirement (Xgen) is directly based on the frontier shift TC. It is integrated in the revenue cap for all operators to create a sharing ex ante of past productivity gains from e.g., AI and new construction methods.*

Efficiency Change EC (Catch-Up)

7.47 EC refers to how well an individual TSOG closes the gap between itself and the best-performing TSOGs over time. When technology improves, some TSOGs may lag if they adapt more slowly than their peers.

7.48 *Example: the catch-up speed in the sector can be applied as a lower bound to the efficiency requirement per year for inefficient operators to level the playing field.*

7.49 Figure 7-5 illustrates the decomposition of the two metrics in a simplified example.

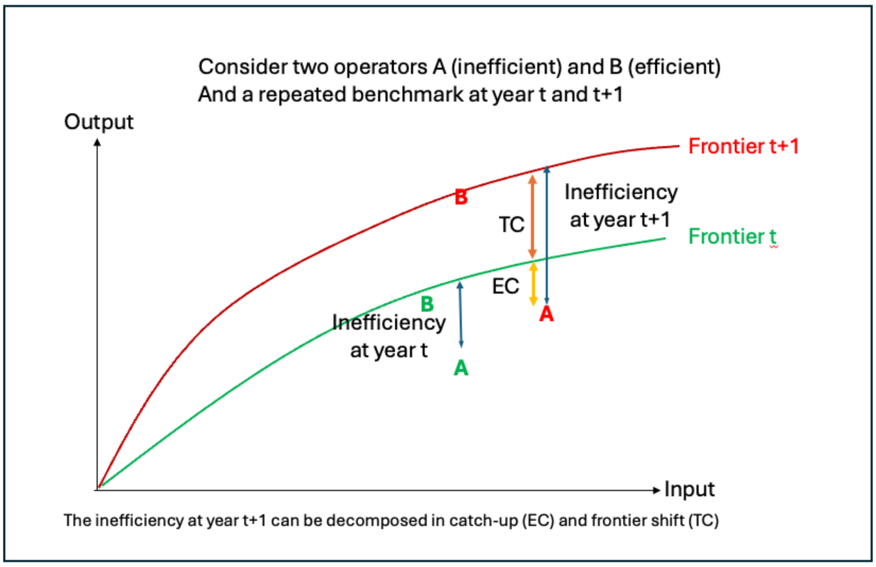


Figure 7-5 Dynamic efficiency and decomposition in EC and TC

7.50 In regulatory models, technological change (TC) is used to compute the industry-specific efficiency factor (X or Xgen), efficiency change (EC) is used to compute the firm-specific one (Xi).

7.51 The overall productivity development, the *Malmquist index* (M) of the sector is the product of frontier moves TC through innovation and catch-up EC by inefficient firms:

$$M = TC * EC$$

(sector productivity) = (frontier shift) * (catch-up)

7.52 *In TCB21, the findings were very similar: across the two final models used over 2016-2020, the productivity development was close to 0%, the frontier shift (technical change) was 0.5%-0.9% per year, and the efficiency change was negative, -0.7% to -1.3% per year. This underlines the importance to challenge inefficient operators to improve, the efficient peers seem to do well internationally.*

How can it inform regulation?

7.53 The application of dynamic efficiency is straightforward in a regulatory framework using revenue caps: the TC frontier shift is used as the sharing parameter forward in the revenue cap: all operators are expected to meet this productivity requirement. Traditionally, the TC is called the *X-factor* defined above.

7.54 Regulators are facing some observations as mentioned in the earlier chapters:

1) Cost, investment and unit-cost changes over time are important for NRAs in their regulations that often are different.
2) Fully efficient TSOG should be provided incentives to continue process improvements.
3) Catch-up of lagging TSOGs should be analyzed to understand the reasons (opex, capex, staff, growth etc)
4) Dynamic effects are intuitive for external stakeholders.

7.55 Thus, the *dynamic efficiency analysis* is important for the NRA useability of the model results:

1) Important to choose a method with a sound and consistent dynamic decomposition of results, like in DEA and SFA.
2) This analysis integrates exogenous and contextual variables over times, such as regulation regimes, interest rates and demand (Battese and Coelli, 1992).
3) Dynamic analyses enable a better interpretation of TSOG performance since the analysis may differentiate between transient (temporary) and permanent (structural) inefficiency, see Filippini and Greene (2016) or Badunenko and Kumbhakar 2016).
4) Data definitions and metrics should be kept as constant as possible over time to allow for dynamic analyses and learning
5) AEC could initiate dynamic analyses by requesting not only the reference year, but the full period 2021-2025 to cover the entire period since TCB21. Future applications could also cover the time between AEC runs.

7.56 We propose to integrate and facilitate dynamic efficiency, including the use of earlier data from AEC to provide longer data series with consistent data definitions. See also previous section.

7.9 Conclusion

7.57 The following metrics are proposed for use in AEC:

7.58 The main efficiency metric is **Cost efficiency (CE)** for all operators. The use of CE requires a standardized definition of inputs across operators, and we discuss in the next sections the options that will ensure the cost comparability of all the TSOGs. The **decomposed cost efficiency CE(2)** in opex and capex-inputs is added for further understanding of the efficiency profile of the TSOGs.

7.59 Use of **Scale efficiency (SE)** as a secondary metric. In order not to penalize operators with different scale, imposed by the size of country or concession, the assumptions for SE will be tested with the options for constant, increasing and variable returns to scale[6] In increasing (or non-decreasing)[7] returns to scale, used in the last two CEER projects, scale diseconomies are corrected for smaller TSOG. E.g., if the analysis shows that the technology of the sector exhibits increasing returns to scale, then the scores will be calculated using IRS (NDRS) as assumption. If another return to scale is shown to prevail, this is recommended to be applied.

[6] CRS, IRS/NDRS and VRS.

[7] The difference between increasing (IRS) and non-decreasing (NDRS) returns is that NDRS also includes (sections) of constant returns to scale (CRS).

7.60 **Dynamic efficiency DE**, decomposed in efficiency change (EC, catch-up) and technological change (TC, frontier shift) for the most efficient units. Dependding on model, DE may also include the evolution of Capex, Opex and Totex as a function of volume and peak load, including contextual and environmental factors over time. For the dynamic efficiency analysis, it is necessary to use several years of data for the benchmark (the so-called panel data).

8. Methods and models

In this Chapter we review existing major methods for benchmarking and present the choice of methods for AEC and their application. Two main methods are proposed for the AEC with complementary strengths. The Chapter also proposes three models for estimation and the appropriate methods for their calculation. Illustrations are provided using gas transmission data from USA.

The reader familiar with efficiency analysis may skip sections Efficiency and productivity measurement methods

8.1 Outline

8.01 This summary is non-technical and concise. For an in-depth discussion of the techniques used with examples, see e.g. Coelli et al. (2005), Bogetoft and Otto (2010), O'Donnell (2018), Sickles and Zelenyuk (2019).

8.02 There are different available frontier methods that can be used for the AEC:

1) Parametric methods (e.g. SFA, COLS, MOLS),
2) nonparametric (e.g. DEA, Robust DEA, CrossEfficiencyAnalysis CEA),
3) semiparametric (e.g. StoNED),
4) engineering methods (e.g. reference networks),
5) index number approaches (e.g. unit cost).

8.03 These methods are evaluated for the purpose of the AEC, based on the following restated preliminaries from earlier chapters on energy transition, regulatory needs in section 6.4, and metrics in section 7.9:

1) The due diligence requirement to select methods with a particular attention to their techno-economic soundness for the purpose of AEC as not to produce information that could be misinterpreted or distorted (see §6.27).
2) The need from NRA to address in detail cost efficiency for TSOG in a given year for an existing asset based (the execution-focus), with full control of heterogeneity (see §6.28).
3) The common interest of ACER, NRA, TSOG and other stakeholders to use methods that produce results that correspond to best possible scientific practice in the area, along with the relevant measures to estimate and control for intrinsic uncertainty in these calculations (see §6.29).
4) The need from NRA to analyze dynamic utilisation phenomena using models that can cater uncertainty and exogenous variables, potentially stochastic (see §6.30-6.31).
5) The priority for metrics directly relevant to the Regulation, that is, cost efficiency CE, CE(2) and dynamic efficiency DE (see Section 7.9).

8.2 Productivity, efficiency and effectiveness

8.04 Assume that a TSOG k has reported inputs $x(k)$ and outputs $y(k)$. We recall the three conceptual definitions of productivity, effectiveness and efficiency

1) **Productivity** is "converting inputs to outputs" (absolute rate of resource conversion).

$$Productivity(k) = \frac{y(k)}{x(k)}$$

2) **Efficiency** is about "doing things right" (applying best practice to minimize waste and maximizing resource use). This metric is fundamentally linked to the concept of observed *best practice*, meaning the productivity (or cost) for a set of structurally comparable operators.

$$Efficiency(k) = \frac{Productivity(k)}{\max \{Productivity\}}$$

3) **Effectiveness** is about "doing the right things" (achieving desired outcomes, regardless of resources used). Here we additionally assume a subset output $Y(k)$, like share of renewable energy in a country compared to a target objective (TargetY) for this primary superior output.

$$Effectiveness(k) = \frac{Y(k)}{\text{TargetY}}$$

8.05 If TSOG *k* achieves a planned conversion of its network in coordination with interfacing networks, ports and generators, at a given time, it may be said to be effective. However, the operations might not neither productive (if a lot of assets are unused), nor efficient (if the achievement of the objective has been made at excessive investment and operating costs). In practice, effectiveness for TSOG is expressed as a binary (true/false) condition for fulfilling their statutory role, safety standards and compliance with national and European regulation.

8.06 In AEC, we are interested in measuring productivity over time and for a reference year for the purpose of estimating (cost) efficiency. Given the complexity and multi-output character of gas transmission operations, a mathematical efficiency analysis method is necessary.

8.3 Benchmarking methods

8.07 *Efficiency measurement methods* differ along several key dimensions: how they treat *randomness and noise*, whether they assume a *specific functional form* for the cost or production relationship, the *degree of flexibility* in modelling the frontier, and whether they rely on *empirical performance data or engineering benchmarks.*

8.08 At one end of the spectrum, **parametric methods** specify a functional form and allow for stochastic variation, while **nonparametric methods** are purely data-driven and deterministic. **Semiparametric methods** combine features of both. Separately, **engineering methods** rely on technical standards or simulations, and **ratio-based approaches** use straightforward performance metrics like unit costs. Each method has strengths and limitations, and the choice depends on data availability, regulatory context, and the objectives of the benchmarking exercise.

8.09 The families are graphed on two major distinguishing dimensions, noise separation (e.g. data error handling) and model flexibility (e.g. technical assumptions needed on the cost function) in Figure 8-1 below.

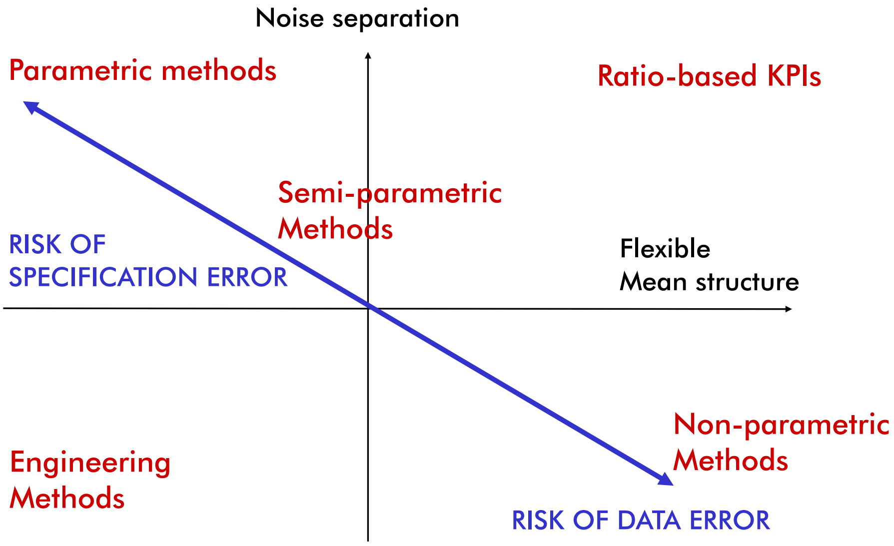


Figure 8-1 Families of methods vs noise separation and flexibility.

Engineering methods

8.10 Engineering or **bottom-up benchmarking** relies on technical specifications, engineering models, and process simulations to estimate what performance "should be" under ideal or best-practice conditions. For TSOGs, this may involve simulating gas flow dynamics, compression efficiency, and energy loss using engineering software or guidelines. These methods are particularly valuable when detailed operational data is available, or when heterogeneity in geography or network complexity needs to be explicitly accounted for. While they provide a deep technical benchmark, they are resource-intensive and may lack comparability across TSOGs unless standardized.

8.11 In Figure 8-2 below we illustrate a hypothetical engineering cost function: it is not based on the observations but on an *a priori* technical cost function. This means that the sample may have no efficient peers (here, TSOG 2 is inefficient), all efficient and/or some infeasible (the point to the extreme right). The validity of the engineering cost function entirely depends on the scope of the estimated system compared to the actual observations. In practice, brownfield estimations of cost are extremely difficult to make with precision, and the model can be very heavy to establish and maintain.

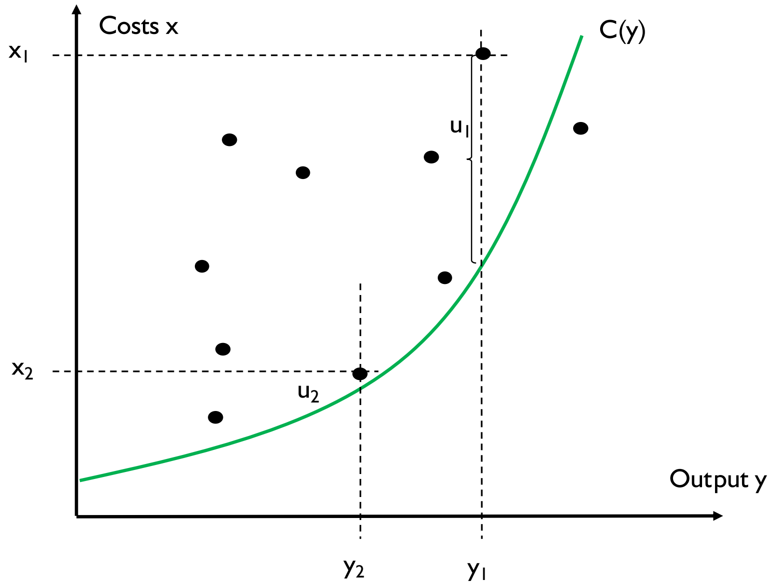


Figure 8-2 Engineering cost function.

8.12 Since engineering methods require substantial investments both to conceive and to operate, their usage is limited in network regulation. A precursor is Chile (cf. Rudnick and Donoso, 2000, Recordon and Rudnick, 2002) where electricity regulation has been managed by ideal network models since 1982. In Spain (cf. Grifell-Tatje and Lovell, 2003), the electricity DSO regulation is assisted by an engineering model (BULNES). For the period 2003-2009, there was an engineering model in operation for electricity DSO in Sweden (cf. Agrell and Grifell-Tatje, 2015). Finally, the German incentive regulation (ARegV) foresees the use of reference network models in the case a benchmarking (with DEA and SFA) cannot be performed with reliable results (e.g. due to size of data set or degree of heterogeneity).

KPI approaches

8.13 Unit cost benchmarking and Key Performance Indicators (KPIs) are widely used in regulation and performance comparison due to their transparency and ease of interpretation. These methods calculate **cost per unit of output** or **efficiency ratios**, such as:

1) €/GWh of gas transported
2) Operating cost per km of pipeline
3) Compressor fuel consumption per GWh transported

8.14 In TSOG benchmarking, unit cost measures serve as straightforward indicators of cost efficiency, especially in **cross-sectional comparisons (e.g., for 2024)**. They are often used by regulators and stakeholders for their simplicity and direct linkage to financial and operational performance.

8.15 **KPIs** extend unit cost analysis by incorporating **non-cost metrics**, such as:

1) Network utilisation rates
2) Maintenance cost (or any specific function) per km pipeline
3) Leak rates per km
4) Emergency response times

8.16 These indicators can be used in both **descriptive benchmarking** and as part of a **composite scoring system**, ranking TSOG by weighted performance across multiple dimensions, see e.g. Agrell and Wikner (2011) for the integrated partial efficiency (IPE) metric.

8.17 While accessible, unit cost and KPI-based methods have limitations:

1) They do not adjust for differences in operating environments (e.g., terrain, support, etc)
2) They lack a formal inefficiency model or stochastic treatment of data variation
3) Comparability is sensitive to accounting or reporting standards (simple ratios)
4) Ratios assume single output and constant returns to scale

8.18 KPI are partial measures and beyond the deficiencies in lack of control for scale, environment, multi-output etc, there are also obvious methodological flaws in using them for evaluating complex activities.

8.19 Consider the example in Table 8-1 as illustration. Two TSOG (A and B) are evaluated on a partial cost, maintenance cost, for two types of assets (pipelines MP and HP). E.g., TSOG A has maintained 20 km of MP pipelines for 10 k€, leading to a unit cost for maintenance of MP-pipelines of 10/20 = 0.50 k€/km. In this example, TSOG A is the "KPI-leader" since the partial unit cost is lower for both MP and HP pipelines.

Table 8-1 KPI partial Unit Cost Opex-M per pipeline, example.

	Opex-Maintenance		Pipelines(km)		UnitCost	
	MP	HP	MP	HP	UC(Opex,MP)	UC(Opex,HP)
TSOG A	10	10	20	40	**0.50**	**0.25**
TSOG B	2	21	3	80	0.67	0.26

8.20 However, when analyzing the total cost per pipeline in Table 8-2, the conclusion is reversed: TSOG B maintains 83 km at a cost of 23 k€ whereas TSOG A only manages to maintain 40 km at a cost of 20 k€. Thus, the partial KPI give useless information for the overall assessment of the TSOG[8]. Naturally, this would be even worse in a real data situation where A and B had different investment cycles, nominal values, varied overhead rates, staff costs and productivity and environmental conditions.

Table 8-2 KPI Total Unit Cost Opex-M per pipeline, example.

Total	Opex-Maintenance	Pipelines(km)	UC(Opex,km)
TSOG A	20	60	0.33
TSOG B	23	83	0.28

[8] The well-known finding is called the *Fox paradox* (cf. Karagiannis, 2012), the example is adapted from Bogetoft and Otto (2010, p. 10).

8.21 Even consultants for the industry raise serious critique to the use of KPI and partial measures in regulation, e.g. Shuttleworth (2005) on the use of separate Opex-benchmarking by OFGEM for the 1999-rate review:

"In 1999, Ofgem applied benchmarking to opex, and found that the Southern distribution network was on the frontier, whilst the Seeboard distribution network lay some way above it (Ofgem, 1999b). Ofgem gave Seeboard a revenue allowance for opex that demanded a rapid reduction. However, it became apparent during the review that Southern had favoured the use of capex (e.g. asset replacement) rather than opex (e.g. asset maintenance), whilst Seeboard had favoured opex rather than capex. The difference between their respective opex performance therefore reflected different choices over the mix of inputs, rather than differences in efficiency. In its final proposals, therefore, Ofgem had to undo the effects of its opex benchmarking, by awarding Seeboard additional revenue (for capex efficiency, among other items), and penalising Southern for excessive investment, thereby rendering the benchmarking largely irrelevant." Shuttleworth (2005, p. 314).

8.22 None of the best-practice guides include index numbers or KPIs stand-alone as a basis for regulatory cost assessments, see e.g., Lovell (2006), Lowry and Getachew (2009).

Parametric methods

8.23 Parametric efficiency analysis relies on a pre-specified functional form that defines the relationship between inputs and outputs, often assuming a particular distribution for inefficiency and noise. The most widely used parametric method is **Stochastic Frontier Analysis (SFA)** by Aigner et al. (1977) which separates random errors from inefficiency in performance. However, we start with the simplest parametric methods, corrected ordinary least squares (COLS) and modified ordinary least squares (MOLS), both derived directly from linear regression and with use in regulation.

Corrected Ordinary Least Squares (COLS)

8.24 The classical regression (OLS = Ordinary Linear Regression) framework is the basis for **Corrected Ordinary Least Squares (COLS)** by assuming that any inefficiency is deterministic. After estimating the cost or production function using OLS, COLS **shifts the regression line upward (or downward)** so that the most efficient observation lies on the estimated frontier. This shift assumes that **all deviations from the frontier reflect inefficiency**, effectively treating the residuals as inefficiency. While simple and transparent, COLS do not allow for statistical noise and may therefore **overstate inefficiency** in the presence of random variation. It is often used as a benchmark or diagnostic tool in regulatory settings.

8.25 In Figure 8-3 below, TSOG 2 is the only efficient unit in the sample and TSOG 1 has a slack term of u_1 that explains all the difference.

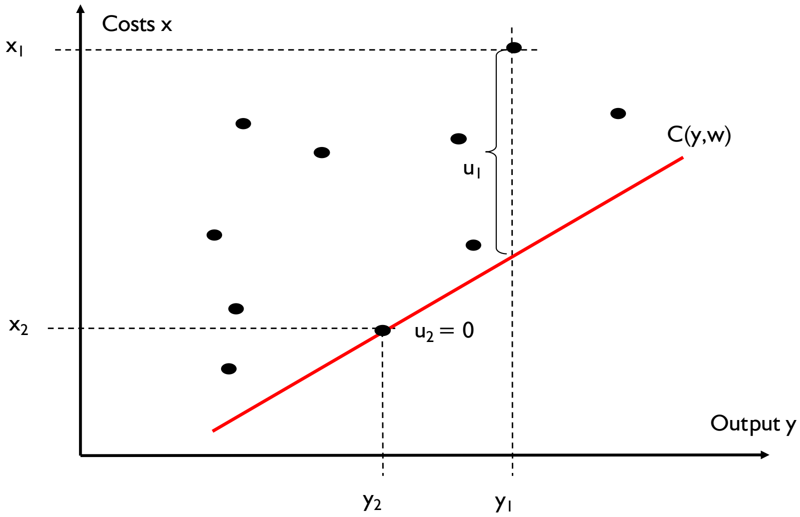


Figure 8-3 Cost function in COLS.

8.26 COLS (Richmond, 1974) is "classical" and simple method with an initial popularity. OFGEM used COLS for their 1999 tariff reviews and Haney and Pollitt (2012) report that NRA in Finland, Portugal, Poland and Austria were testing or considering COLS in 2010 or before. However, the method has been severely criticized (Shuttleworth, 2005) for its extreme dependency on a single peer, lack of flexibility and methodological inconsistency (error estimate for slope but deterministic frontier). In consequence, the method is no longer in regulatory use.

Modified Ordinary Least Squares (MOLS)

8.27 Modified Ordinary Least Squares (MOLS) improves upon COLS by introducing a basic adjustment for statistical noise. Instead of assuming all deviation from the OLS frontier is due to inefficiency, MOLS subtracts an estimate of the statistical noise (typically the standard deviation of the residuals) from the COLS shift. This results in a more conservative frontier that recognizes some part of the residual variation is random rather than purely inefficient. MOLS is a pragmatic compromise between the simplicity of COLS and the statistical rigor of SFA, making it useful when sample sizes are small, or data quality is limited.

8.28 In Figure 8-4 below, TSOG 1 is labelled as inefficient with slack u_1 compared to a frontier that is shifted upwards with a standard deviation *v*, meaning that the efficient TSOG 2 now has part of its performance explained as a result of random events, or simply "luck".

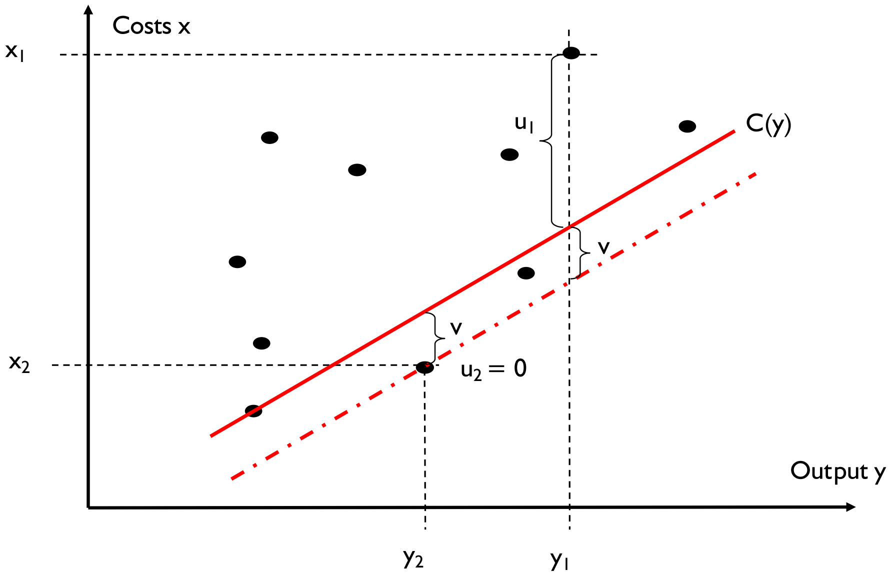


Figure 8-4 Cost function in MOLS.

8.29 MOLS addresses the obvious inconsistency of COLS while staying simple as regression expression without separation of inefficiency and noise. Agrell and Bogetoft (2017) reports a certain popularity for the method in DSO-regulation in Europe with four NRA using or testing it in 2016, thereof at least Austria is maintaining it in combination with DEA for DSO reviews.

8.30 MOLS is rarely subject of scientific studies and comparisons, like the methods DEA and SFA, as the method is a "little sister" to the more powerful sibling SFA below that offers a statistical estimate for the inefficiency term, separate from the noise.

Stochastic Frontier Analysis (SFA)

8.31 **Stochastic Frontier Analysis (SFA),** Aigner, Lovell and Schmidt (1977) is a fully parametric method that models both **inefficiency and random noise** in the estimation of a production or cost frontier. SFA assumes a specific functional form for the underlying relationship and decomposes the residual into two parts: a **two-sided random error** (representing noise, measurement error, or shocks) and a **one-sided inefficiency term** (reflecting deviation from the best practice frontier). This allows SFA to statistically separate managerial inefficiency from uncontrollable variation. SFA is especially powerful when applied to **panel data**, enabling time-varying inefficiency estimation and control for firm-specific heterogeneity. However, it requires distributional assumptions (e.g., half-normal, exponential)[9] and can be sensitive to model misspecification.

8.32 In short, the cost function in SFA is more advanced with three elements:

$$C(y) = g(y) + v + u$$

[9] The inefficiency is coming from an one-sided (positive) random distribution, whose parameters (mean, variance etc) are estimated from the data, but the family of distribution (half-normal, exponential, gamma) need to be defined by the analyst prior to the runs. The assumptions can be tested using different or combined versions, where the significance of the inefficiency term indicates the fit of the distribution.

where u is a one-sided inefficiency term (typically half-normal, truncated normal, exponential), v is the random noise (symmetric, typically assumed normal), and $g(y)$ is cost function with arguments y in form of an output vector. The cost function g can have any shape, (linear, loglinear, quadratic…), the terms and coefficients in g will be estimated using the data.

8.33 The elementary form of SFA is easily extended to consider dynamic effects (by year or period), environmental factors (fixed or dynamic), interaction effects (environment-output), explanatory factors (regime, tenure, …) and other extensions. The computation of the SFA model may also be made by different methods, notably Maximum Likelihood (ML) or Method of Moments (MOM).

8.34 In Figure 8-5 below, the cost function is estimated as C(y) and the inefficiency for TSOG 1 is split into (i) the general noise term (v_1) and the random inefficiency effect (u_1). Typically, the SFA scores are higher than corresponding non-parametric scores for inefficient units, since noise is explaining part of the observation, but lower than non-parametric scores for efficient units, since the cost function is not as flexible. In a deterministic model, the difference $x_1 - C(y_1)$ would be classified as inefficiency, in SFA it is only the term u_1.

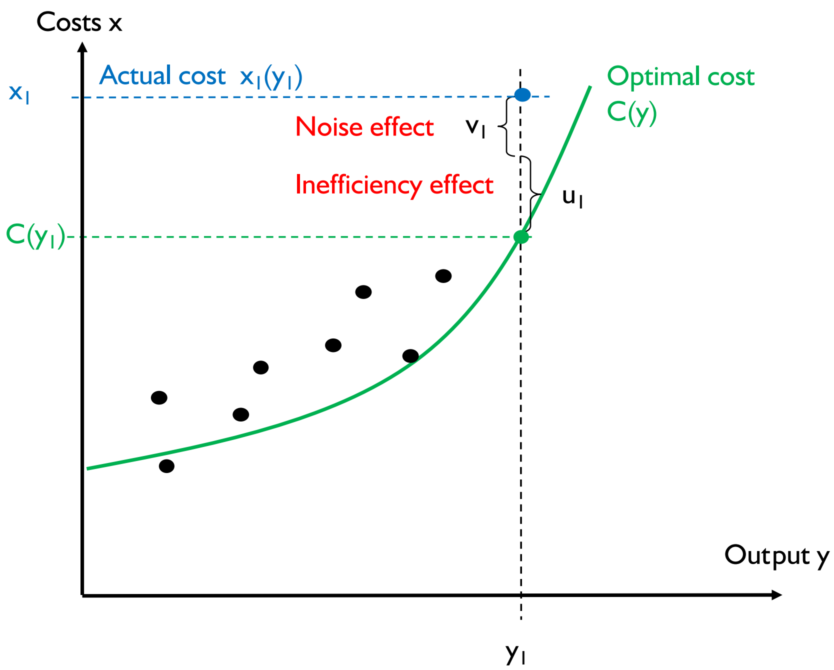


Figure 8-5 Cost function in SFA.

8.35 In the context of TSOGs, SFA is particularly relevant when modelling energy transport under external factors such as weather volatility or regulatory shifts that introduce noise into cost or output data. SFA is an ideal platform for controlling heterogeneity due to environmental factors (cf. Coelli et al., 2003) but also addressing endogeneity (inefficiency linked to model parameters), cf. Hou (2025) for an example in energy networks.

8.36 SFA has two principal drawbacks: (i) it relies on (non-trivial) assumptions regarding the distribution of the inefficiency (usually half-normal or gamma distributed, see footnote above), and (ii) it performs poorly for small data sets, especially with low noise (i.e. good data quality or deterministic variables). In the latter cases, Ahn et al. (2023) reports infeasibilities for SFA from around 23% of the cases with 25 observations, 11% with 50 observations and 0.5% for 100 observations (1 input and 3 outputs). For these reasons SFA is normally applied to panel data (see above) or to large crossections (e.g the German energy networks) with adequate model specifications. A further discussion of the sample size impact on SFA (and DEA) is found in section 8.8.

Nonparametric methods

8.37 Nonparametric methods like **Data Envelopment Analysis (DEA)** (Charnes et al., 1978) enable the construction of efficiency frontiers based solely on the data, without assuming a functional form. These are deterministic methods that attribute all deviation from the frontier to inefficiency. DEA is particularly useful for TSOG benchmarking when there's a need to evaluate relative performance across operators using a wide range of input and output variables—such as pipeline length, compressor power, and connection points—without imposing restrictive model assumptions. With panel data, **dynamic efficiency** or **window analysis** can be used to track performance evolution. However, nonparametric approaches are sensitive to outliers and measurement error, and they do not account for statistical noise.

8.38 The cost function in DEA is simple and requires no assumption of shape:

$$C(y) = C^*(y) + u$$

where $C^*(y)$ is a piece-wise linear cost function with arguments y in form of an output vector and u is the inefficiency (simply the difference $u = C(y) - C^*(y)$). The estimation of the optimal cost is done by linear programming by enveloping the observed cost-output combinations (see e.g. Bogetoft and Otto, 2011).

8.39 In Figure 8-6 below, the cost function in red (solid is with variable returns to scale, dashed is with constant returns to scale) would distinguish TSOG 2 as full efficient and TSOG 1 as inefficient with a slack that depends on the returns to scale. Note that for two or several outputs, the slopes (marginal costs) can vary in DEA: for instance, a transit TSOG can have a different marginal cost for pipeline expansion then a TSOG with a high number of connection points, for instance.

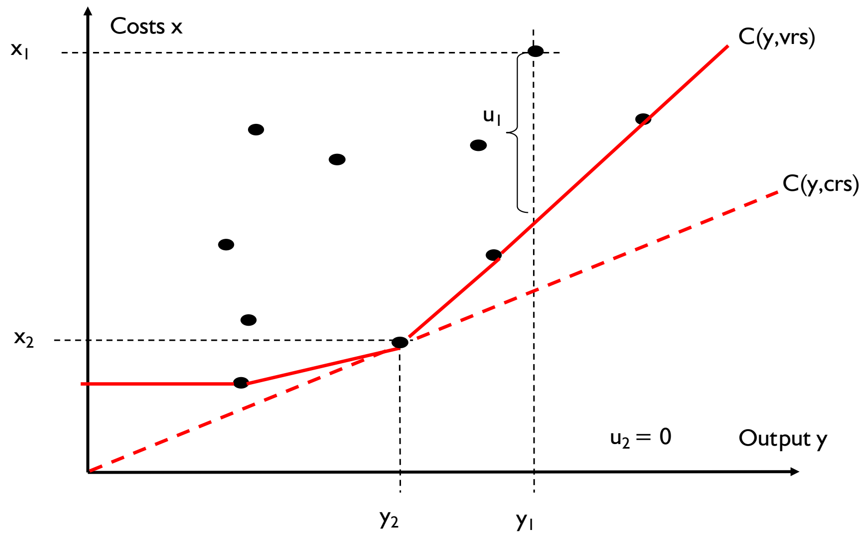


Figure 8-6 Cost function in DEA.

8.40 As seen in the review above, DEA is most popular benchmarking for regulatory applications in Europe and elsewhere. The advantages for regulation are (i) the absence of assumptions on the functional form of the frontier or the inefficiency distribution, (ii) construction of cost function "targets" based on identified real observations, (iii) convergence and discriminatory ability already for small datasets without any numerical "tuning" or "tweaks".

8.41 The weaknesses of DEA are linked to its deterministic nature: sensitivity to peer data, necessitating a careful data validation and outlier detection. In addition, DEA does not provide estimates for the uncertainty of the estimation (like confidence intervals) without specific methods (Simar and Wilson, 2001) like SFA, or direct testing of environmental effects, e.g. over time. To distinguish it from the parametric methods that pad the frontier to compensate for noise, DEA acts like a multi-item auction or recommendation system finding the lowest supplier for any given output profile.

Semiparametric

8.42 Semiparametric methods combine the flexibility of nonparametric frontier estimation with the noise-filtering benefits of parametric techniques. For instance, the **StoNED (Stochastic Non-Smooth Envelopment of Data)** method (Kousmanen & Kortelainen, 2012) blends a shape-constrained regression (like Convex Nonparametric Least Squares) with a stochastic noise component. This is well-suited to benchmarking, especially with panel data, where it captures both random variation and managerial inefficiency while enforcing economic properties like monotonicity and concavity. Semiparametric methods can provide a balance between flexibility and robustness, yielding interpretable results even in the presence of limited sample sizes and noisy data.

8.43 The cost function in StoNED is then a combination of DEA and SFA:

$$C(y) = C^{*}(y)e^{(z+u+v)}$$

where $C^{*}(y)$ is a piece-wise linear cost function with arguments y in form of an output vector (like in DEA) u is the inefficiency (with assumptions as in SFA), v is random noise (with assumptions as in SFA) and z is (an optional) vector of environmental factors. As seen, if the noise is zero and there are no environmental factors, the model approaches the DEA result.

8.44 In Figure 8-7 below the (inverse) cost functions for 10 simulated observations in DEA and StoNED. The main difference is in the importance of efficient peers ("corner points") in the graph: in DEA these units are given weight as cost targets without deduction, in StoNED the frontier is smoothed below and many units are in fact "superefficient" meaning that they have a lower cost than StoNED would predict (here: above the graph). Note that the point A in Figure 8-7 an efficient peer in DEA partially setting the targets for most inefficient points: in StoNED the frontier is flat and A represents simply a randomly good outcome, like a winning a lottery. On the other hand, point B in Figure 8-7 is inefficient in DEA and the horizontal distance between it and the blue DEA-frontier is the potential cost saving for its ouput level. In StoNED, point B is only marginally inefficient, the point is almost on the red StoNED frontier. In Andor and Hesse (2014) it is noted that the StoNED method has a tendency to overestimate the efficiency of units when calculated using a specific method (pseudolikelihood, PL). Our figure illustrates the same pattern.

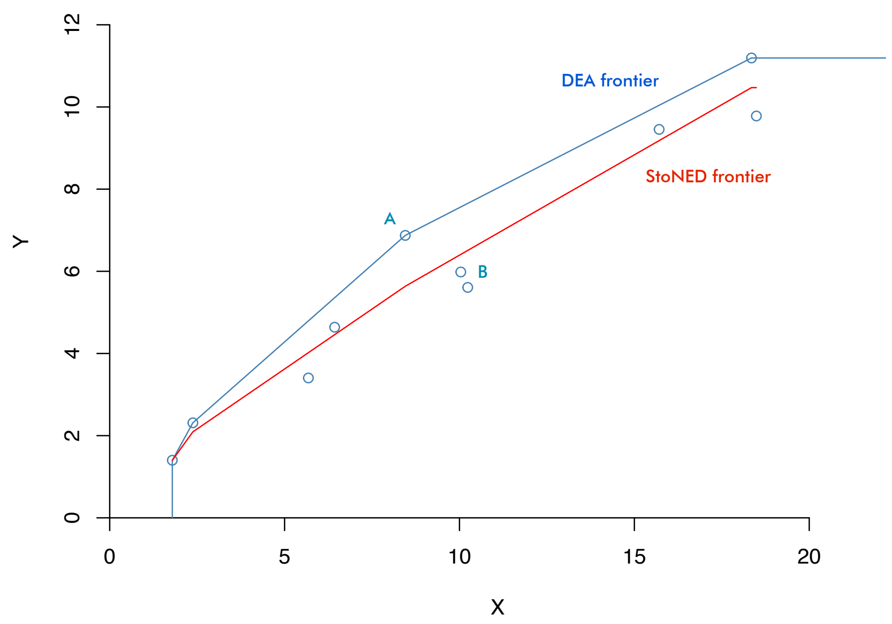


Figure 8-7 Cost functions (inverse) in DEA and StoNED. Simulation, 10 observations.

8.4 Comparisons of methods

8.45 The methods differ in the following characteristics, listed by method in Table 8-3 below:

1) *Noise separation:* Can the method deal with random errors in data?
2) *Specification errors:* How does the method deal with errors or omissions in the model specification (e.g., missing variables)?

3) *Inefficiency model:* How does the model derive *inefficiency*, i.e. assumptions behind non-frontier observations?
4) *Cost function consistency:* The primary metric is Totex-efficiency, with secondary metrics for partial efficiencies (opex, capex). Is the method compatible with this?
5) *Cost function flexibility:* Is the form of the cost function asserted by the analyst or derived from the data?
6) *Environmental factors:* Can environmental factors be integrated in the analysis, and if so, how?
7) *Data requirements:* What are the data requirements to achieve an adequate differentiation of performance?
8) *Calculations:* What is the computational burden, the software basis and the requirements for coding?
9) *Precision:* How accurate is the method and how can this be estimated (if at all)?
10) *Regulatory use:* How frequently is the method used by NRA in regulation?

Table 8-3 Overview efficiency measurement methods.

	SFA	COLS	MOLS	DEA	StoNED	UC	EngM
Data errors	Random	No	Random	Outlier det	Random	No (outliers)	No
Specification error	Random	No	No	Minimal assump	Random	No	?
Inefficiency model	Random	Frontier	Frontier	Frontier	Random	(Partial)	Norm
C(y) consistent	Yes	Yes	Yes	Yes	Yes	No	(No)
C(y) flexible	No	No	No	Open	Open	Linear	No
Environmental factors	Integrated, with noise	No	No	Integrated, deter	Integrated, with noise	No	Flexible
Data requirement	High	Low	Low	Low	High	Very low	Very high
Calculations	Complex	Simple	Simple	Simple	Complex, non-std	Simple	Very complex
Precision	High	Low	Medium	High (det)	Medium	Low (partial)	Uncertain
NRA use	XX	(X)	XX	XXX	X	(X)	X

8.5 Published studies on gas transport efficiency

8.46 In Table 8-4 below a selection of the most important published studies 1978-2025 (see the bibliography) concerning the cost efficiency of gas transport (primarily transmission) by academic authors. For each study we note the year, the region or country of origin of the data, the method(s) applied, the scope (T= transmission, D = distribution) and the horizon of the data (panel or single-year data).

8.47 The methods listed in (for TSOG studies) and (for DSOG studies) are summed by type in Table 8-6 to reveal some tendencies. As seen, most of the studies (19 of 36) are based on DEA (alone or with other methods), followed by SFA in 7 studies, thereof 2 in combination with DEA. We find 7 general econometric applications, almost all cost-function analyses of different properties (scale, profit margin, resource use) before 2000. COLS is represented twice with other methods. Concerning the data, we note a strong dominance for panel data (28 of 36 papers), which is partially motivated by research questions related to dynamic changes in efficiency, productivity or effects of regulatory changes in time. Finally, given the systematic data collection and early transparency by FERC, it is not surprising to note that 15 of the 36 studies are using US data (at least partially). However, the recent projects on EU-data (since 2014) are more detailed than any of the earlier international papers. Note that Lawrence et al. (2011) is work for the Commerce Commission of New Zealand on gas transmission and distribution, using data primarily from New Zealand and Australia, as well as secondary sources from US and EU. The project Jamasb et al. (2007) is exploratory study for CEER with 43 US TSOG and data for 4 EU TSOG, it was not integrated in the general workplan.

8.48 The state of the art in gas transmission analysis follows closely that in the methodological field of efficiency analysis, with a continuous increase in the use of more advanced methods and tests to control of heterogeneity and endogeneity, as with recent integration of techniques such as latent class methods (LCM) and clustering techniques (Romano et al., 2022), variable reduction in DEA (Toloo & Babaee, 2015), robust super efficiency (Sadjadi et al. 2011) and methods to control for heterogeneity and endogeneity in SFA (Hou et al., 2025).

Table 8-4 Published studies on TSOG network cost efficiency, **bold** = regulatory projects, T= transmission, D = distribution.

AuthorTable 8-4	Year	Country	Method	Data	T/D
Callen	1978	US	Econometrics	Panel	T
Aivazian et al	1987	US	Econometrics	Panel	T
Sickles & Streitwieser	1992	US	SFA	Panel	T
Ellig & Giberson	1993	US	Econometrics	Panel	T+D
BIE	**1994**	**International**	**DEA**	**1994**	**T+D**
Kim & Lee	1995	IT	Econometrics	Panel	T
Boussofiane et al.	1997	UK	DEA	Panel	(T)
Granderson & Linvill	1997	US	DEA	Panel	T
Sickles & Streitwieser	1998	US	TFP	Panel	T
Kim et al.	1999	International	DEA	Panel	T
Bernard et al	2002	CA	Econometrics	Panel	T
Hawdon	2003	International	DEA	1998	T
Gordon et al.	2003	CA	Econometrics	Panel	T
Jamasb et al.	**2007**	**US+EU**	**DEA,SFA,COLS**	**2004***	**T**
Jamasb et al.	2008	US	DEA	Panel	T
E2GAS	**2014**	**EU**	**DEA, SFA**	**2014(2010)**	**T**
Storto	2018	US	DEA	2012	T
Vikas & Bansal	2018	IN	DEA	Panel	(T+D)
TCB18	**2019**	**EU**	**DEA**	**2017**	**T**
Ajayi & Pollitt	2022	UK	DEA	Panel	T
TCB21	**2025**	**EU**	**DEA**	**2020**	**T**

Table 8-5 Published studies on DSOG network cost efficiency.

Author	Year	Country	Method	Data	T/D
Hollas & Stansell	1994	US	Econometrics	Panel	D
Kim & Lee	1995	KR	Econometrics	Panel	D
Waddams-Price & Weyman-Jones	1996	UK	DEA	Panel	D
Rossi	2001	AR	SFA	Panel	D
Hollas et al.	2002	US	DEA	Panel	D
Farsi et al.	2007	CH	SFA	Panel	D
Erbetta & Rappouli	2008	IT	DEA	Panel	D
Gonacharuk	2008	UA+US	DEA	2005	D
Lawrence et al.	2011	NZ+AU	TFP	Panel	D
Casarin	2014	AR	TFP	Panel	D
Aleifar et al.	2014	CH	Econometrics	Panel	D
Tovar et al.	2015	BR	SFA	Panel	D
Gugler & Liebensteiner	2019	AT	TFP	Panel	D
Capece et al.	2021	IT	DEA	Panel	D
Romano et al.	2022	IT	DEA, LCM	Panel	D
Kasiri & Mirnezami	2023	IN	SFA, COLS	Panel	D

Table 8-6 Methods and data in gas pipeline efficiency studies (TSOG and DSOG).

Method	Single-year	Panel	Sum
DEA	6	10	16
DEA, LCM		1	1
DEA, SFA	1		1
DEA, SFA, COLS	1		1
SFA		4	4
SFA, COLS		1	1
TFP		4	4
Econometrics		8	8
Total	8	28	36

8.6 Evaluation criteria for methods

8.49 There is no single method that fits all applications of productivity and efficiency analysis. The choice of method can be guided by a set of criteria related to the specific application in AEC, the objectives of the study, the type and quality of the data available and the features required by the results and their reliability.

8.50 A few authors have proposed selection criteria for frontier analysis methods and processes in regulatory applications. We note here the consistency conditions by Bauer et al. (1998) applied in Rossi and Ruzzier (2000), the best practice summary by Lowry & Getachew (2009), the cost norm criteria in Agrell and Brea-Solis (2017) and the code of practice proposed by Biggar (2025).

8.51 Bauer et al. (1998) present a set of condition for frontier analysis methods in financial regulation, without loss of generality here. The conditions are (verbatim):

Internal consistency

1) The efficiency scores generated by the different approaches should have comparable means, standard deviations, and other distributional properties;
2) the different approaches should rank the [units] in approximately the same order;
3) the different approaches should identify mostly the same [units] as "best-practice" and as "worst-practice;"

External consistency

4) All of the useful approaches should demonstrate reasonable stability over time, i.e., tend to consistently identify the same [units] as relatively efficient or inefficient in different years, rather than varying markedly from one year to the next;
5) the efficiency scores generated by the different approaches should be reasonably consistent with competitive conditions in the market; and
6) the measured efficiencies from all of the useful approaches should be reasonably consistent with standard non-frontier performance measures, /../.

8.52 Bauer et al. (1998) concentrates on the 'scores' per se, as the cost function concept is less relevant in ranking financial institutions. The *internal consistency* guarantees inter-method consistency within the benchmarking, the *external consistency* is aimed at the plausibility of the results facing evidence outside of the benchmarking. In considering the concrete choice of DEA vs SFA for (financial) regulation, Bauer et al. (1998) states:

> *"The data also show a high degree of consistency within the parametric methods and within the nonparametric methods. /../ Rather, the only choice that appears to matter greatly for regulatory policy considerations is the choice between the parametric and nonparametric methods."*

8.53 Rossi & Ruzzier (2000) apply the conditions above in the context of energy network regulation using empirical considerations for data, model specification and regulatory applications. Rossi and Ruzzier (2000) conclude with a schematic process that contains both a parametric and non-parametric phase, as of below:

1) Identify a reference set of comparable operators.
2) Define the 'core' model (cost or production function) and define relevant input and output candidates, collecting reliable data for all.
3) Select and collect any environmental variables that may impact the assessment.
4) Run a parametric regression (average cost or production function) with the selected inputs and outputs, dropping any non-significant variables until a model with only valid regressors has been obtained.
5) Run a non-parametric DEA model with the inputs and outputs from step (2) to identify efficient and inefficient firms.
6) Regress the results from (5) on the final regression model in (4).
7) Apply the Bauer consistency conditions on the scores from (5), after corrections if necessary from (6).

8.54 Lowry & Getachew (2009) review the choice of DEA, SFA or COLS for utility regulation. After a critical review, also addressing the general objections raised by Shuttleworth (2005) against benchmarking in tariff reviews, the authors conclude that frontier analysis is important for the economic utility regulation to bridge information asymmetry, that both DEA and SFA correspond to best practice methods with strengths that are complementary, and in particular that the use of DEA to identify peers and outliers may be useful even for regulators using low-powered methods (cost-plus or partial cost regulation) to guide prudence reviews.

8.55 Agrell and Brea-Solis (2017) focus at the more technical properties of cost functions for incentive regulation of energy networks. They derive a set of criteria to make sure that the chosen methods and model produce results that are compliant with due regulatory process (here reformulated without notation):

1) *Feasibility*: any prediction for cost of operation should be feasible for an operator.
2) *Neutrality*: there should be no bias in the cost norm for some group(s) of operators.
3) *Robustness*: the cost norm should be insensitive to irrelevant, flawed or inefficient data added to the reference set.
4) *Repeatability*: if the cost norm requires technical constants, assumptions or numerical parameters, they should be endogenous or set a priori based on practice and science. The calculations should be repeatable by other analysts.

8.7 Method evaluation

8.56 A common factor in the selection of benchmarking methods in the literature is *the ability to estimate the metric(s) selected in a consistent and scientifically sound manner for the technologies and reference sets that are present in the data.*

Engineering methods

8.57 The engineering methods provide an estimate of optimal cost compared to some a priori assumptions without any link to the reference set, dynamics or changing tasks. The methods are intrinsically 'black-box' and based on specific installations without any structured assessment of uncertainty or feasibility of the obtained estimates for other types of situations. For these reasons, the engineering methods fail the selection criterion for AEC.

KPI-methods

8.58 Partial measures such as KPI and ratios do not form a consistent cost function for multi-output production such as gas transport. The methods cannot reliably estimate any sound cost function, nor control for heterogeneity. For these reasons, the KPI-methods fail as eligible methods for AEC.

Parametric methods: COLS and MOLS

8.59 The parametric methods are based on sound principles in statistically-based econometrics. Regression techniques allow the estimation of production and cost functions with determined uncertainty. Corrected OLS regression (COLS) is a special variant where the frontier (best practice) estimation depends on a single point and the intercept of the regression lines is shifted to run through this point. However, this forces two strong assumptions on the maximum (or minimum for a cost function) and the variance of the cost (or output) that is one-sided. This means that the statistical properties are complex. Further, the interpretation is unclear of best practice: why is optimal cost (a single point) perfectly controllable if the cost function itself is stochastic? For these reasons, COLS is not a consistent cost function apt to capture the technology in AEC.

8.60 The Modified OLS (MOLS) is similar to COLS, but with an added uncertainty to the frontier. However, some statistical measures may turn out to meaningless (the variance may be negative depending on the position of the maximum point and the variance of the average cost function) for MOLS, cf. Mastromarco (2008). Moreover, MOLS assumes that the cost function is fully defined by the average units, but the best-practice is defined by a single point. This is contradictory to production theory the 'best practice' technology is expected to the different from the mean technology, potentially also at the different scale or profile, when there is inefficiency in the sample. MOLS confounds noise and inefficiency without any flexible functional form, and for the reasons above MOLS fails as a method for AEC.

Parametric methods: SFA

8.61 The Stochastic Frontier Analysis method is a powerful and popular (as seen above) for panel data analysis, especially when controlling for exogenous conditions or time effects. The properties in this area are well documented and will be evaluation below. However, for a crossection or smaller data set, SFA performs poorly or, at best, equally to DEA (Ondrich and Ruggerio, 2001, Ruggerio, 2007). For these reasons we note that SFA alone cannot fulfil all requirements for the models desired in the AEC.

Non-parametric methods: DEA

8.62 Data Envelopment Analysis is the primary empirical method in gas network efficiency analysis (see above) for multiple reasons: flexible functional form, stable with various size of reference set, clearly identified technology and peers for the optimal cost function, full decomposition of static and dynamic efficiency. However, DEA is sensitive to model specification errors and outliers in the reference set. Differences in exogenous conditions and corrections are most likely occurring when comparing across time. For these reasons, DEA cannot replace a method like SFA in dynamic time-series analysis with potentially stochastic data.

Semi-parametric method: StoNED

8.63 The new method StoNED appears to combine the two major methods DEA and SFA. The method is sound and well documented, meaning that there is no reason to dismiss it based on the exclusion criterion above.

Final evaluation: DEA, SFA and StoNED

8.64 We proceed to a detailed evaluation of the three non-dominated methods in Table 8-7 below.

Table 8-7 Detailed comparison SFA, DEA, StoNED.

	SFA	DEA	StoNED
Noise separation	Modeled as random variable with given distribution	None, deterministics	Modeled as random variable with given distribution
Data errors	Stochastic errors in production and inefficiency	Deterministic variables, outlier test for observations	Stochastic errors in production and inefficiency
Specification error	(Variance of estimation, see data error)	Minimal assumptions on production function	(Variance of estimation, see data error)
Inefficiency model	One-sided random variable with given distribution	Deterministic residual, defined by data	One-sided random variable with given distribution
Cost function consistency	Consistent, production and cost functions	Consistent, production and cost functions	Consistent, production and cost functions
Cost function flexibility	Given functional form, identical for all TSO	Fully flexible, returns to scale can be defined	Fully flexible, returns to scale can be defined
Environmental factors	Given factors, marginal effect same across TSO	Marginal effects can be different across TSO	Marginal effects can be different across TSO

8.65 StoNED is as noted essentially a compromise between the stochastic SFA and the deterministic DEA, with many of the advantages. As discussed above, the differences are not decisive from a qualitative perspective. A good approach in such situations is to evaluate the power of the methods to correctly estimate the efficiency for simulated data with known results, e.g. using the Mean Absolute Deviation MAD or Mean Square Error. However, the design of a simulation may have a strong bias on the outcome, it is therefore important to vary as many characteristics as possible in a systematic pattern. We refer here to the large-scale comparison in Andor and Hesse (2014) using 188 variants (sample size, degree of noise in data, distribution for noise and inefficiency, calculation method, omitted variables, heteroskedasticity, functional form for production function, number of inputs, collinearity and input distribution). The main results are summarized in Table 8-8 below. Low MAD and MSE are preferred. In short, SFA using Maximum Likelihood estimation (ML) is coming out as the strongest method for both high and low noise for stochastic models, followed by DEA for models with no noise (deterministic). StoNED does not outperform neither DEA for low noise, nor SFA for medium noise, it performs well only for settings with very high noise level, where the random variance ("chance") is at least twice as important as the variance for inefficiency ("skills"). As the primary models for AEC are focused on deterministic capacity and grid provision, using asset or asset-proxies, the noise level by design should be very low.

8.66 In addition, Andor and Hesse (2014) notices that StoNED has a relatively slow calculation convergence and few standard software packages for calculation and validation[10]. For regulatory benchmarking, this is an additional risk, since the

[10] The calculation of StonED in the R-package notes that "*No non-commercial solver at this time of writing is able to solve the NLP formulation required for the multiplicative StoNED. Therefore, the NLP is approximated with a QP formulation with some transformation of the objective function.*" However, a

results must be able to be reproduced with any software and not depend on a specific implementation, just like DEA using Linear Programming. It is likely that researchers will propose, test and enhance the software base for StoNED as well as for other new methods in the coming years.

8.67 For these reasons, we recommend DEA and SFA as the two preferred methods for AEC. DEA is perfectly adapted to crossection and subset analysis, flexible and stable. SFA is the preferred method for panel data analysis and dynamic models. The methods are complementary and irreplaceable for the purposes of the AEC, well documented and supported by practice and the scientific community.

Table 8-8 Evaluation of SFA, DEA, StoNED (Andor and Hesse, 2014, Tables 2 and 3).

Table 2 Overview of the performance criteria for all 188 settings

	DEA	SFA MoM	SFA ML	StoNED MoM	StoNED PL
MD	−0.0696	−0.0496	−0.0230	−0.0385	0.0295
MAD	0.1101	0.0850	0.0659	0.0862	0.0719
Rank (MAD)	*3.63*	*3.31*	*2.14*	*3.49*	*2.42*
MSE	0.0262	0.0124	0.0106	0.0132	0.0106
Rank (MSE)	*3.58*	*3.23*	*2.15*	*3.56*	*2.49*
MRC	0.6868	0.7163	0.7295	0.6613	
Rank (MRC)	*3.21*	*1.98*	*1.36*	*3.46*	

Table 3 Overview of the performance criteria in the subsample without noise ($\rho_{nts} = 0$)

	DEA	SFA MoM	SFA ML	StoNED MoM	StoNED PL
MD	0.0127	−0.0530	−0.0023	−0.0368	0.0227
MAD	0.0426	0.0682	0.0232	0.0701	0.0535
Rank (MAD)	*2.33*	*4.01*	*1.30*	*4.06*	*3.31*
MSE	0.0051	0.0074	0.0023	0.0084	0.0066
Rank (MSE)	*2.31*	*3.80*	*1.39*	*4.19*	*3.34*
MRC	0.8710	0.9001	0.9290	0.8268	
Rank (MRC)	*2.84*	*2.30*	*1.23*	*3.64*	

8.8 Reference data

8.68 Both DEA and SFA can, and should be used, to calculate estimates for a specific year (static) and for efficiency development over time (dynamic efficiency).

8.69 The development, validation and estimation of efficiency models crucially depend on the size and quality of the dataset. It is therefore planned that the default dataset is the full *panel* for all TSOGs and all years (2021 to 2025). Ideally, the dataset is balanced, meaning that all TSOGs report valid data for all years. In the case this is not fulfilled, provisions should be made to control for an unbalanced panel as to avoid bias in model specification and outlier detection.

8.70 *In TCB21, the panel was unbalanced due to German data available only for one year. Specific measures and tests were performed in Screen B Robustness tests (Sumicsid and CEER, 2024, Section 5.9) to test whether the model results depended on a specific year or country.)*

Python code (pystoned) for the calculation of StonED has been released 2021, https://pystoned.readthedocs.io/en/latest/

Cross-sectional data

8.71 If data from a single year is used separately to estimate effects, it is called a *cross section*. With the size of the data set for AEC, this is possible using DEA, but hardly feasible using SFA unless the model is very compact (few outputs). Cross-sections have the advantage of avoiding yearly effects and shocks, but a model cannot be reliably specified using just a single year of data. Moreover, such approach would be inconsistent with the assumption of going concern and stationary transmission technology.

Multi-year data

8.72 For the calculation of dynamic efficiency, the full data set is used as a compilation of years, with results calculated for each year. This means that each participant not only receives one year's result, but one score for each of the years 2021 to 2025. In addition, the Malmquist metrics is calculated as described above for the horizon.

8.73 DEA is the natural method for dynamic efficiency as it is based on the deterministic execution model, but SFA may also be used for control or as complement (see also Jamasb et al, 2008, for an analogous conclusion).

Panel data

8.74 In panel models, the models depend explicitly on the time effects (fixed per year or random effects), and the variables can be defined as changes to base levels rather than cumulative measures ("stocks"). In this manner, one may study impact of recent investments, demand changes and marginal effects of environmental changes. This type of model avoids the endogeneity problem that is intrinsic in asset-based deterministic models: the Totex varies very little year by year and cumulative efficiency is strongly predicted by the past.

8.75 The panel models are proposed to be estimated using SFA under Maximum Likelihood and also involve stochastic utilisation data, depending on availability also continuous-valued environmental variables (e.g. the raw slopes, volume of stones etc) or discrete level variables (e.g. shares of landuse categories).

8.76 *An example of the sensitivity of SFA to sample size is given in Figure 8-8 below. SFA is applied to two SFA models (ED and UD) for panel data that will be presented in next section, varying the number of TSOGs from 5 (meaning 5 TSOG x10 years = 50 observations) to 20 (200 observations). Unfilled dots in the figure indicate runs in which SFA cannot obtain a significant inefficiency term (Gamma). As seen, this happens for one instance for model UD, but three times (for 7,8 and 14 TSOGs) for the execution model ED. When the inefficiency term in SFA is not significant, the model cannot be used to show any inefficiency, although the effect is entirely due to the sample set, not to the frontier. See also the discussion on sample size above.*

8.77 *The behaviour of SFA is not different than any stochastic estimation method. Assume for instance that we try to prove that a student is better than the expected mean student at a test. As the number of students passing the test decreases, the standard deviation in confidence interval increases until finally no effect can be proven, although the number (score) is the same.*

8.78 *In AEC the expected usable dataset will be around 40 x 5 = 200 observations, which should be sufficient to estimate both an execution and utilisation model with SFA (rightmost observations in Figure 8-8).*

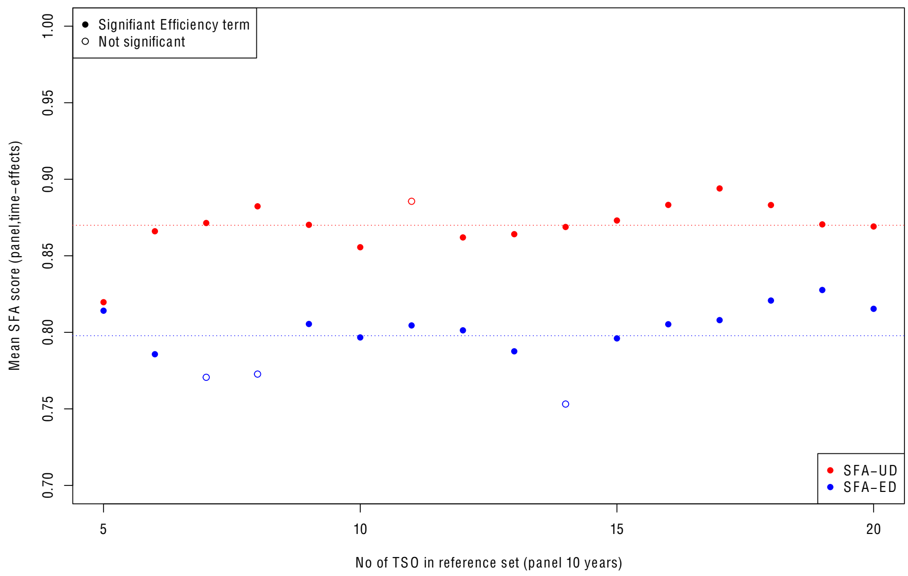


Figure 8-8 Mean SFA score (unfilled=insignificant) per number of TSOG included in panel.

8.9 AEC Core Models

8.79 As concluded in section 6.4 above, the NRA needs require AEC to produce both an execution-based and a utilisation-based model. In discussing the temporal dimension, it has also been argued that AEC should be able to provide both results for any single year (static) as well as dynamic analyses for the development of efficiency and grid utilisation.

8.80 In consequence, we propose three core models for the purposes of AEC, presented with their focus in Table 8-9, Table 8-10 Table 8-11 and in Figure 8-9:

1) An execution-focused (E) model for each year (static=S) in the period (ES)
2) An execution-focused (E) model for dynamic (D) efficiency (ED)
3) A utilisation-focused (U) model using stochastic dynamic (D) panel data (UD)

Table 8-9 Model and use in regulation.

Model	Execution ES	Execution ED	Utilisation UD
Functional focus	Procurement, operations, maintenance	Cost and staff control, continuous efficiency improvements	Planning, capacity management, decommissioning and repurposing
Regulatory mechanism (ex)	Revenue cap (Xi), cost review	Revenue cap (Xgen), improvement targets	Investment and decommission approvals,
Perspective	Short term	Medium term	Long term
TSOG risk sharing	Low	Medium	High

Table 8-10 Model, method and orientation in AEC.

Model	Execution ES	Execution ED	Utilisation UD
Method	DEA, SFA	DEA	SFA
Timescale	Static	Dynamic	Dynamic
Data set	Crossection	Panel data	Panel data
Objective	Min Cost for a given grid system in a year	Efficiency and cost development over time	Min Cost to cater for transport demand
Input	TOTEX	TOTEX	TOTEX
Outputs	Grid provision	Grid provision	Capacity use
	Capacity provision	Capacity provision	Volume transported
	Service provision	Service provision	Area covered

Table 8-11 Model and efficiency assumptions in AEC.

Model	Execution ES	Execution ED	Utilisation UD
Data generation	Deterministic	Deterministic	Stochastic
Efficiency distribution	Deterministic, controllable	Deterministic, controllable	Stochastic, eg half-normal
Environmental effects	Adjusted outputs	Adjusted outputs	Separate control variables (few)
Time effects	(separate)	none	Fixed or random effects
Outlier detection	Superefficiency,	Superefficiency,	Cook's distance
	Dominance,	Dominance,	
	Peeling	Peeling	

Deterministic model (ES, ED)

8.81 The deterministic models (execution-focused, ES and ED) are multi-year applications of one or several models that are based on variables related to the provision of assets, capacity and other services to the users, by definition controllable and deterministic actions, excluding exogenous non-controllable shocks such as peakload, utilisation factors, weather (heat year) and natural disasters.

8.82 The deterministic model ES is recommended for use in the assessment of incumbent total cost efficiency for any TSOG for any year, e.g. for informing revenue-cap parameter setting.

8.83 Model ED is proposed to be solved by DEA to produce scores year by year (static) as well as dynamic efficiency results (efficiency change and technological change) using a consistent decomposition (basically ES models over time).

8.84 SFA may also be used, if necessary, after transformations of variables to cater to the parametric specification (logs or norming, for instance) for ES using one of the years in a panel data study.

Stochastic model (UD)

8.85 The stochastic utilisation model UD is a panel data application of one or several models that also include usage-based outputs (volumes, peak load) as well as yearly fixed or random effects. This model will serve several purposes from the sector:

1) Relative efficiency development with investment, decommissioning and volume changes.
2) Impact of random environmental parameters on variable cost (Opex).
3) Development of global, regional and country-level utilisation levels in natural gas.

8.86 The stochastic model is proposed to be solved using SFA with appropriate assumptions for the distribution of inefficiency, production functions and functional forms.

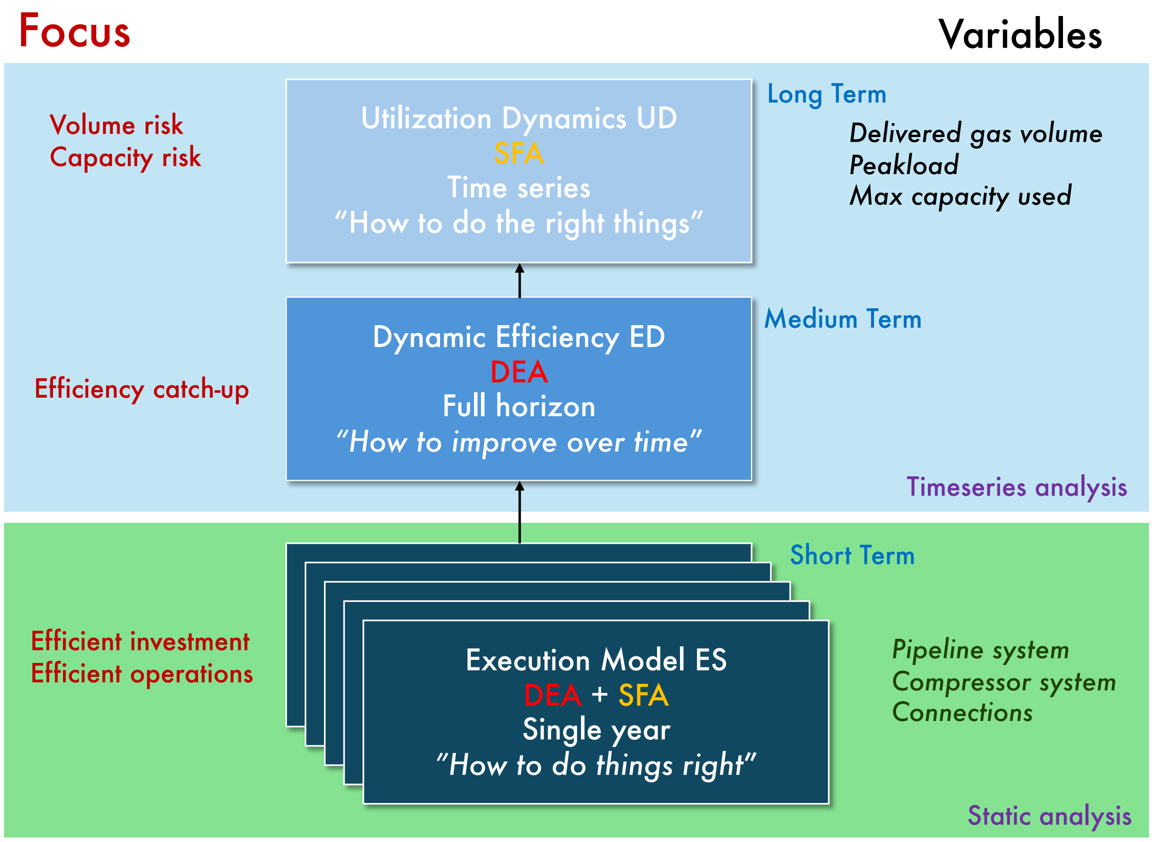


Figure 8-9 Core models ES, ED, and UD.

8.10 Illustration: Model ES

8.87 *Data for 20 US TSOG 1977-1986 (see Appendix F) are used to develop models using the ideas in the note for each model, starting from ES and ED, then extending to UD.*

8.88 *We start with the core model ES (static execution) that could be highly informative for regulatory rulings in the medium term. The model focuses on cost efficiency CE in a specific reference year and is independent of events or changes in other years of the data.*

8.89 *As an example of an execution-based model, we select an asset-based core, that is the pipeline capital (vPipe) and compressor volume (xCap.comp). The dependent variable is Totex, in real terms using standardized input prices for capital. Here is USA is considered a homogenous labour region, so no LCI is applied to staff costs for the operators.*

8.90 *The execution model ES is presenting good properties as an average cost function*[11]*.*

8.91 *For the illustration, we arbitrarily chose 1986 as the reference year and apply both DEA and SFA to the model as discussed above.*

8.92 *The ES results for SFA in the reference year are graphed in Figure 8-10 below. The mean score for 1986 is 94.3% and the median score is 95.2%, indicating an inefficiency level of about 5% of Totex with a maximum of 20% for one TSOG. The distribution of scores is very even except the lowest performer (outlier analysis is discussed below and not implemented for this run). The SFA results are significant*[12] *for the model using the panel data as shown in Table 8-14.*

8.93 *The ES results for DEA in the reference year 1986 are shown in Figure 8-11 below, for both IRS (NDRS) correcting for size-effects and CRS (using optimal scale). The mean efficiency for IRS is 91.5% and the median score 92.7%. The CRS scores are close: 90.2% (mean) and 91.2% (median). The minimum score for both is 71.1% for the UNITED GAS, scoring worst also in SFA. Six TSOGs are characterized as fully efficient peer units, they will be discussed below in the Peer Analysis. As for SFA, the outlier analysis for DEA will be discussed below, scores after removal of outliers will be increased.*

8.94 *The full results for both DEA and SFA along with scale efficiency results (SE) are listed in Table 8-15 for each TSOG. Here, individual scores by TSOG and method can be compared. As seen, for the worst and best, the ranking order is very similar*[13]*, The middle column in Table 8-15 gives the scale efficiency (DEA-ES-SE) for model*

[11] See the regression results for level (untransformed) in Table 8-12 and for the logarithmic transformation in Table 8-13. All coefficients are positive and significant at the 0.01-level, the model explains 96% (97% in log) of the Totex-variance and has the expected positive skew (the residual error is more positive than negative, meaning that there is potentially inefficiency in the sample). The model does not suffer from collinearity and performs well both in levels and logarithms.

[12] All coefficients are significant, have correct signs and the inefficiency term gamma is significant. If run with a time-effect (per year) as in the table, the effect is positive, indicating an increasing cost per year, raising a caveat to use a pooled frontier (observations from different years) to estimate the efficiency in e.g. DEA.

[13] Spearman rank order correlation between the DEA-IRS and the SFA scores is 80.3%.

ES. Note that the TSOGs are mostly in the optimal range, mean SE is 98.6%, but there is a group of smaller TSOG that will be discussed below in the Peer analysis.

Table 8-12 OLS (average cost function) Execution, model ES-ED (panel, level,FERC-US)[14].

Variable	Estimate	Std. Error	t-value	Pr($>\|t\|$)
$vPipe$	0.141	0.007	20.14	< 0.001 ***
$xCap\_comp$	184.484	10.053	18.35	< 0.001 ***
Constant	17,861,064	3,381,942	5.28	< 0.001 ***
Observations				220
R^2 / Adjusted R^2				0.958 / 0.958
Residual Std. Error				29,800,694 (df = 217)
F Statistic				2,497.6 *** (df = 2; 217)

Notes: Dependent variable is $dTotex.real.std.y$.
Significance codes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table 8-13 OLS (average cost function) Execution, model ES-ED (panel, log,FERC-US).

Variable	Estimate	Std. Error	t-value	Pr($>\|t\|$)
$vPipe$	0.513	0.023	22.30	< 0.001 ***
$xCap\_comp$	0.411	0.019	21.63	< 0.001 ***
Constant	3.373	0.279	12.09	< 0.001 ***
Observations				200
R^2 / Adjusted R^2				0.973 / 0.972
Residual Std. Error				0.117 (df = 197)
F Statistic				3,514.9 *** (df = 2; 197)

Notes: Dependent variable is $dTotex.real.std.y$. Standard errors are heteroskedasticity-robust.
Significance codes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

[14] The regression tables contain the columns variable names, estimate (value of the coefficient in the model), Std.Error (standard error for the estimate), t-value (significance value for the estimate), Pr(>|t|) (significance level as probability). The table also includes the number of observations, the model fit (adj.R2), the residual Std.Error (the positive square root of the mean square error, a measure of fit), F-statistic (a measure of overall significance of the model compared to a model with no independent variables, just a constant, higher is better), df = degrees of freedom = number of observations – number of variables estimated – 1.

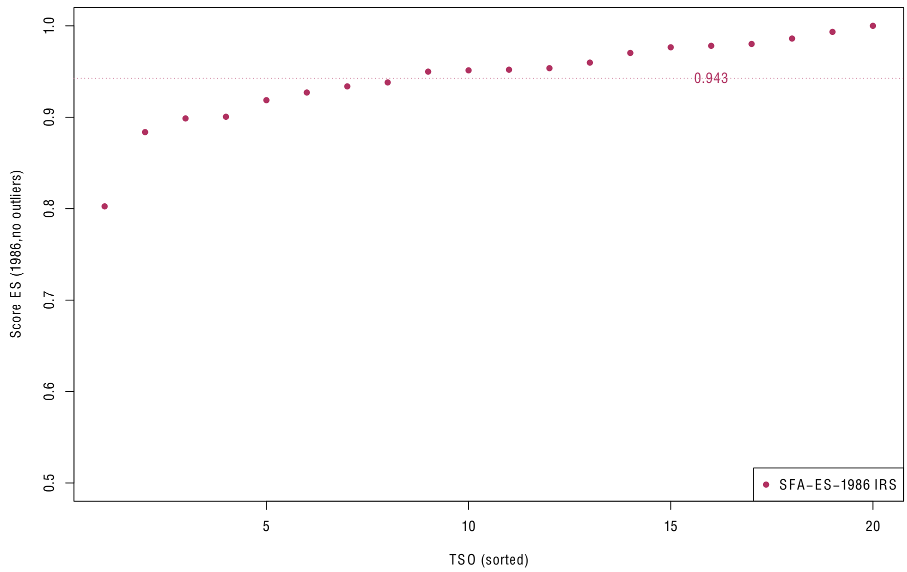


Figure 8-10 Scores SFA-ES-IRS, 1986, no outlier removal.

Table 8-14 SFA results, model ES-ED panel, Battese-Coelli (1992) specification.

Variable	Estimate	Std. Error	z-value	Pr($>|z|$)
(Intercept)	2.8077	0.9153	3.068	0.0022 **
$vPipe$	0.5261	0.0501	10.504	$< 2.2 \times 10^{-16}$ ***
$xCap\_comp$	0.4179	0.0506	8.266	$< 2.2 \times 10^{-16}$ ***
σ^2	0.0411	0.0157	2.623	0.0087 **
γ	0.8749	0.0513	17.050	$< 2.2 \times 10^{-16}$ ***
$time$	0.0382	0.0100	3.805	0.00014 ***
Log-likelihood				209.785
Convergence				after 16 iterations
Panel size				$N = 200$ (20 cross-sections $\times$ 10 years)

Notes: The model is an Error Components Frontier where inefficiency increases the endogenous variable (cost frontier). The dependent variable is in logarithmic form. Maximum-likelihood estimation followed the iterative procedure of Battese and Coelli (1992), converging after 16 iterations when changes in the log-likelihood and parameter estimates fell below tolerance limits.
Significance codes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

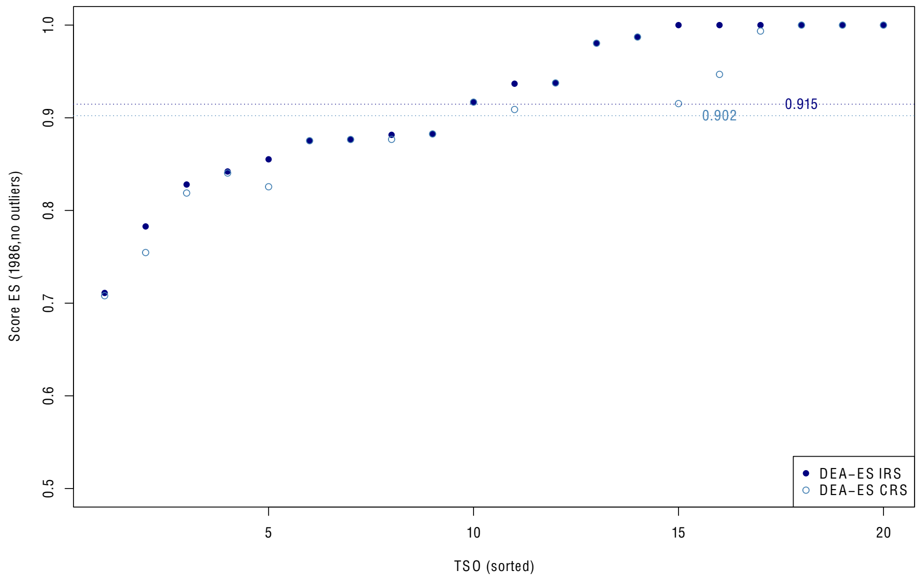


Figure 8-11 Scores DEA-ES-IRS/CRS, 1986, no outlier removal.

Table 8-15 ES-IRS-1986 results, DEA and SFA.

	ID	DEA-ES-IRS	DEA-ES-SE	SFA-ES
ARKLA	101	0.937	0.970	0.927
COLORADO	102	1	0.915	0.970
COLUM GAS	103	0.842	0.998	0.884
COLUM GULF	104	0.917	1	0.978
CNG	105	0.783	0.964	0.899
EL PASO	106	0.987	1	0.952
FLORIDA	107	0.828	0.989	0.919
MISS RIVER	108	1	0.947	0.986
NAT GAS	109	0.877	1	0.934
NORTHERN	110	0.875	1	0.901
NORTHWEST	111	1	0.994	0.993
PANHANDLE	112	1	1	0.938
SOUTHERN	113	0.882	0.994	0.954
TENNECO	114	0.938	1	0.960
TEXAS EAST	115	0.980	1	0.980
TEXAS GAS	116	0.883	1	0.950
TRANSCON	117	1	1	0.977
TRANSWEST	118	0.855	0.965	0.951
TRUNKLINE	119	1	1	1
UNITED GAS	120	0.711	0.996	0.802

8.11 Illustration: Model ED

8.95 *The dynamic efficiency model ED is based on the ES model, but focused the relative improvement of the efficiency of each operator (EC) and the best practice development in the sector (TC), as well as the overall productivity progress (Malmquist index M).*

8.96 *The Malmquist index is calculated and decomposed in Technological Change (TC) and Efficiency Change (EC) as explained in Section 7.8 above. The results are shown in Figure 8-12 below. As seen, the solid grey line for productivity growth (Malmquist index) shows a stationary period until 1982 and then a strong increase, denoting a good productivity development. The growth can be explained in the graph almost entirely through the golden curve for technological change TC), whereas the dashed red curve (efficiency change or catch-up) is weak and declining at the end. This means that the TSOGs improved during this period in productivity through changes in the frontier, not by catching up incumbent inefficiency.*

8.97 *What was then the substance of the frontier change? We explore this as customary using a second-stage analysis in Figure 8-13 below. The graph shows that the totex-unit cost, indexed to the initial year, declines rapidly in 1983-1987, landing at 67% of the initial value. This impressive change is enabled, among other factors, by a high investment volume in pipelines in 1986-1987, amounting to 16% and 12% per year. The technological change is therefore likely linked to new efficient investments in pipelines and compressors.*

8.98 *Indeed, the gas transmission sector went through two major reforms in this period. First in 1984, the introduction of Order 380 (Federal Energy Regulatory Commission, 1984) that stopped the "minimum-bill" or "take-or-pay" contracts that linked integrated local utilities to the TSOGs with guaranteed coverage of asset costs, whether the volume was needed or not. This created the premises for a type of TSOG-competition for gas volumes and prompted restructuring. This was followed in 1986 by Order 436 (Federal Energy Regulatory Commission, 1985) that created to Open Access Rule where clients freely could choose their gas transport provider and contract type. As the result of Order 380 had flooded the spot market with gas cancelled in TSOG-contracts, the spot market prices were lower than ever. This in turn accelerated the shift of clients from fixed long contracts with their incumbent TSOG to spot market interactions and renegotiated transport fees. The positive productivity growth is the result of both these processes: relevant and new constructions and restructuring of stranded assets.*

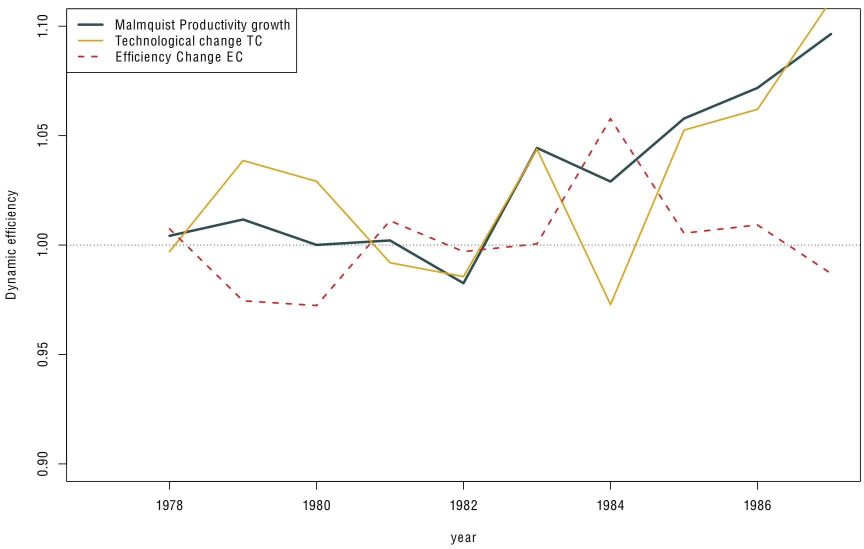


Figure 8-12 Results Model ED (Malmquist, TC, EC), FERC (TSOG, 1977-1987).

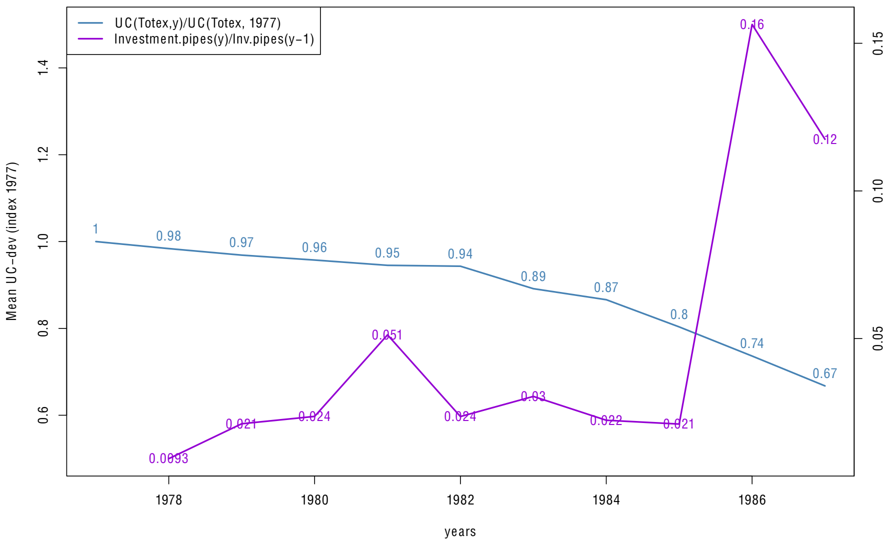


Figure 8-13 Mean UC development and investment volume in pipelines, TSOG 1977-1987.

8.12 Illustration: Model UD

8.99 *For the utilisation-focus model UD, we draw on the execution model ES in Table 8-13 to which we then add indicators related to the utilisation of the grid. The FERC dataset does not contain any parameters related to peakload or contractual capacity, meaning that we (for the example) need to concentrate on transported volume, for which we dispose of different parameters linked to total, mainline deliveries, transit transport and volume resold to neighbouring grids. After analysis, the UD model covers gas transport in two variables (sales to downstream load in grid (yVol.sales.main) and the sum of transport for transit, storage and losses (yVol.ex.main), as well as the ES-variables grid provision (pipeline capital), and compressor capacity (xCap.comp).*

8.100 *The OLS results for the model are graphed in Figure 8-14 for 1986 and listed for all years in Table 8-16 below. As seen in the table, the model is highly significant (high F-value), all coefficients are as expected positive and significant (low p-values). Thus, the UD model is already for average cost, without the efficiency estimation, a strong estimator of total expenditure for the TSOGs in the sample. Model UD here is an example, as it illustrates utilisation profiling with the data available.*

8.101 *Detailed and conventional econometric analysis[15] shows that the model is good for an application with the SFA method and that the results are significant.*

8.102 *The detailed results for the SFA-UD are listed in Table 8-18 below. The colour scale is made to illustrate magnitude with green being the highest and red the lowest in the sample. The stability of the SFA results per TSOG is striking, although the volume and investments vary over time. The highest performing TSOGs are 102, 104 and 119 (see also below in the Outlier Analysis for a discussion of 119), in each year.*

8.103 *Note the difference in analysis with respect to the models ES (only assets) and UD (using exogenous volume), where the latter allows a global dynamic assessment of the market development whereas ES is focused on pure execution (investment in anticipation of demand), but not the outcome. UD is therefore richer in its ex-post analysis of the events and causes linked to the results found.*

8.104 *To illustrate the differences with respect to ES and the sources of inefficiency in UD, consider the three TSOGs with scores below 80% in Table 8-18: PANHANDLE*

---

[15] The skew is correct for a cost function (positive), but the split of the volume variable creates a collinearity that would need to be addressed. The model has constant variance (homoskedasticity), passing the Breusch-Pagan test.

We use the panel data (10 years, 20 TSOG) and let the statistical Hausman test determine whether fixed effects by year or random effects (by operator) should be used. The test rejects the random effect variant, so we use fixed effects by year.

The assumption for the distribution is tested between half-normal and truncated normal. An LR-test does not reject the half-normal assumption that also has a marginally better BIC-score. A half-normal distribution is a conventional choice is many scientific studies using SFA.

The results for an SFA run of model UD under half-normal inefficiency and fixed effects is given in Table 8-17 below. We estimate SFA with the Error Components specification of Battese and Coelli (1992). Results show significant positive effects of pipeline volume, capital composition, and sales volume on cost. The variance ratio parameter $\gamma = 0.70$ ($p < 0.001$) indicates that about 70% of residual variation is attributable to inefficiency rather than noise. The hypothesis that noise explains all variation in cost can therefore be rejected. No significant time trend in cost efficiency was detected. This means that in the utilisation model there are no systematic trends to compensate for.

(Panhandle Eastern), COLUM GAS (Columbia Gas Transmission) and TRANSWEST (Transwestern). The three have considerably lower efficiency than in the ES model, where PANHANDLE is fully efficient. The reason is primarily linked to anticipatory investments faced with demand downturn when the US natural gas market dislocated in the early 1980s, see Figure 8-15 below. All three expanded capacity aggressively in the late 1970s under expectations of continued demand growth and long-term take-or-pay supply contracts. Following the economic downturn and unusually mild winters of 1982–83, gas demand growth slowed sharply, while contractually committed supplies and fixed network capacity remained in place. This created a sustained oversupply and underutilisation of transmission assets, compounded by high legacy gas purchase costs. The regulatory transition marked by FERC Orders 380 (1984) and 436 (1985) further shifted volume and price risk onto pipelines. In the SFA UD model that explicitly combines throughput and capacity, this structural imbalance shows up as persistently weak utilisation efficiency, different from the earlier examined execution cost efficiency. E.g. the cost of the pipeline construction and its opex is optimal for PANHANDLE, but the pipeline is redundant and over dimensioned.

Table 8-16. OLS (average cost function) Utilisation focus, UD (panel, FERC-US).

Variable	Estimate	Std. Error	t-value	Pr(> $\|t\|$)
$I(yVol.tot - yVol.sales.main)$	0.164	0.021	7.81	< 0.001 ***
$vPipe$	0.496	0.021	23.62	< 0.001 ***
$xCap\_comp$	0.339	0.018	18.83	< 0.001 ***
$yVol.sales.main$	0.174	0.021	8.29	< 0.001 ***
Constant	2.367	0.264	8.97	< 0.001 ***
Observations				200
R^2 / **Adjusted** R^2				0.981 / 0.981
Residual Std. Error				0.097 (df = 195)
F Statistic				2,556.6 *** (df = 4; 195)

Notes: Dependent variable is $dTotex.real.std.y$. Standard errors are heteroskedasticity-robust.
Significance codes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

Table 8-17 SFA results, model UD-panel, Battese-Coelli (1992) specification.

Variable	Estimate	Std. Error	z-value	Pr($>|z|$)
(Intercept)	2.6784	0.5654	4.737	2.17×10^{-6} ***
$I(yVol.tot - yVol.sales.main)$	0.2042	0.0292	6.991	2.74×10^{-12} ***
$vPipe$	0.4569	0.0338	13.537	$< 2.2 \times 10^{-16}$ ***
$xCap\_comp$	0.3261	0.0491	6.646	3.02×10^{-11} ***
$yVol.sales.main$	0.2093	0.0288	7.256	4.00×10^{-13} ***
σ^2	0.0281	0.0115	2.447	0.0144 *
γ	0.8472	0.0681	12.446	$< 2.2 \times 10^{-16}$ ***
$time$	0.0094	0.0114	0.822	0.411
Log-likelihood				232.146
Convergence				after 18 iterations
Panel size				$N = 200$ (20 cross-sections $\times$ 10 years)

Notes: Error Components Frontier estimated by iterative maximum likelihood following Battese & Coelli (1992). Inefficiency increases the dependent (logged) cost variable. Convergence was achieved after 18 iterations when the log-likelihood and parameter estimates changed less than the tolerance threshold. All 200 observations belong to balanced panels (20 TSOs observed for 10 years each).
Significance codes: *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$.

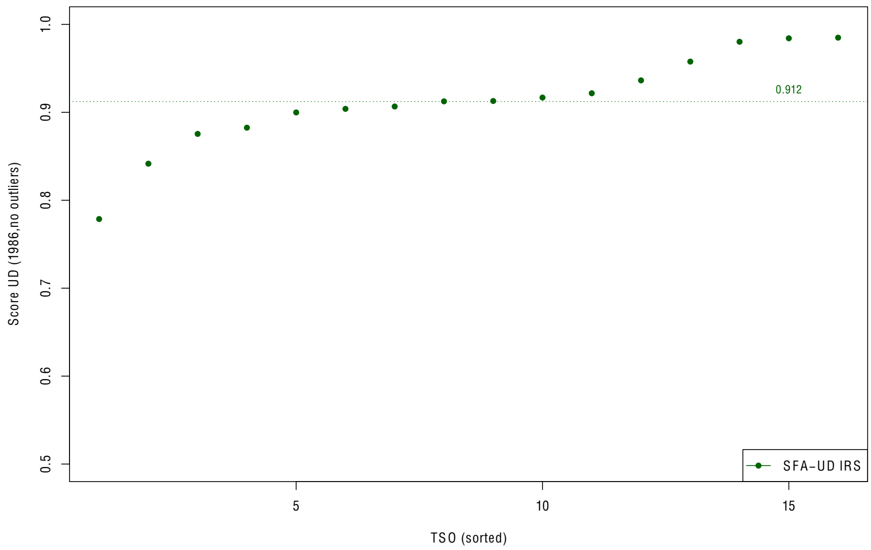


Figure 8-14 Scores SFA-UD-1986, no outliers.

Table 8-18 Results SFA-UD, 1977-1986.

tso	tso.name	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986
101	ARKLA	0.85	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.86	0.87
102	COLORADO	0.98	0.98	0.98	0.98	0.98	0.99	0.99	0.99	0.99	0.99
103	COLUM_GAS	0.80	0.80	0.80	0.81	0.81	0.81	0.81	0.81	0.81	0.82
104	COLUM_GULF	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
105	CNG	0.88	0.88	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89
106	EL_PASO	0.85	0.85	0.85	0.85	0.85	0.86	0.86	0.86	0.86	0.86
107	FLORIDA	0.83	0.83	0.83	0.84	0.84	0.84	0.84	0.84	0.84	0.84
108	MISS_RIVER	0.91	0.91	0.91	0.91	0.92	0.92	0.92	0.92	0.92	0.92
109	NAT_GAS	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93
110	NORTHERN	0.84	0.84	0.84	0.84	0.85	0.85	0.85	0.85	0.85	0.85
111	NORTHWEST	0.94	0.94	0.94	0.94	0.95	0.95	0.95	0.95	0.95	0.95
112	PANHANDLE	0.79	0.80	0.80	0.80	0.80	0.80	0.81	0.81	0.81	0.81
113	SOUTHERN	0.81	0.81	0.81	0.82	0.82	0.82	0.82	0.82	0.82	0.83
114	TENNECO	0.85	0.85	0.85	0.85	0.86	0.86	0.86	0.86	0.86	0.86
115	TEXAS_EAST	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.89	0.89
116	TEXAS_GAS	0.89	0.89	0.89	0.89	0.89	0.90	0.90	0.90	0.90	0.90
117	TRANSCON	0.84	0.84	0.84	0.84	0.85	0.85	0.85	0.85	0.85	0.85
118	TRANSWEST	0.76	0.76	0.76	0.76	0.77	0.77	0.77	0.77	0.77	0.78
119	TRUNKLINE	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
120	UNITED_GAS	0.86	0.86	0.86	0.86	0.86	0.86	0.87	0.87	0.87	0.87

Figure 8-15 Volume and pipeline growth 1977-1986 for PANHANDLE, COLUM GAS and TRANSWEST, obtaining lowest results in model SFA-UD.

8.13 Combining methods

8.105 As initially stated, it is important to determine whether the use of a single method is sufficient or whether the used of several methods in combination could bring additional information or robustness, see section 8.1.

8.106 NRA practice (DE, AT, …) and academic research (cf. Bauer et al., 1998, Jamasb and Pollitt, 2003, Nillesen and Pollitt, 2010, Andor et al., 2019, Rita et al., 2023) show clear advantages from combining parametric and non-parametric methods in regulatory applications:

1) Cross validation of results, mitigating effects of model assumptions.

2) Balancing noise sensitivity, SFA filters the impact of potential data errors and validates the average cost function quality.
3) Functional flexibility, DEA can provide fair estimates for operators in specific settings with fewer observations to estimate the necessary functional form in SFA.
4) Legal robustness: the use to two landmark methods hedges against appeals by TSOGs using alternative methods.
5) Robustness of results: DEA has a better small subsample performance, whereas SFA is stronger in the development and testing of environmental effects and dynamic impacts.

8.107 Bauer et al. (1998) states e.g. in their conclusions on regulatory use of frontier methods:

> *"As a final policy conclusion, our results suggest that when performing regulatory analysis—or really any other analysis which depends on frontier efficiency measurement—the use of multiple techniques and specifications is likely to be helpful."*

8.108 Normally, sample size and heterogeneity[16] are two reasons not to apply SFA. In AEC, the envisaged dataset of 43 TSOGs over 5 years (around 200 observations) provides support for an optimistic view on the feasibility of SFA, at least for a dynamic analysis, see also the analysis in §5.51. However, the maturity of the sector, the heterogeneity of the operators and the potential lack of balance (number of observations per operator) in the initial run seriously challenge the prospects of using SFA for a cross-section analysis (that is: calculating the scores for a single year).

8.109 Multimethod approaches can be based on two levels of method usage: *complementarity*, for which the models provide different perspectives of efficiency that together provide a more complete performance assessment, and *validation methods* that are used to cross validate, compare and analyse the results from another method. In AEC and in the literature, DEA (deterministic) and SFA (stochastic) are generally seen as complementary methods; the strength of one is the weakness of the other and vice versa.

8.110 The parallel calculation of ES with SFA includes the full dataset, output variables corresponding to those for DEA, and in addition environmental and contextual variables that are relevant and significant in the explanation of the sources and uncertainty of the efficiency scores. Given the complementary features of the methods (discussed above), this approach brings further robustness, as also suggested by the Expert Review.

8.111 Over time in a repeated analysis with increasing maturity, convergence in terms of operation and asset standards, as well as size of data set, the two methods will continue to increase the precision of the assessments for the benefit of the NRAs and TSOGs.

8.112 The application of two types of models and two methods leads to the question about which score(s) should be used in regulatory applications. From a legal viewpoint, the use of the information generated in the AEC project should be adapted to the national regulatory framework and provisions, which may put different emphasis on static (single-year) versus dynamic (multi-year) performance, as well as absolute

[16] Sample size was discussed above. Heterogeneity of data may require a high number of control variables or require very flexible functional forms that a sample may not permit to estimate.

(e.g., partial unit costs) versus relative (best-practice) scores. Regulatory practice of using multiple methods show different solutions:

1) Single unique score (by defining a primary model and/or method)
2) Average score (across defined models and/or methods)
3) Maximum score (across defined models and/or methods)

8.113 Methodologically, solution (1) assumes that there are good reasons (model, data or method) to limit the focus to a single model or method. Albeit simple to communicate, its use in frontier-based regulation could be challenged by operators evoking alternative models on the basis of its selection reasons.

8.114 Solution (2) can be defended methodologically if the models or methods are of the same character or specification, e.g., two DEA-models with equally plausible and fitting output parameters, or two stochastic models with different assumptions for the distribution of inefficiency (which is empirically non-verifiable). However, the solution is more fragile[17] if the models have different character, specifications or validity, e.g., an average of a crossectional DEA model (one year, 43 observations) and a dynamic SFA model with 172 observations and additional variables.

8.115 Finally, solution (3) the use of max(DEA, SFA) or "best-of-two", used in the German network regulation ordinance ARegV[18], can be seen as an "insurance" against data and specification errors across different categories of operators, where they are impacted differently by one or the other depending on profile (see Agrell, 2017).

8.116 *Above, we have shown the ES scores for DEA and SFA. We now turn to a practical example in which the two scores are combined, following the application in e.g Germany (ARegV) by the Bundesnetzagentur.*

8.117 *Define the maximum score as the following formula:*

$$Max(ES{:}DEA,SFA)_{kt} = \max\{DEA(ES,k,t),SFA(ES,k,t)\}$$

where DEA(M,k,t) is the DEA score for TSOG k in year t for model M, SFA(M,k,t) is the score SFA score for TSOG k in year t for model M, and 'max' is the maximum operator (e.g., max{1,3} = 3).

8.118 *Below in Figure 8-16 we combine the earlier two scores for model ES for the year 1986, ES-DEA and ES-SFA using the maximum, i.e the score marked by the green line is the Max(ES:DEA,SFA)kt as above. The DEA scores are marked by blue dots, SFA scores are red dots and the max(DEA,SFA) is a green line.*

8.119 *E.g., TSOG PANHANDLE 112 has a DEA score of 1 (fully efficient) and an SFA score of 93.8% in Table 8-15. TSOG 112 is one of the TSOGs in Figure 8-16, namely the*

---

[17] If a model has a higher statistical validity (e.g. more observations or suitable methodology), the overall estimate becomes more fragile if an average value is calculated with a model of lower validity. If a deterministic DEA and a parametric MOLS are seen as equally valid an average between the two would not be more fragile. However, if one model is seen as more valid (e.g. DEA due to flexible cost function and higher precision) an average between the two would be more fragile. If an input-oriented DEA is seen as more valid, an average with an output-oriented DEA would also increase the fragility of the score.

[18] In fact, ARegV prescribes "best-of-four" since both standardized and book-value Totex bases for both DEA and SFA are used.

third from the right. The max(DEA,SFA)-score for PANHANDLE is max(1.00, 0.938) = 1.00, which is marked by the green line at 1.00 (the blue DEA score).

8.120 *Another example, TSOG UNITED GAS 120 has a DEA score of 71.1% and an SFA score of 80.2% in Table 8-15. TSOG 120 is the leftmost unit in Figure 8-16. The max(DEA,SFA)-score for UNITED GAS is max(0.711, 0.802) = 0.802, which is marked by the green line at 0.802 (the red SFA score).*

8.121 *Notice the interaction between the two methods in Figure 8-16. For the 12 lowest TSOG scores, 11 are 'saved' by a higher SFA score. These units take profit of the assumption behind SFA that the variables are stochastic and uncertain. Part of their poor performance is therefore explained by random noise, not managerial action. On the other hand, for the 7 best performers in the set, DEA assigns 6 of them the status of 'peers' for the other with different profiles. These operators are not as highly ranked in SFA with up to 5% inefficiency as part of their performance is explained by noise since the variables are stochastic.*

8.122 *See also Figure 8-17 for an example of the scores in DEA and SFA in a joint application (DE, DSOE, BNetzA, 2015). The method is identical to the one described above and the results are qualitatively similar. For highly efficient units with specific profiles (to the right), DEA is the method yielding the highest score, for units with generally poor performance (to the left), SFA is more "forgiving" as it deducts a part of the cost difference as a result of noise (random uncontrollable events). Taken together, it hedges against uncertainty in the reporting of poor units (for who the outlier analysis in DEA would not intervene, see below) and in parallel it does not devaluate the performance of operators with specific profiles, for whom the functional form flexibility of DEA is important.*

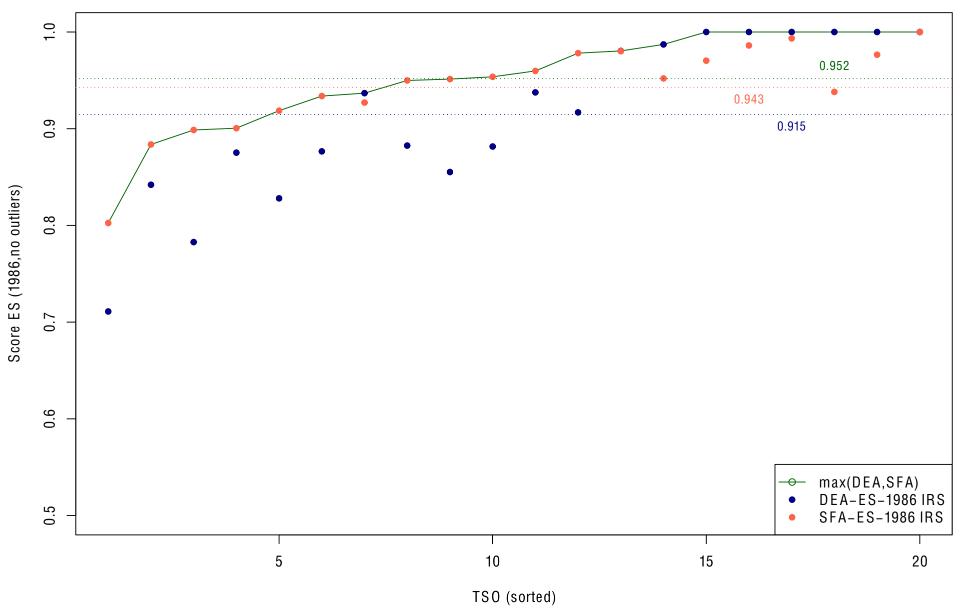


Figure 8-16 Max(SFA,DEA) ES-IRS-1986, no outlier removal.

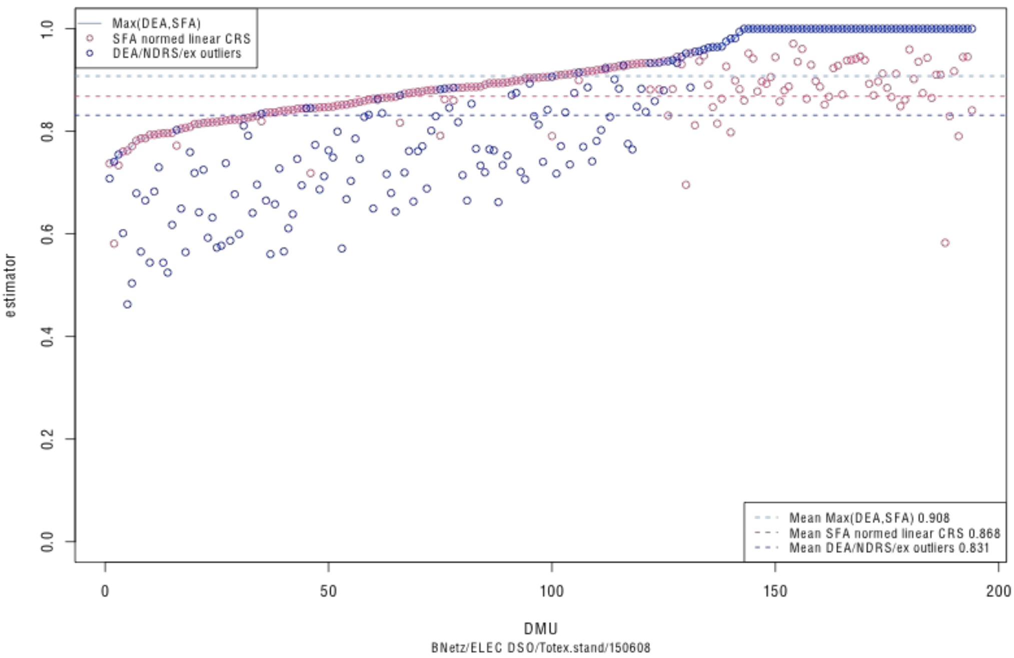


Figure 8-17 Example of DEA-SFA combination (DSOE, BNetzA, 2015).

8.14 Outlier analysis

8.123 Data validation (see below) involves several methods to identify, analyze and validate data prior to the release of the final analysis data set.

DEA

8.124 Established methods such *as super-efficiency filters, peeling and dominance tests* (see TCB21 Methods, Section 7.1 and Bogetoft and Otto, 2011) should be used in the DEA runs on the full data set to identify and exclude outliers from the data that otherwise would perturb the frontier with observations that potentially lack reproducibility.

8.125 In DEA, outliers are considered fully efficient and assigned the score of 100% for the year concerned. Note that DEA is invariant and insensitive to data errors or outliers that *worsen t*he unit's performance by increasing cost and/or decreasing output. Thus, adding an inefficient unit (perhaps under major restructuring) has no impact on the quality or precision of the assessment of the other units.

8.126 We recommend following the current scientific practice for outlier detection and carefully examine outliers in data reviews, involving both consultants, TSOG and NRA, to determine whether reporting errors could be a potential cause.

SFA

8.127 In AEC, the data set use should be subject to parametric outlier filters, such as Cook's distance, identifying and excluding observations that fall outside the standard acceptance levels. Contrary to the outlier process in DEA, where peers are explicitly defined and can be treated, the parametric review in SFA is aiming at protecting the results of the cost function shape and the noise estimation. This means that a very large unit with a specific profile may be identified as parametric outlier, not due to data errors, but for the leverage the unit exerts on the estimation process due to its size and position. SFA is not invariant to data errors: introducing poor performance data may have an impact on the shape of the cost function (less likely) and the size of the noise term for the cost estimation (very likely). This effect is natural and intended: presence of poor data is interpreted as stochastic (random) noise and assumed endemic in the dataset.

8.128 As for DEA, the exclusion in an SFA model should lead to an information and validation process.

8.129 *We apply the super-efficiency filter to the dataset above for the model ES (DEA), and the parametric test Cook's Distance for the SFA runs of models ED and UD. The results with ID for identified outliers are listed in Table 8-19 below. We note that the number of outliers for this data is low: 0 or 1 outlier (≤ 5% of observations) except for one year in DEA and only one outlier for 11 years in SFA. This shows high stability of the data, which is consistent with the stability of the scores in the models.*

8.130 *The outliers identified for the reference year 1986 can also be confirmed by other data[19]. The outlier analysis methods therefore have face validity in this data set.*

Table 8-19 List of outliers in DEA-ES ,SFA-ED and SFA-UD.

	Outliers (ID)		
Year	DEA-ES-IRS	SFA-ED	SFA-UD
1977	-	-	-
1978	-	-	-
1979	-	-	-
1980	108	-	-
1981	108	-	-
1982	-	-	-
1983	102, 104, 108	-	119
1984	102	-	-
1985	102	-	-
1986	108	119	-
1987	108	-	-

8.131 We recommend applying parametric outlier detection also to SFA runs and to document these results for further analysis.

[19] We note that in the illustration for 1986, TSOG 108 is detected as an outlier in DEA-IRS. TSOG 108 (Mississippi River Transmission, MRT, MISS RIVER) is a relatively small interstate TSOG (about 1,968 miles of pipelines) that operates from Perryville, Louisiana, through Arkansas, to St. Louis (Missouri), with branches into Illinois. They run 10-12 compressor stations along the main line with a total power of about 86 kHP. In 1985-1987 there was a FERC litigation against MRT in a rate case concerning tariff methodology and rate design. MRT was ultimately sold to ARKLA in 1988. Although identified automatically by the method, there are clear reasons to confirm this exclusion also based on prima facie evidence from the period.

For SFA the only unit identified in 119 TRUNKLINE due to its size and leverage. However, removing the TSOG only affects the cost function in SFA-ED marginally. TRUNKLINE (119) is a clear peer both in model ED and in the utilisation model UD.

8.15 Peer analysis

8.132 Each run is followed and documented with a peer analysis. At this stage, the profiles of inputs and outputs per group, non-peers, peers and outliers are determined and compared. The geographical and structural origins of the TSOGs are also examined and documented.

8.133 The peer analysis may reveal potential bias in the model, such as an unnatural prevalence of peer units from certain regions or organisational forms, or with specific exogenous structures that should have been controlled for. If such bias is revealed, the model specification is reviewed to ensure that the assessment is free from any a priori bias towards certain operators or structures.

8.134 *The following example shows how peer analysis could work in our application for FERC 1986 (Appendix F).*

8.135 *In the ES-model for 1986, there are 5 peers and 1 outlier, leaving 14 non-peer TSOGs. The descriptive statistics for the three groups are given in Table 8-20 below. The initial peer analysis shows that the peers and the outlier generally are smaller than the non-peers, both in assets; pipeline length and compressor capacity; and in volume delivered. The peers are dependent on the IRS (NDRS) assumption, meaning that the larger efficient TSOG do not dominate the smaller. For the very smallest TSOG, the number of potential comparators-peers is very limited, and this is attributing two of them the status of peers. (Note that the outlier was not removed from the set for the run, which would have been the case in a real application.)*

8.136 *As such, this effect is not a bias per se, but an effect of the IRS assumption. However, at the next stage we study how many TSOG each peer dominates in Table 8-21 below. Here, the analysis shows that most of the non-peers (11) are dominated by a single peer, 119 TRUNKLINE, which is the second largest in the sector at the time. This information would of course prompt an in-depth investigation of the nature and sources of the efficiency of TRUNKLINE to ensure a representative peer.*

8.137 *In the next step, we investigate the geographical origin of the peers compared to the TSOGs in Table 8-22 below. The regions are defined in Appendix F. We see that 3 peers are from the Gulf-Texas Core, which is a dense region with in all 8 interstate TSOGs. The expected proportion would be 2 peers for GLF, but one additional is not strange. However, we also note that both TSOGs from WST (Western-Mountain) are peers, whereas the expected proportion would only be 0.5. The regions EAS (Eastern-Appalachian) and SEA (Southeast-Florida) have no peer, although the expected proportion would be 1.25. Same proportion for MWD (Midwest-Central), where the only eligible peer (smallest TSOG in the set) is classified as outlier and will be removed. We summarize the locational and size findings in Figure 8-18 below where the particular status of peer 119 in region GLF is confirmed.*

8.138 *The preceding analysis indicates a potential geographical bias to a mature infrastructure region (GLF) and one geographical zone (WST) for smaller TSOGs. In AEC, we use environmental correction factors to address this problem in run R2, for the illustration we do not have such data. The outcome of the peer analysis in the illustration is therefore suggesting environmental differentiation. The procedure and the two findings (size and locational bias) are general in many studies.*

Table 8-20 ES-input-output data for peers, non-peers, outliers, 1986.

Variable	Statistic	Non-peer TSO	Peer TSO	Outliers
	n	14	5	1
dTotex.real.std.y	Min	67,670,794	57,316,589	49,842,971
	Mean	212,595,104	169,960,862	49,842,971
	Median	174,901,166	121,619,542	49,842,971
	Max	544,135,698	439,065,387	49,842,971
yPipe.length	Min	3,088	2,797	1,997
	Mean	7,954	5,464	1,997
	Median	7,503	4,020	1,997
	Max	15,420	10,315	1,997
yComp.power	Min	89,895	170,320	85,795
	Mean	557,015	459,801	85,795
	Median	428,243	341,000	85,795
	Max	1,307,485	983,739	85,795
yVol.tot	Min	284,168	342,818	145,656
	Mean	769,400	644,470	145,656
	Median	807,661	452,527	145,656
	Max	1,353,127	1,490,656	145,656
ROR	Min	0.0759	0.1141	0.1509
	Mean	0.1265	0.1274	0.1509
	Median	0.1245	0.1278	0.1509
	Max	0.1562	0.1398	0.1509

Notes: Values are grouped by frontier status from DEA-ES-IRS-1986 scores.

Table 8-21 Peers with state, region and number of TSOG dominated.

tsoid	tsoname	state	region	status	#dominated
102	COLORADO	CO	WST	PEER	2
108	MISS_RIVER	MO	MWD	OUTL	5
111	NORTHWEST	UT	WST	PEER	7
112	PANHANDLE	TX	GLF	PEER	7
117	TRANSCON	TX	GLF	PEER	5
119	TRUNKLINE	TX	GLF	PEER	11

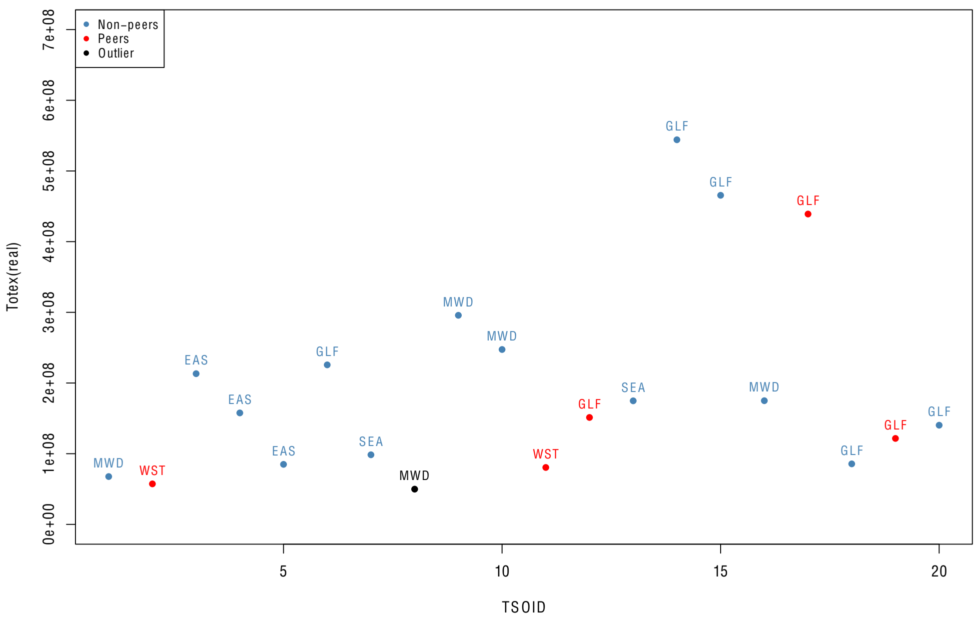


Figure 8-18 Size (Totex) for peers, non-peers and outliers in DEA-ES-IRS-1986.

Table 8-22 Regions for peers, non-peers and outliers, 1986.

Region	Non-peers	Peers	Outliers	Total TSO
EAS	3			3
GLF	5	3		8
MWD	4		1	5
SEA	2			2
WST		2		2

8.139 We recommend performing in-depth validation of the peers obtained in the AEC with respect to technology, scale, country of origin and profile for asset and services. In the interest of transparency, we propose that AEC discloses the identity of the peers in the respective models, at least for the participants.

8.16 Planning of runs and sensitivity analyses

8.140 The output from the AEC to the participants will be differentiated into *runs* and *sensitivity analyses.*

Runs

8.141 A *run* is a complete implementation of all calculations necessary for the determinations of scores for all TSOGs , including data processing, standardisation of capex and opex, normalisation of assets, outlier detection, determination of reference sets, calculation of scores in chosen methods for chosen returns to scale, individual feedback on scores for each TSOG, peer analysis with disclosure of composition, dual variable analysis, documentation of the distribution of scores in graphs, mean and median. The individual information is disclosed in a confidential report, the aggregate information is documented in the open reporting.

8.142 The list of the AEC runs is found in Table 8-23 below. Note that run R5 is on request only and after definition of a national WACC (if necessary, per year) to be used in the run.

Table 8-23 List of specific runs R.

ID-R	Name	Parameter values
R1	Base case	Base case
R2	Base incl environmental corrections	Environmental corrections
R3	Base excl legacy investments	Legacy investments
R4	Base with premium decommissioning	Flat deduction for decommissioned assets
R5	Base with national WACC	WACC replaces standard interest rate

8.143 We propose to organize five runs, thereof one optional at request, to fully separate estimation depth (green field vs environmental correction) from policy parameters (legacy investments, decommissioning incentives, local WACC).

Sensitivity analyses

8.144 The application of results in regulation requires a full assessment of the sensitivity to parameters that may be set differently, country specific, or changing over time. For this reason, it has been the standard in projects as TCB21 to provide and report on extensive *sensitivity analyses*.

8.145 A sensitivity analysis is a calculation of the impact of changing a parameter, variable or reference set on the base score (R1). The set of planned sensitivity analyses SA is given in Table 8-24 below. The information from a sensitivity analysis is intended to illustrate the robustness of the base assessment to the assumption with regard to a choice of standardisation, normalisation, indexation or exclusion/inclusion. The analysis is made *ceteris paribus* meaning that no other factors are changed in the setup, such as reference sets, outliers or methods.

8.146 The information disclosure in Table 8-24 can be of two types: (i) *aggregate scores*, meaning a graph and a table that indicates how minimum, maximum, median and mean scores R1 are impacted by the change, and (ii) *individual scores* that show the detailed impact of a change on the score for an individual TSOG, in addition to the aggregate results.

8.147 Aggregate scores are documented in the final AEC report, individual scores are provided in the individual reports for each TSOG.

Table 8-24 List of sensitivity analyses SA.

ID-SA	Parameter/variable	Range/series	Report	Impact metrics
SA-PI	Inflation index PI	Alternative PI	Changes in score	Min, max, mean
SA-LC	Labour cost index LCI	Alternative LCI	Changes in score	Min, max, mean
SA-RA	Interest rate, r	0.5% … 5%	Changes in score	Min, max, mean
SA-OA	Overhead allocation key	Alternative keys	Individual scores	Individual impact
SA-AP	Asset ages: pipelines	10… 50 years	Changes in score	Min, max, mean
SA-AC	Asset ages: compressors	0… 30 years	Changes in score	Min, max, mean
SA-NG	NormGrid weights	Pipeline/compressor	Changes in score	Min, max, mean
SA-EN	Environmental factors	Inclusion/exclusion	Individual scores	Individual impact
SA-MS	Alternative output var	Inclusion/exclusion	Individual scores	Individual impact

8.148 We propose to prescribe sensitivity analyses for relevant parameters as above and to provide the general results (mean and ranges min-max) in the common report and the detailed individual results (score, mean, min-max) for each TSOG in an individual report.

8.17 Conclusions

8.149 The proposal to use both DEA and SFA in the efficiency analysis for AEC based on two types of models:

8.150 For the deterministic execution-based model primarily DEA is recommended, but SFA should be tested. Both static and dynamic efficiency calculations

8.151 For the stochastic utilisation model, SFA is recommended

8.152 The use of two methods aims at using the data to the maximum and provide complementary results. DEA and SFA are different methodologies, based on different assumptions and the efficiency scores with the two methods are therefore likely to differ.

8.153 DEA is a strong reliable workhorse for heterogenous data of medium size, without technical assumptions on cost functions. DEA is proposed as primary method for the static (one year) efficiency calculations, using primarily deterministic, execution-based model(s).

8.154 SFA is a powerful analysis tool to understand the causes and development of efficiency in a sector, requiring somewhat larger datasets and analysis resources. SFA is proposed as primary method for the dynamic (multi-year) analysis, using stochastic models with both assets, usage and contextual-environmental factors included. The use of SFA for the stochastic utilisation model is appropriate as usage-based outputs display some randomness that is better controlled with SFA.

8.155 The two methods complement each respective analysis by acting as secondary methods for cross validation: SFA contributes to validating the DEA method results and the environmental adjustments therein, DEA (Malmquist) serves as to comparator for the dynamic SFA analysis.

8.156 As informational complement in e.g. investment and cost reviews by NRAs, we propose index number approaches (Unit Cost, KPIs) to be reported.

8.157 All final calculations are recommended to undergo stringent outlier filtering.

8.158 It is recommended to use the full data set whenever possible, if necessary, with measures taken to control for possible unbalanced years.

8.18 Comparing AEC to international best practice

8.159 The determination of efficient cost levels and their improvement over time is but one of the components of an effective regulatory strategy. Nevertheless, well performed and with the required features corresponding to NRA needs and with a robustness for legal challenges, a regular cost efficiency study could be a fundamental pillar of incentive regulation, as discussed e.g. in Agrell (2015) and references. Remains to objectify what such a study should contain. In Haney and Pollitt (2009) the authors develop a Best Practice Index for regulatory benchmarking (BPI), reviewing both regulatory practice and methodological advances relevant to the tasks at hand. BPI includes metrics, cost definitions, data sets, methods, measures for cross validation and uncertainty, as well as competence levels. The eight dimensions of the index (Table 8-25) were assessed through a questionnaire to NRAs in Haney and Pollitt, (2009) and further analyzed in Haney and Pollitt (2011).

8.160 We grade the AEC proposal above in accordance with the BPI criteria in the right-most column in Table 8-25. The AEC receives highest points on 7 of 8 dimensions, including method choice, number and type of methods, cost model, panel data, management of uncertainty, management on environmental effects, and data sets. An aggregate score in Haney and Pollitt (2011) groups scores between 6 and 8 points into the category H(igh). ACER can therefore trust that the proposal relies upon scientific and regulatory best practice in its methodological design.

8.161 The final dimension (8) concerns the internal vs external resources dedicated to the modelling and calculations. We recommend that ACER pursues the building of a network of regulatory analysts to gradually participate and extend the methodological work enabled by the regular AEC.

Table 8-25 Best practice index for regulatory benchmarking (Haney and Pollitt, 2009, 2011).

Dim	Indicator	Pts	Score AEC[20]
1	Current use of DEA, COLS, SFA and/or process/activity benchmarking; 0.5 for concrete plans to use one or more of these techniques	1	1
2	Use of more than one of above benchmarking techniques in most recent price review	1	1
3	Totex modelling (i.e. total expenditure: operating costs plus a measure of capital costs)	1	1
4	Use of panel data	1	1
5	Dealing with uncertainty: full score for DEA, SFA, COLS or process/activity if tests for well-behaved functional form, CIA or specific adjustment	1	1
6	Incorporation of environmental factors	1	1
7	Use benchmarking techniques and have either ≥ 30 companies, or 30 companies and use of international data (large dataset)	1	1
8	Mixture of both external and internal analysis =1; sophisticated internal analysis (i.e. using one of advanced benchmarking techniques) = 0.5; external analysis only = 0	1	0-1
	Total best practice score (max)	**8**	**7-8**

[20] Assessment by the authors of this report in accordance with the definitions in Haney and Pollitt (2009).

9. Benchmarking gas under transition

In this Chapter, we discuss how the new challenges discussed in Chapter 5 can be addressed in the AEC to obtain valid values for the metrics from Chapter 7 using the methods and models in Chapter 8. The Chapter results in a series of adaptations and extensions to a regular cost-efficiency study to fully capture the new challenges.

9.1 Changes in gas flows

Impact on costs, outputs and investments

9.01 The development of biogas, G2P and P2G has an impact on outputs, leading to a higher load on selected pipe sections, new entry and exit points, and new connecting pipes. Opex and Capex increase both with additional connection points. Local outputs (energy injected, energy delivered, peakload) increase, and connecting pipes increase pipeline length/volume. Similarly, the LNG terminals led to new entry points and connected pipelines to these entry points. New investments were made in the reverse flow capability of pipelines. The implementation of reverse flow capabilities may require new compressors.

Implication for benchmarking

9.02 The standard is to reflect new entry and exit points as any other existing entry or exit point, with corresponding effects on assets and outputs. If new connections are subsidized, however, they must be treated consistently. This can be done either by including the network-related subsidies in the benchmarked cost base or by an appropriate reduction of outputs. Benchmarking models should account for the growing role of renewable gases and the corresponding infrastructure changes. It is recommended to differentiate, in data collection, natural gas (H and L) vs biogas connection points and other relevant assets. Extension of unit-cost measures for these new types of assets could be provided.

9.2 Decommissioning of underused assets

9.03 A consequence of the decreasing gas volume is that some network assets might be decommissioned. They may still have a residual book value or RAB-value at the date of the decommissioning, becoming *stranded assets*. This could be mitigated if assets could be refitted for other purposes (e.g. hydrogen). However, it is unlikely that all assets will be redeployed.

9.04 It is an important regulatory issue to determine who will pay for these stranded assets, as the write-off could be shared between consumers, firms, the financial system, owners and the state. For instance, regulators can pass the cost to consumers by increasing the depreciation rate, with the consequence of accelerating the transition from natural gas to other modes of energy.

9.05 In the benchmarking context, it is important to ensure a coherent treatment of stranded assets as the real annuity of the investment stream is derived from the RAB, which should be comparable across operators. Attention should be paid to situations where TSOGs could pass from inefficient to efficient peers simply by a public or regulatory intervention in this manner.

9.06 Likewise, a neutral policy of matching exact values could lead to perverse incentives for keeping 'investment-efficient' assets and dismantling sunk inefficient assets. Removing an efficient asset from the asset base may lead to a decreased Capex-efficiency although it is a socially desirable action.

9.07 In the AEC, the Capex related to decommissioned assets is cancelled, decommissioning cost in Opex is identified and differentiated as a separate category, and corresponding output is cancelled from the year of decommissioning. The full or standardized investment value is deducted from the Capex-base (see section 10.5 on legacy investments) for the decommissioned asset.

9.08 To assure that incentives are provided for economically rational decommissioning, a specific run (R4 in Table 8-23) is to be made where for each asset removed the maximum of the actual investment and a mean norm value is deducted from the capital (and thus reducing Capex). This means that all decommissioning of all pipelines is incentivized (see footnote 4 above) in the benchmarking in run R4.

9.09 The parametrisations for run R4, that is the level of incentives provided, will be determined after analysis of unit costs.

9.3 Preparation of assets for conversion to H2 pipelines

9.10 Some of the existing gas pipes will be decommissioned, but others may be repurposed for H2. The value of the first category of asset depends on the regulatory treatment of stranded assets (see above), the value of the second category depends on the possibilities to transfer gas pipes to hydrogen, and the price (potentially regulated) that will be charged for that. As for stranded assets, the regulatory treatment of refitted assets is important for understanding the TSOGs ' incentives and for making a proper comparison between them. Article 5 of Regulation 2024/1789 provides some rules on the asset transfer value, but their application may vary between countries.

9.11 When transferring an asset from natural gas to hydrogen, the hydrogen operator should pay for this asset. The default value is, again, the residual asset value in the RAB (the non-depreciated value at the time of the transfer). However, this could vary. Some NRAs could opt for a higher or a lower value, which could result in users of natural gas paying for the difference.

9.12 Furthermore, for comparison, it is important to clearly distinguish at this stage the assets that are used for gas and those that are used for H2. In the case of H2-ready pipes, the costs that should be allocated to gas should be separated from those that should be allocated to H2 and that H2 premium should not be included in the benchmark. For this reason, the data request (see note T03C and synthesis D03) is proposed to be enhanced by an indicator for H2-readiness along with an (transitionary) investment category in the financial reporting sheet. Furthermore, it is important to determine the timing of the transfer to H2 as most new pipelines are H2 ready.

9.13 For the benchmarking, it is important to differentiate, in data collection, natural gas pipelines from other assets that are transferred from natural gas transmission to H2 transmission operators with equivalent removal of Capex and Opex relating to these assets.

9.14 A *greenfield* investment made by an H2 operator has no impact on the TSOG assessment.

9.15 However, in the case of a *brownfield* investment in H2 networks, a TSOG is ceding a site or route already permitted and prepared for the deployment of new infrastructure. If the H2 operator pays for the brownfield equipment, the site or part of the land preparation, this transaction should not intervene to bias the assessment of the efficiency of the operators. A first rule is that transactions of this type, before inclusion in a TSOGbenchmarking, must have equity and comparability. Negotiated or arbitrary transfer prices may create efficiency frontier targets that are unattainable and dependent on national political decisions on cost recovery.

9.16 An investigation was proposed to ENTSOG to carry out a small pilot group of TSOGs to determine whether these risks are material and, if so, how these issues are currently managed among TSOGs. The TSOGs did not respond to the request.

9.17 There is consensus among NRAs to consider that the default not leading to cross-subsidisation between operator is the residual asset value, see ACER(2021b). However, if there are deviations from this value, TSOGs could make a profit when selling assets, or it could be users that pay the difference, and it will have an impact on the efficiency. For that reason, the rule for transferring assets is recommended to be neutral. For the purpose of the benchmarking exercise, the rule for transferring values is proposed to be uniform across TSOGs , e.g., based on residual values or book values, and be accurately translated into RAB/Totex of TSOG, if these costs are not borne by or transferred to the H2 TSOG. In this case, the operation has no impact on the efficiency score of the operators.

9.4 Decrease in gas volumes

9.18 Following the above discussion on the energy transition and its impact on gas transport, we could expect a decrease in the demand of gas, that is a decrease in the volume of gas transmitted and, in the longer run, a decrease in the number of exit points. The evolution of demand and entry points depend on many factors, most of them being uncontrollable by the TSOG. TSOGs have to size their network to match the actual network use. This implies, in the long run, a decrease in network size and in the network capacity.

9.19 But assets are usually installed for long periods, with some assets having a life cycle over 50 years, so matching capacity to actual use cannot be done easily in the short run. Therefore, TSOGs with older assets are likely to have over-capacity in a context of decreasing demand while TSOG with more recent assets can dimension the assets to the actual demand. In the efficiency analysis, if the actual demand (gas volume or entry point) is used as an output, the TSOGs with the most recent assets is likely to appear more efficient, as their asset capacity more closely match the actual output.

9.20 In an environment of decreasing use of infrastructure having utilisation metrics in a benchmark implies that TSOGs are rewarded (or penalized) for bringing the size/capacity of the network in line with decreasing demand.

9.21 For that reason, the deterministic capacity-based efficiency model ES in AEC is proposed to use only asset-based outputs instead of utilisation-based outputs to compare like with like.

9.22 However, to provide information on, and incentives for, the correct asset intensity facing fuel substitutions, AEC will also incorporate a stochastic panel model UD, covering several years of operations. In this case, to monitor the volume transported and peak load development in the natural gas sector and the adaptation of assets to outputs, usage-based outputs will also be used. The inclusion of usage-based outputs in the stochastic model may capture how TSOG can adapt to changing circumstances.

10. Harmonisation of capital expenditure

In this Chapter, we discuss the measures to control for heterogeneity of assets and investments in the regulatory environments and in the accounting practices. The Chapter creates a catalogue of capital structural adjustments to harmonize the benchmarked capital expenditure data for a fair comparison.

10.1 Scope of investment

10.01 We will define (see Section 4.3) the scope of activities that enter in the perimeter of the benchmark. Assets related to activities that are out of the scope will not be included in the benchmark. Similarly, we described above that investments and capital equipment related to H2 will be excluded from the benchmark.

10.02 *Acquisition of land* and settlements of leases and compensations to landowners for grid installations differ among the countries depending on legal framework, real estate prices, type of land required and timing. To address this issue of heterogeneous and country-specific costs, these costs are excluded from the benchmarked cost.

10.03 For the *scope of investments,* the question is whether to include all or part of the investment stream related to the relevant activities.

10.04 Intangible investments, land and right-of-way and investments in plants for out-of-scope activities (H2, LNG terminals, storage,…) are proposed to be excluded from the investment-base in the benchmarking.

10.2 RAB

10.05 Regulators have different treatments for different categories of assets (working capital, assets under construction and leased assets and assets partially financed by third parties). As a result, the composition of the RAB varies from one country to another. For the benchmarking exercise, it is necessary to have a uniformized measure for the RAB and define what should be and what should not be included in the RAB since it is the basis for the comparable Capex input.

10.06 To address these issues of comparability across TSOG, we propose to define (1) a method to evaluate real annuities (Capex), (2) the scope of investments considered in the perimeter of the benchmark and (3) to determine an appropriate value for the investments included in the perimeter.

10.3 WACC

10.07 The WACC results from a methodological approach within each NRA to assess, determine or adjust the appropriate level of renumeration for the regulated asset base. There are different possible methods to calculate the WACC (determination of the risk-free interest, gearing factors, beta, etc.) and to index the WACC to financial reference variables, influencing the total allowed revenue, the composition of the capital and the size of the RAB.

10.08 For the benchmarking exercise, for the purpose of comparison, we propose to use a standardized WACC based on European sector data applied to all operators.

10.09 In addition, to control for potential distortions in the opex/capex mix we propose a specific run (R5 in Table 8-23) made using national WACC for the TSOG: this means that the national WACC is used as capital cost for all TSOG in the reference set when calculating the target cost, corresponding to the use of local factor prices.

10.4 Real annuities

10.10 The TSOG renew their assets at different times, leading to different actualisation of the investment costs. If nominal (book) values are used, the TSOG with the oldest asset will appear cost-efficient up until the time when reinvestment happens, then fall back as inefficient. This "roller-coaster" pattern is an artificial heterogeneity that does not relate to managerial inefficiency but is caused by the investment cycle timing.

10.11 The nominal value of an investment undertaken in 2000 say at 1.0 MEUR corresponds to about 1.6 MEUR in 2025 (German inflation). Using nominal book values and straight-line depreciation creates a strong bias against new investments and reinvestments, where the peer units by definition will be the TSOG with the oldest and most depreciated assets. Nominal values also create a strange distribution of capex where the capex decreases over time although the asset has the same techno-economic usage value, thereby penalizing new investments.

10.12 *Example: An asset has a techno-economic life of 10 years, acquired in 2000 at 1 M€. The average inflation rate is 2% per year and the real capital cost is 3% per year. The asset is replaced in 2010 by an identical asset at a nominal investment of 1,289 M€ (1 M€ at 2000 value). The Capex for the asset is depicted in Figure 10-1 in blue for the nominal capex (depreciation and cost of capital), in black the nominal capex after adjustment to CPI in 2000 and the real annuity (3% WACC over 10 years) in orange. As seen, the use of nominal values creates a "roller coaster" where the lowest capex is found just before the replacement, then followed by a sharp hike. Assuming a constant Opex, then the Totex will vary accordingly. Thus, such a method will penalize investments, younger grids and confuse the analysis. The real annuity method results in a constant real term per year (since the replacement has the same value) and is indifferent to the investment cycle: two TSOG in different cycles with the same investment value in real terms will have the same efficiency, which is the purpose of the analysis.*

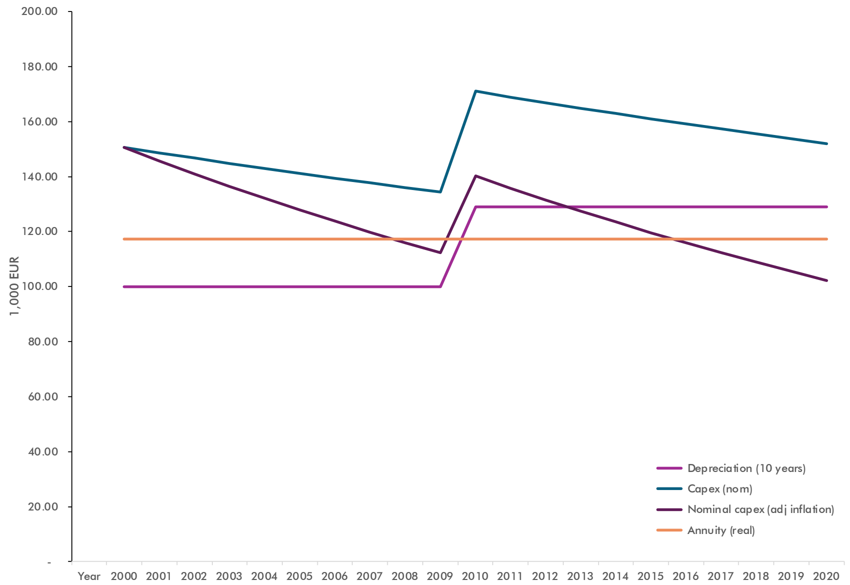


Figure 10-1 Nominal investments with inflation adjustments vs real capex annuities, example.

10.13 *Conclusion on definition of Capex:* to achieve comparability across time and the irrelevance of investment cycle, the Capex used must be based on a real annuity spread over the standard techno-economic life of the asset. In this manner, the Capex for a 10-year-old asset and a new asset will be equal and comparable. This is possible by using the original investments by year in applying an inflation correction to the reference year and using a standardized method for depreciation.

10.14 The TSOG have assets with different mean ages. AEC is proposed to collect data on asset ages, overage assets and share of new/old assets by category. These indicators will be used to investigate analytically the impact of aging on Totex and Opex over time. If a dependency is revealed, it will be considered for a correction of the normalized grid output as to create a lifecycle-invariant assessment.

10.5 Legacy investments

10.15 Past investments, prior to deregulation, were not always undertaken with an efficiency focus. For instance, pre-deregulation decisions may have been prompted by other owners and for national or non-economic reasons. Furthermore, investments in transition states prior to EU-membership were in some cases subject to hyperinflation or non-market prices for labour or equipment

10.16 These observations, which are valid for TSOG and important for incentive regulation of future investments and operations, call for a periodized analysis of the past capex. The relevance of determining the exact investment efficiency for assets that date more than 30 years ago and to assure the comparability of their market conditions are likely less important than the comparability of recent investments and new assets. It would therefore be informative to analyze the impact of these **legacy investments** in the benchmark by neutralizing their impact in one of the runs.

10.17 A similar problem is found for opening balances that may be biased for several reasons. First, some operators may have opening balances without the actual investment cost (missing or arbitrary value), which means that they potentially have a large asset base without the corresponding cost of capital in their books. Second, some operators may have artificially low opening balances, for instance because the assets were ceded to the TSOG at deflated price. This will bias Capex downwards to potentially create a peer with an unattainable cost target. Third, some operators may have excessively high opening balances, therefore having a very high level of Capex, making them potentially inefficient because their assets were over-evaluated. The use of biased estimations for the opening balances in the efficiency analysis will lead to misestimation of the efficiency with, for instance, operators having an artificially high efficiency score because of underestimated assets.

10.18 To address these problems, we need additional tools to study the investment efficiency using comparable recent data as well as the full asset base.

10.19 The AEC is proposed to include all the investments, legacy and recent, in the base run R1.

10.20 A new approach is proposed implemented in run R3 where the AEC will calculate the same model but with legacy investments neutralized to standardized values to understand the impact on efficiency. All investments for all operators, prior to a break year are reset to normalized values that correspond to unit costs for an average operator. Opex is not affected as all assets in use are part of the physical output. With such a method all the assets that have been installed prior to the break year have a comparable value and the comparisons are not biased. This permits an analysis of efficiency under comparable regulation.

10.21 The break year may be defined as either the entry in force of the first gas directive 1998 or the accession year to the EU.

10.22 *Example: Consider two TSOG, A and B for the period 1980-2024. Both have invested constantly in 10 km pipeline each year throughout the period, but at varying unit costs (Figure 10-2). Prior to deregulation (2000), TSOG A benefitted from subsidized investments in an integrated structure with biased raw material prices. On the other hand, TSOG B was subject to random investment shocks, imposing overpriced take-overs of existing assets from acquired grids and using internal low-productivity labour. From deregulation in 2000, both TSOG undertook investments with random unit cost but equal mean (1,000).*

10.23 *Including the full investment stream yields an overall unit-cost (proportional to Capex) that renders TSOG B hopelessly inefficient towards TSOG A (Figure 10-3).*

10.24 *We now apply the suggested method: all investments during the period 1980-2000 are revalued using a standardized unit cost (here assumed to 500). This means that the unit cost Capex for both TSOG at the beginning of the deregulation (2000) is identical (i.e., 500) whereas the real investment impact the rest of the horizon. The resulting efficiency score (or unit cost Capex) in*

10.25 *Figure 10-4 are then almost equal, which is indeed the correct assessment of the managerial efficiency from 2000 to 2024.*

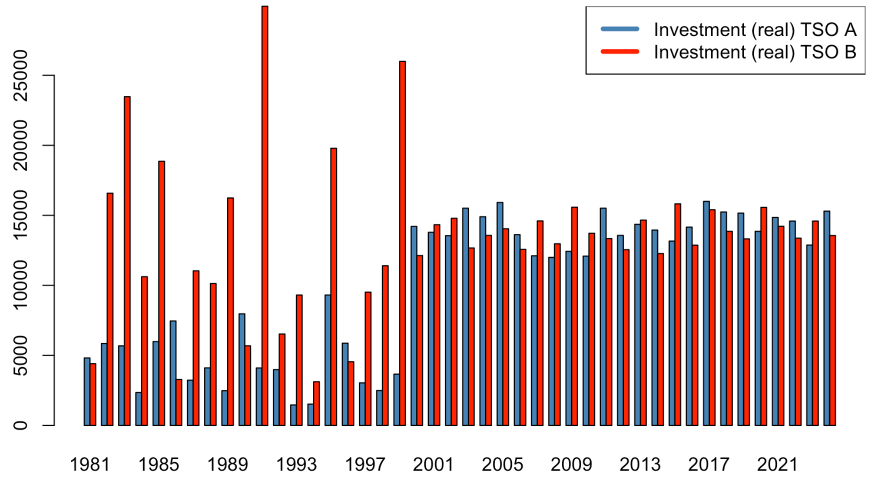


Figure 10-2 Real investment by year, TSOG A and B.

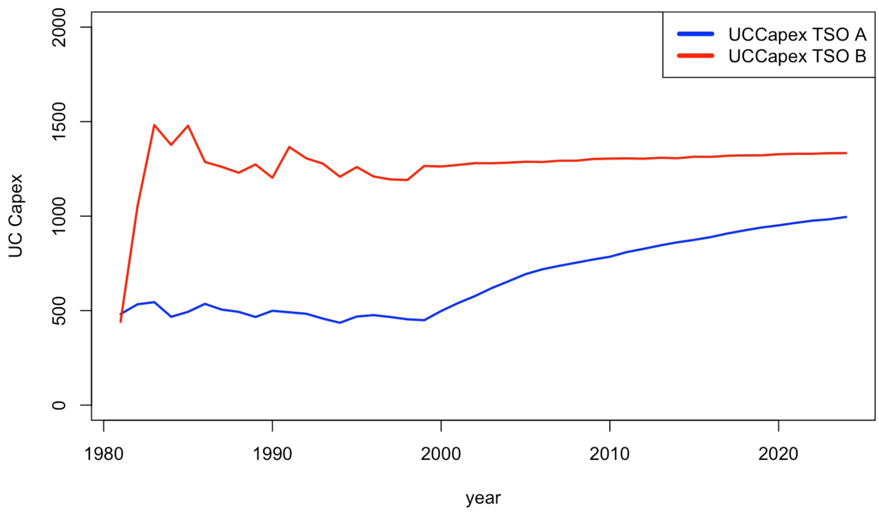


Figure 10-3 Unit cost Capex with full investment stream, TSOG A and B.

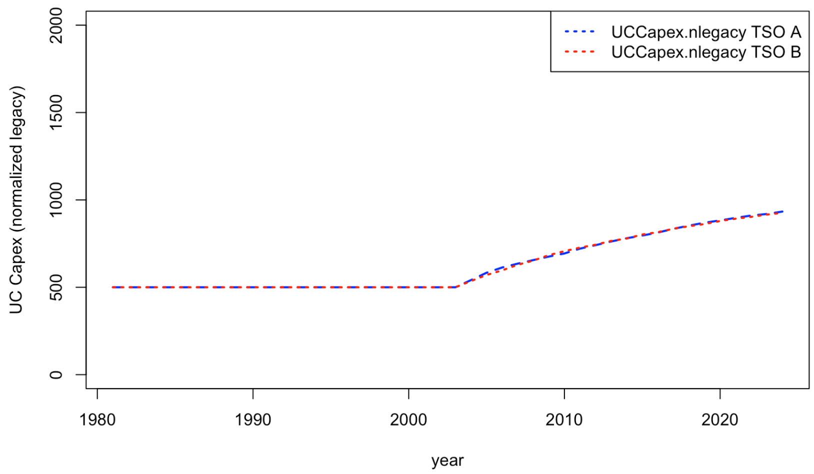


Figure 10-4 Unit cost Capex with neutralized legacy investments 1980-2000, TSOG A and B.

11. Measures for structural comparability

In this Chapter we address the general question of structural comparability between TSOG in terms of operations, asset profile and operating expenditure using the models defined in Chapter 9. The Chapter is illustrated with several examples of standardisation for comparability.

11.1 Sources of heterogeneity

11.01 The objective of an international regulatory assessment of transmission system operations is to obtain robust estimates for the efficient cost of structurally comparable operators. Given fact that the overall number of operators is low and that each is normally confined to a separate national service area without any overlap in assets or exact service conditions, this type of study immediately faces multiple dimensions of heterogeneity: investment tenure and activation, non-standardized task, environmental and topological factors, climate, financial structure, current and past technical regulation, etc. Further, as the reference set is limited and the asset base dimensionality is high, the intrinsic heterogeneity cannot be solved with direct statistical techniques. Hence, a substantial effort in international benchmarking is devoted to ex ante activity analysis and cost allocation using activity-based accounting systems.

11.02 In the following we will address the most important sources of heterogeneity and the measures taken to accommodate them in the benchmark project. We identify the following sources of heterogeneity:

1) Perimeter of activity
2) Prices and currency differences
3) Randomness
4) Environmental heterogeneity

11.2 Scope of activity

11.03 TSOGs have different assets and perform partially different activities, some of which are unregulated or national. For the efficiency analysis, we need to have a homogeneous perimeter of activity.

11.04 We may distinguish ten core activities for transmission system operators:

In-scope AEC

1) **T** Transport; the day-to-day task of operating the energy transport network
2) **M** Grid maintenance: activities linked to the preventive and reactive maintenance.
3) **P** Grid planning; activities occurring prior to the commissioning of the assets
4) **I** Indirect expenses; overhead and support costs

Out-of-scope AEC:

5) **S** System operations: e.g. balancing, capacity management, ancillary services
6) **X** Market facilitation: energy exchanges, auctions and support schemed
7) **TO** Offshore; activities related to offshore transport assets

8) **SF** Storage Facility: activities related to gas storage
9) **L** LNG facility
10) **O** Any other activity; residual, non-regulated activities

11.05 The TSOG in the AEC is essentially a pipeline operator, operating assets for the transport of gas using high pressure equipment. The AEC puts a spotlight on this subset of the transmission tasks, namely the construction and physical operations and maintenance of the grid. Activities that are intrinsically (O,X), technically (L,SF) or empirically (S) heterogenous among the operators are not susceptible to a cost benchmarking study of regulatory quality.

11.06 The allocation of operating costs and staff between the core functions T, M and P may be different depending on internal organisation and legacy. To avoid distortions in the performance assessment, the three functions are assessed together without partial measures when it comes to the cost efficiency.

11.07 For the efficiency analysis, we will follow TCB21 and exclude S, X, TO, SF, and L from the AEC general scope and treat T, M, and P together. Activity I is partially included through standard overhead rates. For simplicity, we use the abbreviation from TCB21 for the standard scope, i.e., TMP (transport, maintenance, planning).

Regional transmission networks

11.08 Some TSOGs are managing regional gas transmission assets and there is a question of including or not these regional activities in the scope of the benchmark. This type of assets and operations (transport, maintenance, planning, service) are in-scope and are compatible with transmission assets and operations. The AEC is designed to take heterogeneity into account, so it is compatible with some TSOGs having regional assets. There are no reasons to exclude those assets. Furthermore, excluding regional assets might make it difficult to establish a criteria differentiating regional and transmission assets. The definitions provided in the EU regulation[21] might not be detailed enough to identify the category to which each asset belongs Separating transmission and distribution would open a grey zone for those who can exclude regional networks. This distinction can be considered where some distinction is applied in national legislation between the two types of services, transmission and regional networks (e.g. FR). In such cases, a specific run at the request of the concerned NRAs can be calculated distinguishing regional and transmission assets. The remaining calculations of the AEC are proposed to be carried out considering all assets, regional and distribution, falling within each TSOG.

11.09 All in-scope assets that are generating transmission revenue are proposed to be included in the AEC. The cost functions control for the heterogeneity of the operations by environmental variables or dimensional weighting. If regional pipes have for instance a smaller diameter, this will be controlled for in the output definition. An NRA may request a sensitivity analysis to establish the efficiency for a subset of its TSOG, e.g., by removing assets and/or operations that are considered of regional character.

[21] Directive (EU) 2024/1788 defines "*transmission*" as "*the transport of natural gas through a network, which mainly contains high-pressure pipelines, other than an upstream pipeline network and other than the part of high-pressure pipelines primarily used in the context of local distribution of natural gas, with a view to its delivery to customers, excluding supply.*"

Example A: comparing asset-heterogenous TSOGs

11.10 *To illustrate the statement above that asset heterogeneity can be managed by the AEC data normalisation, we present a stylized example with two TSOGs.*

11.11 *Consider the TSOGs R and S, where R has a network of regional importance and S is a pure transit operator. R has 1,000 km pipelines whereof 500 km diameter 34'' at HP and 500 km diameter 10'' at MP , respectively. S has 1,000 km pipelines only of diameter 34'' operating at HP.*

11.12 *The difference in profile is evident in the benchmarking by three means:*

11.13 *The different assets could be scaled using the TCB21 NormGrid or another scaling function (see 13.4), each will achieve the purpose of bridging heterogeneity. We present three options mentioned in 10.4: NormGrid, Yepez (2008) and pipeline volume:*

11.14 *The normgrid output for R is 802,8 k€ and for S 1,188 k€. This reflects the fact that the smaller pipelines are less costly in equipment and construction. S is here 1.48 times larger than R.*

11.15 *If Yepez (2008) is used*[22] *rather than the NormGrid, the corresponding values would be: R = 86,853,236 and S = 141,453,264. With this formula (including capex and opex), the relationship S/R is 1.63.*

11.16 *The simplest comparison parameter, but with a very high explanatory power for Totex, is pipeline volume*[23]*. The values are V_R = 470,722.5 and V_S = 868,306.8. The ratio S/R is 1.84.*

11.17 *Thus, depending on which granularity we choose, we can determine that S is at least 48% and at most 84% larger than R.*

11.18 *The pipeline length per pressure level (HP and MP) are {500, 500} vs {1,000, 0} for TSOG R and S. This indicates in e.g. DEA that the two are operating two different technologies.*

11.19 *If other assets than pipelines are differentiated, the model may separate normgrid assets at HP and MP levels, with in this case would be {594, 208,8,} and {1,188, 0}, respectively. This could include compressors, regulators and meters.*

11.20 *The efficiency of R and S is determined by the Totex per output. This means that a TSOG with assets of type regional networks may very well be fully efficient, but also inefficient depending on their investment and operation costs, just as for TSOG with assets with other dimensions. Note that the dimensions of the asset base vary also among "pure" TSOGs , which may have pipelines of 42'' or 25'' depending on needs and location, the method handles their comparison just as well as for regional networks.*

[22] Assuming wall thickness of 0.75 in (HP) and 0.35 in (MP), respectively, in accordance with the minimum wall thickness in Yepez (2008). The formula is given in §13.20 below.

[23] The pipeline volume is $V = \pi L \left(\frac{D-t}{2}\right)^2$ with L = length, D = diameter, t= wall thickness.

11.3 Overhead allocation

11.21 TSOG are organised differently with respect to overhead and management functions, some are separate with internal staff, whereas others are organised in holding companies or groups. The cost for support is therefore heterogenous. A benchmarking would not be reliable if each TSOG could decide individually on how indirect cost should be allocated to in-scope benchmarked cost. Thus, to ensure comparability overhead (indirect) costs are therefore allocated to direct benchmarked cost using a uniform allocation key for all operators.

11.22 Support activities (I) are proposed to be collected and assigned to benchmarked cost (personnel, services, depreciation for assets and vehicles, leasing fees for assets) with an allocation key that is calculated for all based on a defined basis, e.g. share of cost personnel and services, excluding energy, taxes and fees.

11.23 The sensitivity of the allocation key may be explored using sensitivity analyses for the final results globally and per TSOG and year.

11.4 Outsourcing

11.24 TSOG may decide to outsource some of their activities to external providers. This is a classical make or buy decision and outsourcing of some activities can contribute to a higher efficiency. According to EU directives, competitive tenders should be organised at the EU level which does not raise in principle any question regarding the comparability as all TSOGs can outsource in the same EU-wide market.

11.25 There are, however, some systematic restraints that forces some TSOGs to outsource locally. If it is the case, it may result in a lower efficiency, but this is very difficult to validate. TSOGs that are not efficient will claim that they were forced to contract more expensive labour in their own Member States. The high labour costs in some Member States are often used as a waiver to justify inefficiencies.

11.26 As a robustness check, it is possible to consider that a given percentage of labour cannot be outsourced outside of the country, and this part should be corrected for. However, this should be based on an *objective assessment* by the engineers of the costs that cannot be outsourced outside a Member State. In addition, new data may be collected for this purpose as neither the share of external labour in cost of external services nor the share of labour for external services that must be performed in the country of operation can be extracted from IFRS or GAAP public reports.

11.27 We propose that information is collected in AEC on external services with labour contents, their origin and extent, for each OPEX year. Validation will be made of the submitted data to determine whether the services are required to be locally staffed or not.

11.28 A neutral treatment is recommended for internal and external services when concerning services that are considered localized. E.g., facility services and security, but not data cloud services. See section 12.7 for detailed proposals.

11.5 Correction of inflation, prices and currency

Currency (EUR)

11.29 For operators in non-EUR countries, the values in local currency should be converted to EUR using an average exchange rate.

Inflation

11.30 The inflation is different in all countries, even with the same currency, and across baskets of goods and services. There is no specific price index for the natural gas transmission, but we suggest using a reliable and harmonised inflation index such as HICPIGNE (HICP for non-energy industrial goods) for investments. The selection of an inflation index will be subject to a sensitivity analysis. The exact list of indexes available is to be established at the start of the data collection of T0 data.

Price level differences (labour)

11.31 The cost of labour in Europe varies by location, profile and market and operators cannot freely hire workers at the European average or lowest wage level. Therefore, labour costs are heterogeneous and partially uncontrollable. In earlier projects such as TCB21, a labour cost index such as LCI-UTIL (cost index for utilities) published by EUROSTAT, was used for in-scope staff costs.

11.32 Therefore, to ensure the comparability of labour costs, we propose using a correction factor for differences in national salary cost levels. The factor is preferably based on EUROSTAT indexes for labours cost (LCI). The correction applies to all in-house labour costs related to in-scope functions (transport, maintenance and planning) as well as to indirect support service. The same type of correction will be applied to service bought (if reported) that could exhibit similar differences across countries. The exact list of indexes available is proposed to be established at the start of the data collection of T0 data.

11.33 The extended correction for labour cost is discussed in more detail in 11.4.

Fiscal rules and taxes

11.34 TSOGs in Europe operate under different fiscal and accounting regimes. To neutralize the influence of these different rules, we proceed as follow:

1) Exclude all taxes, fees and levies from the benchmarking exercise,
2) Use of a standard interest rate,
3) Use a standard lifetime for each class of assets.

11.6 Environmental heterogeneity

11.35 The external environment for grid construction and operating varies considerably in Europe, in terms of natural, artificial terrain and infrastructure features. These features are exogenous and heterogenous, with different impact on investment and operating expenditure (soil, weather, topography, etc.). Therefore, it is of prime importance to control for these heterogeneous environmental conditions, to maintain the comparability between TSOGs operating in different environmental conditions.

11.36 The factors for correction are based on public databases that are linked to GIS locations, such as CORINE and LUCAS. The list of these factors and their collection are discussed in Section 13.6 below.

11.37 These aspects are assessed by collecting data for the conditions related to the locations of assets for each operator and applying an appropriate correcting factor for the cost.

11.38 Previous benchmarking exercises have already incorporated environmental variables in the model, either as control factors for the efficiency or to normalize inputs and outputs. Based on previous work, we conclude that it is possible to collect cost-relevant data from verifiable sources and to estimate relative complexity factors from independent engineering data for all necessary asset types and operators. By doing so, we will consider explicitly the impact of environmental variables on the cost of the operators. This exercise has been done extensively in TCB21, and the model may use a similar approach and environmental variables. See Sumicsid-CEER (2023).

Randomness

11.39 There are two potential sources of randomness, the first resulting from errors in the reporting of the data, the second from fluctuations of outputs that are out of control of the operator, like weather, fluctuations in prices, the Ukrainian crisis, the energy transition, etc.

11.40 To avoid randomness resulting from errors in the data reporting, there are several steps in the data validation procedure by both the operators and the NRAs. This is designed as to minimize the risk of data errors in the process. The independent auditing of TSOG-reports and a careful cross validation of incoming data ensure that reporting errors do not prevail.

11.41 To avoid the second source of randomness, the choice of outputs is primarily directed to variables that are deterministic or bounded by deterministic variables. We will make the choice of having asset-based outputs to eliminate the risk linked to demand forecasting and fluctuations. The model should be neutral with respect to the actual demand level, capacity investment policy and the utilisation of the assets.

12. Robustness means

In this Chapter we discuss means to assure legal and practical robustness of results for regulatory purposes, interpretation, transparency, resistance to gaming and economic soundness. The Chapter also contains a list of additional adjustments to the benchmarking to obtain extra robustness against potential local differences in operating conditions and costs.

12.1 Facilitating interpretation of efficiency scores

12.01 The interpretation of composite scores such as cost efficiency CE can be complex without insights in modelling. To facilitate this assimilation of the model results, it is recommended that the AEC offers the following information as part of the reporting of results:

1) Individual reports of scores for each operator.
2) Decompositions of CE results into CE(2) and scale (SE).
3) Graphs for unit cost measures (capex and opex) per key assets.
4) Graphs showing cost development over a period (e.g. Staff cost per volume) compared to peers and average operators.
5) Analysis of the features for each operator, highlighting particular aspects (e.g., asset age or outsourcing) that may have an explanatory impact.

12.2 Transparency of process and results.

12.02 To enable transparency over the process and the results it is recommended that the data collected and produced for the AEC is classified in a tier structure (Appendix A) with the following levels:

1) T0 Public parameters and variables derived thereof in the AEC
2) T1 Variables and data, aggregated, inputs and outputs, indicators.
3) T2 Detailed TSOG data for costs, investments, assets and services produced.
4) T3 Sensitive data, subject to national laws for security.

12.03 Without a transparent process for data collection, methodology, calculations and reporting, the process value for NRA will be low. To ensure transparency, it is recommended that the AEC maintains the following principles:

1) Clear established data definitions and data specifications prior to the data collection.
2) Independent audits of financial and asset data for each TSOG
3) Open access to all T1 data (non-commercially sensitive)
4) Full references and access to any non-TSOG reported parameters (T0 data)
5) The methodology, methods, models and tests deployed should be documented to enable reproduction and analysis by external parties.
6) Calculations and reporting of results should be independently audited by third party.
7) The individual reporting should specify each step in the process as to enable each NRA to reproduce the input data.
8) Find ways to protect sensitive data without getting in conflict with the principles of transparency and reproducibility of the analysis

12.3 Comparability of results across time

12.04 Keeping a stable database transforms the repeated AEC over time into a limited exercise of reporting operating cost for the given year and updating the RAB with new and decommissioned assets and investments. ACER should maintain the earlier data in a safe manner and facilitate the integration of new updates to perform the calculations.

12.05 The establishment of a common AEC database is then equivalent to the common use of accounting data to review the economic performance and risk of firms but adapted to regulated infrastructure. This initiative greatly facilitates transparency and process efficiency.

12.06 The data specifications are recommended to be kept stable over time to limit reporting errors, decreasing process and auditing cost and, foremost, enable the calculation of comparisons across time in terms of sector innovation and catch-up.

12.4 Controlling for uncertainty in the AEC data

12.07 As discussed in the data and method sections in Chapter 8, the ES and ED models will primarily include variables that are deterministic by nature, such as output parameters derived from objective and controllable grid features, such as length, volumes, and surface of pipelines and the power and number of compressor stations, etc. However, there could be misreporting or errors in transmission of data introducing uncertainty in the incoming data

12.08 The proposed AEC addresses the potential incoming uncertainty in data by several important design features:

1) Data validation prior to runs, rigorous and performed by TSOG, NRA, and consultants, in several documented rounds with feedback on any detected potential issues.
2) Pre-run outlier detection with follow up to TSOG and NRA for incoming data.
3) Combined use of SFA and DEA in the ES runs, where the complementary nature of SFA vs DEA address potential data errors in the sample. The SFA scores may be presented by confidence intervals to provide NRA with added security.
4) Post-run outlier detection with a round of global validations, including identification of the output or input factors that resulted in the classification.

12.5 Use of the AEC results in NRA decisions

12.09 Regulatory decisions at European and national levels can be challenged in court. The provision to appeal decisions of public authorities is a fundamental right in an open society, the use of which disciplines agencies to be compliant and clear in their application of regulations. It is also intrinsic to the process of regulation that the interest of operators, network users and regulators differ in certain questions. It is important for the AEC that NRA decisions based on AEC are legally implementable and perceived as robust and fair by all stakeholders. These principles are key to ensure the credibility of AEC, see Agrell and Grifell-Tatjé (2016).

12.10 To support the use of the AEC by NRAs[24], the following elements are recommeneded as due diligence without prejudice to any specific issue that may be challenged:

1) *Alignment with EU and national regulatory frameworks.*
2) *Data used in the process should be validated systematically and with equal effort across TSOG data submissions.*
3) *Data processing should be as transparent as possible, and data should be made available for legal review if requested.*
4) *The methods used for the estimation should correspond to best scientific practice and be performed in a well-documented and professional manner*
5) *The operators should have the opportunity to review the data as well as the resulting efficiency targets(Opex and Capex) they are compared against.*
6) *The operators should have the opportunity to introduce additional claims for operator-specific conditions if the model(s) do not address these conditions, and to have them reviewed objectively.*

12.11 In TCB18, conditions (2) and (3) were not fully met as the data validation was handled largely outside the project and the raw data submitted by TSOGs was confidential, in particular the data underlying the output calculations. For TCB21, the data validation was reinforced and documented, but condition (3) was not fulfilled.

12.12 We now provide more detail for the elements above:

Alignment with EU and national regulatory frameworks

12.13 The objectives and criteria for AEC, the cost-efficiency orientation, the attention to the energy transition (see Chapter 6, in particular 6.2), asset heterogeneity (Chapter 10) and the organisation of multiple runs for specific applications and national prerogatives as well as the runs (see Section 8.16) are aligned with the EU and national regulatory frameworks.

12.14 The use of multiple model (ES, ED and UD in section 8.9) facilitate both static analyses for regular revenue-cap settings as well as dynamic analyses of efficiency and utilisation, for regulators that might require this information for their tariff reviews.

Data validation effective and with equal effort

12.15 The data collection and validation process is planned in detail from the specification, templates, verification, submission control, NRA-validation, cross-validation, audit to the post-run validation (See Chapter 11 and in particular section 14.5.).

12.16 The data validation process is organised in detail, at multiple levels with several loops and TSOG, NRA, auditors and consultants collaborating towards a strict data quality guideline. (See section 15.4.).

[24] Here we omit legal requirements that pertain to any operator, authority or agent for penal or civil liability related to respect of contracts, confidential information, absence of corruption and capture. By 'legal robustness' we refer primarily to due diligence criteria that any court or instance of appeal is likely to require as fundamental to accept the produced information as a fundamental part of the decision process in a national regulatory proceeding.

Transparency of data and process

12.17 Transparency is a key concept in AEC that differentiates towards earlier studies by CEER. The three-tier data structure ensures that all essential data for the replication of runs and understanding of the data processing are available, before and after the runs, but after the data validation. Transparency is proposed for all T0 and T1 data, including inputs and outputs for all operators in the AEC. Exceptions are made for T2 data of commercially sensitive nature and T3 data concerning national security. (See section 12.2 and Appendix D).

Economic and scientific soundness of the methods

12.18 The assignment of methods and their combinations, along with the tests applied to model specifications, average cost functions and frontier results, all correspond to established scientific practice as validated by an external group of scientific experts, complemented by results from a literature review, international studies and comparisons of regulatory best practice. (See Chapter 8, specifically section 8.18, and the Expert Review.).

Disclosure of comparison data and peers

12.19 The use of DEA as one of the methods calculating the scores for model ES enables the identification of named and dated peer TSOGs for which data are available in accordance with the earlier evoked transparency policy. Same conditions apply to outliers, detected by the methods listed in section 8.14, that can be reviewed objectively.

Introduction of operator-specific conditions

12.20 AEC is designed to address a multitude of dimensions of heterogeneity (Chapter 10) and to guarantee structural comparability (Chapter 11), including individual environmental conditions based on objective and verifiable data (Section 13.6). In addition, the AEC process provides for a structured collection, processing and implementation of operator-specific conditions in the cost comparison (Section 1)

12.6 Hedging against gaming

12.21 There are several ways for the operator to game the system and the efficiency analysis should be sufficiently robust to prevent strategic behaviour to occur. We identify four areas of concern.

12.22 *Cost reallocation*. Operators have some discretion in their cost allocation. As some costs will be included in the efficiency analysis while other will not, there is a temptation to reallocate costs from categories that are included to those who are not. In such a case, the operator may look more efficient than it is. To prevent this strategic reallocation of expenses, it is needed to carefully monitor the data. To that end, the team will specify clear, detailed and understandable rules for data reporting. Furthermore, the data will be checked and verified by the NRA, limiting the risk of strategic reallocation.

12.23 *Collusion*. There is a risk that operators collude and, for instance, decide to coordinate on their cost reduction effort or on their investment. By colluding they may shift downward the frontier and increase their efficiency. Collusion is particularly a concern when the comparators are few and operate under similar regulation. However, this is not the case in the current benchmarking exercise, and the risk of collusion is considered as low. Given that the AEC likely only informs the

NRA for a part of the incentive regulation, it further seems unlikely that a TSOG would risking an individual loss of revenue for a speculation on the impact on the cost frontier for other TSOGs. In addition, collusion assumes common sharing of the benefits which is complicated for regulated entities such as TSOGs , unless they are under common owners. For the latter category, joint reporting may be required unless the data can be clearly identified as relevant and locally caused.

12.24 *Sensitivity of performance assessment to output choices*. If efficiency measurements are sensitive to output choices i.e. different model specifications lead to (radically) different performance measurements, and to different ranking of operators or different peers, then there is a risk that operators strategically shape their investment to target those who will contribute the most to the measured efficiency, given the output choices of the model (Waidelich, Haug & Wieshammer, 2022). Therefore, it is of prime importance to have output choices that are meaningful from both a technical and an economic point of view, and these choices should be endorsed by operators and regulators. In such a case, the 'manipulation' of the model i.e. different output specifications will be considered as irrelevant.

12.25 *Predictability*. If peers can be predicted in advance, firms can manipulate the frontier to their advantage. Firms may strategically merge to improve their (apparent) efficiency (Agrell and Teusch, 2020) but this risk does not rank high in case of TSOG.

12.7 Extended corrections for local conditions

12.26 The main proposal for AEC is based on CEER projects such as E2GAS, TCB18 and TCB21. The scope and level of detail within these projects can be seen as a point of reference, feasible for repetition without excess work.

12.27 Some participants have requested adjustments of the level details below and we review these options:

Reporting expenditure for services with adjustment of labour cost depending on origin of providers.

1) Pro: outsourced services will be equivalent to inhouse provision in terms of labour cost indexation. E.g., maintenance in Germany would be considered as expensive as inhouse staff in Germany.
2) Con: the outsourced staff may not be from the country of the service provider, several countries could be involved, labour contents in services (IT, repair,..) may be unknown. In addition: no incentives for reducing high-cost services.

12.28 Proposal: extend reporting for labour cost part of external services related to in-scope activities, intensity and origin are requested to obtain correction subject to validation. Materiality criteria should also be established to avoid minor adjustments.

Labour cost adjustments in investments

1) Pro: capitalized investments contain not only physical equipment and material, but also labour services related to the construction, installation and testing of the assets. These services are performed at the TSOG site locations, potentially by service providers using staff from the TSOG country or from some other countries(s). To be neutral towards capitalized investments using in-house staff, the labour cost should be indexed to European average levels.

2) Con: a capitalized investment is competitively tendered and would normally comprise the optimal level and composition of labour, temporarily working with the construction. Often grid construction is made with international labour at costs that are competitive at European level, not national. National cost adjustments may give perverse incentives to TSOG in high-cost countries to opt for more labour-intensive processes from unidentified origins to benefit from high correction factors. However, low-cost countries have no incentives to report labour cost contents as their investment values would be deflated.

12.29 Proposal: extend reporting for labour cost part of external services related capital investments in grid assets and land preparation, intensity and origin are requested to obtain correction subject to validation.

Rehabilitation and upgrades of assets

1) Given the transition objectives and the expected phase-out of natural gas, it is not likely that replacement of worn-out gas transport assets would be fully used until the end of their techno-economic life. The rational response from network operators has been to look at partial rehabilitation and upgrading of assetsto prolong their life and to avoid replacement.
2) Pro: investments in asset-sweating prolongs the life of the assets without new assets in RAB, optimizing the net present value of the asset system given early dismantling.
3) Con: the significant rehabilitations are not always clearly identifiable in the accounting systems, some TSOG use Opex without capitalisation. There is a risk that any maintenance cost could pass for upgrading if the costs are reclassed as investments with a lower annuity.

12.30 Proposal: enhance reporting with significant rehabilitation as option, increasingly relevant as gas networks may not need full renewal. The reporting should be part of the third-party audit material to avoid strategic action.

Detailed environmental correction of specific infrastructure, e.g., intersections.

1) Pro: some TSOG propose to report individually detailed information for corrections of higher costs related to specific infrastructure crossings that are not covered through LUCAS and CORINE data bases for land use cover zones. In principle, this could be extended to any condition for which TSOGs perceive that the granularity is not sufficiently high.
2) Con: ad hoc corrections for certain locations and conditions will only be introduced by TSOG anticipating higher compensation than by the default resolution. However, an average condition also means that an equal number of TSOGs (perhaps the same) also enjoy over-corrections for conditions that are below the average but not detected. The reporting will also break equity in treatment: some TSOG with extra resources may obtain additional complexity factors whereas others would be assigned average factors.

12.31 Proposal: equal conditions for all: the precision of existing GIS databases determines the level of environmental corrections, no ad hoc corrections.

12.8 Criteria for model and method selection

12.32 The selection of a method and a model specification for regulatory efficiency analysis require a different set of criteria to be fulfilled, compared to a purely exploratory or academic study. The model specification depends naturally on the data material collected; it cannot be postulated a priori. In Agrell and Brea-Solis

(2017), the following criteria are formulated and justified for the optimal design of a regulatory frontier analysis framework:

Robustness

12.33 Robustness for a cost model is defined as absence of (statistical) contradictions (model signs and/or significance) when the reference set changes. It also means that the model specification and results should not unduly depend on the presence of specific observations, time periods, technical assumptions or ad hoc constants. The sensitivity of the model results to key parameters should be explored and quantified transparently.

12.34 Means to test:

1) Statistical tests of average cost functions (coefficients, F-test of models)
2) Tests for stability of estimates when reference set is changed.
3) Tests for stability for varying time horizons in dynamic models.
4) Sensitivity tests for technical and economic parameters (interest rates, distributions, etc)

Verifiability

12.35 The model specification should lead to an efficiency model that potentially could be validated with other methods or variations of the specification. This means that the results of a model should be qualitatively aligned with those of other scientifically accepted and used frontier analysis models.

12.36 Means to test:

1) Test for the rank-order correlation of scores from a method using a secondary established method (e.g., SFA vs DEA)
2) Comparisons of earlier and published model specifications to the specific used.

Unambiguity

12.37 The model's definitions must be clear and free of ambiguity to withstand challenges related to conflicting interpretations, e.g., over time and organisational levels. Here, penalizing variables that are inherently stochastic (energy flow) or potentially covering other asset classes (metering stations vs meters).

12.38 Means to test:

1) Cross validation of model variables and their definitions with NRA and TSOG as to determine stability.
2) Critical review of aggregations (sums of assets or costs) in categories (pressure levels, dimensions, etc.) that could penalize TSOGs with different organisations
3) Review of cost allocation methods (in-house vs outsources) to prevent bias related to assumptions on organisational design
4) Analysis of the variance of variables to detect potential stochastic influence.

Relevance

12.39 Output parameters should make techno-economic sense to stakeholders, correlate to one or several output dimensions, and together form a reasonably complete coverage of the service dimensions. The model should be able to treat multiple outputs without resorting to ad hoc scaling or aggregation.

12.40 Means to test:

1) Critical review of model variables and their meaning with NRA and TSOG as to determine relevance.
2) Comparison of the techno-economic interpretation of the model specification to that of other published models
3) Critical analysis of the use of scaling or calibration in the modelling.

Unbiased

12.41 Variables should be defined to enable the model to capture efficiency fairly across operators with varying scale, organisation, and output profiles. Models that are skewed in favor of a subset of operators with a specific asset profile, transport tasks or organisation, should be rejected.

12.42 Means to test:

1) Review of data results and model features that indicate and support different types and modes of operation, if relevant.
2) Statistical analyses (Tobit) of scores towards variables that are not part of the model to detect potential bias towards a group or technology,
3) Cross validation with NRA and TSOG operating under different regimes to ensure that the model allows for full efficiency for all possible profiles.

Precision

12.43 Given the conditions stated above, the selected method should be able to yield a result at an acceptable level of precision and be accompanied by an assessment of its model-based and data-based uncertainty. The model specification should be scientifically and economically sound with interpretations that are fully aligned with the regulatory focus, such as cost efficiency, and subject to relevant decomposition to provide stakeholders with explanations of the sources of differences towards the frontier (as specified elsewhere for the AEC).

12.44 Means to test:

1) Statistical test for the average cost and frontier models, including the ability to explain Totex under the earlier robustness conditions.
2) Complementary analyses using e.g., Robust DEA or confidence intervals for SFA to explore the level of certainty of the scores.
3) Interpretations should be based on solid economic concepts such as comparable and controllable costs and observable assets, not purely statistical predictions.

13. Variables and parameters

In this Chapter we discuss the variables (inputs and outputs) and the parameters (the common factors that are used to construct variables) that will be collected or calculated from TSOG data or public sources. A specific issue in the Chapter is how to construct comparable sums of assets with different dimensions using scaling functions such as the NormGrid.

13.1 Categories of variables

13.01 A comparison of the efficiency of gas transmission system operators using the methods of DEA and SFA requires different types of variables. Figure 13-1 shows the three categories of variables which are required to carry out benchmarking:

1) **Inputs** denoted by X, which represent the controllable resources (e.g. Totex),
2) **Outputs** denoted by Y, which represent the service provision (e.g. transport work, capacity provision or customer service) and
3) **Environmental factors** denoted by Z, which represent structural (environmental) conditions and include proxies for e.g. land use, slope, soil type or climate.

13.02 Variables and parameters are distinguished as follows:

1) Variables: Inputs and outputs used in a model, e.g. Totex or compressor power
2) Parameters: Factors applied to adjust variables if needed, e.g. labour cost index

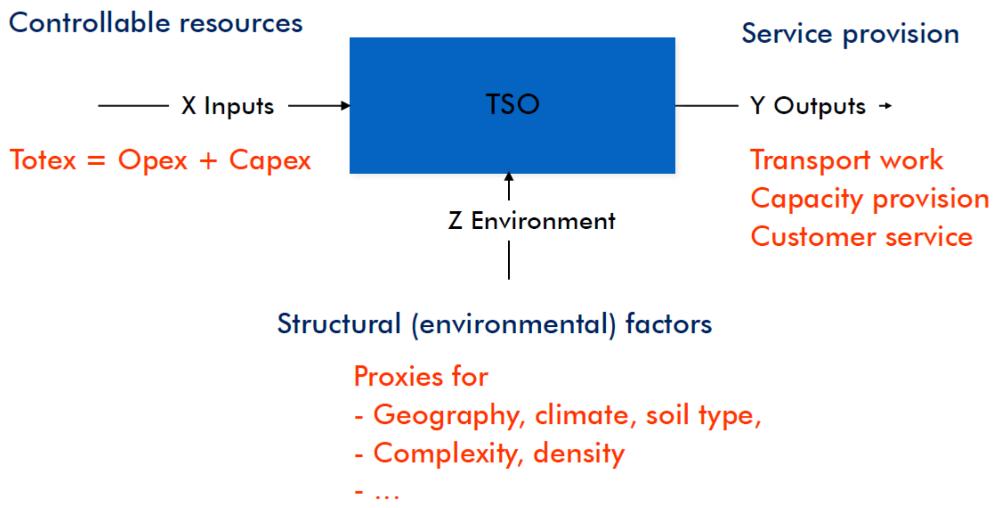


Figure 13-1: Categories of variables

13.03 Figure 13-2 provides further information on these categories of variables, namely their idea, implicit assumptions, illustrating examples, relevant data sources and complementing parameters to correct the variables to ensure comparability. Note that Y variables are further divided into four subcategories, ranging from input- and cost-oriented outputs to true output-oriented, dynamic variables. Output-oriented variables are more exogenous, i.e. less controllable by the TSOGs as opposed to more input-oriented variables.

	X Inputs	→ Y Outputs				Z – Environment
	Costs	**Input-oriented, cost-oriented**	**Input-oriented, asset characteristics**	**Output-oriented, static**	**Output-oriented, dynamic**	**Structural Factors**
		more endogenous/controllable outputs			more exogenous	
Idea	Controllable costs as inputs for benchmarked activities	Normalized costs of asset structure as an output	Asset characteristics provided as an output	Capacity provided of TSO is justified and should serve as a benchmark	Demand of consumers is the competitive benchmark	Reflect unavoidable structural differences
Implicit assumption	(task: derive cost efficiency)	Assets are necessary, cost-weights of assets are applicable	Assets are necessary, no weighting	Fully acknowledge capacity provided even if not needed	Acknowledge requested outputs of consumers only	TSOs may have specific structural challenges
Examples	Totex, Opex, Capex	Normgrid Normgrid Pipelines	Pipeline volume, pipeline length	Connections (also unused ones), Area served, transport capacity	Connections in operation, energy delivered, peakload, transport moment	Subsoil and topsoil characteristics
Sources	TSO financial data, TSO asset data	TSO asset data Engineering based weights	TSO asset data	TSO asset data TSO GIS data	TSO indicators TSO asset data TSO GIS data	Public sources and/or TSO GIS data
Corrections	Currency, inflation, asset age structure, etc.	Own usage share, structural factors	Own usage share			

Figure 13-2: Overview of categories of variables and parameters

13.04 It is noted that the figure intends to illustrate the types of Y variables; it does not imply that all these Y variables will be used in every analysis. The choice of variables for the applied benchmarking model will be taken in AEC once the efficiency model is decided. It is hence not part of PAEC. We recommend however to collect data to be in a position to build parameters for all Y categories, thereby increasing validation possibilities (i.e. validate peakload by information from connection points).

13.05 Using different colours, the figure also highlights the kind of **TSOG data sources** relevant for the project:

1) TSOG financial data: Information on investments and expenses,
2) TSOG asset data: Information on the operated asset structure and properties,
3) TSOG indicators: Information on outputs provided,
4) TSOG locational data: Information on exact locations of assets (pipelines etc.).

13.2 Inputs (X)

Overview

13.06 A shown in Figure 13-1, the aim is to build Totex as sum of Capex and Opex. The idea is to derive Capex and Opex each year based on an **audited financial statement** of a TSOG. IFRS is used as a basis, but there is also the possibility to apply other standards. Figure 13-3 outlines how TOTEX is derived.

13.07 A consistent definition and determination of the benchmarked TOTEX is essential. This includes a definition of which costs are “in scope”. The Investments, expenses/revenues, and personnel costs are therefore differentiated along with the **activities** explained above (in scope, out of scope, in part) with a focus on Gas transport (T), grid maintenance (M), grid planning (P).

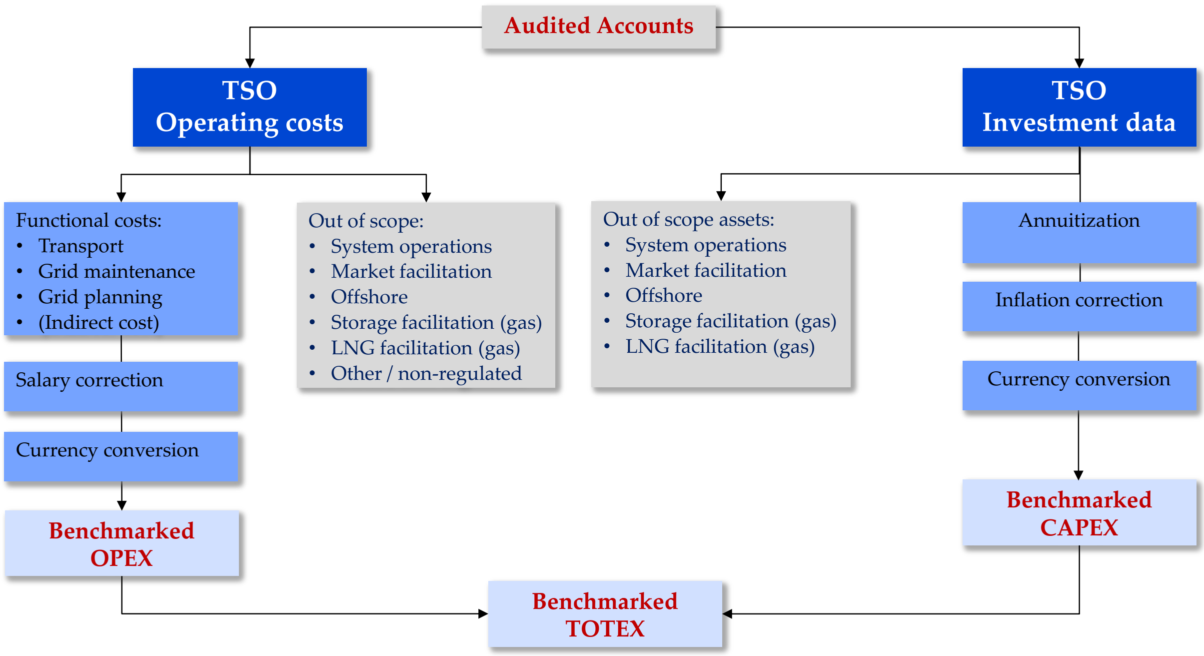


Figure 13-3: Cost standardisation for benchmarked TOTEX

Capex: Investment stream

13.08 The reporting of investments from 1971 onwards on yearly basis is asked but note the new method in AEC for a shorter differentiation of Capex. A net investment stream based on gross investments (no deductions for subsidies or third-party contributions) and disinvestments (for the year an asset is put into operation) must be reported. Capitalized borrowing cost and capitalized land cost are reported separately. The share of investments in H2-ready assets attributed to H2 will be in a separate stream. Investment costs are transformed into **annuities,** applying several factors such as asset lifetimes. This means that book value depreciations according to the financial statement are not used.

Opex: Expenses and revenues

13.09 The total expenses reported for a given year should be equal to the expenses in the audited financial statements of the TSOG for that year. Depreciation, interest, taxes and extraordinary incomes/expenses are excluded from Opex/Capex. For personnel costs, monetary values are further differentiated and complemented by FTE numbers (full staff equivalents) for direct TSOG staff and staff in subsidiaries. As mentioned in section 5.3 of D02, Opex related to H2 assets will be separated.

Cost standardisation

13.10 Figure 13-3 also shows the necessary steps to compute benchmarked TOTEX from the operating costs and the investment data reported by the TSOG. This includes also the topics mentioned above to ensure the comparability of Capex/Opex: investment cycle and asset age, standard lifetimes, WACC / real interest rate, opening balances, taxes, non-grid assets (land), operators with asset separation, scale effects, randomness and environmental heterogeneity.

13.11 Concerning the cost standardisation with price level differences and the possibility of outsourcing (see 12.27), the share of external labour in cost of goods and services

(COGS) is a further parameter already discussed for extended reporting in section 12.7. As discussed, detailed data on external labour is not easy to identify. Notwithstanding, as the attribution of labour cost index (LCI) only to in-house labour creates a bias towards TSOG using outsourced labour for certain services (potentially with a good economic rationale), there is a legitimate interest in finding a pragmatic solution. To address the problem and correct if it is prevalent, AEC is proposed to collect data on labour in external services, amount and origin, as well as type of service.

13.3 Outputs (Y)

Classification along supply tasks and controllability

13.12 Taking EFG4, the German benchmarking of TSOG [25], as an example Figure 13-4 classifies the possible output variables along the three basic **supply tasks** of TSOGs (Transport, Capacity Provision and Service Provision) and **controllability** (output-oriented, less controllable or input-oriented, more controllable). In addition to the variables tested in TCB21 also GIS-based variables tested in Germany in EFG4 are listed (underlined), namely transport moment, transport capacity, area, branches, and meshes. How they have been computed is illustrated in Figure 13-5 further below. Of these, area was selected for the chosen benchmark model together with connections, pipeline volume and compressor power.

Classification of groups of variables based on EFG4	3 basic supply tasks		
	Transport Transport gas as requested	**Capacity Provision** Providing appropriate capacity/pressure	**Service Provision** Connection of consumers to network
Y **Output-oriented; more exogenous; less controllable**	Energy injected/delivered	Peak load	
	Transport moment	Transport capacity[1]	
			Connections
X(Y) **Input-oriented; more endogenous, more controllable by operators**			Pipeline length
		Pipeline volume and surface	
	Compressor energy	**Compressor power**	
		Regulators (nominal flow)	Regulators (quantity)
			Area
			Branches, Meshes

Note:
Variables tested in TCB21
Variables selected in Germany in EFG4
GIS-based variables calculated in Germany

[1] In Germany called "Transportwurzelmoment"

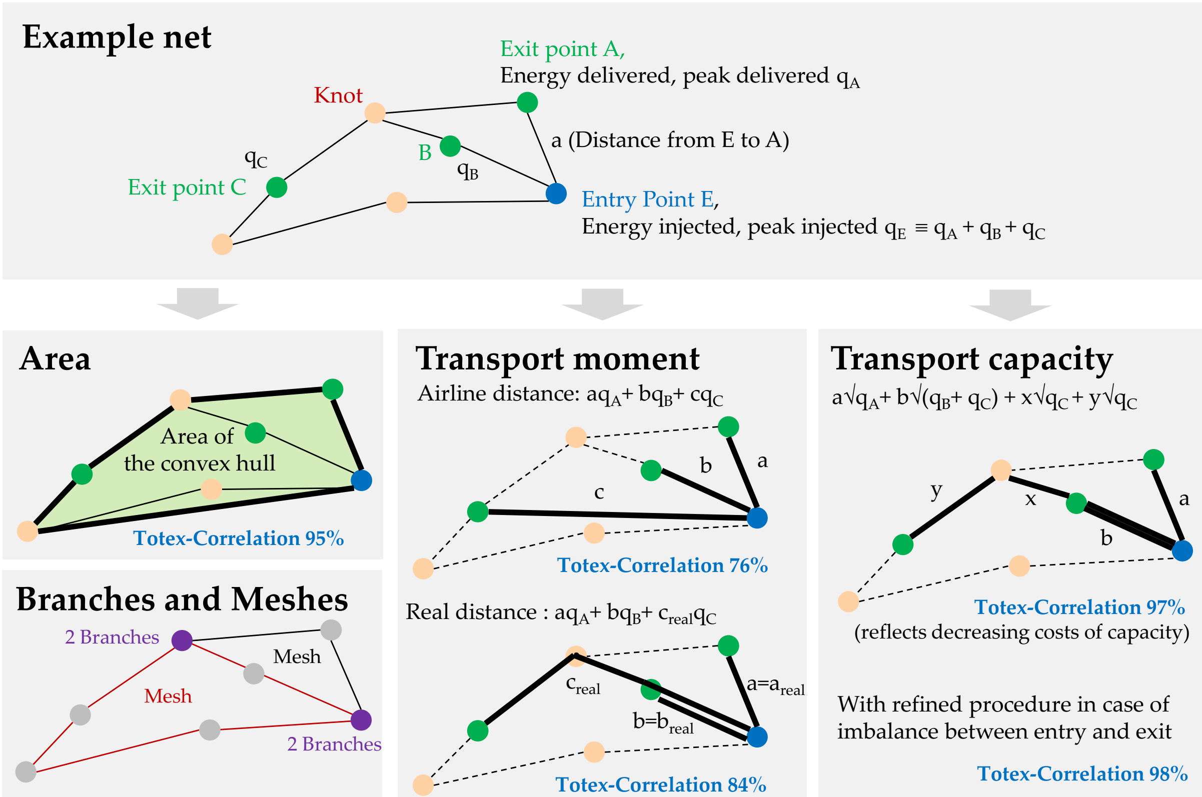
Figure 13-4: Classification of outputs

GIS-based variables

13.13 The GIS-based variables offer a valuable addition, with the advantage that transport moment and transport capacity are output-orientated variables with potentially high explanatory power. Transport moment reflects the transport needs to serve consumers at their exit points. Transport capacity indicates the pipeline capacity needed to serve all exit points. The formal derivation is given in Figure 13-5. As illustrated in the figure, in EFG4 transport capacity had a 98% correlation to Totex.

[25] See Swiss Economics, Sumicsid & IAEW (2023).

To calculate these variables, the exact location and the relevant properties of pipes and nodes, entry and exit points, compressors and regulators are needed.



Note: Indicated Totex-correlations relate to data of EFG4.

Figure 13-5: Illustration of GIS-based variables.

13.14 In earlier benchmarks of CEER, some TSOGs have not been willing to provide the exact location of their assets based on security concerns. For the AEC benchmark it must therefore be decided whether TSOGs should be required to share this information to allow a better selection of possible variables and if so, how the data could be processed to fulfil stricter security requirements. In line, access and confidentiality of data must be decided beforehand.

H2 information

13.15 As set out in Report D02, information related to H2 initiatives, e.g. H2-readyness of pipelines that are still used for gas transport, may become of relevance for a reliable benchmark, requiring specific extensions of the reporting requirements / asset template.

13.4 Scaling functions and NormGrid

13.16 The assets of gas transmissions can be divided into six asset categories:

1) Pipelines
2) Compressors
3) Regulators
4) Metering stations
5) Connection points

6) Control centers

13.17 With each category, there are several data fields. For instance, pipeline data include characteristics like commissioning year, length, diameter, usage share, etc.

13.18 The TSOG have different pipeline systems, including material choices, dimensions, and potentially joint ownership. When creating output variables, a natural question is how to create scalars (sums) that are fair for a comparison. An intuitive solution may be to add pipelines of different dimension by their length. However, this would ignore the fact that 16′′ pipeline must be less onerous in investments than a 56′′ pipeline per km. How many km of 16′′ would then correspond to 1 km of 56′′ pipeline with the same material and installation conditions?

13.19 To answer this question, we need a *scaling function*. Theoretically, a scaling function can have any basis, e.g. weight, volume, capacity, etc. In practice, the relevant basis for our purposes is investment and operating cost, meaning that we look for a *cost function.*

13.20 There is a well-developed literature on cost functions for gas transmission, since Chenery (1949), looking at the optimal combination of two assets: gas pipelines and compressors. A recent development is Yepez (2008), determining that a cost function for a gas pipeline per km (green field) is fully determined by diameter D, thickness **τ** and material choice/steel grade (omitted here). Yepez (2008) shows that the replacement value (investment) is found by the following relation (USD, 2008):

$$C_1(D,\tau) = aD^{\alpha}\tau^{\rho} = 37563.56D^{0.881}\tau^{0.559}$$

where we assume steel as material, the distance is 1 mile, D is the pipeline diameter and **τ** is the wall thickness in inches. Yepez (2008) notes that transport costs are correlated to volume and weight, both given by these variables, as are excavation cost and land preparation. For thickness there is also a minimal bound that depends to the operating pressure and material, both can be derived. Finally, by using an annuitisation of the investment and adding annual operating costs through a statistical study, Yepez presents the following relationship for the total annual expenditure (Totex) for 1 mile of a pipeline as:

$$\begin{aligned} C_D &= (d_L + r)C_1(D,\tau) + C_2(D) = aD^{\alpha}\tau^{\rho} + bD^{\delta} \\ &= 7144.59D^{0.881}\tau^{0.559} + 317.61D^{0.809} \end{aligned}$$

where dL is the depreciation percentage for pipelines, *r* is the cost of capital, the constants a and b are costs coefficients in USD, D is the pipeline diameter and **τ** is the wall thickness in inches.

13.21 The cost function above has a normative value in the very specific situation where we are at a greenfield site in the USA in 2008. In such cases, the normative cost function is a sort of substitute for a benchmarking, one could simply rank investments and TSOGs using the ratio to projected cost. However, this is not the purpose of the AEC. The scaling function here serves instead to answer the question stated above, and this by using a relative system with an arbitrary reference asset.

13.22 Consider the initial investment for the 16′′ and 56′′ pipelines with thicknesses 15 and 25 mm, respectively:

$$CN(\{d,\tau\},\{D,T\}) = \frac{aD^{\alpha}T^{\rho}}{ad^{\alpha}\tau^{\rho}} = \frac{7144.59\ D^{0.881}T^{0.559}}{7144.59\ d^{0.881}\tau^{0.559}}$$

$$= \left(\frac{D}{d}\right)^{0.881} \left(\frac{kT}{k\tau}\right)^{0.881} = \left(\frac{56}{16}\right)^{0.881} \left(\frac{25}{15}\right)^{0.559} = 4.012$$

Where k is the conversion from mm to inch (redundant here). Thus, for cost equivalence, 1 km of 56′′ weighs as much as 4 km of 16′′. A TSOG with 1000 km of 56′′ pipelines therefore has a gross asset value that is not twice as much as that of a 500 km 16′′ pipeline system, but 8 times higher.

13.23 The DNV (2024) Normgrid system is an alternative to e.g. Yepez, developed with European standards EN1594 for TCB21. There are also cost catalogues developed for NRAs that are equivalent to discrete cost functions. In addition, there are cost functions developed for other markets with the same variables, e.g. Massol (2011) or Oliver (2015). As is evident from the demonstration, the calibration constants to a specific currency, value, depreciation and time are irrelevant for the use as a scaling function: they are cancelled for reference asset.

13.24 In practice, the use of the scaling function ("NormGrid") in TCB21 was limited to a residual output for assets that were not covered by physical outputs. Thus, the unit of measurement is not important as the output is not added to any other output.

13.25 In addition to comparing (scaling or normalizing) assets of the same type or category, a scaling function denominated in a common unit (e.g., investment or Totex cost) can also be used to make aggregations of assets of multiple types, e.g. pipelines and compressor stations. The TCB21 NormGrid system (DNV, 2024) is unit consistent, meaning that all scaling functions for gas transmission assets have been denominated in the same currency and year. The same condition applies to other systems as well, e.g. Yepez (2008) in USD summing pipelines and compressors. This allows the construction of *monotonous* cost functions that are essential in regulatory economics for execution-based models, meaning that all relevant assets are included and not only assumed proportional to main cost-driving assets. Notwithstanding, as shown in Yepez (2008), these cost functions may also be used to evaluate capacity expansion choices (increasing pressure through more compressor power vs building parallel pipelines).

13.26 We propose that a scaling function may be applied for a fair and relevant comparison of assets with different dimensions. The DNV(2024) NormGrid is one alternative, AEC could also use the cost functions in Yepez (2008) or others. The scaling function is not an absolute norm, it is relative measure.

13.27 The robustness of the chosen scaling function may be explored through a sensitivity analysis using one or several alternative cost functions.

13.5 Regional networks

13.28 Based on the provisional view above, all assets which are allocated as part of the TSOG revenue (including regional networks) should be part of the comparison. This also means that the variables should include outputs of the corresponding regional networks. As experience shows, various outputs can potentially account for particularities of regional networks. For example, connections, branches or meshes can serve as a proxy of the granularity and complexity of the service task, and the diameter of pipes allows to compute outputs that reflect varying capacities of pipes.

13.6 Environmental / structural factors (Z)

13.29 Environmental factors are conditions that may have an impact on the cost of providing services (Capex and/or Opex). There are two perspectives on how these factors influence cost: the environment is causing higher asset intensity (e.g. building a pipeline around a lake) or the environment is increasing the cost for existing assets of a particular type (corrosion, etc.).

13.30 There are three approaches to how a correction for environmental factors can be implemented:

1) Increase multiplicatively existing outputs (E2GAS, TCB18, TCB21)
2) Use second-stage analysis to scan for environmental effects (EFG4)
3) Add the factor in parallel with other outputs ("Density" in e3GRID, pipeline corridors of specific soil type in EVG4)

13.31 The first approach is flexible since it applies to the relevant outputs or assets only, the factors can therefore be tested and validated by engineers for their prima facie validity as relevant cost-increasing factors. It does not increase the dimensionality of the model (in DEA) but is tested as an additive logarithmic model in e.g SFA or OLS. Single-dimension impacts for pipelines are found in some empirical works, cf. Rui et al. (2011).

13.32 The second method is standard and used in most published work. It is recommended to be performed in AEC as well. However, the single factor focus and the general impact limit the analysis of the conditions, relatively few will be found if the model has a high explanatory power (like the example models in Chapter 8).

13.33 The third approach is problematic since it assumes that all outputs are equally affected by the condition, potentially also with an arbitrarily high impact (in DEA). In e3GRID it was mitigated with weight restrictions, but these ad hoc bounds must be set prior to the analysis, contradicting the flexibility and data-driven nature of DEA.

13.34 In TCB21 the following structural factors were explored:

1) Land use cover (type, density, access),
2) Topography (slope),
3) Subsoil structure,
4) Wetness and humidity,
5) Subsoil water regime,
6) Topsoil structure,
7) Depth to rock,
8) Climate,
9) Population and connection density.

13.35 There are different levels of accuracy of the environmental factors. As shown in Figure 13-6 the precision has steadily increased in past benchmarks, enabling a higher correction power. In TCB18, open sources were used to set correction factors at country level, which did not exactly reflect the environment along the pipelines of TSOGs. In TCB21, the approach was improved by mapping the TSOG assets based on publicly available GIS data on pipelines (OpenInfraMap) to the environmental factors in these corridors. Mapping proprietary GIS-data of TSOs, as it is the case in EFG4 in Germany, would lead to a further increase in precision of the environmental factors, being based on verified exact asset locations.

13.36 We propose to test the environmental factors from TCB21 on GIS data for their power to explain Totex, Capex or Opex by asset and condition, based on an average cost function.

Information requirements

13.37 Correction for environmental conditions is a key feature in explaining residual cost differences after the initial adjustments for structural, economic and technical heterogeneity. The GIS-approach is a powerful and transparent method to compensate asset by asset type for factors that increase Opex, Capex or both.

13.38 Missing GIS data can be replaced (in priority) with zonal (NUTS3) or national (NUTS1) data for locations. In this case, the mapping of environmental conditions with follow the average over larger zones, which generally will increase the complexity factors as pipeline routes normally would following least-cost paths and not cross particularly difficult terrain.

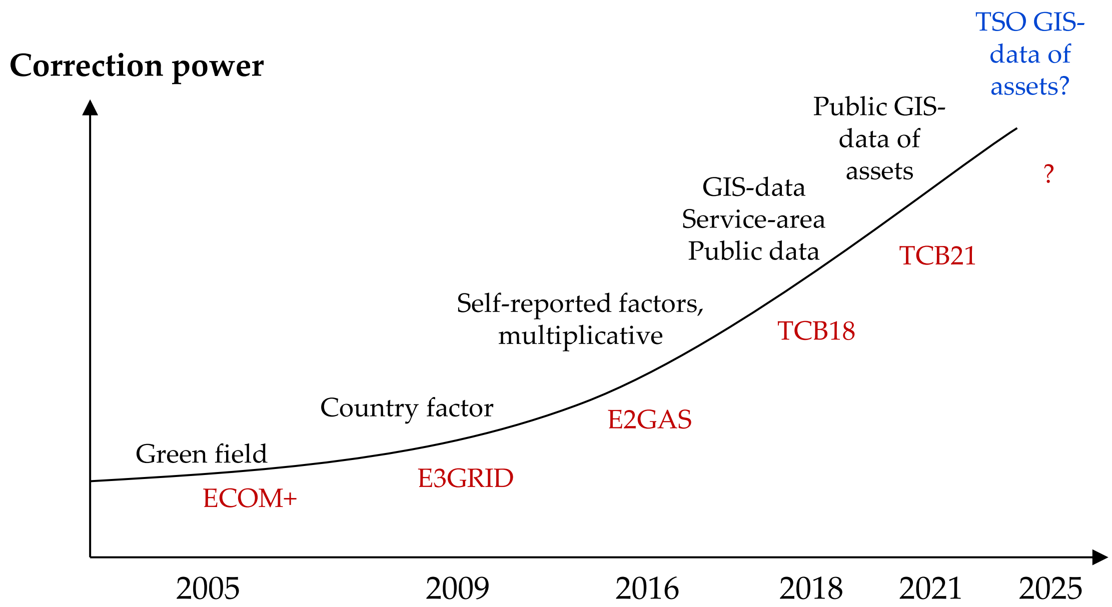


Figure 13-6: Correction power of environmental data

13.39 X, Y, and Z variables and parameters used in previous European gas benchmarks are still relevant, particularly those derived in TCB21.

13.40 To account for investments in H2 capabilities of assets that are still used for gas transport, we recommend expanding data collection in a targeted manner.

13.41 We propose to collect GIS data for TSOG assets to build powerful Y variables and to increase the precision of the environmental factors Z. TSOG delivered GIS-based asset data is proposed to be considered as the highest information tier (T3) given its potential importance for energy supply security. Consequently, it is recommended to ensure storage and processing of the data at ACER or elsewhere under the strictest confidentiality conditions.

13.42 If GIS data cannot be collected for some TSOG, we recommend collection of data at the resolution of at least NUTS3-level.

14. Data request and validation

To compute the necessary variables and parameters explained in Chapter 13 for the models presented in Chapter 8, TSOGs have to provide financial data, asset data, and indicators. This Chapter provides an overview of the proposed data collection and validation process for the AEC.

14.1 Data quality strategy

14.01 Figure 14-1 provides an illustration of the **proposed approach** to achieve an accurate data set for use in an international benchmark. It is the same approach as already applied in TCB21. Starting at the bottom, verified information from TSOGs forms the basis of the data that is requested by structured templates accompanied by clear guides. A process is foreseen for interaction between TSOGs and ACER and their consultants to facilitate the understanding. Once delivered, NRAs ensure the national data validation, followed by a cross validation across the various TSOGs. The final step consists of the modelling phase, where the data is analyzed in detail for use in the benchmark.

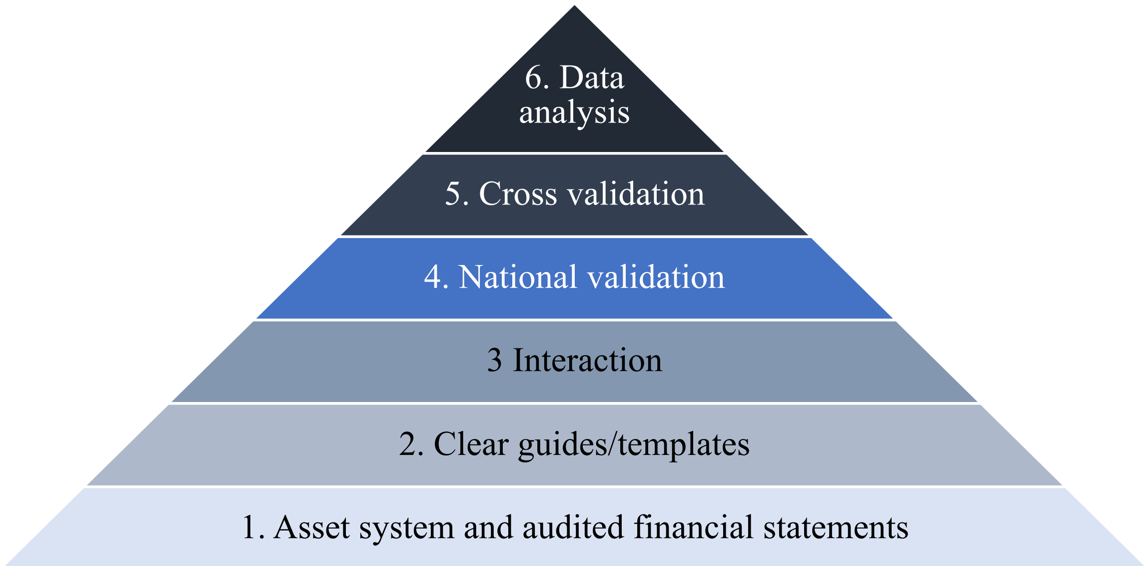


Figure 14-1: Proposed data quality strategy

14.2 Reporting templates

14.02 As a highlighted in Figure 14-2, based on past benchmark experiences, the data will be collected by two templates per participating TSOG: An **asset reporting template** to report assets (which may include GIS locations) as well as indicators, and a **financial reporting template** containing investments and expenses to compute benchmarked Totex.

	X Inputs	→ Y Outputs				Z – Environment
	Costs	**Input-oriented, cost-oriented**	**Input-oriented, asset characteristics**	**Output-oriented, static**	**Output-oriented, dynamic**	**Structural Factors**
		more endogenous/controllable outputs			more exogenous	
Idea	Controllable costs as inputs for benchmarked activities	Normalized costs of asset structure as an output	Asset characteristics provided as an output	Capacity provided of TSO is justified and should serve as a benchmark	Demand of consumers is the competitive benchmark	Reflect unavoidable structural differences
Implicit assumption	(task: derive cost efficiency)	Assets are necessary, cost-weights of assets are applicable	Assets are necessary, no weighting	Fully acknowledge capacity provided even if not needed	Acknowledge requested outputs of consumers only	TSOs may have specific structural challenges
Examples	Totex, Opex, Capex	Normgrid Normgrid Pipelines	Pipeline volume, pipeline length	Connections (also unused ones), Area served, transport capacity	Connections in operation, energy delivered, peakload, transport moment	Subsoil and topsoil characteristics
Sources	TSO financial data, TSO asset data	TSO asset data Engineering based weights	TSO asset data	TSO asset data TSO GIS data	TSO indicators TSO asset data TSO GIS data	Public sources and/or TSO GIS data
Corrections	Currency, inflation, asset age structure, etc.	Own usage share, structural factors	Own usage share			

Financial reporting template: Costs

Asset reporting template: Assets, indicators, (GIS data)

Figure 14-2: Data request needed to compute variables

14.03 In TCB21, the two templates were in Excel format, accompanied by two explaining guides. The instructions are documented in the "**Gas asset reporting guide**" for the asset data and in the "**Financial reporting guide**" for the financial data.

14.3 Asset template

14.04 The content of the asset template will be based on the template for TCB21 to facilitate the delivery for TSOGs that have already participated in past benchmarking exercises.

14.05 Based on the challenges identified in Chapter 5 and experiences with TCB18 and TCB21, it is proposed to extend the asset template by the following information:

- For all assets (pipes, pressure regulator units, compressor units, connection points, metering units):
  - Rehabilitation year (as used in TCB18) to account for lifetime prolongations
  - H2 readiness and year planned for repurposing to H2 to account for tasks related to H2 for assets, that are still in use for gas
  - Investment value, gross,
  - Status (active, decommissioned, repurposing, mothballed)
- For specific assets the following information to get relevant information that describes the installed capacity of assets:
  - Regulator units: G-size (or nominal diameter)
  - Connection points: maximum entry/delivery capacity
  - Metering units: reporting of flow and/or G-size

14.06 Based on the explanations in section 13.6, it is a question whether **GIS-data** should be collected for all assets. The GIS-delivery could be set up similar to EFG4 in Germany by extending the existing asset template from TCB21:

- For any asset but pipes, one location ID is to be given (every location ID belongs to a unique coordinate that is reported in a new sheet)

- For pipes:
  - At least 2 location IDs, one for start point and one for end point
  - Possibility to provide location IDs of supporting points ("Stützpunkte") to reflect the exact route of a pipeline between start and end point.
  - Alternatively, a GIS-file (shape file) can be provided (6 of 16 TSOGs opted for this possibility in Germany).

14.4 Financial template

14.07 The financial template is also based on the template for TCB21, where the basic strategy is to start from audited accounts and incorporate checks for consistency with audited accounts.

14.08 The investment stream and expenses are differentiated along **activities**:

1) In scope of benchmarking are TMP (transport, grid maintenance, grid planning) and in part I (indirect costs)
2) Not in scope are TO (offshore), X (market facilitation), S (system operations), SF (storage), L (LNG) and O (other activities)

14.09 The sum of activities should be consistent to audited financial statements, and it should be taken care of a proper allocation of costs to the activities.

14.10 The **investment stream** is important to calculate annuities (from 1960 onwards for consistency with TCB21 template, needs to be provided from 1971 onwards only). Note the change of method in D02 for a shorter differentiating investment period. Challenges in the reporting of the investment stream are the separation of capitalized borrowing costs and land costs and the rather disaggregated information for gross investments, disinvestments (financial year, acquisition year) and net investment stream (rows: individual years; columns: investment type along activities). Investment subsidies and direct contributions to investments from third parties (without co-ownership) are listed separately by year. The net investment stream is thereby given by the gross investments minus subsidies and disinvestments (acquisition year). The disinvestments financial year are used for consistency purposes with audited accounts. A check with audited financial statements is implemented.

14.11 **Expenses and revenues** are disaggregated along activities and given for the past five years. All activities are covered to enable consistency checks with audited statements. Due to substantial differences, cross-comparisons are more difficult. Since the activities TMP are of special importance, they are more disaggregated than other cost categories. They are also connected with P&L to ensure consistency.

14.12 The amount of data for financial data is demanding but reasonable. **Facilitations** could be achieved by broadening "other activities" O and could include, as an example, market facilitation (X) as well. This is not expected to make much of a difference but would lose some consistency with previous benchmarks, as well as a slight loss of precision in some consistency checks.

14.5 Data verification and validation

14.13 To ensure a high-quality dataset for use in benchmarking, a careful and multi-level data validation is necessary. There is a special focus on extreme values that might be of relevance in setting the frontier in the benchmarking.

14.14 In TCB21 an iterated procedure was used for data verification and validation (see also right side of Figure 14-3), with many iterations for most TSOGs:

1. TSOG own validation
    - Own validation.
    - Use of validation tool[26] if provided by the NRA (NRA received a validation tool that they could provide to TSOGs).
    - Upload of templates to secured platform.
2. NRA validation and approval
    - Validation of the data delivery by the country-specific NRA, consistency with audited accounts.
    - Use of validation tool provided to NRAs.
    - Green light by NRA (marking a template as "NRA approved").
3. Consultant cross validation
    - Team DATA: Various formal, analytical and statistical[27], delivery of issues list to TSOGs with the request to explain or change.
    - Team TECH: Written word report for response by TSOG.
    - Presentation of validation results at workshop.
4. Data analysis: Implicit validation by team ECON
    - OLS / Frontier modelling.
    - An extensive outlier and model validation is carried out to identify conspicuous or influential values.

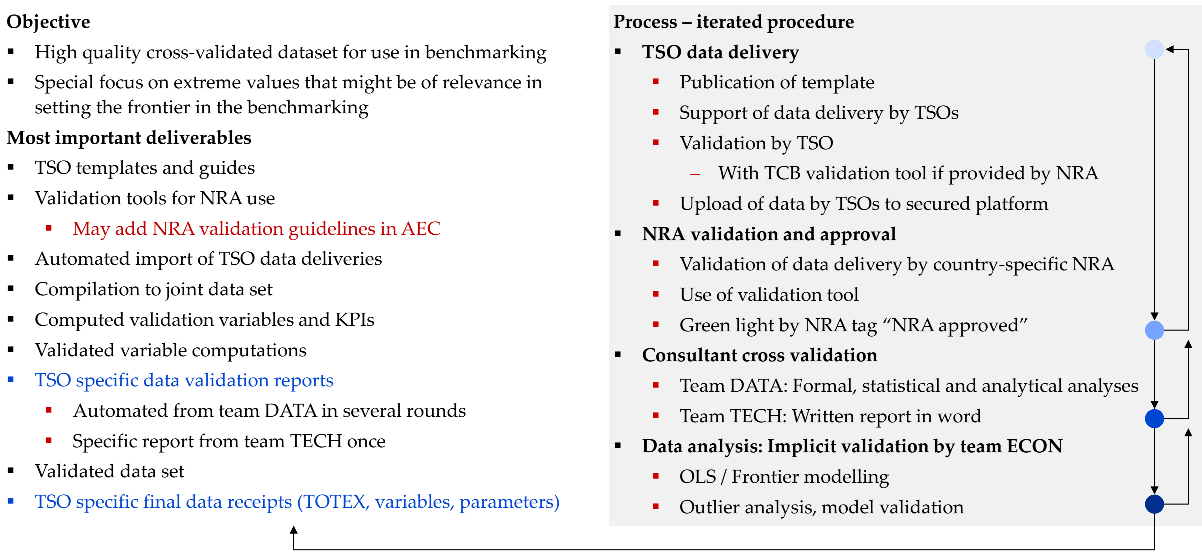


Figure 14-3: Objectives, deliverables and procedure in TCB21

14.15 The experience of TCB21 was that not all NRAs were validating with the same effort. The cross-validation could be complemented with a brief final report that documents the final cross validation steps and the underlying rationale. In addition, a separate audit of TSOG deliveries could be requested. The technical validation may provide

---

[26] The asset validation tool provided a summary of reported assets and sheet integrity, listed several indicators and KPIs including range validation and illustrated how the data is coded for further use. The financial validation tool checked for integrity, provided a summary of reported values and illustrated how the data is coded for further use.

[27] Distribution analyses to identify outliers within a given variable (low and high extremes per variable) included boxplots, IQR-outliers (inter-quartile ranges), Rosner-outliers (>25 obs.), Dixon-outliers (for normally distributed variables). Across variables, analyses included ratios of outputs/outputs, output/inputs and inputs/inputs, scatterplots, outlier analyses such as Cooks-Distance relative to Totex, correlation analysis, and special evaluations with a view to DEA, see Section 8.14 (control of unit cost leaders for each relevant variable/output).

some additional information, however, once established, may not need to be undertaken with every delivery.

14.16 The table below provides a discussion of **typical validation elements per validation step** and concludes on the frequency of their application.

Table 14-1 Description of validation elements.

	National validation by NRA	Audit TSOG data	Statistical cross-validation	Technical cross-validation
Typical Elements	Review of TSOG delivery according to guidelines with the help of validation tool (includes formal checks, consistency checks, value range checks), comparison with audited financial statements, comparison with previous delivery, comparison with previous year, approval	Review TSOG financial data delivery, may compare TSOG asset data with RAB details, check alignment with validation tool	Same checks as national validation, in addition: comparison of data across TSOGs: IRQ analyses, outlier analyses, correlation analyses, UC-analysis; more complex checks with various variables; issues reporting and tracking	Setting of value ranges, (setting of normgrid-weights), review of asset data consistency by engineers, comparison with publicly available data and own experience
Discussion	Important with every data delivery	Increases accountability, may overcome NRA information asymmetries, additional external costs (likely to be borne by TSOGs if not random sample)	Important step to detect data errors and outliers, important information to efficiency modeller	Setting value ranges rather robust over time, may be applied with every review of normgrid weights
Conclusion	**Mandatory** with every delivery	**Optional** (but same across TSOGs)	**Mandatory** for every year of data	**Optional**, every 4 years

14.6 Operator specific conditions and validation

14.17 Each TSOG participating will have the right and opportunity to present evidence for operator-specific conditions that might not be fully covered by the standardizationstandardisation and harmonizationsation measures implemented in AEC to bridge the heterogeneity of TSOG.

14.18 The reporting will take place during Phase II in parallel with the regular data collection and the procedures and data format will be specified by a project note at the beginning of the project.

14.19 *An example of an operator-specific reporting guide can be found in the TCB21 material.*

National validation

14.20 We recommend that each NRA considers the relevance of their TSOG(s) claims, including proposals for the inclusion or exclusion of investment for the run without legacy investment (R3 in Table 8-23). The ruling on such claims under common instructions should be submitted to ACER for decision. Since the default run includes all investments, there is always a viable assessment for overall cost efficiency. As long as consistency is maintained across TSOGs, the specific run R3 can be defined by NRA.

Final implementation

14.21 Following the AEC, upon receiving the results and complementary analyses, each NRA will reflect upon the adjustments that may be necessary to apply the findings in national regulatory instruments. Upon request, the project staff will assist the NRA and TSOG in this phase to assure full understanding of the results and their limits.

14.7 Summary

14.22 The planned data collection must be feasible. In the past it included details of the assets operated, investment data since 1960, and revenue and expenses of the past five years.

14.23 We deem the amount of data collected in TCB21 reasonable. We propose a light extension of the data collection by further information on assets that would allow build or amend variables regarding the expansion of the **hydrogen** infrastructure and repurposing. In addition, selected properties related to the **capacity of assets** should be added. Finally, the **GIS-locations of assets** would allow to build powerful variables with a potentially high Totex-correlation. Besides, the reliability of Z-parameters could be improved. These advantages must be weighed against the possible burden placed on TSOGs, the proposed offline processing system, and the availability of such data across TSOGs.

14.24 Important decisions should be made with respect to data **validation process**, where further adjustments could improve data quality and time compliance. To improve the iterative process, we propose the following measures:

1) Establish transparency rules before any data is provided.
2) Extend the templates as outlined above.
3) Provide validation tool to TSOGs from the beginning.
4) Provide NRAs a validation guide, describing the scope of the national validation step.
5) Add third-party audit of first TSOG data delivery to ensure consistency between audited accounts and data deliveries as well as compliance with the AEC data guides.
6) Reserve resources in project budget for more extensive review of short investment streams.
7) Try to increase accountability of data deliveries, decide upon the following questions: Consequences for incorrect data deliveries of TSOGs? Consequences for late deliveries of TSOGs? Time available for data revision? To strengthen responsibility for data, add final written data approval by TSOG and NRA officials?

14.25 As for other key features of the AEC, ACER will decide upon the data validation requirements prior to the start of Phase II.

15. Process

This Chapter explains in more detail how the data collection, validation, calculation and reporting of the models in AEC, based on the data described in the preceding Chapter, may work in a multi-year process.

15.1 Timing and frequency

15.01 Given the method, parameters and data to be used for the AEC, this chapter proposes a process which could be implemented for the AEC. Figure 15-1 illustrates the proposed process in six steps and distinguishes between the following elements:

1) **Fixed elements** with few or no changes over the iterations, such as the frequency of AEC (every 4 years), the general timelines for every iteration, methods that should be applied at least, or criteria applied to select across various model candidates.
2) **Work performed every 4 years**, examples include the definition of the data templates, the definition of variables and parameters, and search to find an updated frontier that serves as a benchmark.
3) **Work ideally performed every year**, such as data collection, validation, and individual reporting. A lower frequency, e.g. every two or four years is also possible, however, yearly reporting is expected to increase process efficiency and data reliability.

15.02 In the Chapter, we outline a proposed process for the case where full flexibility is required, that is, data is proposed to be collected every year for an annual update of the efficiency development. However, ACER will weigh the benefits of this approach against the costs of the data collection, processing and validation.

15.03 The basic idea is that every four years a frontier model is determined and validated, to be applied yearly based on the most accurate data to enable regulatory application of the AEC with different lengths of regulatory periods as well as different base year across member states.

15.04 For example, a frontier model valid for the years t to $t+3$ may be applied to determine a revenue cap in country A in year $t+1$ and in country B in year $t+2$.

15.05 In the following, we describe the proposed process for each of the six process steps.

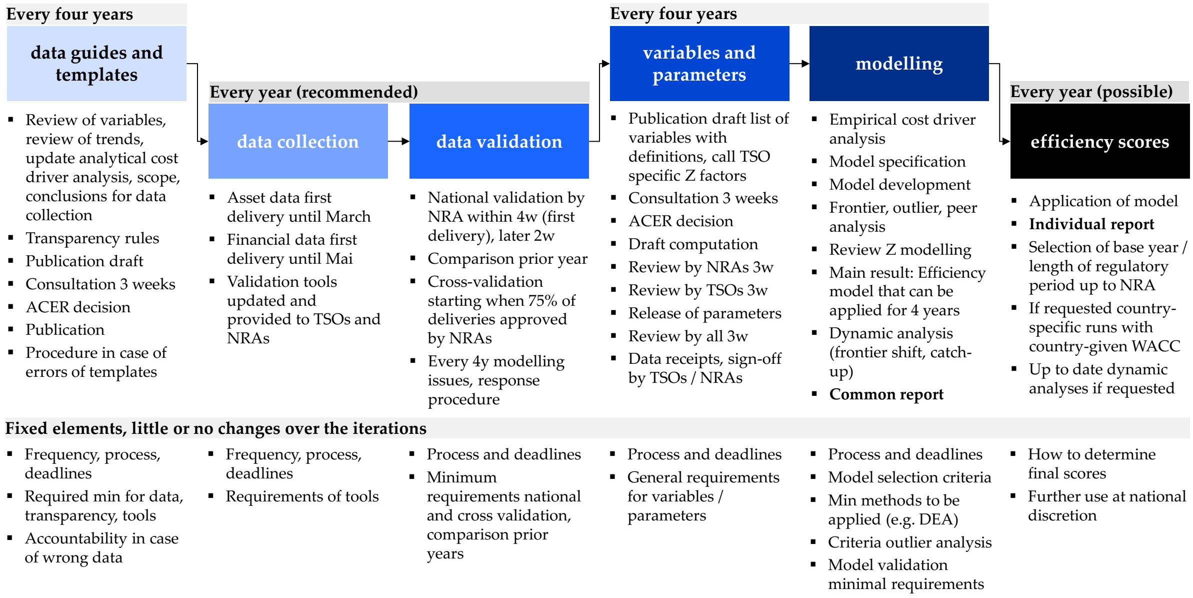


Figure 15-1: Overview of the proposed process for PAEC

15.2 Step 1: Data guides and templates

15.06 Every four years, a review of existing variables and future trends as well as an update of the analytical cost driver analysis is carried out. Based on developments in the gas industry, the scope of benchmarking is defined (e.g. activities TMP) and conclusions are drawn for the data collection (proposals for amended templates and guides). Fixing the data collection for four years reduces process costs for TSOGs and increases data consistency as well as validation possibilities.

15.07 Together with the proposed data guides and templates, corresponding transparency and confidentiality rules are defined. Guides, templates and transparency rules are consulted by ACER for a specified time (e.g. 3 weeks). Considering the responses of the stakeholder, ACER decides on the procedure and final data guides and templates. These are then published.

15.08 It is possible that the amended templates exhibit errors. Therefore, ACER should consider including a procedure to correct such errors. This allows for gradual improvements over time, as both NRAs and TSOGs have more experience in filling in and validating the templates.

15.09 The elements which are maintained over each iteration in this step are the frequency, the basic process and the deadlines. Minimum requirements for data, transparency and tools should also be fixed. In addition, the accountability in case of wrong data deliveries could be defined. Related to the transparency and confidentiality of data, Appendix D provides a note on a possible classification of the data.

15.3 Step 2: Data collection

15.10 The data is collected every year (note disclaimer in §15.02) according to the applicable guidelines and templates of step 1. A yearly process has the advantage that it can be implemented in the regular reporting processes of TSOGs and NRAs with a resulting higher consistency and efficiency. Also, the necessary know-how is more likely to be retained. Moreover, asset extensions can be reported much sooner, with better means to validate by NRAs. Also, photo year effects (strategic shifts of costs from or to other years by a TSOG) can be detected more likely.

15.11 To ensure enough time for the following steps, the deadline for the delivery of the asset data should be early in the year, for instance March. Since the financial data must be based on audited financial statements, it may take longer to have this data available. A first delivery could be required by May.

15.12 Validation tools are updated based on learnings and feedback by TSOGs and NRAs from previous years. This helps in reducing errors and improving data quality from early on.

15.13 The elements which are maintained over each iteration of the comparison in this step are the yearly frequency, the processes for alignment between TSOGs and NRAs and the corresponding deadlines, as well as minimal specifications of the validation tool.

15.4 Step 3: Data validation

15.14 National validation and cross-validation takes place **every year** (note disclaimer in §15.02)

15.15 The **national validation** by NRAs takes place within 4 weeks after receipt of the initial asset and financial data according to validation guidelines and with the help of the validation tools. The validation includes comparisons with the previous year, providing another means to identify anomalies and irregularities in data over time.

15.16 For later iterations, e.g. if errors are detected, TSOGs and NRAs get each a maximum of 2 weeks for resubmission / revalidation. The narrow time windows helps avoiding late deliveries and to ensure that the data across TSOGs is complete enough by August, when cross-validation should take off.

15.17 The national validation is followed by a **cross-validation** at ACER. Experience has shown that there may be longer delays if the cross-validation is delayed until all data has been accepted by all NRAs. It therefore makes sense to start cross-validation as soon as a high proportion of data deliveries (e.g. 75%) have been accepted by NRAs. Anomalies detected in this preliminary cross-validation can already be checked. Once all data deliveries have been accepted, the cross-validation is performed again with the data from all TSOGs.

15.18 The cross-validation results in reported issues and questions that should be addressed by the TSOGs and validated by NRAs. Response times should again be specified in advance to ensure a smooth process. Cross-validation ends with approved templates by both NRAs and ACER.

15.19 In addition, every four years **modelling issues** may arise from step 5. The response procedure is to be set in advance with similar deadlines as for national validation and cross-validation.

15.20 The fixed elements over time (within every 4-year period and across all periods) includes the process, deadlines with the aim of minimizing delays. Fixed are also the minimum requirements for national and cross validation to ensure a reliable data basis and consistence over multiple years. The requirements include guidance for the comparison with prior years.

15.21 The financial data auditing is proposed every year (i.e. for each reporting). A less appealing alternative could be to require a full financial audit of the four years reporting at the end of each cycle, as otherwise the question could arise what happens to the benchmarking if errors are discovered in the subsequent audit of peer companies.

15.5 Step 4: Variables and parameters

15.22 Every four years, based on the data guides and the templates in step 1, variables and parameters are specified. The following stages are carried out for this purpose:

1) A draft list of all variables with definitions is published, including a call for TSOG specific Z factors.
2) During a consultation period of 3 weeks, NRAs and TSOGs can submit comments on the variables and parameters published in the draft.
3) ACER takes these comments into account and decides on a final list of all variables including Z factors
4) Based on the final list, a draft calculation of variables and parameters will be prepared and communicated to the NRAs
5) Within three weeks the NRAs review the draft of the variable calculation, afterwards the same for TSOGs
6) Once all responses are processed, release of all variables and parameters, final review by NRAs and TSOGs within three weeks
7) After successful review, data receipts of the final data are prepared and delivered to NRAs/TSOGs for sign-off and formal approval by NRA/TSOG representatives.

15.23 As fixed elements over time, again the process and the deadlines should be fixed for each iteration. Deadlines are important to get the necessary inputs to calculate yearly efficiency scores in time. In addition, general requirements for variables / parameters should be fixed over time.

15.6 Step 5: Modelling

15.24 The modelling takes place every four years with the following elements being revised:

1) An **empirical cost driver analysis** to determine the correlation between the variables computed in the previous step. Each new iteration allows to analyse the cost impact of new variables as well as changing correlations of previously used variables in the last four years
2) **Model development** takes place considering the results of the empirical cost driver analysis and models of previous iterations (see e.g. TCB21 Model Specification GAS)
3) **Model specification for** deterministic and stochastic (utilisation) models.
4) A **frontier** is modelled and a corresponding outlier and peer analysis conducted
5) **The modelling of the environmental factors** is reviewed to ensure that heterogeneity due to environmental influences and related changes are sufficiently considered
6) The main result is **a frontier model** that can be applied for 4 years in DEA and SFA.
7) A model for a **dynamic analysis** is selected and a corresponding analysis, which can separate between a frontier shift and a catch-up is carried out. This provides results about increase of the total factor productivity in the last 4 years as well as

the catch up to best practice of inefficient TSOGs. The time frame of the 4-year period reduces the influence of short-term fluctuations and can better reflect the general trend.

8) A **common report** including the previous working steps, the results of the empirical cost driver analysis, the model specification and development, the estimated frontier including an outlier and a peer analysis, as well as the final efficiency scores is published. The exact contents of the reporting (common and potentially individual) should be decided in Phase II for implementation in Phase III.

15.25 For the modelling step, certain elements are maintained for each iteration of the comparison:

1) **Process and deadlines** is consistent for each iteration. Deadlines are important to have the necessary inputs to calculate yearly efficiency scores and to avoid a delay of the entire 4-year frequency.
2) **Model selection criteria** (e.g. robustness, relevance, unbiasedness) stays consistent as these requirements do not change over the iterations and are important to obtain suitable and robust models.
3) A **minimum requirement of methods** to be applied stays unchanged. This could for instance include the Data Envelopment Analysis (DEA). Further methods could still be added in an iteration.
4) **The criteria of the outlier analysis** remains unchanged over time. The chosen outlier criteria should be selected with precaution and be able to select identify conspicuous or influential values independent of the selected model specification. Minor adjustments are possible in the event of scientific progress.
5) **Minimal Requirements for the model validation** are fixed to ensure comparability of the models between different iterations and robustness of the models within the iterations.

15.26 Academic and regulatory practice in efficiency measurements show that:

1) Some methods require larger datasets to adequately estimate all coefficients.
2) Certain model specifications (deterministic, linear with low or no noise) converge poorly in parametric methods.
3) Random variables are inappropriate for use in deterministic methods.
4) Random errors in certain variables across all TSOG cannot be detected through outlier detection.

15.27 Thus,

1) Data validation should be closed prior to model specification in the modelling
2) Model specification(s) across the two methods before calculating any scores
3) AEC should cross validate all results across methods, but also over time.

15.7 Step 6: Efficiency calculations

15.28 Based on the model selected in step 5, the model for the corresponding 4-year period can be applied in each year with little effort, allowing the determination of efficiency scores.

15.29 It is proposed to provide each TSOG and NRA with an individual report including an individual efficiency score. The content of the individual report is decided in collaboration with NRA and TSOG at the outset of the project. Parts that are kept confidential are decided by ACER at the outset of phase III.

15.30 The selection of the base year and the length of the regulatory period in which the calculated efficiency scores are applied is up to the NRA. If it is requested by an NRA, a country-specific run (R5 in Table 8-23) with a country-given WACC can be carried out. Also, up to date dynamic analyses, based on the model determined in step 4 could be used by request of an NRA.

15.31 As fixed elements over time, the calculation of efficiency scores should be fixed and, if possible, harmonised.

15.8 Summary

15.32 A transparent and well-documented process is of crucial importance to a successful benchmarking.

15.33 ACER is strongly recommended to aim at fixing elements that do not change over time, such as frequency of data collection and modelling, as well as selected aspects of the methods and criteria that should be applied. The same relates to topics that are decided every four years and yearly.

15.34 A regular, yearly repetition of the data process induces learning effects and reduces errors and collection efforts. NRAs and TSOGs get a valid information for their tariff reviews irrespective of where in the cycle or length of period they are. Frequent feedback also reinforces the information value of the exercise. Finally, photo year effects can be detected more accurately.

References

ACER (2018) *Methodologies for target revenues of gas TSOs*, ACER Report.

ACER (2021) *Transporting Pure Hydrogen by Repurposing Existing Gas Infrastructure: Overview of existing studies and reflections on the conditions for repurposing*. ACER Report.

ACER (2021b) *Position Paper on the Key Regulatory Requirements to Achieve Gas Decarbonisation*, ACER-CEER Report. https://www.ceer.eu/wp-content/uploads/2024/04/ACER-CEER-Position-paper-on-gas-decarbonisation_final.pdf

Afriat, S (1972), Efficiency estimation of production functions, *International Economic Review* 13, 568–598.

Afsharian, M., & Bogetoft, P. (2025). Outliers and Rewards in Regulation: Insights from German Electricity Benchmarking. *Omega*, 103426.

Afsharian, M., Ahn, H., & Kamali, S. (2022). Performance analytics in incentive regulation: A literature review of DEA publications. *Decision Analytics Journal*, 4, 100079.

Agrell, P J (2015) Incentive Regulation of Networks: Concepts, Definitions and Models., *Reflets et Perspectives de la Vie Economique,* **LIV**(1), 103-132.

Agrell, P J (2017) *Twenty years of frontier analysis in the service of regulatory economics: perspectives and open questions*, Invited keynote lecture, EWEPA 15, Loughborough University, UK, June 12-15, 2017.

Agrell, P. J. and P. Bogetoft (2004) *Evolutionary Regulation: From CPI-X towards contestability,* ENCORE position paper, Amsterdam. https://dial.uclouvain.be/pr/boreal/object/boreal:18397/datastream/PDF_01/view

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.

Agrell, P. J. and P. Bogetoft (2009) *International Benchmarking of Electricity Transmission System Operators.* Final report for project e3GRID 2009. 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., & 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., & Grifell-Tatjé, E. (2016). A dynamic model for firm-response to non-credible incentive regulation regimes. *Energy Policy*, 90, 287-299.

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

Agrell, P. J., & Teusch, J. (2020). Predictability and strategic behavior under frontier regulation. *Energy Policy*, 137, 111140.

Agrell, P. J., Dehaybe, H., & Herrera, M. (2025). Balancing supply security and decarbonisation: Optimizing Germany's LNG port infrastructure under the European Green Deal. *Energy Policy*, 198,

Agrell, P. J., P. Bogetoft, C. Riechmann, A. Rodgarkia-Dara, C. Zimmer (2013) *e3GRID 2012 – European TSOG Benchmarking Study.* Final report for project e3GRID 2012.

Agrell, P. J., P. Bogetoft, Trinkner, U. (2016). *Benchmarking European Gas Transmission System Operators*. Final report for project e2GAS.

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.

Ahn, H., Clermont, M., & Langner, J. (2023). Comparative performance analysis of frontier-based efficiency measurement methods–A Monte Carlo simulation. *European Journal of Operational Research*, 307(1), 294-312.

Aigner, D., Lovell, C. A. K., and Schmidt, P. July 1977. Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics* 6(1):21–37.

Ajayi, V., & Pollitt, M. G. (2022). *Changing times: Incentive regulation, corporate reorganisations, and productivity in the Great Britain's gas networks*. ERPG Working Paper 2214, University of Cambridge.

Alaeifar, M., Farsi, M., & Filippini, M. (2014). Scale economies and optimal size in the Swiss gas distribution sector. *Energy Policy*, 65, 86-93.

Amaral, A. L., Martins, R., & Dias, L. C. (2022). Efficiency benchmarking of wastewater service providers: An analysis based on the Portuguese case. *Journal of Environmental Management,* 321, 115914.

Amirteimoori, A., Despotis, D. K., & Kordrostami, S. (2014). Variables reduction in data envelopment analysis. *Optimisation*, 63(5), 735-745.

Andor M, Parmeter CF, Sommer S. (2019) Combining uncertainty with uncertainty to get certainty? Efficiency analysis for regulation purposes*. European Journal of Operational Research* 274(1):240–52.

Andor, M., & Hesse, F. (2014). The StoNED age: the departure into a new era of efficiency analysis? A monte carlo comparison of StoNED and the "oldies"(SFA and DEA). *Journal of Productivity Analysis*, *41*(1), 85-109.

Athawale, R., & Felder, F. A. (2022). Electric utility death spiral: Revisited in the context of tariff design. *The Electricity Journal,* 35(1),

Badunenko, O. and Kumbhakar, S.C. (2016). When, Where and How to Estimate Persistent and Transient Efficiency in Stochastic Frontier Panel Data Models, *European Journal of Operational Research*, 255(1), 272–287.

Badunenko, O. and Mozharovskyi, P. (2016), Nonparametric Frontier Analysis using Stata, *Stata Journal*, 163, 550--89,

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

Battese, G. E., & Coelli, T. J. (1992). Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India. *Journal of Productivity Analysis*, 3, 153-169.

Bauer, P. W., Berger, A. N., Ferrier, G. D., & Humphrey, D. B. (1998). Consistency conditions for regulatory analysis of financial institutions: a comparison of frontier efficiency methods. *Journal of Economics and Business*, 50(2), 85-114.

BIE (Bureau of Industry Economics). (1994e). International Performance Indicators: Gas Supply. Research, Report 62, AGPS, Canberra, December.

Biggar, D. (2025). The role of cost benchmarking in public utility regulation. *Journal of Regulatory Economics*, 1-33.

Bjurek, H., Førsund, F. R., & Hjalmarsson, L. (1998). Malmquist productivity indexes: an empirical comparison. In *Index numbers: Essays in honour of Sten Malmquist* (pp. 217-239). Dordrecht: Springer Netherlands.

Bogetoft, P., & Otto, L. (2010). *Benchmarking with DEA, SFA, and R*. Springer Science & Business Media.

Boussofiane, A., Martin, S., & Parker, D. (1997). The impact on technical efficiency of the UK privatisation programme. *Applied Economics*, 29(3), 297-310.

Bretz, F., Hothorn, T. and Westfall, P. (2010), *Multiple Comparisons Using R*, CRC Press, Boca Raton.

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

Callen, J. L. (1978). Production, efficiency, and welfare in the natural gas transmission industry. *The American Economic Review*, 68(3), 311-323.

Campos, M. S., Costa, M. A., Gontijo, T. S., & Lopes-Ahn, A. L. (2022). Robust stochastic frontier analysis applied to the Brazilian electricity distribution benchmarking method. *Decision Analytics Journal*, 3, 100051.

Capece, G., Costa, R., & Di Pillo, F. (2021). Benchmarking the efficiency of natural gas distribution utilities in Italy considering size, ownership, and maturity. *Utilities Policy*, 72, 101277.

Carbon Limits and DNV (2021) *Re-Stream — Study on the reuse of oil and gas infrastructure for hydrogen and CCS in Europe*. Final Report. https://www.carbonlimits.no/wp-content/uploads/2021/10/Re-stream-report-October-2021.pdf

Casarin, A. A. (2014). Productivity throughout regulatory cycles in gas utilities. *Journal of Regulatory Economics*, 45(2), 115-137.

CEER (2023) *Report on regulatory frameworks for European energy networks 2022.* CEER Report C22-IRB-61-03.

CEER (2024) *Report on Regulatory Frameworks for European Energy Networks 2023*, Report C23-IRB-70-03.

Cerniauskas, S., Junco, A. J. C., Grube, T., Robinius, M. & Stolten, D. (2020). Options of natural gas pipeline reassignment for hydrogen: Cost assessment for a Germany case study. *International Journal of Hydrogen Energy*, 45(21), 12095-12107.

Cervigni, G., Conti, I., Glachant, J. M., Tesio, E., & Volpato, F. (2019). *Towards an efficient and sustainable tariff methodology for the European gas transmission network*. EUI/Florence School of Regulation, Technical Report, https://doi:10.2870/013490

Charnes, A., Cooper, C. A., Lewin, A., and Seiford, L. (1994). *Data Envelopment Analysis: Theory, Methodology and Applications*. Boston: Kluwer Academic Publishers.

Charnes, A., Cooper, W. W., and Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research* 2(6):429–444.

Čížek, P., Härdle, W., Weron, R. (2011). *Statistical tools for finance and insurance*. Berlin: Springer.

Coelli, T. J., Rao, D. S. P., O'Donnell, C. J., & Battese, G. E. (2005). *An introduction to efficiency and productivity analysis*. Springer.

Costello, K. and R. Hemphill (2014) Electric utilities' 'death spiral': Hyperbole or reality? *The Electricity Journal*, 27(10), 7-26.

Diewert, W. E. (2002). *Harmonised Indexes of Consumer Prices: Their Conceptual Foundations*, ECB Working Paper 130. http://dx.doi.org/10.2139/ssrn.357342

DNV (2018). *Future Role of Gas from a Regulatory Perspective*. Report for CEER.

DNV (2022). *Future Regulatory Decisions on Natural Gas Networks: Repurposing, Decommissioning and Reinvestments*. Report for ACER.

DNV and CEER (2024). *NormGrid Weights*. Report for CEER-TCB21.

Doyle, J., & Green, R. (1994). Efficiency and cross-efficiency in DEA: Derivations, meanings and uses. *Journal of the operational research society*, 45(5), 567-578.

Enerdata (2024). *Energy system infrastructures and investments in hydrogen*. Study requested by the ITRE committee.

ENTSOG (2020). *Ten-year Network Development Plan*. Executive Summary.

ENTSOG and ENTSOE (2024). *TYNDP 2024 Draft Scenarios*

Erbetta, F., & Rappuoli, L. (2008). Optimal scale in the Italian gas distribution industry using data envelopment analysis. *Omega*, 36(2), 325-336.

Farsi, M., Filippini, M., & Kuenzle, M. (2007). Cost efficiency in the Swiss gas distribution sector. *Energy Economics*, 29(1), 64-78.

Federal Energy Regulatory Commission (1984). *Order No. 380: Elimination of Minimum Bill Provisions Fromfrom Certain Natural Gas Pipeline Rate Schedules* (48 Fed. Reg. 52491, June 28, 1984). Washington, DC: U.S. Government.

Federal Energy Regulatory Commission (1985). *Order No. 436: Regulation of Natural Gas Pipelines After Partial Wellhead Decontrol* (50 Fed. Reg. 42408, October 18, 1985). Washington, DC: U.S. Government.

Felder, F. A., & Athawale, R. (2014). The life and death of the utility death spiral. *The Electricity Journal*, 27(6), 9-16.

Filippini, M. and Greene, W.H. (2016). Persistent and transient productive inefficiency: A maximum simulated likelihood approach. *Journal of Productivity Analysis*, 45 (2), 187–196.

Frontier Economics (2018). *Effizienzvergleich der Verteilernetzbetreiber Gas (4.Regulierungsperiode). Bericht im Auftrag der Bundesnetzagentur* [Gutachten EVG4]. (In German).

Goncharuk, A. G. (2008). Performance benchmarking in gas distribution industry. *Benchmarking: An International Journal,* 15(5), 548-559

Gordon, D. V., Gunsch, K., & Pawluk, C. V. (2003). A natural monopoly in natural gas transmission. *Energy Economics*, 25(5), 473-485.

Granderson G. and C. Linvill (1996) The Impact of Regulation on Productivity Growth: An Application to the Transmission Sector of the Interstate Natural Gas Industry, *Journal of Regulation Economics*, 10, pp. 291-306.

Granderson, G., & Linvill, C. (1999). Parametric and nonparametric approaches to benchmarking the regulated firm. *Journal of Productivity Analysis*, 12(3), 211-231.

Greenblatt, J. (2015). *Opportunities for efficiency improvements in the US natural gas transmission, storage and distribution system*, Report LBNL-6990E, ERNEST ORLANDO LAWRENCE BERKELEY NATIONAL LABORATORY,

Grifell-Tatje E., Lovell, C.A.K. (2003). The managers versus the consultants. The *Scandinavian Journal of Economics* 105 (1), 119–138.

Gugler, K., & Liebensteiner, M. (2019). Productivity growth and incentive regulation in Austria's gas distribution. *Energy Policy*, 134, 110952.

Haney, A. B., & Pollitt, M. G. (2011). Exploring the determinants of "best practice" benchmarking in electricity network regulation. *Energy Policy*, 39(12), 7739-7746.

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.

Haney, A.B., Pollitt, M.G. (2009). Efficiency analysis of energy networks: an international survey of regulators. *Energy Policy* 37 (12), 5814–5830.

Hawdon D. (2003) Efficiency, performance and regulation of the international gas industry – a bootstrap DEA approach, *Energy Policy*, 31, pp. 1167-1178.

Heesche, E., & Asmild, M. (2022). Incorporating quality in economic regulatory benchmarking. *Omega*, 110, 102630.

Hoff, A. (2007). Second stage DEA: Comparison of approaches for modelling the DEA score. *European journal of operational research*, 181(1), 425-435.

Hollas, D. R., Macleod, K. R., & Stansell, S. R. (2002). A data envelopment analysis of gas utilities' efficiency. *Journal of Economics and Finance*, 26(2), 123-137.

Hollas, D., & Stansell, S. (1994). The Economic Efficiency of Public vs. Private Gas Distribution Utilities. *Annals of Public & Cooperative Economics*, 65(2).

Hou, Z., Ramalho, J. J., & Roseta-Palma, C. (2025). Dealing with endogeneity in stochastic frontier models: A comparative assessment of estimators. *Energy Economics*, 108922.

Jamasb T, Pollitt M. (2003) International benchmarking and regulation: an application to European electricity distribution utilities. *Energy Policy*, 31(15):1609–22.

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

Karagiannis, G. (2012). More on the Fox paradox. *Economics Letters,* 116(3), 333-334.

Kasiri, M., & Mirnezami, S. R. (2023). How can the compensation structure of Iran's natural gas distribution services be modified based on incentive-based regulations?. *Energy*, 285, 129457.

Kasiri, M., & Mirnezami, S. R. (2023). How can the compensation structure of Iran's natural gas distribution services be modified based on incentive-based regulations?. *Energy*, 285, 129457.

Kim, T. Y., & Lee, J. D. (1994). Cost analysis of gas distribution industry with spatial variables. *J. Energy & Dev.*, 20, 247.

Kim, T. Y., Lee, J. D., Park, Y. H., & Kim, B. (1999). International comparisons of productivity and its determinants in the natural gas industry. *Energy Economics*, 21(3), 273-293.

Kim, T.-Y., J.-D. Lee, Y. H. Park and B. Kim (1999) Industrial comparisons of

Kumbhakar, S. (1990), Production frontiers, panel data, and time-varying technical inefficiency, *Journal of Econometrics* 46, 201–212.

Kumbhakar, S. and Lovell, C. (2000), *Stochastic Frontier Analysis,* Cambridge University Press, Cambridge.

Kuosmanen, T. (2012). Stochastic semi-nonparametric frontier estimation of electricity distribution networks: Application of the StoNED method in the Finnish regulatory model. *Energy Economics*, 34(6), 2189-2199.

Kuosmanen, T., & Kortelainen, M. (2012). Stochastic non-smooth envelopment of data: semi-parametric frontier estimation subject to shape constraints. *Journal of Productivity Analysis*, 38(1), 11-28.

Kuosmanen, T., Saastamoinen, A., & Sipiläinen, T. (2013). What is the best practice for benchmark regulation of electricity distribution? Comparison of DEA, SFA and StoNED methods. *Energy Policy*, 61, 740-750.

Lawrence, D., Fallon, J., Cunningham, M., Zelenyuk, V., & Hirschberg, J. (2017). *Topics in efficiency benchmarking of energy networks: Choosing the model and explaining the results*. Report for ACM, Economic Insights Ltd.

Lawrence, D., Kain, J., & Coelli, T. (2011). *Regulation of Suppliers of Gas Pipeline Services–Gas Sector Productivity.* Report for Commerce Commission, Economic Insights Ltd.

Lo Storto, C. (2018). A nonparametric economic analysis of the US natural gas transmission infrastructure: Efficiency, trade-offs and emerging industry configurations. *Energies*, 11(3), 519.

Lowry, M. N., & Getachew, L. (2009). Statistical benchmarking in utility regulation: Role, standards and methods. *Energy Policy,* 37(4), 1323-1330.

Massol, O. (2011). A cost function for the natural gas transmission industry: further considerations. *The Engineering Economist*, *56*(2), 95-122.

Mastromarco, C. (2008). *Stochastic frontier models*. Department of Economics and Mathematics-Statistics, University of Salento. http://www.camillamastromarco.it/CIDE/STFR.pdf

Maziotis, A., & Molinos-Senante, M. (2025). Benchmarking technical efficiency of water utilities in Chile under heterogeneity: A latent class frontier approach. *Utilities Policy*, 97, 102076.

Medal-Bartual, A., Molinos-Senante, M., & Sala-Garrido, R. (2016). Assessment of the total factor productivity change in the Spanish ports: Hicks–Moorsteen productivity index approach. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, 142(1), 04015013.

Mocholi-Arce, M., Sala-Garrido, R., Molinos-Senante, M., & Maziotis, A. (2022). Performance assessment of the Chilean water sector: A network data envelopment analysis approach. *Utilities Policy*, 75, 101350.

Molinos-Senante, M., & Maziotis, A. (2021). Benchmarking the efficiency of water and sewerage companies: Application of the stochastic non-parametric envelopment of data (stoned) method. *Expert Systems with Applications*, 186, 115711.

Molinos-Senante, M., Marques, R. C., Pérez, F., Gómez, T., Sala-Garrido, R., & Caballero, R. (2016). Assessing the sustainability of water companies: A synthetic indicator approach. *Ecological indicators*, 61, 577-587.

Molinos-Senante, M., Maziotis, A., Sala-Garrido, R., & Mocholi-Arce, M. (2022). An investigation of productivity, profitability, and regulation in the Chilean water industry using stochastic frontier analysis. *Decision Analytics Journal,* 4, 100117

Molinos-Senante, M., Mocholi-Arce, M., & Sala-Garrido, R. (2016). Efficiency assessment of water and sewerage companies: a disaggregated approach accounting for service quality. *Water resources management,* 30(12), 4311-4328.

Monsma, V., Ilson, T., Thodla, R. & Hussain, A. (2024). *Repurposing Onshore Pipelines for Hydrogen*. Whitepaper DNV, https://www.dnv.com/focus-areas/hydrogen/repurposing-pipelines-for-hydrogen-guiding-operators-through-the-re-evaluation-process/

Nillesen P, Pollitt M. (2010) Using regulatory benchmarking techniques to set company performance targets: the case of US electricity. *Compet Regul Netw Ind* 11(1):50–84

O'Donnell, C. J. (2011). The sources of productivity change in the manufacturing sectors of the US economy. *Working Paper WP07/2011.* CEPA, University of Queensland, ISSN 1932-4398.

O'Donnell, C. J. (2018*). Productivity and efficiency analysis*. Singapore: Springer Singapore.

OFGEM (1999). *Reviews of public electricity suppliers 1998- 2000*. Distribution price control review, final proposals. https://www.ofgem.gov.uk/sites/default/files/docs/1999/12/reviews-of-public-electricity-suppliers-1998-to-2000---dpcr---02-12_0.pdf

Oliver, M. E. (2015). Economies of scale and scope in expansion of the US natural gas pipeline network. *Energy Economics*, 52, 265-276.

Ondrich, J., & Ruggiero, J. (2001). Efficiency measurement in the stochastic frontier model. *European Journal of Operational Research*, 129(2), 434-442.

Pointon, C., & Matthews, K. (2016). Dynamic efficiency in the English and Welsh water and sewerage industry. *Omega*, 58, 86-96.

Price, C. W., & Weyman-Jones, T. (1996). Malmquist indices of productivity change in the UK gas industry before and after privatisation. *Applied Economics*, 28(1), 29-39.

Ramsey, J.B. (1969) Tests for Specification Error in Classical Linear Least Squares Regression Analysis. *Journal of the Royal Statistical Society*, Series B 31, 350-371.

Rasmussen, K. E. (2025). Comparison of cost efficiency among electricity distribution companies in Northern Europe: A panel data stochastic frontier approach. *Energy Economics*, 108654.

Recordon, E., Rudnick, H., (2002). Distribution access pricing: application of the OFTEL rule to a yardstick competition scheme. *IEEE Transactions on Power Systems* 17 (4),1001–1007.

Richmond, J. (1974), Estimating the efficiency of production, *International Economic Review* 15, 515–521.

Rita, R., Marques, V., Bárbara, D., Chaves, I., Macedo, P., Moutinho, V., & Pereira, M. (2023). Crossing non-parametric and parametric techniques for measuring the efficiency: Evidence from 65 European electricity Distribution System Operators. *Energy*, 283.

Romano, T., Cambini, C., Fumagalli, E., & Rondi, L. (2022). Setting network tariffs with heterogeneous firms: The case of natural gas distribution. *European Journal of Operational Research*, 297(1), 280-290.

Rossi, M. A. (2001). Technical change and efficiency measures: the post-privatisation in the gas distribution sector in Argentina. *Energy Economics*, 23(3), 295-304.

Rudnick, H., & Raineri, R. (1997). Chilean distribution tariffs: Incentive regulation. De) *Regulation and Competition: The Electric Industry in Chile*, 223-257.

Rudnick, H., Donoso, J. (2000). Integration of price cap and yardstick competition schemes in electrical distribution regulation. *IEEE Transactions on Power Systems* 15 (4), 1428–1433.

Ruggiero, J. (2007). A comparison of DEA and the stochastic frontier model using panel data. *International Transactions in Operational Research*, 14(3), 259-266.

Rui, Z., Metz, P. A., Reynolds, D. B., Chen, G., & Zhou, X. (2011). Historical pipeline construction cost analysis*. International Journal of Oil, Gas and Coal Technology*, 4(3), 244-263.

Sadjadi, S. J., Omrani, H., Abdollahzadeh, S., Alinaghian, M., & Mohammadi, H. (2011). A robust super-efficiency data envelopment analysis model for ranking of provincial gas companies in Iran. *Expert Systems with Applications*, 38(9), 10875-10881.

Sala-Garrido, R., Mocholí-Arce, M., Maziotis, A., & Molinos-Senante, M. (2023). Benchmarking the performance of water companies for regulatory purposes to improve its sustainability. *Npj Clean Water*, 6(1), 1.

Schmidt, P. and Sickles, R. (1984), Production frontiers and panel data, *Journal of Business and Economics Statistics* 2, 367–374.

Shuttleworth, G. (2005). Benchmarking of electricity networks: Practical problems with its use for regulation. *Utilities Policy*, 13(4), 310-317.

Sickles, R. C., & Streitwieser, M. L. (1992). Technical inefficiency and productive decline in the US interstate natural gas pipeline industry under the Natural Gas Policy Act. *Journal of Productivity Analysis*, 3(1), 119-133.

Sickles, R. C., & Streitwieser, M. L. (1998). An analysis of technology, productivity, and regulatory distortion in the interstate natural gas transmission industry: 1977–1985. *Journal of Applied Econometrics*, 13(4), 377-395.

Sickles, R. C., & Zelenyuk, V. (2019). *Measurement of productivity and efficiency.* Cambridge University Press.

Smith, A. S. (2012). The application of stochastic frontier panel models in economic regulation: Experience from the European rail sector. *Transportation Research Part E: Logistics and Transportation Review*, 48(2), 503-515.

Stephens, M.A. (1986) *Tests based on EDF statistics*. In: D'Agostino, R.B. and Stephens, M.A., eds.: *Goodness-of-Fit Techniques*. Marcel Dekker, New York.

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

Sumicsid-CEER (2019a) *Pan-European cost-efficiency benchmark for gas transmission system operators*, TCB18 Main report V1.2, 2019-07-12.

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

Sumicsid-CEER (2021a). *Financial reporting guide*. Final version for CEER-TCB21.

Sumicsid-CEER (2021b). *Gas asset reporting guide*. Final version for CEER-TCB21.

Sumicsid-CEER (2023a). *Note on Comparability*. Technical Report for CEER-TCB21.

Sumicsid-CEER (2023b) *Note on Environmental Modelling*, Technical Report in CEER-TCB21, version 1.4, 2023-01-13.

Sumicsid-CEER (2024). *Model Specification Note Gas*. Final Report for CEER-TCB21.

SUMICSID, Consentec and Frontier Economics. Report for CEER. https://www.researchgate.net/publication/265592571_E3GRID2012_European_TSOG_Benchmarking_Study

Swiss Economics, Sumicsid & IAEW (2023). *Kostentreiberanalyse und Effizienzvergleich der deutschen Fernleitungsnetzbetreiber zur vierten Regulierungsperiode* [EFG4]. Gutachten im Auftrag der Bundesnetzagentur. (In German).

Toloo, M., & Babaee, S. (2015). On variable reductions in data envelopment analysis with an illustrative application to a gas company. *Applied Mathematics and Computation*, 270, 527-533.

Tovar, B., Ramos, Real, F. J., & Fagundes de Almeida, E. L. (2015). Efficiency and performance in gas distribution. Evidence from Brazil. *Applied Economics*, 47(50), 5390-5406.

Vikas & Bansal, R. (2019). Efficiency evaluation of Indian oil and gas sector: data envelopment analysis. *International Journal of Emerging Markets*, 14(2), 362-378.

Waidelich, P., Haug, T., & Wieshammer, L. (2022). German efficiency gone wrong: Unintended incentives arising from the gas TSOs' benchmarking. *Energy Policy*, 160, https://doi.org/10.1016/j.enpol.2021.112595.

Walker, N. L., Styles, D., Gallagher, J., & Williams, A. P. (2021). Aligning efficiency benchmarking with sustainable outcomes in the United Kingdom water sector. Journal of Environmental Management, 287, 112317.

Winsten, C. (1957). Discussion on Mr. Farrells paper, *Journal of Royal Statistical Society*, Series A 120, 282–284.

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

Yépez, R. A. (2008). A cost function for the natural gas transmission industry. *The Engineering Economist*, *53*(1), 68-83.

SUMICSID GROUP
Rue Maurice Lietart 56
B-1150 Bruxelles, BELGIUM
info @ sumicsid.com

http://www.sumicsid.com