PROJECT GERNER/AS6

Development of benchmarking models for German electricity and gas distribution

FINAL REPORT

Per Agrell
Peter Bogetoft

2007-01-01

Disclaimer

This is a first report on model specification and variable selection in the project on the development of an efficiency measurement model for electricity and gas distribution, commissioned by Bundesnetzagentur (*BNetzA*), under the supervision of the authors, professors Per AGRELL and Peter BOGETOFT for SUMICSID AB.

The modeling and calculations in the report are based on confidential data material submitted to and obtained by the Bundesnetzagentur (*BNetzA*), but it may contain errors in the original reporting, transmission or interpretation. All individual and/or sector calculations of performance are intended to identify appropriate model structures and not to make individual estimates of firm efficiency. No numeric estimates of efficiency, costs or revenues in this report constitute regulatory rulings or the official viewpoint of the Bundesnetzagentur.

The recommendations in the report have not been subject to review from the Bundesnetzagentur and express only the viewpoint of the authors, who exclusively bear the responsibility for any possible errors.

Title: Development of benchmarking models for German electricity and gas distribution
Final report, open. Project no: 730
Release date: Jan 1, 2007

Sumicsid AB
Tunbyn 502
S-85590 SUNDSVALL, SWEDEN
www.sumicsid.com

Summary

Bundesnetzagentur has embarked upon a reform process for network regulation in electricity and gas in Germany. The new incentive regulation marks a distinct break with the previous low-powered cost-recovery regulation as it relies on individual, high-powered incentives based on the demonstrated performance of the network operators with respect to their specific exogenous operating conditions. The reform is divided into three stages, gradually introducing higher importance to the yardstick mechanism that establishes a pseudo-competition on the network operations.

The performance of the network operators is assessed with three different instruments: a deterministic frontier analysis model, a stochastic frontier analysis model and an analytic cost model. This report concerns the methodology, verification and validation of the two former models, based on a cost-function approach on to detailed data for the electricity and gas distributors. It also includes the documentation of the variable specification phase, including descriptive regression results for both sectors.

In particular, three issues have been addressed empirically.

First, the structural choice of model integration is informed by a test for subadditivity in the cost function for the electricity data. The results, combined with an analysis of the task description in the IEM directive, indicate that an integrated model for the voltage levels up to high voltage is preferable. The frontier approach in this report is not recommended for the electricity transmission system operators due to their specific status and their low number.

Second, the impact of asset age on cost and efficiency estimates has been investigated to detect and possibly control for the cyclical effects of lumpy network investments, vintage of capital equipment and operating cost. The preliminary results indicate that the selected models are relatively robust to the age effects in the data, none of which are significant determinants of efficiency.

Third, the capital valuation problem has been addressed with a series of tests to determine whether potential "invisible" capital in the networks due to heterogeneous depreciation practices can be detected in the data. The results are not conclusive, but there are sufficient indications as to recommend further work on the definition and validation of the capital input variables for both sectors. A standardized asset measure will not only neutralize the capital evaluation problem, but it will also be effective to the previously mentioned investment cycle problem.

The underlying tradeoff between deterministic and stochastic models can be expressed as the relative weight assigned to model specification errors versus errors in data. On the one hand, the preliminary results presented in the report demonstrate the feasibility of economically relevant benchmarking models with

high consistency in both level and ranking order as measured by Pearson and Spearman correlations. This means that irrespective of the data and model maturity, as measured by the risk of errors in data and specification, respectively, the obtained scores and the ranking of the operators are similar. On the other hand, the differences in data quality between the sectors, in particular the variability of the output data for gas distribution, provide support for the hypothesis that there are likely residual problems in data quality. This finding, supported by the second-stage analysis and the obtained fit for the stochastic model, prompts for cautious use of the level estimates obtained by the deterministic DEA models.

The impact of structural conditions has been the subject of particular attention in this work. For electricity distribution, the cost impact of differentiated area definitions (industry, residential, dense residential, vegetation cultivated, park, rocks, waterways, infrastructure), soiltype, decentralized generation, density and number of buildings, multi-utility operations, historic energy delivered and peak load, absolute height level, height differences, average slope, calculated model network lengths have been tested.

The resulting model reflects the significant exogenous cost drivers in the three dimensions capacity provision, customer service provision and transportation work, including new measures related to the service area and decentralized generation.

For gas operations the cost impact of soiltype, quality performance (frequency and volume of leakages), density and number of buildings, multi-utility operations, historic energy delivered and peak load, absolute height level, height differences, average slope, proportion of sealed area have been tested.

The resulting base model for gas is a minimal and robust model on which further development can be done in the next phase.

In summary, the presented deterministic DEA and stochastic SFA models for electricity and gas distribution, validated on actual full-scale data, constitute an already stable ground for further development along the lines of the regulatory reform that has been initiated. Given the fast and promising increase in the quality of the data along with demonstrated commitment of Bundesnetzagentur to spare no effort to find unbiased models for the performance evaluation of the sector, the German regulation is clearly heading towards a bright, new future.

Table of contents

1. Introduction

1.01 This final report is commissioned by the German Network Agency (Bundesnetzagentur, BNetzA) within a project (AS6) on efficiency analysis for network regulation, internal project name GERNER.

1.02 The work has been closely integrated in the overall regulatory reform by section manager Dr. Joachim Muller-Kirchenbauer. Project coordinator at BNA has been Advisor Dr. Maria Krogias. Considerable data collection work and assistance has been provided by several members of the department, in particular advisors Dr. Sascha Lukac and Stefan Albrecht, Dipl-Wirt.

1.03 Project leader at SUMICSID has been Prof.dr. Per Agrell, Senior Associate. The project team has also included Prof.dr. Peter Bogetoft, Senior Associate SUMICSID, Consultant Mathias Lorenz, Dipl.Wirt-Ing., Helle Grønli, Mag and Svein Sandbakken, cand.oecon. EC Group AS.

Background

1.04 The IEM directive 2003/54/EC was implemented with the new Electricity Act (EnWG), entering into force July 13, 2005, which clearly defines the regulation as based on *yardstick competition* (§21:2) to determine *efficient costs* of *structurally comparable operators,* as well as the setting of *incentives* through performance targets (§ 21a:5) that are *reasonable, possible to attain and surpass*. The regulatory authority is the Federal Network Agency, Bundesnetzagentur (BNetzA). BNetzA performs tasks and executes powers which under the EnWG have not been assigned to the federal state regulatory authorities. The State regulatory authorities are responsible for regulating power supply companies with fewer than 100,000 customers connected to their electricity or gas networks und whose grids do not extend beyond a federal state's borders.

1.05 BNetzA's central task is to create the prerequisites for functioning competition on the upstream and downstream markets by unbundling and regulating the power and gas supply grids. The regulatory task of the BNetzA covers ensuring non-discriminatory network access and the control of the network usage rates levied by the power supply companies. BNetzA supervises anti-competitive practices and monitors the regulations concerning the unbundling of network areas and the system responsibility of the supply network operators.

1.06 With respect to the regulatory reform, the BNetzA was called upon by the EnWG (§112a:1) to submit to the Federal Government a report about the introduction of incentive-based regulation as set out in § 21a until July 1, 2006. The development process was done in consultation with the Ministry of Economics and Technology, regulatory authorities of the federal states and representatives for network utilities and customers. The official report, called BNetzA (2006) below, is the authoritative source concerning the future plans for implementation and the definition of data.

Objectives

1.07 The objectives of the project are to develop benchmarking models for electricity and gas distribution, to specify the output, input and structural variables of the model in cooperation with parallel projects on expert assessment of cost drivers in German electricity and gas distribution (AS2) and analytical cost models (AS3), to specify the technical assumptions underlying the chosen models, to validate and demonstrate the models using German data and to propose areas for further model development to assure regulatory application.

1.08 The further use of the models developed in this report is discussed in more detail in Chapter 13. The implementation of the methodology, the models or other recommendations in the report is subject to the approval of BNetz and the Ministry of Economic Affairs.

Project Process

1.09 The project proceeded in 6 phases, namely

1) Criteria definition
2) Data analysis
3) Method development
4) Model specification
5) Model implementation
6) Reporting

1.10 The activities were essentially carried out as planned in i.e. sequentially, although important feedback took place in the data analysis phase. In addition, BNA performed primary data collection, data validation and specific data collection for structural variables.

Project Deliverables

1.11	Presentation P1: (2006-02-28)	Model specification, regression results.
1.12	Presentation P2: (2006-04-11)	Model development electricity
1.13	Preliminary report PP: (2006-04-28)	Preliminary project results
1.14	Final report FP: (2006-06-15)	Final project results

Other project activities

1.15 Presentation of the project and some preliminary results at consultation meetings with industry and customers (Konsultationskreis, KK) in 2006-03-21 and 2006-04-11 and federal states' regulatory authorities (Arbeitskreis, AK) in 2006-03-14 and 2006-04-04. The results from the project were also presented at an international scientific conference on incentive regulation, organized by BNA in Bonn, 2006-04-25 – 26, and an industry seminar organized by VDEW/VDN in Bonn 2006-04-20.

Outline

1.16 This final report (FP) covers the methodology, model and variable specification of the project. Chapter 2 gives an overview to different benchmarking models, Chapter 3 provides the criteria for model structural choice and Chapter 4 covers the model specification methodology. The results are presented in Chapter 5 for the electricity model and in Chapter 6 for the gas model. The report can be seen as a technical appendix to BNetzA (2006) where all general discussion on model orientation, the use of the model in the setting of revenue targets and the data collection are covered.

2. Benchmarking models

2.1 Types of benchmarking models

2.17 In this chapter, we discuss the choice of model type and the model specification methodology used.

2.18 We first discuss the different types of benchmarking models and we briefly summarize their pros and cons. At a general level, one can distinguish between parametric and non-parametric models on the one hand and between stochastic and non-stochastic models on the other.

Parametric versus non-parametric

2.19 In the modern benchmarking literature (as opposed to traditional statistics), parametric models are characterized by being defined a priori except for a finite set of unknown parameters that are estimated from data. The parameters may refer to the relative importance of different cost drivers or to the parameters in the possibly random noise and efficiency distributions. Non-parametric models are characterized by being much less restricted a priori. Only a broad class of functions – or even production sets – is fixed a priori and data is used to estimate one of its elements. The classes are so broad as to prohibit a parameterization in terms of a limited number of parameters.

Deterministic versus stochastic models

2.20 In stochastic models, one make a priori allowance for the fact that the individual observation may be somewhat affected by random noise, and tries to identify the underlying mean structure stripped from the impact of the random elements. In non-stochastic elements, the possible noise is suppressed and any variation in data is considered to contain significant information about the performance of the unit and the shape of the technology.

Taxonomy

2.21 The two dimensions leads to a 2x2 taxonomy of methods as illustrated in table 3.1 below. A few original key references are included.

Table 2-1 Model taxonomy

	Deterministic	Stochastic
Parametric	Corrected Ordinary Least Square (COLS) Greene(1997), Lovell(1993), Aigner and Chu(1968)	Stochastic Frontier Analysis (SFA) Aigner, Lovel and Schmidt (1977), Batesee and Coelli (1992), Coelli, Rao and Battesee (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), Weyman-Jones(2001)

2.22 We emphasize that for each class of model, there exist a large set of model variants corresponding to different assumptions about the production technology, the distribution of the noise terms etc. We will discuss the key assumption below. Here, we simply stress that the non-parametric models are the most flexible in terms of the production economic properties that can be invoked while the stochastic models of course are the most flexible in terms of the assumptions one can make about data quality etc.

2.23 We presume a basic knowledge of these models here and are not going to explain them in any details. We simply recall the differences in a simple cost modeling context. The setting then is that we seek to model the costs that results when best practice is used to produce one or more outputs. We have data from a set of production units as indicated in Figure 3.1 below. Now, COLS corresponds to estimating an ordinary regression model and then making a parallel shift to make all units be above the minimal cost line. SFA on the other hand recognizes that some of the variation will be noise and only shift the line – in case of a linear mean structure – part of the way towards the COLS line. DEA estimates the technology using the so-called minimal extrapolation principle. It finds the sample production set (i.e. the set over the cost curve) containing data and satisfying a minimum of production economic regularities. Assuming free disposability and convexity, we get the DEA model illustrated in figure 3.1. Like COLS, it is located below all cost-output points, but the functional form is more flexible and the model therefore adapts closer to the data. Finally, SDEA combines the flexible structure with a realization, that some of the variations may be noisy and only requires most of the points to be enveloped.

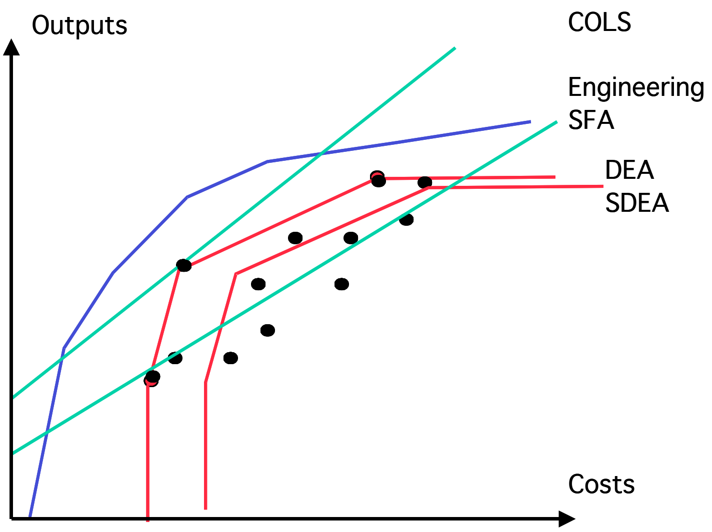

Figure 3.1 Benchmarking methods (example)

2.24 In Figure 3.1 we have included a fifth frontier, termed engineering. The idea is to base the modeling on data from engineers about best possible performance, perhaps in idealized settings. We will discuss engineering approaches later in this chapter.

Pros and cons

2.25 We will now focus on the pros and cons of these methods in general, and in particular their relative merits in a regulatory context. Again, it goes beyond the scope of this project to explain and prove the pros and cons in any details. It is our experience however that good scholars will tend to share our views below. In the cases where there may be different opinions among the specialists in these methods, we will give a few more explanations.

2.26 Some of the strengths of non-parametric methods like DEA include

- Requires no or little preference, price or priority information
- Requires no or little technological information
- Makes weak *a priori* assumptions
- Handles multiple inputs and multiple outputs
- Provides reel peers
- Identifies best practice
- Cautious or conservative evaluations (minimal extrapolation)
- Supports learning and in some cases planning and motivation
- Game theoretical foundation of the industry-regulator relation

2.27 Some of the strengths of parametric methods like SFA are

- Strong theory of significance testing (sensitivity, re-sampling, bootstrapping, asymptotic theory)
- Separates noise and efficiency
- Smoothes out some dynamic differences
- May leave lower rents when functional form known
- Creates anonymous peers, may be relevant in regulation

Basic trade-offs

2.28 As indicated, the different approaches have different advantages and disadvantages. From a regulator's viewpoint, the relative importance of these merits depends on the overall regulatory approach (cf. Agrell and Bogetoft, 2003a), i.e., the role assigned to the model among the regulatory instruments.

2.29 In our view, however, a fundamental difference from a general methodological perspective and from regulatory viewpoint is the relative importance of *flexibility* in the mean structure vs *precision* in the noise separation. The inevitable tradeoff is illustrated in Figure 2.2 below.

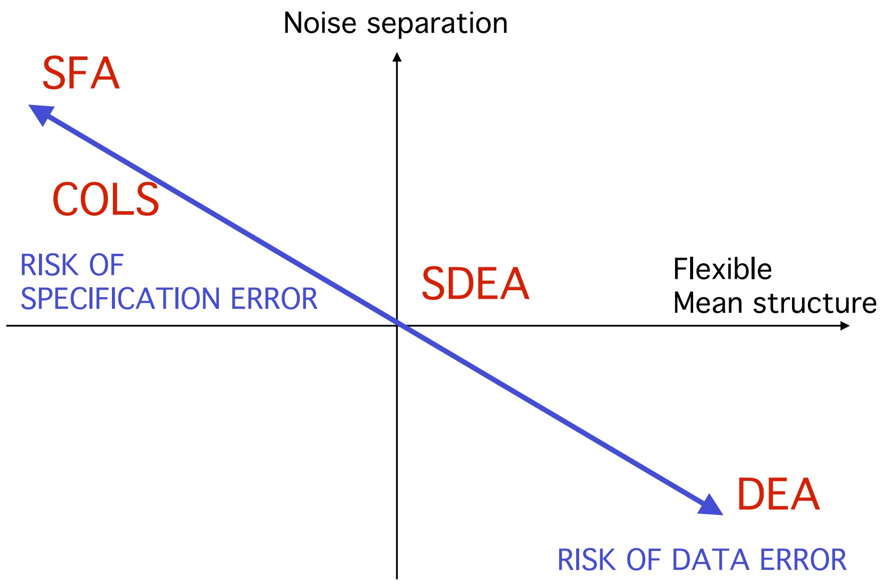

Figure 2.2 Tradeoffs in technology and noise specification, respectively.

2.30 An important property of a benchmarking approach is its ability to reflect and respect the characteristics of the industry. This requires a *flexible* model in the wide sense that its shape (or its mean structure to use statistical terms) is able to adapt to data instead of relying excessively on arbitrary text book assumptions. This is particularly important in attempts to support learning, individual motivation and coordination. It is probably less important in models aimed at evaluating system wide shifts, e.g. in non-individualized motivation and incentive provision. The non-parametric models are by nature superior in terms of flexibility.

2.31 Another important property of a benchmarking approach is its ability to cope with noisy data. A *robust* estimation method gives results that are not too sensitive to random variations in data. This is particularly important in individual benchmarking and perhaps learning – and probably less important in industry wide motivation and coordination studies. The stochastic models are particularly useful in this respect.

2.32 Ideally, then, we would like to use flexible models that are robust to random noise. The problem however is that all of this comes at a cost. The estimation task becomes bigger, the data need larger and still we cannot avoid a series of strong assumptions about the distributions of the noise terms. Coping with uncertainty requires us to dispense somewhat with flexibility and vice versa.

2.33 SDEA is essentially an extension to the DEA program where the envelopment (the linear hull) is not exactly covering the observations, but subject to an apriori distribution. The resulting efficiency estimate then becomes a confidence interval around the frontier. However, the arbitrariness of the distributional assumption of the production space that, contrary to the functional specification of the error term in SFA, has a major impact on the distribution of the efficiency estimates, makes the method poorly adapted for regulatory use.

2.34 We will continue our discussion of the pros and cons of parametric versus non-parametric and between stochastic and non-stochastic models below.

2.2 Steps in a benchmarking study

2.35 The value of benchmarking tools – as most tools – depends on how skillfully they are used. With the forthcoming of professional computer codes, the ease of efficiency analyses has increased – and hereby also the risk of un-reflected misuses of the frontier approaches. A particular problem in the business of frontier modeling is the lack of simple warning indicators and model specification tests. The risk increases when the modelers do not have string methodological training. Textbooks seldom contain detailed guidelines for proper uses of the tools they describe. A safeguard against misuses is sound application procedures. We now outline a series of relevant steps in such a procedure.

2.36 The model development includes the following steps: 1) Analysis of regulatory interface with benchmarking (preference structure and application), 2) Choice of model structure, orientation and evaluation horizon, 3) Choice of production technology (returns to scale and disposability), 4) Choice of variables and environmental proxies, 5) Choice of estimation approach (parametric or non-parametric)

2.37 The steps are illustrated in figure 4.3 below. We now comment on the individual steps.

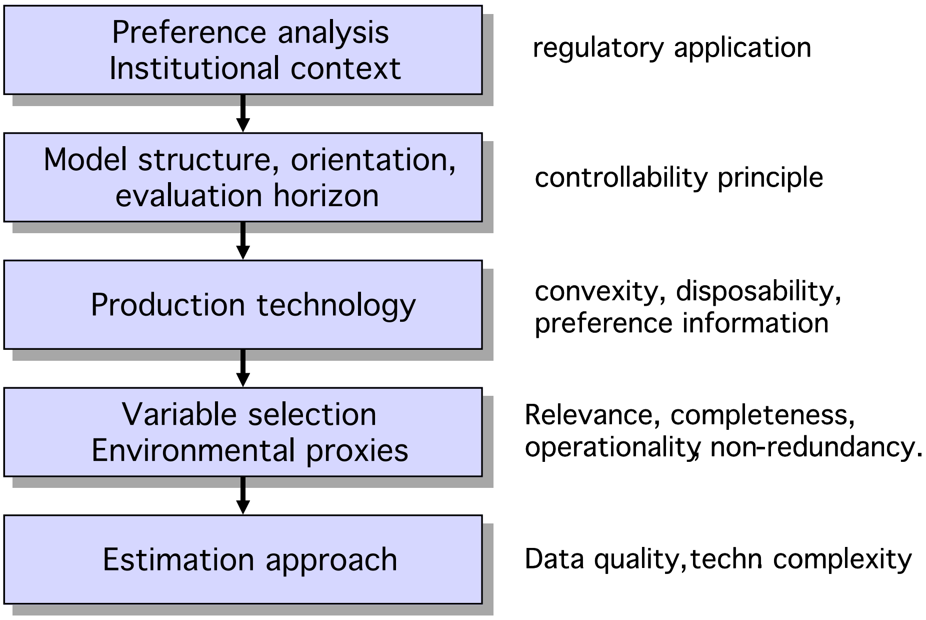

Figure 4.3 Model development steps

Regulatory interaction

2.38 The regulatory approach and the benchmarking model are closely interdependent. This is the general theme of this report. Here we simple remind that the scope, frequency and scale of the regulation regime shall ideally guide the choice of optimal benchmarking method. In repeated moderately incentivized settings with audited data collection, deterministic non-parametric methods, such as data envelopment analysis DEA, are often selected as primary benchmarking tools. In one-shot assessments of incumbent inefficiency and settings with high-powered regimes and potentially noisy data, parametric approaches, such as stochastic frontier analysis SFA or multi-output econometrics, are appropriate.

2.39 It is also important to adapt the benchmarking approach to current legislation, as well as to the long-term vision of the regulator. The methodology for this sort of *dynamic regulatory trajectory* has been subject to study in Agrell and Bogetoft (2001, 2003a) and Estache and Martimort (1998).

Model structure

2.40 The modeling proceeds to investigate the activity under the *controllability principle*. The idea is to tailor the evaluation horizon with the degree of controllability over the activity, if necessary splitting the comprehensive model in a long-run and a short-run model. In distribution regulation, this corresponds to the need to incentivize efficient infrastructure investments as well as efficient grid operation (Sweden, Finland and Norway). However, regulation may also start by assessing stranded cost due to inefficient investments and then operate a comprehensive long-run model with the adjusted capital input (Holland, UK and Denmark).

2.41 The *orientation* is normally given by the controllability principle as well. That it, the discretionary and non-discretionary variables are identified and discretionary inputs (or outputs) are reduced (or expanded). In transmission benchmarking, the focus is usually on cost minimization in an unbundled cost structure, but more advanced solutions may be relevant for integrated utilities. The recent development of directional distance functions offers a flexible approach that can take into account both the *controllability* of different resources and the preferences towards alternative directions.

2.42 The *preferences* for alternative improvement directions may reflect the regulator's trade-offs, say between economic and environmental concerns. Alternatively, the preferences may reflect intra-firm improvement strategies, e.g. as they are settled depending on the relative power of different employee groups.

Production technology

2.43 Non-parametric as well as parametric models usually convexity assumptions, disposability assumptions and return to scale assumptions.

2.44 Most models use a global *convexity* assumption. That is they assume that any weighted average of any pair of feasible productions plans are feasible as well. Although it is widely used and can be motivated in some cases, it is fair to say that it is traditionally assumed for technical convenience to simplify the duality between the production and cost space. Also, in efficiency studies it is done to increase the discriminatory power by extending the production possibility set. On the other hand, there is by now a series of models invoking less convexity assumptions, e.g. Agrell and Tind(2001), Bogetoft(1996), Bogetoft, Tama, Tind(2000), Borger and Kerstens(1996), Deprins, Simar and Tulkens(1984), Petersen (1990), Tulkens(1993). These models are theoretically appealing as they rely less on a priori assumptions and they are in general easier for the industry to accept as they rely less on the idea of mixed organizations – and of course tend to put everyone in a better light.

2.45 In terms of *disposability, i.e.* whether or not the production space is characterized by congestion constraints, rather strong assumptions are usually imposed, say strong free disposability where more inputs can always produce less outputs.

2.46 In terms of *return to scale*, the traditional models either make no assumptions or presume a – possibly local – version of the constant return to scale hypothesis. There are several common motivations to use a constant return to scale assumption like in Norway, i.e. to assume that if we adjust inputs up-ward or down-wards with a given factor, we can do the same on the output side and vice versa. One is that one can always use multiples of smaller units. This prohibits decreasing return to scale were more inputs generates small and smaller increases in the output. A second is – as with convexity – to retain sufficient discriminatory power. A third is that we would like companies to work on the constant return to scale parts of a technology to ensure that they have the right scale.

2.47 The basic assumptions of an efficiency analysis model should ideally be tested. Validation with *statistical tools* allows the analyst to settle on the right model with arguments that withstand industry challenge. There is a growing literature on statistical test but an early approach by Banker(1996) can be used to test all of the above models, i.e. the validity of a constant return to scale assumption, a free disposability assumption and a convexity assumption.

Variables and environmental proxies

2.48 The choice of variables for a given model structure involves looking for a set that is *relevant, complete, operational* and *non-redundant.*

2.49 *Relevance* means that the set of variables should reflect the industry's and the authority's comprehension of the system. The variables should be defined such that decision makers and legislators can relate to and refer to them in the regulation. In the modeling, a compromise is found in the interval between the industry's process-oriented desire to capture the details of the process and the authority's tendency to aggregate to increase comparability.

2.50 *Completeness* means that the set of variables fully capture the objectives (or regulated costs/revenues) of the decision making units. Non-modeled activities are to be explicitly acknowledged to avoid opportunistic action.

2.51 *Operationality* makes it preferable to use variables that are unambiguously defined and measurable. Qualitative indexes, subjective assessments of utility or service value are inadequate in this sense.

2.52 *Non-redundancy* is another word for Occam's razor, prescribing the least complicated means that achieves the end. Overlapping and partially redundant variables may interfere and introduce avoidable noise in the analysis.

2.53 The *model's degree of freedom* is a technical concept that relates the number of observations to the dimensionality of the model. The lower the dimensionality of the model, the higher it's discretionary ability. In the parametric, statistical model, the concept is related to the power of subsequent hypothesis tests. In the non-parametric models, heuristic upper limits on the number of variables have been proposed as well. They require that the number of observations must exceed 3*(no of inputs + no outputs) or (no inputs)(no outputs). With a fair number of distribution companies, this allows for rather flexible non-parametric models.

2.54 Regulatory benchmarking is the art of ensuring a fair treatment of all firms without leaving excessive rents. The proper use of *environmental variables* in the benchmarking models assures these two conflicting objectives. Categorical variables are related to climate, topology, density or other imposed regional heterogeneity in operating conditions. In particular mountainous regions such as Austria, Sweden and Norway are subject to such conditions, whereas models for the fairly homogenous countries in Western Europe have ignored this aspect. E.g., the final regulatory models for Sweden 2000 in Agrell and Bogetoft (2002), included four control variables (climate zone[1], transforming capacity/interconnection station, subscribed capacity in MW, minimal spanning net-length). However, we recommend that the final choice of environmental variables be made after exhaustive pilot-runs with alternative configurations and statistical tests like above. In this manner, the regulator has access to convincing evidence to various objections to the benchmarking by the regulated firms.

2.55 The choice of variables for the model need not be unique. It can in many cases be useful to have an *arsenal of complementary models.* First of all, it gives more

[1] The climate zone was suppressed in 2002 after statistical tests to coordinate with the Network Utility Model, which does not control for climate.

credibility to the results if they are verified in a series of models. Secondly, to the extent that the different specifications lead to contradicting results, one can let the benefit of the doubt protect the evaluated – like it is done in a Norwegian context with respect to alternative capital measures. The idea of picking the best result fits particularly nicely with the DEA idea of putting everyone in their best possible light. In fact, DEA results can be interpreted as the best results one can obtain using linear (or convex) cost functions, cf. Bogetoft(2000). Thirdly, using a spectrum of specification can be useful to understand the nature of the inefficiency and to decompose the differences among them. Again this has a nice theoretical basis as several types of inefficiency, e.g. technical, scale and allocative inefficiencies are defined precisely from the effects of using one or another model assumption. The use of models with different variables is probably less common than the use of different model assumptions like return to scale assumptions. Still, it is routinely done in second stage analysis of the results. Also, in a Swedish context we have found it very useful to work with cost models that have either direct operating expenditure or direct consumer charges as inputs, cf. chapter 6. By comparing the outcome of the two, one can identify if more efficient companies simply generate more profit to the owners and one can better identify possible strategic behavior

Estimation approach

2.56 In the principal choice of an estimation method, a number of issues can be used to evaluate the appropriateness of a particular method. We discussed four classes of estimation methods above, and summarized some of their strengths and weaknesses. Here we simply remind that one of the important concerns from the point of view of designing incentives based on benchmarking models is the (partial) choice between flexibility and robustness to noise.

2.3 Steps in model implementation

2.57 In applied research, it is of course not enough to have good theoretical models and good scientific procedures for how to chose models etc. It is also important to have a good implementation of whatever model is decided on. We discuss a few issues that we find particularly important in the implementation of benchmarking models in a regulatory context.

2.58 The model implementation phase includes steps like 1) Choice of software platform, 2) Implementation of the chosen model using the software, 3) Data validation and review, 4) Experimental design for pilot-runs, 5) Execution of runs, 6) Output analysis. The steps are illustrated in figure 4.4 and commented on below.

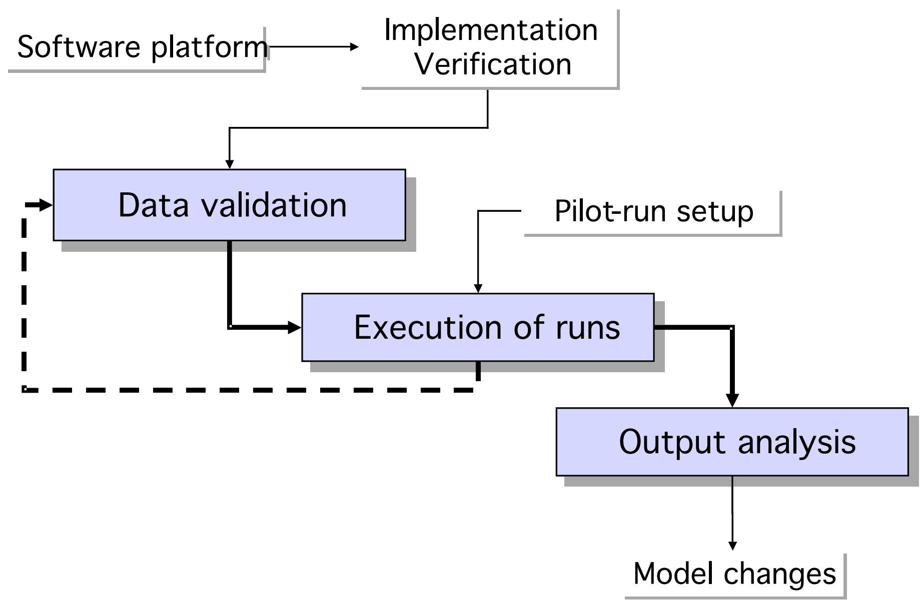

Figure 4.4 Stages in model implementation

Software platform

2.59 For the calculations in this round have been used the statistical platform R and the SFA code Frontier 4.1. The open platform has the advantage of constructing time saving customized solutions (scripts) to perform even complex estimation tasks, including the outlier detection.

Implementation and verification

2.60 In order to minimize implementation errors, it can be useful to cross-validate preliminary models with other software. For many of the dedicated software various types of numerical problems have been demonstrated.

2.4 Data validation

2.61 *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.

2.62 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 in mitigated with the enlargement of the sample size. Thus, the number of validated units in the analysis for both electricity and gas has continuously been increased during the project. More on the actual implementation of these tests is found in section 4.11 below.

2.63 In parametric methods, we are concerned with observations for which the x-value is extremely large, meaning that they have a potential leverage in influencing the shape and slope of the regression and that are off-center in the meaning that they actually exercise their leverage. Below, we briefly describe four diagnostic functions that are used in automated outlier analysis in the regression phase: Cook's distance, DFFits, DFBetas and covariance ratios. These re-normalize the residuals to have unit variance, using an overall and leave-one-out measure of the error variance respectively. Values for generalized linear models are approximations, as described in Williams (1987) (except that Cook's distances are scaled as F rather than as chi-square values). All functions are integrated in the statistical software package R.

Cook's distance

2.64 Cook's distance is a measurement of the influence of the i'th data point on all the other data points. In other words, it tells how much influence the i'th case has upon the model. The formula to find Cook's distance, Di, is the squared error term divided by the product of the number of parameters in the model and the Mean Square Error.

2.65 Using the F distribution to compare with Cook's distance, the influence that the i'th data point has on the model can be found. Values in the F distribution table can be used to express the percentage of influence the i'th data point has. A percentage of 50% or more would indicate a large influence on the model. The larger the error term implies that the Di is also larger which means it has a greater influence on the model.

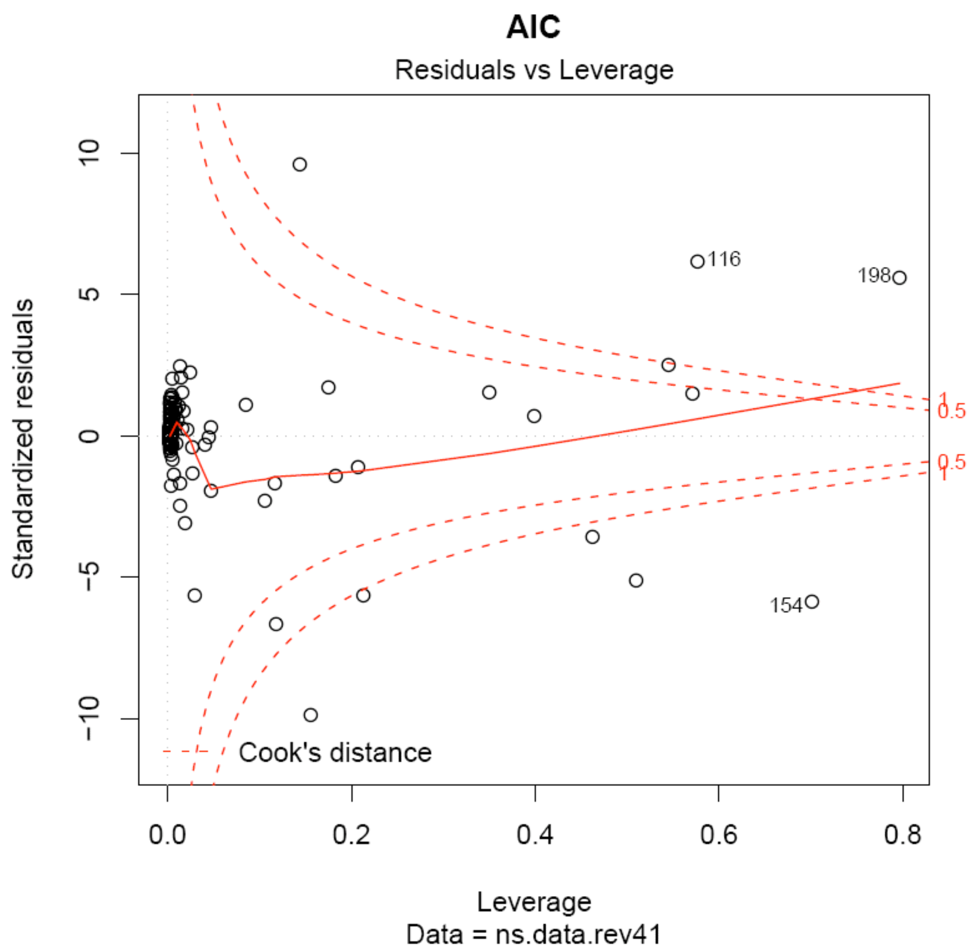

Figure 2-1 Cook's distance test applied to electricity data, low voltage distribution.

DFBETAs

2.66 The DFBETA for a predictor and for a particular observation is the difference between the regression coefficient calculated for all of the data and the regression coefficient calculated with observation i deleted, scaled by the standard error calculated with the observation deleted. The cut-off value for DFBETAs is 2/sqrt(n), where n is the number of observations.

DFFITS

2.67 The DFFITS statistic is a scaled measure of the change in the predicted value for the ith observation. For linear models, the formula is given in a separate document on PP. Large absolute values of Fi indicate influential observations. A general cutoff to consider is 2 or more generally 2 sqrt((number of obs)/number of parameters)

Covariance ratio

2.68 The Covariance ratio statistic measures the change in the determinant of the covariance matrix of the estimates by deleting the ith observation (cf Belsley, Kuh and Welsch, 1980). The cutoff to consider is 3*parameters/number of observations.

2.5 Experimental design

2.69 To assure that the chosen model fulfils the model criteria stated above, in particular with respect to environmental and categorical variables, it is useful to make a set of runs to determine the sensitivity of the model with respect to these parameters. The theory for experimental design and inclusion results for non-parametric methods effectively reduce the necessary runs.

Output analysis

2.70 We note that inefficiency indicated by benchmarking may be due to one or more explanations:

- Technical inefficiency
- Scale inefficiency
- Allocative inefficiency
- Industry efficiency development
- Excluded variables
- Low capacity utilization following investments
- Non-accounted quality differences in outputs/inputs
- Environment: climate, regulation or local market
- Uncertainty or measurement error

3. Criteria for model structure

3.1 Definitions

3.71 Below we outline and motivate the principal structure of the benchmarking models for both electricity and gas.

3.72 A *benchmarking model* is a quantitative representation of the relationship between resource consumptions (*input*) and the supply of demanded services (*output*) based on a set of comparable observations.

3.73 The model is used to estimate *cost efficiency*, defined as the ratio of actual cost to minimal cost to produce an equal or higher amount of products and services under equal or more severe environmental conditions. The minimal cost calculated by the benchmarking model for each output profile and structural specification is said to be the *efficient cost* for that condition.

3.74 To enable application of the efficiency concept in regulation, i.e. to obtain feasible and realizable targets, two conditions have to be fulfilled: (i) minimal extrapolation from the data and (ii) minimized data error.

3.75 Based on the definitions above, we may already infer certain important properties of a benchmarking model for use in e.g. incentive regulation.

3.2 Output orientation

3.76 Although a detailed input description, i.e. inclusion of resource variables such as the type of assets deployed, the number and qualification of staff employed, and the type of costs incurred in the operations, may bring interesting information to managers, the previously discussed endogeneity makes the model useless for incentive provision. In fact, a too detailed input description not only limits the incentives that a regulator can put on the results, but it may also misinform firms about the true latent efficiency potential by introducing artificially high substitution costs between inputs in the benchmarking. Modern incentive-based regulation is output-based, meaning that the input side is characterized by maximal aggregation to permit the firm to realize any beneficial substitutions, and that the effort in modelling is invested in finding good indicators for the services really demanded and delivered by the final customers. All intermediate steps are seen as internal processes that the regulatory benchmarking model has no interest to measure or influence, only their final impact on cost matters to the efficiency estimation.

3.77 Which are then the relevant output dimensions in energy networks in general? Customer tariff calculations can not be directly used, as until now they have been based on accounting costs for the activities. The dimensions are instead found by a purely logical reasoning. In the liberalized market, the final client buys his energy from a supplier in an open competition. To exercise the purchase the client turns to a system operator and requests a seamless and uninterrupted transportation and distribution of the energy from the supplier to the connection point. This creates three service dimensions for the distribution task; *transportation work*, *capacity provision* and *customer service*.

3.78 In electricity, the volume transferred gives rise to physical losses that are proportional to the network configuration, the distance between supply and connection and the outside temperature. In gas, the volume transported also relates to the energy costs for pressurizing the network. This dimension, which is perfectly measurable at final client level, is called *transportation work.* The relevant variables in our dataset include total delivered electricity (MWh) per voltage level, measured or non-measured, as well as the total energy contents in the gas delivered (MJ) or the gas volume delivered (nm^3). For gas, the normalized volume has been retained, rather than the energy contents, as the difference in the two is related to e.g. temperature and quality that are not relevant for the costs of the gas distributor, but related to the final use.

3.79 The demand from the client for electricity connection cannot be refused under the Universal Service Obligation (USO) and the client is not contractually requested to remain connected, or to guarantee any specific load profile. In gas, we note moreover that the client has the right *not* to be connected, even if the network extends to his house or property. USO is a great privilege, but as there are no free privileges, the corresponding obligation of the system operator to foresee, provide and adjust the capacity of the network to the demand carries a cost that corresponds to the *capacity provision* the client enjoy. In this dimension we find variables that help to explain the dimensioning, the capacity utilization and thus also the economies of scale in the networks. In our dataset, we may mention coincidental and non-coincidental peak load (MW) per voltage and transformer level, peak feed and outtake rates (nm^3/h), but also the delivery area itself (km^2) and its population are USO drivers for capacity provision.

3.80 Once connected, the customers require both specific capital assets (meters, connections, lines and cables, transformers, compressors) with their associated direct and indirect costs, in addition to some direct operating costs (connection, metering, billing, fault detection and correction, disconnection, inspection). We label all these connection point proportional services as *customer services*. Relevant variables in our data to capture this dimension are the number of metering points per voltage and transformer level and the number of customers per pressure, voltage and transformer level. Alternative variables used in regulation include the number of connection points per voltage and pressure level and the number of terminal blocks in gas.

3.81 The output side of the model should adequately represent the three service dimensions. The selection of the exact parameters within dimension and their functional form is left to the extensive statistical tests reported elsewhere. The core model is estimated prior to the selection of structural correction parameters.

3.3 Structural correction

3.82 Comparability between operators with the same output profile, but different costs, is assured through the successive test of plausible structural parameters on the results of the previously estimated core model. The candidates are selected among indicators for particular operating contexts that are likely to significantly deviate from that of the firms ranked efficient by the core model. In our datasets, we have solicited, constructed and used indicators for asset age and profile, slope, soil type (electricity), joint operation of other utilities and historical load growth patterns (gas). All tested variables have been documented and their capacity to explain cost deviations has been recorded. The choice to include or not a significant structural correction factor has been made using information-criteria from statistical theory. Some details of the tests in this respect are found in this report.

3.4 Integration

3.83 The network distribution of energy can be modeled by two principally different approaches with regard to structure: the separable model and the integrated model.

3.84 The vertically separable model (Figure 3-1) is based on separate estimations of cost functions and corresponding efficiency assessments for each vertical level, e.g. medium voltage distribution of electricity. Costs from superior levels are regarded as uncontrollable and no consideration is made to the joint ownership and/or operation of multiple vertical levels. As the choice of variables can be tailor made for each vertical level, a high fit can usually be achieved.

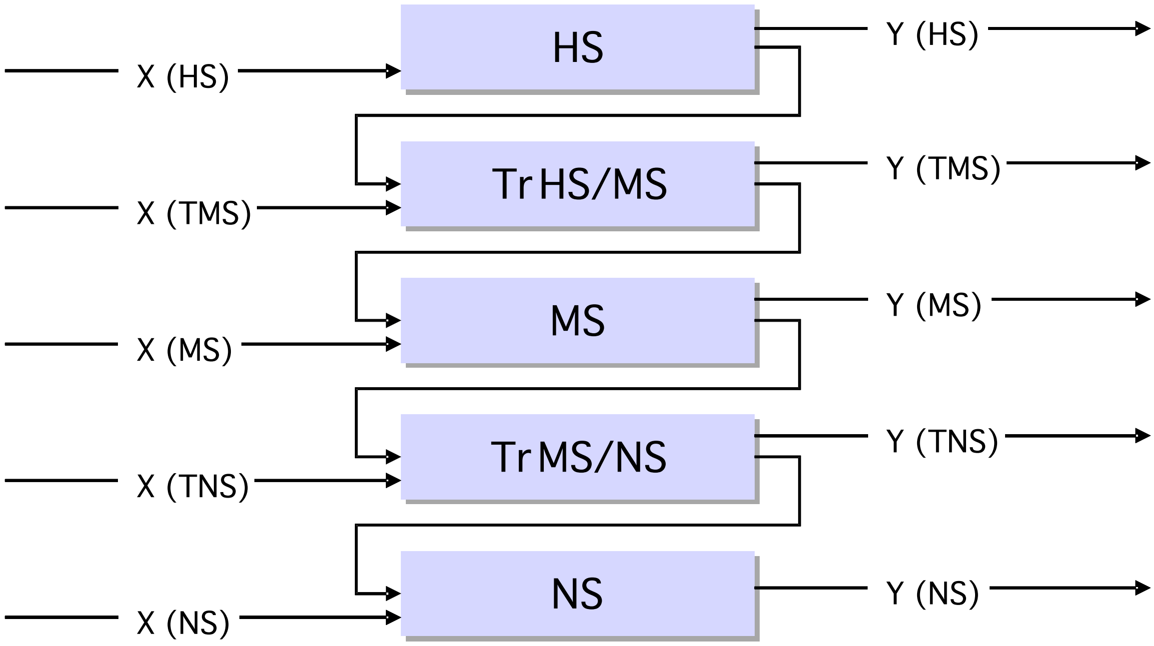

Figure 3-1 Separable model (example electricity distribution).

3.85 The integrated model (Figure 3-2), however, a more aggregated perspective on the activity as to take into account the synergies between vertical levels. Thus, internal monetary and energy flows are ignored as the integrated system is benchmarked. As all enterprises do not share the same vertical segments, the resulting model will have to select a set of variables that adequately characterizes all firms. The advantage is a greater stability to "accounting creativity" in regulated firms and an explicit recognition of the expected integration synergies.

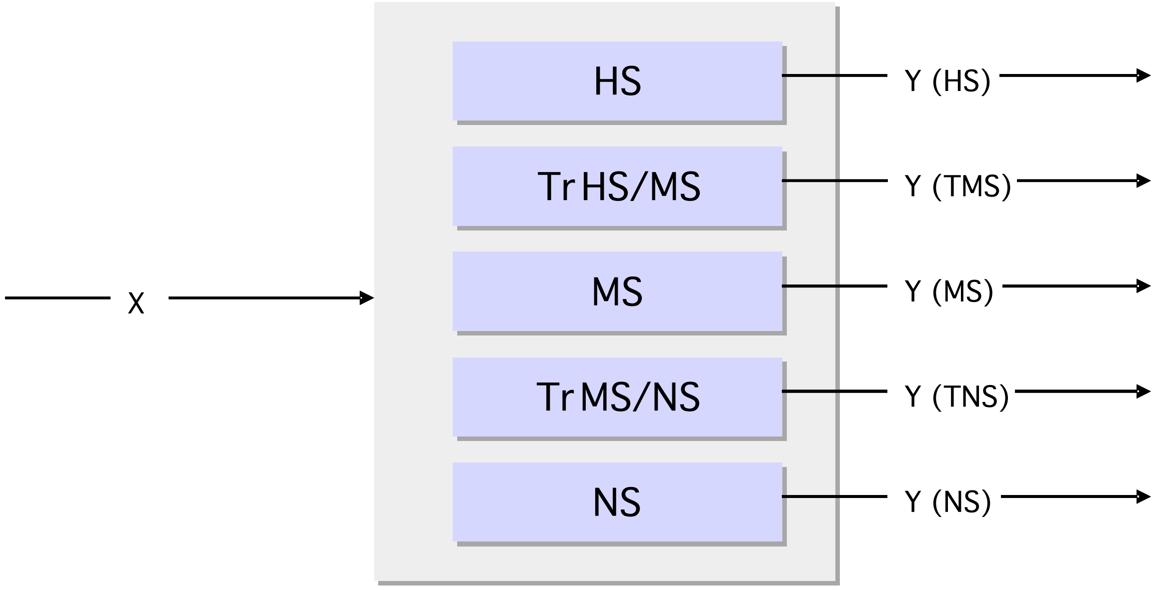

Figure 3-2 Integrated model (example electricity distribution).

3.5 Summary

3.86 Our approach for regulatory benchmarking to support incentive regulation is based on the concept of a core model with an aggregated input specification, here total direct cost less taxes and a detailed output specification along the final client dimensions transport work, capacity provision and customer service. The core model is augmented with environmental variables when they are shown to significantly explain cost deviations using both statistical and analytical techniques. The scope of the core model should correspond to the economies of scale and scope in the real operations, as to reduce model complexity and diversity, to avoid distortions due to bias in number of observations and model specification, and to acknowledge operating synergies rather than accounting creativity in the efficiency calculations. In the current dataset we have found support to use models that are integrated from low voltage up to high voltage for electricity and completely integrated in gas.

4. Methodology for model specification

4.1 Background

4.87 The benchmarking model is pivotal in incentive based regulation of natural monopolies. Benchmarking is relative performance evaluation. The performance of a DSO is compared against the actual performance of other DSOs rather than against what is theoretically possible. In this way, benchmarking substitutes for real market competition. A regulation where the DSO rights and obligations have already been allocated and where no service segment is covered by multiple DSOs, one cannot have competition for or on the market, but one can imitate the competitive situation by a benchmarking contest.

4.88 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.

4.89 From a benchmarking perspective, the German situation is unique in several ways. There is a large number of DSOs and a large set of data for each of these. Moreover, our preliminary results suggest that the data in general is of good quality. This means – in an international perspective - that one is to a lesser degree forced to use simplistic models and to rely on regulatory discretion to protect against noise in the data.

4.90 BNA has - in part as a consequence hereof - rather ambitious goals when it comes to the benchmarking. In particular, it is the goal of BNA to protect against methodological diversity by exploring several possible model specification (choice of benchmarking parameters) and by exploring several possible state-of-the-art estimation techniques (DEA, SFA etc). In this way, the methodological risk can be limited.

Preliminary recommendations

4.91 The analyses of the data that has been collected so far suggest that data it is of relatively good quality and that the ambitious goal of having results that are not too sensitive to methodological diversity can indeed be reached.

4.92 Although the analysis undertaken so far are the results of a first data collection phase, the empirical results are no worse than in other European countries even after several rounds of data collection and benchmarking model applications. This suggests that if BNA is given the discretionary power to continue the refinement of the benchmarking models, and if the specific model choice is not limited by an ordinance, Germany will be able to set new standards in regulatory benchmarking.

4.2 Variable selection criteria for regulatory use

4.93 *Robustness.* The model specification and results must be robust to foreseeable cost, technology and institutional changes to guarantee stable incentive provision and minimization of the regulatory risk. Specifications that rely heavily on specific process information, e.g., may become obsolete with technological progress.

4.94 *Verifiability.* An efficiency measurement model used in incentive regulation must be based on verifiable information. Use of poorly defined or private information is directly encouraging opportunistic action. Worse, in yardstick regulation, distorted information may directly affect the incentives of complying firms.

4.95 *Unambiguous.* The model's definitions have to be unambiguous to withstand challenges related to conflicting interpretations, e.g. over time and organizational levels.

4.96 *Output (correlated).* As discussed at length in Agrell and Bogetoft (2003a) *Dynamic Regulation*, the most robust and least long-run costly regulation regime will be implemented with close definition at the output side and high aggregation on the input side. We pursue this strategy in this report by allocating more effort to the long-run output specification than the (transitional) input problems. The output orientation also yields a process independent model, which strengthens the robustness condition above and creates clear signals of regulatory non-involvement in the operations.

4.97 *Minimal structural impact.* The variables should be defined as to enable the model to capture efficiency with equal precision across firms with varying size, organization and output profile. This means that *bias* should be avoided if possible already in the variable specification phase.

4.98 *Feasibility.* A regulatory model must show feasible results for any imaginable outcome to limit regulatory discretion. In incentive regulation, we note the problem of superefficiency, where the DEA program may fail to find an efficiency estimate for certain production profiles. Superefficiency is not a problem in a SFA context and we refer the reader to the discussion in Agrell and Bogetoft(2004), Agrell, Bogetoft and Tind (2005).

4.3 Variable classification

4.99 The classification of variables and parameters for the models is illustrated in Figure 4-1 below. With input X or controllable resources we primarily mean the direct costs C(X) for the level or activity estimated (e.g. medium-voltage electricity distribution), but also all variables related to the operating costs and assets deployed, e.g. route length of overhead lines (electricity) and installed power for compressors (gas) The class of outputs Y is made of exogenous indicators for the results of the regulated task, such as typically variables related to the transportation work (energy delivered etc), capacity provision (peakload, coverage in area etc) and service provision (number of connections, customers etc). 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 find indicators of geography (topology, obstacles), climate (temperature, humidity, salinity), soil (type, slope, zoning) and density (sprawl, imposed feed-in locations). However, it is common to find the effects of a particular control in Z correlated with Y and/or cancelled with other controls in Z.

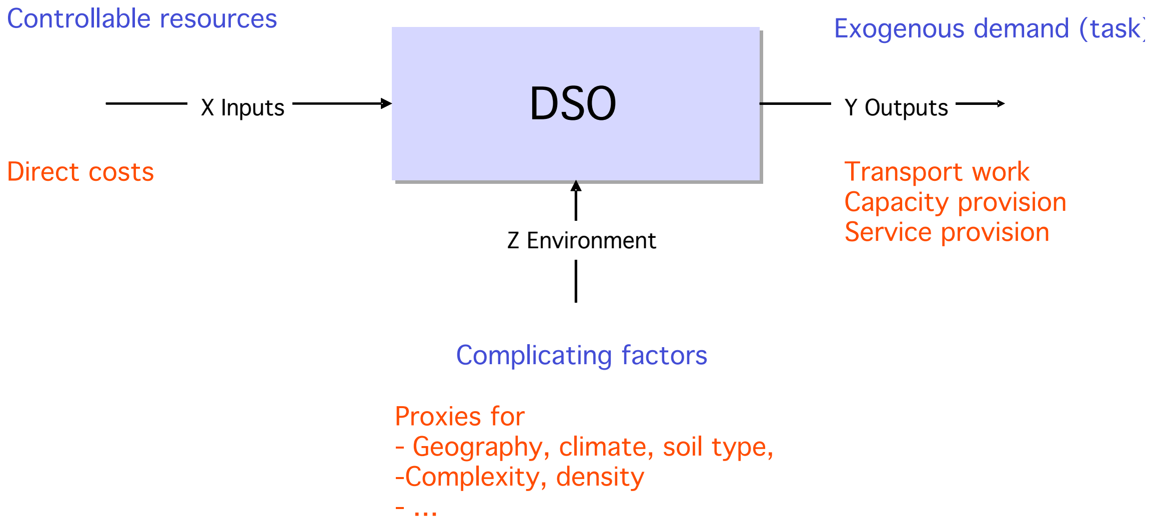

Figure 4-1 Variable classification.

4.4 Estimation procedure

4.100 The model estimation procedure is an interesting case of combined used of econometric and analytical cost functions.

4.101 The first phase (OLS 1) investigates the relationship between any resources (controllable variables X) and the resulting costs C(X). These models are of limited interest for benchmarking, as the resources are controllable, but serve to validate the data.

4.102 The second phase (OLS 2) aims at the validation of the explanation of the intermediate variables (controllable) through the exogenous variables that characterize the service output. These models serve a validation purpose only and to the extent the intermediate variables can be adequately explained by the exogenous variables, they can be omitted from the further analysis, which is a prerequisite for a regulatory application.

4.103 The third phase (OLS/DEA/SFA) estimates the cost function as a function of exogenous outputs only, potentially a subset of the variables used in the preceding phase. This step constitutes the core of the benchmarking modeling as it has to be effectuated for each type of model and estimation approach. In the process of estimation these benchmarking models the data is validated and cleaned once more, with a focus on possible outlier detection. This is done to reflect the sensitivity of benchmarking results to extreme performances.

4.104 The fourth phase (Structural) is devoted to augment the model(s) from the third phase with controls (environmental variables) to assure comparability and explain revealed inefficiency. This phase involves the test of a large number of potential candidates for structural variables to determine whether they add relevant information. The phase has in this project lead to the construction and data collection of new variables for e.g. soil type, urban sprawl and age structure of the firms.

4.105 Following the four phases, we have a balanced benchmarking model that is a function of the relevant outputs from the regulated task, subject to the assurance of a structural comparability.

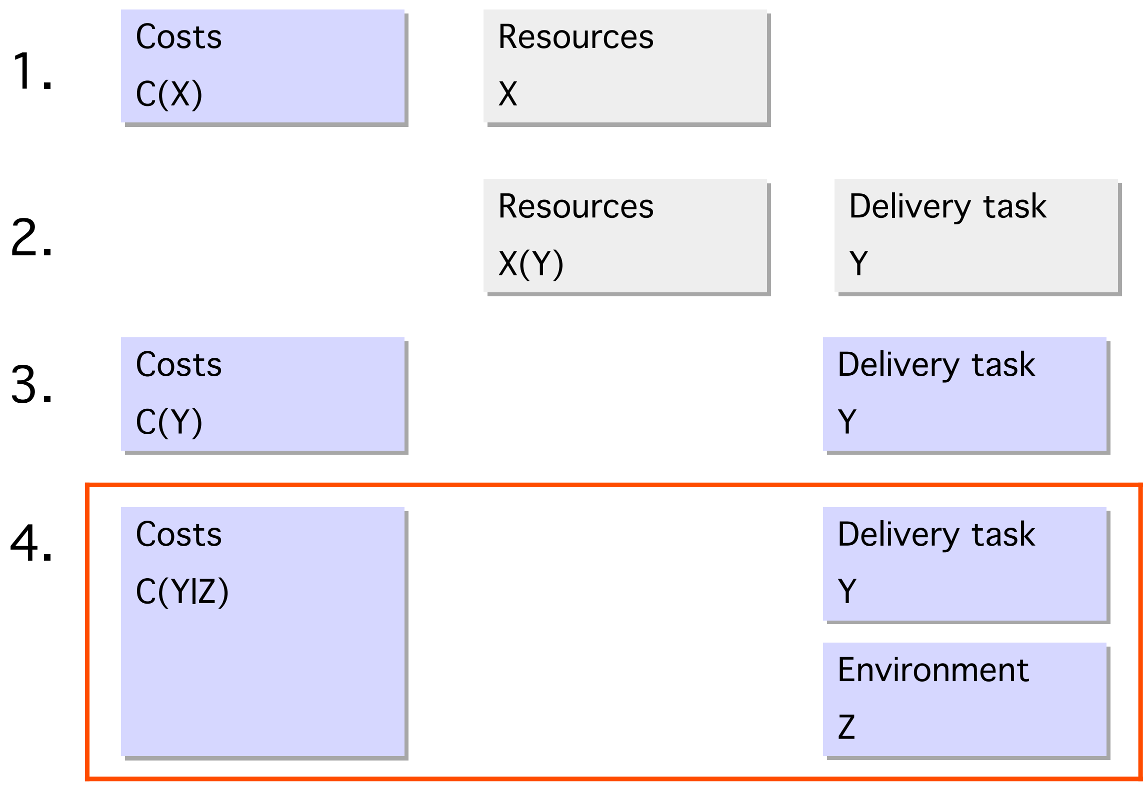

Figure 4-2 Four-stage statistical estimation approach.

4.5 Model specification procedure

4.106 The variables proposed for both the delivery task and the structural characterization come from two sources in the development phase as illustrated Figure 4-3. First, the ranked results of the AS2 empirical expert assessment from a panel of electricity and gas experts is used as a first list of candidate variables for validation. The cost drivers from this phase (OLS 1) are validated in the first stage of the statistical process above, since they constitute of mix of endogenous (controllable) and exogenous (non-controllable) cost drivers.

4.107 The next phase (OLS 2) requires a clear distinction regarding the causality of potential variables for X, Y and Z. Here the analytical results of the MNA project intervene to contribute to the classification, since the engineering model more closely can validate the causality of the variables.

4.108 In the third phase (OLS/DEA/SFA) the MNA analytical cost models provide the model specification with a qualified set of primary variables to characterize the distribution task and in addition testable hypotheses for the relationships between the variables. The empirical phase validates the relationships to obtain as close fit as possible in the parametric estimation. Possible contradictions are investigated in several levels as they may indicate problems in the cost accounting and allocation. The nonlinear relationship between the number of connections and the network length is an example of information obtained that contributes to a better fit for the empirical estimates.

4.109 The expert assessments and the analytical cost models also contribute to the fourth phase (Structural) to select variables for both the parametric and non-parametric models. While the analytical relationships from MNA serve to find some structural relationships, the ideal network estimates from selected network provide comparators to the uncorrected efficiency estimates that are used as validation sample for selected sets of structural variables Z. Obtained coefficients for the parametric model are also compared to the analytical estimates from MNA to validate the soundness of the model and to aid in the screening of correlated variables.

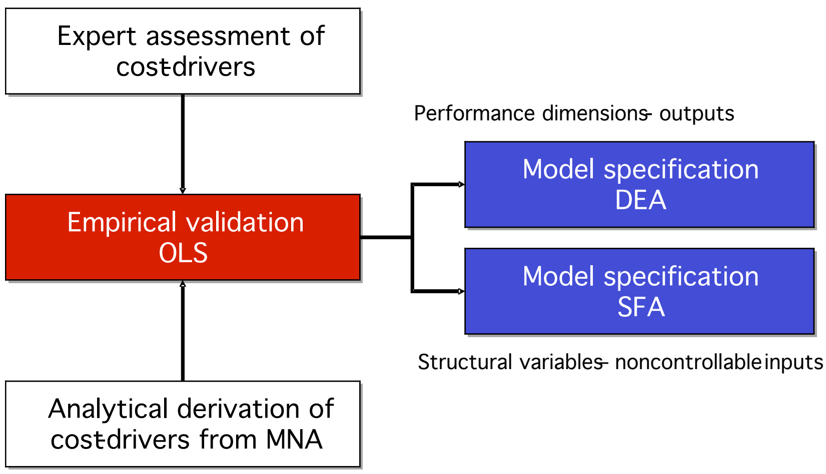

Figure 4-3 Interactions between expert assessments, analytical cost models and model specification for DEA and SFA models.

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

4.6 Banker test of model specification

4.111 Banker (1996) developed statistics to test hypothesis about the characteristics of a cost or production function, including return to scale and the choice of one variable specification against another. The general structure of these test statistics are based on the idea that a change in the model specification should not lead to too large changes in the distribution of the inefficiency scores. 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 DSO 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*.

4.7 Three general model specification issues

4.112 *T1 Subadditivity test.* Joint models were estimated on a validated, representative sample of 268 electricity distributors. The coverage of the electricity data over the levels is good from lower voltage up to medium voltage. Vertically separated models for each voltage/pressure level were then estimated using relevant subsets of the variable set for the joint model. The statistical results were weakly in favor of the joint model in both sectors. The interpretations to the findings are twofold. For electricity, the technology is relatively linear from a fairly modest scale, which is also confirmed by later estimations. This implies that the joint model also exhibits additive properties rather than strong subadditivity. Secondly, several arguments of regulatory and model robustness speak in favor of a joint model. To meaningfully implement separate models would require a convincing accounting separation of costs and observability of intermediate outputs. As this would be not only costly, but also unnecessarily contradict the stated ambition to create an incentive regulation without micro-management, the joint model structure is strictly preferred for both electricity and gas. As with regard to TSO, see previous comment.

4.113 *T2 Age effect tests: electricity.* The age effect is tested in electricity using a standardized age indicator that is constructed from a validated, representative sample of 268 electricity distributors. Investment data from 1955-2004 divided into lines, cables, transmission equipment and distribution equipment were used to create an inflation-adjusted weighted age indicator for the non-depreciated asset base. The mean value is 18 years, which roughly corresponds to half the asset life for typical depreciation lengths of 30-40 years. The weighted age, as well as the specific estimates for the age of the other asset groups are submitted for second-stage tests against DEA and SFA scores to determine potential age bias.

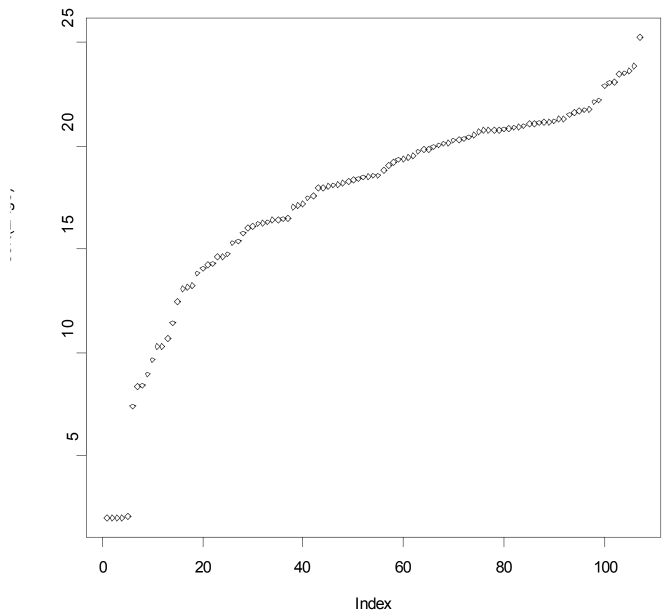

Figure 4-4 Age indicators for electricity distributors (sample).

4.114 *T2 Age effect tests: gas.* The age effect in gas has been tested using an age measure based on the reported age of pipelines for high, medium and low pressure. Four categories each were reported (<5, 6-20, 21-45 ,>45 years) and the age metric was weighted with the physical length of the pipelines, cf. Figure 4-5 below. The age measure, as well as the reported asset classes themselves, were part of the structural variable candidates in the OLS estimations and are submitted for second-stage tests against DEA and SFA scores to determine potential age bias.

4.115 *T3 Capex-Opex test.* The potential impact of asset valuation methods (book or replacement value) on the efficiency assessment has been estimated for both gas and electricity using reconstructed asset bases using standardized annuity-based capex measures. A second-stage regression is used to investigate to which extent the choice of depreciation and/or valuation method introduces statistically significant distortions in the subsample.

4.8 T1 Subaddivitity test

4.116 To give some theoretically sound basis for the selection of the level of aggregation above, we have conducted a test on the subadditivity hypothesis of the underlying cost function. Simply put, a subadditive cost function $C(a,b,\ldots c) < C_1(a) + C_2(b) .. + C_n(c)$ indicates that there is a cost advantage to perform an activity jointly. The test is fairly standard in regulation, cf. Evans and Heckman (1984)[2] when considering e.g. the level of vertical integration allowed in a natural monopoly. As other European regulators have chosen only to distinguish up to three vertical segments in electricity (TSO, DSO-MV, DSO-LV, including transforming activities) and two in gas (TSO, DSO), we might expect to find support for some integration in electricity

4.9 T2 Age effect test

4.117 The age structure of the asset base may potentially have two effects that contradict the two purposes of the benchmarking, the efficiency estimation and the structural comparability.

4.118 First, the incumbent regulation in Germany is cost/input-based which means that the costs and tariffs reflect accounting costs rather than real annuities. Naïve estimations of cost efficiency of firms at different stages in the investment cycle would result in the infeasible finding that firms ideally should remain in the state with assets (almost) fully depreciated. To correct this problem, the impact of an age-corrected capital measure must be employed.

4.119 Second, as has been hypothesized for e.g. electricity TSOs, asset age may have an unfavorable impact on operating cost. Although not an issue under annuity pricing, this may disturb the comparability of firms using different activation policies.

4.120 The age effect test T1 is defined as the construction of weighted asset age parameters.

[2] David S. Evans, James J. Heckman (1984) A Test for Subadditivity of the Cost Function with an Application to the Bell System, American Economic Review, 74(4), pp. 615-623.

4.121 In gas, the asset ages are estimated from the given proportion of pipelines in age classes, weighted with the proportion of lines (in km), cf. Figure 4-5 above.

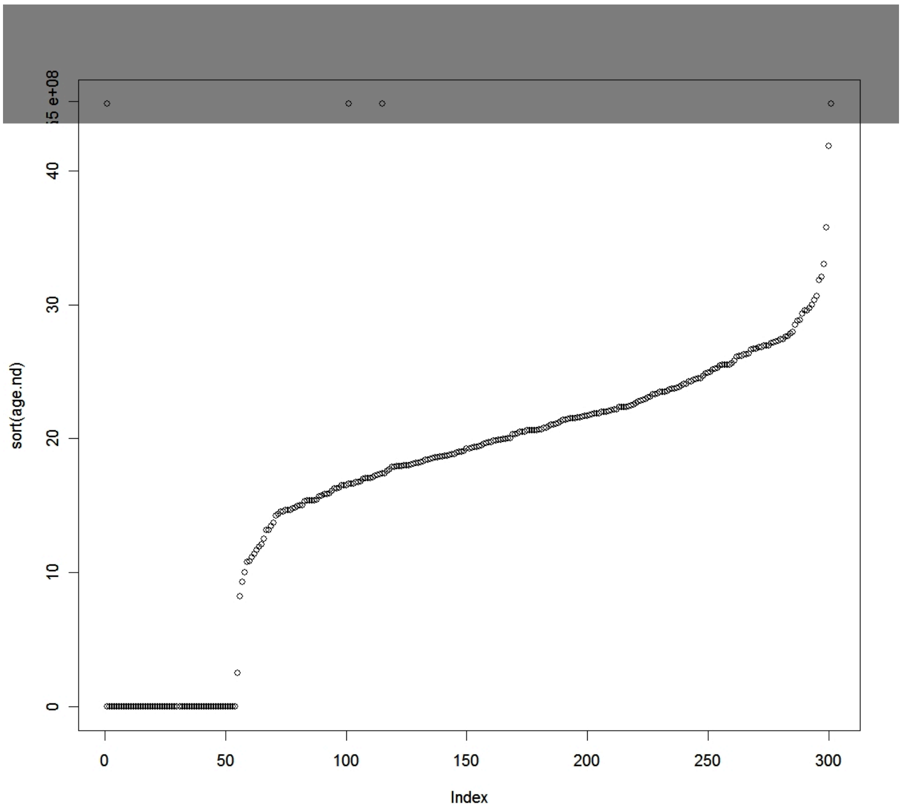

Figure 4-5 Example of age estimator: average age for low pressure gas pipelines

4.122 For electricity, an asset age parameter was constructed from the proportion of residual book value (depreciated) to total (non-depreciated) book value for five asset classes. The underlying assumption is that the firms practice linear depreciation.

4.123 The asset age estimators have been included as potential explanatory variables in class Z to determine whether age has a impact on (i) cost, (ii) uncorrected efficiency estimates under DEA and SFA.

4.10 T3 Capex-opex test

4.124 *T3 Capex-opex test.* To test for the potential distortion of the capital cost component of the direct cost through heterogeneous activation and depreciation schemes, a standardized total cost measure was created for a subsample of 232 electricity distributors. Investment data for 1955-2004 divided into lines, cables, transmission equipment and distribution equipment were adjusted using a producer price index (Index für Erzeugnisse der Investitionsgüterproduzenten) and carried forward as annuities using a spread of real interest rates and technical life lengths as in the ECOM+ method by Agrell and Bogetoft (2003). The obtained annuity-based capital cost measure is used in a second-stage analysis to detect potential bias, in regression studies against the total cost measure, and as component of an annuity-based total cost measure with operating cost in comparative runs of analogous DEA and SFA models.

4.125 The scope of the tests are twofold: (i) to test the hypothesis that the potential impact of the heterogeneous capital valuation has an explanatory value for efficiency estimates in either model, and (ii) to explore alternative and robust future remedies to the asset valuation problem in the benchmarking. The first task is relatively straightforward, as described through the second-stage regression above in line with the methodology in international TSO benchmarking (Agrell and Bogetoft, 2004b) and the international regulatory guidelines by the World Bank (cf. Coelli et al, 2003). The second task is more demanding and may involve the AS3 project on analytical cost models in a subsequent project to validate the obtained annuity corrected measure, especially for frontier units and/or firms that have exceptionally low annuity-based capital intensity per e.g. weighted grid asset unit (km of cable, line etc) due to extremely low activation of investments in the past. Note that the intention of this investigation primarily has the purpose to create comparability of performance to guarantee feasibility (*Erreichbarkeit*) though the future recalculation of the capital intensity of the firm in the benchmarking, without direct implications with respect to the reporting, accounting and tariff regulation in force.

4.126 The current investigation is limited to a *standardization* of the capital valuation and consequently the capital costs for comparative purposes. This measure does not involve an investigation regarding the appropriate weighted average cost of capital (WACC) of the sector and, in particular, of the efficient firms, which falls outside of the scope of the benchmarking. In this context, it is sufficient to note that the benchmarking can be made robust also when the reference set includes firms with artificially low capital costs (cf. NEMESYS, 2005)

4.127 The standardized capital cost measure for gas (based on asset and investment data from 1940 to 2004, adjusted for inflation and using standardized technical lifetimes of 45 years for pipelines and 20 years for meters, average real interest rate 3.5%) correlates well with reported CAPEX for the available sample (Pearson 0.946, cf test below). In Figure 4-6 we note a corresponding strong relationship with direct cost, meaning that, at least, differentiation in depreciation regimes has not created any substantial distortions in the reported cost.

```
data:  xCAPEX and xCAPEX.stand
t = 35.7853, df = 149, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.9268405 0.9609150
sample estimates: cor 0.9464533
```

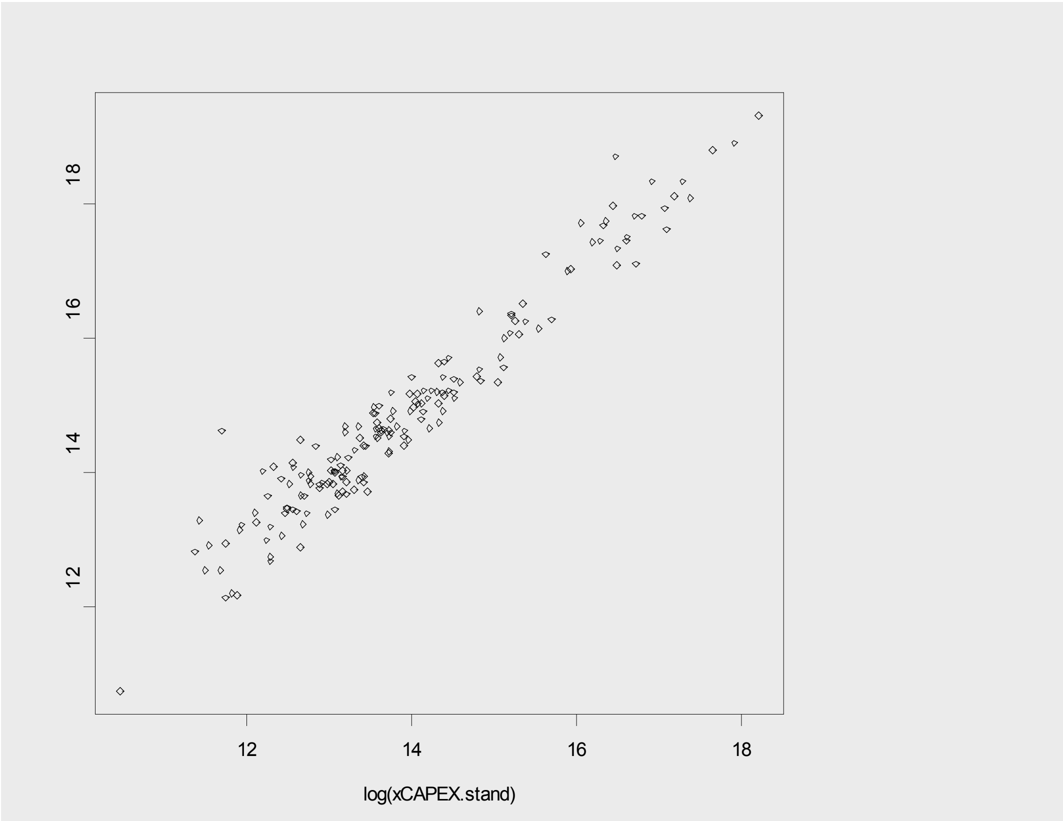

Figure 4-6 Direct cost as a function of standardized CAPEX, both log, gas model 8.

4.11 Outlier detection

4.128 Frontier models in general and DEA in particular are sensitive to possible data problems in frontier units. This suggests that extra care is needed in the control of the data from such units. In this project we have applied an iterative process of doing trial estimations, to detect DSOs that might be outliers and set the frontier. Such units have then been fed back to data validation and in some cases this has lead to the identification of reporting problems. Having fixed them, the process has been repeated in up to three rounds to make sure that the data of the frontier units are of good quality. This is a time consuming process that has to be repeated whenever new data is collected. It is however an important task and not one that can be done effectively without the benchmarking since the benchmarking gives a clear indication of the relevance of different units.

4.129 It is important to emphasize that we do not eliminate data simply to trim the model. Indeed, the preliminary results presented below rely on all the observations that the benchmarking team had access to. This means also that the results are based on a representative sample of German gas and electricity DSOs

4.130 The identification of DSOs to check more carefully has used in particular four approaches.

4.131 One is to identify the number of times a DSO serves as a peer unit for other DSOs. If a DSO is the peer for an extreme number of units, it is either a very efficient units – or there are possibly some mistakes in the reported numbers.

4.132 The other approach is to investigate the impact on average efficiency from unilateral elimination of the DSOs. If the elimination of one DSO leads to a significant increase in the efficiency of sufficiently number of units, there are again good reasons to check this more carefully.

4.133 Thirdly, we have done so-called shell analysis where the idea is to study the impact of groups of DSO, like the ones in the first shell, the second shell etc, cf also Agrell and Bogetoft(2002). 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.

4.134 Finally, we have used super-efficiency calculations to determine units with extreme super-efficiencies that are often associated with outliers, c.f. Banker and Chang(2005). Other outlier detection methods designed with particular focus on frontier models have also been considered, for example Wilson(1993).

4.135 In addition to these three methods, the data analysis has benefited from so-called order-m efficiency. The idea here is to evaluate a given DSO against the costs of the least costs DSO in a sample of m units that all provided at least the same of all services under the same or more difficult structural conditions. Of course, the order-m efficiency will depend on the sub-sample selected and for each unit, we have therefore done Monte-Carlo simulations picking the m units at random from the full population some 1000 times.

4.12 Bias correction

4.136 DEA models provide cautious estimates of the saving potentials and cost inefficiencies. This is one of the attractive features of DEA and is part of the theoretical foundation for the optimality of DEA based yardstick competition, cf. Bogetoft(1997,2000). If the model structure and the variables are chosen correctly, it means that no-one will be required to produce at costs that are below the truly minimal ones. In the terminology of incentive theory, the outcome is *individually rational*.

4.137 The back-side of the cautiousness is that the cost efficiency is biased upwards. On average, the units will look more efficient than they really are. This may be a problem in terms of the sharing of benefits between consumers and firms. It is however the price one must pay to avoid that some units are faced with too hard terms. Moreover, the bias for some units may be larger than for others. Hence, the division of benefits between consumers and distribution companies may differ from one area to another.

4.138 Likewise, some distribution companies may be facing tougher norms than others. This is an unattractive feature in regulation since it may prohibit a fair treatment of the different units. Units with many comparators will usually be compared to tougher norms than units with few comparable units.

4.139 Before getting too concerned it is important to understand the implications. All norms should suffice by the cautiousness property, but for some units the extra benefits from a biased evaluation is more lucrative than for others. Hence, in summary, *all units benefit from the bias in the DEA model – but the benefit is not equally distributed among the firms.*

4.140 Fortunately, recent theoretical developments have taught us how to correct for this bias, namely via boot-strapping. In consequence, we can – following the literature initiated by Simar and Wilson (2000) - determine bias corrected efficiency scores, i.e. scores that are not biased. Moreover, using bootstrapping, we can determine confidence intervals around the bias corrected efficiencies. We shall rely on these advances in the empirical analysis. The procedure involves the simulation of for example 2000 different versions of a model with variable and properties fixed but with the inefficiencies drawn from an efficiency distribution determined by a so-called kernel estimation of the empirical efficiency distribution (i.e. non-naive bootstrapping).

4.141 Another way to at least partially cope with the unequal treatment of units due to different estimation bias is to impose some general regularity on the structure, viz via parametric models, via weight restrictions or via enhanced possibilities to rescale.

4.142 A third approach is of course to rely on increased international comparisons. The large number of German DSO in electricity and gas and the general problems of international benchmarking make this less attractive but it may be worthwhile for very large units and the units working on the higher voltage levels. In particular, international benchmarking of TSOs seems obvious.

4.13 Structural corrections

4.143 The relative performance evaluation of the DSOs shall ideally be corrected for difference in the structural conditions. Thus, a DSO forced to live with the difficulties of special climate, topology, particularly dispersed costumers etc shall not be compared directly with DSO without these challenges.

4.144 There are several ways to correct for such differences and to test their importance.

4.145 In a DEA framework, ordinal structural variables can be used to group the DSO and to only compare a DSO against DSOs working under less favorable conditions. If the structural variables are interval scaled, we can instead include them as pseudo non-controllable inputs or outputs. The only difficulty with this approach is that several of the variables one can think of have a different nature than the usual resource oriented inputs and outputs. This means that the underlying assumptions about DEA in terms of convexity and scale may not make sense. This is often the case with ratio oriented structural variables. In such cases, one will instead have to rely on a second stage analyses. In a second stage analysis, the efficiency scores are regressed against the structural variable to determine the general impact of these. Next the efficiencies are corrected for the impact of these variables using the regression model. Often, an OLS estimation is used and in large samples with small efficiency levels like in the German context, this is a quite safe approach. Nevertheless, if one analyzes ordinary efficiencies as opposed to superefficiencies, one should ideally take into account the truncated nature of the dependent variable, the efficiency, i.e. one should use a truncated regression a la Simar and Wilson (2004) or a TOBIT regression following Tobin(1958).

4.146 In an SFA framework, the correction for structural variables even if they are captured by ratios, can be handled as an integral part of the maximum likelihood estimation by parameterzing the inefficiency distribution with such variables, cf. Batese and Coelli(see FRONTIER notes). Of course, one can also use a second stage approach like in DEA but there seem to be some advantages of the former approach, cf. Coelli and Perelman(EJOR; Airlines paper).

4.147 In some case, we are not interested to correct the efficiency scores for the structural variables but we are interested to know if the model is biased against one type of DSO rather than another, e.g. DSOs with old rather than new grid assets due to the accounting cost of capital. In such cases, one can – in addition to the second stage regressions – make non-parametric tests – like classical Mann-Witney(1947) and Kruskal and Wallis(1952) tests.

4.14 Robustness

4.148 The invariance of the model to stochastic changes in input data constitutes the numeric robustness of the model. A numerically robust model smoothes or at least does not amplify small perturbations of the reference data for performance. However, in regulation, the robustness of the model specification to changes in the operation processes, accounting reporting standards and financing policy is an equally important issue. A model specified in such a way that it draws on a process- or accounting specific specification will either unnecessarily hamper innovation, which might increase cost, or prompt for periodic revision of the model specification, which lowers regulatory commitment. Likewise, a variable specification that implicitly assumes a certain financial policy, such as full ownership of network asset as opposed to leasing or renting agreements, will also be vulnerable to future changes that may be potentially cost efficient. This aspect of robustness implies naturally that a regulatory model should opt for aggregation on the input side as to remain independent of changes in process, but should be disaggregated on the output side to correctly record the performed services and avoid dependency on implicit service specifications. A robust model for regulation is a model that consistently gives feasible and reliable, if not optimal, estimates of efficiency with a model specification (e.g. variables, methods, parameters) that stays constant or that is updated only gradually in a systematic fashion (inflation adjustment, price parameters etc.). The proposed models are robust in the sense that for instance

i) the aggregated cost input corresponds to total relevant cost,

ii) the scope of the model corresponds to all voltage levels, low voltage, medium voltage, high voltage performed by a DSO,

iii) the outputs of the models are fully exogenous and measured in physical terms,

iv) the combined utilization of a deterministic and a stochastic method to assess the efficiency of the same specification hedges against both data specific errors and model specification errors.

Robustness - Method consistency

4.149 The invariance of the efficiency estimates to the regulatory choice of methods is a proof of robustness of the regulatory approach. In the proposal this corresponds to the consistency between the deterministic estimates from the DEA model and those of the SFA. A simple measure of such a consistency might be the classical Pearson correlation between scores. However, such restrictive interpretation of the consistency criteria is overly simplistic. The assumptions behind the deterministic and stochastic models are fundamentally different. Whereas the DEA model gives a deterministic point estimate the SFA outcome is a probability distribution around a mean where the variability depends on both data quality and inefficiency distribution. Thus the traditional evaluation of model consistency between DEA and SFA is assessed using rank order correlation metrics. As suggested by Prof. Weyman-Jones, an additional useful indicator of consistency may be the proportion of DEA estimates included in the corresponding SFA confidence interval. Albeit this indicator correlates positively with the incidence of noise in the data, it has the flavor of statistical hypothesis test, although no threshold value can be defined. The preliminary results from the electricity model indicate rank order correlation in the order of 70-75% which judging by comparable data for internationally used models (Sweden, DSO, 2001 data, regulatory DEA-VRS model, Pearson correlation with SFA 40%) must be considered highly satisfactory. Numerical experiments on the impact of dimensionality on the model consistency suggest that the obtained level consistency needs not to be compromised with the inclusion of further structural variables.

4.15 Economies of scale

4.150 The question of economies of scale for efficiency analysis is intimately related to the time horizon and the licensing instrument of the regulation. An operator awarded a specific concession area cannot be asked to operate beyond scale in an area with little or no growth, and the application of the efficiency evaluation should be adjusted accordingly. The earlier regulation in Norway and Sweden, cf. Agrell and Bogetoft (2003), used a variable return to scale (vrs) specification for the DEA models. However, the logical implication of this for larger enterprises would be a diseconomies of scale which may be disputed in a sector that is characterized by mergers. Consequently later application (Agrell and Bogetoft, 2004) have proposed using non decreasing return to scale (ndrs) as to give leeway for smaller operators with limited possibilities for merger or network growth, while maintaining the requirement of optimal scale of operation for larger operators. This is equivalent to accepting the hypothesis of economies of scale but rejecting the hypothesis of diseconomies of scale. However, a long-term cost efficiency run for the German network regulation should be made using constant return to scale (crs) as no concession areas are predefined, i.e. the scale of operation is endogenous to the firms. Note that such assumption in no way hinders the technical calculation of scale efficiency as to potentially adjust short run efficiency targets. The preliminary empirical results for the electricity model confirm earlier international findings of relatively linear technology which renders the differences between the ndrs and crs specifications minimal. Thus, we argue that a crs specification is an uncontroversial and robust specification for the model.

5. The electricity model

5.1 Parameter selection electricity

5.151 We have experimented with alternative specifications of the electricity distribution cost models. An early specification has been presented at various meetings, cf. above. The second prototype model for electricity, electricity model II, that we present here is estimated on the most recent data and is slightly modified compared to the initial model. Electricity model II is based on a specification with 12 output dimensions in two categories that cover the voltage levels high voltage, medium voltage and low voltage, with the transformations high- to medium voltage and medium to low voltage. The removal of the previous parameters related to energy delivered at HS level and DG power at all three voltage levels decomposed resulted in improvements in model robustness and consistency.

5.152 For all variable and parameter definitions, please refer to BNetzA (2006) to obtain the official translation and conversion to the data specification call.

Input

5.153 Total direct cost (as specified in the regulation)

Outputs

Table 5-1 Parameters in electricity model.

Parameter	Category	Definition
yMeters.hs	Service provision	No of meters high voltage
yMeters.ms	Service provision	No of meters medium voltage
yMeters.ns	Service provision	No of meters low voltage
YArea.hs	Service provision	Total service area for high voltage
YArea.ms	Service provision	Total service area for medium voltage
YArea.1.ns	Service provision	Total service area for low voltage
yPeakload.hs	Capacity provision	Coincidental load high voltage
yPeakload.hs_ms	Capacity provision	Coincidental load transformer HS/MS
yPeakload.ms	Capacity provision	Coincidental load medium voltage
yPeakload.ms_ns	Capacity provision	Coincidental load transformer MS/NS
yPeakload.ns	Capacity provision	Coincidental load low voltage
yDg.power.total	Capacity provision	Feed-in power of decentral generation

5.154 From a conceptual point of view, the omission of transportation work from the electricity model is somewhat unusual. The transportation work will in principle generate costs due to losses in the network and given this – and the general idea of capturing the services provided to the customers - it is included in many models. There are however several reasons why we have excluded it here.

5.155 First of all, the model fits and the comparability of the models are worsened by the inclusion of different measure of the transport work in the German context. This is actually not surprising since the statistical significance of this cost driver is also not impressing in other international models.

5.156 Discussions with engineers also suggest that amount of energy is not a primary cost driver as long as the model includes an appropriate measure of the capacity provision, i.e. peakloads. Truly, the losses will depend on the energy transported but these losses are not large and are not varying very much compared to other cost elements.

5.157 Lastly, we note that the detailed measures of meters used in the model may proxy also for amount of energy.

Structural tests

5.158 Structural parameters have been included in a series of second stage runs, see below.

Data

5.159 The preliminary estimates below are made using a validated sample of 328 firms operating in the relevant voltage levels, covering 99% of the total direct costs in the available dataset with a total of 349 firms. The eliminated firms are identified using the outlier identification techniques described above, most notably outlier identification based on superefficiency, cf. Banker and Chang(2005). Judging from the results below, the general data quality is reasonably high to permit a first evaluation of model specification.

5.2 Descriptive OLS 2 results for endogenous variables

5.160 Extensive regression runs on endogenous (controllable) variables against exogenous demand factors (cf. phase OLS 2 above) have validated the hypothesis that these variables can be adequately explained by exogenous factors. The methodology is described in section 4.5. The explanatory value is very high as measured in adjusted R-square and the predictors are subsets of the variables selected for the cost-efficiency model. An example is given in Figure 5-1 below, there the adjusted R-square and the significant exogenous cost drivers are given for some nine endogenous variables. After validation of the existence of the relationship, no further attention is given to the direct use of the intermediate variables in the estimations.

low-voltage	yCustomers	yMeters	yEnergy	yPeakLoad	yArea	yArea1	yArea2	yArea3	yArea4	yArea5	yPop		adj.r.squared
xCablesCircuit	x		x	x	x	x	x						0.98
xLinesCircuit	x		x	x	x	x					x		0.87
xCablesRoute	x	x	x	x	x	x					x		0.94
xLinesRoute	x		x	x	x	x	x				x		0.86
xTowers	x	x	x			x	x				x		0.79
xCostDIR		x	x	x	x	x	x				x		0.99
xCostIN	x	x	x	x	x	x	x				x		0.99
xRevenues		x	x	x	x	x	x						1.00
xTOTEX	x	x	x	x	x						x		1.00

Figure 5-1 Example of OLS 2 results for endogenous variables (NS electricity distribution). Extract from KK 28.03.2006.

5.3 Results electricity model II

5.161 The model specification has been tested under the three model categories COLS, DEA and SFA (linear/ log-linear/ translog/ normedlinear/ normedloglinear / effectslinear function; truncated normal distribution for the inefficiency error term), and in particular for DEA using the scale assumptions CRS (constant), DRS (decreasing), NDRS (non-decreasing) and VRS (variable) returns to scale. Specific runs were also made with a bias-corrected DEA model, included here as a confidence interval [c1, c2]. Only Farrell (multiplicative scores) results are presented here. A summary of the preliminary results for the sample data is provided in Table 5-2.

Table 5-2 Results for electricity model II using various assumptions (328 firms).

Model	Mean	Std.dev	# (≥1)	Min
d_cols_far	-49.045	74.514	1	-639.628
d_fdh_far	0.982	0.055	237	0.604
d_dea_far_vrs	0.825	0.142	53	0.401
d_dea_far_drs	0.822	0.143	51	0.400
d_dea_far_ndrs	0.798	0.140	33	0.401
d_dea_far_crs	0.796	0.140	31	0.400
d_dea_sup_far_vrs	1.011	2.530	48	0.401
d_dea_sup_far_drs	1.007	2.530	46	0.400
d_dea_sup_far_ndrs	0.826	0.263	33	0.401
d_dea_sup_far_crs	0.822	0.262	31	0.400
d_dea_far_vrs_biascorr	0.766	0.122	0	0.377
d_dea_far_vrs_biascorr_c1	0.728	0.113	0	0.356
d_dea_far_vrs_biascorr_c2	0.812	0.140	0	0.395
d_dea_far_crs_biascorr	0.749	0.124	0	0.377
d_dea_far_crs_biascorr_c1	0.721	0.118	0	0.359
d_dea_far_crs_biascorr_c2	0.783	0.136	0	0.395
d_orderm_far	1.072	0.126	251	0.669
d_dea_shell_far[, s]	0.987	0.685	96	0.451
d_dea_shell_far[, s]	0.825	0.142	53	0.401
d_dea_shell_far[, s]	1.114	0.769	138	0.558
d_sfa_linear_far	0.849	0.276	189	0.026
d_sfa_loglinear_far	0.778	0.120	0	0.001
d_sfa_translog_far	0.860	0.110	0	0.386
d_sfa_normedlinear_far	0.968	0.144	259	0.025
d_sfa_normedloglinear_far	0.980	0.000	0	0.978
d_sfa_effectslinear_far	0.855	0.275	187	0.022
d_dea_far_se	0.966	0.049	36	0.596

5.4 Descriptive OLS results

Standard model

5.162 The OLS results for electricity model II are given below, indicating an adjusted R2 of 0.9837 on 328 valid observations, corresponding to 99% of delivered volume.

```
Residuals:
      Min        1Q    Median        3Q       Max
-95821484   -966152   -374847    116481  93509838

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      124933.21  723461.39   0.173  0.86301
yMeters.hs       -72116.53   32226.59  -2.238  0.02593 *
yMeters.ms         9466.31    2221.88   4.260 2.70e-05 ***
yMeters.ns          126.86      12.41  10.223  < 2e-16 ***
yPeakload.hs      25606.00    8285.31   3.091  0.00218 **
yPeakload.hs_ms  -32632.88    6196.28  -5.267 2.58e-07 ***
yPeakload.ms     111804.68   34046.93   3.284  0.00114 **
yPeakload.ms_ns -113076.19   73123.72  -1.546  0.12302
yPeakload.ns      36297.85   65337.00   0.556  0.57891
yDg.power.total    4484.21    3727.34   1.203  0.22986
yArea.hs           2055.94     795.09   2.586  0.01016 *
yArea.ms          16214.28    7124.67   2.276  0.02353 *
yArea.1.ns       -11175.90    7303.50  -1.530  0.12697
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11920000 on 315 degrees of freedom
Multiple R-Squared: 0.9843,     Adjusted R-squared: 0.9837
F-statistic:  1648 on 12 and 315 DF,  p-value: < 2.2e-16
```

5.163 Below, we test the sensitivity of the electricity model above by adding one-by-one candidate structural variables from the set of groups of structural variables.

Soiltype

5.164 Below are the results for OLS integration of soiltype variables, here transformed into dummy variables for the dominant soiltype for the operator (from the variable zSoil.1m.maxNEW). As seen, none is significant.

Add-one term for group Soiltype

Df	Sum of Sq	RSS	AIC	F Value	Pr(F)
<none>	NA	4.47E+16	10701	NA	NA
zSoiltype.1	2.42E+14	4.45E+16	10702	1.70	0.193
zSoiltype.2	2.64E+11	4.47E+16	10703	0.00	0.966
zSoiltype.3	1.53E+14	4.46E+16	10702	1.08	0.300
zSoiltype.4	5.77E+14	4.42E+16	10699	4.10	0.044
zSoiltype.5	1.71E+14	4.46E+16	10702	1.20	0.274
zSoiltype.6	4.56E+12	4.47E+16	10703	0.03	0.858
zSoiltype.7	2.77E+11	4.47E+16	10703	0.00	0.965

Building

5.165 Below are the results for OLS integration of variables related to the number of buildings in the NS area served. The variables zBuildings.ns is significant at the 0.001 level. However, it is essentially a complement to the variable yArea.1.ns, which was judged to be more reliable.

Add-one term for group Building

Df	Sum of Sq	RSS	AIC	F Value	Pr(F)
<none>	NA	4.47E+16	10701	NA	NA
zBuildings.ns	1.34E+16	3.13E+16	10587	134.16	0.000

Energy Input

5.166 Below are the results for OLS integration of energy variables related to delivered energy. As seen, all are significant, but the corresponding coefficients are low and the underlying data material is not consistently of high quality.

Add-one term for group Energy

Df	Sum of Sq	RSS	AIC	F Value	Pr(F)
<none>	NA	4.47E+16	10701	NA	NA
zEnergy.del.lower_grid.hs	1.69E+16	2.78E+16	10548	190.75	**0.000**
zEnergy.del.lower_grid.hs_ms	7.55E+14	4.40E+16	10698	5.39	0.021
zEnergy.del.lower_grid.del.ms	2.60E+16	1.87E+16	10417	437.51	**0.000**
zEnergy.del.lower_grid.ms_ns	2.82E+16	1.66E+16	10377	534.08	**0.000**
zEnergy.del	3.15E+16	1.33E+16	10305	744.18	**0.000**
zEnergy.del.hs	1.91E+16	2.57E+16	10521	233.53	**0.000**
zEnergy.del.hs_ms	3.15E+16	1.33E+16	10304	745.42	**0.000**
zEnergy.del.ms	4.78E+15	4.00E+16	10666	37.54	**0.000**
zEnergy.del.ms_ns	8.22E+15	3.65E+16	10637	70.68	**0.000**
zEnergy.del.ns	1.57E+14	4.46E+16	10702	1.11	0.294

Peakload (other)

5.167 Below are the results for OLS integration of the peakload variable that is not included in the base model, zPeakload.max, significant at the 0.001-level.

Add-one term for group Peakload

Df	Sum of Sq	RSS	AIC	F Value	Pr(F)
<none>	NA	4.47E+16	10701	NA	NA
zPeakload.max	3.12E+16	1.36E+16	10312	719.94	**0.000**

Height and tilt

5.168 Below are the results for OLS integration of the variables related to absolute height, height difference and slope. None is significant.

Add-one term for group Heigh

Df	Sum of Sq	RSS	AIC	F Value	Pr(F)
<none>	NA	1.79E+16	10401	NA	NA
zHeight	7.01E+10	1.79E+16	10403	0.00	0.972
zHeightdiff	7.86E+12	1.79E+16	10403	0.14	0.711
zTilt	2.82E+13	1.79E+16	10403	0.50	0.482

Location East-West

5.169 Below are the results for OLS integration of a dummy variables related to the location in the new or old Bundesländer. The variable is not significant.

Add-one term for group Loc Loc

Df	Sum of Sq	RSS	AIC	F Value	Pr(F)
<none>	NA	7.97E+15	10135	NA	NA
zLoc.hs	8.43E+13	7.88E+15	10134	3.36	0.068
zLoc.hs_ms	8.43E+13	7.88E+15	10134	3.36	0.068
zLoc.ms	8.43E+13	7.88E+15	10134	3.36	0.068
zLoc.ms_ns	8.43E+13	7.88E+15	10134	3.36	0.068
zLoc.ns	8.43E+13	7.88E+15	10134	3.36	0.068

DG power

5.170 The impact of a further decomposition of the decentralized generation is tested through the variables zDg.power below. The variables for HS and MS are are significant, the first indicating a cost-decreasing effect on HS-level, the second negative. Note that the sum of the DG power is included in the base model.

Add-one term for group Seal

Df	Sum of Sq	RSS	AIC	F Value	Pr(F)
<none>	NA	4.47E+16	10701	NA	NA
zDg.power.hs	3.62E+15	4.11E+16	10676	27.60	**0.000**
zDg.power.ms	4.12E+15	4.06E+16	10672	31.84	**0.000**
zDg.power.ns	2.95E+13	4.47E+16	10703	0.21	0.649

Meters

5.171 The residual variables related to the number of meters, namely the total number of meters and the number of meters at transforming levels HS/MS and MS/NS, respectively, are significant at the 0.001-level. However, as the number of meters at voltage levels HS, MS and NS already are included, the effects are mainly absorbed by the former variables.

Add-one term for group Meter

Df	Sum of Sq	RSS	AIC	F Value	Pr(F)
<none>	NA	4.47E+16	10701	NA	NA
zMeters.tot	4.91E+15	3.98E+16	10665	38.72	0.000
zMeters.hs_ms	1.47E+16	3.00E+16	10573	153.54	0.000
zMeters.ms_ns	5.44E+15	3.93E+16	10661	43.45	0.000

Population

5.172 The population in the area of HS, MS, and HS/MS operations is shown to be significant at the 0.001-level, whereas a corresponding set of density variables (naturally) are insignificant in the OLS-setting. Varying signs indicate the high level of multi-colinearity in the group.

Add-one term for group Population

Df	Sum of Sq	RSS	AIC	F Value	Pr(F)
<none>	NA	4.47E+16	10701	NA	NA
zPopulation.ms	6.37E+15	3.84E+16	10653	52.16	0.000
zPopulation.hs_ms	7.06E+15	3.77E+16	10647	58.82	0.000
zPopulation.hs	7.13E+15	3.76E+16	10646	59.57	0.000
zPopulationDensity.hs	1.69E+14	4.46E+16	10702	1.19	0.277
zPopulationDensity.hs_ms	2.98E+12	4.47E+16	10703	0.02	0.885
zPopulationDensity.ms	1.35E+14	4.46E+16	10702	0.95	0.330
zPopulationDensity.ms_ns	1.36E+14	4.46E+16	10702	0.96	0.328
zPopulationDensity.ns	1.39E+14	4.46E+16	10702	0.98	0.322

Connection points

5.173 Four indicators of connection density at voltage levels HS, MS and NS are tested, none is significant.

Add-one term for group Connect

Df	Sum of Sq	RSS	AIC	F Value	Pr(F)
<none>	NA	4.47E+16	10701	NA	NA
zConnpoidens.a2..ns	9.02E+13	4.47E+16	10703	0.63	0.426
zConnpoidens.a1..ns	6.15E+13	4.47E+16	10703	0.43	0.511
zConnpoidens.ms	1.82E+14	4.46E+16	10702	1.28	0.259
zConnpoidens.hs	9.43E+13	4.46E+16	10703	0.66	0.416

Area measures

5.174 Detailed indicators of landuse (dense city, industry, infrastructure, rocky ground, waterways, other) have been tested in addition to the three area variables aleary in the base model. zAreaTraffic, zAreaWhatev and zAreaWater turn out to be significant at the 0.001-level. However, in a full inclusion in an OLS model, only zAreaWater maintain its significance (the others have significance levels above 0.5), acting as a negative correction factor to the other area-variables.

Add-one term for group Area

Df	Sum of Sq	RSS	AIC	F Value	Pr(F)
<none>	NA	1.074E+16	10233	NA	NA
zAreainnercity	8.83E+12	1.074E+16	10235	0.26	0.612
zAreacity	1.49E+12	1.074E+16	10235	0.04	0.835
zAreaIndustry	3.97E+13	1.071E+16	10234	1.16	0.281
zAreaTraffic	1.06E+15	9.689E+15	10202	34.21	**0.000**
zAreaWhatev	1.46E+15	9.288E+15	10188	49.24	**0.000**
zAreaVeget	1.06E+12	1.074E+16	10235	0.03	0.861
zAreaRock	6.08E+11	1.074E+16	10235	0.02	0.894
zAreaWater	3.08E+15	7.663E+15	10125	126.26	**0.000**

Model network length

5.175 For a smaller subsample of 35 firms, the calculated model network length showed to be a significant regressor at NS-level, not at MS-level.

Add-one term for group Model

Df	Sum of Sq	RSS	AIC	F Value	Pr(F)
<none>	NA	2.32E+12	7464	NA	NA
zModel.LengthNS.detailed	5.25E+11	1.79E+12	7382	92.05	**0.000**
zModel.LengthNS.aggr	1.94E+11	2.12E+12	7438	28.71	**0.000**
zModel.LengthMS.detailed	5.88E+09	2.31E+12	7465	0.80	0.372

Calculated network length

5.176 Indicators of calculated network length showed significance at the 0.001-level.

Add-one term for group Length

Df	Sum of Sq	RSS	AIC	F Value	Pr(F)
<none>	NA	4.47E+16	10701	NA	NA
zLength.calc.net.a2.ns	2.04E+16	2.44E+16	10504	262.74	**0.000**
zLength.calc.net.a1.ns	1.59E+16	2.89E+16	10560	172.45	**0.000**
zLength.calc.net.ms	3.13E+16	1.35E+16	10310	728.77	**0.000**
zLength.calc.net.hs	1.11E+16	3.36E+16	10609	104.14	**0.000**

5.5 T1 Results

5.177 The subadditivity hypothesis has been tested using an extension of the Bogetoft and Wang (2005) method for estimation of synergies in multi-output production. The obtained Farrell efficiency estimates for voltage-separated operation vs integrated operation are given in Table 5-3 below for the same set of electricity DMU as in the model estimation in Chapter 6. The model framework is equivalent to the model used in Chapter 6 also except that to adequately capture likely cost separations, we have extended the framework with a decomposition of DGpower according to voltage levels. In this way, the total services of a DSO can be split into hs, ms and ls activities, and comparing the efficient cost level of a joint operation with the efficient cost level of the three separate operations, as it is done in the last column, we can evaluate the potential savings from synergies. As seen below, the non-parametric estimates indicate substantial subadditivity. This is despite of the fact that the underlying model does not allow for global economies of scale. Indeed, since the technology exhibits close to crs properties, the potential saving is driven mainly by economies of scope. In addition, the bias towards larger cost levels is larger in the separate models than in the joint model. Still, the considerable difference suffices to confirm subadditivity in the DSO activities.

Table 5-3 T1 Subadditivity test Electricity

	d_dea_far_vrs_full	d_dea_far_vrs_hs	d_dea_far_vrs_ms	d_dea_far_vrs_ns	dea_joint_to_sep_cost
mean	0.789	0.171	0.635	0.626	0.572
std dev	0.163	0.212	0.203	0.199	0.093
max	1.000	1.000	1.000	1.000	1.000
min	0.350	0.002	0.000	0.000	0.333

5.178 The use of a joint model covering all voltage levels may raise the question if such an approach is biased against or in favor of DSOs with multiple levels. Conceptually, this should not be the case. Subadditivity suggests that a DSO that is large and/or covers multiple levels may have a cost advantage but if the model is well-specified, this is captured by the estimated frontier and we should see no systematical difference in the distribution of efficiency scores for DSO covering more or less levels. (Of course, there may be some relationship related to other factors that are correlated with the number of levels covered. Thus for example, if the larger DSOs are in general having less variation in efficiency than the smaller units, we may see smaller variations also for DSOs covering more levels if they are in general larger.)

5.179 To check for the impact of levels, we have done several parametric and non-parametric tests. Figure 5-2 illustrates how the DEA superefficiency crs efficiency scores depend on the number of levels covered.

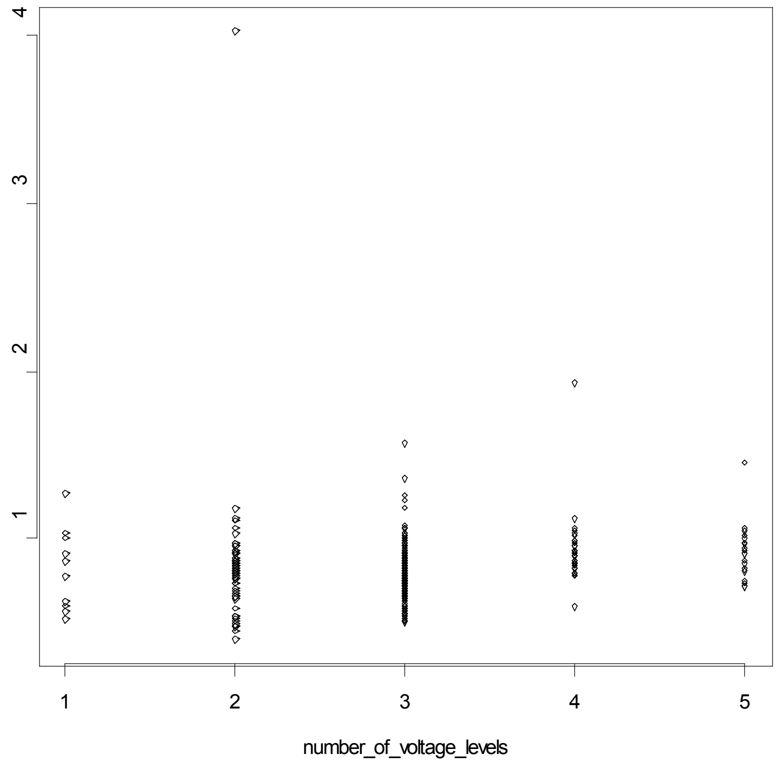

Figure 5-2 Impact of number of voltage levels

5.180 We see that the number of levels may have some although no very systematic impact on the efficiencies. This is confirmed also by several tests.

5.181 A linear regression of the efficiencies on the number of voltage levels show no strongly significant impact:

```
Coefficients:
                         Estimate Std. Error t value Pr(>|t|)
(Intercept)               0.72104    0.05192  13.887   <2e-16 ***
number_of_voltage_levels  0.03344    0.01686   1.983   0.0482 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2488 on 326 degrees of freedom
Multiple R-Squared: 0.01192,    Adjusted R-squared: 0.008893
F-statistic: 3.934 on 1 and 326 DF,  p-value: 0.04816
```

5.182 This confirms that the relationship is at the least not a simple monotone one with more levels leading to higher efficiencies.

5.183 To look for an arbitrary relationship between the number of levels and the efficiency levels, we can do a one-side ANOVA or a non-parametric Kruskal-Vallis test for differences in the mean efficiency in the different groups:

```
Coefficients: (1 not defined because of singularities)
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    0.924718   0.054977  16.820   <2e-16 ***
one.levels   -0.105580   0.095223  -1.109   0.2684
two.levels   -0.091996   0.062984  -1.461   0.1451
three.levels -0.138060   0.057661  -2.394   0.0172 *
four.levels   0.009555   0.069285   0.138   0.8904
five.levels         NA         NA      NA       NA
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2459 on 323 degrees of freedom
Multiple R-Squared: 0.04387,    Adjusted R-squared: 0.03203
F-statistic: 3.705 on 4 and 323 DF,  p-value: 0.005767


Kruskal-Wallis rank sum test

data:  d_dea_sup_far_crs by number_of_voltage_levels
Kruskal-Wallis chi-squared = 35.6209, df = 4, p-value = 3.463e-07
```

5.184 Again, this confirms that there may be some although weak and certainly not a monotone relationship. We see that DSOs with three levels may be slight disfavored.

5.185 A possible explanation of the found relationships – although they are weak – could be the bias resulting from the number of comparable observations. We have therefore repeated the above using DEA bias corrected efficiencies. The results however are almost identical and are therefore not included in the report.

5.6 T2 Results

5.186 The age test is performed both against the dependent variable (xCosts.dir) here and against the obtained score below. Age is not significant for electricity model II

Add-one term for group Age

Df	Sum of Sq	RSS	AIC	F Value	Pr(F)
<none>	NA	1.14E+16	10252	NA	NA
zAge	7.25E+12	1.14E+16	10254	0.20	0.655

5.7 T3 Results

5.187 For all used model specifications, the explanatory value of using true capital expenditure or a standardized measure of capital expenditure (zCAPEX.stand) has been investigated in Table 5-4 below. The first table shows the impact of CAPEX on the scores, none of the specifications has any coefficient larger than 0.02 significant at the 0.001-level. The relevant metric is the t-value for zCAPEX and zCAPEX.stand when applied to the superefficiency specifications, for other specifications truncated regressions have to be used. Besides the superefficiency results, CAPEX does seem to correlate positively for the normal DEA models, as opposed to CAPEX.stand, which means that larger DSO would tend to higher scores. The occurrence of these indications for all returns to scale prompts at the interest to further investigate the CAPEX as to eradicate any risk of bias.

Table 5-4 T3 CAPEX test electricity.

Model	Mean xCAPEX* Estimate	Std. Error	t value	Pr(>|t|)
d_dea_sup_far_vrs	0.057	9.865E-10	2.762	**0.006**
d_dea_sup_far_drs	0.058	9.870E-10	2.785	**0.006**
d_dea_sup_far_ndrs	0.015	2.726E-10	2.530	0.012
d_dea_sup_far_crs	0.015	2.728E-10	2.580	0.011
d_sfa_linear_far	0.005	4.277E-10	0.582	0.562
d_sfa_loglinear_far	0.006	1.716E-10	1.736	0.084
d_sfa_translog_far	0.008	1.605E-10	2.366	0.019
d_sfa_normedlinear_far	-0.012	2.336E-10	-2.442	0.016
d_sfa_normedloglinear_far	0.000	6.842E-13	2.178	0.031
d_sfa_effectslinear_far	0.012	4.145E-10	1.325	0.187

Table 5-5 T3 CAPEX.standard test electricity.

Model	Mean xCAPEX.stand* Estimate	Std. Error	t value	Pr(>|t|)
d_cols_far	63.819	1.766E-05	0.520	0.604
d_fdh_far	-0.045	1.052E-08	-0.615	0.539
d_dea_far_vrs	0.045	3.177E-08	0.205	0.838
d_dea_far_drs	0.057	3.179E-08	0.257	0.797
d_dea_far_ndrs	0.095	3.140E-08	0.434	0.665
d_dea_far_crs	0.106	3.132E-08	0.488	0.626
d_dea_sup_far_vrs	-2.086	8.151E-07	-0.368	0.713
d_dea_sup_far_drs	-2.072	8.152E-07	-0.366	0.715
d_dea_sup_far_ndrs	0.109	3.774E-08	0.415	0.679
d_dea_sup_far_crs	0.123	3.756E-08	0.472	0.638
d_dea_far_vrs_biascorr	-0.015	2.691E-08	-0.079	0.937
d_dea_far_vrs_biascorr_c1	-0.040	2.444E-08	-0.233	0.816
d_dea_far_vrs_biascorr_c2	0.035	3.117E-08	0.163	0.871
d_dea_far_crs_biascorr	0.067	2.758E-08	0.350	0.727
d_dea_far_crs_biascorr_c1	0.065	2.604E-08	0.360	0.720
d_dea_far_crs_biascorr_c2	0.087	3.057E-08	0.409	0.683
d_orderm_far	0.033	2.817E-08	0.170	0.865
d_dea_shell_far[, s]	0.045	3.177E-08	0.205	0.838
d_dea_shell_far[, s]	0.209	8.190E-08	0.368	0.714
d_dea_shell_far[, s]	-0.149	1.090E-07	-0.196	0.845
d_sfa_linear_far	0.294	6.670E-08	0.635	0.526
d_sfa_loglinear_far	0.194	2.887E-08	0.968	0.335
d_sfa_translog_far	0.117	2.396E-08	0.704	0.482
d_sfa_normedlinear_far	0.104	2.618E-08	0.573	0.567
d_sfa_normedloglinear_far	0.002	8.948E-11	3.549	**0.001**
d_sfa_effectslinear_far	0.251	6.283E-08	0.574	0.567

Ratio OPEX/CAPEX: Electricity model

5.188 To test the potential bias introduced by abnormal proportions of capital costs, the ratio of capital costs to total costs was used as a regressor to SFA scores, DEA superefficiency scores for all units and for frontier units in DEA-CRS. Results are present in Table 5-6. We test a one-way Analysis of variance (ANOVA) where the ratio of Capex to total cost is the predictor and the superefficiency score is the response variable. ANOVA is a standard technique to determine whether a predictor has a specific influence at another variable. None of the coefficients are significant at the 0.01-level and as visible in Figure 5-3 there is no specific pattern for the frontier units. However, some indicative results for the normal DEA model e.g. DEA-CRS model, point at a negatively correlated relationship irrespective of scale assumption, meaning that firms with higher proportions capital costs have lower scores, the coefficient is about -0.22 (CRS, NDRS) and -0.30 (VRS, DRS). One may interpret this as a signal that activation policy, affecting the share of capital costs to total costs, needs to be taken into account to correct for current and future bias in the benchmarking. A standardized (re)activation policy is one way to address this problem.

Table 5-6 ANOVA test ratio CAPEX/TOTEX to scores electricity model II.

Supereff DEA-NDRS	Df	Sum Sq	Mean Sq	F value	Pr(>F)
I(xCAPEX/(xCAPEX + xOPEX))	1	0.0894	0.0894	2.6671	0.1041
Residuals	184	6.1643	0.0335		

Supereff DEA-CRS	Df	Sum Sq	Mean Sq	F value	Pr(>F)
I(xCAPEX/(xCAPEX + xOPEX))	1	0.0908	0.0908	2.7042	0.1018
Residuals	184	6.1774	0.0336		

SFA_translog_fa	Df	Sum Sq	Mean Sq	F value	Pr(>F)
I(xCAPEX/(xCAPEX + xOPEX))	1	0.03445	0.03445	2.9864	0.08564
Residuals	184	2.12259	0.01154		

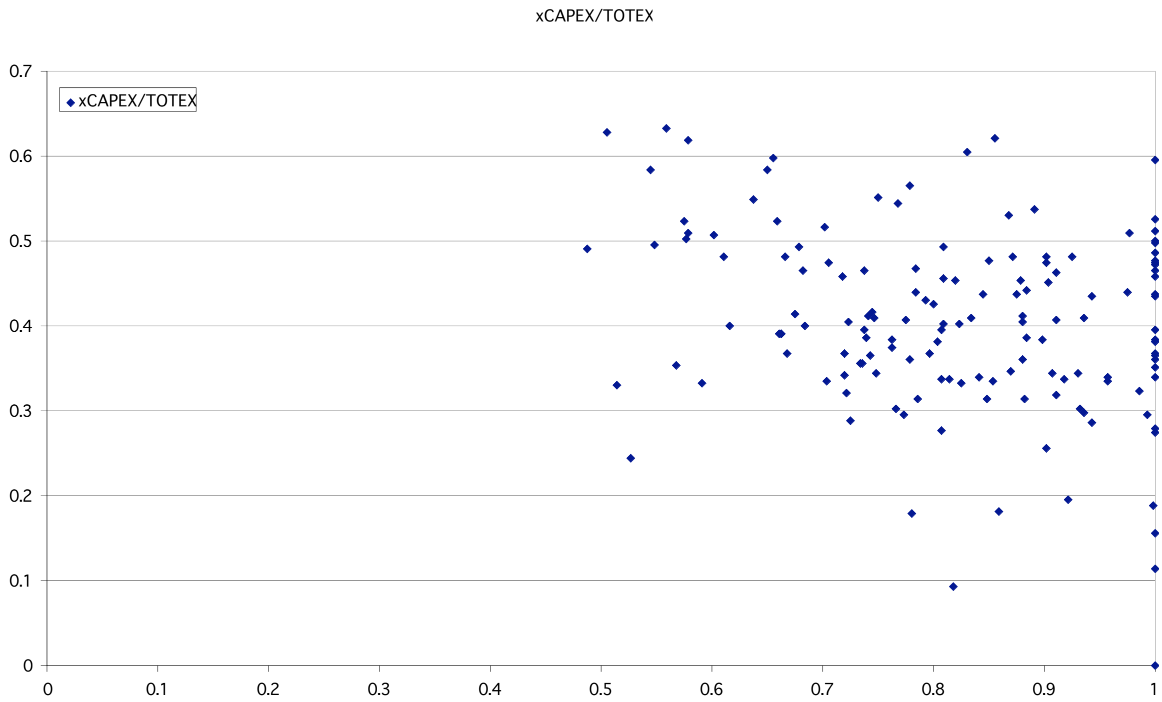

Figure 5-3 Correlation ratio CAPEX/TOTEX and DEA-CRS scores electricity.

Table 5-7 T3 CAPEX/TOTEX ratio test electricity model II.

Model	Estimate	Std. Error	t value	Pr(>|t|)
d_dea_sup_far_vrs	-0.426	0.278	-1.532	0.127
d_dea_sup_far_drs	-0.427	0.278	-1.535	0.126
d_dea_sup_far_ndrs	-0.185	0.113	-1.633	0.104
d_dea_sup_far_crs	-0.187	0.114	-1.644	0.102
d_sfa_linear_far	-0.078	0.176	-0.440	0.661
d_sfa_loglinear_far	0.029	0.071	0.405	0.686
d_sfa_translog_far	-0.115	0.067	-1.728	0.086
d_sfa_normedlinear_far	0.110	0.097	1.127	0.261
d_sfa_normedloglinear_far	0.000	0.000	-0.569	0.570
d_sfa_effectslinear_far	-0.145	0.171	-0.849	0.397

Asset intensity test (Normalized CAPEX)

5.189 An alternative approach that is sometimes mentioned is to consider the capital intensity measured over some proxy for assets. Here we implemented this as CAPEX per km of circuit length (lines and cables, equal weight). Two outliners were removed (DMU 218 and 220) with extremely short networks and incomparable values. The distribution of asset values is illustrated in Figure 5-4 below. The full results in Table 5-8 refute any suspicion of influence, the regressors are insignificant for all relevant models.

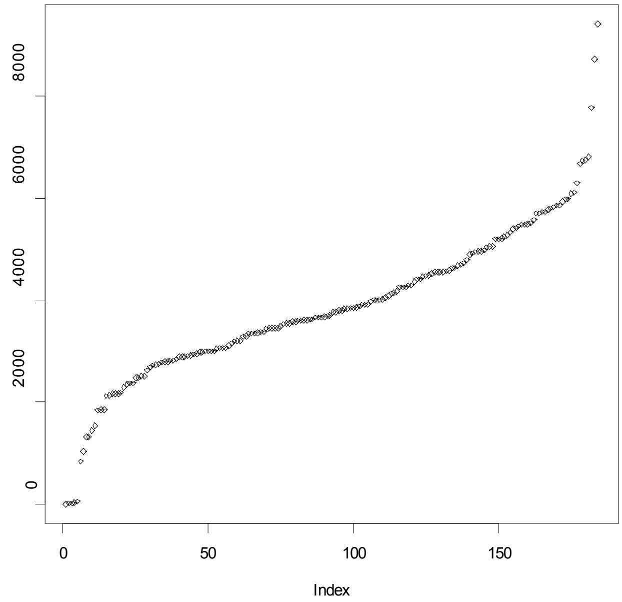

Figure 5-4 Normalized CAPEX per total circuit length (lines and cable) electricity model II.

Table 5-8 T3 Normalized CAPEX intensity test electricity model II.

Model	Estimate	Std. Error	t value	Pr(>|t|)
d_dea_sup_far_vrs	-1.791E-07	1.079E-06	-0.166	0.868
d_dea_sup_far_drs	-2.566E-07	1.079E-06	-0.238	0.812
d_dea_sup_far_ndrs	-6.846E-08	4.548E-07	-0.151	0.881
d_dea_sup_far_crs	-1.461E-07	4.553E-07	-0.321	0.749
d_sfa_linear_far	3.743E-07	7.016E-07	0.533	0.594
d_sfa_loglinear_far	3.256E-07	2.827E-07	1.152	0.251
d_sfa_translog_far	-2.482E-07	2.665E-07	-0.931	0.353
d_sfa_normedlinear_far	1.114E-07	3.892E-07	0.286	0.775
d_sfa_normedloglinear_far	4.037E-10	1.136E-09	0.355	0.723
d_sfa_effectslinear_far	3.102E-07	6.827E-07	0.454	0.650

Frontier units electricity

5.190 The frontier units in DEA-CRS do not have a statistically different proportion of capital cost to total cost, as documented in the test Table 5-9 below. The t-test determines the difference of the means from two samples that are assumed to have the same variance in the distribution. Qualitatively equivalent results can be obtained using z-test assuming different variances.

Table 5-9 t-test for means of CAPEX proportion for frontier vs non-frontier units, electricity model II

	CAPEX/TOTEX DEA-CRS=1	*CAPEX/TOTEX DEA-CRS<1*
Mean	0.420000	0.392643
Known Variance	0.008162	0.013915
Observations	21	164
Pooled variance	0.013287	
Hypothesized Mean Difference	0	
df	183	
t Stat	1.024023	
P(T<=t) one-tail	0.153588	
t Critical one-tail	1.653223	
P(T<=t) two-tail	0.307176	
t Critical two-tail	1.973012	

5.191 We have examined also if the cleaning of data based on outlier detection by extreme super-efficiencies may be related to the CAPEX evaluation. Data allows us to calculate CAPEX proportions for 10 out of the 21 units eliminated. The average CAPEX ratio, i.e.. CAPEX / (CAPEX+OPEX) among these is 0.369 while it is 0.393 in the full sample used. The corresponding t statistic is 0.31 with a two-sided p value of 0.76, i.e. there is no significant difference between the eliminated and included units. A non-parametric Kruskal-Wallis gives a chi-squared test statistic for the null hypothesis of similar means of 0.786 with df = 1, such that the p-value i.e. 0.375, i.e. again no indication of a biased outlier sample.

5.8 DEA results

5.192 More detailed results of some of the DEA models are illustrated in the figures below. The distributions of scores are comparable to the findings in other DEA models of electricity distribution internationally. The Salter diagrams show a fairly even mix in operating scale across the efficiency spectrum. Note also that the economic significance of the most inefficient firms is minor.

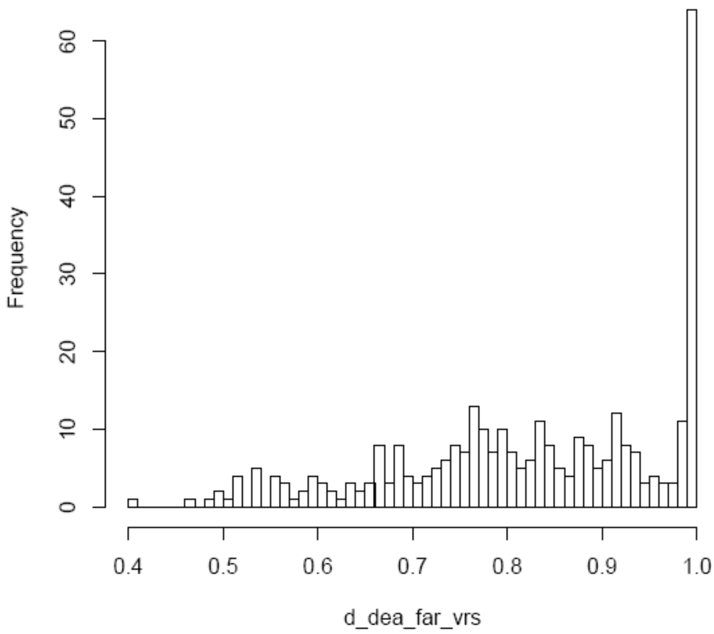

Figure 5-5 Histogram DEA-VRS results electricity model II.

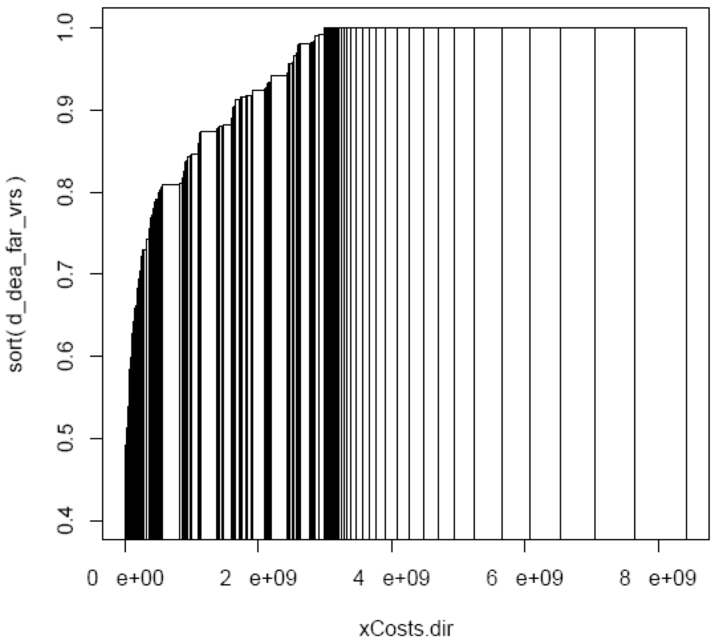

Figure 5-6 Salter diagram electricity model II under DEA-VRS

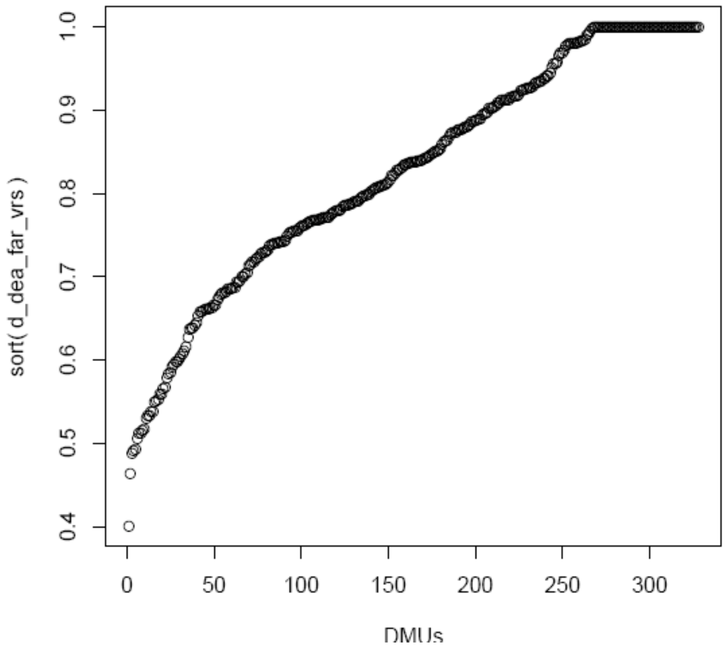

Figure 5-7 Distribution of scores DEA-VRS for the electricity model II

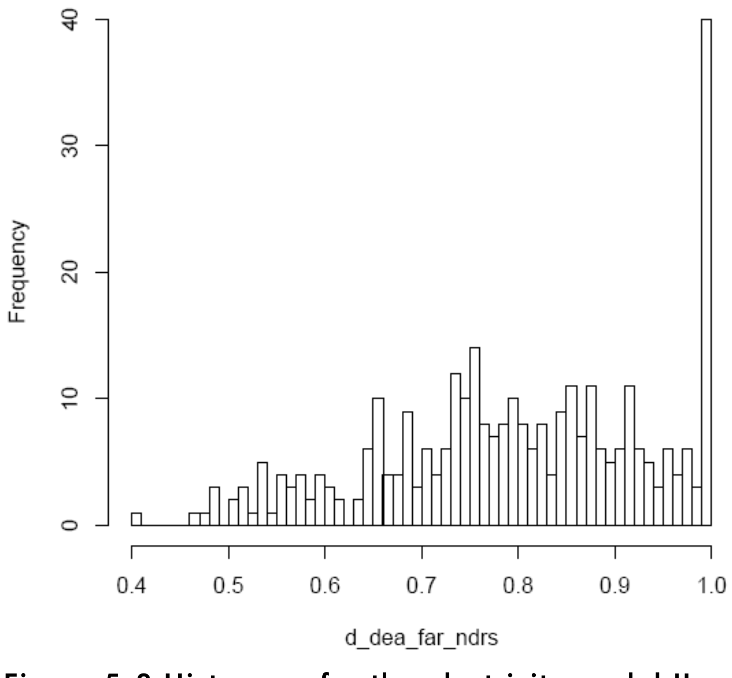

Figure 5-8 Histogram for the electricity model II under DEA-NDRS

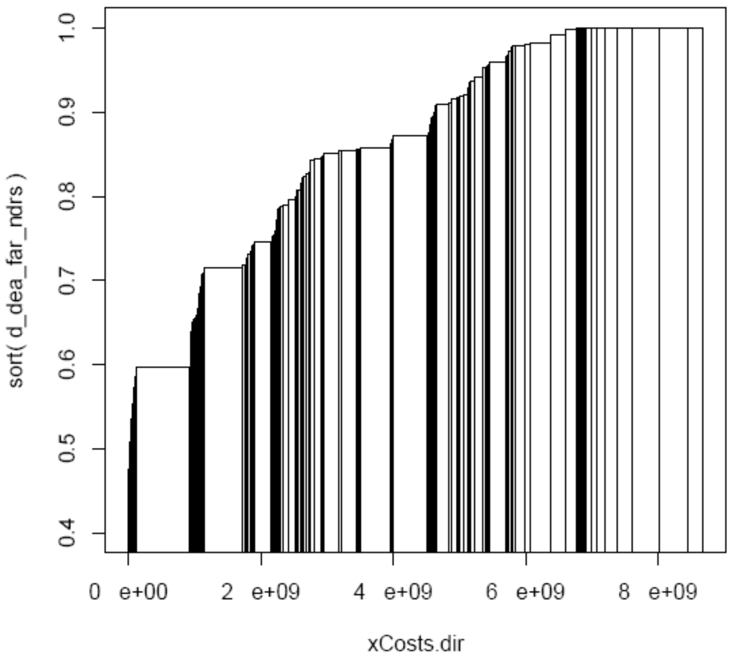

Figure 5-9 Salter diagram for the electricity model II under DEA-NDRS

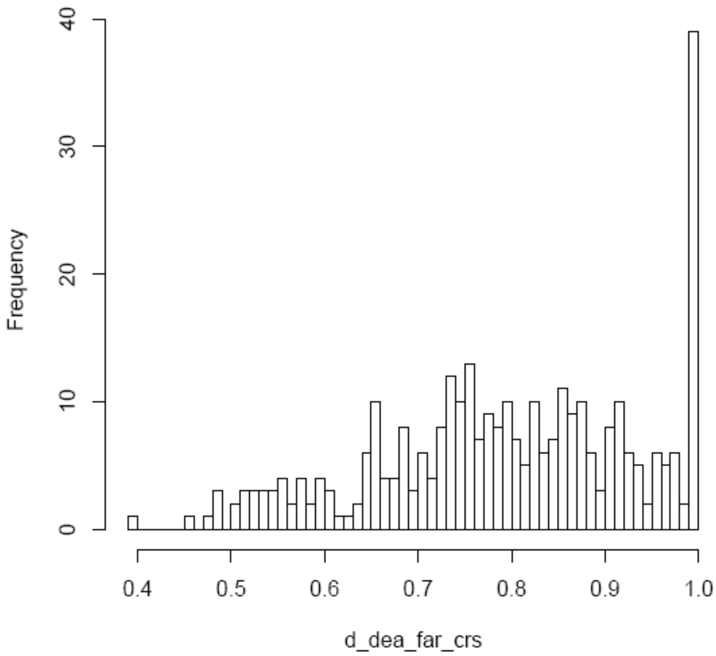

Figure 5-10 Histogram for the electricity model II under DEA-CRS.

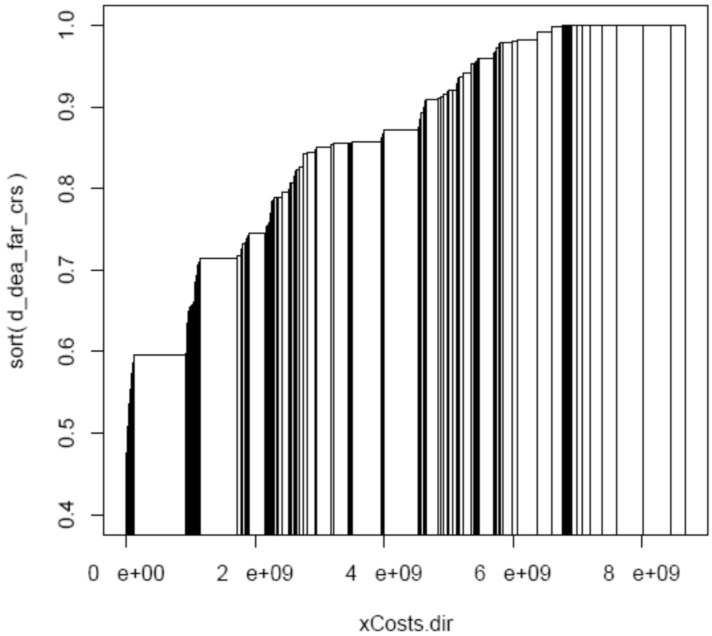

Figure 5-11 Salter diagram electricity model II DEA-CRS.

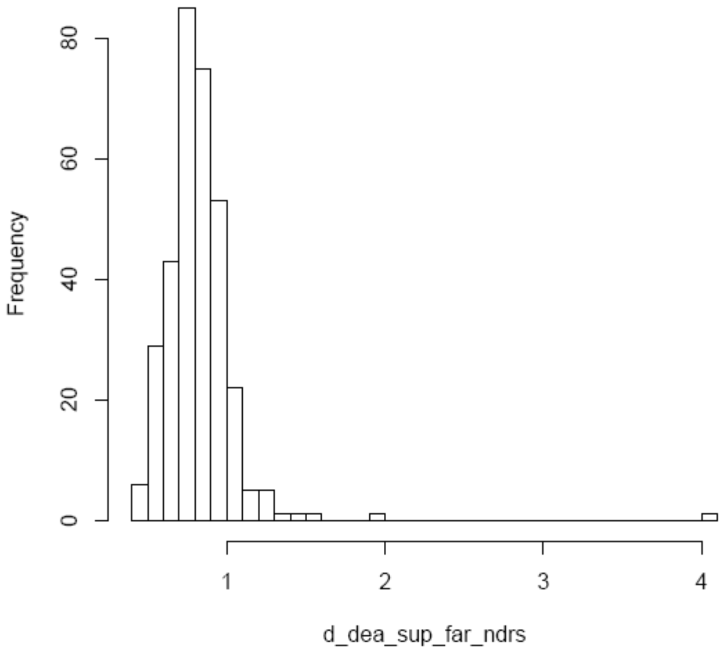

Figure 5-12 Histogram for the superefficiency DEA-NDRS model for electricty

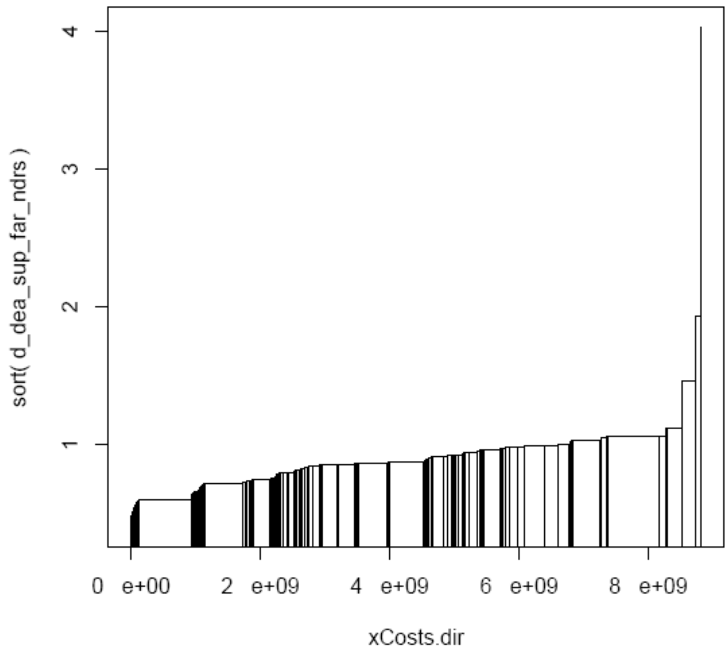

Figure 5-13 Salterdiagram for the superefficiency DEA-NDRS model for electricity

Scale efficiency results

5.193 The scale efficiency results in Population NS in Figure 5-14 indicate a rather linear technology with efficient units both for very small and very large firms. Firms operating over a span of scales (here measured as population (NS)) are detected as fully efficient. The scale efficiency concept is always somewhat complicated to interpret in a multi-output technology with different types of inputs. Variables that are strongly correlated to the service provision, such as meters, connections and population, explain well a fairly linear technology in e.g. line assets. However, variables related to capacity provision, such as peakload, DG power etc, relate to the dimensioning of the assets, which include some non-continuities in e.g. transformer sizes and cross-section areas. For larger DSOs, this may wash out as averages, but for smaller units the impact might be more visible.

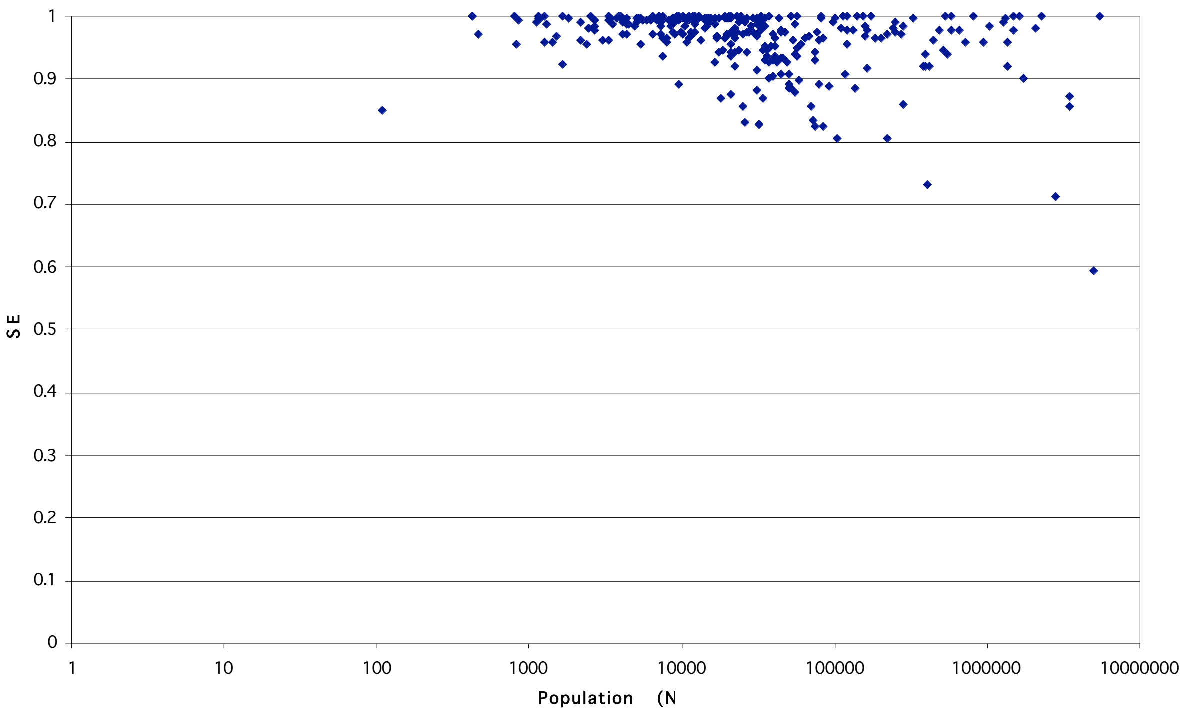

Figure 5-14 Scale efficiency estimates (in Population NS, logarithmic scale) for electricity model II DEA.

Bias corrected DEA results

5.194 Bias corrected scores are illustrated in Figure 5-15 - Figure 5-19 below.

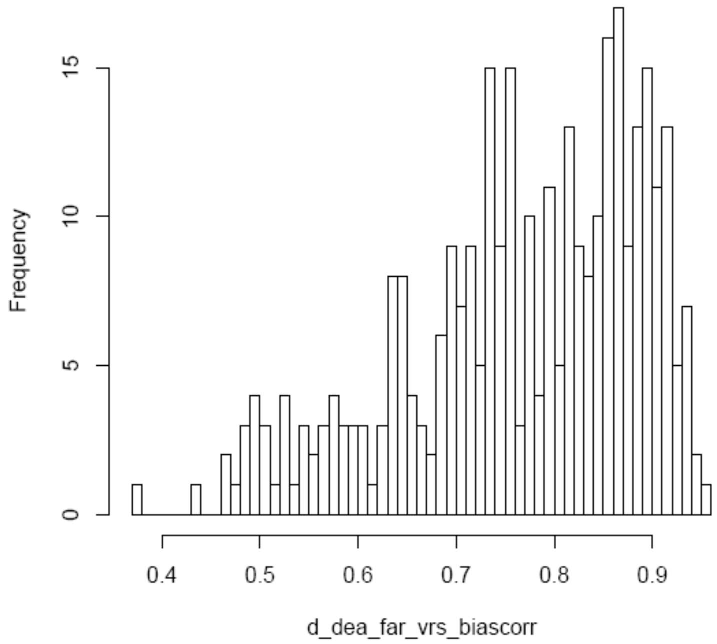

Figure 5-15 Histogram for DEA-VRS bias corrected scores for electricity

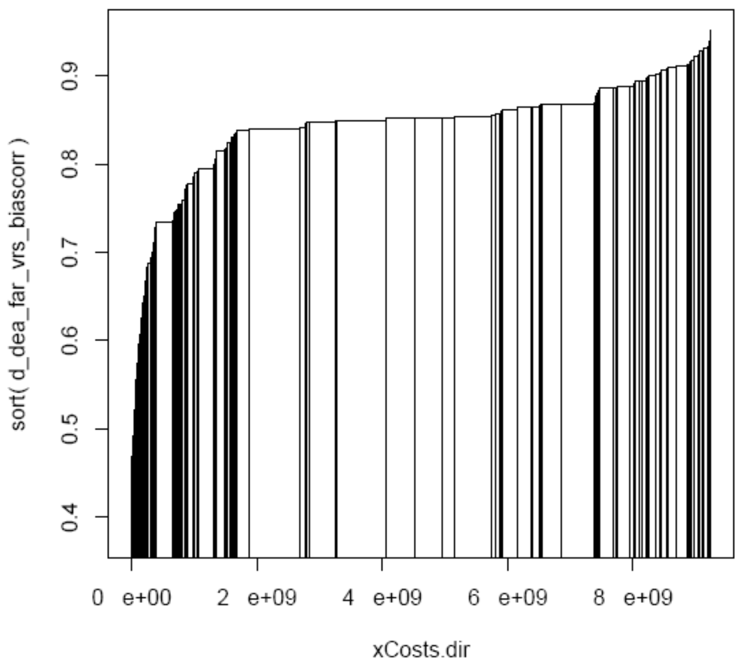

Figure 5-16 Salter diagram for the DEA-VRS biascorrected electricity model II

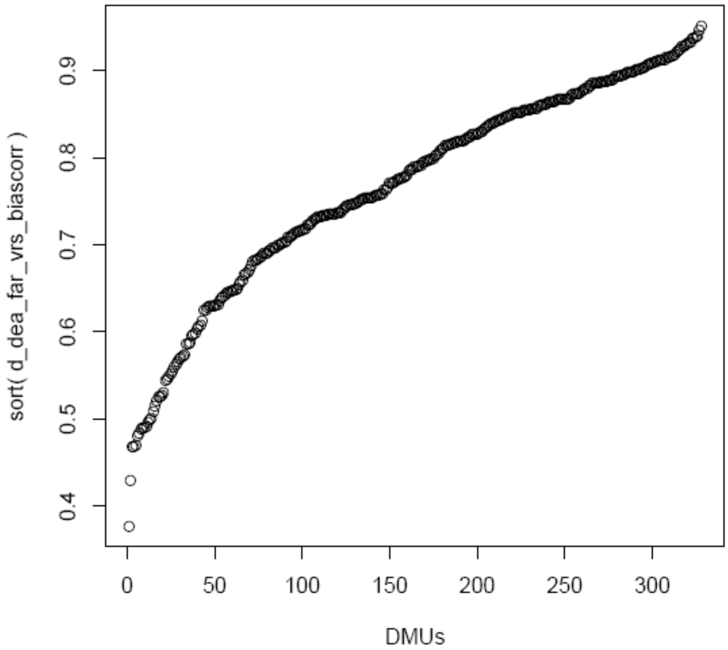

Figure 5-17 Distribution of biascorrected DEA-VRS scores for electricity

5.195 The correspondence between the normal DEA and the bias corrected model is good, as shown in Figure 5-19. However, as one would expect, the scores are lower in the bias corrected model, in particular for units with odd profiles.

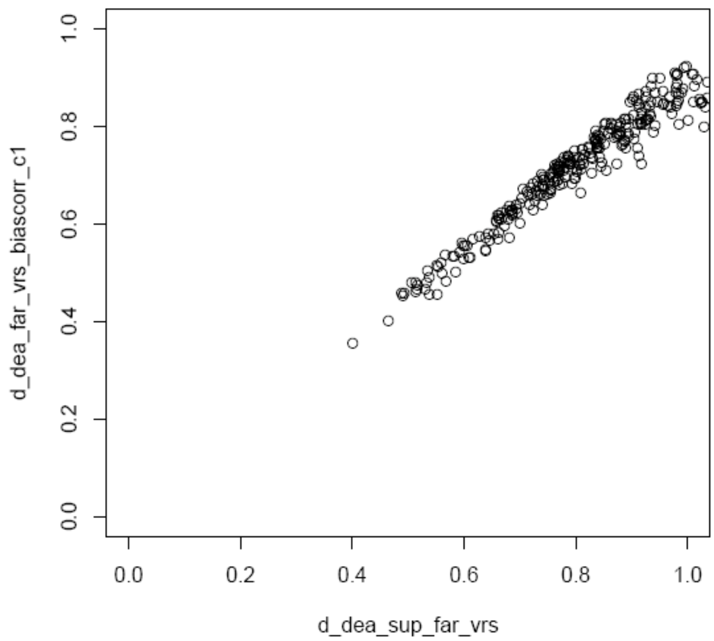

Figure 5-18 Correlation between the lower bound of the biascorrected DEA-VRS model and the superefficiency model DEA-VRS for the electricity model II

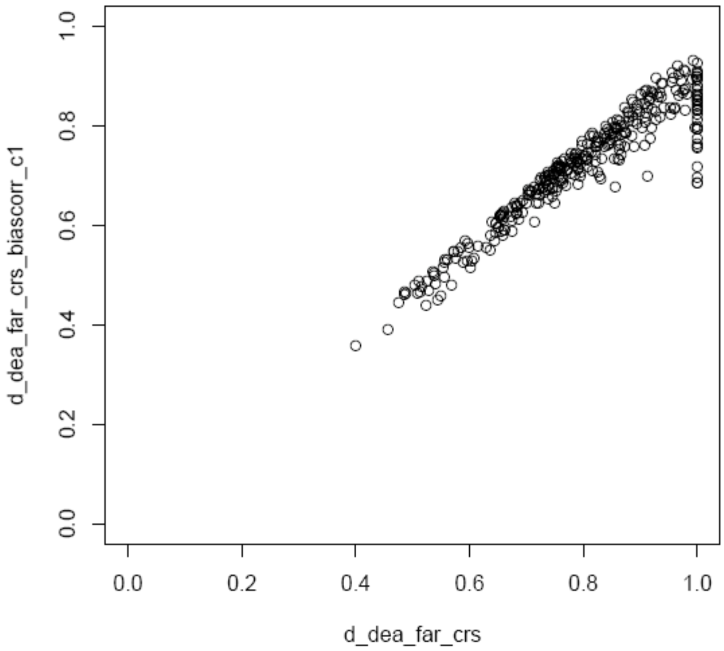

Figure 5-19 Correlation between Farrel scores for DEA-CRS and lower bound of bias corrected DEA.

Other non-parametric scores

5.196 The distribution of super-efficiency and order-m and shell scores are illustrated in Figure 5-20 - Figure 5-25below.

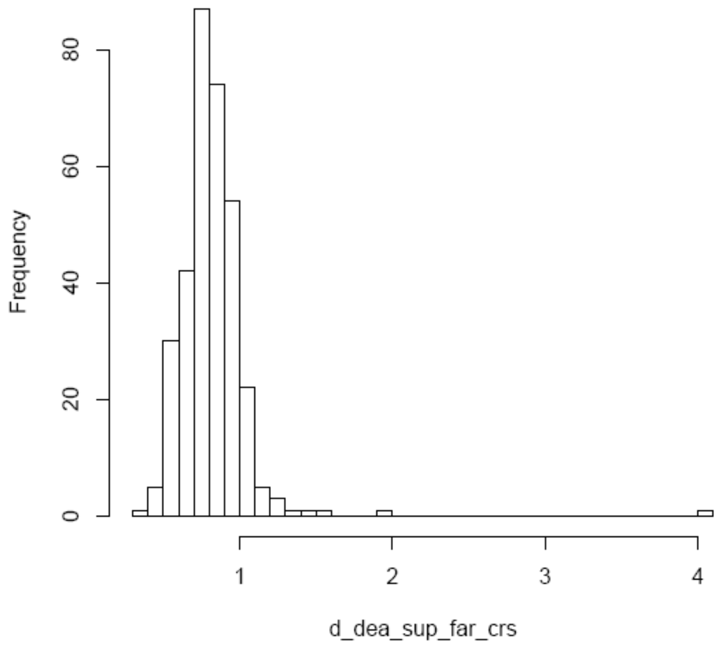

Figure 5-20 Histogram for the superefficiency DEA-CRS model for electricity

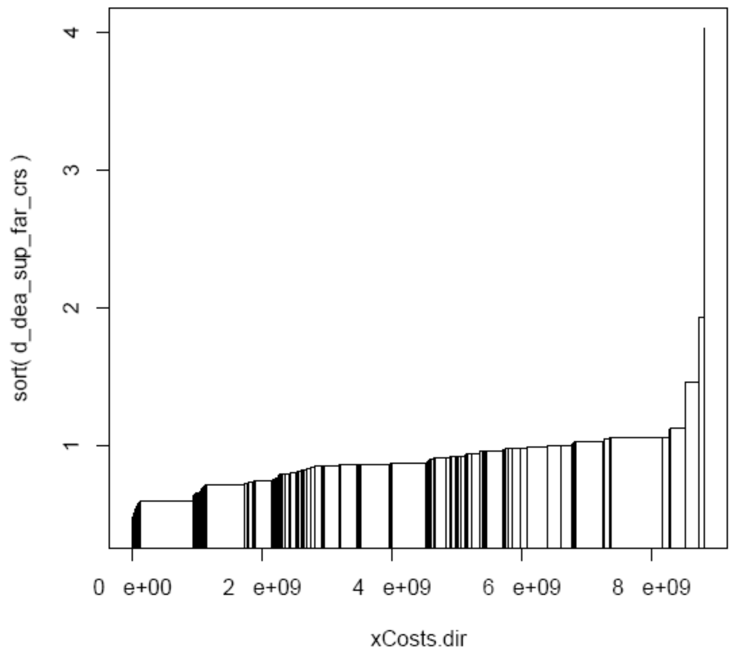

Figure 5-21 Salter diagram for the superefficiency DEA-CRS model for electricity

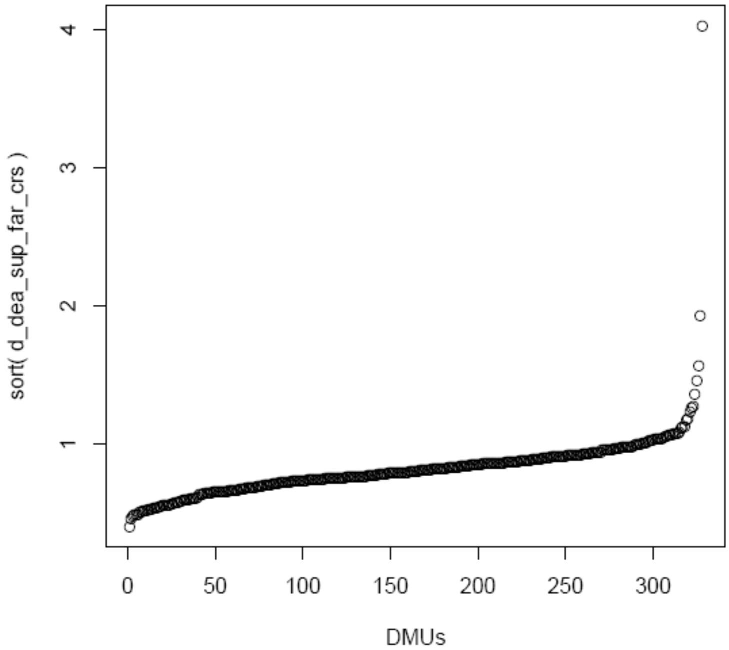

Figure 5-22 Distribution of scores for superefficiency DEA_CRS for electricity

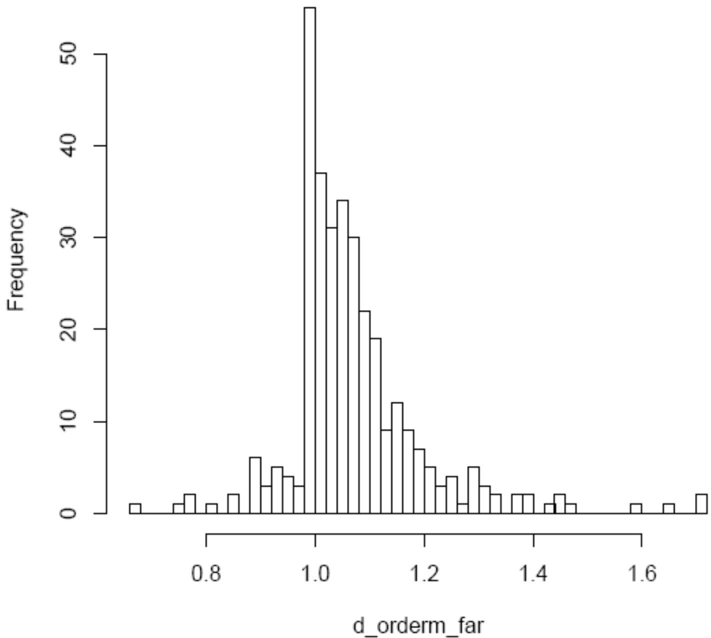

Figure 5-23 Histrogram for DEA order-m results for electricity

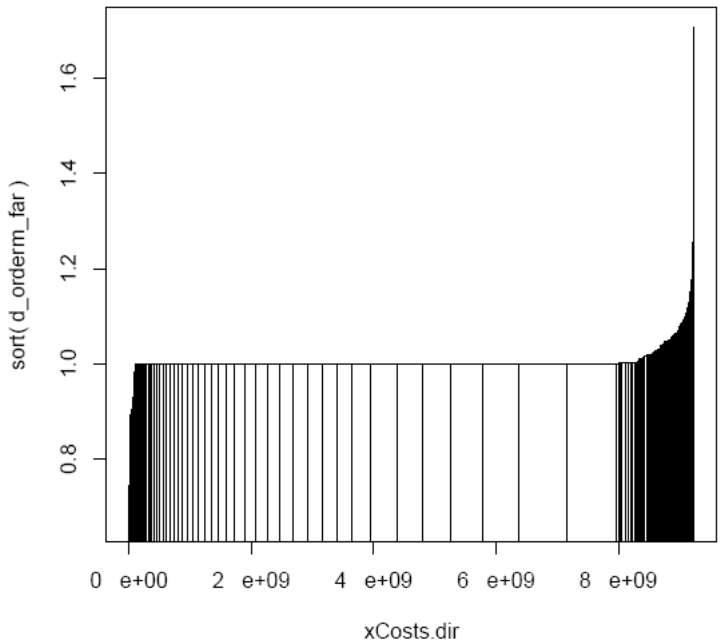

Figure 5-24 Salter diagram for DEA order-m results for electricity

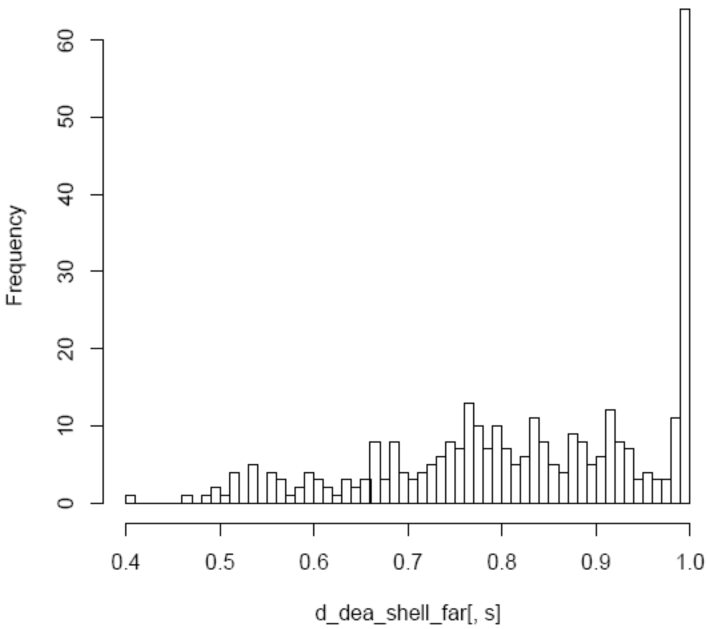

Figure 5-25 Histogram for DEA shell 1 results for electricity

5.9 SFA results

5.197 The results for the stochastic frontier analysis using a translog functional form in Figure 5-26 and Figure 5-27 show a uni-modal distribution with modes around 0.92. No unit is classified as fully costefficient, since the error term is fairly large for the first data set. Note that the general error term in SFA is monotonously decreasing in the number of observations (if they have the same variance).

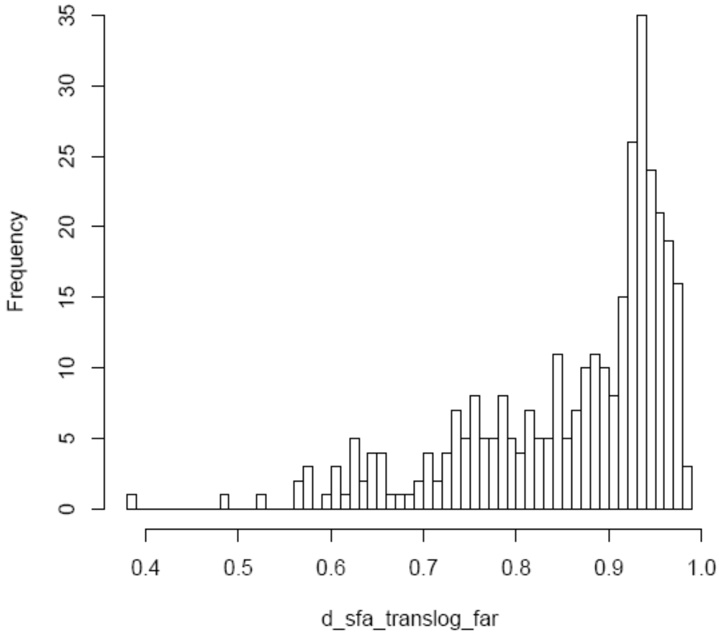

Figure 5-26 Results for electricity model II SFA translog.

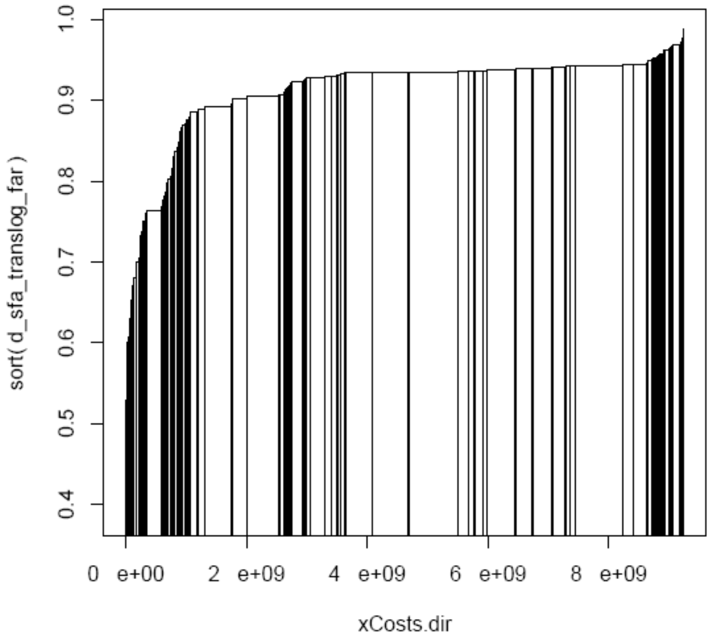

Figure 5-27 Salter diagram for electricity model II SFA translog.

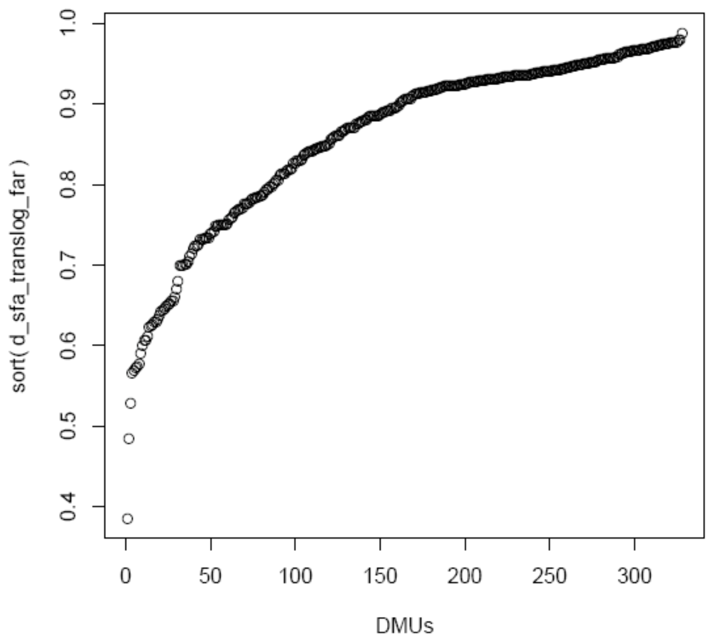

Figure 5-28 Distribution of SFA translog scores for the electricity model II.

5.10 Model consistency

5.198 The consistency of the model specification II can be deduced from the correlations between scores and rank orders across various models in Table 5-11, Table 5-10 and Table 5-12 below for some relevant models. Pearson correlation measures the level of dependency between the scores. Kendall's rank correlation provides a distribution free test of independence and a measure of the strength of dependence between two variables. Spearman's rank correlation is similar to Kendall, with a lower usefulness in case of rejected dependency (which is clearly not the case at hand here). Two findings are important. First, the specification gives unusually high correlation both in terms of scores and rank order between the non-parametric deterministic DEA and parametric stochastic frontier SFA. Second, the scale assumptions CRS or NDRS are close and mainly correct for some specific small operators.

Table 5-10 Pearson correlation of electricity (exerpt from Table 6-13)

PEARSON	d_dea_far_ndrs	d_dea_far_crs	sfa_translog_far
d_dea_far_ndrs	1.000	0.997	0.733
d_dea_far_crs		1.000	0.736
sfa_translog_far			1.000

5.199 The Pearson correlation is validated using the correlation test;

```
data:  d_dea_far_crs and d_sfa_translog_far
t = 19.6395, df = 326, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.6822705 0.7821125
sample estimates:
      cor
0.7361713
```

Table 5-11 Kendall's rank correlation electricity (exerpt from Table 6-14)

KENDALL	d_dea_far_ndrs	d_dea_far_crs	sfa_translog_far
d_dea_far_ndrs	1.000	0.983	0.534
d_dea_far_crs		1.000	0.541
sfa_translog_far			1.000

5.200 The Kendall test below validates the rank-order correlation between NDRS and SFA, and by extension the CRS-SFA as well.

```
   Kendall's rank correlation tau

data:  d_dea_far_ndrs and d_sfa_translog_far
z = 14.3369, p-value < 2.2e-16
alternative hypothesis: true tau is not equal to 0
sample estimates:
      tau
0.5335241
```

Table 5-12 Spearman's rank correlation electricity (exerpt from Table 6-15)

SPEARMAN	d_dea_far_ndrs	d_dea_far_crs	sfa_translog_far
d_dea_far_ndrs	1.000	0.997	0.717
d_dea_far_crs		1.000	0.724
sfa_translog_far			1.000

5.201 The Spearman rank order correlation test validates the relationship between DEA-NDRS and SFA as below.

```
   Spearman's rank correlation rho

data:  d_dea_far_ndrs and d_sfa_translog_far
S = 1664549, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.7169715
```

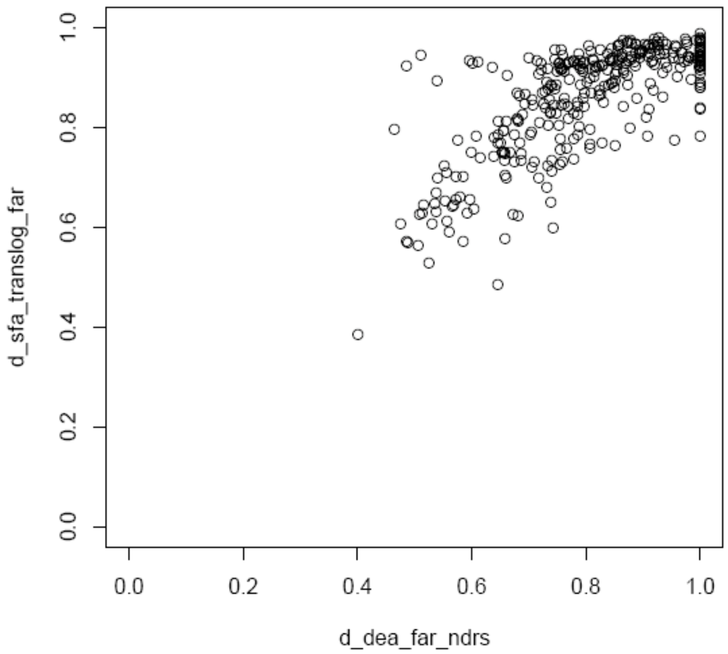

Figure 5-29 Correlation between DEA-NDRS and SFA results.

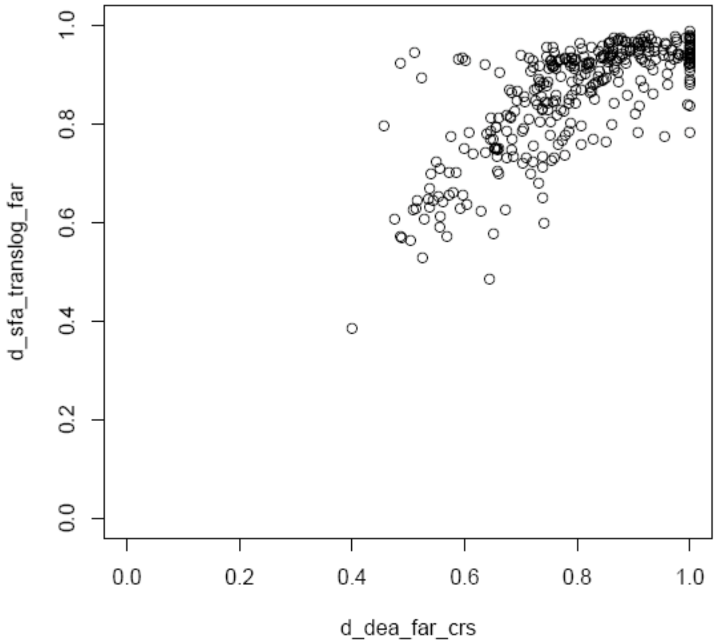

Figure 5-30 Correlation between DEA-CRS and SFA results.

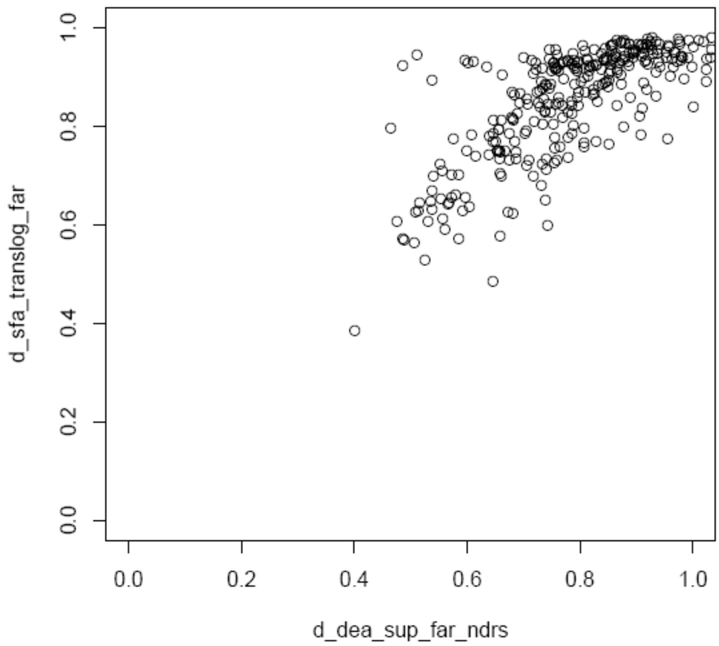

Figure 5-31 Correlation between superefficiency DEA-NDRS and SFA results.

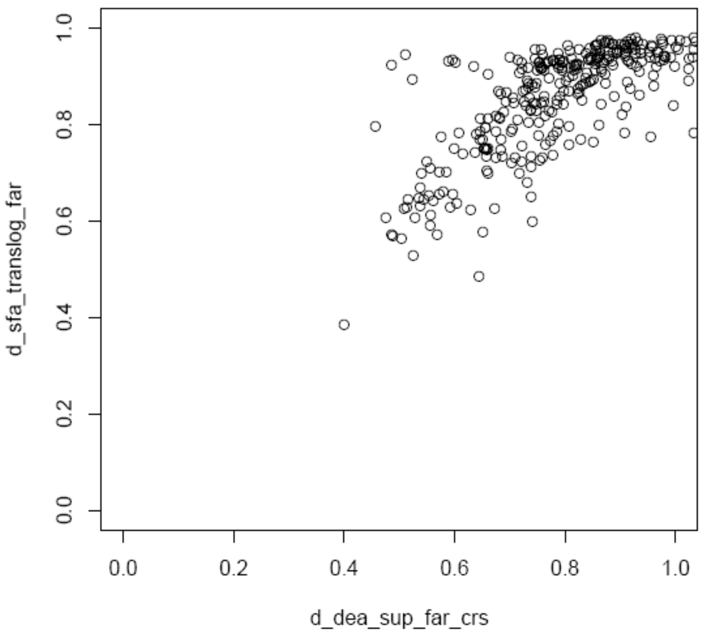

Figure 5-32 Correlation between superefficiency DEA-CRS and SFA results.

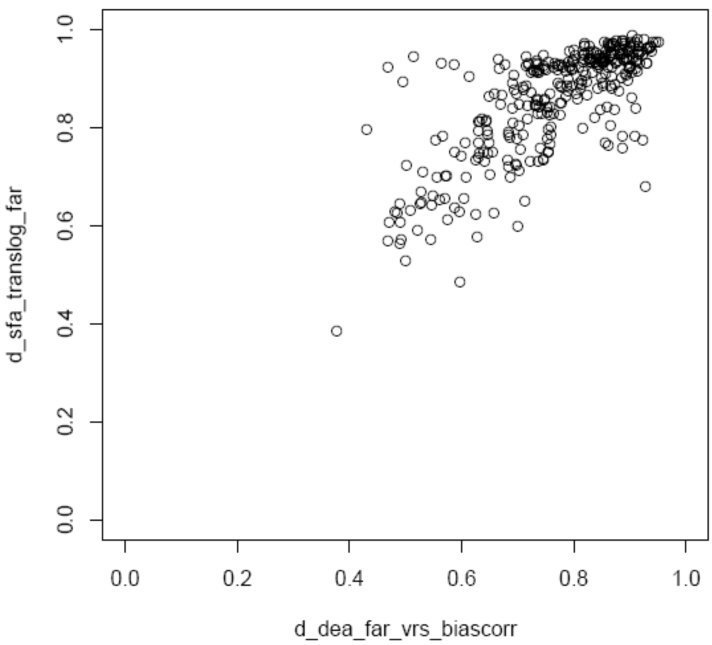

Figure 5-33 Correlation between biascorrected DEA-VRS and SFA results.

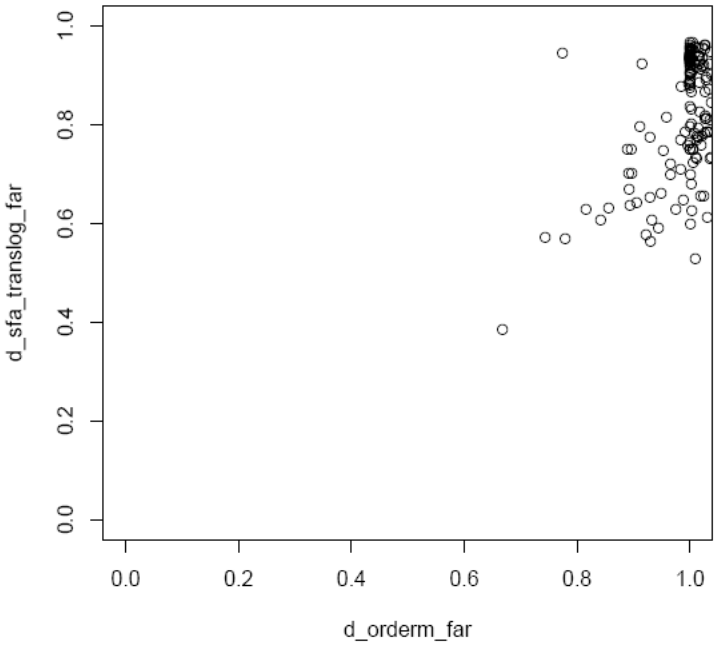

Figure 5-34 Correlation between DEA order-m and SFA results.

5.11 Second-stage analysis electricity model II

5.202 The statistical analysis of the benchmarking model with respect to structural variables is performed using an OLS regression. The only valid OLS model f-test is made through the superefficiency, since the other scores come from truncated distributions. It is possible to make regressions on truncated data using e.g. Tobit, bootstrapping or other adapted techniques. Although these functions are poorly or not at all integrated in the R software used for the calculations, we suggest the use of such techniques selectively after the first analysis for all relevant models.

5.203 The presentation is made for DEA-NDRS, CRS, the superefficiency DEA-CRS and SFA translog models. We test the fit of a model of the type

$$E = c^{T}z + e$$

5.204 The z-variables are tested groupwise in logical groups, soiltype, buildings etc, as to avoid interaction effects. For each test, we report the estimate c, the standard error for the coefficient, the t-value and the probability that the coefficient is identified as noise. The latter p-value is marked in bold face black when lower than 0.01 and bold face red if it is significant at the 0.001 level.

5.205 In addition to the superefficiency results, indicative OLS analyses have also been made on the other models. The DEA-NDRS model is very robust, the results below indicate influence only Connection density and maximum Peakload, but with rather low impact.

5.12 Second-stage results for Superefficiency DEA-CRS

5.206 The second-stage analysis for the superefficiency model under CRS shows significant variables at the 0.001-level in the groups Soiltype (1), DG power (2), calculated network length (1), Meters (3), Population (2), Connection density (1). All coefficients are in the order of 0.001 or lower, except for zSoiltype.1 and zMeters.hs_ms. The Soiltype dummy is based on a subsample of 10 observations with very particular profile, giving them very high superefficiencies. A number of these observations are pinpointed by the Banker-test as potential outliers. The zMeters.hs_ms variable has a coefficient of 0.003, which must be considered as negligible. Hence, we conclude that the relatively simple electricity model II is robust and unbiased to the identified sample of structural variables.

Second stage estimates for model d_dea_sup_far_crs Soiltype

Soiltype

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.762	0.071	10.795	0.000
zSoiltype.1	0.375	0.105	3.584	0.000
zSoiltype.2	0.003	0.130	0.020	0.984
zSoiltype.3	0.024	0.078	0.310	0.757
zSoiltype.4	0.066	0.073	0.899	0.369
zSoiltype.5	0.027	0.076	0.348	0.728
zSoiltype.6	0.095	0.095	0.998	0.319

Second stage estimates for model d_dea_sup_far_crs Dg.power

Dg.power

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.794	0.011	70.190	0.000
zDg.power.hs	0.000	0.000	13.245	0.000
zDg.power.ms	0.000	0.000	-2.317	0.021
zDg.power.ns	0.000	0.000	4.136	0.000

Second stage estimates for model d_dea_sup_far_crs Buildings

Buildings

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.814	0.014	56.961	0.000
zBuildings.ns	0.000	0.000	1.648	0.100

Second stage estimates for model d_dea_sup_far_crs Energy

Energy

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.800	0.011	71.275	0.000
zEnergy.del.lower_grid.hs	0.000	0.000	-0.337	0.736
zEnergy.del.lower_grid.hs_ms	0.000	0.000	0.776	0.438
zEnergy.del.lower_grid.del.ms	0.000	0.000	-0.369	0.712
zEnergy.del.lower_grid.ms_ns	0.000	0.000	0.556	0.579
zEnergy.del	0.000	0.000	-1.126	0.261
zEnergy.del.hs	0.000	0.000	1.117	0.265
zEnergy.del.hs_ms	0.000	0.000	1.373	0.171
zEnergy.del.ms	0.000	0.000	1.451	0.148
zEnergy.del.ms_ns	0.000	0.000	1.099	0.273
zEnergy.del.ns	0.000	0.000	2.806	**0.005**

Second stage estimates for model d_dea_sup_far_crs Length

Length

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.805	0.011	75.225	0.000
zLength.calc.net.a2.ns	0.000	0.000	1.432	0.153
zLength.calc.net.a1.ns	0.000	0.000	-1.951	0.052
zLength.calc.net.ms	0.000	0.000	1.493	0.136
zLength.calc.net.hs	0.000	0.000	16.480	0.000

Second stage estimates for model d_dea_sup_far_crs Meters

Meters

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.802	0.012	68.397	0.000
zMeters.tot	0.000	0.000	-4.622	0.000
zMeters.hs_ms	0.003	0.000	11.954	0.000
zMeters.ms_ns	0.000	0.000	5.416	0.000

Second stage estimates for model d_dea_sup_far_crs Peakload

Peakload

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.786	0.011	70.666	0.000
zPeakload.max	0.000	0.000	14.199	0.000

Second stage estimates for model d_dea_sup_far_crs Height

Height

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.820	0.016	49.989	0.000
zHeight	0.000	0.000	0.587	0.557
zHeightdiff	0.000	0.000	-1.561	0.120

Second stage estimates for model d_dea_sup_far_crs Tilt

Tilt

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.826	0.014	58.801	0.000
zTilt	-0.006	0.004	-1.590	0.113

Second stage estimates for model d_dea_sup_far_crs Loc

Loc

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.928	0.097	9.596	0.000
zLoc.hs	-0.020	0.077	-0.252	0.802

Second stage estimates for model d_dea_sup_far_crs Population

Population

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.794	0.015	53.261	0.000
zPopulation.ms	0.000	0.000	-4.312	0.000
zPopulation.hs_ms	0.000	0.000	1.772	0.077
zPopulation.hs	0.000	0.000	5.558	0.000
zPopulationDensity.hs	0.000	0.000	-0.425	0.671
zPopulationDensity.hs_ms	0.000	0.000	2.528	0.012
zPopulationDensity.ms	0.000	0.001	-0.468	0.640
zPopulationDensity.ms_ns	0.000	0.001	0.174	0.862
zPopulationDensity.ns	0.000	0.000	1.352	0.177

Second stage estimates for model d_dea_sup_far_crs Conn

Conn

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.998	0.053	18.829	0.000
zConnpoidens.a2..ns	-0.001	0.000	-3.995	0.000
zConnpoidens.a1..ns	0.001	0.000	3.062	**0.002**
zConnpoidens.ms	-0.002	0.005	-0.351	0.726
zConnpoidens.hs	0.203	0.266	0.762	0.446

Second stage estimates for model d_dea_sup_far_crs Area

Area

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.770	0.017	45.287	0.000
zAreainnercity	0.018	0.023	0.757	0.450
zAreacity	0.000	0.001	0.162	0.871
zAreaIndustry	0.007	0.004	1.726	0.085
zAreaTraffic	0.013	0.006	2.136	0.034
zAreaWhatev	-0.001	0.001	-0.549	0.584
zAreaVeget	0.000	0.000	-1.357	0.176
zAreaRock	0.006	0.018	0.316	0.753
zAreaWater	0.000	0.002	-0.187	0.852

Second stage estimates for model d_dea_sup_far_crs Age

Age

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.751	0.062	12.071	0.000
zAge	0.004	0.003	1.401	0.164

Second stage estimates for model d_dea_sup_far_crs Model

Model

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.816	0.057	14.199	0.000
zModel.LengthNS.detailed	0.000	0.001	-0.246	0.808
zModel.LengthNS.aggr	0.000	0.001	0.537	0.598
zModel.LengthMS.detailed	0.000	0.002	0.015	0.988

5.13 Second-stage results for SFA translog

5.207 The translog SFA model of electricity model II shows a remarkable stability and robustness to the structural variables below, partially due to its highly flexible functional form. The one and single variable that is significant for the untruncated regression is the slope, shown to have a slight efficiency decreasing effect. For recall, the second-stage analysis tries to find variables that may have a systematic influence on the scores, which would then mean that the scores can be explained by some omitted factor.

Second stage estimates for model d_sfa_translog_far Soiltype

Soiltype

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.823	0.032	25.940	0.000
zSoiltype.1	0.068	0.047	1.446	0.149
zSoiltype.2	0.012	0.058	0.213	0.832
zSoiltype.3	0.043	0.035	1.234	0.218
zSoiltype.4	0.044	0.033	1.353	0.177
zSoiltype.5	0.009	0.034	0.266	0.791
zSoiltype.6	0.076	0.043	1.791	0.074

Second stage estimates for model d_sfa_translog_far Decentralized generation (DG)

DG Power

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.855	0.006	136.314	0.000
zDg.power.hs	0.000	0.000	1.492	0.137
zDg.power.ms	0.000	0.000	1.333	0.184
zDg.power.ns	0.000	0.000	0.705	0.481

Second stage estimates for model d_sfa_translog_far Buildings

Buildings

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.856	0.006	135.971	0.000
zBuildings.ns	0.000	0.000	2.350	0.019

Second stage estimates for model d_sfa_translog_far Energy delivered

Energy delivered

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.855	0.007	119.168	0.000
zEnergy.del.lower_grid.hs	0.000	0.000	-1.114	0.266
zEnergy.del.lower_grid.hs_ms	0.000	0.000	1.271	0.205
zEnergy.del.lower_grid.del.ms	0.000	0.000	-0.713	0.476
zEnergy.del.lower_grid.ms_ns	0.000	0.000	0.713	0.477
zEnergy.del	0.000	0.000	-0.437	0.663
zEnergy.del.hs	0.000	0.000	0.406	0.685
zEnergy.del.hs_ms	0.000	0.000	0.526	0.599
zEnergy.del.ms	0.000	0.000	0.684	0.494
zEnergy.del.ms_ns	0.000	0.000	0.748	0.455
zEnergy.del.ns	0.000	0.000	0.310	0.757

Second stage estimates for model d_sfa_translog_far Length

Length

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.854	0.006	133.625	0.000
zLength.calc.net.a2.ns	0.000	0.000	1.447	0.149
zLength.calc.net.a1.ns	0.000	0.000	-0.902	0.368
zLength.calc.net.ms	0.000	0.000	0.466	0.642
zLength.calc.net.hs	0.000	0.000	0.652	0.515

Second stage estimates for model d_sfa_translog_far Meters

Meters

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.854	0.006	134.077	0.000
zMeters.tot	0.000	0.000	1.774	0.077
zMeters.hs_ms	0.000	0.000	0.769	0.442
zMeters.ms_ns	0.000	0.000	0.607	0.544

Second stage estimates for model d_sfa_translog_far Peakload

Peakload

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.857	0.006	137.869	0.000
zPeakload.max	0.000	0.000	2.188	0.029

Second stage estimates for model d_sfa_translog_far Height

Height

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.878	0.010	86.884	0.000
zHeight	0.000	0.000	-0.725	0.469
zHeightdiff	0.000	0.000	-1.622	0.106

Second stage estimates for model d_sfa_translog_far Tilt

Tilt

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.879	0.009	102.175	0.000
zTilt	-0.008	0.002	-3.389	0.001

Second stage estimates for model d_sfa_translog_far Location East-West

Location East-West

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.873	0.033	26.106	0.000
zLoc.hs	0.030	0.027	1.139	0.261

Second stage estimates for model d_sfa_translog_far Population

Population

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.847	0.009	93.644	0.000
zPopulation.ms	0.000	0.000	0.255	0.799
zPopulation.hs_ms	0.000	0.000	0.846	0.398
zPopulation.hs	0.000	0.000	-0.484	0.629
zPopulationDensity.hs	0.000	0.000	0.301	0.764
zPopulationDensity.hs_ms	0.000	0.000	0.581	0.562
zPopulationDensity.ms	0.000	0.000	-0.719	0.473
zPopulationDensity.ms_ns	0.000	0.000	0.673	0.501
zPopulationDensity.ns	0.000	0.000	0.240	0.810

Second stage estimates for model d_sfa_translog_far Connection points

Connection points

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.890	0.024	37.370	0.000
zConnpoidens.a2..ns	0.000	0.000	-1.824	0.069
zConnpoidens.a1..ns	0.000	0.000	2.048	0.041
zConnpoidens.ms	-0.001	0.002	-0.606	0.545
zConnpoidens.hs	0.150	0.120	1.256	0.210

Second stage estimates for model d_sfa_translog_far Area

Area

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.837	0.011	77.142	0.000
zAreainnercity	0.015	0.015	0.988	0.324
zAreacity	0.000	0.001	0.167	0.868
zAreaIndustry	0.004	0.003	1.490	0.137
zAreaTraffic	0.000	0.004	-0.036	0.972
zAreaWhatev	-0.001	0.001	-0.778	0.437
zAreaVeget	0.000	0.000	0.234	0.815
zAreaRock	0.000	0.012	0.031	0.976
zAreaWater	0.001	0.002	0.454	0.650

Second stage estimates for model d_sfa_translog_far Age

Age

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.847	0.039	21.528	0.000
zAge	0.001	0.002	0.540	0.590

Second stage estimates for model d_sfa_translog_far Model network length

Model network length

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.860	0.031	27.356	0.000
zModel.LengthNS.detailed	0.000	0.001	-0.613	0.547
zModel.LengthNS.aggr	0.000	0.000	0.665	0.514
zModel.LengthMS.detailed	0.001	0.001	0.588	0.563

6. The gas model

6.1 Gas model specifications

6.208 The results for the gas model are more preliminary than those of the electricity model for two reasons. First, the data collection has been terminated later than for electricity with comparatively higher input variability and data error. Second, the state of the art in conceptual gas regulation models is less advanced than for electricity with more uncertainty regarding techno-economic relationships. Below we present the result of gas model 8, which is the result of the model development process described above.

6.209 For all variable and parameter definitions, please refer to BNetzA (2006) to obtain the official translation and conversion to the data specification call.

Parameters model 8

Table 6-1 Parameters in gas model 8.

Parameter	Category	Definition
yAreaServed	Service provision	Total service area
yConnections.total	Service provision	Total number of connections
yEnergy.out.2004	Transportation work	Total distributed energy
yPeakload.sync.out.2004	Capacity provision	Synchronized load output

Data

6.210 Model 8 has been estimated on a sample consisting of 488 firms. The original dataset contained 500 firms and the reduced sample used covers more than 99.6% of the direct cost in the dataset. The eliminated firms are identified as outliers using the Banker test for outlier identification based on superefficiency, cf. Banker and Chang (2005).

6.2 Descriptive OLS results

Standard model

6.211 The OLS results for gas model 8 are given below, indicating an adjusted R2 of 0.9875.

```
Residuals:
      Min        1Q    Median        3Q       Max
-47615551   -600287   -274845    360206  39008328

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)              3.366e+05  2.346e+05   1.435   0.1520
yEnergy.out.2004        -1.295e-02  5.022e-03  -2.578   0.0102 *
yAreaServed              1.791e+03  4.109e+02   4.358 1.60e-05 ***
yPeakload.sync.out.2004  2.240e+02  1.762e+01  12.714  < 2e-16 ***
yConnections.total       2.021e+01  9.619e+00   2.101   0.0361 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4699000 on 483 degrees of freedom
Multiple R-Squared: 0.9436,     Adjusted R-squared: 0.9431
F-statistic:  2018 on 4 and 483 DF,  p-value: < 2.2e-16
```

6.212 Below, we test the sensitivity of the base model 8 above by adding one-by-one candidate structural variables from the groups Soil, Soiltype, Buildings, Operations, Energy Input, Quality, Peakload, Height and tilt, Location East-West, Degree of sealed ground, Meters and Connection points.

Soil

6.213 Below are the results for OLS integration of soiltype variables, here transformed into ordinal variables for the dominant soiltype for the operator multiplied with the area. As seen, zSoilAreaAsset is significant at the 0.001-level.

Add-one term for group Soil

Df	Sum of Sq	RSS	AIC	F Value	Pr(F)
<none>	NA	9.75384E+15	14955.55	NA	NA
zSoilAreaAsset	2.75E+14	9.47869E+15	14943.58	13.99	0.000
zSoilAreaServed	4.09E+11	9.75343E+15	14957.53	0.02	0.887

Soiltype

6.214 Below are the results for OLS integration of soiltype variables, here transformed into dummy variables for the dominant soiltype for the operator. As seen, none is significant.

Add-one term for group Soil

Df	Sum of Sq	RSS	AIC	F Value	Pr(F)
<none>	NA	1.06638E+16	14999.08	NA	NA
zSoiltype.1	1.31744E+13	1.06506E+16	15000.47	0.60	0.440
zSoiltype.2	1.01028E+12	1.06628E+16	15001.03	0.05	0.831
zSoiltype.3	9.85397E+11	1.06628E+16	15001.03	0.04	0.833
zSoiltype.4	8.92279E+11	1.06629E+16	15001.03	0.04	0.841
zSoiltype.5	2.01925E+13	1.06436E+16	15000.15	0.91	0.339
zSoiltype.6	1.90938E+13	1.06447E+16	15000.20	0.86	0.353
zSoiltype.7	463702840	1.06638E+16	15001.08	0.00	0.996

Buildings

6.215 Below are the results for OLS integration of variables related to the number of buildings in the area served and in the area for which the operator has assets, respectively. Both variables zBuildingAreaServed and zBuildingsAreaAsset are significant at the 0.001 level. However, they are essentially complements for the variable yAreaServed (population in the area served), which was judged to be more reliable.

Add-one term for group Building

Df	Sum of Sq	RSS	AIC	F Value	Pr(F)
<none>	NA	1.02447E+16	14979.51	NA	NA
zBuildingsAreaServed	2.77E+15	7.47948E+15	14827.98	178.20	0.000
zBuildingsAreaAsset	1.49E+15	8.75284E+15	14904.71	82.15	0.000

Operations

6.216 Below are the results for OLS integration of dummy variables for multi-utility operation, as operation of district heating, gas storage, sewerage or water distribution. As seen, none is significant.

Add-one term for group Operator

Df	Sum of Sq	RSS	AIC	F Value	Pr(F)
<none>	NA	1.06638E+16	14999.08	NA	NA
zOperator.dis.heat	2.76E+13	1.06363E+16	14999.81	1.25	0.264
zOperator.gas.hol	5.02E+12	1.06588E+16	15000.85	0.23	0.634
zOperator.sew.net	5.96E+11	1.06632E+16	15001.05	0.03	0.870
zOperator.water	3.45E+13	1.06293E+16	14999.49	1.57	0.211

Energy Input

6.217 Below are the results for OLS integration of historic energy variables related to feed-in of gas. As seen, all are significant.

Add-one term for group Energy

Df	Sum of Sq	RSS	AIC	F Value	Pr(F)
<none>	NA	8.93584E+15	14912.80	NA	NA
zEnergy.in.2000	1.53E+15	7.40267E+15	14822.95	99.83	**0.000**
zEnergy.in.2001	1.22E+15	7.71347E+15	14843.02	76.38	**0.000**
zEnergy.in.2002	2.17E+14	8.71932E+15	14902.83	11.97	**0.001**
zEnergy.in.2003	5.73E+14	8.36245E+15	14882.44	33.05	**0.000**

Quality

6.218 Below are the results for OLS integration of quality indicators to the number and severity of leakages at pressure levels ND, MD and HD. As seen, only the zQ.Leakage.nd (volume of leakage at low pressure network level) is significant at the 0.05 level, none at 0.01 level. The sign of zQ.Leakage.nd is positive, indicating that quality problems are signs of some related problem and not outcomes of a trade-off cost-quality.

Add-one term for group Q

Df	Sum of Sq	RSS	AIC	F Value	Pr(F)
<none>	NA	1.0663E+16	14999.04	NA	NA
zQLeakage.md	2.29E+14	1.04338E+16	14990.43	10.58	**0.001**
zQ.Leakage.nd	6.83E+13	1.05946E+16	14997.90	3.11	0.079
zQ.Disturbance.hd	8.65E+12	1.06543E+16	15000.64	0.39	0.532
zQ.Leakage.hd	2.21E+12	1.06607E+16	15000.93	0.10	0.752

Peakload (other)

6.219 Below are the results for OLS integration of the peakload variables that are not included in the base model. None is significant.

Add-one term for group Peakload

Df	Sum of Sq	RSS	AIC	F Value	Pr(F)
<none>	NA	5.95296E+15	14714.59	NA	NA
zPeakload.async.out.2000	7.85E+12	5.9451E+15	14715.94	0.64	0.425
zPeakload.async.out.2004	8.86E+12	5.94409E+15	14715.86	0.72	0.397
zPeakload.out.async.2001	8.23E+12	5.94472E+15	14715.91	0.67	0.414
zPeakload.out.async.2002	8.37E+12	5.94458E+15	14715.90	0.68	0.410
zPeakload.out.async.2003	8.55E+12	5.9444E+15	14715.88	0.69	0.405

Height and tilt

6.220 Below are the results for OLS integration of the variables related to absolute height, height difference and slope. None is significant.

Add-one term for group Height

Df	Sum of Sq	RSS	AIC	F Value	Pr(F)
<none>	NA	6.61108E+15	14765.76	NA	NA
zHeight	5.6E+11	6.61052E+15	14767.72	0.04	0.840
zHeightdiff	3.41E+12	6.60766E+15	14767.51	0.25	0.618
zTilt	1.07E+12	6.61001E+15	14767.68	0.08	0.781

Location East-West

6.221 Below are the results for OLS integration of a dummy variable related to the location in the new or old Bundesländer. The variable is not significant.

Add-one term for group Loc

Df	Sum of Sq	RSS	AIC	F Value	Pr(F)
<none>	NA	1.06638E+16	14999.08	NA	NA
zLoc	6.58E+13	1.05981E+16	14998.06	2.99	0.084

Degree of sealed ground

6.222 The impact of sealing is tested with two variables expressing the area sealed for with the DSO has assets and is serving, respectively. The variables are significant.

Add-one term for group Seal

Df	Sum of Sq	RSS	AIC	F Value	Pr(F)
<none>	NA	1.02447E+16	14979.51	NA	NA
zSealedAreaAsset	9.02E+14	9.3431E+15	14936.55	46.51	0.000
zSealedAreaServed	1.81E+15	8.43175E+15	14886.47	103.64	0.000

Meters

6.223 The numbers of meters at HD and MD levels are not significant additions to the model.

Add-one term for group Meter

Df	Sum of Sq	RSS	AIC	F Value	Pr(F)
<none>	NA	1.06638E+16	14999.08	NA	NA
zMeters.hd	5.88E+11	1.06632E+16	15001.05	0.03	0.871
zMeters.md	7.46E+13	1.05892E+16	14997.65	3.40	0.066

Connection points

6.224 The number of connection points at MD level is significant at the 0.001-level. However, the number is only a subset of the variable included (yConnections.total).

Add-one term for group Connect

Df	Sum of Sq	RSS	AIC	F Value	Pr(F)
<none>	NA	1.06638E+16	14999.08	NA	NA
zConnectpoint.out.dso.hd	2.75E+13	1.06363E+16	14999.82	1.24	0.265
zConnectpoint.out.dso.md	4.78E+15	5.88362E+15	14710.87	391.61	0.000

6.3 T1 Results

6.225 The available data on gas does not allow us to make a reasonable subadditivity test; it is only possible to split up connections according to pressure level which does not suffice to make a reliable model of the costs of the individual pressure levels. As indicated by the gas model analyzed below, the area conditions, the peak capacity and the energy delivered are important cost drivers that cannot be detailed according to pressure level.

6.4 T2 Results

6.226 The age test is performed both against the dependent variable (xCosts.dir) here and against the obtained score below. As seen, inclusion of age-proxies in the OLS model cannot be defended at the 0.01-level.

Add-one term for group Age

Df	Sum of Sq	RSS	AIC	F Value	Pr(F)
<none>	NA	1.06593E+16	14998.87	NA	NA
zAge	9.25E+13	1.05668E+16	14996.61	4.22	0.040
zAge.hd	6.59E+13	1.05934E+16	14997.84	3.00	0.084
zAge.md	7.82E+12	1.06515E+16	15000.51	0.35	0.552
zAge.nd	3.88E+13	1.06205E+16	14999.09	1.76	0.185

6.5 T3 Results

6.227 The results for gas model 8 resemble those of the electricity model in some significant instances. To illustrate the marginal sensitivity with respect to capital, the coefficients are scaled with mean CAPEX and standardized CAPEX, respectively. Note that the relevant results, those for the superefficiency model, come out insignificant for the used NDRS and CRS specifications both for CAPEX and standardized CAPEX. We also note that the coefficients of the SFA model occasionally are negative, albeit insignificant. Indeed, the results are actually somewhat counterintuitive, as the costefficiency score is non-increasing in cost, thus CAPEX.

Table 6-2 T3 CAPEX test gas model 8.

Model	Mean xCAPEX* Estimate	Std. Error	t value	Pr(>|t|)
d_dea_sup_far_vrs	0.03	0.000	2.705	**0.008**
d_dea_sup_far_drs	0.04	0.000	3.716	**0.000**
d_dea_sup_far_ndrs	0.00	0.000	0.233	0.816
d_dea_sup_far_crs	0.00	0.000	0.801	0.424
d_sfa_linear_far	-0.01	0.000	-1.938	0.054
d_sfa_loglinear_far	0.00	0.000	-0.631	0.529
d_sfa_translog_far	0.00	0.000	0.313	0.754
d_sfa_normedlinear_far	-0.02	0.000	-2.966	**0.003**
d_sfa_normedloglinear_far	0.01	0.000	1.836	0.068
d_sfa_effectslinear_far	-0.01	0.000	-1.943	0.054

Table 6-3 T3 CAPEX.standard test gas model 8.

Model	Estimate	Std. Error	t value	Pr(>|t|)
d_dea_sup_far_vrs	0.03	2.81E-09	2.944	**0.004**
d_dea_sup_far_drs	0.04	2.34E-09	4.036	**0.000**
d_dea_sup_far_ndrs	0.00	1.76E-09	-0.014	0.989
d_dea_sup_far_crs	0.00	1.38E-09	0.497	0.620
d_sfa_linear_far	-0.02	1.67E-09	-2.206	0.029
d_sfa_loglinear_far	0.00	1.02E-09	-1.093	0.276
d_sfa_translog_far	0.00	8.24E-10	-0.228	0.820
d_sfa_normedlinear_far	-0.02	1.46E-09	-2.917	**0.004**
d_sfa_normedloglinear_far	0.01	8.61E-10	1.830	0.069
d_sfa_effectslinear_far	-0.02	1.67E-09	-2.209	0.028

Ratio OPEX/CAPEX: Gas model

6.228 To test the potential bias introduced by abnormal proportions of capital costs, the ratio of capital costs to total costs was used as a regressor to SFA scores, DEA superefficiency scores for all units and for frontier units in DEA-CRS. Results are present in Table 6-4. The strongest results (at the 0.002-level) are obtained for the Superefficiency CRS model, followed by the SFA loglinear model (0.01-level) and lastly the superefficiency model under NDRS (0.02-level). As suggested in Figure 6-1, all coefficients are positive, meaning that a higher proportion would predict a higher score. A 10% deviation from the average proportion would imply about 0.01 higher score (in units) for DEA, whereas the SFA impact is about half as big. The full coefficients are provided in Table 6-5 below.

Table 6-4 ANOVA test ratio CAPEX/TOTEX to scores gas model 8.

Supereff DEA-NDRS	Df	Sum Sq	Mean Sq	F value	Pr(>F)
I(xCAPEX/(xCAPEX + xOPEX))	1	0.3380	0.3380	6.2042	0.0137
Residuals	170	9.2626	0.0545		

Supereff DEA-CRS	Df	Sum Sq	Mean Sq	F value	Pr(>F)
I(xCAPEX/(xCAPEX + xOPEX))	1	0.3367	0.3367	10.241	0.001639
Residuals	170	5.5887	0.0329		

SFA_loglinear_fa	Df	Sum Sq	Mean Sq	F value	Pr(>F)
I(xCAPEX/(xCAPEX + xOPEX))	1	0.1194	0.1194	7.1426	0.00826
Residuals	170	2.8406	0.0167		

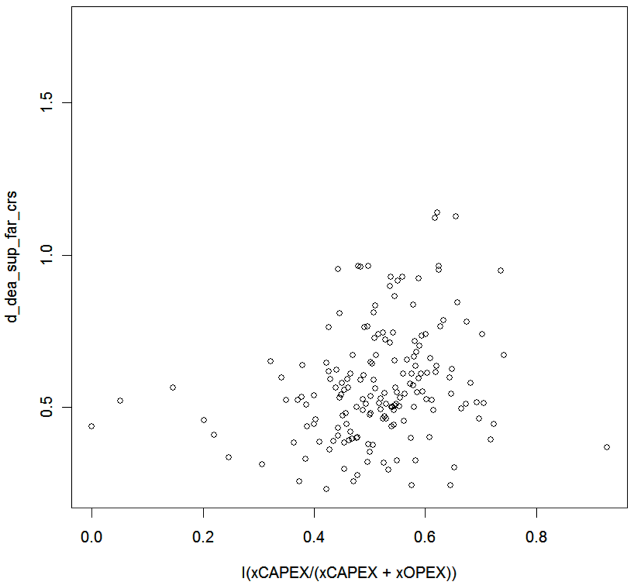

Figure 6-1 T3 CAPEX-ratio for DEA-Superefficency CRS result gas model 8.

Table 6-5 T3 CAPEX/TOTEX ratio gas model 8.

Model	Estimate	Std. Error	t value	Pr(>|t|)
d_dea_sup_far_vrs	0.421	0.174	2.422	0.016
d_dea_sup_far_drs	0.421	0.147	2.875	**0.005**
d_dea_sup_far_ndrs	0.378	0.152	2.491	0.014
d_dea_sup_far_crs	0.378	0.118	3.200	**0.002**
d_sfa_linear_far	0.241	0.143	1.679	0.095
d_sfa_loglinear_far	0.225	0.084	2.673	**0.008**
d_sfa_translog_far	0.149	0.068	2.198	0.029
d_sfa_normedlinear_far	0.012	0.140	0.085	0.933
d_sfa_normedloglinear_far	0.292	0.076	3.867	**0.000**
d_sfa_effectslinear_far	0.241	0.144	1.677	0.095

Asset intensity test (Normalized CAPEX): gas

6.229 We define an asset intensity proxy for gas using CAPEX over the weighted pipeline length. Using the OLS estimates below, we use the weights 1, 2, 4 per km for MD, ND and HD pressure levels and also an unweighed sum. The distribution of the regressor is illustrated in Figure 6-2 below and the full results for the weighed model are presented in Table 6-6 and for the unweighed model in Table 6-7. The weighed model confirms the results from the electricity model, meaning that asset intensity is not a predictor of efficiency score. The asset intensity with the weights above is also shown to be independent of average asset age and location, cf regression results. However, the results are somewhat sensitive to the weights as shown by the unweighed model, where the factor is significant at the 0.02-level for e.g. supereffiency NDRS. The unweighed proxy is actually directly proportional to age (cf Figure 6-3 and the regression results below it), meaning that there is, if anything, higher capital intensity for the older networks and networks in the new Bundesländer (zLoc). The results can be interpreted in two ways: (i) there is no support in the data for simple investment cycle behavior ("rollercoaster") in the benchmarking, since older networks have, if anything, higher asset intensities, (ii) given the very simple proxy construction above, there might be interesting to further investigate if there is any difference in capital evaluation between age and location.

```
Call:
lm(formula = modl)

Residuals:
      Min        1Q    Median        3Q       Max
-24656426  -1760049  -1436458   -372450  29386491

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 1871954.5   517461.9   3.618 0.000391 ***
xLength.hd    12562.1     2952.0   4.255 3.43e-05 ***
xLength.nd     6094.7     1592.0   3.828 0.000181 ***
xLength.md     2558.4      486.3   5.261 4.24e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6233000 on 171 degrees of freedom
Multiple R-Squared: 0.8348,     Adjusted R-squared: 0.8319
F-statistic:   288 on 3 and 171 DF,  p-value: < 2.2e-16
```

Table 6-6 T3 Asset intensity test gas model 8, weighed pipeline length.

Model	Estimate	Std. Error	t value	Pr(>|t|)
d_dea_sup_far_vrs	-5.445E-06	8.762E-06	-0.621	0.535
d_dea_sup_far_drs	-7.041E-06	7.447E-06	-0.946	0.346
d_dea_sup_far_ndrs	-5.070E-06	7.648E-06	-0.663	0.508
d_dea_sup_far_crs	-6.730E-06	5.999E-06	-1.122	0.264
d_sfa_linear_far	-3.672E-06	7.154E-06	-0.513	0.608
d_sfa_loglinear_far	-5.722E-06	4.236E-06	-1.351	0.179
d_sfa_translog_far	-4.221E-06	3.396E-06	-1.243	0.216
d_sfa_normedlinear_far	-5.080E-06	6.903E-06	-0.736	0.463
d_sfa_normedloglinear_far	-8.758E-06	3.862E-06	-2.267	0.025
d_sfa_effectslinear_far	-3.680E-06	7.164E-06	-0.514	0.608

Table 6-7 T3 Asset intensity test gas model 8, unweighed pipeline length.

Model	Estimate	Std. Error	t value	Pr(>|t|)
d_dea_sup_far_vrs	-1.084E-05	4.680E-06	-2.316	0.022
d_dea_sup_far_drs	-8.535E-06	3.992E-06	-2.138	0.034
d_dea_sup_far_ndrs	-9.695E-06	4.079E-06	-2.377	0.019
d_dea_sup_far_crs	-7.471E-06	3.210E-06	-2.328	0.021
d_sfa_linear_far	-4.373E-06	3.861E-06	-1.133	0.259
d_sfa_loglinear_far	-5.870E-06	2.262E-06	-2.595	0.010
d_sfa_translog_far	-4.497E-06	1.815E-06	-2.478	0.014
d_sfa_normedlinear_far	-2.959E-06	3.736E-06	-0.792	0.429
d_sfa_normedloglinear_far	-7.988E-06	2.033E-06	-3.929	**0.000**
d_sfa_effectslinear_far	-4.403E-06	3.866E-06	-1.139	0.256

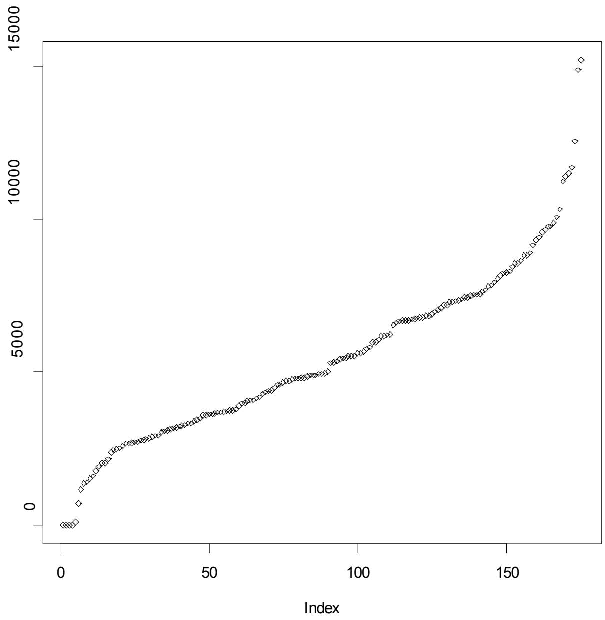

Figure 6-2 CAPEX per weighted pipeline length, gas model 8.

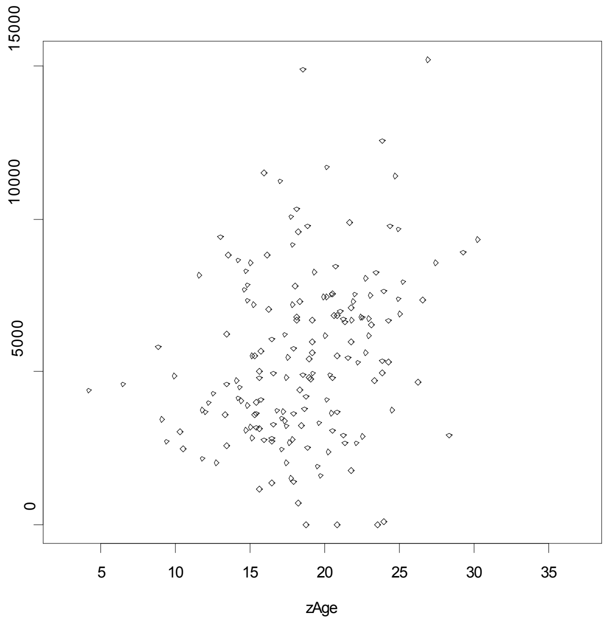

Figure 6-3 Relationship between zCAPEX.norm and zAge, gas model 8.

```
Call: weighed zCAPEX.norm
lm(formula = testm.norm)

Residuals:
    Min      1Q  Median      3Q     Max
-4973.8 -1392.4  -204.5  1039.0  9217.7

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5512.12     861.91   6.395 1.53e-09 ***
zAge          -28.79      42.86  -0.672    0.503
zLoc          210.54     392.22   0.537    0.592
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2319 on 168 degrees of freedom
Multiple R-Squared: 0.006291,   Adjusted R-squared: -0.005539
F-statistic: 0.5318 on 2 and 168 DF,  p-value: 0.5885
```

```
Call: unweighed zCAPEX.norm
lm(formula = zCAPEX.norm ~ zAge + zLoc)

Residuals:
    Min      1Q  Median      3Q     Max
-9686.5 -2603.4  -143.5  2332.9 20905.3

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5474.07    1567.91   3.491 0.000614 ***
zAge          179.25      77.96   2.299 0.022719 *
zLoc         1392.53     713.49   1.952 0.052637 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4218 on 168 degrees of freedom
Multiple R-Squared: 0.03995,    Adjusted R-squared: 0.02852
F-statistic: 3.495 on 2 and 168 DF,  p-value: 0.03257
```

Frontier units gas

6.230 There are only 7 frontier units in DEA-CRS and the sample of CAPEX standardized data only comprises 3 of those. Thus, the t-test for means differences under unequal variances in Table 6-8 below rejects the hypothesis of equal means, but based on a very narrow sample. We consider the result to be without value.

Table 6-8 t-test for means under unequal variances of CAPEX proportion for frontier vs non-frontier units, gas model 8

	CAPEX/TOTEX DEA-CRS=1	*CAPEX/TOTEX DEA-CRS<1*
Mean	0.63126	0.517346
Variance	0.000413	0.012298
Observations	3	168
Hypothesized Mean Difference	0	
df	5	
t Stat	7.84227	
P(T<=t) one-tail	0.000271	
t Critical one-tail	2.015049	
P(T<=t) two-tail	0.000541	
t Critical two-tail	2.570578	

6.231 Again, we have examined if the outliers detected in the gas model are related to the capital intensity or the possible capital reporting problem. Data allows us to calculate CAPEX proportions for only 6 out of the 12 units eliminated. The average CAPREX ratio, i.e.. CAPEX / (CAPEX+OPEX) among these is 0.516 while it is 0.543 in the full sample used. The corresponding t statistic is -0.9021 with a two-sided p value of 0.401, i.e. there is no significant difference between the eliminated and included units. A non-parametric Kruskal-Wallis gives a chi-squared test statistic for the null hypothesis of similar means of 0.242 with df = 1, such that the p-value i.e. 0.623, i.e. again no indication of a biased outlier sample.

Preliminary conclusions from T3

6.232 The indication from our tests above is that there may be some relationship between efficiency and the amount of CAPEX reported - with more CAPEX reported in some cases being associated with lower efficiencies and vise versa. This suggest that the capital evaluation issue should remain a focus area in the development of the regulation and that the introduction of standardized asset registers, standardized annuity based depreciations and potentially the use of standard costs to cope with historical differences may be useful.

6.233 The purely statistical conclusions we can draw on the available data are however not fully certain. Although higher CAPEX (or CAPEX/TOTEX ratios) seem to give lower efficiencies, it does not seem to be striking at the frontier, i.e. the frontier units does not seem to get better scores simply because of underreported capital. The lower performing units may therefore also be inefficient due to an allocative efficiency problem - relying perhaps too much on CAPEX and too little on OPEX or because of a purchasing inefficiency – perhaps buying its inputs at to high prices. As always, the challenge is to distinguish between (different types of inefficiencies) and misreported / biased cost data. A use of standardized registers, depreciation and perhaps prices should improve this inference).

6.234 In particular for gas, the findings using normalized proxies and their age- and location dependencies suggest that there may be some differences in the (re)valuation of gas networks that may explain part of the range and/or the level of the efficiency results. We suggest that further work should be done to remedy any potential problems in the capital measure for gas.

6.6 Results

6.235 The model specification has been tested under the three model categories COLS, DEA and a linear/ loglinear/ translog/ normedlinear/ normedloglinear/ effectslinear SFA, and in particular for DEA using the scale assumptions CRS (constant), DRS (decreasing), NDRS (non-decreasing) and VRS (variable) returns to scale. Only Farrell (multiplicative scores) results are presented here. A summary of the preliminary results for the sample data is provided in Table 5-2

Table 6-9 Sample results for gas model 8 using various assumptions (294 firms).

Model	Mean	Std.dev	#(≥ 1)	Min
d_cols_far	-37.552	64.40	N/A	-1005.28
d_fdh_far	0.820	0.19	190	0.32
d_dea_far_vrs	0.592	0.19	20	0.23
d_dea_far_drs	0.585	0.19	19	0.23
d_dea_far_ndrs	0.557	0.17	8	0.23
d_dea_far_crs	0.550	0.17	7	0.23
d_dea_sup_far_vrs	0.601	0.23	18	0.23
d_dea_sup_far_drs	0.591	0.22	17	0.23
d_dea_sup_far_ndrs	0.563	0.20	8	0.23
d_dea_sup_far_crs	0.553	0.18	7	0.23
d_dea_far_biascorr	0.539	0.16	0	0.22
d_dea_far_biascorr_c1	0.511	0.15	0	0.21
d_dea_far_biascorr_c2	0.572	0.18	0	0.23
d_orderm_far	0.909	0.24	20	0.35
d_dea_shell_far[, s]	0.592	0.19	20	0.23
d_dea_shell_far[, s]	0.687	0.30	41	0.25
d_dea_shell_far[, s]	0.738	0.31	30	0.26
d_sfa_linear_far	0.875	0.26	369	0.00
d_sfa_loglinear_far	0.623	0.13	0	0.31
d_sfa_translog_far	0.661	0.11	0	0.42
d_sfa_normedlinear_far	0.952	0.18	441	0.02
d_sfa_normedloglinear_far	0.283	0.12	0	0.05
d_sfa_effectslinear_far	0.875	0.26	369	0.00
d_dea_far_se	0.936	0.08	7	0.52

6.236 The level of efficiency is different in the non-parametric DEA models and in the parametric SFA model. One reason may be that the data is not sufficiently cleaned and that the DEA model is particularly sensitive to outliers. Both will tend to make the DEA scores low and – if it has any effect on the SFA scores, it is likely to make them higher since the SFA model will attribute more to the noise term.

DEA-CRS results model 8

6.237 The results for the CRS gas model 8 under DEA are illustrated in Figure 6-9, Figure 6-10 and Figure 6-11 below. The distribution shows a trimodal shape with mean scores around 0.4, 0.6 for a frontier shaped by some 15 units. Although no conclusions are ready concerning the specific validation of these few units, a peeling analysis where layers of efficient units are removed shows increases in average efficiency estimates. The shape of Figure 6-10 also supports this finding, where the frontier visibly is determined by a low number of fairly small units. Otherwise, the Salter diagram shows no apparent bias in operating scale across the efficiency spectrum.

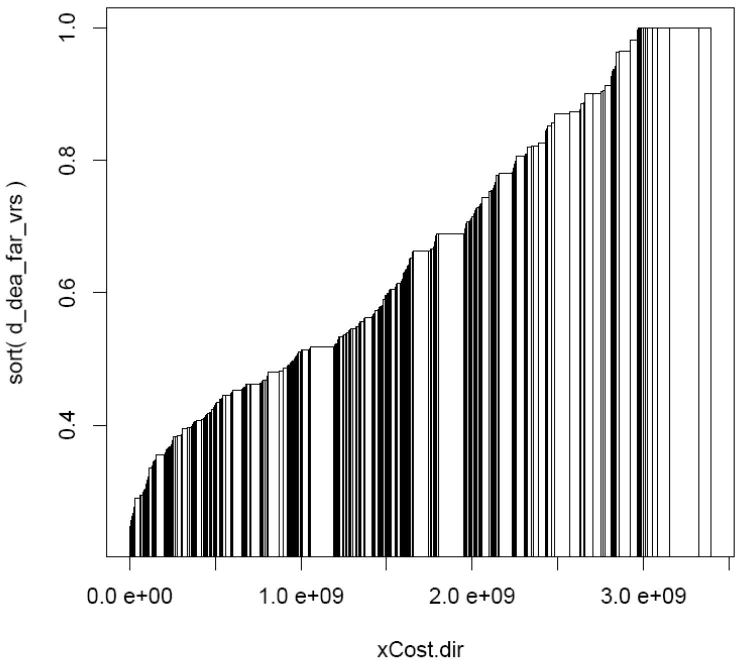

Figure 6-4 Salter diagram DEA-VRS gas model 8

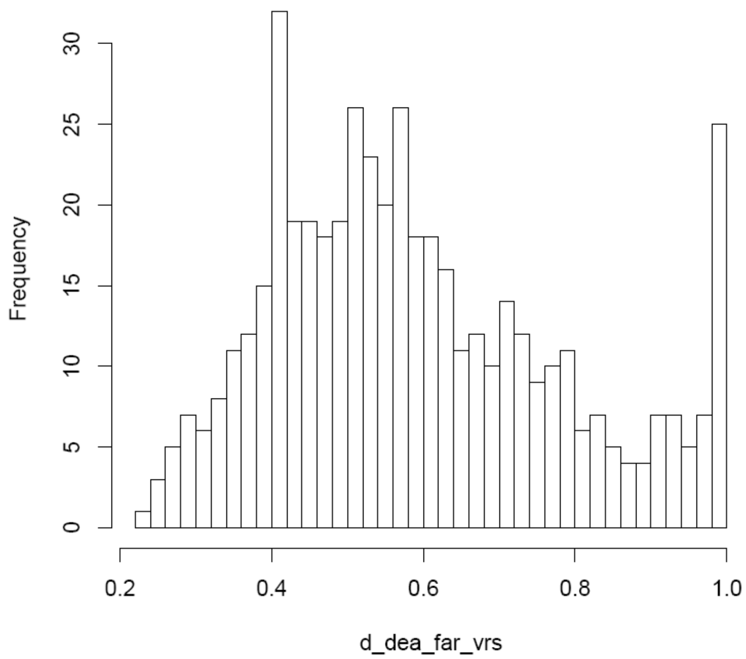

Figure 6-5 Histogram DEA-VRS gas model 8

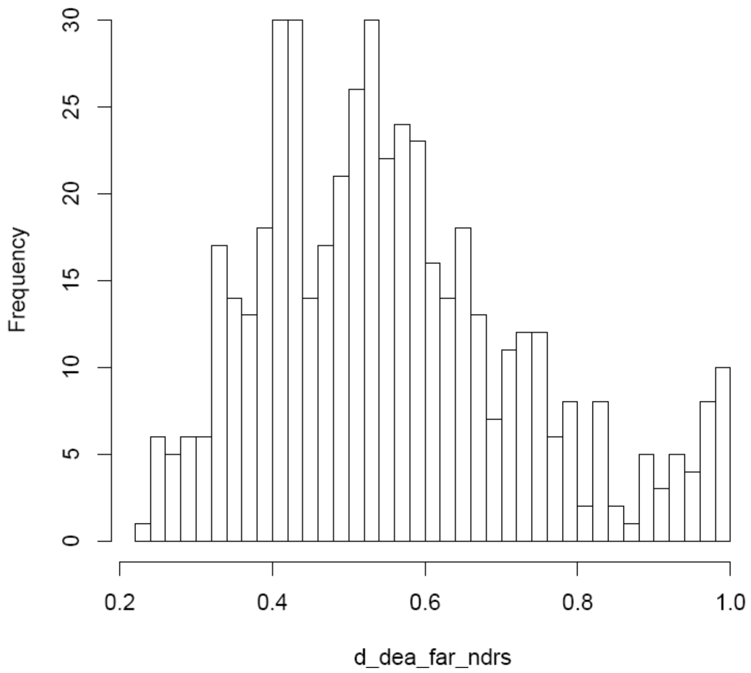

Figure 6-6 Histogram DEA-NDRS gas model 87

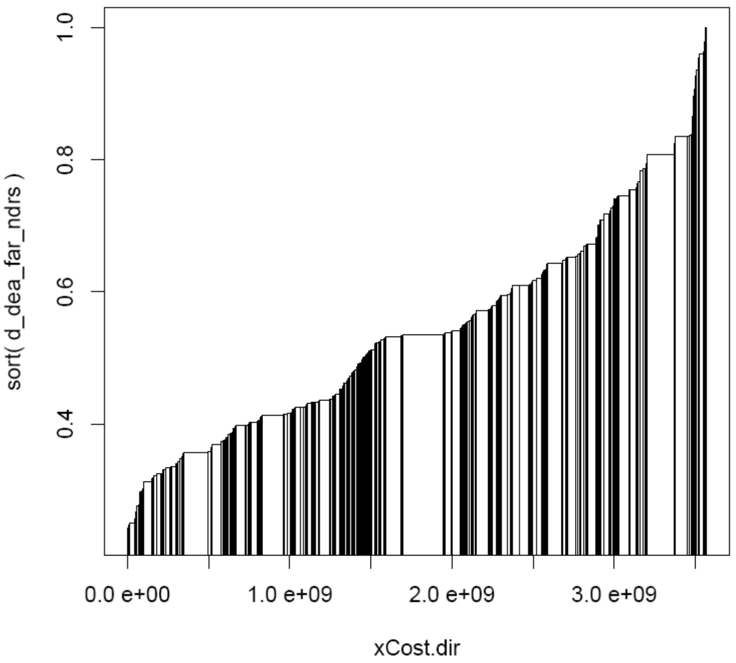

Figure 6-7 Salter diagram DEA-NDRS gas model 8

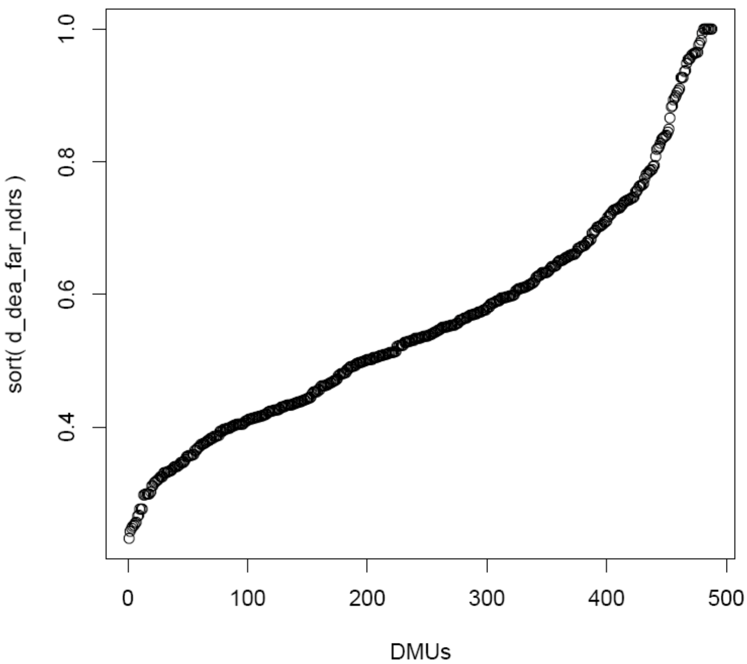

Figure 6-8 Distribution of scores DEA-NDRS gas model 8

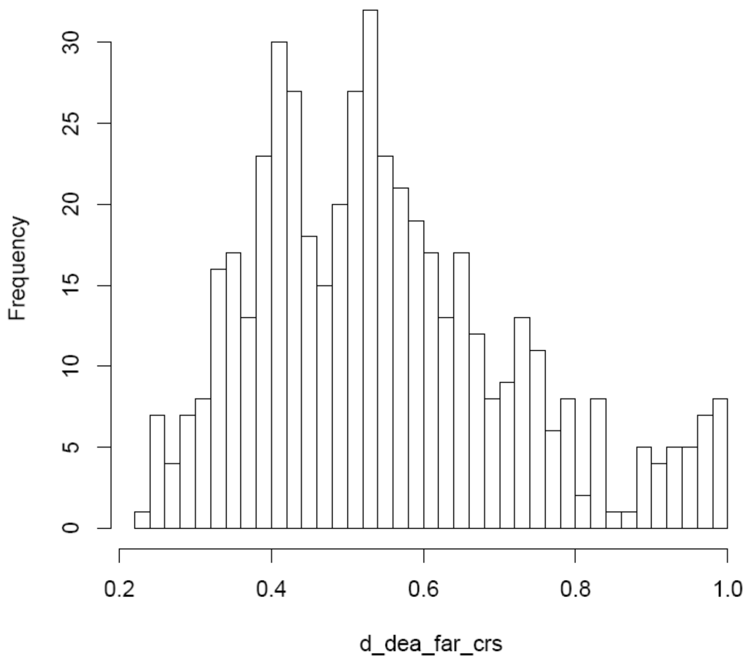

Figure 6-9 Histogram gas model 8 DEA-CRS

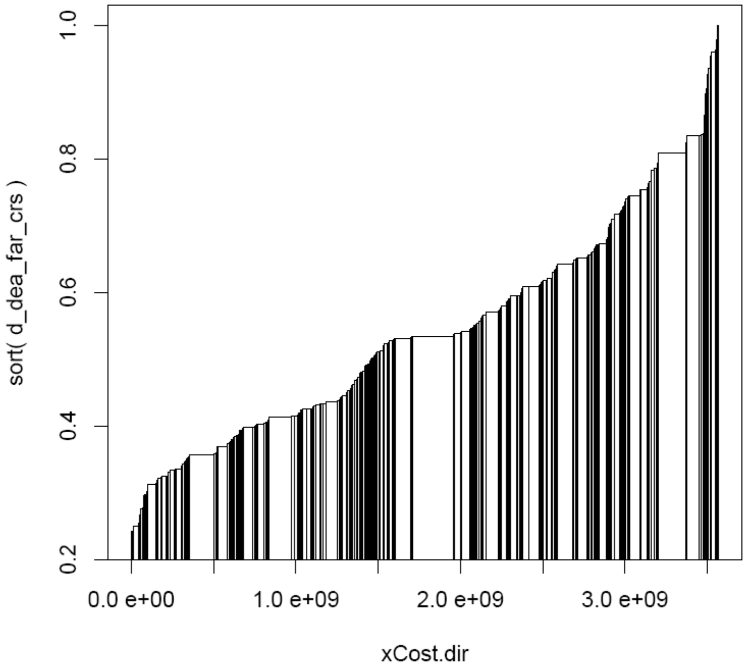

Figure 6-10 Salter diagram gas model 8 DEA-CRS

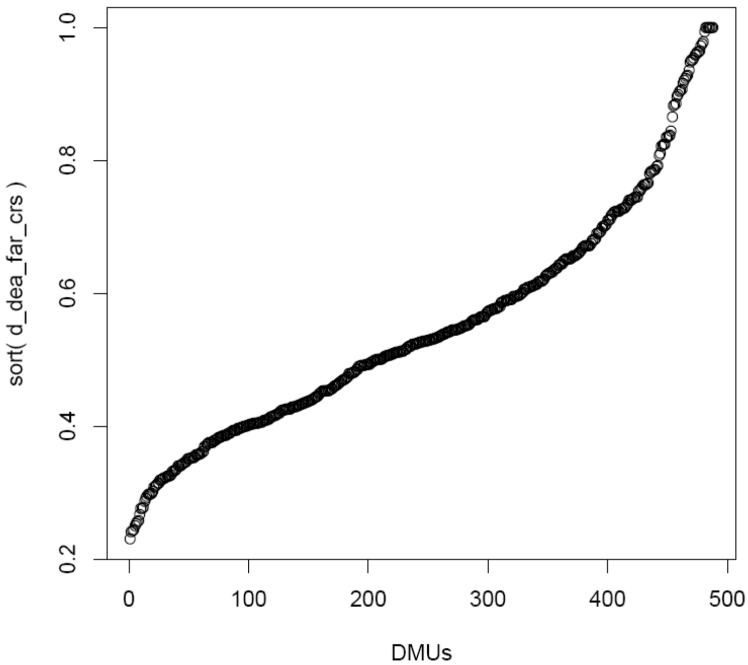

Figure 6-11 Distribution of scores gas model 8 DEA-CRS

Scale efficiency results

6.238 The average scale efficiency is 0.936 with a standard deviation of 0.08. The results in delivered energy (nm3) in Figure 6-12 indicates diseconomies of scale from some 100,000,000 nm3 in model 7. As visible in the Figure, all benchmarking peers are below average size (in delivered energy), which suggests that the larger units do not offset any subadditivity gains in their actual costs.

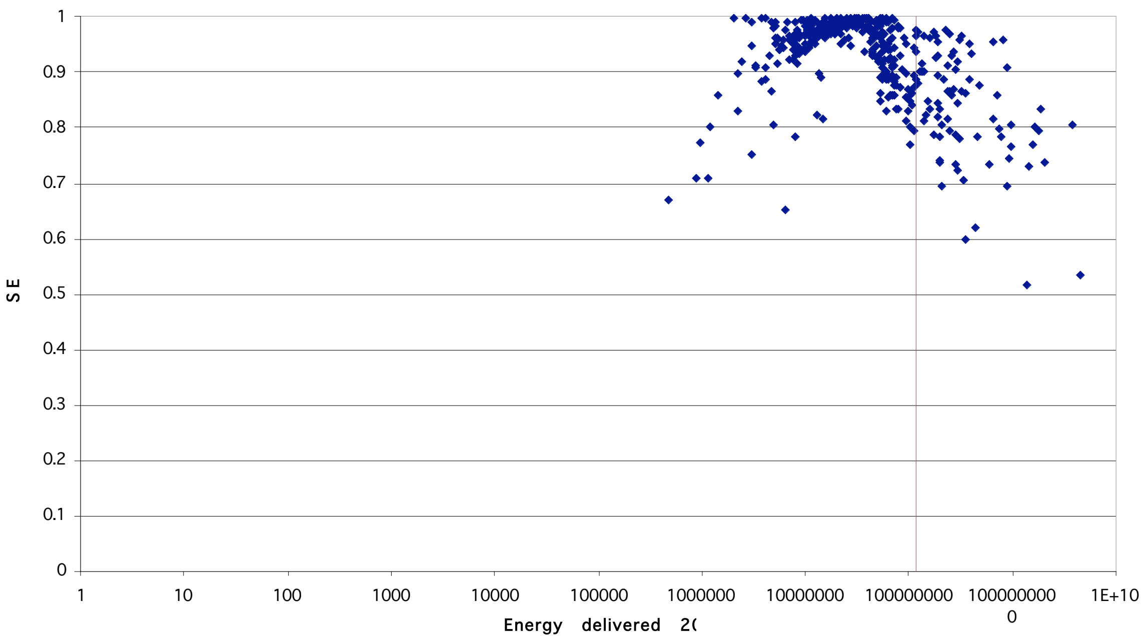

Figure 6-12 Scale efficiency estimates for gas model 8 DEA (red line depicts average size in delivered energy (nm3, 2004), logarithmic scale.

SFA results

6.239 The results for the stochastic frontier analysis in Figure 6-13 and Figure 6-13 show a fairly smooth distribution with broad mode around 0.5-0.8. No unit is classified as fully costefficient, since the error term is fairly large for the first data set. The Salter diagram shows a smooth envelopment and besides a small set of insignificant observations below 0.4, the stochastic model performs in a much more robust manner to the data errors than the deterministic DEA model above.

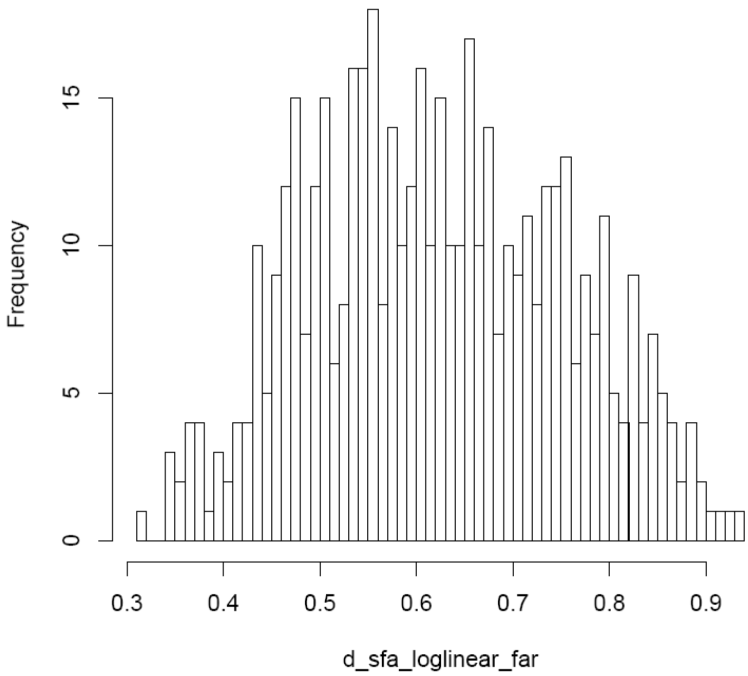

Figure 6-13 Histogram scores SFA loglinear for gas model 8

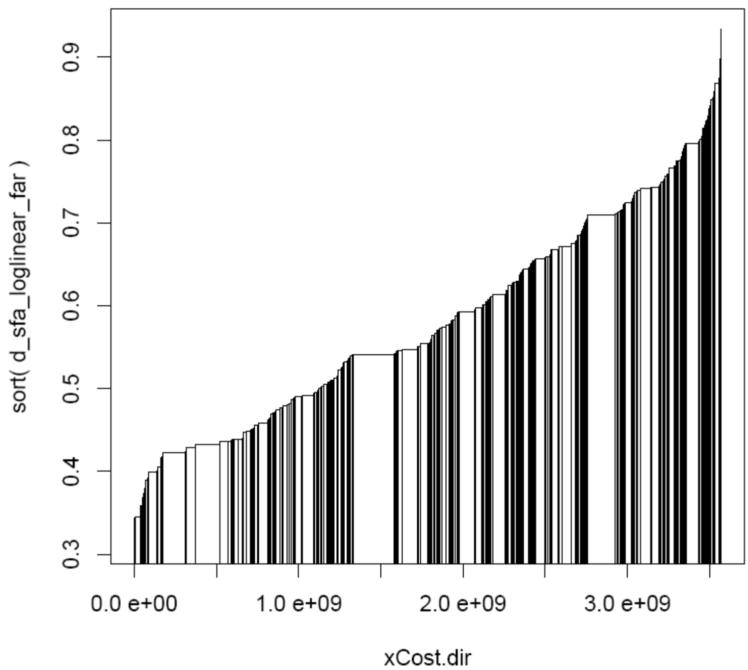

Figure 6-14 Salter diagram for SFA loglinear for gas model 8

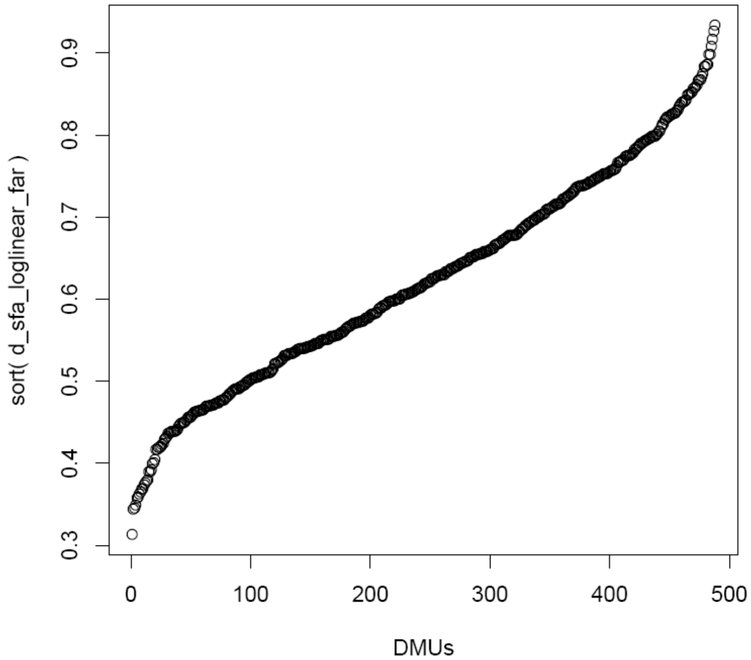

Figure 6-15 Distribution of scores SFA loglinear for gas model 8

6.7 Model consistency

6.240 The consistency of the model specification can be deduced from the correlations between scores and rank orders across various models in Table 6-10, Table 6-11 and Table 6-12 below for some relevant models (full results in Table 6-16 and below). The rank order consistency between deterministic and stochastic models for model 8 is exceptionally high (about Spearman 0.937), combined with a very high Pearson correlation (0.922) This means that the two methods rank the units in a similar order, albeit at somewhat different level. This relationship is somewhat illustrated in Figure 6-16 and Figure 6-17 where the model results for SFA and DEA exhibit a fairly continuous concave shape.

Table 6-10 Pearson correlation of scores gas (exerpt from Table 6-16)

PEARSON	d_dea_far_ndrs	d_dea_far_crs	sfa_loglinear_far
d_dea_far_ndrs	1.000	0.992	0.922
d_dea_far_crs		1.000	0.922
sfa_linear_far			1.000

6.241 The Pearson product-moment correlation test provides information on whether the scores come from related data generating processes or not. In the case above, the Pearson test is significant (d_dea_far_ndrs and sfa_loglinear_far), t = 52.544, df = 294862, p-value = 2.220e-16, alternative hypothesis: true correlation is not equal to 0, 95 percent confidence interval: [0.9076560 0.9344078]. However, since the classical Pearson test assumes that one variable is normally distributed and moreover is highly sensitive to nonlinearities in the sample, we opt for two non-parametric tests; Kendall and Spearman.

Table 6-11 Kendall's rank correlation gas model 8 (exerpt from Table 6-17)

KENDALL	d_dea_far_ndrs	d_dea_far_crs	sfa_loglinear_far
d_dea_far_ndrs	1.000	0.957	0.797
d_dea_far_crs		1.000	0.796
sfa_linear_far			1.000

6.242 Kendall's rank-order correlation is a non-parametric test that applies to data after an ordinal transformation. The rank correlations above are significant, data, the lowest being DEA-CRS vs SFA with metric z = 26.2754, risk of accidental correlation of rank p-value < 2.2e-16.

Table 6-12 Spearman's rank correlation gas model 8 (exerpt from Table 6-18)

SPEARMAN	d_dea_far_ndrs	d_dea_far_crs	sfa_loglinear_far
d_dea_far_ndrs	1.000	0.993	0.937
d_dea_far_crs		1.000	0.938
sfa_linear_far			1.000

6.243 Spearman's rho is a 'quasi-ordinal' correlation coefficient. It is equivalent to the Pearson's r correlation, computed on the variables, after they have been transformed into an ordinal scale of measurement. The test is similar to Kendall and varies primarily in terms of interpretation of the metric. All rank-order correlations between DEA-NDRS, CRS and SFA translog are significant for gas model 8. Example for lowest correlation, DEA-NDRS vs SFA, metric S = 1218943, risk of accidental correlation of rank (p-value) < 2.2e-16. The test only assumes that the correlation between the variables is linear, and that the sample has been selected randomly from the population, i.e., no normality is assumed.

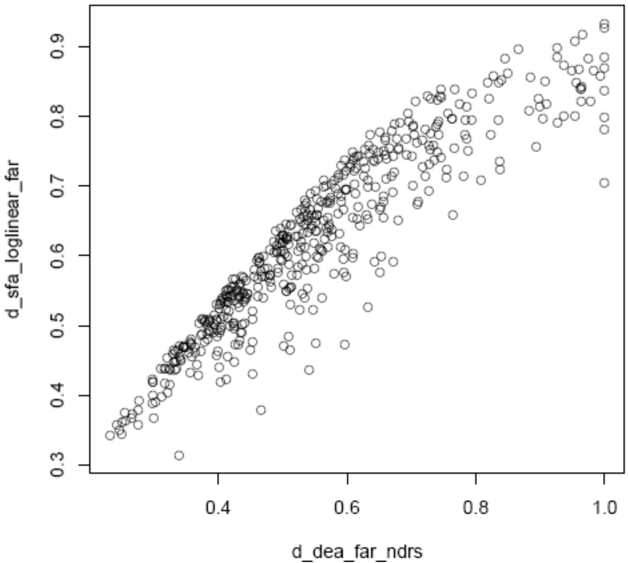

Figure 6-16 Correlation between DEA-NDRS and SFA results, gas model 8.

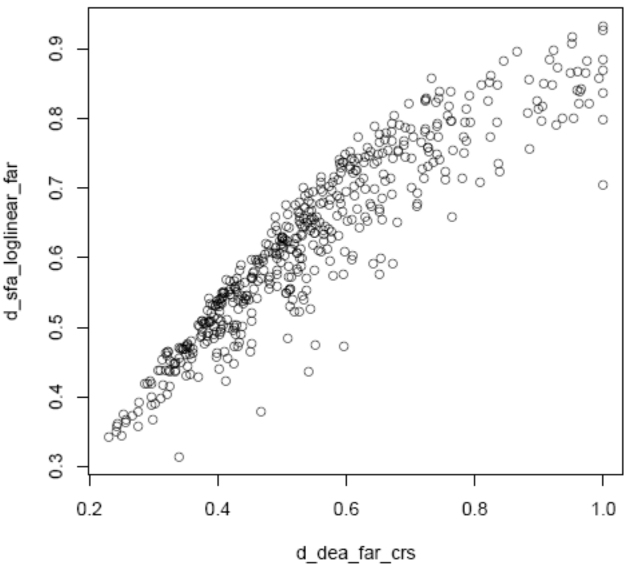

Figure 6-17 Correlation between DEA-CRS and SFA results, gas model 8.

6.8 Second-stage analysis gas model 8

6.244 The statistical analysis of the benchmarking model with respect to structural variables is performed using an OLS regression. We only present non-truncated results from an OLS model f-test on a normal NDRS model, although the qualitative findings are validated using extensive analyses on superefficiency models under both CRS, NDRS and VRS assumptions.

6.245 The presentation is made for DEA-NDRS and SFA loglinear. We test the fit of a model of the type

$$E = c^{T}z + e$$

6.246 The z-variables are tested groupwise in logical groups, soiltype, buldings etc, as to avoid interaction effects. For each test, the estimate c, the standard error for the coefficient, the t-value and the probability that the coefficient is identified as noise. The latter p-value is marked in bold face black when lower than 0.01 and bold face red if it is significant at the 0.001 level.

6.9 Second-stage results for DEA-NDRS

6.247 The only significant second-stage variables for DEA-NDRS is the weighted age of ND pipeline (zAge.hd) at the 0.001-level with a negative impact. The result (that is confirmed using a superefficiency framework) suggests a bias in favor of younger ND networks for the NDRS model.

Second stage estimates for model d_dea_far_ndrs SealedArea

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.555	0.009	63.960	**0.000**
zSealedAreaAsset	0.000	0.000	-0.193	0.847
zSealedAreaServed	0.000	0.000	0.670	0.503

Second stage estimates for model d_dea_far_ndrs SoilArea

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.556	0.009	64.483	**0.000**
zSoilAreaAsset	0.000	0.000	0.027	0.978
zSoilAreaServed	0.000	0.000	1.014	0.311

Second stage estimates for model d_dea_far_ndrs Soil

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.524	0.045	11.744	**0.000**
zSoiltype.1	-0.075	0.079	-0.942	0.346
zSoiltype.2	0.005	0.097	0.053	0.958
zSoiltype.3	0.077	0.049	1.588	0.113
zSoiltype.4	0.018	0.046	0.384	0.701
zSoiltype.5	0.054	0.048	1.115	0.265
zSoiltype.6	0.034	0.051	0.658	0.511

Second stage estimates for model d_dea_far_ndrs Operator

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.594	0.014	42.757	0.000
zOperator.dis.heat	-0.037	0.016	-2.298	0.022
zOperator.gas.hol	-0.033	0.029	-1.170	0.243
zOperator.sew.net	0.015	0.027	0.566	0.571
zOperator.water	-0.029	0.017	-1.684	0.093

Second stage estimates for model d_dea_far_ndrs Energy

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.550	0.009	62.192	0.000
zEnergy.in.2000	0.000	0.000	-0.978	0.329
zEnergy.in.2001	0.000	0.000	-0.966	0.335
zEnergy.in.2002	0.000	0.000	1.727	0.085
zEnergy.in.2003	0.000	0.000	0.896	0.371

Second stage estimates for model d_dea_far_ndrs Q

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.555	0.008	65.510	0.000
zQLeakage.md	0.001	0.000	2.298	0.022
zQ.Leakage.nd	0.000	0.000	-1.463	0.144
zQ.Disturbance.hd	0.000	0.000	0.455	0.649
zQ.Leakage.hd	-0.001	0.001	-0.718	0.473

Second stage estimates for model d_dea_far_ndrs Peakload

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.546	0.011	48.212	0.000
zPeakload.async.out.2000	0.000	0.000	-0.545	0.586
zPeakload.async.out.2004	0.000	0.000	-0.402	0.688
zPeakload.out.async.2001	0.000	0.000	0.318	0.751
zPeakload.out.async.2002	0.000	0.000	0.134	0.893
zPeakload.out.async.2003	0.000	0.000	0.290	0.772

Second stage estimates for model d_dea_far_ndrs Heigh

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.568	0.013	45.092	0.000
zHeight	0.000	0.000	-1.337	0.182
zHeightdiff	0.000	0.000	0.548	0.584

Second stage estimates for model d_dea_far_ndrs Tilt

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.568	0.012	48.953	0.000
zTilt	-0.005	0.004	-1.100	0.272

Second stage estimates for model d_dea_far_ndrs
Loc

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.563	0.009	63.101	0.000
zLoc	-0.024	0.019	-1.283	0.200

Second stage estimates for model d_dea_far_ndrs
Building

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.556	0.009	63.935	0.000
zBuildingsAreaServed	0.000	0.000	0.404	0.687
zBuildingsAreaAsset	0.000	0.000	-0.174	0.862

Second stage estimates for model d_dea_far_ndrs
Meter

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.558	0.008	69.238	0.000
zMeters.hd	0.000	0.000	0.027	0.979
zMeters.md	0.000	0.000	-0.570	0.569

Second stage estimates for model d_dea_far_ndrs
Age

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.526	0.032	16.406	0.000
zAge	0.005	0.003	2.109	0.035
zAge.hd	0.000	0.001	0.127	0.899
zAge.md	0.001	0.001	0.811	0.418
zAge.nd	-0.005	0.001	-4.345	0.000

6.10 Second-stage results for SFA loglinear

6.248 As for the DEA-NDRS specification, the only significant second-stage variables for the stochastic frontier analysis model using a loglinear functional form is the average age of ND pipelines (zAge.nd) at the 0.01-level.

Second stage estimates for model d_sfa_loglinear_far
SealedArea

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.625	0.007	94.792	0.000
zSealedAreaAsset	0.000	0.000	-0.496	0.620
zSealedAreaServed	0.000	0.000	-0.211	0.833

Second stage estimates for model d_sfa_loglinear_far
SoilArea

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.624	0.007	95.186	0.000
zSoilAreaAsset	0.000	0.000	-0.256	0.798
zSoilAreaServed	0.000	0.000	0.164	0.870

Second stage estimates for model d_sfa_loglinear_far
Soil

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.588	0.034	17.237	0.000
zSoiltype.1	-0.023	0.060	-0.373	0.709
zSoiltype.2	0.032	0.074	0.436	0.663
zSoiltype.3	0.065	0.037	1.747	0.081
zSoiltype.4	0.029	0.035	0.834	0.405
zSoiltype.5	0.038	0.037	1.044	0.297
zSoiltype.6	0.033	0.039	0.840	0.401

Second stage estimates for model d_sfa_loglinear_far
Operator

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.888	0.021	41.801	0.000
zOperator.dis.heat	-0.019	0.025	-0.768	0.443
zOperator.gas.hol	-0.066	0.044	-1.505	0.133
zOperator.sew.net	-0.001	0.041	-0.036	0.971
zOperator.water	0.002	0.026	0.069	0.945

Second stage estimates for model d_sfa_loglinear_far
Energy

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.625	0.007	92.414	0.000
zEnergy.in.2000	0.000	0.000	-1.237	0.217
zEnergy.in.2001	0.000	0.000	-0.550	0.583
zEnergy.in.2002	0.000	0.000	1.004	0.316
zEnergy.in.2003	0.000	0.000	1.236	0.217

Second stage estimates for model d_sfa_loglinear_far
Quality

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.625	0.006	96.804	0.000
zQLeakage.md	0.000	0.000	1.362	0.174
zQ.Leakage.nd	0.000	0.000	-1.510	0.132
zQ.Disturbance.hd	0.000	0.000	-0.327	0.744
zQ.Leakage.hd	-0.001	0.001	-1.097	0.273

Second stage estimates for model d_sfa_loglinear_far
Peakload

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.620	0.008	74.253	0.000
zPeakload.async.out.2000	0.000	0.000	-0.995	0.321
zPeakload.async.out.2004	0.000	0.000	-0.357	0.721
zPeakload.out.async.2001	0.000	0.000	0.677	0.499
zPeakload.out.async.2002	0.000	0.000	-0.083	0.934
zPeakload.out.async.2003	0.000	0.000	0.659	0.510

Second stage estimates for model d_sfa_loglinear_far
Heigh

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.636	0.010	66.472	0.000
zHeight	0.000	0.000	-0.491	0.624
zHeightdiff	0.000	0.000	-0.822	0.411

Second stage estimates for model d_sfa_loglinear_far
Tilt

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.632	0.009	71.620	0.000
zTilt	-0.004	0.003	-1.193	0.233

Second stage estimates for model d_sfa_loglinear_far
Loc

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.627	0.007	92.375	0.000
zLoc	-0.017	0.014	-1.167	0.244

Second stage estimates for model d_sfa_loglinear_far
Building

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.626	0.007	94.860	0.000
zBuildingsAreaServed	0.000	0.000	-0.373	0.709
zBuildingsAreaAsset	0.000	0.000	-0.504	0.614

Second stage estimates for model d_sfa_loglinear_far
Meter

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.558	0.008	69.238	0.000
zMeters.hd	0.000	0.000	0.027	0.979
zMeters.md	0.000	0.000	-0.570	0.569

Second stage estimates for model d_sfa_loglinear_far
Connection points

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.621	0.006	101.043	0.000
zConnectpoint.out.dso.hd	0.000	0.000	0.742	0.458
zConnectpoint.out.dso.md	0.000	0.000	1.001	0.317

Second stage estimates for model d_sfa_loglinear_far
Age

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.577	0.025	23.416	**0.000**
zAge	0.005	0.002	2.360	0.019
zAge.hd	0.000	0.001	-0.142	0.887
zAge.md	0.000	0.001	0.436	0.663
zAge.nd	-0.002	0.001	-2.861	**0.004**

6.11 Comments on the gas results

6.249 The relatively low efficiency scores in the gas model 8 raise the question if the efficiency in this sector is really that low and in particular if it is substantially lower than in the electricity sector.

6.250 The first thing to observe is that although the efficiency is low, with an average DEA crs efficiency of 0.549, this is quite a bit higher than the efficiency one gets using more traditional and simplistic measures of variation. If for example we use the direct cost per energy delivered as the performance measure and calculate efficiencies hereon (as the minimal (xCost.dir/yEnergy.delivered) divided by the realized (xCosts.dir/yEnergy.delivered)) we obtain an average efficiency of 0.395.

6.251 The second thing to observe is that the level of efficiency cannot directly be compared between models with a different number of observations and a different number of outputs. The gas model is estimated on more observations and it is based on fewer outputs than many international models and the German electricity model in particular. Both factors will tend to lower the efficiencies, in particular the DEA-efficiencies, in the gas model. The reason is that the DEA cost model approximation is biased towards higher costs and that this bias is larger the fewer observations and the more outputs one uses since both tend to make possible comparisons fewer.

6.252 This bias was first analyzed systematically in Zhang and Bartels (1998) to correct for sample size effects in international comparisons of electricity distribution models but it is a well-known and often observed phenomenon among users of these techniques. Zhang and Bartels (1998) also provide simulations to evaluate this impact under some more specific conditions of the number of observations. In a one-output two-input model, they find for example that the efficiency can decrease with some with 7.6 % point (from 0.685 to 0.609) if we move from a sample with 20 to a sample with 200 observations. The impact of the number of dimensions is only partially investigated in Zhang and Bartels (1998) but the simulations show that also this will have an impact although smaller. Zhang and Bartels(1998) also find that the mean efficiency is particularly sensitive to the number of observations for small values of the sample, i.e. the rate of decrease in the mean technical efficiency slows down as the number of firms increases. Since both the electricity and the gas models are based on sample sizes that are large in an international perspective, this would tend to make the impact less dramatic.

6.253 In a further analysis of the gas and electricity models, it would be natural to investigate the impact of the number of units and number of dimensions further. The first part is quite straightforward. We could draw a random sample of say 100 gas units and calculate the efficiency in this model. Doing this say 2000 times we would get a good impression of the expected efficiency in the gas model estimated on 100 units. We could then repeat this with sample of sizes of 150, 200 etc. This would allow corrections for sample size differences in the two models much like corrections based on a second stage analysis.

6.254 The analysis of the impact of the number of variables in a German context is more difficult since changing the number of variables also changes the nature and interpretation of the model. We can nevertheless get some impression of the impact by the robustness analysis we have done using the loops estimating efficiencies for some 3000-5000 model specifications. We have for example found that as the number of outputs in electricity models increases from 2 to 9, the level of the efficiency score tends to increase 20% point, cf Figure 6-18 below. These are of course rough comparisons and the lower number of observations in the electricity than in the gas model may tend to make the impact of the number of dimensions more extreme.

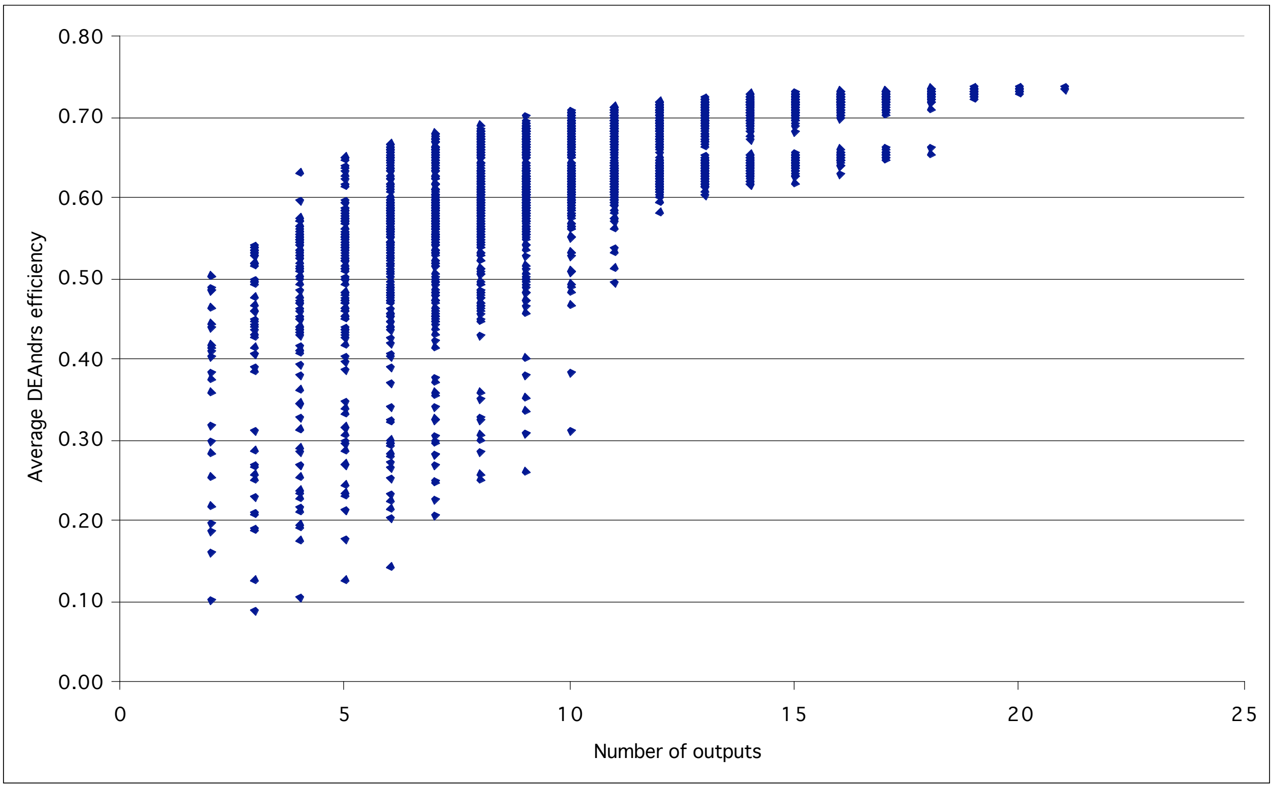

Figure 6-18 Relationship between average DEA score (NDRS) and the number of output dimensions, electricity.

6.255 Another way to get an impression of the importance of the number of comparators is via the bias corrections calculated in the two models using bootstrapping. For the electricity models the average bias increases efficiency with 4.8% point while the bias in the gas model is only 1.1% point.

Model revisions

6.256 The T3 results for gas show some indications of capital valuation influencing the results. Further work on the standardization of the capital cost component, including the potential revaluation of "hidden" capital may increase the scores if the abnormal units form a large part of the frontier, which may be the case in the current data set.

6.257 As a final observation we note that although the gas model specification is effective as a measure of static efficiency, there are ongoing discussions as to the importance of not only present but also future or potential customers. Refinement of the model including to capture the growth potential, as proxies of the potential customer base, may increase the efficiency level and make the model closer to actual decision making.

Summary

6.258 Considering the overall design, significance level, and consistency of the model in relation to detected issues related to data and model structure, we find convincing evidence of material cost inefficiency for the German gas DSO sector. Although the present estimates should be regarded as an upper bound of the inefficiency, there is no reason to believe that the true scores after correction of the issues mentioned above would exceed the efficiency scores obtained for electricity distribution, i.e. we would expect inefficiencies of at least 20% in average direct cost.

Table 6-13 Pearson correlations for the electricity model (sample 328 firms)

PEARSON	d_cols_far	d_dea_far_vrs	d_dea_far_drs	d_dea_far_ndrs	d_dea_far_crs	d_dea_sup_far_vrs	d_dea_sup_far_drs	d_dea_sup_far_ndrs	d_dea_sup_far_crs	d_sfa_translog_far
d_cols_far	1.000	0.005	0.048	-0.082	-0.038	0.018	0.025	-0.067	0.004	0.020
d_dea_far_vrs	0.005	1.000	0.997	0.945	0.946	0.149	0.148	0.648	0.645	0.735
d_dea_far_drs	0.048	0.997	1.000	0.940	0.945	0.150	0.149	0.646	0.646	0.736
d_dea_far_ndrs	-0.082	0.945	0.940	1.000	0.997	0.161	0.159	0.701	0.695	0.733
d_dea_far_crs	-0.038	0.946	0.945	0.997	1.000	0.162	0.161	0.700	0.698	0.736
d_dea_sup_far_vrs	0.018	0.149	0.150	0.161	0.162	1.000	1.000	0.345	0.350	0.066
d_dea_sup_far_drs	0.025	0.148	0.149	0.159	0.161	1.000	1.000	0.343	0.350	0.066
d_dea_sup_far_ndrs	-0.067	0.648	0.646	0.701	0.700	0.345	0.343	1.000	0.996	0.473
d_dea_sup_far_crs	0.004	0.645	0.646	0.695	0.698	0.350	0.350	0.996	1.000	0.474
d_sfa_translog_far	0.020	0.735	0.736	0.733	0.736	0.066	0.066	0.473	0.474	1.000

Table 6-14 Kendall's rank correlations for the electricity model (sample 328 firms)

KENDALL	d_cols_far	d_dea_far_vrs	d_dea_far_drs	d_dea_far_ndrs	d_dea_far_crs	d_dea_sup_far_vrs	d_dea_sup_far_drs	d_dea_sup_far_ndrs	d_dea_sup_far_crs	d_sfa_translog_far
d_cols_far	1.000	0.226	0.242	0.110	0.126	0.202	0.221	0.109	0.126	0.091
d_dea_far_vrs	0.226	1.000	0.983	0.837	0.837	0.985	0.969	0.832	0.832	0.525
d_dea_far_drs	0.242	0.983	1.000	0.819	0.837	0.968	0.986	0.814	0.832	0.530
d_dea_far_ndrs	0.110	0.837	0.819	1.000	0.983	0.843	0.825	0.994	0.977	0.534
d_dea_far_crs	0.126	0.837	0.837	0.983	1.000	0.843	0.842	0.977	0.994	0.541
d_dea_sup_far_vrs	0.202	0.985	0.968	0.843	0.843	1.000	0.980	0.847	0.845	0.522
d_dea_sup_far_drs	0.221	0.969	0.986	0.825	0.842	0.980	1.000	0.827	0.846	0.526
d_dea_sup_far_ndrs	0.109	0.832	0.814	0.994	0.977	0.847	0.827	1.000	0.981	0.530
d_dea_sup_far_crs	0.126	0.832	0.832	0.977	0.994	0.845	0.846	0.981	1.000	0.537
d_sfa_translog_far	0.091	0.525	0.530	0.534	0.541	0.522	0.526	0.530	0.537	1.000

Table 6-15 Spearman's rank correlations for the electricity model (sample 328 firms)

SPEARMAN	d_cols_far	d_dea_far_vrs	d_dea_far_drs	d_dea_far_ndrs	d_dea_far_crs	d_dea_sup_far_vrs	d_dea_sup_far_drs	d_dea_sup_far_ndrs	d_dea_sup_far_crs	d_sfa_translog_far
d_cols_far	1.000	0.323	0.347	0.159	0.183	0.293	0.323	0.157	0.183	0.132
d_dea_far_vrs	0.323	1.000	0.997	0.931	0.932	0.997	0.995	0.930	0.931	0.699
d_dea_far_drs	0.347	0.997	1.000	0.924	0.931	0.994	0.998	0.924	0.930	0.705
d_dea_far_ndrs	0.159	0.931	0.924	1.000	0.997	0.950	0.944	0.999	0.997	0.717
d_dea_far_crs	0.183	0.932	0.931	0.997	1.000	0.951	0.950	0.996	0.999	0.724
d_dea_sup_far_vrs	0.293	0.997	0.994	0.950	0.951	1.000	0.997	0.952	0.952	0.703
d_dea_sup_far_drs	0.323	0.995	0.998	0.944	0.950	0.997	1.000	0.944	0.951	0.707
d_dea_sup_far_ndrs	0.157	0.930	0.924	0.999	0.996	0.952	0.944	1.000	0.997	0.716
d_dea_sup_far_crs	0.183	0.931	0.930	0.997	0.999	0.952	0.951	0.997	1.000	0.722
d_sfa_translog_far	0.132	0.699	0.705	0.717	0.724	0.703	0.707	0.716	0.722	1.000

Table 6-16 Pearson correlations for gas model 8 (sample 488 firms)

PEARSON	d_cols_far	d_dea_far_vrs	d_dea_far_drs	d_dea_far_ndrs	d_dea_far_crs	d_dea_sup_far_drs	d_dea_sup_far_ndr	d_dea_sup_far_crs	d_sfa_loglinear_far	d_sfa_translog_far
d_cols_far	1.000	-0.114	-0.014	-0.208	-0.098	-0.050	-0.456	-0.143	-0.115	-0.149
d_dea_far_vrs	-0.114	1.000	0.994	0.946	0.947	0.958	0.868	0.924	0.844	0.811
d_dea_far_drs	-0.014	0.994	1.000	0.933	0.947	0.964	0.830	0.925	0.838	0.799
d_dea_far_ndrs	-0.208	0.946	0.933	1.000	0.992	0.898	0.925	0.973	0.922	0.873
d_dea_far_crs	-0.098	0.947	0.947	0.992	1.000	0.912	0.891	0.981	0.922	0.867
d_dea_sup_far_drs	-0.050	0.958	0.964	0.898	0.912	1.000	0.828	0.922	0.796	0.755
d_dea_sup_far_ndrs	-0.456	0.868	0.830	0.925	0.891	0.828	1.000	0.909	0.828	0.789
d_dea_sup_far_crs	-0.143	0.924	0.925	0.973	0.981	0.922	0.909	1.000	0.896	0.844
d_sfa_loglinear_far	-0.115	0.844	0.838	0.922	0.922	0.796	0.828	0.896	1.000	0.978
d_sfa_translog_far	-0.149	0.811	0.799	0.873	0.867	0.755	0.789	0.844	0.978	1.000

Table 6-17 Kendall's rank correlations for gas model 8 (sample 488 firms)

KENDALL	d_cols_far	d_dea_far_vrs	d_dea_far_drs	d_dea_far_ndrs	d_dea_far_crs	d_dea_sup_far_drs	d_dea_sup_far_ndr	d_dea_sup_far_crs	d_sfa_loglinear_far	d_sfa_translog_far
d_cols_far	1.000	-0.008	0.030	-0.118	-0.077	0.023	-0.118	-0.077	-0.138	-0.151
d_dea_far_vrs	-0.008	1.000	0.960	0.859	0.857	0.959	0.859	0.856	0.701	0.667
d_dea_far_drs	0.030	0.960	1.000	0.819	0.863	0.999	0.819	0.863	0.694	0.652
d_dea_far_ndrs	-0.118	0.859	0.819	1.000	0.957	0.822	1.000	0.957	0.797	0.744
d_dea_far_crs	-0.077	0.857	0.863	0.957	1.000	0.865	0.956	1.000	0.796	0.734
d_dea_sup_far_drs	0.023	0.959	0.999	0.822	0.865	1.000	0.822	0.865	0.699	0.654
d_dea_sup_far_ndrs	-0.118	0.859	0.819	1.000	0.956	0.822	1.000	0.956	0.797	0.744
d_dea_sup_far_crs	-0.077	0.856	0.863	0.957	1.000	0.865	0.956	1.000	0.796	0.734
d_sfa_loglinear_far	-0.138	0.701	0.694	0.797	0.796	0.699	0.797	0.796	1.000	0.887
d_sfa_translog_far	-0.151	0.667	0.652	0.744	0.734	0.654	0.744	0.734	0.887	1.000

Table 6-18 Spearman's rank correlations for gas model 8 (sample 488 firms)

SPEARMAN	d_cols_far	d_dea_far_vrs	d_dea_far_drs	d_dea_far_ndrs	d_dea_far_crs	d_dea_sup_far_drs	d_dea_sup_far_ndr	d_dea_sup_far_crs	d_sfa_loglinear_far	d_sfa_translog_far
d_cols_far	1.000	-0.008	0.047	-0.175	-0.114	0.036	-0.175	-0.115	-0.204	-0.223
d_dea_far_vrs	-0.008	1.000	0.994	0.959	0.961	0.994	0.959	0.961	0.864	0.838
d_dea_far_drs	0.047	0.994	1.000	0.946	0.961	1.000	0.946	0.961	0.860	0.829
d_dea_far_ndrs	-0.175	0.959	0.946	1.000	0.993	0.948	1.000	0.993	0.937	0.903
d_dea_far_crs	-0.114	0.961	0.961	0.993	1.000	0.963	0.993	1.000	0.938	0.899
d_dea_sup_far_drs	0.036	0.994	1.000	0.948	0.963	1.000	0.948	0.963	0.865	0.832
d_dea_sup_far_ndrs	-0.175	0.959	0.946	1.000	0.993	0.948	1.000	0.993	0.937	0.903
d_dea_sup_far_crs	-0.115	0.961	0.961	0.993	1.000	0.963	0.993	1.000	0.938	0.899
d_sfa_loglinear_far	-0.204	0.864	0.860	0.937	0.938	0.865	0.937	0.938	1.000	0.979
d_sfa_translog_far	-0.223	0.838	0.829	0.903	0.899	0.832	0.903	0.899	0.979	1.000

References

Agrell, P. J and P. Bogetoft (2000), *Ekonomisk Nätbesiktning*. Final Report, Swedish Energy Agency STEM, SUMICSID AB. (In Swedish)

Agrell, P.J. and P. Bogetoft, (2001), *On Dynamic Regulation of Natural Monopolies Using Frontier Analysis*, invited talk at CREPP, Université de Liège, 2001

Agrell, P.J. and P. Bogetoft, (2002), *Ekonomisk Nätbesiktning 2000, Report* ER 7:2002, Swedish Energy Agency STEM (in Swedish), SUMICSID AB.

Agrell, P.J. and P. Bogetoft, (2003a), *Dynamic Regulation, Report* AG2-V1, Norwegian Energy Directorate NVE, SUMICSID AB.

Agrell, P.J. and P. Bogetoft, (2003b), *Norm Models, Report* AG2-V2, Norwegian Energy Directorate NVE, SUMICSID AB.

Agrell, P.J. and P. Bogetoft, (2003c), *Use of Menu of Incentive Contracts, Report* AG2-V4, Norwegian Energy Directorate NVE, SUMICSID AB.

Agrell, P.J. and P. Bogetoft, (2003d), *Benchmarking for Regulation, Report* FP4, Norwegian Energy Directorate NVE, SUMICSID AB.

Agrell, P.J. and P. Bogetoft, (2003e), *Integrated, Parallel Energy Regulation, Report* FP5, Norwegian Energy Directorate NVE, SUMICSID AB.

Agrell, P.J. and P. Bogetoft, (2004a), *NVE Network Cost Efficiency Model, Report*, Norwegian Energy Directorate NVE, SUMICSID AB.

Agrell, P.J. and P. Bogetoft (2004b), *Note on methodology: ECOM+ Project*, Note 2004-10-01, SUMICSID AB.

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 J. Tind, (2001) A Dual Approach to Nonconvex Frontier Models *Journal of Productivity Analysis*, vol. 16, pp. 129-147.

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

Banker, R.D., and H. Chang (2005), The super-efficiency procedure for outlier identification, not for ranking efficient units*, European Journal of Operational Research*, In Press.

Bogetoft, P. (1990), Strategic Responses to DEA Control, Working Paper, CBS.

Bogetoft, P. (1994), Incentive Efficient Production Frontiers: An Agency Perspective on DEA, Management Science, 40, pp.959-968.

Bogetoft, P. (1995), Incentives and Productivity Measurements, International Journal of Production Economics, 39, pp. 67-81.

Bogetoft, P. (1996), *DEA on Relaxed Convexity Assumptions*, Management Science, 42, pp. 457-465.

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

Bogetoft, P. (2000), DEA and Activity Planning under Asymmetric Information, *Journal of Productivity Analysis,* 13, pp. 7-48.

Bogetoft, P. and K. Nielsen (2005), Internet Based Benchmarking, *Journal of Group Decision and Negotiation*, 14 (3), 195-215.

Bogetoft, P. and D. Wang, (2005), Estimating the Potential Gains from Merger, *Journal of Productivity Analysis*, 23, pp. 145-171.

Bogetoft, P. and J.L. Hougaard (2003), Rational Inefficiency, *Journal of Productivity Analysis, 20, pp. 243-271.*

Bogetoft, P. J. Tama, J.Tind (2000), *Convex Input and Output Projections of Nonconvex Production Possibility Sets*, Management Science, vol. 46(6), pp. 858-869.

Brännlund, R., R. Färe and S. Grosskopf (1995), *Environmental Regulation and Profitability: An Application to Swedish Pulp and Paper Mills,* Environmental and Resource Economics, 6, pp. 23-36.

Brännlund, R., Y. Chung, R. Färe and S. Grosskopf (1998), *Emission Trading And Profitability: The Swedish Pulp and Paper Industry*, Environmental and Resource Economics, 12, pp. 345-356.

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

Charnes, A., W. Cooper, A.Y. Lewin and L.M. Seiford (1994), *Data Envelopment Analysis: Theory, Methodology and Application*, Kluwer Academic Publishers.

Charnes, A., W.W. Cooper and E. Rhodes (1978), Measuring the Efficiency of Decision Making Units, *European Journal of Operational Research*, 2, pp. 429-444.

Charnes, A., W.W. Cooper and E. Rhodes (1979), Short Communication: Measuring the Efficiency of Decision Making Units, *European Journal of Operational Research*, 3, pp. 339.

Coelli, T.J., Estache, A., Perelman, S. and Trujillo, L. (2003), *A Primer on Efficiency Measurement for Utilities and Transport Regulators*, World Bank Publications, 129 pp

Cooper, W. W., Seiford, L. M. and Tone, K. (2000), *Data Envelopment Analysis*. Kluwer Academic Publishers.

Deprins, D., D. Simar, and H. Tulkens (1984), *Measuring Labor Efficiency in Post Offices*, pp. 243-267 in M. Marchand, P. Pestieau, and H. Tulkens, "The Performance of Public Enterprises: Concepts and Measurements", North Holland.

Edvardsen, D. F., and F. R. Førsund (2003), *International Benchmarking of Distribution Utilities*. Resource and energy Economics, 2003..

Estache, A. and D. Martimort (1999), *Politics, Transaction Costs, and the Design of Regulatory Institutions*. Policy Research Working Paper No. 2073, World Bank.

Farrell, M.J. (1957), The measurement of productive efficiency, *Journal of the Royal Statistical Society*, Series A, III, pp. 253-290.

Farrier Swier (2002), *Comparison of Building Blocks and Index-Based Approaches.* Utility Regulators Forum, Australia.

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

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

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

NEMESYS (2005), NEMESYS Project: *A Pan-Nordic Approach to Regulation of Distribution System Operations*. Final report for Nordenergi, SUMICSID AB

Petersen, N.C., *Data Envelopment Analysis on a Relaxed Set of Assumptions*, Management Science, 36, 1990, pp. 305-214.

Simar, L. and P. Wilson(2000), A general methodology for bootstrapping in nonparametric frontier models, *Journal of Applied Statistics*, 27, 779-802.

Simar, L. and P.W. Wilson (2004), *Estimation and Inference in Two-Stage, Semi-Parametric Models of Production Processes*, WP, Department of Economics, University of Texas at Austin, Austin, Texas, USA.

Tobin, James (1958), Estimation for relationships with limited dependent variables. *Econometrica* **26** (1), 24–36.

Tulkens, H., (1993), *On FDH Efficiency Analysis: Some Methodological Issues and Applications to Retail Banking, Courts, and Urban Transit*, The Journal of Productivity Analysis, 4, , pp. 183-210.

Wilson, P.W. (1993), Detecting outliers in deterministic nonparametric frontier models with multiple outputs, *Journal of Business and Economic Statistics* 11, 319-323.

Zhang, Y. and Bartels, R. (1998), The Effect of Sample Size on the Mean Efficiency in DEA with an Application to Electricity Distribution in Australia, Sweden and New Zealand, *Journal of Productivity Analysis*, 9, pp. 187–204.

SUMICSID

SUMICSID AB (GROUP)
Tunbyn 502
S-855 90 Sundsvall, SWEDEN

Tel:	+46 60 56 51 41
Fax:	+32 27 06 53 58

sweden @ sumicsid.com

SUMICSID SPRL (BELGIUM)
Rue de la Piété 11
B-1160 Bruxelles, BELGIUM

Tel:	+32 26 75 15 23
Fax:	+32 27 06 53 58

belgium @ sumicsid.com

SUMICSID APS (DENMARK)
Fru Ingesvej 19
DK-4180 Sorø, DENMARK

Tel	+45 57 83 15 18
Fax:	+32 27 06 53 58

denmark @ sumicsid.com

http://www.sumicsid.com