

Commentary on *An approach for benchmarking European gas transmission system operators* INTERIM REPORT R1, PROJECT PE2GAS, dated 2014-11-09, SUMICSID for the Council of European Energy Regulators, CEER,

A report to National Grid Gas

Tom Weyman-Jones

Emeritus Professor of Industrial Economics

School of Business and Economics,

Loughborough University, UK

t.g.weyman-jones@lboro.ac.uk

November 24, 2014

Introduction

The interim report PE2GAS-R1 by SUMICSID

“presents a feasibility analysis for a potential pan-European benchmark of gas transmission system operators on behalf of the Council of European Energy Regulators”

In this commentary on the report PE2GAS-R1, I offer views on the design and possible implementation of the project, its feasibility and probable robustness and the nature of the model specifications and modelling techniques. At the end of the commentary I believe readers will have a clear idea of what constitutes a feasible, robust and successful benchmarking exercise. I will present a template for the phases, agenda and content of a benchmark amongst benchmarking exercises. I believe it will be apparent that the design offered in PE2GAS-R1 is seriously deficient and that it will produce only arbitrary and unhelpful results presented in a top-down non-interactive way, despite claims in PE2GAS-R1 to the contrary.

Some of the analysis in this paper draws on experience with previous SUMICSID work on benchmarking European electricity TSOs SUMICSID (2013) and my comments on that work, Weyman-Jones (2013).

Questions raised by SUMICSID

This commentary proceeds by analysing by chapter the SUMICSID PE2GAS R1 report dated 2014-11-09 and the feedback questions that are presented in the report. The commentary consists of possible responses to these chapter questions raised by SUMICSID as well as a free-standing commentary on regulatory benchmarking exercises. One objective of the commentary is to indicate the essential requirements of a benchmark for regulatory benchmarking projects and to assess whether the SUMICSID proposals meet some, all or none of these requirements. In particular, by drawing on experience of admired regulatory benchmarking projects both in the UK and in other member states of EU-28, this commentary will suggest three critical ingredients of a benchmark for regulatory benchmarking projects:

1. The choice of model specification, including the range of GTSO inputs and outputs, and its validation against available data should be developed during the benchmarking process by interactive involvement between the project consultants and the NRAs and GTSOs. The model specification should not be chosen before the project implementation nor should it be the sole property of the project consultants
2. A wide variety of modelling techniques reflecting the current state of the art in frontier efficiency and productivity analysis should be used with efficiency scores published for each model specification and each modelling technique. There should not be reliance on a single modelling technique, especially one that is relatively outdated, ignores sources of errors in the data and is rarely used as a sole strategy by NRAs.
3. The project must be transparent and open (a mantra used widely by SUMICSID), therefore there should be full involvement of NRAs and GTSOs in all stages of data checking, model specification, model estimation and evaluation of the results that are published for each specification and each modelling technique. The interaction should not be only one-way, retaining all the data in the hands of the project consultants and closing down methodological discussion, experimentation and post-project interpretation of all the results.

Finally, to incorporate a response to the shortcomings in the SUMICSID report, an alternative template for a 'good benchmarking' exercise is offered in my response to chapter 7

COMMENTS ON QUESTIONS RAISED IN PE2GAS-R1FOLLOW

Chapter 2 Gas Transmission in Europe

2.14 *The current analysis focuses at TSOs rather than regional transmission operators (RTOs) – do you agree with this limitation?*

A problem is that some are both TSOs and RTOs – how are they to be separated and where are the demarcation lines to be drawn?

2.15 *Do (sic) what extent are the European (GTSOs?) more or less similar than operators outside of EU-28?*

An unrecognised issue here is the differences among European GTSOs themselves. They differ substantially in three principal ways in addition to their geographical and population environments. These are the *ownership structures*, the *regulatory environment* and the relation to local and global *capital markets*. The overview in chapter 2 of the SUMICSID report paints a picture of a homogeneous industry structured and regulated in a similar manner in all member states and where EU-Commission directives and policy statements are the principal organizational constraints. This is far from accurate as a description of the industry.

Across the EU, there are different ownership forms from complete state ownership to complete investor owned status. The nature of the energy regulatory regimes in individual member states and the different national energy policy objectives shape the industry differently in each country, and the national regulatory authorities, which each have different incentive rules and output and service definitions, have precedence over the EU-Commission in determining the GTSOs' behaviour. Finally some of the GTSOs are much more dependent on meeting global capital market disciplines than others. All of these factors mean that the SUMICSID description of a homogenous industry differing only in geographical and population aspects is not at all realistic, and should not be used as support for a homogeneous modelling design. It also means that ownership structures, regulatory environment and capital market oversight are sources of heterogeneity that must be accounted for in determining the specification of the appropriate inputs and, particularly, outputs.

Chapter 3 Benchmarking Scope

3.52 The chapter argues that the initial scope should be limited to a subset in order to assure comparability. Do you agree with this statement?

It is not clear how the total costs are to be allocated amongst all GTSO operations. What is the detailed auditing process so that different GTSOs do not allocate these subset costs in different ways. How transparent is this auditing process? Will each GTSO be able to access published data to check the cost allocation to activities of different GTSOs?

3.53 The chapter is negative with respect to the feasibility of comparing system operations among GTSOs. Do you agree with this assessment? If not, what information should be used to achieve comparability in this regard?

A key aspect (failure) of this treatment is that it does not distinguish between variables representing exogenous demand as a determinant of costs and the variable representing accumulated assets at replacement cost. This issue – the use of **assets as outputs** – is one that continually causes regulatory problems and yet it fails to be addressed by the consultants. One cause of this confusion is the popular use of the management-speak term “cost driver” which has no economic meaning or content. This point is developed at length in the comments on chapter 6.

Chapter 4 Data Collection

Individual GTSOs are the only entities qualified to answer questions raised in this chapter.

Chapter 5 Heterogeneity

5.26 Is there any aspect (cost driving) of grid construction that you believe is not represented in the approach in this chapter?

Yes – the regulatory and market (including capital market) environment. In addition, this chapter and chapter 6 raise serious problems of specification and modelling which SUMICSID has failed to address. All of my comments on these are combined with my responses to chapter 6.

5.28 Is heterogeneity primarily an issue for CAPEX or OPEX differences in your opinion?

Both

In particular refer to the comments on chapter 2 where it is clear that SUMICSID ignores a major aspect of heterogeneity – the different ownership, regulatory and capital market exposures of different EU GTSOs leading to differences in relevant company objectives and behavioural responses.

Chapter 6 Model Approaches

A problem with this chapter is that the feedback questions posed by SUMICSID avoid confronting major issues in their whole analysis. Therefore my comments on this chapter are by necessity more free-standing.

Very few questions are raised in chapter 6 although it is fundamental to the whole of SUMICSID's approach to benchmarking. This chapter describes an approach to the benchmarking model which is virtually identical to that used in the e3grid project by the same consultants. That project was one of the most widely criticised of all European benchmarking studies and remains the subject of court proceedings in at least one member state.

6.59 The Chapter argues that frontier analysis is more suited for regulatory benchmarking than other methods, such as unit-cost analysis. Do you agree with this statement?

A variety of methods should be used – see below

6.60 DEA is advocated to be a good alternative for a frontier model, provided an activity model is developed. Do you agree with this position?

No – see below

SUMICSID proposes to pre-determine the modelling technique as well as the specification. It does so by choosing a technique DEA (data envelopment analysis) which is by now considerably out of date and which has been largely abandoned for good reason by many if not most NRAs in the EU. The principal reason is that DEA does not allow for errors in the data and ascribes all variation in measured performance to inefficiency. SUMICSID in addition propose some statistical tests of the data robustness based on DEA efficiency scores. These Banker (1996) tests have very low power in small samples and therefore are very weak at detecting possible differences in the nature of the data.

The project must investigate a wide range of techniques and a summary of these is contained below.

Analysis of chapter 6

Section 6.2 discusses model specification and validation methods. Inputs (X) are typically OPEX, CAPEX and TOTEX (p. 36). SUMICSID comments that careful cost reporting is essential “to make sure out-of-scope is interpreted uniformly and that differences ... are neutralized”. However, a key criticism of the e3grid project was the lack of transparent and publicly verifiable auditing of these cost allocations, and the same criticism is likely to be levelled at the proposed PE2GAS project.

This section also discusses the outputs (Y) and correctly identifies these with exogenous indicators of the regulated task, typically transportation work, capacity provision and service provision. SUMICSID mentions the proxy variables that might be used for outputs including the use of input assets as outputs. SUMICSID mentions that there may be GTSO gaming (strategic) behaviour in response to the measures of output, especially assets as outputs, but completely fails to recognise that it is the role of different incentive mechanisms in EU NRA regulation which causes the problem. They believe that this problem could be small, but offer no evidence for this strong statement. This argument arises again in section 6.3. A solution is simple: experiment with different model specifications for a wide range of output (Y) variables, but SUMICSID implicitly rejects this approach without further consideration.

SUMICSID also comments that with small samples (they assume on p. 51 that 15 sample points might be available) proper allowance for heterogeneity through the structural variables (Z) can only be handled by individual adjustment of sample points. In view of the comments on chapter 2 that SUMICSID has greatly underestimated the heterogeneity of the European GTSOs this is likely to be ineffective in ensuring the comparability of data across the sample.

SUMICSID next considers the issue of statistical tests and robustness and propose (p.37) to use on the sample of 15 points Banker (1996) tests based on the distribution of the efficiency scores under different assumptions. These large sample tests are all derived on the assumption that testable differences will shift the asymptotic theoretical probability density functions of the efficiency scores. However, they are known to lose statistical power in small samples (a fact which Banker himself cautions researchers to note). This means that in very small samples the tests will discover no significant differences amongst widely different specifications of the models.

SUMICSID discusses outlier detection but limit their discussion of this to the application of the DEA technique only – presuming that their later argument for using only one modelling technique i.e. DEA, can be accepted without objection. DEA outlier tests are seriously restricted in their applicability. In fact they are not designed to test the homogeneity or heterogeneity of the sample in the way that regression based econometric outlier tests do. Instead, so-called DEA outlier tests consider only whether specific frontier efficient sample points (100 per cent efficient) can be excluded from the sample without significantly altering the remaining efficiency scores. Inefficient sample points which may be entirely due to sample heterogeneity are completely ignored. SUMICSID's defence of their recommendation not to use econometric tests (p.37) is that they have found that the "general idea of robustness is more important". It is impossible to relate this vague subjective comment to the ideas of statistically robust testing. However, SUMICSID suggest a comparison with a set of conceptually meaningful alternative specifications. This would be helpful

were it not the case that subsequently SUMICSID proposes to limit its modelling to one particular specification and one modelling technique, DEA.

In section 6.3, SUMICSID discusses static and dynamic efficiency measures, identifying the three most relevant as the unit cost approach, DEA and SFA. In discussing the Unit Cost approach (p.39) it is important to note that SUMICSID's version of this idea is very different from the unit cost approach to benchmarking widely used and known amongst engineers and engineering based benchmarking. Engineering based benchmarking uses a Unit Cost approach to compare the unit cost of separate infrastructure assets and processes on an individual asset or process basis. This is not the concept of unit cost envisaged by SUMICSID, which instead is a re-use of a controversial ratio from the e3grid project. They begin by stating that the main cost drivers are typically the different assets. In fact this is a vague and loose usage of the meaningless term 'cost driver' which can be understood as follows.

Costs are defined in economic theory as:

$$\begin{aligned}
 C &\equiv \sum_{k=1}^K w_k x_k \equiv \mathbf{w}'\mathbf{x} \\
 &= \min_x(\mathbf{w}'\mathbf{x}: y = f(\mathbf{x})) \\
 &= c(y, \mathbf{w})
 \end{aligned}$$

- The first line of this expression is an identity that defines total cost (C) as expenditure on inputs (x) at exogenous input prices (w)
- The second line states that the objective is to minimise expenditure on inputs subject to meeting the demand target (y) given the current technology (f(x))
- The third line of the expression predicts that the determinants of total cost are the exogenous demand target and exogenous input prices.

Where do accumulated assets at replacement cost fit into the picture? Assets already accumulated and currently being accumulated are clearly inputs. They are the subject of input expenditure but not its determinant. Assets which are the subject of future acquisition plans are designed to meet the demand target and may be a proxy for demand if the demand forecasts are accurate and no inefficiency is present. We can conclude that, at all stages, costs are determined by input prices and demand targets. If historical data are used, accumulated assets are not a proxy for demand and are not useful for measuring cost efficiency. Some or all of these assets will be endogenous and correlated with current costs, and although some may be predetermined they are not exogenous.

SUMICSID's unit cost concept is an aggregated ratio:

$$UC = \text{Actual Cost (€)}/\text{Sum of netvolumes (€)}$$

Both numerator and denominator are aggregated expressions denominated in currency units (€). The numerator is aggregate TOTEX at constant prices and market exchange rates. The denominator is a specialised concept favoured by SUMICSID to present a single number to measure the size or complexity of grid infrastructure. This number is also in currency units at constant prices and market exchange rates. It is computed as follows (p.39)

For a given group of infrastructure assets, e.g. pipelines, LNG terminals etc., SUMICSID will construct the following number: *Netvolume* , (here I have corrected an obvious typographical error in the SUMICSID report which inserts K into the expression to be summed instead of putting it above the summation sign).

$$Netvolume = \sum_{k=1}^{k=K} N_k v_k$$

This represents the size of the grid in currency units. N_k is the number of assets of type k in this infrastructure category while v_k is unit cost to be applied to that asset type. This unit cost v_k is determined separately by SUMICSID and is applied uniformly across all GTSOs. The unit cost v_k is one of two types. For CAPEX purposes, it is

$$v_k^C = w_k a(r, T_k)$$

Here w_k is the raw unit replacement cost of asset type k and this is multiplied by the annuity factor (loan repayment equivalent) that will write off this cost at interest rate r over T_k years, the assumed life of the asset type. The parameters r and T_k are to be chosen by SUMICSID and applied uniformly across all GTSOs. For OPEX purposes, unit cost is

$$v_k^O = u_k$$

Here u_k is the raw unit OPEX cost of asset type k . In e3grid work, SUMICSID assumed that v_k^O was a fixed proportion of v_k^C .

SUMICSID then argues (p.39) that in more advanced work such as DEA, this *Netvolume* measure, i.e. a currency weighted aggregated value of input infrastructure assets can play the role of output – they use the term “cost-driver” which I have already shown to be meaningless in this context.

The academic literature, some of which is cited in SUMICSID’s R1 feasibility study, uses a much wider definition of outputs in the gas transmission industry and considers several different specifications. In Table 1 below, I indicate several different approaches that could be taken based on this academic literature and I contrast this with SUMICSID’s approach.

Table 1 Possible Specifications

Production efficiency

gas delivery = $f^P(\text{gridsize, operating inputs; heterogeneous characteristics}) + \text{errors in data}$

OPEX efficiency

OPEX = $f^O(\text{gas delivery, prices of operating inputs; gridsize, heterogeneous characteristics}) + \text{errors in data}$

TOTEX efficiency

TOTEX = $f^T(\text{gas delivery, prices of capital and operating inputs; heterogeneous characteristics}) + \text{errors in data}$

SUMICSID efficiency, DEA based without errors in data

$$TOTEX = f^S(\text{gridsize})$$

Definitions:

Operating inputs: labour, energy, materials

Grid-size: Net-volume on SUMICSID definition, or weighted asset complexity using engineering weights

Gas delivery measures: annual gas volumes, gas off-take at time of system peak demand, gas volumes and distance, gas volumes corrected for greenhouse gas emissions

Heterogeneous characteristics: geographical and geological features of grid area and terrain, population and urbanization density characteristics, applicable regulatory incentives, ownership characteristics, capital market constraints

An essential property of the benchmarking is that the consultants who are given the task should test and measure all of these and other possible specifications and present the efficiency scores associated with each different specification. These may then be submitted to the NRAs who have the regulatory authority to approve which specification is relevant to the grids within their jurisdictions. It should never be the consultants who determine the appropriate regulatory benchmarking specification for different NRAs.

SUMICSID present a very brief summary of DEA and SFA methods for benchmark modelling. They fail to emphasise the major difference between these, which are that SFA allows for errors in the data whereas DEA does not. They also fail to mention that further methods have been developed, notably StoNED, stochastic non-parametric envelopment of data, Kuosmanen et al (2014). StoNED is now the methodology of choice in distribution benchmarking in Finland and is under serious consideration for benchmarking by BundesNetzAgentur in Germany. It is also the case that most NRAs throughout Europe that use frontier benchmarking have abandoned the use of DEA as a sole methodology, and in some cases have abandoned it completely, for the reason that it is unable to handle errors in the data. It is used still but usually only as a comparator with more econometrically justified methods. It is clear that SUMICSID should consider more than one type of benchmarking methodology, but it recommends that DEA should be the predetermined methodology before the benchmarking commences.

In discussing DEA, SUMICSID advocates (p.42) the use of weight-restricted DEA methods. These are controversial since they are designed to insert more variation into a range of standard DEA efficiency scores by penalising sample points where the shadow prices implied by the model solution do not appear to fall in a pre-specified range; this range is determined by expert opinion, although in the e3grid project it was never made clear what these weight restrictions were nor how they were derived.

SUMICSID's treatment makes it clear that their preferred recommendation is the DEA methodology even though this unable to handle potential errors in the data in the same way that other methods can.

In table 2 below, I document the differences among the various frontier benchmarking methodologies that are available. It is apparent that SUMICSID is wrong to advocate a single methodology, especially one that is so deficient compared with the others in handling the major issue of errors in the data. In an appendix I describe in schematic form the differences amongst the procedures and it is demonstrated there that DEA and SFA are both subsets of the broader non-parametric stochastic approach represented by the StoNED methodology.

Table 2 summary of modelling approaches

Assumptions in different modelling approaches		Modelling approaches	
What is explained and how	Variation in observed TOTEX is modelled by:	<p>Parametric Stochastic</p> <p>stochastic frontier analysis (SFA) (regression based, with stochastic errors to allow for measurement, sampling, specification errors and inefficiency)</p>	<p>Non-parametric Deterministic</p> <p>data envelopment analysis (DEA)(mathematical programming based, without stochastic errors, all variation must be inefficiency)</p>
Can be explained by the data	<p>Differences in Cost determinants (outputs delivered): $Y_1 \dots Y_R$</p> <p>Differences in Non-discretionary characteristics $Z_1 \dots Z_L$</p>	<p>Marginal frontier effects (the technology) assumed to be the same across all TSOs</p>	<p>Non-parametric Stochastic</p> <p>Stochastic non-parametric envelopment of data (StoNED)</p> <p>(mathematical programming based, with stochastic errors to allow for measurement, sampling, specification errors and inefficiency)</p>
		<p>Marginal frontier effects (the technology) can be different across all TSOs</p>	<p>Marginal frontier effects (the technology) can be different across all TSOs</p>
		<p>Marginal frontier effects are assumed to be the same across all TSOs</p>	<p>Marginal frontier effects can be different across all TSOs or can be assumed to be the same</p>

	Differences in input prices amongst TSOs: $W_1 \dots W_K$ (Often ignored by regulators)	Marginal frontier effects are assumed to be the same across all TSOs;	Marginal frontier effects can be different across all TSOs;	Marginal frontier effects can be different across all TSOs or can be assumed to be the same
Modelled as a residual	Measurement error	Statistically modelled as an additional random variable	Not explicitly modelled but some outlier tests applied (with weak statistical power in small samples)	Statistically modelled as an additional random variable
	Model specification error	Statistically modelled as an additional random variable	Not explicitly modelled but few parametric model assumptions made, minimal model specification	Statistically modelled as an additional random variable
	Sampling error	Statistically modelled as an additional random variable	Not explicitly modelled but can be treated by re-sampling with replacement	Statistically modelled as an additional random variable
	Residual inefficiency	Modelled by a skewed distribution	Calculated as a residual and assumed to be sole source of variation in performance	Modelled by a skewed distribution
Comment	Principal disadvantages	Large sample needed	Combines errors and inefficiency into single measure of inefficiency	Computationally intensive

In summary, it can be seen that there is a wide range of methodological approaches which offer much more robust benchmarking results than is possible through reliance on DEA as a single methodology.

There are two major deficiencies in SUMICSID's suggested benchmarking approach. These are:

The decision to recommend only a single measure of output instead of creating a number of model specifications of the GTSO activity with different outputs including multiple outputs such as gas delivered, or gas multiplied by distance. Each model should be the subject of a set of experiments to determine stability and robustness. Efficiency scores for each model should be available for replication by the GTSOs

The decision to recommend only a single modelling technique or methodology, in particular one that cannot adequately handle errors in the data and that has long ago lost favour amongst EU regulators as a sole means of measuring efficiency. I have listed here a number of different methodologies and explained their relative strengths and weaknesses. All should be used in a proper benchmarking exercise.

However, one issue remains: the small sample that is the only database available to evaluate this massive and heterogeneous industry. If the size of the dataset is preventing robust estimation, the answer is for CEER to set about building a better database, not to apply an over-simplified model to inadequate data.

6.61 The last section argues that a set of comparable non-European TSOs could be used to estimate dynamic effects, e.g. through productivity improvement rate. Is this a feasible and sound approach in your (sic) view

Only if made the subject of a separate distinct modelling exercise. The suggestion can only be easily implemented if the modelling is restricted to DEA, which I have already argued is a serious error. Productivity decompositions that suffer from the same problem of failure to handle errors in the data can be developed from the DEA methodology but they would suffer from all of the same problems as the efficiency analysis that is solely reliant on DEA

Chapter 7 Process Planning

7.06 Are the requirements above all necessary and complete for the project organization?

No – see below

7.16 The section assumes that transparency is important and feasible using a combination of workshops and project platforms. Do you agree with the assumption and the assessment?

Yes I agree with the assumption. No I do not agree with the assessment that SUMICSID's proposed process will meet the transparency requirement

7.32 This section outlines a procedure with two rounds of calculations, both providing feedback to the TSOs. Is this a good approach?

The description of the feedback process is inadequate, and the process seems to fail to deliver the required procedures, see below.

7.37 To what extent is auditing a prerequisite for you to assign credibility to the results?

Essential – see below

7.38 Is there a better way of organizing the data validation of the incoming data?

Yes – see below

Analysis of Chapter 7

The process described by SUMICSID is deficient for meeting the needs of a robust and trustworthy benchmarking exercise. To explain this more clearly I have suggested, based on many years of experience of regulatory benchmarking exercises, a template for a frontier benchmark of benchmarking. It will be easily seen how far short SUMICSID's proposals are of this standard.

Benchmark template for a good benchmarking exercise

Phase 1 Propose the study

- Consult NRAs and GTSOs about their views of (i) relevant GTSO outputs in particular how they are affected by the regulatory environment (ii) relevant inputs (iii) data availability (iv) sources of heterogeneity (v) relevant model design (vi) relevant modelling techniques, in particular procedures for handling data errors
- Feedback all responses publicly (suitably anonymised if necessary)
- Determine the widest range of variables to be measured

- Series of Interactive workshops with the stakeholders: NRAs, GTSOs, – these should be open for discussions; they are not lectures by the consultants. Keep a verified record of the discussion and the stakeholders' submissions.

Phase 2 Data collection

- Collect all data on full range of variables discussed in phase 1 including all relevant GTSO outputs
- Confirm the auditing process to validate and verify the allocation of costs to categories of GTSO operations
- Publish all anonymised data to all participating NRAs, GTSOs and other interested parties
- Consult all participants on the perceived reliability of all of the data
- Interactive workshops with a verified record of the discussion and the stakeholders' submissions.

Phase 3 Modelling

- Review widest possible range of model specifications
- Consult NRAs and GTSOs on different specifications
- Review widest possible range of modelling techniques
- Consult NRAs and GTSOs on modelling techniques
- Implement wide range of models by specification and technique
- Publish results of all experiments with anonymised data so that NRAs and GTSOs are able to duplicate and verify the results by specification and technique
- Interactive workshops with a verified record of the discussion and the stakeholders' submissions

Phase 4 Feedback

- Review all CEER consultants' work with sensitivity analysis as suggested by NRAs and GTSOs
- Review all contributed GTSO versions of model specifications and techniques using the publicly available anonymised data from phase 2
- Interactive workshops with a verified record of the discussion and the stakeholders' submissions

Phase 5 Report

- Present a review and summary of all results and feedback interaction as a recommendation to the regulatory body CEER without pre-judging the regulatory outcomes
- Publish all audited and verified material from the project in anonymised form for all participants – recognise that this material may be used in judicial court proceedings

- Hand over the feedback platform to NRAs for further interaction with their GTSOs

From this template, it becomes clear that the consultants' role is to systematise and review modelling procedures in term of specification and technique in the widest and most up-to-date form. It is not the consultants' role to do any of the following:

- Retain data without publication
- Send each GTSO a single final efficiency score without elaboration
- Have a “top-down expert lectures” attitude to participants
- (Above all) attempt to set the regulatory outcome

The SUMICSID process outlined in PE2GAS-R1 compares unfavourably with this template in several ways.

- a. Phase 1's components are mostly absent, but it is not too late to repair the damage.
- b. The data collection (phase 2) process outlined in PE2GAS-R1 is one-way only with no discussion or interaction.
- c. Phase 3 in the template does not exist in the PE2GAS-R1 process since SUMICSID plans to adopt the single model specification (assets as output) and the sole technique (standard DEA) as pre-determined before any data collection or modelling and testing begins, contrary to all accepted norms of benchmarking.
- d. Phase 4 of the template does not exist in the PE2GAS-R1 process.
- e. Phase 5 is one-way only in the PE2GAS-R1 process and puts the consultants in a regulatory role by allowing them to judge the performance of GTSOs according to the pre-determined model and by implication to judge the competence of NRAs.

Chapter 8 Feasibility analysis

8.21 *Do you share this assessment? In particular, is it likely that you would retain valuable information from a benchmark performed along the lines in Chapter 6?*

No

Assuming that PE2GAS would be implemented similarly to the e3grid project, I believe that the benchmark would yield no useful information for GTSOs or NRAs. In the e3grid project, after submitting data, attending lecture workshops at which no public record was kept of any input except SUMISCID's opinions, the only substantive feedback to the individual NRAs and TSOs, apart from a copy of the general report, was a one-line email to the TSO reporting the single final efficiency score of the TSO in question, with no access to any supporting data, modelling results or other TSO scores. This was meaningless for evaluation purposes.

8.28 *Are there other risks or contingencies that should be mentioned and addressed here?*

Yes – on the basis of the PE2GAS-R1 report, the main risk is that GTSOs will be asked to devote huge efforts to a non-robust study using a minute sample with very restricted methodology, no access to data for verification and a meaningless single efficiency score at the end of it.

References

Banker, Rajiv (1996) Hypothesis tests using data envelopment analysis, *Journal of Productivity Analysis*, 7, 239-59

Kuosmanen, T., A.L. Johnson, and A. Saastamoinen (2014), "Stochastic nonparametric approach to efficiency analysis: A Unified Framework", in J. Zhu (Ed) *Handbook on DEA Vol. 2*, Springer

SUMICSID (2013) *E3GRID2012 European TSO Benchmarking Study A report for European Regulators*, SUMICSID-CONSENTEC-Frontier Economics, Frontier Economics Ltd, London

Weyman-Jones, T (2013) *The e3grid2012 project of the Council of European Energy Regulators (CEER)*, Report to National Grid

APPENDIX

I show below a schematic representation of the different methodologies for frontier efficient benchmarking

Kernel Technology,
e.g. outputs

Other Exogenous
Factors, e.g. input prices, regulatory &
operating environment

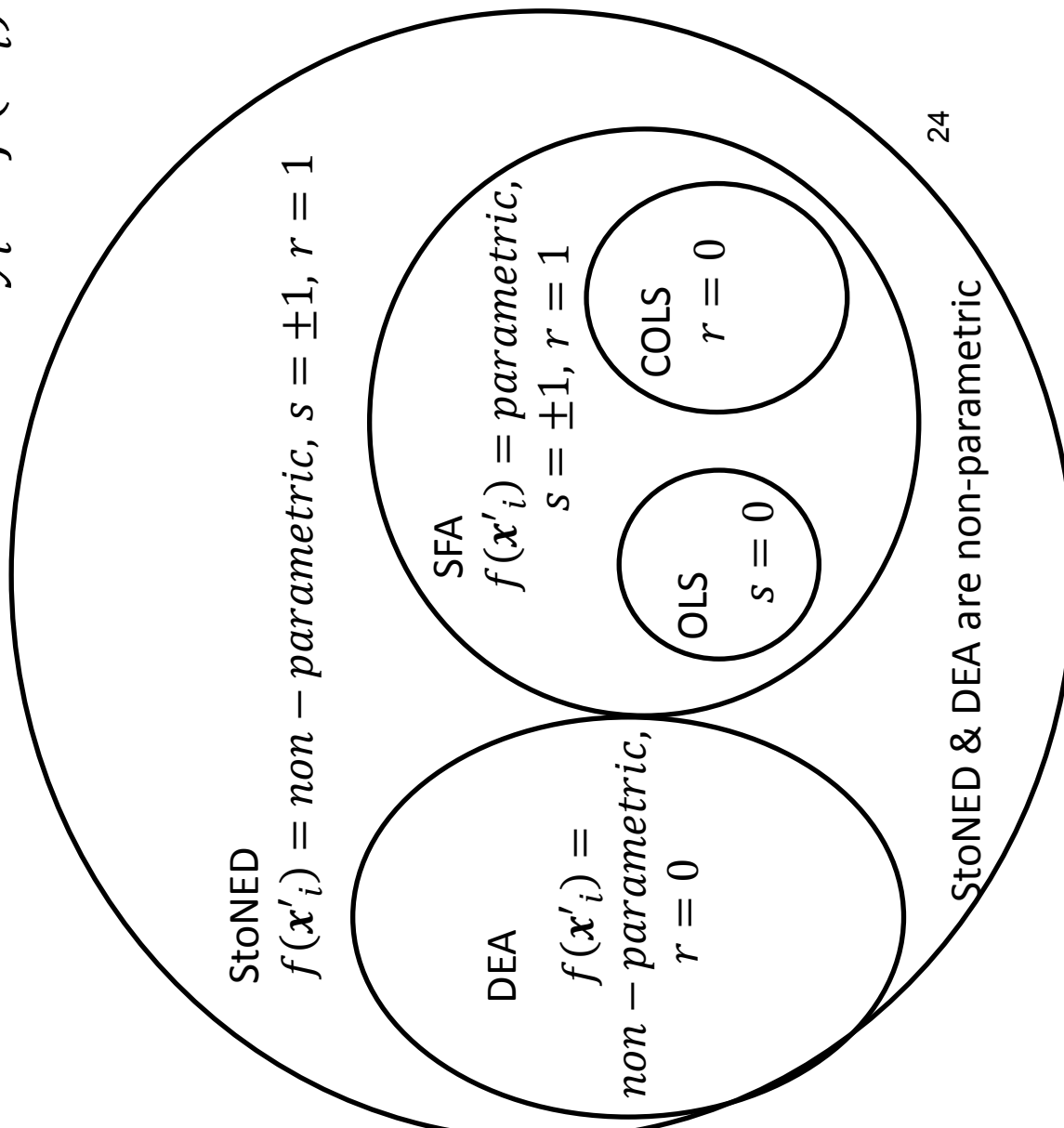
$$y_i = f(x'_i) + \sum \delta_k z_{ki} + r v_i + s u_i$$

Performance variable, e.g. cost



Idiosyncratic error
Inefficiency error component

- Sampling error
- Specification component
- Measurement



StoNED

$f(x'_i) = \text{non-parametric}, s = \pm 1, r = 1$

DEA

$f(x'_i) = \text{non-parametric}, r = 0$

SFA

$f(x'_i) = \text{parametric}, s = \pm 1, r = 1$

OLS

$s = 0$

COLS

$r = 0$

StoNED & DEA are non-parametric

$s = \begin{cases} -1, & \text{production inefficiency} \\ 0, & \text{no inefficiency} \\ +1, & \text{cost inefficiency} \end{cases}$

$r = \begin{cases} 0, & \text{deterministic frontier, no errors} \\ 1, & \text{stochastic frontier} \end{cases}$

$f(x'_i) = \begin{cases} \text{parametric} \\ \text{non-parametric} \end{cases}$

