PROJECT E2GAS

Benchmarking European Gas Transmission System Operators

FINAL REPORT

Per J. AGRELL
Peter BOGETOFT
Urs TRINKNER

OPEN

2016-06-02

Disclaimer

This is the final report for the pan-European efficiency benchmarking of gas transmission system operations commissioned by the Netherlands Authority for Consumers and Markets (ACM), Den Haag, on behalf of the Council of European Energy Regulators (CEER) under the supervision of the authors: project team leaders professors Per AGRELL and Peter BOGETOFT from SUMICSID and Urs TRINKNER from Swiss Economics. The project acronym is E^2GAS (Economic Efficiency analysis of GAS transmission operators).

The report is open. No part of the final report has been formally reviewed by the Commissionaires and expresses only the viewpoint of the authors, who exclusively bear the responsibility for any possible errors.

This version contains the final DEA and SFA results for an updated run and replaces an earlier report issued 2016-05-17.

Benchmarking European Gas Transmission System Operators
Final report. Project no: 361 / e2GAS
Release date: 2016-06-01

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

Executive Summary

The European efficiency benchmarking for gas transmission system operators (E^2GAS) is an initiative by a group of national regulatory authorities (NRAs) in the Council of European Energy Regulators (CEER). The project is the largest regulatory efficiency benchmarking so far in gas transmission, delivering reliable cost-efficiency estimates and establishing a new regular cooperation project among the NRAs, just as in electricity transmission. The project, following a feasibility study in 2014, is to be seen as a pioneer study where the data definitions, data collection and methodological aspects are emphasized.

Comparability

The primary challenge of any benchmarking is assuring comparability among observations emanating from operators with differences in organization, task scope and asset base. Drawing upon the analysis in the feasibility study, this challenge is addressed by (i) limiting the scope to comparable activities in transport and capacity provision, (ii) controlling to systematic differences in labor costs, (iii) standardizing the asset life-times and capital costs to equal conditions, (iv) excluding country-specific cost factors (land, taxes), (v) controlling for joint assets and cost-sharing, (vi) adjusting capital costs for inflation effects.

Reliability

A prerequisite of the results to be reliable is the use of state-of-the art benchmarking techniques, in particular Data Envelopment Analysis (DEA). DEA is the predominant regulatory method, the validity of which has been confirmed several times in judicial and regulatory appeals. It is routinely used in Germany, Norway, the Netherlands, Sweden, and other countries. The use of DEA is clearly motivated in infrastructure provision, since the method can handle multiple outputs (e.g. asset categories, capacity metrics and service metrics) without prior weights. It also provides a conservative estimate based on interior estimations with real peers, rather than e.g. linear projections based on statistical estimates of cost-factors. Moreover, the DEA models can be decomposed into partial productivities, down to unit cost measures, leading a rich information source for operators and NRAs. The reliability and replicability of DEA results are immediate, since the method does not depend on any *ad hoc* parameters, but relies on the input data and linear programming. The sensitivity analysis shows that the results are robust and show the expected signs for all coefficients.

Verifiability

The quality of the data material in the project is a key determinant of the precision of the project results. The project addresses this criterion (i) by issuing specific data collection guides and templates to avoid the use of incomparable data sources, (ii) by defining an external control process for the submitted data through the NRAs, (iii) by organizing an internal data validation process for both technical and economic data, (iv) by fully disclosing all processed data to each respective operator for control and confirmation to avoid misinterpretations and error, (v) by organizing interactive workshops to enable questions, and (vi) by providing online support on the project platform for submitting operators.

Confidentiality

The data involved in the study go deeply into the operational efficiency of the participating operators. As this data are of crucial economic importance to the enterprises, the integrity and confidentiality of the data are taken seriously in the project both from structural, procedural and

organizational viewpoints. The balance between procedural transparency and data confidentiality is always a delicate matter. The pioneer E^{2}GAS project has chosen a cautious policy in order to assure a secure participation of all operators, irrespective of economic stakes at hand. This implies full confidentiality of data and results for all participants, open and transparent dissemination of all methods, public parameters and average results.

Approach

The methodological approach in E^{2}GAS enhances that of similar benchmarking studies in electricity (e^{3}GRID) in two ways.

The electricity benchmarking in transmission starts at a general activity model that is estimated with only few complexity factors, later to be complemented through a second stage of data submission. In E^{2}GAS, the main data collection already takes into account the density-related, topological, geological and meteorological factors leading to cost differences, based on the engineering analysis in the feasibility study. This means that the core activity model will at the outset control for most asset-related cost differences, as well as the structural cost differences through the activity decomposition. However, for completeness, the project also includes an analogous second submission process for operator-specific conditions that are not detected through the general engineering cost function. This latter optional process does not involve any general assessment of structural differences between operators, but opens for the correction of known conditions of imposed, material and durable cost-increasing procedures and assets that affect the assessment.

Normalized grid function

Regulatory benchmarking is in principle measuring how well an operator meets a set of exogenous output targets with minimal resources. For gas TSOs, the grid in itself is then only a means, not an end in itself. If the delivery task is known at a given moment, one could in principle use the grid as an input variable to be minimized. The most efficient operator would be the grid with the highest annual utilization; the others would be classified as inefficient. The problem is that this exposes the grids to an investment risk that contradicts other objectives, such as security of supply. A grid may have lower utilization because it is preparing for future growth (good forecasting), or because it has been exposed to fluctuations in gas prices or business cycles (exogenous risks). In both circumstances, the provision of the grid in itself is a service put at the disposal of the grid users, shippers and traders. Thus, for a TSO, the grid is indeed an essential output. However, this does not mean that any investment at any cost is efficient. An efficient operator is able to offer a given system size at a lower cost than an inefficient cost, assuming structural comparability. Thus, we need a metric to measure the 'technical size' of the grid. We observe that a grid consists of a several classes of components (pipelines, regulators, compressors, stations, ...) each with many technical parameters for specifications of dimensions, material, power etc. Thus, a naïve summing over the assets would not be a good indication of the 'technical size'. Likewise, a statistical estimate of the relative costs for each type and class of asset would require very large samples and be subject to identification problems. Thus, we have developed a technical measure called the Normalized grid that basically is nothing else than a weighted sum of the grid assets that serve the in-scope activities. The weights for the normalized grids correspond to the relative costs of components, transportation, installation and operation for the different assets, based on engineering expertise. New to this project, it also includes the cost factors resulting from environmental complexities (land use, topography, soil structure, humidity). The normalized grid is not derived from the accounting data from the operator, it is calculated directly from the technical asset data. Thus, the desired investment efficiency can be obtained by simply observing the ratio of benchmarked (standardized) capital costs to the normalized grid.

Activity model

The benchmarking model in E^2GAS has total expenditure in real euros at 2014 value as input, net of energy costs and after structural corrections for labor costs, inflation and outsourcing. The three outputs are (i) a normalized grid measure to represent the grid provision, basically a weighted sum of all activity-relevant pipeline, regulator and compressor assets, (ii) the maximum of peak capacity for delivery and injections, to represent capacity provision and (iii) the total number of connections to the pipeline system, to represent customer service. The resulting model shows strong statistical properties; the specification explains 92% of the variance in total expenditure and is robust to linear or log-linear functional forms. Extensive tests with over 100 other variable candidates showed no significant omitted variables with correct signs for inclusion.

Benchmarking results

The primary benchmarking model is a DEA model under increasing returns to scale, meaning that operators that are below best-practice scale are not penalized for their size but larger operators are always compared to best-practice. The dataset consisting of 22 operators in a crossection for two reference years (2010 and 2014) was reduced to 21 operators by identifying one extreme outlier using standard techniques.
The results give an average cost efficiency of 79% in the sample, where two operators are classified as frontier outliers and six operators constitute the best-practice peers. This means that on average there is a 21% efficiency difference in terms of total expenditure, corrected for inflation and labor cost differences.

Robustness

The DEA run using the standard model has been validated with an equivalent SFA model, yielding a rank-order correlation of 74% and an average efficiency level of 78%. This indicates a high inter-method consistency of the results. A series of alternative specifications for total cost and parameter settings have been tested to investigate the robustness of the results. The average results are very robust to changes in the model specification, ranging from 78% to 81%. For individual operators, the average range of highest to lowest results across specifications is 6%-units (median range 3%-units); the maximum range is 27%-units.

Caveats

As mentioned, the e2GAS project is a pioneer and the first of its scope and kind in gas transmission. During a relatively fast process of seven months, three interactive workshops and three data collections, a group of European regulators and gas transmission operators underwent a benchmarking leading to a convincing model. The data collection and activity decomposition formats, the technical pipeline asset characteristics are lasting elements that could serve in future projects, irrespective of benchmarking technique used. The improvements in future benchmarking project could address the 'passive' operators that did not fully use the reporting format, the need for specific workshops on cost reporting to collectively solve allocation problems and the collection of data for a longer time horizon to provide dynamic estimates of productivity changes.

Table of Contents

1. Project objectives and organization

In this chapter we state the project objectives and the organization. At the end of this chapter, we summarize the objectives and provide an outline for the rest of the report.

1.01 This project plan concerns *Pan-European gas TSO benchmark* for CEER under the leadership of SUMICSID, commissioned by the Netherlands Authority for Consumers and Markets (ACM), Den Haag, acting as representative for a undefined group of European NRAs, active members of CEER and regulating at least one gas transmission system operator (TSO). The project acronym is e^2GAS (**E**conomic **E**fficiency analysis of **GAS** transmission operators).

1.02 The project leader for this mission is prof. dr. Per AGRELL, Senior Associate. The project also included contributions from Senior Associate prof. dr. Peter BOGETOFT, Consultant Daniele BENINTENDI, Senior Engineers Henri BEAUSSANT and Jacques TALARMIN from SUMICSID and from Dr. Urs TRINKNER, Dr. Martin KOLLER, Ms Isobel OXLEY and Mr Matthias HAFNER from Swiss Economics, respectively.

1.03 The 22 participating gas transmission system operators (TSOs) and the associated national regulatory authorities are listed in Table 1-1 below. In addition, the NRAs from two countries (Austria and Greece) participated as observers in the process without TSO participation.

Table 1-1 e2GAS Participants by country, NRA and TSO.

Country	NRA	TSO
BE	CREG	Fluxys
DE	BNetzA	bayernets
DE	BNetzA	Fluxys Tenp
DE	BNetzA	GASCADE Gastransport
DE	BNetzA	Gastransport Nord
DE	BNetzA	Gasunie Deutschland Transport Services
DE	BNetzA	GRTgaz Deutschland
DE	BNetzA	jordgasTransport
DE	BNetzA	terranets bw
DE	BNetzA	Thyssengas
DE	BNetzA	Nowega
DE	BNetzA	Ontras - VNG Gastransport
DE	BNetzA	Open Grid Europe
DE	BNetzA	DONG Energy Pipelines
DK	DERA	Energinet.dk
ES	CNMC	ENAGAS
ES	CNMC	REGANOSA
FI	EMV	Gasum
HR	HERA	Plinacro
NL	ACM	GTS
PT	ERSE	REN - Gasodutos
UK	OFGEM	National Grid Gas Transmission

1.1 Main objectives

1.04 The overall objective of the study was to provide reliable static cost efficiency estimates for all participating TSOs. In particular, the project aimed at the following elements:

1) Restricted initial benchmarking scope
The scope of the benchmarking is defined in the Tender document, i.e. (i) transport services including transit, (ii) asset provision: pipeline system with control equipment, and (iii) support activities in grid planning, grid financing, - construction, - maintenance and – metering.
2) Comprehensive initial data request
The project draws directly on the recommendations in the feasibility study in 2014 PE2GAS. Thus an initial data call was followed by a consultation process for the templates and the data collection period in the first month of the project.
3) Well-defined and relevant TSO benefits
A new additional service for TSOs was given in the form of interactive benchmarking, providing anonymized data on other variables and indicators that have managerial importance to the TSOs, but that are not necessarily included in the core benchmarking.

1.2 Project management

1.05 The Commissioning NRAs appointed a steering group for the project consisting of NRAs in the CEER taskforce. The Steering Group organized teleconferences every three weeks, inviting the project leader from the consultant whenever relevant.

1.06 During the project workshops with NRAs and TSOs, the consultant made minutes available on the project platform from all workshops under the Chatham House Rule (comments recorded without identity or affiliation of speaker provided).

1.3 Project deliverables

1.07 The project produced two deliverables to document the results and the process:

1.08 **TSO-specific report, R1:**
Clear and informative report on all used data, parameters and calculations leading to individual results, decomposed as useful for the understanding. The report only contains data, results and analyses pertaining to a single TSO. The confidential report was uploaded in an electronic version to each authorized NRA on the platform.

1.09 **Final report, R2**:
This document constitutes the final report documenting the process, model, methods, data requests, parameters, calculations and average results, including sensitivity analysis and robustness analysis. The report is intended for open publication and does not contain any data or results that could be linked to individual participants.

1.10 This version of the final report is the result of a revised run 28/05/2016 to correct some data and code errors. The specification for this rerun is presented in art 5.36 below.

1.4 Reading guide

1.11 In Chapter 2 the process is described. The data collection is documented in Chapter 3. The model is presented in Chapter 4, followed by the results in Chapter 5. The report is closed with some conclusions in Chapter 6. Further information about the different topics can be found in the supplementary documentation mentioned in the next section.

1.5 Appendixes

1.12 The following documents were issued during the project process for the participants and included in Appendix they form an integral part of the final report:

1)	Cost Reporting Guide (Appendix A)	V1.0	2015-11-11
2)	Asset and Output Reporting Guide (Appendix B)	V1.2	2015-12-01
3)	Operator Specific Conditions (Appendix C)	V1.0	2015-12-16

2. Benchmarking process

In this Chapter the benchmarking process is summarized, including the points of decision and interaction with the project participants.

2.1 Overall process

2.01 The project, as any large undertaking involving the coordination of numerous countries and different stakeholders requires a careful organization. To facilitate the organization and coordination of the project, the process was divided into different work packages and phases, and clear milestones shall be defined. Below we describe the parts of the process that are of interest for an external audience. More detailed information about the internal project organization can be found in the project plan.

2.02 The project planning of the study aimed at a flexible, constructive and interactive process not only to meet the specified deadlines, but also to inspire support and promote interest in the final study among the stakeholders.

2.03 The objectives of the process planning were explained in the feasibility study PE2GAS, cf. Agrell et al. (2014, art 7.02). Thus, the project organization must assure:

1) Data quality assurance
2) Confidentiality of data
3) Clear, fast and complete communication
4) Well explained, documented and justified results

2.04 The project process contained six types of components that partially overlapped, thereof three that were already initiated through the PE2GAS project:

a. Methodological work based on econometrics, convex analysis, and efficiency and productivity analysis to solidify the underpinnings of the models. (Started in PE2GAS)

b. Asset and cost data definition guides to ensure precise understanding and assure comparable date amongst the TSOs. (Started in PE2GAS)

c. Data collection routines between the coordinators and the involved TSO, including clear routines for submitting and evaluating TSO specific claims. (Outlined in PE2GAS)

d. Interactive workshop process, each introducing one element of the methodology towards the final result. Final release versions of any project documents were posted at the project platform with notifications issued for each such upload or revision.

e. Data validation and verification with internal partners.

f. Final reporting, detailed confidential and individualized reports to each participant.

2.2 Project Team assignments

2.05 The TSO assigned a project team at the start of the e^{2}GAS round. The project team has three functions: (1) *project manager*, (2) *data manager* and *(3) technical manager*.. Only the named project members from the TSO had access to the project platform. The TSO project team interacted on the e^{2}GAS project platform and took part in (at least some) e^{2}GAS Workshops. Designated members of the project team make sure that documentation that is published at the platform to the workshop is read and, if necessary, commented as to participate actively in the workshop.

2.06 Internally, the consultant's staff was organized into a data management team, an econometric team and a technical team.

2.3 Project documentation

2.07 In addition to the contractual project deliverables, the project plans a certain number of internal documents, some of which are listed below and were used in the project planning.

2.08 Data Guides (draft) G1:
Draft data definition guides and templates for Assets, Costs and Output data. Presented at Workshop 1 (W1).

2.09 Data Guides G2:
Final data definition guides and templates for Assets, Costs and Output data.

2.10 Data Call Operator Specifics G3:
Reporting guide and form for operator specific factors. Presented at Workshop 2 (W2)

2.11 Rulings on Operator Specifics G4:
Motivated decisions on the submissions for operator specifics DS2. Included the option to submit complementary data DS3 for approved factors/conditions

2.4 Milestones

2.12 The main project events are listed in Table 2-1 below, including mention on deliverables from the consultants, input required from NRAs and input required from TSOs.

Table 2-1 Project calendar.

Date	Event	Event	Deliverable	Review	Data
2015-10-06	Project launch	W1	G1		
2015-11-03	Review G1 end			G1	
2015-11-11	Final release G2		G2		
2015-11-16	Release draft G3		G3		
2015-12-03	Review G3 end			G3	
2015-12-16	Operator specifics	W2			
2015-12-11	Submission DS1				DS1
2016-01-17	Submission DS2				DS2
2016-02-16	Release G4 rulings		G4		
2016-02-29	(Submission DS3)				(DS3)
2016-03-01	Model run				
2016-03-16	Final results	W3			
2016-04-29	Final individual reports		R1		
2016-05-31	Final report release		R2		

2.5 Workshops

2.13 The Workshops, three open project meetings at crucial milestones (cf. Table 2-2) were important to advance the project at crucial points. Here important information was released, explained and discussed.

Table 2-2 Project workshops.

Activity	Date	Where
Workshop 1 (project launch)	2015-10-06	Brussels
Workshop 2 (operator specifics)	2015-12-16	Berlin
Workshop 3 (final results)	2016-03-16	Vienna

W1: Project launch for data collection

2.14 At the first workshop with all TSO and NRA participants invited, the project objectives and planning were presented, the data collection was initiated and practical procedures were explained. The draft Data Definition Guides (G1, Cost, Asset, Output) were presented and the participants were invited to submit their opinions on the review versions of the following the workshop.

W2: operator specifics

2.15 The second workshop was designed to provide feedback from the data collection, to present benchmarking methods to be used and to explain how TSOs may submit additional data (operator specific data) to improve the comparability of the standard estimates. Discussions with TSOs at and after W2 lead the technical team to improve the data collection for environmental conditions and thereby reduced the workload for the TSOs in the operator-specific data collection DS2.

W3: final results

2.16 The third workshop closed the project by presentations of the general results, a summary of methods and parameters used and the sensitivity analysis. Constructive feedback at W3 from TSOs and NRAs lead to the replacement of one output parameter, (weighted) pressure difference, with a capacity-related output parameter.

2.6 Reporting

Single point of contact

2.17 Communication in the project was structured to assure efficiency, transparency and to comply with the strict confidentiality requirements. The procedure used in e2GAS is through the Sumicsid platform http://sumicsid.worksmart.net. All NRAs and TSOs assigned to the project were given personal user names for relevant access to the platform by informing the project manager. Project management, document and data handling were handled using the secure online platform. In this way, all relevant parties have access to the same information at the same time. It also allows for traceability of information release, access and exchange throughout the project. All releases of pre-run data for validation, interim or final results were made through the platform at the appropriate level of authority. All access to data was logged and monitored. This procedure guaranteed that confidential data stays so, which is not the case when email is used for project communication.

2.7 Data reporting

2.18 The data submission was organized in three stages.

2.19 The first data set (DS1) contained raw data on grid assets and audited costs from annual reports; followed by decomposed cost data. The basis for this data collection is the Data Guides (Call C, Appendix A, and Call XY, Appendix B) submitted for review at W1. The format of submission was in form of self-explanatory Excel-templates, presented to the NRAs and TSOs at W1. The collection of DS1 was terminated at 11/12/2015.

2.20 An optional second data set (DS2) contained additional material for operator specific cost drivers. The basis for this data collection was the Operator Specific Data Guide (Call Z, Appendix C), released for review and presented at W2. The submission of DS2 closed at 17/01/2016. The validation of DS2 was made by the data team of cost data and by the technical team for technical data, final motivated responses were communicated to the NRAs for further dissemination to the TSOs.

2.8 Operator-specific conditions (Call Z)

2.21 As mentioned above, the data collection DS3 concerned five specific allowances that could be deducted from OPEX and/or CAPEX. These adjustments are implemented on the gross amounts prior to the calculation of benchmarked OPEX and CAPEX.

2.22 In addition to information collected through the central calls (C, X, Y), the participants could complement the data for the study by submitting candidate variables and factors (claims) for potential inclusion as *operator specific corrections*. This process and the confidentiality provisions regarding the claims were more thoroughly described in the guide *Operator Specific Conditions* (Appendix C) *(Call Z)*. The Call Z process was overseen by the steering group and the review work was done by the technical team and the econometric team under the direction of the project manager.

2.23 The process for the operator specific conditions is structured to decrease the work for the operators, to increase transparency and to maintain the time plan for the project. The process was initially presented at W1, a draft guide for reporting was submitted to the participants for consultation 16/11/2016 with dead line 04/12/2016. The final provisions in the guide were presented at W2 and the final version of the guide was published 17/12/2016.

2.24 The process for the data collection DS2 followed the steps below

1. First submission of preliminary claims, DS2 (17/12/2016 – 17/01/2016)
2. Processing of claims DS2 (18/01/2016 – 12/02/2016)
3. NRA review of analysis on DS2 (12/02/2016 – 16/02/2016)
4. Disclosure of conditions DS2 (16/02/2016 – 29/02/2016)
   –In motivated decisions on each submitter on Worksmart (16/02/2016)
   –Process presented at W3
   –Approved categories in specific document on Worksmart (16/02/2016)
5. Complementary submission Z (DS3) (16/02/2016 – 29/02/2016)
   –Inclusion in the final run after W3

2.25 The review process is depicted in Figure 2-1 below. Initially, the submitted claims were reviewed for eligibility. The initial filter is to assure that the claim concerns a specific condition or cost valid for the operator and that it is documented with some supporting material to constitute a valid basis for correction of cost basis or model outputs.

2.26 Any admissible claim is then reviewed by the appropriate team(s) among the consultants to investigate whether it qualifies for inclusion in the study as an operator-specific allowance. The second step for eligible claims involved the actual review with respect to the three criteria of *exogeneity, materiality* and *duration*. Claims that passed on the three criteria were resubmitted to the NRAs in preliminary assessments with *requests for endorsement* (if relevant) of submitted information, in particular with respect to exogeneity. Endorsed claims from this step were reviewed for possible *inclusion in ongoing revisions* of the general model or data collection. Finally, the approved and endorsed claims that were not subject to model extensions were declared *approved* and announced to the participants for possible resubmission in DS3.

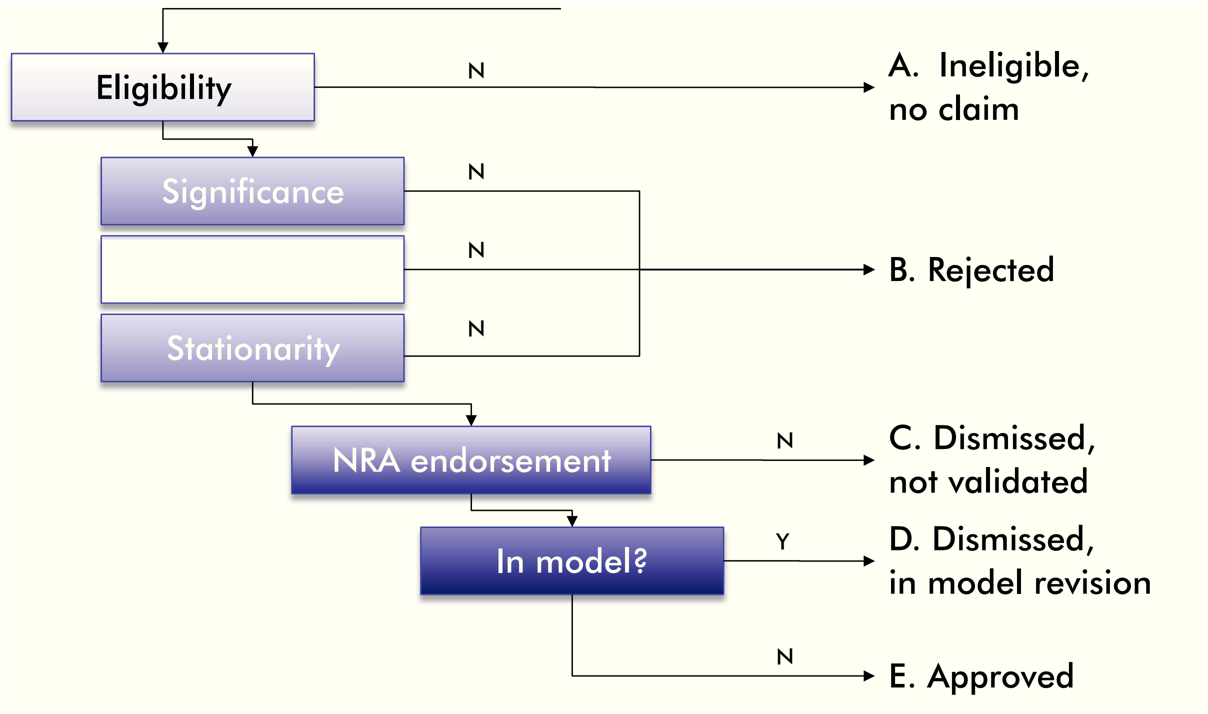

Figure 2-1 Process for the review of operator-specific conditions.

1. Exogeneity

2.27 Claims that referred to conditions, equipment and /or operating standards that may have impact on Capex and/or OPEX but that are the result of an internal decision making process in the firm (e.g. metering standards) were considered endogenous conditions. Crucial for the determination of the exogeneity was the existence of a legal or regulatory obligation to perform a non-standard task/cover the cost *in spite* of an explicit interest on behalf of the operator to adopt a different policy.

2.28 Few claims were rejected on this criterion. Examples included operational choices that might not be cost-optimal (today) such as metering but that did not result from any exogenous involvement. The prerogative of interpretation of exogeneity was given to the NRAs.

2. Significance

2.29 Claims that indeed were exogenous and stationary had to exceed a materiality criterion. The criterion announced was 3% of benchmarked cost by claim (not cumulated). The application was finally by grouped claims. Smaller differences, if indeed relevant when all effects are factored in, may and should be addressed by the NRAs within its applicable regulatory framework.

2.30 Several claims were rejected on this criterion, but primarily with the motivation that there was simply no documentation at all concerning the materiality of the claim. Some claim related to marginal differences between an optimal practice and an imposed suboptimal routine, for which there was only an estimate of the total cost, not the marginal cost increase due to the specific condition.

3. Duration

2.31 Claims that refer to restructuration, accidents, refurbishing, upgrading of assets etc are often related to sporadic events. These costs were included in the data reporting for investments and operating expenditure. Some other claims related explicitly to past policies with the argument that they currently are non-controllable. This motivation was not accepted as such, since by extension this would render all investments non-controllable after the fact.

2.32 Few claims were investigated on this criterion and none was rejected uniquely based on this criterion.

Model inclusion

2.33 A certain number of claims related to outputs and environmental conditions were dismissed on the grounds that they are already included among the model parameters, listed in Call XY documentation or in the final model. In particular this concerned claims for compressor power, asset intensity, soil types, compressor fuel costs and service area differences.

Outcome

2.34 The results of the review process are listed in Table 2-3 below. In all there were 33 claims from 6 TSOs to process. As seen from the table, the evaluation has been in application of the criteria above. As stated elsewhere, this does not mean that the factors in themselves are not complicating or cost-increasing, nor is it a recommendation for NRAs to disregard these costs as imprudently occurred. Some claims were non-unique, meaning that several operators submitted similar claims.

Table 2-3 Outcome of review process for operator specific claims.

Type	#	Decision	Economic	Joint	Technical
A	1	Not eligible	1		
B	16	Rejected on criteria 1-3		12	4
C	0	Dismissed by NRA			
D	6	Dismissed (model feature)	1	5	
E	8	Approved (4 unique)	4	1	3
EP	2	Partially approved (1 unique)		1	1
	33		6	19	8

2.35 In the project, five factors from DS2 were approved for additional reporting from any participant. This request formed data set DS3 limited to the five factors listed below:

1) Specific security and public safety investments imposed by law or regulation
2) Odorization, equipment and operation
3) Sea crossing pipelines
4) Gas chromatographs, equipment
5) Reserve (off line) capacity in regulators and compressors, output value

2.36 The accepted factors above were defined by type of costs, documentation and validation procedure for reporting in DS3.

3. Data collection

In this chapter, the data collection and the data validation process are discussed.

3.1 Procedure

3.01 The actual timeline for the data collection process is already summarized in Table 2-1 above. In this Section, we provide an overview on the operational processes of the data collection performed by the data team.

3.02 As a principle, manual manipulation of TSO specific data by the consultants was avoided as far as possible in order to exclude any source of error. To this end, data from cost [C] and asset and output [XY] templates were imported, cleaned and consolidated automatically.

3.03 The data collection process was organized as follows.

1) TSO upload

- TSO uploads of DS1 templates C and XY mark the starting point for the validation by the data team.

2) NRA approval

- Subsequent to TSO data deliveries, NRAs were asked to verify the data and to approve them by uploading the NRA approval form. The form included the possibility to include comments.

3) Validation procedure by data team

- NRA approved files were tagged by the data team on the platform.
- Subsequent, automated import and consolidation of approved files according to Section 3.2 below. Where necessary, specific requests were addressed to TSOs.
- With imported data, formal and analytical validation as described in Section 3.3 below. Where necessary asked for amendments.
- Final validation based on NRA approved data from all TSOs.

4) Outputs to the econometric team

- Written validation report;
- Automatically generated CSV files (C, CI_i, X, Y), with C, X, Y = cost, asset and output data of all TSOs and CI_i = investment stream of TSO i.

3.2 Data consolidation

3.04 Data consolidation includes *import, cleaning* and *formal validation*.

3.05 Basis of the final data import are NRA approved data templates (C and XY) uploaded by TSOs to the project platform.

3.06 The data of the individual TSOs was imported with customized import procedures into a large consolidated database, separated along C, CI_i, X and Y data. For identification reasons, each observation was assigned a company ID. Some TSOs did not follow the templates structure and therefore information could not be imported properly. As a consequence, either TSOs were asked to amend their data or the import procedure was customized while not editing the original file itself. One TSO did for example report for each single pipe different attribute combinations. In this case every constellation was calculated and the reported pipe was split into different sections accordingly. Another TSO added output information about an additional quality level (“G”). Those values were added to the low quality information (“L”). Some other TSO specific issues which were detected during the data validation process (e.g. correction of ranges) were also corrected during the import procedure (see also Section 3.3 below).

3.07 After successful import, a general harmonization process was performed. There is for example no consistency among the different operators with regard to reporting missing data (e.g. “NA” or ”-“). Hence, all missing data were set empty (=””).

3.08 Finally, the consolidated C, X and Y data were exported to different comma separated (csv) files. In addition, a csv file for each operator’s investment stream (CI_i) was produced and saved. Furthermore, the created cost, output and asset files were imported into a statistical software which then was used for further data cleaning and data validation (see Section 3.3 below).

3.09 Table 3-1 gives an overview about the final data set.

Table 3-1 Overview of collected data.

	Value
# TSOs	22
# Cost files	22
# Output files	22
# Assets	22’632
thereof # pipes	11’911
thereof # controllers / regulators	2’995
thereof # connection points	7’726
Total pipe length (unweighted)	78’565 km
Total pipe volume	31’801’497 m^3

3.3 Data validation

3.10 Different complementing methods were used to validate the data. On the one hand, data was reviewed formally and on the other, plausibility of data was checked with the help of statistical software. Whenever irregularities were detected, they have been corrected as long as it was due to an obvious mistake. In all other cases, respective TSOs or NRAs were asked to verify suspect observations. To ensure transparency, corrections were made within the main file or with help of the statistical software - original data deliveries were never manipulated.

3.11 Subsequently, **format and value ranges** were validated as follows:

1) Completeness of data set,
2) Format of data (e.g. are only numbers in numerical parameters),
3) Sums of parameters,
4) Range / option checks.

3.12 **Completeness:** Some TSO did not report data about all connection points' pressure level. In these cases, values were replaced with the correspondingly maximum pressure value. In other cases, NRAs were asked to complete the data set. The final data is considered to be complete with regard to mandatory information.

3.13 **Format of data:** Several issues were detected. In numerous cases comments were found in numerical parameters. Comments were replaced and NRAs were contacted and / or its value was set empty. Some TSO also reported data within a range instead an exact number. In these cases, values were replaced by its mean or, if guidelines required, by its maximum.

3.14 **Sums**: Sums of main files and sums calculated with the statistical software were compared in order to verify that no error occurred during the process. Issues regarding comments and other strings in numerical parameters were detected and handled appropriately (see also completeness of data). The calculation of sums for the final data set did not indicate any errors of the procedure.

3.15 **Range / option check:** In this step, it was verified that reported values are within a reasonable range (e.g. percentage values are between 0 and 100). Also, it was verified that values from categorical parameters are one of the predefined options (e.g. gas quality level is either "L" or "H"). The check revealed that some parameters were reported in different manner (e.g. shares were reported in decimal or in percentages; some values were reported in millions). In addition, many TSOs did either not report values from the predefined options or labeled them differently (e.g. not in English). It was also detected that some information was reported in the wrong cells in the form. Again, obvious mistakes were corrected or standardized according to the guidelines otherwise TSOs, NRAs or the econometric team were contacted. Another TSO set faulty values for information with respect to ownership of connection points. Whenever the TSO reported a specific render, it was assumed that he owns partially the connection point, if no specific render was stated, value of ownership was assigned "100% own".

3.16 In a next step, it was verified that OPEX (in scope of benchmark) is properly calculated from TOTEX and if cost positions add up to total costs for each activity. In addition, the relations between different parameters were analyzed.

3.17 NRAs were asked to validate if OPEX is calculated accurately from TOTEX. Nevertheless, significant differences were found during the process. In most cases, costs were not assigned to the defined cost position as explained in the guidelines. In consultation with TSOs / NRAs cost assignment were amended. With regard to the summation of cost positions, only minor issues could be found.

3.18 The relation between different parameters of a TSO should be reasonable and to some degree similar to ratios of other TSOs. Hence, if the ratio of one TSO significantly differs compared to the others, this may indicate data related issues. Therefore, the following comparisons were plotted and analyzed:

- Ratios between first and second reported year.
  (eg. total man cost of year 1 vs. total man cost of year 2)

- Ratios of different cost paramaters
  (e.g. total net cost of transportation vs. OPEX).

- Ratios of different output parameters
  (e.g. energy delivered in TWh vs. energy delivered in billion cubic meters).

- Ratios of different asset parametes
  (e.g. pipe section length vs. pipe section volume).

- Ratios amongst differnet cost, output and asset parameter
  (e.g. OPEX vs. total pipe volume).

3.19 Figure 3-1 presents the plots for some examples.

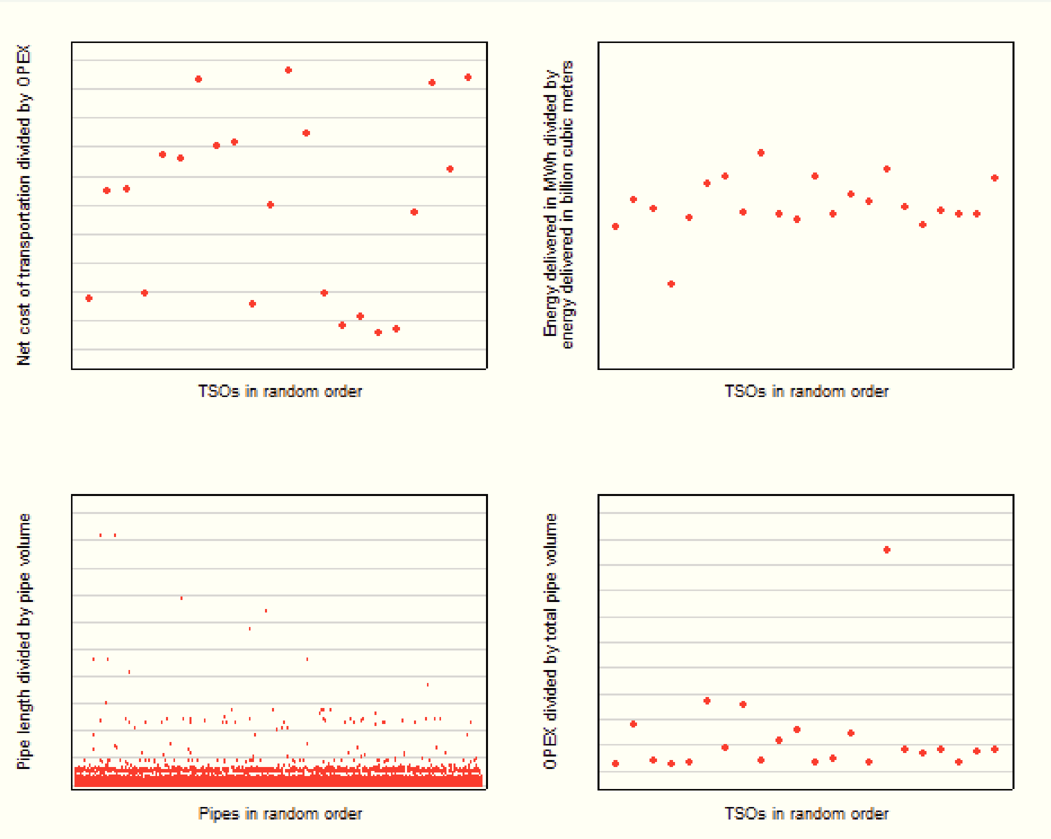

Figure 3-1 Examples of comparisons amongst different parameters.

3.20 The comparison of the different ratios helped to identify irregularities and outliers, which were then analysed in detail.

3.21 In the last step, spot checks were perfomed. Thereby, data was compared to externally available information such as annual reports or TSOs' websites. This verification was conducted especialy with regard to outliers of the previous step, and later with regards to peers of the benchmarking. Some discrepancies could confirmed to be true, others raised concerns about actual data issues and some could not be verified due to missing public information. Again, obvious mistakes were corrected, otherwise the econometric team was informed (validation report).

3.22 For example, the outlier in the upper right plot of Figure 3-1 could be confirmed to be correct by an article of Reuters. The outlier of the plot in the lower right is driven by currency differences. In the final calculations, exchange rates were taken into account.

3.23 In conclusion, the data valdiation process has been important in improving the input data for the benchmarking. Issues regarding data validation were presented at all three workshops W1, W2 and W3. In addition, data validation was an ongoing topic for discussion on the project platform. In the end, the final data are considered to be formally correct, complete and plausible.

3.24 In addition to the internal data validation, the crosschecking of data by NRAs and TSOs is an important control step. Specific data files have been produced and disseminated to the participants.

4. Methodology

This Chapter is devoted to the discussion of the methodological approach that has been used in the TSO benchmarking, including the important preparation in terms of activity analysis, cost standardization, asset aggregation and correction for structural comparability. The Chapter then addresses model specification and method choice.

4.1 Background

4.01 The benchmarking model is pivotal in incentive based regulation of natural monopolies. By essence, benchmarking is a relative performance evaluation. The performance of a TSO is compared against the actual performance of other TSOs rather than against what is theoretically possible. In this way, benchmarking substitutes for real market competition.

4.02 Of course, the extent to which a regulator can rely on such pseudo competition depends on the quality of the benchmarking model. This means that there is no simple and mechanical formula translating the benchmarking results into for example revenue caps. Rather, regulatory discretion – or explicit or implicit negotiations between the regulator, the industry and other interest groups – is called for.

4.03 Different regulatory conditions in different jurisdictions means that the benchmarking approach should ideally support a multiplicity of potential applications. To facilitate this while at the same time creating a coherent benchmarking approach, we start the analysis from a unit cost approach before extending it to the use of more advanced benchmarking methods like Data Envelopment Approach (DEA). The unit cost approach is informative, intuitive and can provide useful information for more process oriented analyses while the use of DEA allows us to do a comprehensive evaluation with less stringent a priori assumptions than a unit cost approach. The use of asset data and assets weights in the unit cost approach is also necessary to cope with the problems of estimating in a small data set.

4.2 Steps in a benchmarking study

4.04 The development of a regulatory benchmarking model is a considerable task due to the diversity of the TSOs involved and the potential economic consequences of the models. Some of the important steps in model development are:

4.05 **Choice of variable standardizations:** Choices of accounting standards, cost allocation rules, in/out of scope rules, asset definitions and operating standards are necessary to ensure a good data set from TSOs with different internal practices.

4.06 **Choice of variable aggregations:** Choices of aggregation parameters, such as interest and inflation rates, for the calculation of standardized capital costs and the search for relevant combined cost drivers, using, for example, engineering models, are necessary to reduce the dimensionality of potentially relevant data.

4.07 **Initial data cleaning:** Data collection is an iterative process where definitions are likely to be adjusted and refined and where collected data is constantly monitored by comparing simple Key Performance Indicators (KPIs) across TSOs and using more advanced econometric outlier - detection methods.

4.08 **Average model specification:** To complement expert and engineering model results, econometric model specification methods are used to investigate which cost drivers best explain cost and how many cost drivers are necessary.

4.09 **Frontier model estimations:** To determine the relevant DEA (and depending on data availability SFA) models, they must be estimated, evaluated and tested on full-scale data sets. The starting point is the cost drivers derived from the model specification stage, but the role and significance of these cost drivers must be examined in the frontier models, and alternative specifications derived from using alternative substitutes for the cost drivers must be investigated, taking into account the outlier-detecting mechanisms.

4.10 **Model validation:** Extensive second-stage analyses shall be undertaken to see if any of the non-included variables should be included. The second-stage analyses are typically done using graphical inspection, non-parametric Kruskal-Wallis tests for ordinal differences and truncated Tobit regressions for cardinal variables. In addition to second stage control for possibly missing variables, it is desirable to perform extensive robustness runs to ensure that the outcome is not too sensitive to the parameters used in the aggregations.

4.11 It is worth emphasizing that model development is not a linear process but rather an iterative one. During the frontier model estimation, for example, we identified extreme observations resulting from data error not captured by the initial data cleaning. In turn this may lead to renewed data collection and data corrections. Such discoveries make it necessary to redo most steps in an iterative manner.

4.3 Activity analysis and scope

4.12 Benchmarking relies crucially on the structural comparability of the operators constituting the reference set. Differences in structure primarily result from differences in (i) assigned transport tasks, (ii) interfaces with other regulated or non-regulated providers and (iii) asset configuration. The identification of the main functions is the first action in a benchmarking context since different operators covers different functions and therefore cannot be directly compared at an aggregate level. The identification is also crucial since different regulations and usages of the performance evaluations may require different perspectives.

4.13 Just as electricity TSOs perform a range of functions from market facilitation to grid ownership, the gas TSOs demonstrate a portfolio of transport and terminal tasks, also including specific functions related to storage, LNG terminals, trading and balancing. The task here is twofold; first to make a systematic and relevant aggregation of the different activities and to map them to existing or obtainable data that could be reliably used in an international benchmarking. Second, the scope must be judged against the types of benchmarking methods and data material realistically available. E.g. if the activity (say planning) yields output for a horizon way beyond the existing data, the activity is not in the relevant scope for a short-term benchmarking.

4.14 The mission in this project is defined as the core tasks of transport and transit using the pipeline assets. More specifically, we focus on (i) services: transport to downstream exit and transit to a cross-border point, (ii) assets: a pipeline network with its control system and (iii) activities: grid planning, - financing/ownership, - construction, - maintenance, and -metering. Other elements, notably storage and LNG services/assets and system operations and market facilitation, activities are out of scope in this project. For more discussion of the definition of relevant scope, see PE2GAS (2014, Chapter 3).

4.4 Grid transmission activities

F Grid ownership

4.15 The grid owner ensures the long-term minimal cost financing of the network assets and its cash flows, including debt financing, floating bonds, equity management, general and centralized procurement policies, leasing arrangements for grid and non-grid assets, management of receivables and adequate provision for liabilities (suppliers, pensions, etc). The purely financial part of grid ownership (the cost of external capital) is not benchmarked here. To compare the financial costs for the operators, a specific analysis would be necessary to control for ownership structure, risk ratings and financial leverage with respect to national regulation. The grid owner function is evaluated through a standardized capital expenditure resulting from the original investments, corresponding to a comparable capital cost for the grid assets.

P Grid planning

4.16 The analysis, planning and drafting of gas network expansion and network installations involve the internal and /or external human and technical resources, including access to technical consultants, legal advice, communication advisors and possible interaction with European, governmental and regional agencies for preapproval granting.

4.17 Grid planning also covers the general competence acquisition by the TSO to perform system-wide coordination, in line with the IEM directive, the TEN corridors and the associated ENTSOG tasks. Consequently, costs for research, development and testing, both performed in-house and subcontracted, related to functioning of the transmission system, coordination with other grids and stakeholders are reported specified under grid planning.

C Grid construction

4.18 The grid construction activity is about implementing the plans from the grid planning once all necessary authorizations have been granted. Construction involves tendering for construction and procurement of material, interactions, monitoring and coordination of contractors or own staff performing ground preparation, disassembly of potential incumbent installations, temporary site constructions and installations, installation of equipment and infrastructure, recovery of land and material, test, certification and closure of the construction site.

4.19 In particular, all expenses related to site selection and environmental impact analyses are classified as grid construction since such expenses normally are capitalized with the asset investment.

4.20 Costs related to the expropriation of land for construction, remodeling or dismantling of grid assets, including direct legal costs for the process and costs potentially paid to claimants as consequences of legal proceedings are excluded as country-specific costs out of scope. These costs are not structurally comparable.

M Grid maintenance

4.21 The maintenance of a given grid involves the preventive and reactive service of assets, the staffing of facilities and the incremental replacement of degraded or faulty equipment. Both planned and prompted maintenance are included, as well as the direct costs of time, material and other resources to maintain the grid installations. It includes routine planned and scheduled work to maintain the equipment operating qualities to avoid failures, field assessment and reporting of actual condition of equipment, planning and reporting of work and eventual observations, supervision on equipment condition, planning of operations and data-collection/evaluation, and emergency action.

T Gas transport and metering

4.22 The transport task includes the operation of the injection, transport and delivery of natural gas through the gas transmission system, from defined injection points to connection points interfacing a client, a downstream network, a storage facility or an interconnection to another transmission network. The transport activity is enabled by the operations of compressors, valves and in-line stations. The assets utilized for transport constitute the pipeline system characterizing the TSO. The operational expenses for transport include staffing control centers, inspections, safety and related activities. The volume of energy (gas or electrical) spent in compressors for transport is also comprised in transport.

4.23 The TSO operates metering of the flow of gas in segments of the pipelines, at stations and at interconnections to other grids or terminals, including the IT-systems and administrative resources necessary for such services. SCADA and control stations are included in the transport and metering activity, both as investments and operating costs.

G Gas storage

4.24 The operation of gas storage facilities, including their maintenance and internal energy consumption, can be considered as separate service of gas storage, analogous to that of non-TSOs. The activity and all specific assets are excluded from the benchmarking.

L LNG terminals

4.25 The operation and maintenance of LNG terminals and peak-shaving plants, the interfaces with ports and other infrastructure, the administration and specific actions necessary to enable such operations are considered part of a specific service that is excluded, if at all existing.

S System Operations

4.26 Within system operations for gas transmission, ancillary services are retained as defined in 2009/73/EC and congestion management (compliant with the ENTSO-G classification). Ancillary services include all services related to access to and operation of gas networks, gas storage and LNG installations, including local balancing, blending and injection of inert gases, but exclude "facilities reserved exclusively for transmission system operators carrying out their functions", 2009/73/EC Art 2(14).

4.27 Day-to-day management of the network functionality, including personnel safety (instructions, training), equipment security including relay protection, operation security, coordination with operations management of the interconnected grids, coupling and decoupling in the network and allowances to contractors acting on the live grid are included in the transport activity T. This entails assets used or leased, own and subcontracted staff and other costs.

X Market Facilitation

4.28 The classification of ENTSO-G for market facilitation services includes capacity allocation mechanisms, congestion management, incremental capacity auctioning mechanisms, balancing and tariff structure. For the purposes of this benchmarking, the market facilitation activity is composed uniquely of direct expenses related to the contractual relations excluding transport and storage, including purchase and sales of natural gas, capacity from interfacing networks or reserves offered to clients. The activity has no eligible assets and no staff costs.

A Administrative Support

4.29 With administration, we refer all costs related to the general management of the undertaking, the support functions (legal, human resources, regulatory affairs, IT, facilities services etc.) that are not directly assigned to an activity above. Central management, including CEO, Board of directors and equivalent is also explicitly included. All residual assets for a gas transmission system operator (e.g. office buildings, general infrastructure) could be considered as assets for Administration. However, to the extent that this entails the incorporation of land, land installations and non-grid buildings in the analysis, all of which are susceptible to be country specific investments, such elements are listed as out of scope costs and hence excluded from the benchmarking.

O Other activities

4.30 Exceptionally, a TSO may have marginal activities that are not covered by the classification above, such as external operator training, field testing for manufacturers, leasing of land and assets for non-transport use. All such revenues, costs and assets should be specified and excluded from the benchmarking.

Summary

4.31 The benchmarking scope includes the main transport function for a gas transmission operator, the assets and costs necessary to maintain a going concern. Whenever relevant, the costs may be standardized to assure comparability, e.g. for labor and energy purchased.

4.32 Activities that are not commonly performed (gas storage and LNG terminals) are excluded.

4.33 Costs in system operations and market facilitation that result from country-specific, time-specific or structurally incomparable processes (ancillary services, capacity reservation contracts, reserves etc) are also excluded from the benchmarking.

4.5 Cost definitions and standardization

4.34 Benchmarking models can be grouped into two alternative designs with an effect on the scope of the benchmarked costs:

a. A short-run maintenance model, in which the efficiency of the operator is judged-based on the operating expenditures (Opex) incurred relative to the outputs produced, which in this case would be represented by the characteristics of the network as well as the typical customer services.

b. A long-run service model, in which the efficiency of the operator is judged-based on the total cost (Totex) incurred relative to the outputs produced, which in this case would be represented by the services provided by the operator.

4.35 From the point of view of incentive provision, a Totex based approach is usually preferred. It provides incentives for the TSOs to balance Opex and Capex solutions optimally. In this study, the focus is therefore on Totex benchmarking.

4.36 The standardization of costs play a crucial role in any benchmarking study, especially, when the study is international. Below we discuss the derivations of the benchmarked operating and capital cost, leading to the final benchmarked dependent variable; the benchmarked Totex.

4.6 Benchmarked OPEX

4.37 There are various steps involved in order to derive the respective benchmarked Opex for the benchmarked functions in scope below, see Figure 4-1 below.

P **Grid planning**
C **Grid construction**
M **Grid maintenance**
T **Gas transport and metering**
A **Administrative support**

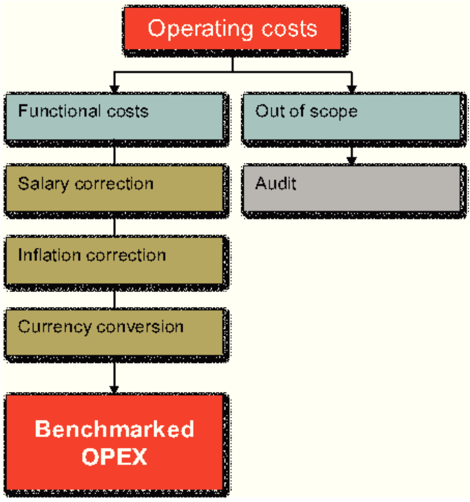

Figure 4-1 Steps in derving benchmarked OPEX.

4.38 The relevant cost items for OPEX, derived directly from the TSOs' data response (DS1, Call C) per function are added together (cf Guide C, art 6.03). This involves the cost items:

- the labor cost of the direct personnel, i.e. the people directly involved in the activity or service that is described. Note that this information is the basis for labor cost adjustments.
- the cost of services purchased externally in order to perform the activities or be able to offer the own services, and the cost of expensed goods, i.e. the cost of goods used to perform the activity or offer the service concerned. For both categories, a distinction is made between the following cases:
  - purchased services/goods that are not capitalized in the books, but that are expensed in the profit and loss accounts of the year in which they were purchased
  - purchased services/goods that are capitalized in the books, and whose cost is then split over several years through the depreciation cost. Here, a further distinction is made between:
    - depreciation cost of grid related assets
    - depreciation cost of other assets (e.g. ICT equipment, cars)
  - For goods, purchased energy is specified separately
  - For goods, leasing fees (excluding buildings and land) are separated.
- indirect costs of management and support services that are situated at the level of the function (and thus not part of the joint support that is to be reported under function A)
- other costs that are, by definition, not included in any of the categories mentioned above
- from this total will be deducted:
  - the capitalized work that was performed for own account (only applicable to the function C – Grid Construction)
  - the revenues that are generated by the sale of the products or services that form the output of the activities considered, such as the sales of working hours to other companies or income related to commercial non-benchmarked services.

4.39 Depreciation of grid related assets is excluded from this list, as this is covered by the benchmarked CAPEX.

4.40 The cost for administrative support (A) is fully allocated to the functions by cost shares of the respective functions.

4.41 In the specification used in the base run, the cost of energy is deducted from benchmarked OPEX at this step.

OPEX: Labor cost adjustments

4.42 In order to make the operating costs comparable between countries a correction for differences in national salary cost levels has been applied. Otherwise TSOs would be held responsible for cost effects, e.g. high wage level, which is not controllable by them.[1] The basis for the labor cost adjustment is the labor cost, not the data collected on FTE (full time equivalent employees) by function, since these data were less reliable.

4.43 The salary adjustment consists of two steps:

1) *Step 1 – adjustment of direct manpower costs* by increasing/decreasing the direct manpower costs of the companies using the respective salary index.
2) *Step 2 – reversal of part of salary* adjustment. Step 1 applies to a gross value, while the Opex entering the benchmarking is a net value after deducting direct revenues (for services outside the scope of the benchmark). Hence, some part of the salary adjustment has to be reversed taking into account that the share of direct manpower costs is proportionally smaller in the Opex used for benchmarking.

4.44 The EUROSTAT EU salary index in Figure 4-2 was used since no other reliable, validated index exists for the countries involved.

[1] We note that there is some simplification involved in the logic of salary cost adjustment. Had the respective operator truly had lower (or higher) salary cost then it may in practice also have chosen a different mix of production factors - e.g. operate less (or more) capital intensively. However, we do not consider this in the context of salary cost adjustments.

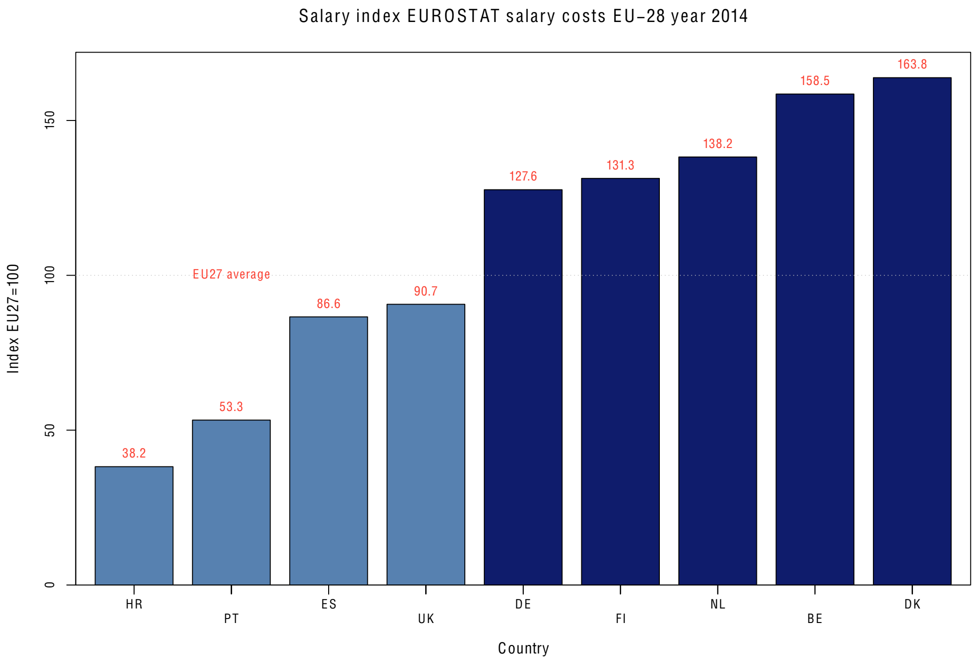

Figure 4-2 Labor cost index (EUROSTAT, EU-28, 2014)

Inflation adjustment

4.45 Opex data has been collected for 2010 (13 observations) and 2014 (9 observations). Hence, an indexation to a base year is necessary to make the costs comparable over the years. As for CAPEX, the consumer price index (CPI) is used, defining 2014 as the base year.

Currency conversion

4.46 All national currencies are converted to EUR in 2014 by the average exchange rate. The sensitivity with respect to this is tested in the robustness analysis when PPP (purchasing power parity) is used to replace exchange rates.

4.7 Benchmarked CAPEX

4.47 The capital expenditure (CAPEX) is basically a long-range real annuity sum corresponding to the function:

F Grid ownership

4.48 As accounting procedures, depreciation patterns, asset ages and capital cost calculations differ between countries and sometimes even between operators depending on their ownership structure, the CAPEX needs to be completely rebuilt from the initial investment stream and up. In addition, a real annuity must be used since the application of nominal depreciations (even standardized) would immediately introduce a bias towards late investments. The steps involved in the calculation of benchmarked CAPEX are given in Figure 4-3 below.

CAPEX: Investment stream data

4.49 The starting point is the full investment stream reported by the operators from 1970 to 2014. The investment stream is divided by type of asset as:

1) Pipelines
2) Controllers, meter stations, compressors
3) SCADA, telecom
4) Other equipment

4.50 The differentiation in investment is subject to different techno-economic life times, i.e. the standard real annuities constituting CAPEX.

4.51 The default category is "Investment: other equipment". The default techno-economic lifetime for investments in this category is the weighted life time for the assets added to the Asset Data Base in the specific year.

CAPEX: upgraded assets

4.52 Investments linked to upgrading assets in order to prolong their life will not lead to new outputs in terms of assets in Call XY but will lead to an adjustment of their annuity (longer life gives a lower annuity value). This value is calculated as the sum of the following columns where the gross investment for assets that are upgraded with respect to life time is reported:

1) Upgraded: pipelines
2) Upgraded: controllers, meter stations, compressors
3) Upgraded: SCADA, telecom
4) Upgraded: other equipment

4.53 Investments linked to upgrading assets that change asset class are counted as new investments. Thus, the original asset is replaced in the asset data with the new asset.

CAPEX: deductions

4.54 The following items are used for the correction of the investment stream prior to the calculation of the annuities:

1) Capitalized labor cost (internal and external)
2) Capitalized costs for out-of-scope assets (see Call C)
3) Capitalized costs for financial costs (construction interest)
4) Capitalized taxes, fees and levies
5) Direct subsidies, exceptional direct depreciation and internal labor as direct expense.

4.55 Capitalized labor cost is adjusted using the same labor cost index as in art. 4.42 above. The share of labor cost is set to 30% for all classes of investments.

4.56 Capitalized cost for out-of-scope assets, financial costs and taxes etc. are deducted from the gross investment stream.

4.57 Direct subsidies and exceptional depreciation are added to the gross investment stream.

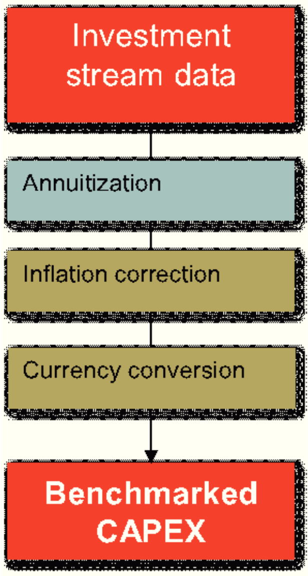

Figure 4-3 Steps in deriving benchmarked CAPEX.

CAPEX: Real annuities

4.58 Capex consists of depreciation and a return on capital. The actual investment streams are annualized using a standard annuity factor α (r,T), where r stands for a real interest rate; and T stands for the average life-time of the investments in the respective year, calculated from the shares in art 4.49. The annual investments from the investment stream data are multiplied with the annual standard annuity factor α (r,T).

4.59 The numerical values for the annuity factors are provided to each TSO in a specific file.

CAPEX: Real interest rate

4.60 The real interest rate in the e2GAS project is set to 3% for the base run. The sensitivity with respect to this parameter is subject to an analysis reported in art 5.61 below.

CAPEX: Standard life times

4.61 The standard life times per asset class are given in Table 4-1 below.

Table 4-1 Standard techno-economic life times.

Asset class	Life time (yrs)
Line	60
Pressure regulators, metering stations	30
Connection points	30
Compressors	30

CAPEX: Inflation adjustment

4.62 The current value of the past investments relative to the reference year is calculated using inflation indexes. Ideally, a sector-relevant index would capture both differences in the cost development of capital goods and services, but also the possible quality differences in standard investments. However, such index does not exist to our best knowledge. Several indexes have been collected from EUROSTAT and OECD. The only generally defined index for the full time horizon for all 22 participating grids is the conventional Consumer Price Index (CPI). All index used are available among the public parameters on the project platform.

4.63 In addition, we have evaluated further indexes (Producer Price Index PPI and Purchasing Power Parity PPP) in the sensitivity analysis in Chapter 5. Sector-specific indexes only exist for a handful of countries and require additional assumptions to be used for countries outside of their definition.

CAPEX: Currency conversion

4.64 As for OPEX, all amounts are converted to EUR values in 2014 using the average exchange rates. The exchange rates used are provided among the public parameter files.

4.8 Benchmarked TOTEX

4.65 Summing up, we obtain the benchmarked Totex as the sum of Opex and Capex

$$\underbrace{C_{ft}}_{\text{OPEX}} + \underbrace{\sum_{s=t_0}^{t} I_{fs} a\left(r, T_f\right)}_{\text{CAPEX}}$$

where C_{ft} is the total OPEX for firm *f* and time *t* after currency correction, I_{fs} is the investment stream for firm *f* and time *s* after inflation and currency correction, and $a(r,T_g)$ is the annuity factor for asset with life time T_g and real interest rate *r*.

4.9 Normalized Grid

4.66 Technically, the relevant scope is provided by an asset base consisting of:

1) Pipeline system
2) Compressor system
3) Pressure regulators and metering stations
4) Connection points

4.67 A very detailed dataset was collected for the four asset categories above. Naturally, it does not make sense just to sum the different asset together since they correspond to different dimensions, pressure levels, material choices and capacities. Likewise, the geographical nature of the pipeline system makes it ideal to capture the environmental challenges through the following factors (see Guide XY):

1) Land use

2) Subsurface features
3) Topography
4) Soil humidity

4.68 The complexity factors are based on detailed engineering construction data, out of sample, that has been averaged to form a multiplicative function for pipeline cost of the shape:

$$w_{lines}(d,L) = C(d,L)\ A_1\ A_2\,A_3\ A_4$$

where $C(d,L)$ is the base case construction cost for a pipeline of dimension class d and length L, and A_k is the complexity factor for dimension k above. The scales for the factors were defined in the asset template (Call X) and its guide, the numeric values were announced at W3. E.g., for a pipeline in diameter class B (say DN 800) the unit cost in base case is 1,234.50 k€ /km. If the pipeline passes through agricultural land (A_1=1.25) in an undulating terrain (slope up to 5%) (A_3=1.15) with soft soil (A_2=1.00), occasionally wet (A_4=1.00), the resulting unit cost would be 1,234.50 * 1.25*1.00*1.15*1.00 = 1,774.59 k€/km.

4.69 The environmental factors apply to the pipelines. For compressors, statistical data for the unit cost of compressor in various configurations yielded an average cost function through an analysis exemplified in Figure 4-4 below.

4.70 The compressor cost function is defined as a cost per station depending on total installed power in kW, P:

$$w_{compressor}(P) = 1{,}330\ P + 28{,}554\ [\text{k€}]$$

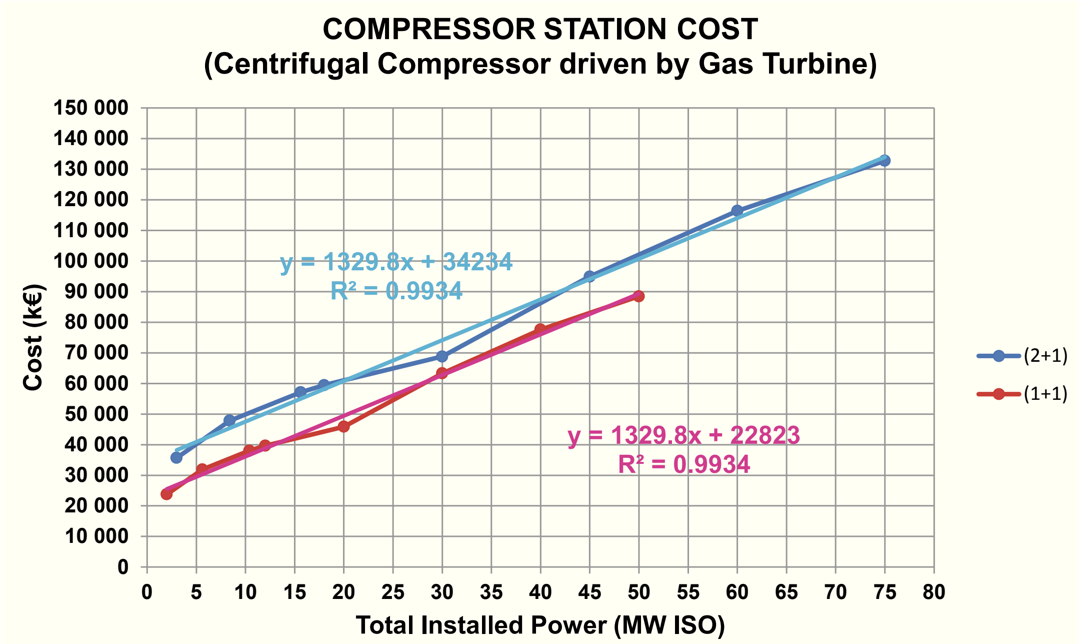

Figure 4-4 Cost function for compressor station cost, blue curve two compressors, red curve a single compressor per station.

4.71 Based on the data specification, a cost-norm for the construction costs for the standard assets above was developed, including the cost increases due to the environmental factors above. The result is an asset aggregate that we call the Normalized Grid (NG). Note that this detailed cost norm is independent of the actual costs and investments of the individual operator; it provides average costs rather than best-practice (or worst-practice) estimates. However, it is more general than a simple cost catalogue since it provides a complete system of complexity factors that explain the ratio of cost between any two type of assets, irrespective of which year, currency or context it is applied to (within reasonable bounds of course).

4.72 The size of the grid as measured by the normalized grid is naturally a key driver for Opex and Capex. A general form of the Size of Grid, often referred to as the Normalized Grid NG, can be written like this

$$\underbrace{\sum_{a} N_{fa} w_{fa}}_{\text{OPEX Grid Size}} + \underbrace{\sum_{s=t_0}^{t} \sum_{a} n_{fas} v_{fa}\, a(r,T_g)}_{\text{CAPEX Grid Size}}$$

where

N_{fa} Number of assets of type *a* that firm *f* operates at time *t*

n_{fas} Number of assets of type *a* acquired by firm *f* in period *s*

v_{fa} Weights (raw) for CAPEX, firm f asset a

w_{fa} Weights (raw) for OPEX, firm f asset a

$a(r,T_g)$ Annuity factor for asset with life time T_g and interest rate r

4.10 Model specification

4.73 Any efficiency comparison should account for differences in the outputs and the structural environment of the companies. A key challenge is to identify a set of variables:

1) that describe the tasks (the cost drivers) that most accurately and comprehensively explain the costs of the TSOs;
2) that affect costs but cannot be controlled by the firm (environmental factors); and
3) for which data can be collected consistently across all firms and with a reasonable effort.

4.74 Conceptually, it is useful to think of the benchmarking model as in Figure 4-5 below. A gas TSO transforms resources X into services Y. This transformation is affected by the environment Z. The aim of the benchmarking is to evaluate the efficiency of this transformation. The more efficient TSOs are able to provide more services using less resources and in environments that are more difficult.

4.75 The inputs X are typically thought of as Opex, Capex, or Totex. In any benchmarking study and in an international benchmarking study in particular, it requires a considerable effort to make costs comparable. We have found in previous studies that a careful cost reporting guide is of outmost importance to make sure that out-of-scope is interpreted uniformly, and that differences in depreciation practices, that taxes, land prices, labor prices etc. are neutralized. We have also found that it is useful to do process oriented models of Opex and Capex efficiency in addition to the theoretical ideal of Totex benchmarking.

4.76 The outputs Y are made of exogenous indicators for the results of the regulated task, such as typically variables related to the transportation work (volume of gas delivered etc.), capacity provision (storage volume, peak load, coverage in area etc.) and service provision (number of connections, customers etc.). Ideally, the output measures the services directly. In practice, however, outputs are often substituted by proxies constructed as functions of the assets base, like km of pipes, number of meters, number of compressors etc. One hereby runs the risk that a TSO could play the benchmarking based regulation by installing unnecessary assets. In practice, however, we have found that this is not a major risk in the early stages of the regulation and that the advantages of using such output indicators outweigh the risk. We shall therefore think more generally of the outputs as the cost drivers.

4.77 The class of structural variables Z contains parameters that may have a non-controllable influence on operating or capital costs without being differentiated as a client output. In this class we may often find indicators of geography (topology, obstacles), climate (temperature, humidity, salinity), soil (type, slope, zoning) and density (sprawl, imposed feed-in locations). One challenge with this class of parameters is that they may be difficult to validate statistically in a small data sample. They role of potential complicating factors will therefore have to be validated by other studies or in a process of individual claims from the TSOs. Another challenge is that in a small dataset, the explicit inclusion of many complicating factors will put pressure on the degrees of freedom in a statistical sense. In small data samples, therefore, we have normally found that individual adjustment of costs or weights to reflect for example difficult terrains is more useful. This is also the approach we have taken in this study. We have used elaborate engineering weight systems of the grid assets to reflect the investment and operating conditions. In this way, Z factors can to a large extent be captured by the traditional Y factors.

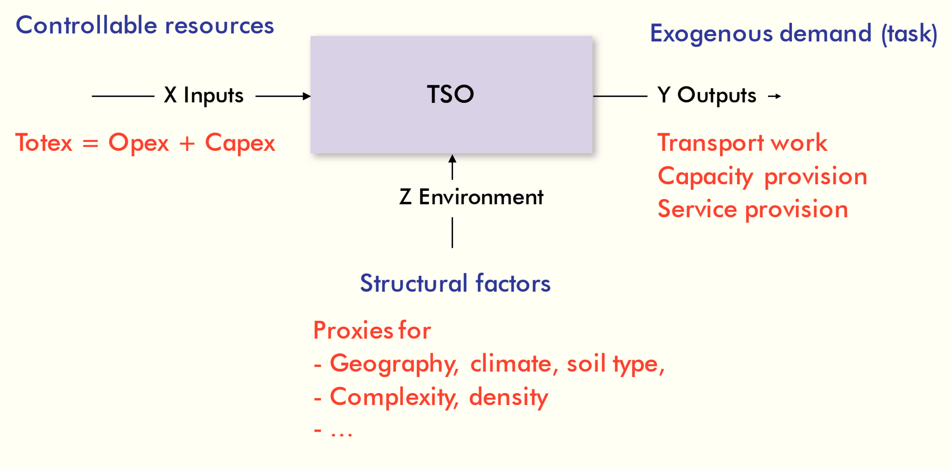

Figure 4-5 Conceptual benchmarking model

4.78 To ensure that the model specification is trustworthy, it is important to decide on some general principles as well as some specific steps. Based on our experience from other projects, we have in this project focused on the following generic criteria:

1) **Exogeneity** – Output and structural parameters should ideally be exogenous, i. e. outside the influence of the TSOs.
2) **Completeness** – The output and structural parameters should ideally cover the tasks of the TSOs under consideration as completely as reasonable.
3) **Operability** – The parameters used must be clearly defined and they should be measurable or quantifiable.
4) **Non-Redundancy** – The parameters should be reduced to the essential aspects, thus avoiding duplication and effects of statistical multi-collinearity and interdependencies that would affect the clear interpretation of results.

4.79 In reality, it is not possible to stick to these principles entirely. In particular, exogeneity most be partly dispensed with since the net assets are endogenous but also in many applications were good approximations of the exogenous conditions. To rely entirely on exogenous conditions would require a projects framework that far exceeds the present both economically and time wise.

4.80 The process of parameter selection combines engineering and statistical analysis. We have in this project used the following steps:

1) **Definition of parameter candidates**. In a first step we established a list of parameter candidates which may have an impact on the costs of TSOs. The relationships between indicators and costs must be plausible from an engineering or business process perspective.

2) **Statistical analysis of parameter candidates**. Statistical analysis was then used to test the hypotheses for cost impacts from different parameter candidates and combinations of parameter candidates. The main advantage of statistical analysis is that allows us to explore a large number of candidate parameters and to evaluate how they individually and in combinations allow us explain as much of the cost variation as possible.
3) **Plausibility checks of final parameters**. The final parameters from the statistical analysis are finally checked for plausibility. This plausibility check is based *inter alia* on engineering expertise.

4.81 These model specification steps have in addition be combined with the benchmarking analyses based on the selected parameters. It may be that some of the parameters that help explain average costs have little explanatory power in the frontier based benchmarking model and vise versa. The model specification steps have therefore been combined with careful second stage analysis to ensure that no frontier relevant have been left out.

4.11 Benchmarking methods

4.82 Econometrics has provided a portfolio of techniques to estimate the cost models for networks, illustrated in Table 4-2 below. Depending on the assumption regarding the data generating process, we divide the techniques in *deterministic* and *stochastic*, and further depending on the functional form into *parametric* and *non-parametric* techniques. These techniques are usually considered state of the art and are advocated in regulatory applications provided sufficient data is available.

Table 4-2 Model taxonomy.

	Deterministic	Stochastic
Parametric	Corrected Ordinary Least Square (COLS) Greene (1997), Lovell (1993), Aigner and Chu (1968)	Stochastic Frontier Analysis (SFA) Aigner, Lovell and Schmidt (1977), Battese and Coelli (1992), Coelli, Rao and Battese (1998)
Non-Parametric	Data Envelopment Analysis (DEA) Charnes, Cooper and Rhodes (1978), Deprins, Simar and Tulkens (1984)	Stochastic Data Envelopment Analysis (SDEA) Land, Lovell and Thore (1993), Olesen and Petersen (1995)

4.83 In a study of European gas TSOs, the number of observations is too small for a full-scale application of SFA as main instrument. We have therefore used DEA as our base estimation approach. As part of the robust check, we have additionally estimated the same model using SFA. Part of the motivation for this is also to discipline the modelling effort. In a good model specification, our experience is that the DEA and SFA approaches lead to comparable results, i.e. the average efficiencies should not deviate too much and the correlation of DEA and SFA efficiencies should be reasonably high.

4.84 Benchmarking methods like DEA and SFA are by now well established in the scientific literature as well as in regulatory applications, and we shall therefore not provide a theoretical outline of these methods. Further details are provided in e.g. Bogetoft and Otto (2011)

4.12 Data cleaning in frontier models

4.85 Data cleaning is a major effort in any regulatory application of the above methods. Likewise, the post analyses and sensitivity analyses are important to correct for any remaining noise and to evaluate the impact of other assumptions made in the estimations. We will briefly outline some important data cleaning and sensitivity analyses techniques in this section.

Outlier analyses

4.86 *Outlier analysis* consists of screening extreme observations in the model against average performance. Depending on the approach chosen (OLS, DEA, SFA), outliers may have different impact. In DEA, particular emphasis is put on the quality of observations that define best practice. The outlier analysis in DEA can use statistical methods as well as the dual formulation, where marginal substitution ratios can reveal whether an observation is likely to contain errors. In SFA, outliers may distort the estimation of the curvature and increase the magnitude of the idiosyncratic error term, thus increasing average efficiency estimates in the sample. In particular, observations that have a disproportionate impact (influence or leverage) on the sign, size and significance of estimated coefficients are reviewed using a battery of methods that is described below.

4.87 In non-parametric methods, extreme observations are such that dominate a large part of the sample directly or through convex combinations. Usually, if erroneous, they are fairly few and may be detected using direct review of multiplier weights and peeling techniques. The outliers are then systematically reviewed in all input and output dimensions to verify whether the observations are attached with errors in data. The occurrence and impact of outliers in non-parametric settings is mitigated with the enlargement of the sample size. However, in the current project, the outlier detection has prompted analyses of the underlying asset base and operating conditions to determine the reasons for the qualification as outlier (see below).

Outlier detection in DEA

4.88 In frontier analysis, the observation included in a reference or evaluation set is called a Decision Making Unit (DMU). A DMU can be an observation of (inputs,outputs) for a firm at a given time (cross section) or at other time periods (panel data). Outlier DMU may belong to a different technology by either errors in data, or unobserved quantities or qualities for inputs or outputs. The identification of DMUs to check more carefully has used in particular four approaches.

4.89 One is to identify the number of times a DMU serves as a peer unit for other DMUs, *peer counting*. If a DMU is the peer for an extreme number of units, it is either a very efficient unit – or there may be some mistakes in the reported numbers.

4.90 The other approach is to investigate the impact on average efficiency from unilateral elimination of the DMUs, *efficiency ladders*. If the elimination of one DMU leads to a significant increase in the efficiency of a sufficient number of units, there are again good reasons to check this unit more carefully.

4.91 Thirdly, we have done so-called *shell analysis* where the idea is to study the impact of groups of DMU, like the ones in the first shell, the second shell etc, cf also Agrell and Bogetoft (2002a). As the cost function is peeled this way, one shall check the shells with a significant impact on efficiency while there is less reason to continue the controls when the average efficiency is only improving slightly when a shell is eliminated.

4.92 Finally, we have used super-efficiency calculations to determine units with extreme super-efficiencies that are often associated with outliers, cf. Banker and Chang (2005). Other outlier detection methods designed with particular focus on frontier models have also been considered, for example Wilson (1993). For an overview and more analysis of how outlier methods are used in regulation, see Agrell and Niknazar (2014).

4.93 The outlier detection used in the final runs follows the German Ordinance for Incentive Regulation and the notion of DEA outliers herein (ARegV, annex 3). The invoked criteria are consistent with the method proposed and used in Agrell and Bogetoft (2007), representing a systematic and useful device to improve the reliability of regulatory benchmarking without resorting to *ad hoc* approaches. The idea is to use a dual screening device to pick out units that are doing extreme as individual observations and that are having an extreme impact on the evaluation of the remaining units. To do so, we use a super efficiency criterion similar to the Banker and Chang(2005) approach, although we let the cut-off level be determined from the empirical distribution of the super efficiency scores. In addition we use a sums-of-squares deviation indicator similar to what is commonly seen in parametric statistics.

4.94 Let I be the set of *n* TSO in the data set and *i* be a potential outlier. Also let E (k,I) be the efficiency of *k* when all TSO are used to estimate the technology and let E (k;I\i) be the efficiency when TSO *i* does not enter the estimation. We can therefore evaluate the impact on the average efficiency by

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

4.95 Large values of this as evaluated in a F (n-1,n-1) distribution, cf. Banker (1996), will be an indication that *i* is an outlier.

4.96 Using also the super-efficiency criteria of the Ordinance (ARegV), we shall classify an entity *i* as an outlier to be eliminated if

$$E(i;I\setminus i) > q(0.75) + 1.5^* (q(0.75) - q(0.25))$$

where *q*(*a*) is the *a*-fractile of the distribution of super-efficiencies, such that e.g. q(0.75) is the super-efficiency value that 75% has a value below. Hence, this criterion indicates if there are units that are having much higher super-efficiencies than the other units. If the distribution is uniform between 0 and 1 in a large sample, for example, all other units are evenly distributed between 0 and 1, a candidate unit must have a super efficiency above 0.75+1.5*(0.75-0.25)=1.5 to be an outlier.

4.13 Model validation

4.97 It is important to understand that there is no mechanical or linear procedure to develop an optimal benchmarking model. Good benchmarking models are typically developed by combining conceptual ideas, analytical results and empirical findings. This entails a process that develops interactively and which requires a good knowledge and understanding of the data available and the pros and cons of the possible estimation techniques.

4.98 The size of the data set limits the possibilities to make numerical model validations.

4.99 In a larger data set, we can compare alternative frontier models using measures of goodness of fit and by testing the if additions or deletion of cost drivers leads to significantly different results. In the DEA models, we can for example rely on the approach of asymptotic hypothesis testing. If it is possible to transform the maintained efficiency distribution into normal or half normal - say by calculating R(Fi)= Fi -1 where Fi is the efficiency of TSO i - then we can use the test statistic

$$\sum_{i=1}^{I} R(F_i)^2 / \sum_{I=1}^{i} R(F_i^*)^2$$

and evaluate this in a F(I,I) distribution with large values critical to test if the Fi estimated under a hypothesis H is reasonable given a maintained hypothesis H*, cf. e.g. Banker (1996).

4.100 In a small data set like the present one, the power of such tests is limited. We therefore only consider such measures of goodness-of-fit as indicative.

4.101 In a larger data set it is also common to make extensive second stage analyses where omitted variables are used to explain efficiency variations via a Tobit regression. Again, however, the size of the data set makes such tests less powerful, and while we have performed them and made sure that there are no such second stage issues in our proposed model, we do note that such analysis are again only indicative in a small sample.

4.102 In our experience, the general idea of robustness is more important that advanced econometric tests. We therefore propose that a model should be compared with a set of conceptually meaningful alternative specifications to document that the results are not too much affected by reasonable model changes.

4.103 In a similar manner, we suggest that the robustness of the model results should be tested by investigating how the model results changes with variations in the model parameters, including changes in the weight systems, the interest rates, the labor cost corrections etc.

The idea of robustness may lead also to the idea of using a best off approach, i.e. to make two or more conceptually sound models and/or to use two or more state-of-the-art benchmarking methods, and to let the efficiency of individual firms be determined as the maximum of the efficiencies.

5. Benchmarking results

This Chapter provides some general and average results from the benchmarking, without providing any information that may lead to the identification of individual operators and their results. The results from the robustness analysis are also included and commented.

5.1 Model specification

5.01 The choice of the model specification is as explained above a multi-criteria problem, where we must balance conceptual objectives with statistical properties taking into account also the availability of data. In this section, we illustrate some of the statistical analysis undertaken.

Correlation analysis

5.02 A good starting point is of course to look at the correlation structure. In Table 5-1 below, we provide an example. The first variable is the cost measure **dxTotex_noenergy_invadj**, i.e. the Totex without energy costs and with adjustments for labor cost differences in Opex and Capex, cf below. The other variables in Table 5-1 are a series of potential aggregate cost drivers. The interesting aspect of the correlation structure is the high correlation between the costs measure and in particular three cost drivers, namely **yNGTotex_adj_ver2**, which is a calibrated version of the raw NGTotex measure, i.e. a measure of the size of grid, the **yConnections_tot** measuring the number of connections, and the **yCapacity_max**, which is a utilization measure calculated as the maximum of injection and delivery peak capacity.

5.03 We see that the **yNGTotex_adj_ver2** is a very strong cost driver and so is **yCapacity_max**. Unfortunately, from a purely statistical point of view, they are also highly correlated internally, so they do to a large extent suggest the same cost variations. **yConnections_tot**, on the other hand has lower direct correlation with costs, but it also has lower correlation with **yNGTotex_adj_ver2** and **yCapacity_max** suggesting that it might in fact be a useful measure.

5.04 Techno-statistical constructions also lead to composite variables with good properties. During the model development, a composite variable (**pressure_difference._flowrate_adj**) that sums the flowrate-weighted pressure difference across all (owned) connection points. Models based hereon in Workshop 3 where TSO representatives expressed concern about the conceptual relevance of this variable, and suggested instead the role of capacity variables. As the explanatory power is equivalent and the techno-economic interpretation value is higher, the latter capacity variable was retained in the final model stage of the model development.

Table 5-1 Part of correlation analysis (colors indicate direction and strength)

	dxTotex_noenergy_invadj	yNGTotex_adj_ver2	yEnergy_injected_kwh	yCompressor_power_sum_mw	yPeakload_injections_h_mw	yConnections_deliv_tot	yConnections_tot	zArea.tot	yPressure_difference_adj	yPressure_diff_flowrate_adj	yCapacity_max
dxTotex_noenergy_invadj	1.00	0.91	-0.03	0.15	-0.13	0.71	0.72	-0.19	0.16	0.32	0.93
yNGTotex_adj_ver2		1.00	0.11	0.29	-0.09	0.55	0.55	-0.15	0.38	0.64	0.89
yEnergy_injected_kwh			1.00	0.88	-0.12	-0.06	-0.07	0.15	0.40	0.34	-0.07
yCompressor_power_sum_mw				1.00	-0.12	0.09	0.08	0.15	0.56	0.49	0.06
yPeakload_injections_h_mw					1.00	-0.17	-0.17	0.10	-0.14	-0.12	-0.13
yConnections_deliv_tot						1.00	1.00	-0.02	0.14	0.16	0.59
yConnections_tot							1.00	-0.03	0.13	0.15	0.61
zArea.tot								1.00	0.03	0.03	-0.27
yPressure_difference_adj									1.00	0.83	0.08
yPressure_diff_flowrate_adj										1.00	0.31
yCapacity_max											1.00

Size of model

5.05 To further analyze what seems to drive the cost differences between the gas TSOs, we can investigate how our ability to explain costs depends on the number of cost drivers. An example of this is given in Figure 5-1 below, where the adjusted R^2 of the best models of different sizes are shown. We see that the explanatory power increases relatively fast when we include 1, 2 and 3 cost drivers. Hereafter, the effects are modest and often negative. Moreover, when the larger models are investigated in details, we see that they often suffer from sign problems, i.e. additional cost drivers may have negative signs. This suggests that from a purely statistical perspective, if our aim is to explain the average cost variations, we should not use more than three cost drivers.

5.06 This is not to say that an ideal gas benchmarking model could not contain more than three cost drivers. If the sample was larger, we are likely to find more relevant cost drivers. This explains why DSO models based on large samples often contain more than three cost drivers. The fact that we only identify three significant cost drivers in this study also implies that there might be non-identified cost drivers for gas TSOs. In turn, this means that part of what we identify as inefficiency may also reflect these non-identified factors. On the other hand, in a small sample, it is also much easier to be close to best practice simply because there are fewer comparators. Thus, we have no reason to believe that the estimation of a larger model with more cost drivers on a larger sample would lead to higher general efficiencies. The two effects have compensating effects: higher fit with a larger model but also higher "competition" with more peers in the sample.

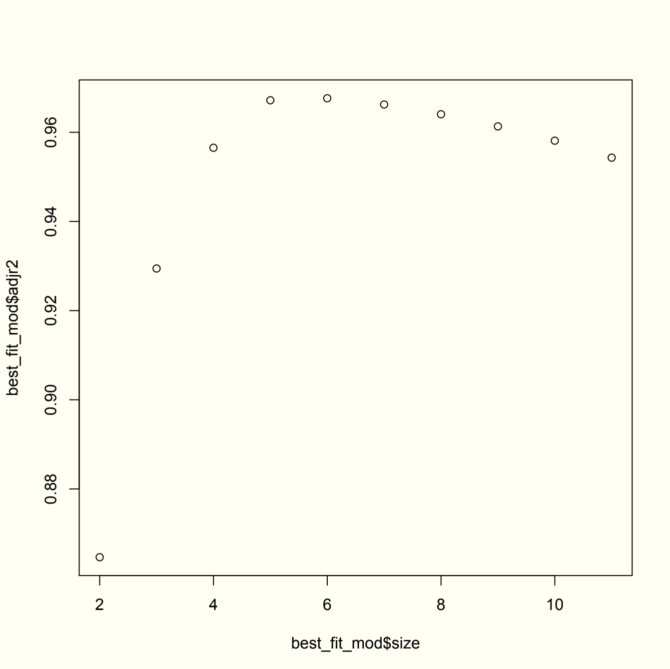

Figure 5-1 Fit (adjusted R^2) as a function of model size (number of coefficients, 2 = intercept + one variable).

5.07 If we investigate which variables are used in the best fitting models of different sizes, we get outcomes that can guide us further in understanding which cost drivers are superior from a purely statistical perspective. An illustration of such an analysis is provided in Figure 5-2 below. At the bottom row we see that the simplest model is one with an intercept and **yCapacity_max**.

5.08 Now, even the purely statistical analysis cannot be carried out in a mechanical fashion. There are at least four reasons for that.

5.09 First, it depends on which variables we choose from. In the example of the statistical analyses we have done given in Figure 5-2, we have tried combinations of are all those on the very bottom. If we choose other subsets, we might get different combinations.

5.10 Second, it depends on the estimation technique we use, and in particular if we use linear or logarithmic specifications and if we use ordinary regressions or robust regressions with outlier elimination

5.11 Third, it is not enough to identify for example three variables that provide an optimal fit. We also need to look at the intercorrelations and coefficients of these variables. Provided the model is not suffering from multicollinearity (meaning that the variables are highly correlated and explaining the same dimension), the coefficients should be positive since they are intended to be interpreted as positive cost drivers in a DEA model where a higher value for the variable should be indicative of a higher cost. This means that some of the proposals that result from of a purely statistical analysis cannot be applied.

5.12 Fourth, it is important to understand that Figure 5-2 illustrates the variable compositions for the models of different sizes that have the largest adjusted R^2 (the conventional measure of regression fit). For a given model size, there may however be other variable compositions that are only marginally worse.

5.13 In summary therefore, the example in Figure 5-2 serves not to justify the later model choice in full. Rather it serves to illustrate part of the underlying statistical analysis in short form. It shows which variables among those on the bottom are particularly interesting candidates. Variables with a high tendency to be applied have a large number of grey/black boxes in their columns.

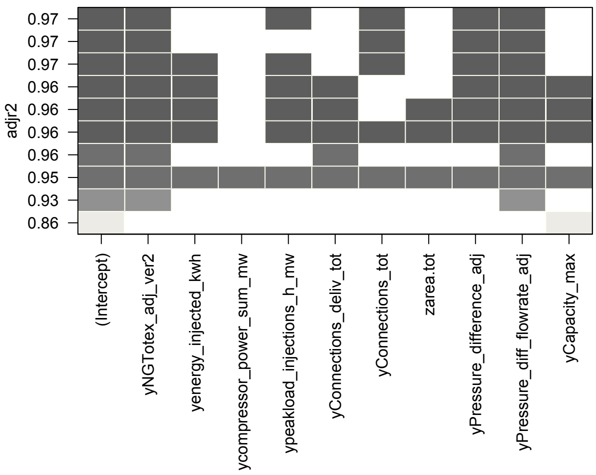

Figure 5-2 Application of cost drivers in different adjusted R^2 models (linear)

5.14 It is important also to note that the statistical analysis is only informing the frontier estimations. Good cost drivers in an econometric average-cost analysis have a tendency to work well in the frontier analysis, but there are not guarantee that they will. This is not surprising considering that the statistical regression models aim to explain average practice while the frontier models aim to explain best practice.

The base models

5.15 Based on conceptual thinking and the statistical analysis partly illustrated above, the final model specification in the e2GAS project includes three cost drivers as shown in Table 5-2 below.

Table 5-2 Model specification: Base model.

Variables	Comment
INPUT	
dxTotex_noenergy_invadj	Totex excl energy, adj for labor in opex AND investments
OUTPUT	
yNGTotex_adj_ver2	Normalized Grid calibrated to reflect opex-capex relation in sample
yConnections_tot	Total number on connections
yCapacity_max	Maximum of injection and delivery peak capacity

5.16 Input in the base model is total expenditure (Totex). It is calculated as standardized capital costs using real annuities and after correcting for inflation and currency differences plus standardized operating costs Opex excluding cost of energy, out-of-scope activities and possible Z-deductions. See the explicit formula in the Method chapter. Labor cost expenditures in Opex are adjusted to average European costs by the EUROSTAT labor cost index. Capex also includes a labor cost adjustment for 30% of the investment amount that is estimated to be activated labor cost.

The base model is using three outputs: normalized grid (weighted sum of all grid components as explained in section 4.9), total number of connection points, and the peak capacity (maximum of injection and delivery peak capacity). These parameters capture both the investment (capital expenditure) dimension through the normalized grid and the operating cost dimension through the connections and peak capacity, leading to good explanatory results for the average cost in the sample. In general, the strongest candidate in the frontier models is the normalized grid. The next strongest cost driver candidate is the connections and the weakest candidate statistically is the max capacity measure. This somewhat contrasts with the average cost models above where max capacity was indeed a very strong candidate. This illustrates the difference between an average model specification and a best practice model specification.

5.17 In the DEA literature there are alternative rules-of-thumbs as to how many inputs and outputs that can be included as a function of the number of observations available. One rule, for example, says that we one needs at least a number of observations that exceed 3 times the number of inputs plus outputs. Since we have 22 observations, this would suggest that we could have at the most 7 inputs and outputs, ie. at the most 6 cost drivers. The base model therefore satisfies this criterion. However, it is fair to say that this rule of thumb is a probably much too optimistic as to what makes sense in terms of the number of cost drivers allowed from a statistical point of view. We are not aware of studies in the academic literature that exploit all "degrees of freedom" allowed by this rule.

yConnections_tot

5.18 The connection points to the transmission grids can be of different type, see Table 5-3. Each of these connections causes certain costs of operation, metering, monitoring etc. Statistically, the sum of the connections, **yConnections_tot** is the best variable, as also explained previously.

Table 5-3 Type of connection points (Guide XY, art 5.30)

T	Description
I	Injection from upstream net/production/injection from biogas/LNG
D	Delivery to downstream network
C	Delivery to customers, direct withdrawal
N	Delivery to neighbouring networks
S	Gas storage

yCapacity_max

5.19 One of the fundamental marketable services for the transmission system is peak capacity for injection and delivery. The capacities for both types are collected by connection point as below:

5.20 **Capacity.injection.peak**: Highest measured hourly concurrent sum of capacities of all physical upstream injections at this connection point of the network operator that has occurred during the relevant year in nm^3/h ("concurrent peak load of the year").

5.21 **Capacity.delivery.peak**: Highest measured hourly concurrent sum of capacities of all physical downstream deliveries/withdrawals at this connection point of the network operator that has occurred during the relevant year in nm^3/h ("concurrent peak load of the year").

This leads to the following definition of the variable:

5.22 **yCapacity_max**: Highest measured sum of hourly concurrent capacities of all physical downstream deliveries/withdrawals and upstream injections at all connection points of the network operator that has occurred during the relevant year in nm^3/h ("concurrent peak load of the year").

yCapacity_max = max {ΣCapacity.injection.peak, Σ Capacity.delivery.peak }

5.2 Unit cost analysis

5.23 In addition to three-cost driver model specification, we have also done Unit Cost analysis where Totex is explained solely by the Normalized Grid, cf. Table 5-4 and the discussion in the Method chapter.

Table 5-4 Model specification: Unit cost analysis

Variables	Comment
INPUT	
dxTotex_noenergy_invadj	Totex excl energy, adj for labor in opex AND investments
OUTPUT	
yNGTotex_adj_ver2	Normalized Grid calibrated to reflect opex-capex relation in sample

5.24 It is possible to look at the unit cost analysis in two ways.

5.25 We may consider the unit cost analysis as an elaborate version of traditional *Key Performance Indicator KPI* analysis. Here we compare one input, Totex, with one (composite) output, Normalized Grid. Large values of the ratio indicates high costs per output. The best practice in this case corresponds to the TSO with the lowest Unit Cost.

5.26 We can also look at the unit cost analysis as a simple DEA model. It is simple by only including one output as opposed to the base model where we have 3 outputs. Taking this perspective, we can replicate the traditional KPI analysis as the efficiency in a DEA analysis assuming so-called constant returns to scale CRS. If however we take the DEA interpretation, there is no reason to restrain outself ex ante to CRS. We can look at data and see if they actually support the CRS assumption, cf below. If this is not the case, then another returns to scale assumption, like IRS, may be applied. The latter is the most cautious approach and our results below are based hereon.

Summary statistics

5.27 Summary statistics of the costs and cost drivers in the base model is shown in Table 5-5 below. (Note that range values cannot be provided for confidentiality reasons). Q1 denotes first quartile, Q3 third quartile and Q2 the median.

Table 5-5 Summary statistics of model variables (full sample, *n* = 22)

	Average	Q1	Q2 (median)	Q3
dxTotex_noenergy_invadj	150,923,933	43,230,694	76,973,677	208,010,336
yNGTotex_adj_ver2	150,923,933	30,106,878	71,569,990	153,077,200
yConnections_tot	344	71	176	389
yCapacity_max	9,056,720	1,699,385	4,482,931	10,363,300

5.28 We see that the gas TSOs in the sample varies in terms of size. The 25% largest gas TSOs are approximately 5 times larger than the 25% smallest TSOs. Also, we see that the average values exceed the median values. This reflects that the size distributions have a relatively long right tail.

5.29 To get an initial understanding also of the ability of these cost drivers to explain the variation in average costs together and individually, Table 5-6 below shows the adjusted R2 (the conventional measure of regression fit) of three ordinary regression models with 1, 2 and 3 cost drivers. We see that the adjusted R2 of a model with only **yNGTotex_adj_ver2** is 81%. Adding **yConnections_tot** as a cost driver brings us to an adjusted R2 of 87%. Finally, when we add also **yCapacity_max**, the adjusted R2 becomes 92%. This emphasize that the **yNGTotex_adj_ver2** measure is a very useful aggregate, and it also shows that the inclusion of the next two cost drivers brings important additional explanatory power to our model.

Table 5-6 Explanatory power in 1, 2 and 3 variables models, linear regressions.

Number of variables	Cost driver(s)	Adjusted R2
1	yNGTotex	81%
2	yNGTotex + yConnections_tot	87%
3	yNGTotex + yConnections_tot+yCapacity_max	92%

Outliers

5.30 The analyses of the raw data as well as the analysis of a series of models specifications, i.e. models with alternative costs drivers, suggest that one of the 22 TSOs almost always is an extreme outlier. This TSO has therefore been permanently removed from the reference set.

5.31 In addition, we have performed model specific outlier detection tests as explained in the Method chapter. This means that depending on which model we analyze, we will investigate which TSOs seem to have a too large impact on the evaluation of the others. We will explain below how many TSOs have been removed in the different runs using these criteria, i.e. the average impact and the super efficiency criteria.

Returns to scale

5.32 For all possible model specifications, we have also tested which of the returns to scale assumptions in the DEA model fit data the best: variable returns to scale (VRS), increasing returns to scale (IRS), decreasing returns to scale (DRS), or constant returns to scale (CRS). We have done so using F-tests based on a goodness-of-fit measure as explained in the Method chapter. The general finding is that the IRS assumption is the best assumption to invoke. There is in general no significant difference between the VRS and the IRS models. In some cases, one can even simplify to a CRS model, but since this happens only occasionally, we have chosen to use IRS as the maintained hypothesis. This is further supported by logarithmic regressions where the sums of the coefficients generally are slightly below 1. This means that if all cost drivers increase by a factor k, the costs only increase by k^{SUM}, where SUM is the sum of coefficients in the logarithmic regression. For an example of such a regression, see Figure 5-3.

```
Call:
lm(formula = log(dxTotex_noenergy_invadj) ~ log(yNGTotex + 1) +
    log(yConnections_tot + 1) + log(yCapacity_max + 1))

Residuals:
     Min       1Q   Median       3Q      Max
-1.22560 -0.17735 -0.05476  0.16632  1.09326

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                5.34915    1.33708   4.001 0.000839 ***
log(yNGTotex + 1)          0.31203    0.07076   4.410 0.000338 ***
log(yConnections_tot + 1)  0.34024    0.09477   3.590 0.002092 **
log(yCapacity_max + 1)     0.34101    0.11689   2.917 0.009193 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.487 on 18 degrees of freedom
Multiple R-squared:  0.9036,    Adjusted R-squared:  0.8876
F-statistic: 56.26 on 3 and 18 DF,  p-value: 2.413e-09
```

Figure 5-3 A log-log regression example for the base model

5.33 The IRS assumption means that it can be a disadvantage to be a small TSO but not to be a large TSO. This is also conceptually appealing. A TSO can be small due to the size of the country or by the service area it has to serve and there may be an element of fixed costs involved in the operation of any TSO. On the other hand, if a TSO is suffering from extra cost of being large, it is likely that a reorganization of the TSO to imitate a combination of smaller TSOs could improve cost efficiency.

Validation through SFA runs

5.34 Our primary estimation approach is DEA vith IRS and excluding outliers. When useful, we refer to this estimation approach as **d_dea_irs_ex_out**.

5.35 In addition we have also used an SFA estimation to validate the model. Due to the small sample size, we consider these simply as part of the model validation. To the extent that DEA and SFA models give results that are similar in size and has a high correlation, it provides additional evidence that the model specification is robust. When we estimate SFA models for comparisons, we generally use a log-linear specification to account for heteroskedasticity and we identify "econometric outliers" using a Cook's distance metric.

5.3 Efficiency scores

5.36 The results in this second reflect a second run made 29/05/2016 after the identification of a labeling error in the normalized grid function. Internally, there was also an analysis of a non-documented calculation in the normalized grid that should be corrected. In all, the rerun included the following corrections of data and adjustments of the calculations:

1) Correction of an error in types 1 and 3 of the landuse characteristics, in which agricultural land (type 3 in the template) was read as urban (type 1) and the corresponding weights were mixed.
2) Correction of the label for the default value for humidity to 'occasionally wet'.
3) Correction of several independent errors in the TSO templates for activated financial costs and investment subsidies (double and omitted deductions)
4) Correction of the opening balance for a TSO (non-peer) at the initiative of a NRA.
5) General adjustment in the calculation of the normalized grid investment part using standard life times per group for all TSOs.

Base model efficiencies

5.37 Summary statistics for the efficiency scores in the base e2GAS model is shown in Table 5-7 below. We see that the DEA model leads to average efficiencies of 79%, i.e. the model suggests that the gas TSOs on average can save 21% in Totex after the removal of energy costs and after corrections for labor cost differences.

5.38 In Table 5-7, we see all the quartiles of the efficiency distribution and we note that there is a longer left tail in the sense that the median is now to the right of the mean value. This is also illustrated in the graph Figure 5-4 below.

5.39 The full distribution of the efficiencies is shown as a bar chart in Figure 5-4. We note also here the relative large number of fully efficient TSOs. This is not surprising since we are using a model with 3 cost drivers on a small sample and with cautious (aggressive) outlier elimination instruments. Indeed, in the base model there are 2 DEA outliers as also show shown in Table 5-7. The decreasing number of outliers (three in the preliminary run) indicates that the data changes have lead to convergence in the dataset.

Table 5-7 Efficiency scores in base model

	DEA (IRS, ex outliers)	SFA (loglin, ex outliers)
Average	0.79	0.78
Q1	0.60	0.66
Q2 (median)	0.82	0.77
Q3	1.00	0.92
Correlation	0.74	
Outliers	2	

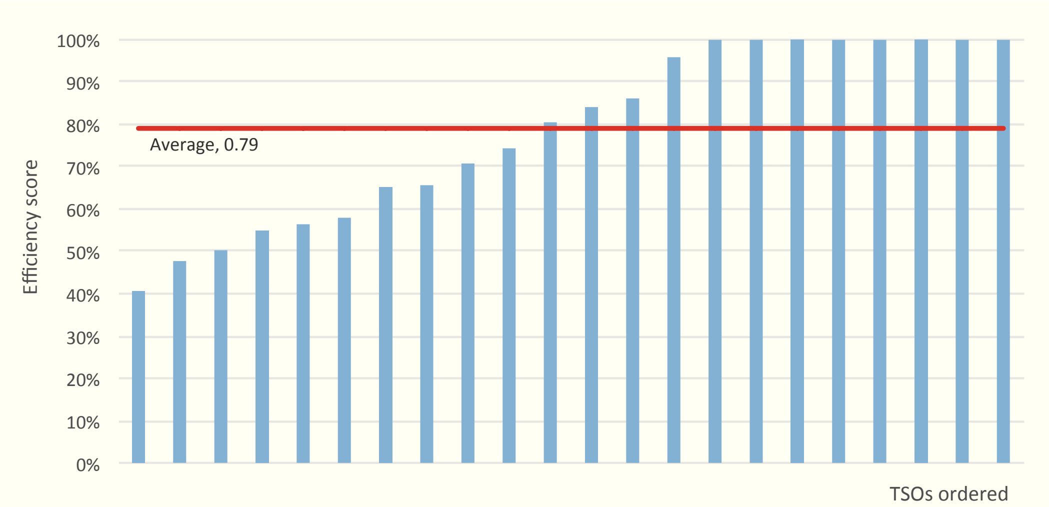

Figure 5-4 Distribution of dea_irs_ex_out scores in base model

5.40 In Table 5-7 we have also given summary statistics for a log-linear SFA estimation of the same base model. We see that the average efficiency score in this is slightly higher and that the distribution of SFA scores is more similar, see also Figure 5-5 below. 50% of all SFA scores are between 66% and 92%. We recall that the data sample is small and that the SFA results therefore shall be interpreted mainly as supplementary information. Still, it is comforting to note the high correlation (80%) between the DEA and SFA estimates. This gives a further indication that the cost drivers are well specified.

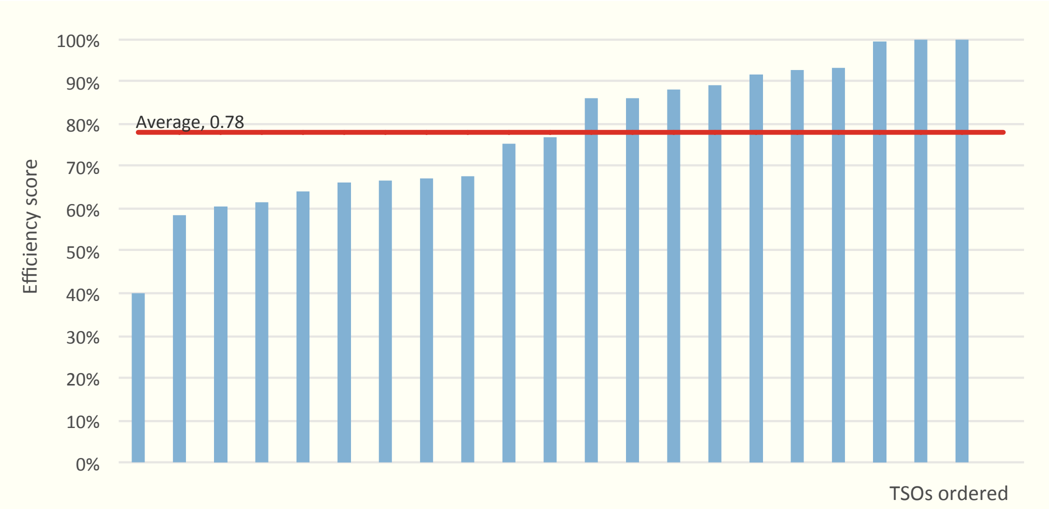

Figure 5-5 Distribution of sfa_loglin_ex_out scores in the base model

Unit Cost efficiencies

5.41 In Table 5-8 we show summary statistics for the Unit Cost efficiencies. As before we use the IRS assumption and we see that in the Unit Cost model, the average efficiency is 60%. Note that the unit cost model has only one outlier under increasing returns to scale (IRS) and no statistically identified outlier under constant returns to scale (CRS).

Table 5-8 Unit Cost efficiencies

	d_dea_irs_ex_out_all	d_dea_crs_ex_out_all
Average	0.60	0.54
Q1	0.38	0.36
Q2 (median)	0.49	0.43
Q3	0.89	0.77
Outliers	1	0

5.42 The full distribution of the efficiencies is shown as a bar chart in Figure 5-6. We note here the relative large number of TSOs with very low efficiencies.

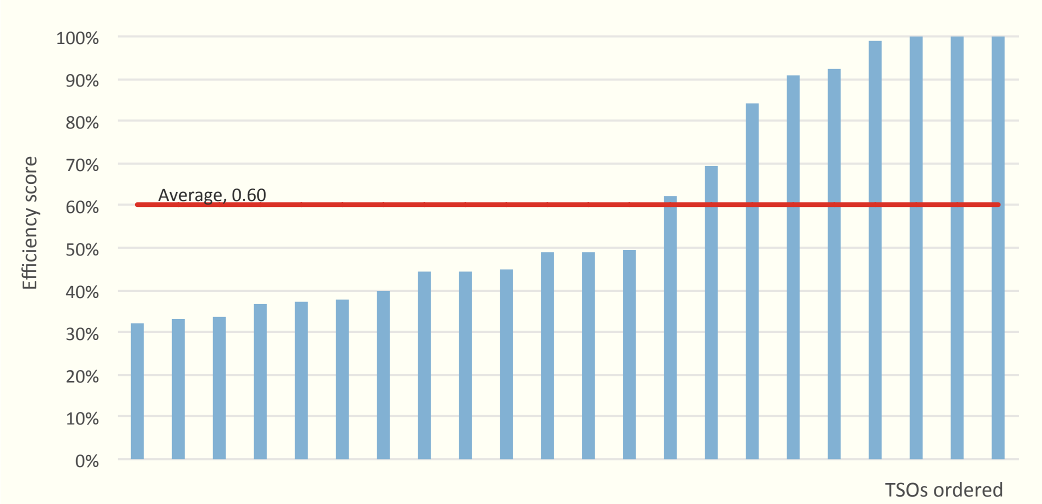

Figure 5-6 Distribution of unit cost scores

5.43 One can - as explained above - also look at UC as a ratio-based *Key Performance Indicator* KPI. In KPI studies it is common to compare entities that vary significantly in size. This corresponds to the CRS approach in 5.2 and leads to an average efficiency of 54%. The reason for the lower score under CRS than under IRS is that we allow larger TSOs to be peer units for smaller TSOs when we invoke CRS. When we invoke IRS, this is not allowed.

5.44 We see that the average UC efficiency (assuming IRS) is 79% – 60% = 19% points lower than in the base model. There are two reasons for this:

5.45 First, when we include more costs drivers, we can make fewer comparisons. In the 3 cost driver base model, we look for a combination of TSOs that has lower costs but larger values of all three cost drivers. In the single cost driver UC model, we look for (combinations of) TSOs that have lower costs but only larger values of the Normalized Grid. This means that we might compare with TSOs that have fewer connections and less max capacity in the unit costs analysis. In the unit costs analysis, we ignore the possible cost impact of the omitted cost drivers while we take them into account in the base model.

5.46 Second, we see that there are less frontier outliers in the unit costs analysis than in the base model analysis. Again, excluding fewer TSOs from the comparisons tends to make the comparisons harsher.

5.4 Opex-Capex efficiency

5.47 Using our base model, it is also possible to get an idea of the relative importance of Opex and Capex in the efficiency scores. More specifically, we can decompose the performance of TSOs according to their saving possibilities in Opex and Capex, respectively.

5.48 To do so, we have developed a two-input variant of our base model. Instead of using Totex as the single input, we now distinguish between Opex and Capex on the input side. Since we now have two instead of one input, comparisons become more restricted. When evaluating a given TSO, we look for (combinations of) other TSOs that have used less of both inputs and have higher values of all the cost drivers. Taking this approach, but otherwise maintaining the assumptions of our base model (IRS and the exclusion of outliers), we get the efficiency distribution summarized in the first columns of Table 5-9 below. Since the comparisons are now more restricted, the average efficiency is some 6% points higher than in our base model.

Table 5-9 Decomposition in Opex and Capex efficiency

	Eff. in two input model	Opex efficiency	Capex efficiency
Average	0.85	0.79	0.80
Q1	0.77	0.53	0.64
Q2 (median)	0.95	0.91	0.89
Q3	1.00	1.00	1.00

5.49 Using the two-input model, we can also calculate Opex and Capex efficiencies individually. They are provided in the last two columns of Table 5-9. The approach used is the so-called sub-vector or conditional approach in the DEA literature. We ask how much the TSOs could save purely in Opex given their Capex levels. This gives Opex efficiency. Likewise, we ask what the TSOs can save on Capex given their actual levels of Opex. This gives the Capex efficiency.

5.50 We note in Table 5-9 that the average Opex efficiency is only 1% points lower than the Capex efficiency and that the median efficiency difference is 2%-units (from 91% and 89%, respectively). In the interpretation of the Totex results of our base model, we can therefore safely conclude that the Totex inefficiency is distributed evenly on Opex and Capex in this model.

5.5 Robustness analysis

5.51 We have undertaken a series of supplementary analysis to ensure that the base model and base results are robust to variations in the underlying assumptions of this study.

Second stage analysis

5.52 First of all, we have used Tobit regressions to see if any of the omitted variables from our list of 100+ indicators should have been included in the modeling. A selection of the variables we have tested against is shown in Table 5-10 below. The result confirms the robustness for omitted variables: we do not find any significantly missing variables.

Table 5-10 Examples of variables tested in second stage analyses

Some second stage analysis	
yNGTotex	yConnections_I_tot
yNGCapex	yConnections_D_tot
yNGOpex	yConnections_C_tot
xNGCg1	yConnections_N_tot
xNGCg2	yConnections_S_tot
xNGCg3	yConnections_deliv_tot
xNGCg4	yConnections_inject_tot
xNGOg1	yConnections_tot
xNGOg2	yConnections_shared_tot
xNGOg3	yPressure_difference
xNGOg4	yPressure_diff_flowrate
yenergy_injected_h_kwh	yPipelines_H_length
yenergy_deliv_h_kwh	yPressure_difference_adj
yenergy_deliv_dsos_h_kwh	yPressure_diff_flowrate_adj
yenergy_deliv_cust_h_kwh	dxTotex_noenergy_ppi
yenergy_deliv_neigh_countr_h_kwh	dxCapex_ppi
yenergy_deliv_own_consump_h_kwh	dxOpex_noenergy_ppi
yenergy_deliv_network_loss_h_kwh	dxTotex_noenergy_invadj
ypeakload_injections_h_mw	dxCapex_invadj
ypeakload_deliveries_h_mw	dxOpex_invadj
ycompressor_h_num	yNGTotex_adj
ycompressor_power_sum_h_mw	yNGTotex_adj_g1
ycompressor_ener_used_sum_h_mwh	yNGTotex_adj_g3
yenergy_injected_h_m3	yNGTotex_adj_g4
yenergy_deliv_h_m3	yNGTotex_adj_ver2
yenergy_deliv_dsos_h_m3	yenergy_injected_kwh
yenergy_deliv_cust_h_m3	ycompressor_power_sum_mw
yenergy_deliv_neigh_countr_h_m3	z_Density_g1
yenergy_deliv_own_consump_h_m3	z_Density_g2
yenergy_deliv_network_loss_h_m3	z_Density_g3
ypeakload_injections_h_m3_per_h	z_Density_g4
ypeakload_deliveries_h_m3_per_h	yNGTotex_adj_ver2_g3_corrected

5.53 It is comforting that there are no omitted variables that are significant in Tobit regressions. On the other hand, we do acknowledge that the power of such tests is low in a small data set like we have here.

Alternative estimation approaches

5.54 In addition to the recommended estimation technique, **d_dea_irs_ex_outliers**, i.e. DEA with IRS and excluding outliers, we have done a series of estimations using alternative assumptions about returns to scale, alternative efficiency measurement directions, alternative weight restrictions etc. These alternative estimations results have served two purposes. First of all, some are necessary to test the returns to scale assumption used. Second, they give information about the robustness of the final recommended model.

Alternative Totex measures

5.55 An important series of robustness runs has been performed by using alternative definitions of the Totex measures. Most notably, we have estimated the base model using the Totex measures in Table 5-11 below.

Table 5-11 Alternative Totex measures

Abbreviation	Interpretation
dxTotex_noenergy_invadj	Totex excl energy, adj for labor in opex AND investments
dxTotex	Totex including energy costs and without adjustment for labor costs differences
dxTotex_noenergy	Totex excl energy costs
dxTotex_noenergy_ppi	Totex excl energy PPI adjusted
dxTotex_noenergy_ppp	Totex excl energy PPP adjusted for OPEX, CPI for capex
dxTotex_noenergy_noadj	Totex excl energy, no labor adj of OPEX

5.56 In the base run, consumer price index (CPI) and the exchange rates to EUR are used to create the reference base (EUR, 2014). In the run based on **dx_Totex_ppp** we use Purchasing Power Parity (PPP) to adjust all Opex, meaning that 1 EUR in Germany is no longer equal to 1 EUR in Portugal. This replaces the exchange rate, but CPI is kept for indexation of Capex.

5.57 In the base run, CPI is used as inflation adjustment index. In the run based on **dx_Totex_ppi**, a producer price index (PPI) is used for inflation adjustment where and for as long as it exists. Missing PPI data or periods are replaced by CPI.

5.58 The base run adjusts for labor cost differences both in Opex and Capex. In a variant based on **dxTotex_noenergy_noadj** we remove all such adjustments to just compare labor costs in Opex and make no corrections for labor in the investments.

5.59 The effects of such changes are shown in Table 5-12. We see that the average impact is minimal. That is, the results are very robust to variations in several of the inflation and labor cost adjustments.

Table 5-12 Impact of changing Totex definition

Totex measure	Mean d_dea_irs_ex_out_all
dxTotex_noenergy_invadj	0.79
dxTotex_noenergy	0.79
dxTotex_noenergy_ppp	0.78
dxTotex_noenergy_ppi	0.80
dxTotex_noenergy_noadj	0.79

Alternative normalized grid measures

5.60 Likewise, we have used alternative definitions of the Normalized Grid NGTotex measure, listed in Table 5-13. Some operators did not submit the environmental characteristics for pipeline-assets in the base run. In the variant based on **yNGTotex.alt**, we have therefore estimated environmental landuse data for all missing data using EUROSTAT data for landuse by country in Table 5-15. Note that the factors only replaced missing data, not reported data, although all countries are listed for confidentiality reasons. All German operators have the average landuse characteristics for Germany. The impact of these adjustments is shown in Table 5-14 below. As seen, the average impact is minimal showing the robustness of the model to the environmental characteristics in the analysis. Data for Croatia were missing, Slovenia was chosen as proxy. Default values were used for the three other environmental parameters (slope, humidity, soil subsurface).

Table 5-13 Alternative NGTotex measures

Abbreviation	Interpretation
yNGTotex_adj_ver2	Normalized Grid calibrated to reflect opex-capex relation in sample
yNGTotex.alt	Like yNGTotex_adj_ver2 but with environmentals estimated for TSOs with missing data

Table 5-14 Impact of adjusting NGTotex measure

NG Totex measure	Mean d_dea_irs_ex_out_all
yNGTotex_adj_ver2	0.79
yNGTotex.alt	0.78

Table 5-15 Complexity weight estimation using EUROSTAT landuse data (% share of surface) by country. HR data (missing) set to SI.

	Complexity weight per landuse type				
	1	1.25	1.5	1.5	
Country	Rural (forestry and other)	Agricultural	Industrial	Residential	Final factor
AT	55.8	38.2	4.0	2.0	1.13
BE	38.1	52.4	6.2	3.3	1.18
CZ	45.4	50.4	3.1	1.2	1.15
DE	41.4	51.7	5.1	1.8	1.16
DK	30.7	64.2	3.5	1.5	1.19
EE	63.6	26.9	2.6	6.9	1.11
ES	38.8	52.9	2.8	5.5	1.17
FI	80.2	7.4	2.1	10.3	1.08
FR	39.8	54.2	3.7	2.3	1.17
GR	56.6	35.4	2.9	5.1	1.13
HR	64.7	30.0	2.9	2.4	1.10
HU	32.7	61.6	3.3	2.5	1.18
IR	17.6	73.2	5.9	3.3	1.23
IT	39.0	51.4	4.5	5.1	1.18
LT	44.6	52.8	2.1	0.6	1.15
LU	39.8	52.4	6.0	1.8	1.17
LV	62.9	31.6	2.2	3.4	1.11
NL	24.1	55.0	12.2	8.6	1.24
PL	43.4	50.9	2.9	2.8	1.16
PT	54.7	37.0	3.6	4.7	1.13
SE	78.8	8.1	2.1	11.1	1.09
SI	64.7	30.0	2.9	2.4	1.10
SK	50.3	42.1	2.2	5.4	1.14
UK	27.9	65.1	3.3	3.8	1.20
EU	48.7	43.0	3.4	5.0	1.15

Alternative interest rents

5.61 Last but not least we have tested the impact of changed the interest rate. In the base run, we have used a real interest rate of 3%, but we have also estimated the model using interest rates of 2% and 4%. The results are shown in Table 5-16 below. As seen, the average

Table 5-16 Impact of interest rate on average DEA score

Interest rate	d_dea_irs_ex_out_all
Low (2%)	0.77
Base level (3%)	0.79
High (4%)	0.78

About the robustness of the e2GAS base model

5.62 The average results throughout the robustness analyses are stable between 78% and 81% cost efficiency.

5.63 Of course, the average score is likely to vary less than the scores for individual TSOs when we look across the different model variants. The average range between the highest and the lowest DEA score in the study across all the operators is 5.6%, the median difference is 3.3% and the maximum range is 26.5%. In the individualized reports, the individual TSO gets information about its sensitivity to the different model parameters.

5.64 Most of the differences can be explained by logical factors such as the relative difference between actual costs and the correction index used. Naturally, the results of the model are proportional to relevant parameters, meaning that e.g. changes in the interest rate lead to a different weight for investment efficiency.

5.65 Robustness, however, is measured as the significance of model variables with respect to technical assumptions and the rank-order consistency among the results. Overall, the model results can then be considered as robust.

6. Summary and discussion

Closing the study, we highlight the innovations, limitations and value of the project.

6.1 What has been done?

6.01 Being the first large-scale European gas transmission benchmarking project and still commissioned to be closed in less than a year, it is clear that e2GAS had to proceed in the footsteps of the proven experience from other international regulatory benchmarks, such as e^3GRID in electricity. Drawing on these insights, the data collection procedures, the activity analysis method and the project time plan were submitted early to the participants at W1 for consultation and review. In all, this introduced some extra time delays, but it also improved the precision of the data collection and lead to a change in the approach for contextual variables.

6.02 In e^3GRID and most other benchmarks, the approach is normally to collect a standardized 'green-field' data set that is then augmented with another set of environmental parameters (e.g population density, climate) as proxies for the actual cost increases observed at OPEX and CAPEX level. The advantage is in data collection time, the drawback is that the estimation of the fit for the proxies requires large data sets to be reliable. In e2GAS, the engineering cost method was used to enhance the normalized grid measure with environmental data as to increase the predictive value of the first cost driver for small data sets and to improve the fit for a range of operating conditions. Considering the results and the high explanatory fit for average cost– independent of the actual investments – this proved to be a successful route.

6.03 Two specific challenges were introduced in the project workshops; (i) the use of joint ventures for part of the operators to share costs and assets in transmission, and (ii) the allegation that the past investments had a high share of capitalized non-competitive local labor costs.

6.04 For the joint ventures, we corrected systematically all output parameters for compressors, pipelines, regulators etc as to benchmark the net share of output against the net share of costs. These data were collected already in DS1 and endorsed by the regulators.

6.05 For the labor cost issue in the investments, we adopted the most cautious policy by fully adjusting for local labor for all investments using the labor cost index. This significantly reduces some CAPEX for high-cost countries, but likewise increases the CAPEX for low-cost countries to create comparability. We believe that the base run in this sense creates an upper bound for the cost efficiency, since an increasing share of investments in gas transmission is made using European contractors at competitive rates. However, for a first estimate of the situation, this assumption seems justified although it should be challenged in subsequent studies.

6.06 Finally, the results obtained using DEA on a well defined and compact activity model without any proxies are stable against parameter changes, functional form assumptions and even correlates well in rank order with SFA. For a first benchmark with a limited data set, this result was beyond our initial expectations and promising for future work in the area. In particular, after a rerun the robustness for the DEA results are striking across the models under various assumptions.

6.07 Although the project was intended to support regulatory oversight of the operators, it is likely that the operators also appreciate the value of the study, the normalized grid system and the chance to obtain relative performance assessments. However, the imposed confidentiality constraints limit the take-away value for the operators in the sense that peer identification and learning are not enabled. Further work may be needed together with the sector to find alternative models to support these aspects of benchmarking.

6.2 What could be done differently?

6.08 As any pioneer project, e2GAS has broken new ground in regulatory benchmarking for gas transmission. As seen above, the results from the study are convincing and clearly ahead of the expected learning curve for international benchmarking. However, a number of observations can be made as to improve future projects of this type, without any claim of being exhaustive.

6.09 First, it would be beneficial to devote an entire workshop just to discuss cost allocation and definitions. Although the cost guides provide instructions and some examples, nothing is more effective than an open, prepared discussion among the stakeholders where problems are voiced, analyzed and solved together. The detection of specific reporting errors late in the process underlines the importance of a thorough understanding of the templates, guides and the calculations.

6.10 Second, the environmental factors for pipelines are a promising model extension in the project that deserves a more systematic data collection procedure. A central collection from grid coordinates, such as in Germany for DSOs, did not work for confidentiality reasons. Other models could be tested, such as the development of software for validation use by NRAs and TSOs without surrendering the sensitive geo-data. This aspect will likely require additional investments.

6.11 Third, the dynamic efficiency dimensions are as important – or even more – as the static cost efficiency. As for electricity, the systematic collection of validated data using stable definitions will enable robust calculations of the annual productivity improvement of the sector, the frontier shift. In doing so, it seems useful to collect data also for functions that currently were not benchmarked as not to block future developments to the initial scope.

6.12 Fourth, the compressor engine data were finally directly used in the normalized grid calculations, although not collected per installation. It would have been useful to collect compressor power and type of engine (gas or electrical) already in the initial data call.

6.13 Fifth, the asset data call was simplified to assets in use a single year. To enable calculation of asset ages it would be preferable to collect asset data per year, as in the electricity TSO benchmarking. Initially this costs more for the TSOs to collect, but the annual updates are simple and it enriches the model specification options.

References

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

Agrell, P. J. and P. Bogetoft (2009b) *International Benchmarking of Electricity Transmission System Operators*. Project e^{3}GRID. SUMICSID AB. 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., P. Bogetoft, H. Beaussant, J. Talarmin (2014) *Benchmarking European Gas Transmission System Operators: A Feasibility Study*. Project PE2GAS. SUMICSID. Report for CEER. https://www.researchgate.net/publication/272094391_Benchmarking_European_Gas_Transmission_System_Operators_A_Feasibility_Study

Agrell, P. J., P. Bogetoft, C. Riechmann, A. Rodgarkia-Dara, C. Zimmer (2013) *e^{3}GRID 2012 – European TSO Benchmarking Study*. Project e^{3}GRID 2012. SUMICSID, Consentec and Frontier Economics. Report for CEER. https://www.researchgate.net/publication/265592571_E3GRID2012_European_TSO_Benchmarking_Study

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

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

Aivazian V. A., J. L. Callen, M. W. L. Chan and D. C. Mountain (1987), Economies of Scale Versus Technological Change in the Natural Gas Transmission Industry, *The Review of Economics and Statistics*, 69(3), pp. 556-561.

Arthur Andersen (2001) *Report on Transco's operating costs for the 2002103 to 2006107 Price Control Period*, Final report to OFGEM.

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

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

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

Callen J. L. (1978) Production, Efficiency, and Welfar in the Natural Gas Transmission Industry, *The American Economic Review*, 68, pp. 311-323.

Carrington R., T. Coelli and E. Groom (2002) International Benchmarking for Monopoly Price Regulation: The Case of Australian Gas Distribution, *Journal of Regulation Economics*, 21:2, pp. 191-216.

Gordon, D. V., K. Gunsch and C. V. Pawluk (2003) A natural monopoly in natural gas transmission, *Energy Economics*, 25, pp. 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. (2000) Regulation, Open-Access Transportation, and Productive Efficiency. *Review of Industrial Organization*, 16, pp. 251-266.

Growitsch, C., & Stronzik, M. (2014). Ownership unbundling of natural gas transmission networks: empirical evidence. *Journal of Regulatory Economics*, pp.1-19.

Hawdon D. (2003) Efficiency, performance and regulation of the international gas industry – a bootstrap DEA approach, *Energy Policy*, 31, pp. 1167-1178.

Kim, T.-Y., J.-D. Lee, Y. H. Park and B. Kim (1999) Industrial comparisons of productivity and its determinants in the natural gas industry. *Energy Economics*, 21, pp. 273-293.

Lawrence, D., Kain, J., & Coelli, T. (2011). *Regulation of Suppliers of Gas Pipeline Services–Gas Sector Productivity*. Report for Commerce Commission, Economic Insights. http://www.economicinsights.com.au/reports/Economic_Insights_NZCC_Gas_TFP_Report_10Feb2011.pdf

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.

MacAvoy, P. W. and R. Noll (1973) Relative Prices on Regulated Transactions of the Natural Gas Pipelines, *The Bell Journal of Economics and Management Science*, 4,(1), pp. 212-234.

Meyrick (2004) *Comparative Benchmarking of Gas Networks in Australia and New Zealand*, Report prepared for Commerce Commission, Wellington

Nieswand, M., Cullmann, A., & Neumann, A. (2010). Overcoming data limitations in Nonparametric Benchmarking: Applying PCA-DEA to natural gas transmission. *Review of Network Economics*, 9(2).

Sickles, R. C. and M. L. Streitwieser (1992) Technical Inefficiency and Productive Decline in the U.S. Interstate Natural Gas Pipeline Industry under the Natural Gas Policy Act, *Journal of Productivity Analysis*, 3, pp. 119-133.

Sickles R. C. and M. L. Streitwieser (1998) An Analysis of Technology, Productivity, and Regulatory Distortion in the Interstate Natural Gas Transmission Industry: 1977-1985, *Journal of Applied Econometrics*, 13, pp. 377-395.

Appendix

SUMICSID

SUMICSID SPRL (GROUP)

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

SUMICSID AB (SWEDEN)

Burs Galtungs 776	Tel:	+46 498 48 31 96
S-623 49 Stånga, SWEDEN	Fax:	+32 27 06 53 58
sweden @ sumicsid.com		

SUMICSID APS (DENMARK)

Fru Ingesvej 19	Tel	+45 57 83 15 18
DK-4180 Sorø, DENMARK	Fax:	+45 57 83 15 18
denmark @ sumicsid.com		

http://www.sumicsid.com