International Benchmarking and Regulation of European Gas Transmission Utilities

Final Report

Prepared for:
The Council of European Energy Regulators (CEER) –
Task Force on Benchmarking of Transmission Tariffs

Tooraj Jamasb
Department of Economics, University of Cambridge
Email: tooraj.jamasb@econ.cam.ac.uk

David Newbery
Department of Economics, University of Cambridge
Email: david.newbery@econ.cam.ac.uk

Michael Pollitt
Judge Business School, University of Cambridge
Email: m.pollitt@jbs.cam.ac.uk

Thomas Triebs
Department of Economics, University of Cambridge
Email: thomas.triebs@econ.cam.ac.uk

TABLE OF CONTENTS

List of Tables

List of Figures

Executive Summary

Aim

The present study by the Electricity Policy Research Group (EPRG) is commissioned by the Task Force on Benchmarking of Transmission Tariffs (BTT) of the Council of European Energy Regulators (CEER). The purpose of the study is to develop a framework for benchmarking of European gas transmission operators and to apply it to actual data.

Terms of Reference

- to help European regulators to develop more knowledge and experience in using benchmarks for their tariff regulation,
- identify differences in cost levels which can then be investigated further,
- to produce efficiency scores for individual European gas transmission operators,
- to perform a sensitivity analysis to assess the robustness of the results,
- to give recommendations on a best approach and model specifications,
- to give recommendations for standardizing European data.

Approach

This study uses a sample of US and European data to analyse cost drivers for gas transmission and produce individual efficiency scores. The benchmarking is done using the three most widely adopted frontier-based benchmarking methods:

- the non-parametric method Data Envelopment Analysis (DEA),
- the deterministic-parametric method Corrected Ordinary Least Square (COLS), and
- the stochastic method of Stochastic Frontier Analysis (SFA).

Data

The analytical part of the study uses data provided by the CEER and the EPRG. The study is based on data for about 40 US interstate gas transmission operators as well as four European operators. The data for the European operators is provided by the participating regulators.

Empirical Results

There is evidence of a consistent pattern across models and techniques for the efficiency scores and their ranking. Such consistency provides support for the validity of the approach and the correctness of the results.

Whereas one European operator performs close to the US average, the other operators perform worse.

Making the data more comparable, gathering more variables, and including further model specifications should lead to even better results.

Recommendations

Whereas comparison with US companies at least produces robust results for relative efficiency across European operators, a common European strategy for data standardization and gathering would be an important step towards European benchmarking that is fit for tariff setting.

Co-operation with other bodies involved in utility regulation such as the US Federal Energy Regulatory Commission (FERC), and the Australian energy regulators, should ensure the inclusion of world best practice in the benchmark, and improve the reliability of European benchmarking.

1. Introduction

Energy sector reforms are transforming the structure and operating environment of the gas and electricity industries in many countries around the world. The central aim of the reforms has been to introduce competition where possible and enhance regulation where necessary. Increasingly, energy sector reforms attempt to improve the efficiency of the natural monopoly segments of the industry, namely distribution and transmission, through regulatory reforms. The context for this study is incentive-based regulation and induced efficiency enhancements for gas transmission operators using benchmarking methods.

Regulatory reform of distribution and transmission utilities generally involves moving away from traditional cost-plus regulation towards incentive regulation. A number of incentive-based regulation models have been proposed in the literature.[1] These models are generally not attributed to theoretical advances in regulatory economics; rather, they reflect dissatisfaction with incentive signals, performance of cost-plus regulation and the need for alternative approaches.

In practice, many regulators have adopted some form of price and revenue cap regulation models based on the RPI-X formula. However, a crucial issue is how the utilities' efficiency requirements or X-factors are to be set. There are different approaches to the setting of X-factors and increasingly the favoured approach is through relative efficiency analysis and benchmarking of utilities.[2] To this end, countries such as the Netherlands, United Kingdom, and Norway have used utility benchmarking in the national context as part of the process of setting X-factors.

The aim of benchmarking is to reveal efficiency shortfalls amongst the regulated utilities. Benchmarking is used for identifying the most efficient firms in the sector and for measuring the relative performance of less efficient firms. Individual X-factors are then assigned based on their relative efficiency. Generally, the greater the inefficiency of a firm, the higher the efficiency requirement assigned to it. The purpose of individual X-factors is to provide firms with a realistic target to close the efficiency gaps between them.

---

[1] See Hall (2000), Hill (1995), Comnes et al. (1995), and Juskow and Schmalensee (1986) for reviews and comparisons of different incentive regulation models.

[2] See for example Jamasb and Pollitt (2002), Jamasb and Pollitt (2003), and DTe (1999) for reviews of alternative approaches.

However, most countries have a limited numbers of utilities, a situation that does not satisfy the data requirements of most analytical benchmarking techniques. Regulators can use cross-country efficiency analysis in order to evaluate performance of their utilities within the larger context of international practice. However, whilst international utility benchmarking has clear advantages, in order to enhance the reliability of the approach, the methodological and practical aspects, as well as the possible implications of this approach, need careful consideration. Empirical studies can be a useful instrument to identify and shed light on some of the issues arising in international benchmarking.

1.1. Terms of Reference

The present study is commissioned to undertake an analysis of the relative efficiencies of European gas transmission operators. The aim of the study is to examine the potential for, and the main issues involved in benchmarking for the purpose of the regulation of gas utilities. The terms of reference for the study are:

- to help European regulators to develop more knowledge and experience in using benchmarks for their tariff regulation,
- identify differences in cost levels which can then be investigated further,
- to produce efficiency scores for individual European gas transmission operators,
- to perform a sensitivity analysis to assess the robustness of the results,
- to give recommendations on model specifications,
- to give recommendations for standardizing European data.

Though cost differences as identified by this study might explain tariff differences this study does not look at actual tariffs in any way. The inefficiencies identified could be used to set tariffs under an incentive regulation model.

1.2. Organization of the Study

This report consists of ten chapters. Chapter 2 summarizes several prior studies on the efficiency measurement of individual gas utilities or relevant methodological issues. Chapter 3 introduces the benchmarking techniques used here. Chapter 4 argues in favour of our overall approach. Chapter 5 introduces the variables and describes the data. Chapter 6 gives an overview of the US market for gas transmission and its regulation. Chapter 7 provides a cost driver analysis prior to the actual benchmarking. Chapter 8 performs the benchmarking. Chapter 9 discusses the results. Chapter 10 indicates which further directions should be explored, and gives recommendations for benchmarking European gas transmission operators.

2. Literature Review

Previous studies of gas distribution and transmission companies as well as studies mainly concerned with benchmarking methodologies provide the literature relevant to the current study. This section gives the number of surveyed studies that include specific variables and use specific methods. Individual summaries can be found in Appendix A. Variable and method counts for a total of 18 relevant studies are shown in Table 1 and Table 2 respectively.

Table 1: Variable count for 18 gas productivity studies

	Input			Output			Non-discretionary	
	Variable	Count		Variable	Count		Variable	Count
	OpEx (physical)			*Physical*				
1	no. of employees	12	19	no. of customers	7	27	climate	1
2	compressor fuel	3	20	compressor fuel	1	28	regional dummies	1
3	gas purchased	1	21	deliveries/throughput	10	29	customer/population density	3
			22	(peak) capacity	1	30	ownership	1
	OpEx (monetary)					31	time	2
4	total O&M	3		*Monetary*		32	age of network	1
5	labor expense	4	23	sales	4	33	sales concentration	4
6	fuel expense	3				34	policy dummies	1
7	overhead/administration expense	2		*Miscellaneous*		35	peak demand	1
8	gas purchases	1	24	volume x distance	1	36	demand growth	1
			25	cost	2	37	supply area size	1
	Capital (physical)		26	unaccounted for gas	1	38	no. of customers	1
9	pipeline length	11						
10	compressor horsepower	4						
11	tons of steel (diameter, length)	4						
12	precentage of cast iron	1						
	Capital (monetary)							
13	capital (unit) cost	5						
14	capital stock	3						
15	replacement value	1						
	Miscellaneous							
16	total cost	1						
17	peak demand/utilization	2						
18	sales	1						

Table 2: Method count for 18 gas productivity studies (right column includes only firm-level studies)

Method		All	Firm-level
Frontier	DEA	11	8
	bootstrapping	2	1
	SFA	2	2
	COLS	0	0
Index	TFP	3	2
Other	Econometric analysis	8	3

2.1. *Lessons*

2.1.1. Choice of Variables

Besides our own data, past studies can give a first indication about which variables to consider. As we opt for cost as the input variable to be minimised, the categorization of variables as inputs or outputs in Table 1 is not necessarily applicable for our purposes. For instance, while 11 studies account for pipeline length as an *input* variable it is a potential *output* variable for the purpose of this report. What is important is that pipeline length seems *an* important variable when modelling the production technology of gas transmission pipelines. Thus, looking at the input and output columns the top counts are deliveries, pipeline length, number of customers and numbers of employees. Cost of labour is accounted for as an input cost in this report. For number of customers we lack the US data. However, industry knowledge indicates that at least for transmission pipelines number of customers is unlikely to be an important cost driver. We have data on deliveries and pipeline length and include it in our cost driver analysis. One variable that surprises by its low count is capacity. The large fixed cost component and actual tariff setting both indicate that capacity is of particular importance. Although, length of mains and horsepower rating implicitly measure capacity, if available a direct technical measure of capacity would be more desirable.

2.1.2. Choice of Technique

In terms of the techniques (Table 2) the literature on firm-level efficiency puts a clear emphasis on DEA. This might be explained by the fact that we included a large number of non-academic (mostly regulatory) reports. It is our impression that academic papers generally give a slight preference to SFA. However, DEA is certainly employed in almost all current regulatory studies and many academic papers. As will become evident later on this might be largely due to the relative robustness, easy application and interpretation of the techniques. COLS is rarely used since it seems to combine the worst of both worlds (no stochastic element and the need to specify a production function). However, Ofgem (2004) in its latest price control review for electricity distribution makes use of this technique. There are also proposals to re-introduce a stochastic element directly into COLS (see for instance Weyman-Jones et al. (2005)).

Last, an important message from the literature (for instance Granderson (2000) and Sickles and Streitwieser (1991)) is that US (and in particular FERC) data is generally fit for academic firm-level benchmarking studies. However, its use for actual benchmarking is limited as it has not been collected for this purpose.

3. The Techniques

3.1. Frontier-Oriented Benchmarking Methods

There are several different techniques for the measurement of the relative efficiency of firms in relation to an efficient frontier or best practice. Each technique is either based on linear programming or econometrics. This section outlines the main features of Data Envelopment Analysis (DEA), Corrected Ordinary Least Square (COLS), and Stochastic Frontier Analysis (SFA) methods used in this study.[3]

Often it is argued that an average benchmark should be employed instead of a *frontier* benchmark. We consider a frontier approach to be superior for several reasons. Most importantly, if the objective of the regulator is to maximize efficiency then there is no alternative to the frontier approach. Second, as the comparators are real companies there is no *a priori* reason to believe that the frontier cannot be reached by any firm. Last, frontier benchmarking allows the regulator to increase the stringency of the regulation over time by employing safety margins (mainly to guarantee financial viability of the firm) in the early stages of regulation without the need to compromise on his general commitment to frontier benchmarking.

All three methods produce efficiency scores in the range from 0 to 1. Fully efficient firms score 1. However, the absolute level of the scores cannot necessarily be compared across different methods. A crucial stage for all three techniques is the choice of appropriate input and output variables. The choice of variables should, as far as possible, reflect the main aspects of resource-use in the activity concerned. Unlike COLS, DEA and SFA can distinguish explicitly between variables over which management has no discretion and variables that management can influence. Generally the former are referred to as environmental or non-discretionary variables.

Once the variables have been chosen, a decision has to be made on how inputs relate to outputs (i.e. the functional form of the production or cost function for the parametric methods). Besides basic theoretical constraints on the production function (e.g. more inputs

[3] This section is largely based on Pollitt (1995) and DTe (1999).

may only produce more or the same amount of outputs) there are no hard rules. Generally, the choice of variables and their relationships can and should be based on industry knowledge, economic theory or statistical significance.

Table 3 gives a broad overview of benchmarking methods. The methods we employ are highlighted. Next we discuss the three benchmarking methods in more detail.

Table 3: Overview of benchmarking techniques[4]

Category	Type	Technique	Main purpose
Programming techniques	Linear programming	Data Envelopment Analysis (DEA)	Firm-level efficiency
		Parametric Programming Analysis (PPA)	Firm-level efficiency
	Index approach	Partial Factor Productivity (PFP)	Industry-level efficiency
		Total Factor Productivity (TFP)	Industry-level efficiency
Econometric (parametric approach)	Deterministic	Corrected Ordinary Least Squares (COLS)	Firm-level efficiency
	Stochastic	Stochastic Frontier Analysis (SFA)	Firm-level efficiency
Process approaches	Engineering Economic Analysis	Engineering economic analysis	Firm-level efficiency
	Process approach	Process benchmarking	Firm-level efficiency

3.2. Data Envelopment Analysis (DEA)

DEA is a non-stochastic method that uses piecewise linear programming to calculate (rather than estimate) the efficient or best-practice frontier of a sample of firms. The firms that make up the frontier envelop the less efficient firms. DEA can be used to calculate the allocative and technical efficiency of the firms. Technical efficiency can be further decomposed into scale, congestion, and pure technical inefficiency.

DEA models can be specified as input or output oriented and each of these can be further specified as constant returns to scale (CRS) or variable returns to scale (VRS) models.

[4] based on CEPA (2003).

Output-oriented models maximise output for a given amount of input factors, while input-oriented models minimise the input factors required for a given level of output. Input-oriented models are generally more appropriate for gas transmission utilities, as demand for their output is a derived demand which is beyond the control of utilities and therefore given.

Figure 1 illustrates the main features of an input-oriented model with constant returns to scale. The figure shows three firms (G, H, R) that use two inputs (capital K, labour L) for a given output Y. The vertical and horizontal axes represent the capital and labour costs per unit of output respectively and the line PP shows the relative price of the two inputs. Firms G and H produce the given output with less inputs and form the efficient frontier that envelops the less efficient firm R.

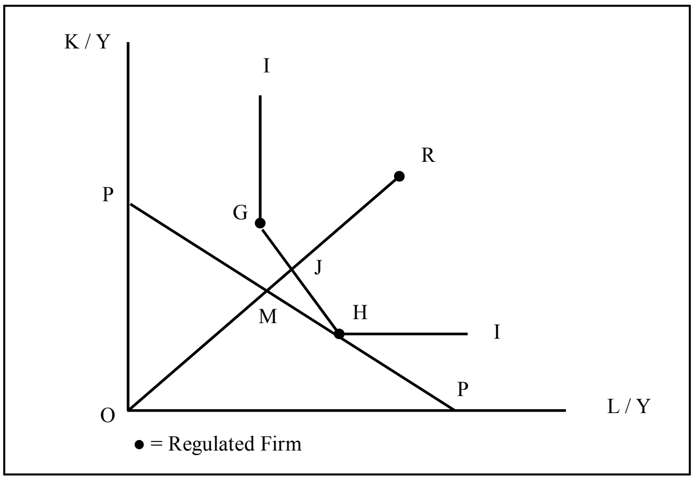

Figure 1: DEA

The technical and allocative efficiency of firm R relative to the frontier are calculated from OJ/OR and OM/OJ ratios respectively. Technical efficiency measures the ability of a firm to minimise inputs to produce a given level of outputs. Allocative efficiency reflects the ability of the firm to optimise the use of inputs given the price of the inputs and the output level. The overall efficiency of firm R is measured from OM/OR.

An advantage of DEA is that inefficient firms are compared to actual firms (or linear combinations of these) rather than to some statistical measure. In addition, DEA does not require the specification of a cost or production function.

A major disadvantage is that the method does not allow for stochastic factors and measurement errors, as it entirely relies on linear optimization. Further, as more variables are included in the model, the number of firms on the efficient frontier increases. Therefore, it is important to examine the sensitivity of the efficiency scores and changes in the rank order of firms to variations in model specification. However, DEA, unlike regression based methods, gives no indication as to whether a certain model specification is to be preferred to any other specification.

3.3. Corrected Ordinary Least Square (COLS)

An alternative to the use of linear optimization is the statistical method, which estimates the best practice frontier and efficiency scores. COLS is one such method based on regression analysis. The regression equation is estimated using the OLS technique and then shifted to the efficient frontier such that the frontier runs parallel to the fitted line and runs through the largest negative estimated error. Note that though commonly done, there is no necessity to fit the frontier through the largest negative error. For this report we fit the frontier through the 90th percentile of errors, which lowers the efficiency shortfall for all firms and increases average performance.

Figure 2 illustrates a COLS model with one cost input C and one output Y. The cost equation $C_{OLS}=\alpha+f_1(Y)$ is estimated using OLS regression and then shifted by CA to $C_{COLS}=(\alpha-CA)+f_1(Y)$ on which the most efficient firm A lies. The example gives the most stringent frontier supported by the actual data. In practice, the OLS regression line could be shifted by any portion of CA, γCA where γ lies between 0 and 1 and a γ of 0 implies a frontier equal to the OLS regression line. The efficiency score for an inefficient firm such as B is then calculated as EF/BF.

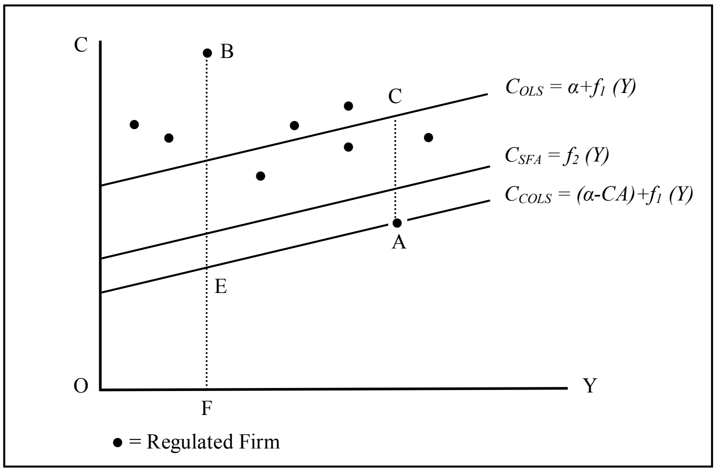

Figure 2: COLS and SFA

The COLS method requires specification of a cost or production function and therefore involves assumptions about technological properties of the firms' production process. A drawback of the method is that the estimated parameters may not make economic or engineering sense (this also applies to SFA discussed below). In addition, the method makes no allowance for stochastic errors and relies heavily on the position of the single most efficient firm or any other arbitrary position of the frontier. Similar to DEA, COLS assumes that all deviations from the frontier are due to inefficiency.

3.4. Stochastic Frontier Analysis (SFA)

SFA is another stochastic method used to estimate the efficient frontier and efficiency scores. The statistical nature of the method allows for inclusion of stochastic errors in the analysis and testing of hypotheses. Similar to COLS, this method requires specification of a cost or production function involving assumptions about the firms' production technologies. Estimation of efficiency scores in SFA is similar to that of COLS. In addition, SFA recognises the possibility of stochastic errors. This reduces reliance on measurements of a single efficient firm.

However, accounting for stochastic errors requires specification of a probability function for the distribution of the errors and the distribution of inefficiencies (e.g. half-normal). As for the result of stochastic factors and their effect on the position of the most efficient firm, the estimated efficiency scores are usually higher than those estimated by COLS. However, the ranking of the scores is identical for SFA and COLS. Figure 2 illustrates (approximately) the estimated cost equation $C_{SFA} = f_2(Y)$ using SFA for the same sample of firms. Another drawback of the method is that even if there are no errors in efficiency measurements, some inefficiency may be wrongly regarded as noise. Moreover the technique might for data or model reasons fail to disentangle the measurement error component and the inefficiency component and produce no useful results.

Thus, though theoretically the most appealing technique SFA is the hardest to apply. A summary of the various characteristics of the techniques is given in Table 4 below.

Table 4: Summary characteristics of DEA, SFA and COLS

Method / Feature	DEA	SFA	COLS
strong assumptions (in particular on production function)	No	Yes	Yes
accounts for measurement error and misspecification (gives confidence intervals)	No	Yes	No
sensitivity to outliers	High	Low	High
sensitivity to sample size	Low	High	High
ease of verifiability/ transparency	High	Low	Low
identification of real comparators	Yes	No	No
extension to multiple outputs and inputs	Easy	Difficult	Difficult
use of environmental factors	Yes	Yes	Not explicitly

4. Our Approach

We have been commissioned to explore ways to conduct an efficiency benchmark for several European gas transmission operators. Any approach that addresses real world policy problems should be defined with a view to the obstacles it faces and the resulting trade-offs. These constraints can be of various kinds. We limit our analysis to five constraints of direct importance to the regulation of European gas transmission companies. In what follows we delineate our approach by tackling one obstacle at a time, moving from the more general to the more specific.

4.1. The Objective of Economic Regulation

The primary aim of economic regulation is the increase of consumer welfare. Moreover, regulation should be restricted to the part of the business that cannot be exposed to competition (yet). For this reason we restrict this benchmark to the transmission function.[5] Thus, a comparable ring-fencing of the transmission function across countries is important.

The objective of consumer welfare requires an economic benchmark and not for instance engineering benchmarks as ultimately consumers benefit from reductions in price and not from an increase in technical efficiency or reduction in cost per se. Only price (i.e. tariff) reductions increase consumer welfare. Assuming equal profit margins, our cost measures reflect tariff efficiency. Additionally, when using revenue as dependent variable tariff efficiency is also given for different profit margins. As revenue is given by price multiplied by output quantity and we control for output our revenue based efficiency scores indirectly make statements about tariff efficiency.

[5] In line with the terms of reference of this study we restrict ourselves to the transmission function only. This excludes gathering, storage and distribution.

4.2. Asymmetric Information Between Regulator and Firm

A large part of the recent literature on economics and regulation focuses on asymmetric information. One important result of this literature is that incentive regulation is a viable strategy to reduce asymmetric information between the firm and the regulator. This underpins the increasingly popular forms of price cap or incentive regulation implemented across Europe. However, these theoretical advances have to be balanced against theoretical and practicable problems with the benchmarking techniques. The techniques are often considered opaque and in-consistent. Our robust and practicable benchmarking approach intends to lessen the burden of this trade-off for gas transmission.

4.3. Number of European Observations

The most substantial practical problem that we face with this study is the insufficient number[6] of comparable European observations to produce robust efficiency scores for the participating gas transmission utilities. However, as Table 5 suggests, Europe as a whole (and in particular over several years) has a sufficient number of transmission operators for the purpose of benchmarking. This report suggests including non-European data to facilitate the benchmarking exercise at this time. In particular, the existence of a large and readily available set of rather well-documented data for US transmission companies is helpful. This report shows how much there is to be gained from the inclusion of US data and what issues their inclusion raises.

[6] Generally, larger data sets are preferable as they increase confidence in the results.

Table 5: Number of gas transmission and distribution companies[7]

Country	Transmission		Distribution
	National	Regional	
Austria	3	4	20
Belgium	1	0	19
Denmark	1	0	4
France	3	0	22
Germany	15	20	730
Ireland	1	0	1
Italy	2	0	557
Luxembourg	1	0	4
Netherlands	1	0	27
Spain	3	6	25
Sweden	1	2	7
UK	1	0	1
Estonia	0	0	11
Latvia	1	0	1
Lithuania	1	0	5
Poland	1	6	68
Czech	1	0	124
Slovakia	1	0	1
Hungary	1	0	11
Slovenia	1	0	17
Sum	**40**	**38**	**1655**

Moreover, it should be possible to rely less on non-European comparators once a sufficient amount of European data is available.

Though the lack of European comparators is the prime reason for undertaking an international benchmark in this report, in the future it might be possible to establish a US/Europe best-practice frontier. This would assist European regulators when setting tariffs using incentive regulation. Today, however, confidence in the comparability of the data is not sufficient to establish a common frontier.[8]

4.4. Comparability

The most obvious practical issue is comparability. Naturally, many concerns are raised around comparability especially because European companies are suspected to be substantially different from US companies. However, our analytical results below show that the differences

[7] European Commission (2005).

[8] For a general discussion on the advantages and disadvantages of international benchmarking see also Jamasb and Pollitt (2002).

within Europe are as large as those between European and US companies. Nevertheless, comparability has to be established regardless of the sources of any differences if the results are to be credible.

A necessary first step is standardizing the data. This is likely to have the greatest positive effect on the relevance of the results. Secondly, there are methodological solutions to overcome any remaining issues with comparability. A lack of comparability is similar in nature to measurement error. Stochastic techniques like SFA are designed to address this and give credibility to results even when measurement errors cannot be avoided entirely.

Last, it should be stressed that achieving comparability amongst the European firms is easier then achieving comparability between US and European firms (as this most likely requires direct communication with FERC). Nevertheless, some relevant results can still be obtained using the US data. If inter-European comparability was sufficient it would be possible to rank the efficiency scores of European firms.[9] However, at this stage, it is not possible to produce reliable measures of the absolute distance to the frontier as this requires comparability among all (i.e. US and European) firms. Thus, the requirements for ranking a *limited number* of European firms are: standardization of European data and inclusion of some US data. Again, without standardization of *all* data it is difficult to measure the distance from the frontier and indicate what the true efficiency shortfalls are. It is in this light that our results are presented. The design of the techniques raised above brings us to the last constraint.

4.5. Issues with the Benchmarking Techniques

We suggest two strategies also used elsewhere, to ensure that we arrive at relevant and consistent efficiency scores.

First we undertake a cost-driver analysis.[10] This analysis is of importance for the various benchmarking techniques. A cost-driver analysis establishes the statistical significance of the available independent variables. Though statistical significance does not have to be the ultimate arbitrator for the inclusion of a variable it gives important guidance especially for DEA, which cannot discriminate between relevant and irrelevant variables itself.

[9] As we illustrate in Section 5.7. the European we use here is actually not sufficiently standardized.
[10] See IPART (1999) for another example of a cost driver analysis in the context of benchmarking.

Additionally, the choice of functional form is of no importance to DEA but might have an effect on the efficiency scores that SFA and COLS produce.

This preparation step before the actual benchmarking is intended to help in selection of model variables. The use of a variety of techniques and subsequent consistency tests helps to make the results more robust.[11] Consistency tests might increase confidence in the results produced by specific models and techniques. However, when following this strategy one has to be aware of the limits of comparison across techniques. Absolute efficiency scores *cannot* be compared across techniques as they differ by construction.

[11] See Bauer (1998).

5. Data

This chapter introduces the variables used, names the data sources, gives descriptive statistics, and discusses comparability. For confidentiality reasons all company data is made anonymous here. This is also the reason why we report only selected descriptive statistics for the underlying data.

5.1. Variables

For the purpose of this study we gathered data on several input and output variables, or costs and cost drivers.

Given our choice of a cost benchmark the dependent variable in all our models is one of four cost measures. All our cost measures are cumulative (i.e. successive cost components are added). For this reason we provide no capital expense (capex) measure here. The cost measures reflect the move from variable (i.e. short-run) costs to fixed (i.e. long-run) costs. The cost measures are listed in Table 6. The second column lists the components that make up each cost measure.

Table 6: Cost measures and their composition

Name	Composition	Remarks
O&M	O&M	includes labour, excludes fuel, taxes and rents
Totex1	O&M, depreciation	
Totex2	O&M, depreciation, cost of capital	Cost of capital equals the written down value (historic asset base less accumulated depreciation) multiplied by a cost of capital percentage (7%)
Revenue	Revenue (less fuel)	Revenue is not built up from components but given as reported.

Often cost measures are only categorized as operating expense (opex) or total expense (totex). Using this distinction our first measure O&M qualifies as the former and the last three measures Totex1, Totex2 and Revenue as the latter.[12]

The one measure that is not common is revenue. Revenue as such is not a cost measure. However it is an important regulatory variable and when profit is considered a cost to the rate payer revenue is the most inclusive proxy for costs. Revenue has the additional appeal that it does not require the determination of a rate for cost of capital. Moreover, as all our models control for output, revenue efficiency allows for general conclusions on tariff levels. As we do not control for the age of networks and other potentially relevant environmental variables, such a tariff comparison across countries can be affected by these omitted variables.

For Totex2 we multiplied the written down value by a rate of cost of capital of 7 percent. This rate does not vary by company or country. Assuming a world market for capital this is justified. Additionally, we check for the sensitivity of the SFA results by using 5 percent and 9 percent instead.

Next, we turn to the independent variables or cost drivers, which are listed in Table 7. The seven variables were selected based on the literature as discussed above and the constraints of data availability with the objective to capture the nature of the production process.

Table 7: List of cost drivers

Name	Composition (unit)	Remarks
Delivery	Total yearly throughput of gas transmitted (m^3/year)	For US, only gas owned by others
Mains	Total length of pipelines (km)	
Horsepower	Total amount of compressor horsepower on pipelines (HP)	
Stations	Total number of compressor stations (#)	
Units	Total number of compressor units (#)	
Capacity	The maximum of all past and present measures of daily peak delivery times the number of days per year (m^3/year)	This measure is not precise as technical capacity is likely to be greater then peak delivery
Load factor	Delivery over capacity (%)	(see remark on capacity)

12 As the names of many of our variables coincide with generic terms we adopt the convention to spell the variables with capital letters.

Capacity and Load Factor are derived from the original data. Generally, we expect all available variables to cover the two most important output dimensions: capacity and throughput. Whereas Delivery represents throughput all other variables represent capacity, Load Factor covers both dimensions at once.

Technical capacity is influenced by various variables like for instance compressor horsepower, entry and exit pressure and pipe diameter. Unfortunately, we do not have data on all of these but expect that our variables are related to the omitted variables and therefore capture the nature of the production process.

A more substantial omission is input price. Our efficiency scores would be distorted if different transmission operators faced different input prices. Besides the fact that input prices are not readily available there are two things that might justify our assumption that input prices do not vary across firms. First, we use Purchasing Power Parities (PPPs), which by construction equalize average domestic price levels.[13] Secondly, especially the totex measures are largely made up of globally traded goods like pipeline steel and therefore equal input prices. Besides a global market, this also requires that assets were bought at the same time. As we do not account for age of the assets, inflation differentials are not accounted for.[14] Two items where this discussion might not entirely be resolved by our use of PPPs are labour and compressor fuel. We decided to include the first but drop the latter because we assume that fuel is a rather constant percentage of deliveries whereas labour expense differentiates firms and constitutes a substantial amount of the O&M expense.

Last, we omit variables relating to quality or the environment both of which are beyond the scope of this report. Moreover, at least for quality we have received no indication that it would impact on efficiency scores.

5.2. Data Sources

All US data is taken from annual regulatory reports that energy utilities have to file with the federal regulator (FERC).[15] All European data is provided by the respective companies in

[13] For a more detailed discussion on labour expense and PPPs see section 5.7. below.
[14] We account only for inflation since the date of reporting not the date of purchase of an asset.
[15] More information and the data are available at the FERC website at:

collaborations with one exception where the data was estimated by the regulator. The data for the European companies was not audited.

For the US data we removed observations that showed missing, zero or negative numbers for our variables. Additionally we removed companies, where the larger part of transmission pipes is offshore. Last we performed several consistency tests to minimize the risk of measurement error.[16] However, as the US data was not gathered for the purpose of this benchmark, it is likely that errors remain and that the comparability with European data is not perfect.

5.3. Number of Observations

Table 8 list the number of utilities that participate in this benchmark. Our dataset constitutes an unbalanced panel. Due to the lack of time series for the European firms it is not possible to construct a panel data set. However, regression-based techniques (i.e. SFA and COLS) produce more robust results when used on panel data. We treat our data as a pooled cross-section.

Table 8: Number of observations

Country	Years	No. of companies	No. of obs.
A	2002-2004	1	3
B	2004-2005	1	2
C	2000-2004	1	5
D	2004	1	1
US	1996-2004	43	317
TOTAL (before any outlier removal)		**47**	**328**

http://www.ferc.gov/docs-filing/eforms/form-2/data.asp.

[16] For instance, if O&M plus depreciation is greater then revenue we dropped the observation.

5.4. Outliers

We only consider outliers in terms of efficiency scores, not in terms of the original data. However, outlying efficiency scores are also caused by the original data.

In order to illustrate the alternative ways to outlying observations we take a different approach for each technique. Note that the term "removal" is misleading here as we do not actually remove observations from the sample for all techniques.

There are numerous ways to account for outliers or to define them in the first place. We found that different techniques lend themselves to different approaches.

In SFA we do not remove any outliers as SFA explicitly accounts for measurement error and outliers have a relatively small effect on the results (this is also a reason why SFA scores might act as a benchmark for other techniques). For COLS we opted for a "lowering" of the frontier by taking the 90th percentile as the frontier instead of the single most outlying observation (the 100th percentile). In order to preserve an upper bound of 1 for the efficiency scores we consider all observations beyond the 90th percentile to be on the frontier. Thus, unlike the standard distribution of COLS scores we have 10 percent of observations scoring 1. This approach has the additional advantage that we keep the information contained in these observations. As DEA neither allows for measurement error nor an arbitrary definition of the frontier we removed the upper 10 percent of observations from the initial run and then did a second run on the reduced number of observations. Note that removing outliers might make certain consistency tests less relevant. For instance, outlier removal effects mean scores and their comparison no longer allows conclusions on consistency.

5.5. Data Transformations

For reasons of comparability and ease of use of the benchmarking techniques the original variables are transformed. By inflation and exchange rate adjustments we make observations comparable across time and countries. Mean correction and taking logs allows for greater ease when applying econometric methods. These latter transformations are, however, not applied for DEA.

5.5.1. Inflation Adjustment

All cost measures (including historic asset values) used are inflation adjusted. The base year is 2004. For asset values we do not account for inflation from the date of purchase, which would amount to current asset valuations, but from the date of reporting. For instance, a firm might report a certain historic asset value in 2002, which was purchased in 1980. The inflation adjustment only covers the period from 2002 to 2004 and not from 1980 to 2004. For the European data we used harmonised annual average consumer price indices published by Eurostat.[17] For the US we used consumer price indices published by the Bureau of Labor Statistics.[18]

5.5.2. Exchange Rate Adjustment

The data on purchasing power parities (PPP) is taken from Eurostat. These international prices are “an artificial common reference currency unit obtained by scaling PPPs so that the aggregate for the EU-25 as a whole is the same whether expressed in euros (ECUs) or in PPS” (Eurostat website). Therefore, in the remainder of the report all expenses are measured in 2004 Purchasing Power Standards (PPS) if not stated otherwise. Though PPS are an artificial currency 1 PPS is roughly of the value of € 1.

5.5.3. Means and Logs

All variables except load factor are further transformed for practical purposes. The objective of these transformations is to make the data fit for the econometric techniques we use. Two steps are taken to lower the variance of the variables and give them a near normal distribution. First variables are mean corrected. That is all observations are divided by their respective sample mean. Secondly, logarithms are taken. After these transformations most variables lie in a range between -4 and +4. This is reflected on the scales of various graphs below.

---

[17] The indices are online available at:
http://epp.eurostat.ec.europa.eu/portal/page?_pageid=1996,39140985&_dad=portal&_schema=PORTAL&screen=detailref&language=en&product=EU_MAIN_TREE&root=EU_MAIN_TREE/basic/YEARLIES_NEW/B/B2/B21/dba10000.

[18] The indices are online available at:
ftp://ftp.bls.gov/pub/special.requests/cpi/cpiai.txt.

5.6. Descriptive Statistics

This section describes the (transformed) variables. In particular it highlights the differences across countries, that is, differences between the group of US companies and European companies.

Table 9 gives the average ratios of the various cost measures to Revenue. These help to put our various cost measures into perspective.

Table 9: Mean percentage of cost measure to revenue by country

Country	O&M	Totex1	Totex2
A	.54	.58	.60
B	.10	.55	.99
C	.44	.64	.96
D	.16	.27	.44
US	.20	.36	.57

By construction the cost measures increase as a percentage of revenue from O&M to Totex2. The difference between Totex2 and Totex1 might be taken as an indication of profits. However, the pronounced differences across countries demand caution when interpreting these figures.

It is interesting to note that for most variables the variance across the European firms is as large as the variance for the US firms. Obviously this might reflect either problems with the comparability of the underlying data or genuine differences across European companies. Certain peculiarities can be explained by our knowledge of the specific circumstances of individual countries. The rather high O&M figures (well above the US mean) for the European countries might relate to higher wages.

Load Factor is interesting because all European countries have almost the same load factor, which is significantly below the US average.

As we make heavy use of box plots (as in Figure 4) in what follows they should be explained shortly. The upper and lower limits of the shaded boxes represent the 25th and 75th percentile of all observations. For instance when looking at the US capacity unit costs we see that in the US 75 percent of all observations have unit costs of about 0 to 0.0015 PPS. The median or the 50th percentile is represented by the line in the middle of this box. The “whiskers” reaching beyond the box have a length of at most 1.5 times the interquartile range (the difference between the 25th percentile and the 75th percentile in the units of the data) and must end at an observed value. Dots represent individual observations that lie beyond. The skew of the box and the whiskers represents the skew of the distribution. Obviously, when a country like D has one observation only, the entire box collapses to one thin line.

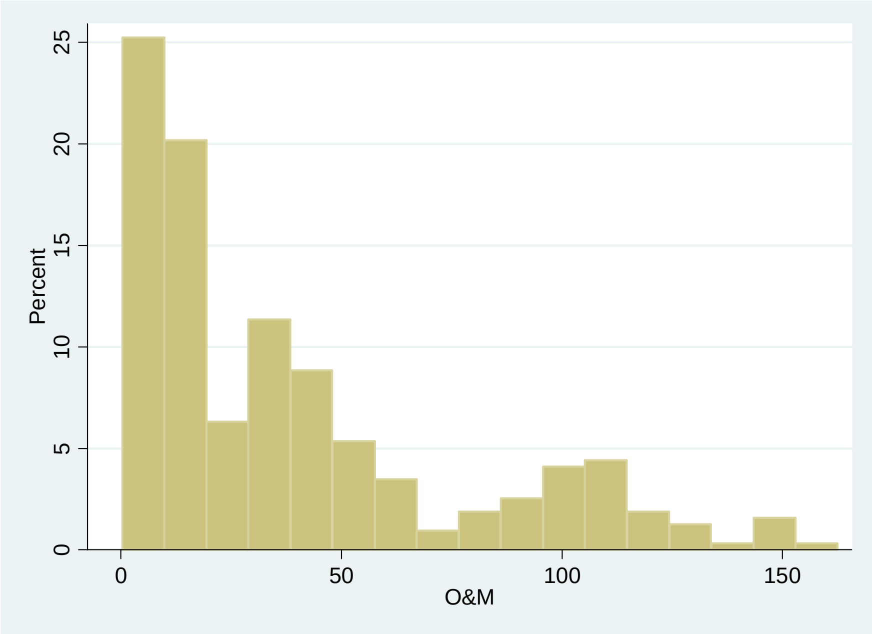

Figure 3: Histogram of O&M expense for US firms (in million PPS)

In other words, box plots represent the same information as is contained in histograms such as in Figure 3 for O&M expense for US firms. Whereas histograms look at the distribution from the “side” the box plot looks at it from “above”. Last, note that all box plots give the median and not the mean as a measure of central tendency. However, we refer to the mean throughout the report for reasons of consistency. This seems justified as generally the two hardly differ.

Having given the interpretation of box plots we continue looking at unit costs measures. Unit cost ratios are often considered “naïve” efficiency scores. Though they do not substitute for proper efficiency scores they can help explain the ranking of individual firms. High unit costs might imply low efficiency scores.

Figure 4 gives unit costs for capacity. Identical figures for other cost drivers are given in Figure 5, Figure 6, and Figure 7 in the chapter’s appendix. Naturally, these ratios reflect the underlying data. Whereas country D consistently performs better then US average most other European countries fare worse on most measures.

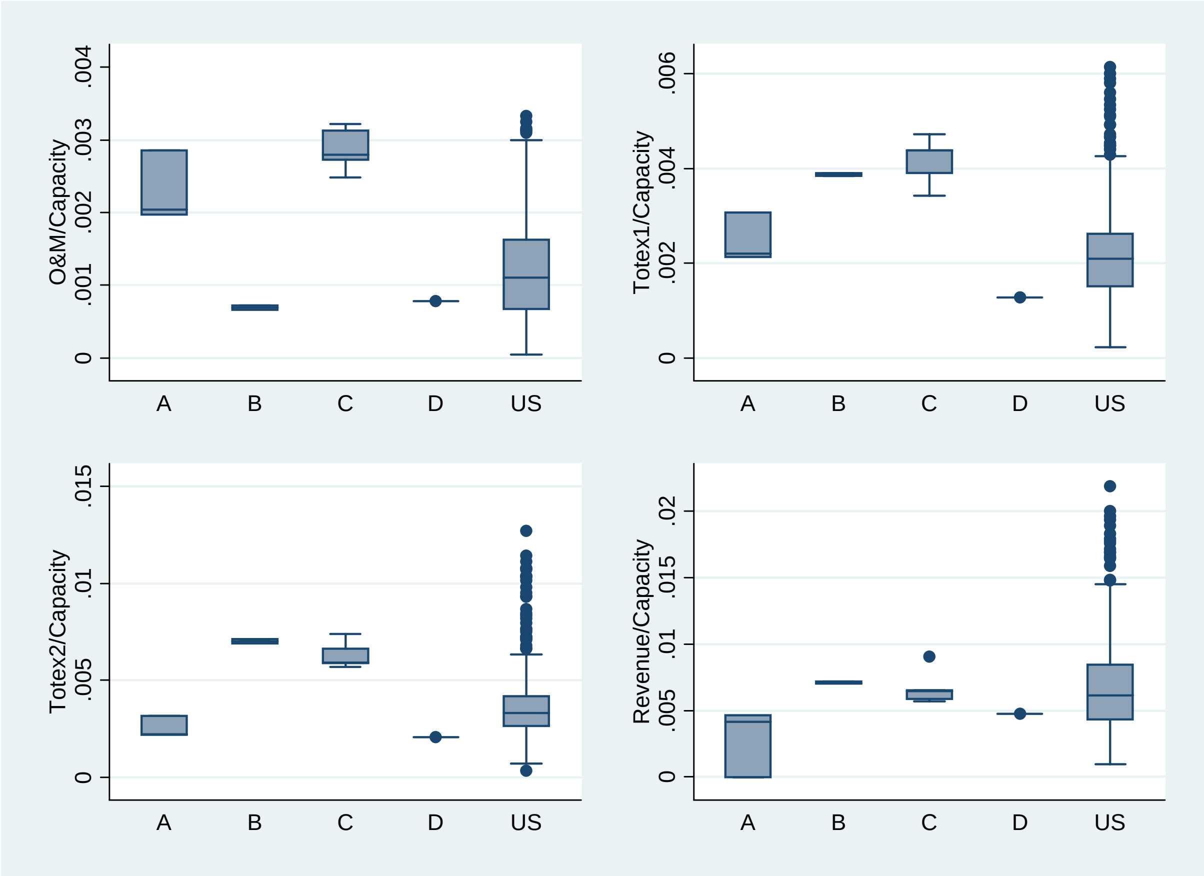

Figure 4: Unit cost for capacity (PPS/m^3)

Given these unit cost measures one would expect the European firms to perform below the US average on the efficiency scores, with the country D being the best performing European country.

5.7. Comparability

We addressed comparability in general terms above. Here we would like to raise detailed issues with the data at hand.

We are aware that there are differences in dealing with transit and export volumes and related costs across European countries. For instance, whereas country D includes all transported volumes countries A and B do not. For A the problem might be mitigated by the separation of domestic and export pipelines. We summarize our current knowledge about the differential treatments of transit and export in Table 10.

Table 10: Comparability of transit/export flows across countries

	US	A	B	C	D
separate transit/export network?	no	yes	no	no	no
throughput (% included)	excludes sales own sales (~80)	only inland (?)	only inland (?)	only inland	all
cost (% included)	all (100)	only inland	all (100)	only inland	all (100)
cost allocation	?	?	?	?	?

Another issue is labour cost. We do not explicitly correct for differences in wage levels across countries. However, we excluded all non-wage labour costs (i.e. pension benefits, taxes paid by the employer). Secondly, PPP conversion indirectly and partially corrects wage levels.

Last, the comparability of asset values and cost of capital is not perfect. First, as US data is measured at historic cost we had no other choice than to use historic cost for this study. We do not standardize depreciation rates or account for the age of assets. For asset age we would expect that the older the assets the lower the capex costs and therefore the lower revenue. However, the higher capital costs of younger pipelines might be balanced by lower operating costs due to better technology. From the tariff benchmark currently conducted by the CEER, we know that for the European companies the network age ranges from 20 to 27 years. We expect most US pipelines to be older.

For the cost of capital as used in our Totex2 measure we applied a cost of capital rate of 7 percent for all companies. Though we are aware that different regulators apply different WACC rates, increasingly integrated capital markets align *actual* cost of capital rates.

We suspect that there are other issues of comparability (e.g. environmental factors) both for the comparison amongst European firms and the comparison with US firms. These have to be addressed before any relevant efficiency scores can be produced.

5.8. Appendix

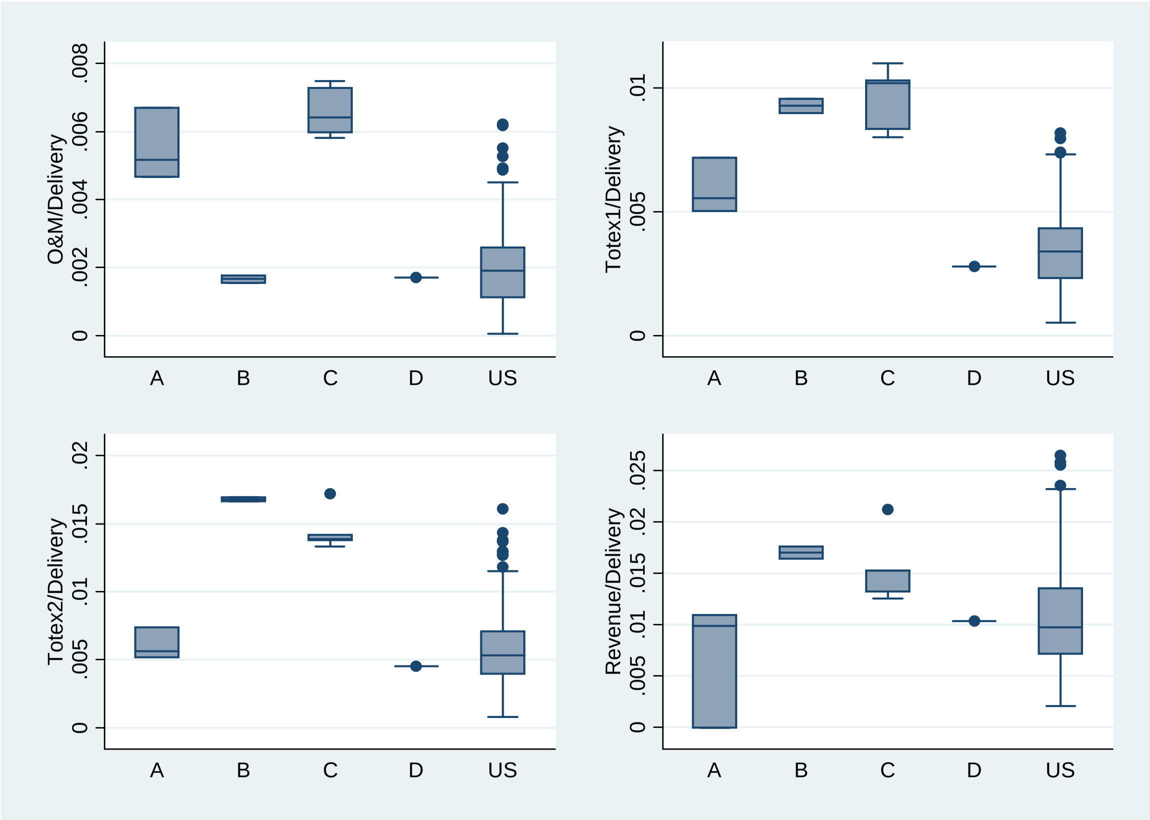

Figure 5: Unit cost for delivery (PPS/m^3)

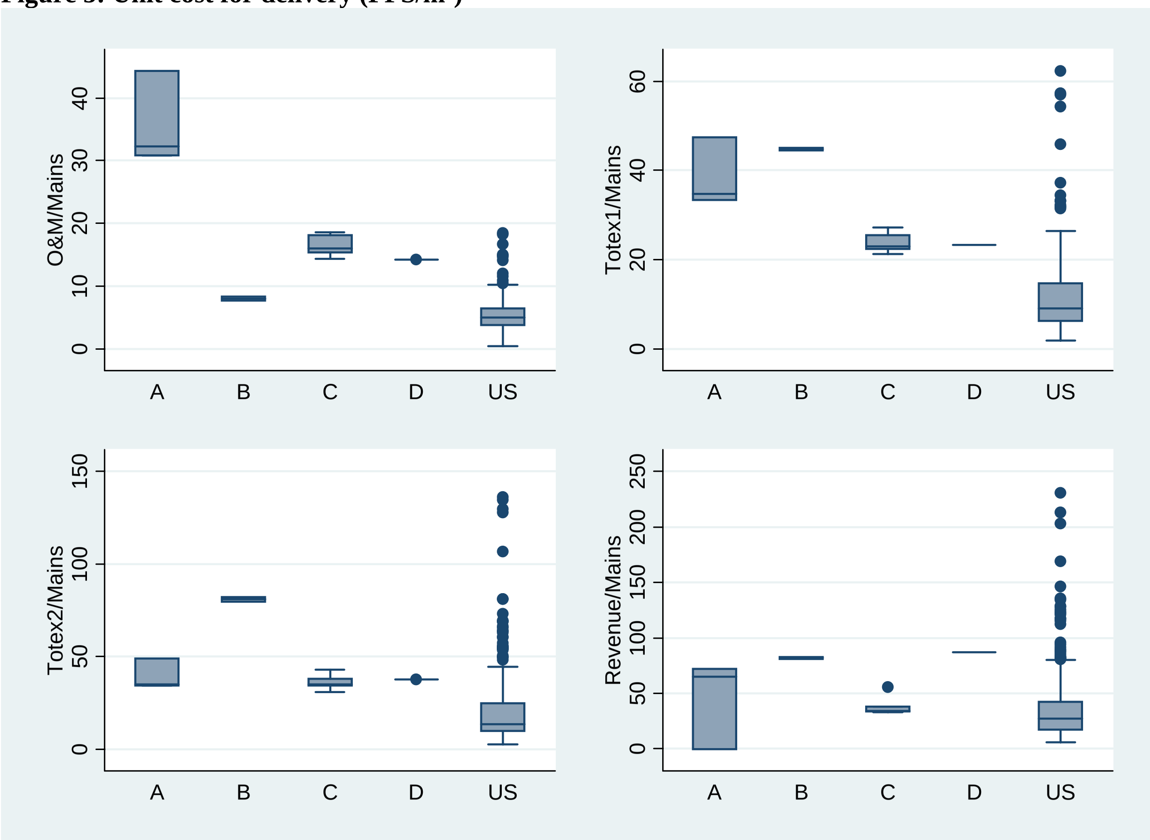

Figure 6: Unit cost for mains (Thd. PPS/km)

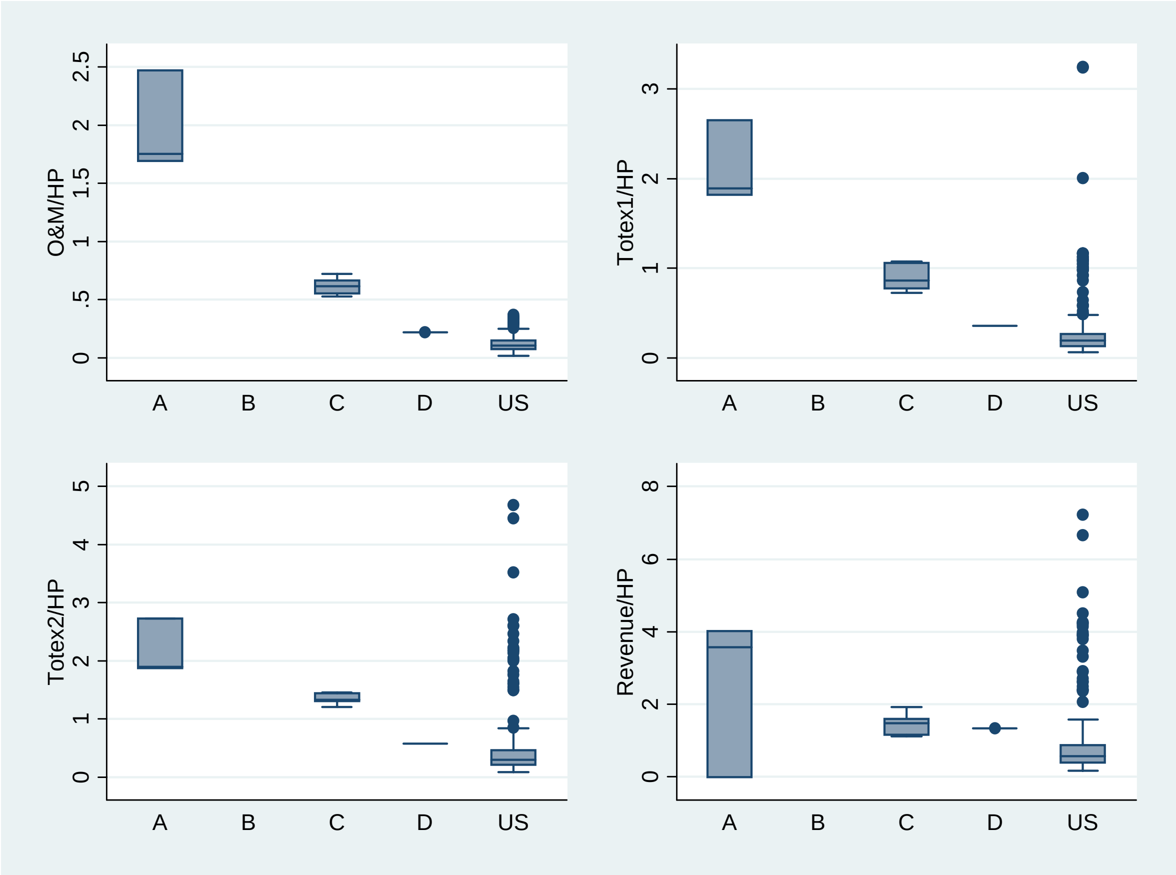

Figure 7: Unit cost for horsepower (Thd. PPS/HP)

6. US Regulation

6.1. Market Players, Structure, and Market Interaction

Today the US has 24 major and 8000 independent producers of gas. Additionally, there are numerous producers in Canada that export to the US. The major producing regions of the Gulf, the Mid-West and Canada are connected to consumers by 190,000 miles of transmission pipelines operated by some 110 companies (of which about half are considered “major”). These companies are privately owned and mostly vertically integrated.

Besides producers and transporters, the value chain for natural gas includes marketers and brokers as intermediaries. Intermediaries are numerous as 2/3 of all gas sales are spot sales.

If not stated otherwise the following refers to interstate regulation, markets etc. Generally, the US distinction between State and Federal level is rather similar to the distinction between the Union and its Member States in Europe.

6.2. Regulation

US gas markets have been regulated since the beginning of the last century. A major reason for the deregulation of interstate transmission was the failed regulation of gas production. When the regulation of production failed and was reversed this had unexpected repercussions for transporters that eventually led to major changes in regulation for these as well. The centre-piece of this re-regulation is Order 636 of 1992, which is also referred to as the “final restructuring”. It completely changed the industry and gave it its current structure. A possible sign of the success of this change in regulation is a drop in transmission cost from $2.20 to $1.40/Mcf between 1987 and 1996 (Gabriel and Minehart (2004)).

The re-regulation of transmission in the decade between 1985 and 1995 introduced three major changes: unbundling, the establishment of lightly regulated secondary capacity markets and a stronger reflection of the ratio between fixed and variable costs in tariffs.

6.2.1. Unbundling

After having experimented with third-party access ("open-access" in US jargon) for several years concerns over discrimination led to unbundling but short of ownership unbundling of transmission. The question remains whether marketing affiliates can abuse the vertical relationship with pipeline operators. There have recently been alleged cases of abuse through information sharing, which prompted FERC to tighten the rules.[19] However, the problem of detection remains.

6.2.2. Capacity Release

"Capacity release" is the catch phrase for the establishment of a secondary capacity market. Shippers have an obligation to make available firm capacity that they have bought but do not use. In practice unused capacity is returned to the pipeline operator, which auctions the capacity and credits the original price to the shipper that bought it first. The crux of this arrangement is that in the secondary market the seller is *not* constrained by the regulated tariffs if he has no market power. Even with market power, the tariff is substituted for by a (higher) cap. This is the crucial channel through which relative scarcity is reflected in prices and incentives for investment are given.

6.2.3. Tariffs

The two measures above went hand in hand with an adjustment of the tariff structure to better reflect the balance between fixed and variable cost. In practice regulated tariffs are regulated unit rates. Thus, they require information on total cost, expected demand and a procedure for a break-down of costs. This procedure includes five steps. First total recoverable costs are determined. These consist of the rate-of-return, O&M expenses, depreciation, income tax allowance and other operating expenses. Second, total cost is allocated to functions (i.e.: transmission and distribution). Third, functional costs are divided into fixed and variable costs as rates are two-part. Fourth, costs are allocated to pipeline services and fifth, unit rates are determined. Generally, the flexible tariff part varies with the time of year and the distance travelled. An illustration is the design of unit rates for interruptible and firm services. Whereas the first are one-part and include only a small fraction of fixed costs the latter are

[19] A prominent case is PUC of the State of California v. El Paso Natural Gas Company (EPNG) and El Paso Merchant Energy (EPME).

two-part and carry the larger amount of fixed costs. Therefore, rates are designed to facilitate efficient use of the pipeline capacity.

For the interpretation of our benchmark results it is important to see what incentives FERC's tariff setting gives for cost reduction. Incentives vary depending on time. Before a rate hearing, companies have an incentive to increase cost as this will lead to a higher absolute return. However, once the tariff is set companies have an incentive to reduce costs as they are the claimant to the difference between tariffs and actual costs. Obviously this distinction is artificial as in reality there is no clear distinction between these two time periods. But it helps to illustrate the two opposing incentives that firms face. A reason why the cost reduction incentives might be dominant is the fact that today FERC regulated transmission operators have no regular rate cases any more. Rate cases only occur when inflation is strong or the company can prove other substantial cost increases.

6.3. Uniform System of Accounts

At the heart of US regulation is a vast and elaborate (the description is about 150 pages long!) set of regulatory accounts named the Uniform System of Accounts. Each regulated utility has to report on a yearly basis. Besides standard financial data much detailed information on the financial and physical specifics has to be provided. The system also provides for a break-down by utility function.

6.4. Comparison with Europe

Regulation and therefore the competitive interaction in the gas markets seems to differ greatly between the US and Europe. One major reason for this might be the great number of supply sources that are available for the US market. In Europe the increasing concentration of supply sources certainly does not help the development of pipe-to-pipe competition as in the US. Our intuition that unbundling and third-party access are stronger in the US seems to be supported by the differences in load factors. Especially, strong unbundling removes any incentives to use the pipeline as a means to reduce competition in downstream markets. Load factors might reflect the strategic use of pipelines.

Last it might be useful to form a hypothesis on whether European or US transmission companies are more efficient. This proves to be rather difficult. One might argue that the US practice of cost-plus regulation leads to "gold-plating" and therefore one would expect European operators to be more efficient. However, the mix of a long history of regulation and recent moves towards competition not only on a particular networks but also pipe-to-pipe might counteract the tendency towards gold-plating. Last, the system of irregular rate-cases as opposed to fixed regulatory periods might increase the incentive for cost reduction. Most academic studies on US gas transportation assume that transporters are cost-minimizing.

7. Cost-driver Analysis

This section analyses the relationship between the main measures of costs and cost drivers in order to establish which variables to include in the efficiency benchmarking. Correlation tables can be used to analyse how strongly related different variables are on a one-to-one basis.

For the cost-driver analysis and the benchmarking we pool the data across years. Certainly, treating the data as a panel would be preferable if several years of observations for all European countries were available.

An alternative to a cost function would be an input distance function.[20] An input distance function would be warranted if there are several inputs and outputs to be included and cost minimization cannot be assumed.

Table 11 provides correlations amongst our four cost measures. All cost measures are highly correlated with one another. Not surprisingly the lowest correlation is between O&M and Revenue. The high correlations might indicate that efficiency scores do not change much with the choice of the cost measure. Also from the perspective of the regulator these cost measures can be considered substitutes for the purpose of cost benchmarking. Thus, preference might be given to cost measures where data is easier to gather and/or standardize.

Table 11: Correlations among cost measures

	O&M	Totex1	Totex2	Revenue
O&M	1.00			
Totex1	0.90	1.00		
Totex2	0.84	0.98	1.00	
Revenue	0.83	0.95	0.96	1.00

Table 12 gives the correlations between the cost measures and various cost components. Again all correlations are high. The compressor energy expense that we exclude has a relatively low correlation with the cost components. Note that all cost measures include all compressor related costs except fuel cost. Looking at the rows, correlations increase (decrease) with an increased (decreased) weight in the cost measure.

[20] A discussion on input distance functions for benchmarking is contained in Coelli et al. (2005).

Table 12: Correlations between cost measures and cost components

	O&M	Totex1	Totex2	Revenue
Depreciation	0.68	0.90	0.93	0.89
Acc. Dep.	0.84	0.89	0.86	0.87
Historic Assets	0.82	0.95	0.97	0.95
Comp. Energy	0.70	0.70	0.70	0.63
Labour	0.88	0.83	0.80	0.82

Table 13 gives the correlations among outputs or cost drivers. All correlations are high except for load factor. Not surprisingly, the size of a network is not correlated with its level of capacity utilization. The high correlations among the independent variables are likely to cause problems for the regression analysis below because the methods cannot disentangle the individual effects.

Table 13: Correlations among outputs

	Delivery	Mains	HP	Stations	Units	LF	Capacity
Delivery	1.00						
Mains	0.76	1.00					
HP	0.87	0.85	1.00				
Stations	0.64	0.88	0.83	1.00			
Units	0.64	0.89	0.81	0.92	1.00		
Load factor	0.38	0.11	0.27	0.04	-0.00	1.00	
Capacity	0.94	0.78	0.84	0.67	0.69	0.04	1.00

Table 14 gives the correlations between inputs and outputs. The high correlations are a first indication that all our output variables (except load factor) *individually* relate to costs.

Table 14: One-way correlations between costs and outputs

	O&M	Totex1	Totex2	Revenue
Capacity	0.84	0.88	0.86	0.86
Delivery	0.79	0.88	0.88	0.87
Mains	0.90	0.83	0.78	0.79
Horsepower	0.85	0.83	0.81	0.83
Stations	0.82	0.68	0.62	0.66
Units	0.88	0.69	0.62	0.64
Load factor	0.05	0.20	0.25	0.23

The first column in Figure 8 visually represents the second column of Table 14. At the same time, Figure 8 gives a sense of the pervasiveness of outliers in the original data. Where the scatter plots show a strong positive correlation extreme outliers would be lying to the North-

West or South-East. Such extreme outliers are hardly present. Figure 9, Figure 10, and Figure 11 in the chapter's appendix give the same graphs for the other cost measures.

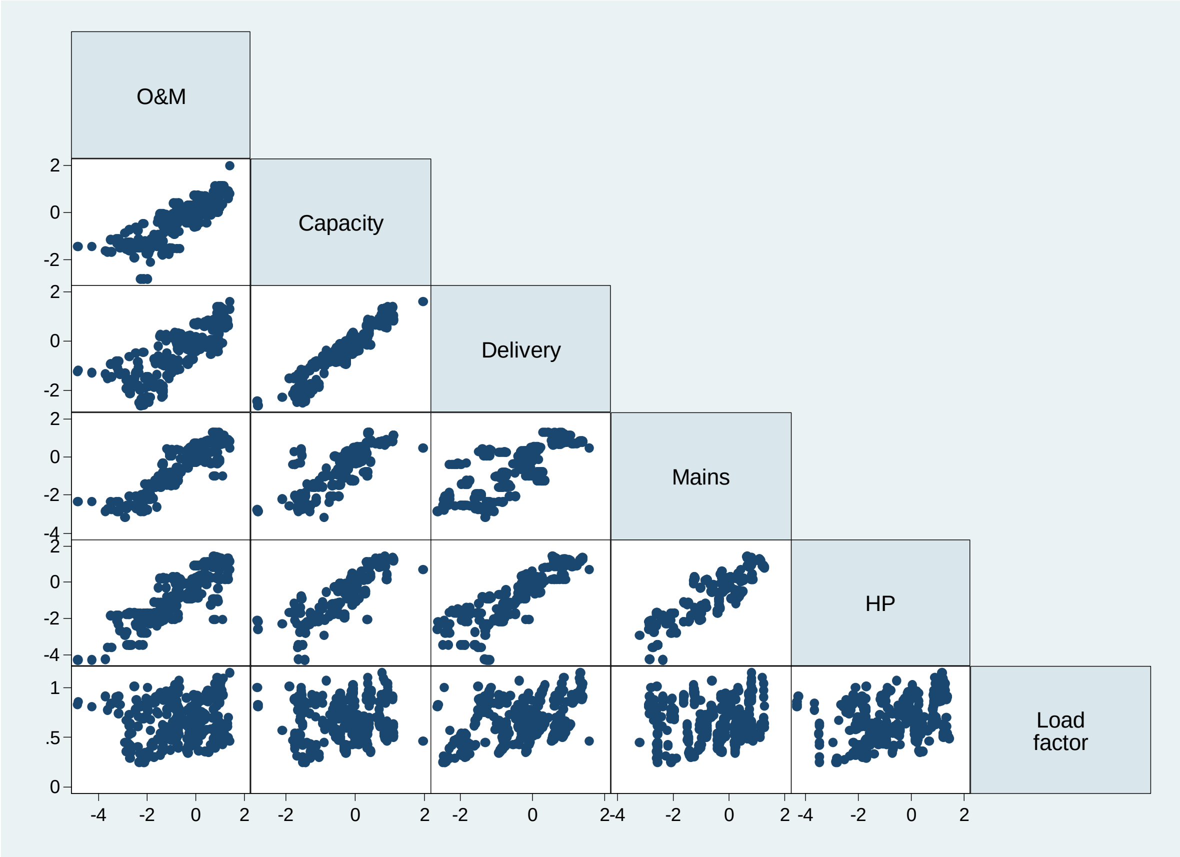

Figure 8: Scatter plots of O&M and various outputs

Next, we make use of Ordinary Least Squares (OLS) regression analysis to determine a representation of the cost function using several cost drivers (outputs). Additionally, OLS would allow the determination of the functional form. Two widely used functional forms are Cobb-Douglas and translog.[21] For ease of interpretation we limit this report to the former. However, using a more flexible functional form such as translog might produce better results and should be performed once standardized data is available.

Table 15 gives the regression results for a model that includes almost all our cost drivers. Each column gives the coefficients and standard errors for one cost measure as the dependent variable. The last row gives the adjusted R-squared, a statistic that expresses the strength of

[21] Whereas the translog model includes the squares of all variables as well as all cross-products the Cobb-Douglas model does not. These additional variables allow a more flexible functional form for the translog model. For instance, it allows costs to increase at different rates for different levels of the cost driver, which is not the case for the Cobb-Douglas model. A drawback is that coefficients are less straightforward to interpret.

the explanatory power of the independent variables. R-squared ranges from 0 to 1. The higher its value, the stronger the common explanatory power of the independent variables.

These models suffer from two shortcomings. First, the high correlation amongst the explanatory variables leads to insignificant coefficients that would be significant individually. Second, the models produce negative coefficients for Units and Delivery, which is inconsistent with the theory. Nevertheless the model serves two important purposes. First, it gives the benchmark R-squared as no other model (given our variables) can achieve a higher overall correlation between inputs and outputs. Secondly, as for each output variable all the coefficients are jointly significant no output variables is completely irrelevant in a statistical sense for the determination of costs.

Table 15: Regression results for full Cobb-Douglas models

	O&M	**Totex1**	**Totex2**	**Revenue**
Capacity	0.989*	0.422	0.382	0.720^
	(0.420)	(0.383)	(0.416)	(0.428)
Delivery	-0.504	0.252	0.310	-0.124
	(0.410)	(0.374)	(0.406)	(0.417)
Mains	0.336**	0.453**	0.413**	0.456**
	(0.054)	(0.050)	(0.054)	(0.055)
Horsepower	0.042	0.033	0.130*	0.191**
	(0.052)	(0.047)	(0.051)	(0.053)
Units	0.317**	-0.146**	-0.255**	-0.270**
	(0.045)	(0.041)	(0.044)	(0.045)
Load factor	0.750	0.142	0.173	0.671
	(0.678)	(0.618)	(0.672)	(0.689)
adj. R Squared	*0.88*	*0.85*	*0.83*	*0.82*

** p<0.01; * p<0.05; ^ p<0.10 two tailed

Having experimented with various model specifications we selected the models shown in Table 16. Given the large number of available cost drivers, using only two might seem odd. However, the high correlations among the cost drivers justify this selection. Also, note that the model excludes throughput entirely. Given economies of scale and the fact that in most countries the tariff component for capacity far outweighs the component for throughput the model seems plausible.

Table 16: Regression results for final Cobb-Douglas models

	O&M	Totex1	Totex2	Revenue
Capacity	0.508**	0.693**	0.779**	0.728**
	(0.048)	(0.043)	(0.050)	(0.050)
Mains	0.688**	0.325**	0.224**	0.278**
	(0.036)	(0.032)	(0.037)	(0.037)
adj. R Squared	*0.86*	*0.82*	*0.77*	*0.77*

** p<0.01; * p<0.05; ^ p<0.10 two tailed

We also omitted from this cost-driver analysis environmental variables like age, time, regulatory regime, soil conditions, etc. Because the passing of time is readily measured by the year variable we included this in a model (not reported) and found that it has a negative coefficient for all cost measures but is significant for revenue only. This might imply that as time passes costs and prices fall due to depreciation and possibly higher factor productivity. Alternatively, demand might have fallen over this period. Environmental variables, given that they have a significant impact should be part of a benchmarking for the direct purpose of tariff setting.

The input variables chosen on theoretical grounds above and the output variables chosen on econometric and theoretical grounds in this section combine to our final input-output variable selections presented in Table 17. The model specifications are chosen in the next section.

Table 17: Input-output variable selections

	Opex specifications	Totex specifications		
Dependent variable	O&M	Totex1	Totex2	Revenue
Independent variable	Capacity	Capacity	Capacity	Capacity
	Mains	Mains	Mains	Mains

To conclude, on economic and statistical grounds capacity both measured in terms of throughput and physical assets is a significant cost driver. The high correlations among the cost drivers make it difficult to disentangle individual effects.[22] We believe that various measures of capacity might be sufficient to perform a benchmark of gas transmission operators based on totex cost measures. When O&M is benchmarked the inclusion of a throughput variable might be warranted. Unfortunately we cannot pass judgment on other variables that are not included in our analysis like diameter and number of entry/exit points.

[22] There exist econometric techniques to solve this problem. However, these rely on substituting the original variables for newly created virtual ones. This affects the transparency of the results and therefore might not be appropriate for a regulatory context.

7.1. Appendix

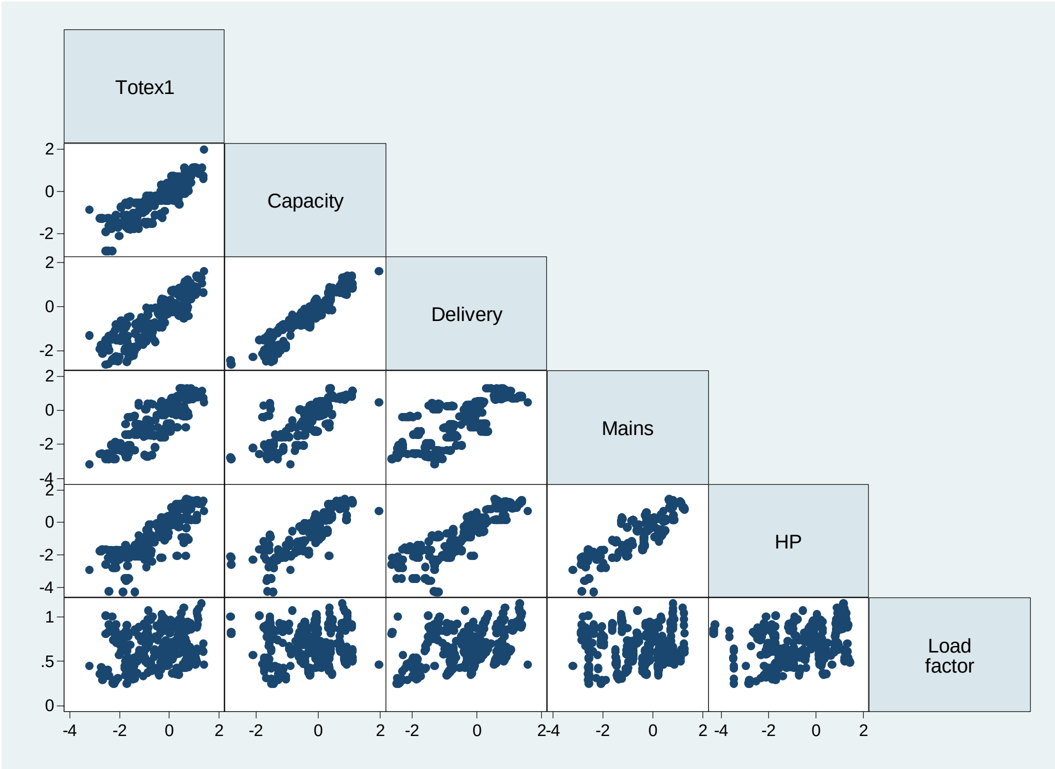

Figure 9: Scatter plots of Totex1 and various outputs

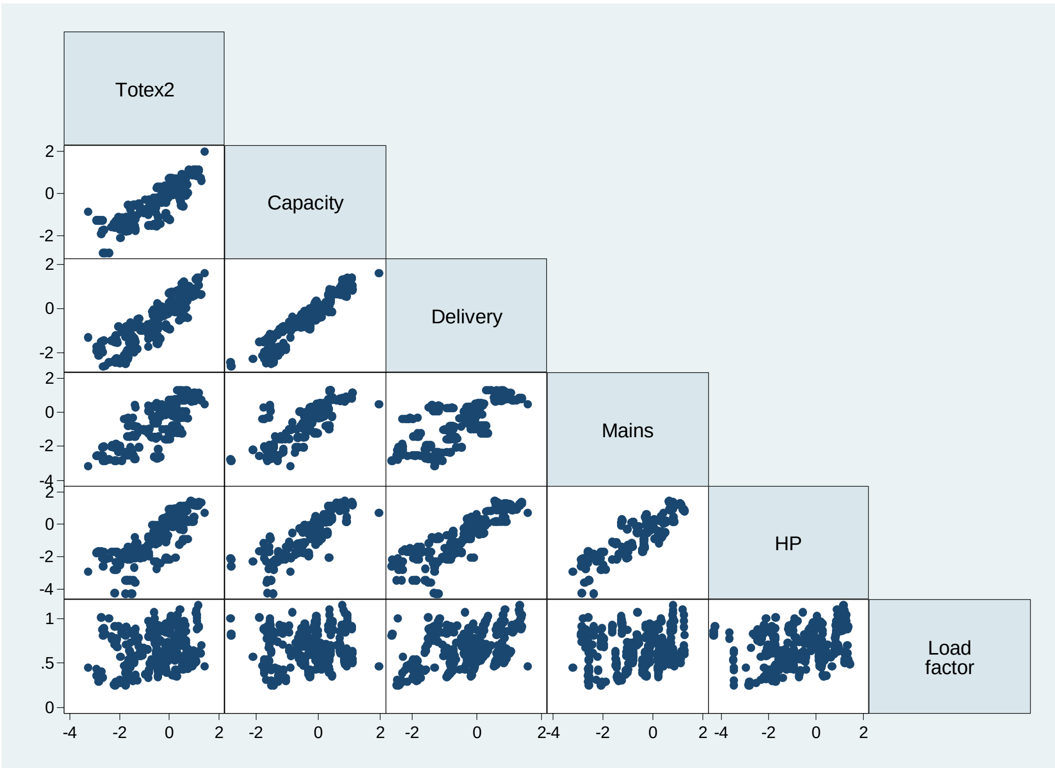

Figure 10: Scatter plots of Totex2 and various outputs

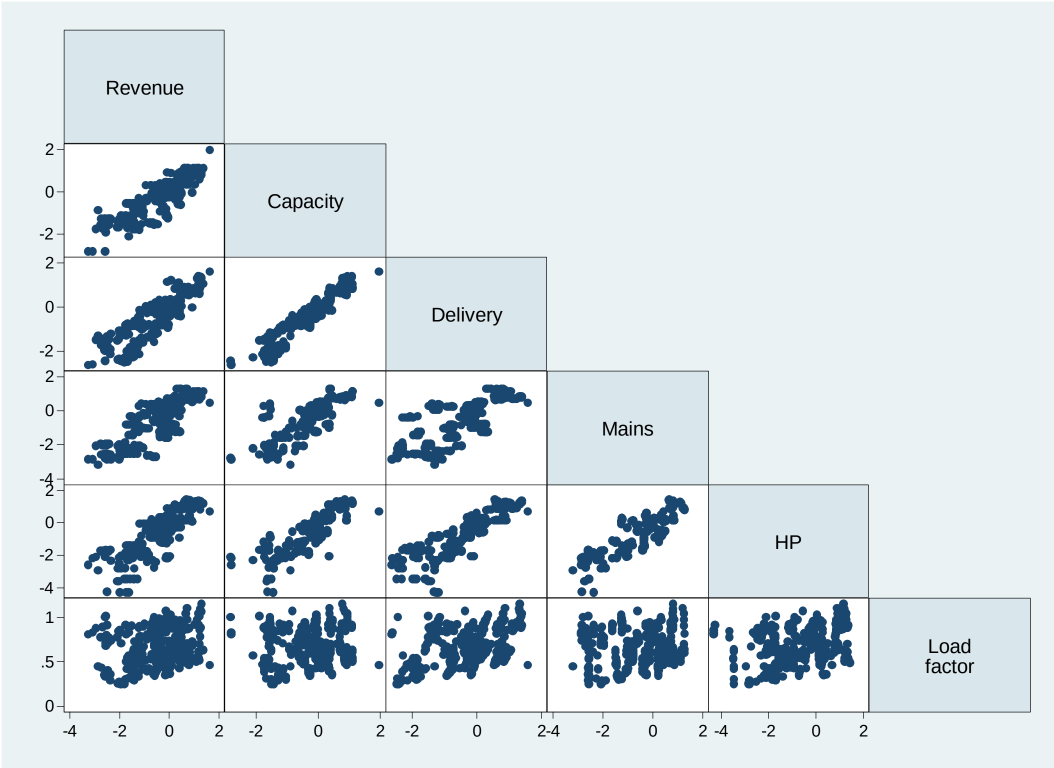

Figure 11: Scatter plots of Revenue and various outputs

8. Efficiency Benchmarking

This section constitutes the core of this report. Here we report on the actual efficiency benchmarking analysis and present individual efficiency scores. The section relies heavily on summary tables and graphics to convey the results.

Before we present the results we need to stress that the emphasis of this report is on the methodology and not on the production of individual efficiency scores. This does not mean that our results do not contain any relevant information, but at this stage we believe that neither the comparability among the European firms nor across the continents is sufficient to take the results presented here as final.

In the remainder we present results for all three techniques discussed above: Stochastic Frontier Analysis, Corrected Ordinary Least Squares and Data Envelopment Analysis. We also perform a series of consistency tests across models and techniques. Note that the input-output variables were selected in the cost driver analysis above. The remaining model specifications are given in Table 18. The four model specifications listed and the four input-output selections from the cost driver analysis sum up to sixteen sets of efficiency scores.

Table 18: Model specifications

Technique	**SFA**	**COLS**	**DEA**	
Functional form	Cobb-Douglas	Cobb-Douglas	n.a.	
Scale	VRS	VRS	VRS	CRS
Orientation	Input	Input	Input	Input
Number of input-output specifications	4	4	4	4
Outlier treatment	None	Frontier through 90th percentile	Removal of 10% of most efficient firms	Removal of 10% of most efficient firms
Other specification	Half-normal or exponential distribution of the inefficiency component	Normal distribution of the inefficiency component	Radial	Radial
Total number of runs: 4 model specifications * 4 input-output selections = 16				

This choice of variables and model specifications seems to be appropriate for the objectives of this report. Undoubtedly, different choices might lead to different results. And for the purpose of actual tariff setting more options should be considered.

8.1. Stochastic Frontier Analysis

SFA is the technique that is the most demanding in practice. Unlike in COLS and DEA there is no guarantee that for a given data set and model specification the technique will produce useful results. Thus, it is almost impossible to set up an SFA tool box independent of the actual dataset at hand. It is worthwhile mentioning again that the SFA literature suggests that panel data models produce results with greater ease. Unfortunately, the limited availability of European data does not allow the use of panel data models here. However, even without panel data we do not advice dropping SFA entirely.

Table 19 gives the mean efficiency scores across cost measures and countries for SFA. Note that the maximum range of SFA, COLS and DEA scores lies between 0 and 1. In order to obtain relevant results we allowed for varying assumptions on the distribution of the inefficiency term. Whereas the distributions for O&M and Revenue are exponential, the distributions for Totex1 and Totex2 are half-normal. As a result the mean scores for the exponential distributions should be somewhat higher as shown in Figure 12. Table 19 shows that *on average* all European countries, except country D perform worse than the US on all cost measures. Also, A performs above US average in Totex2 and Revenue, which might relate to its depreciation practice.

Table 19: SFA mean efficiency scores by cost measure country

Country	O&M	Totex1	Totex2	Revenue
A	.34	.44	.65	.80
B	.72	.31	.34	.72
C	.48	.39	.43	.78
D	.77	.71	.71	.79
US	.79	.67	.65	.79

Figure 12 gives histograms of the relevant scores. These plots illustrate the distribution of the efficiency scores. The distribution for O&M is the most centred. For the totex measures revenue is actually the most centred. For O&M this might be an indication that best practice is more widespread the narrower the cost measure is. For Revenue, as the widest cost measure, the reason might be that the measurement of revenue is more standardized as we relied on revenue as reported and not individual cost components (and assumptions to produce them). However, all totex distributions have a rather large spread. We suspect that this might be caused by insufficient standardization or by the functional form we chose. A functional form like translog that "fits" the data better might produce more centred scores for all cost measures. Additionally, genuine outliers would influence the distribution.

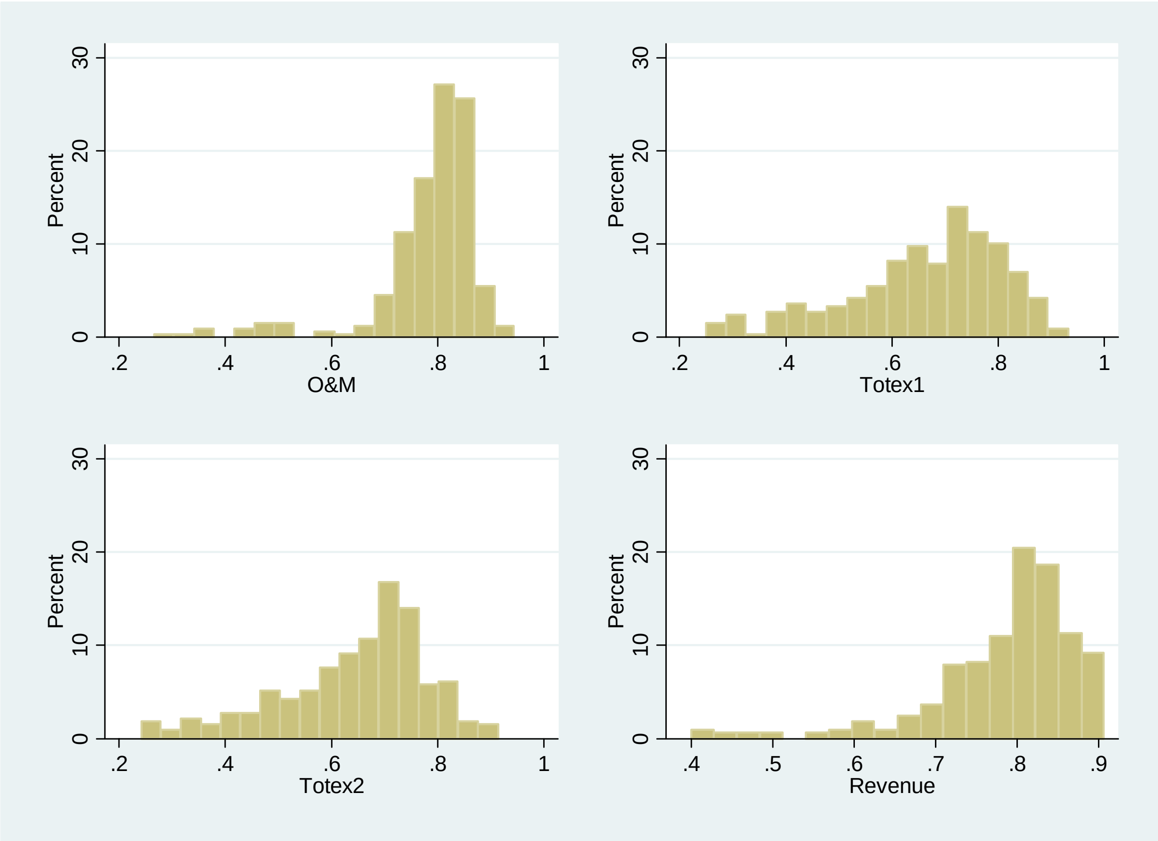

Figure 12: Histograms of SFA efficiency scores by cost measure

The distribution of the efficiency scores by cost measure and country is also given by the box plots in Figure 13. Generally, the European countries rank below the US median but not lower than the US minimum. Country D performs noticeably better then the other European countries. This general pattern can also be found in the efficiency scores for COLS and DEA shown below. We would like to draw the attention to the revenue scores. It is interesting that, the median scores for the widest cost measure have the lowest variance across countries.

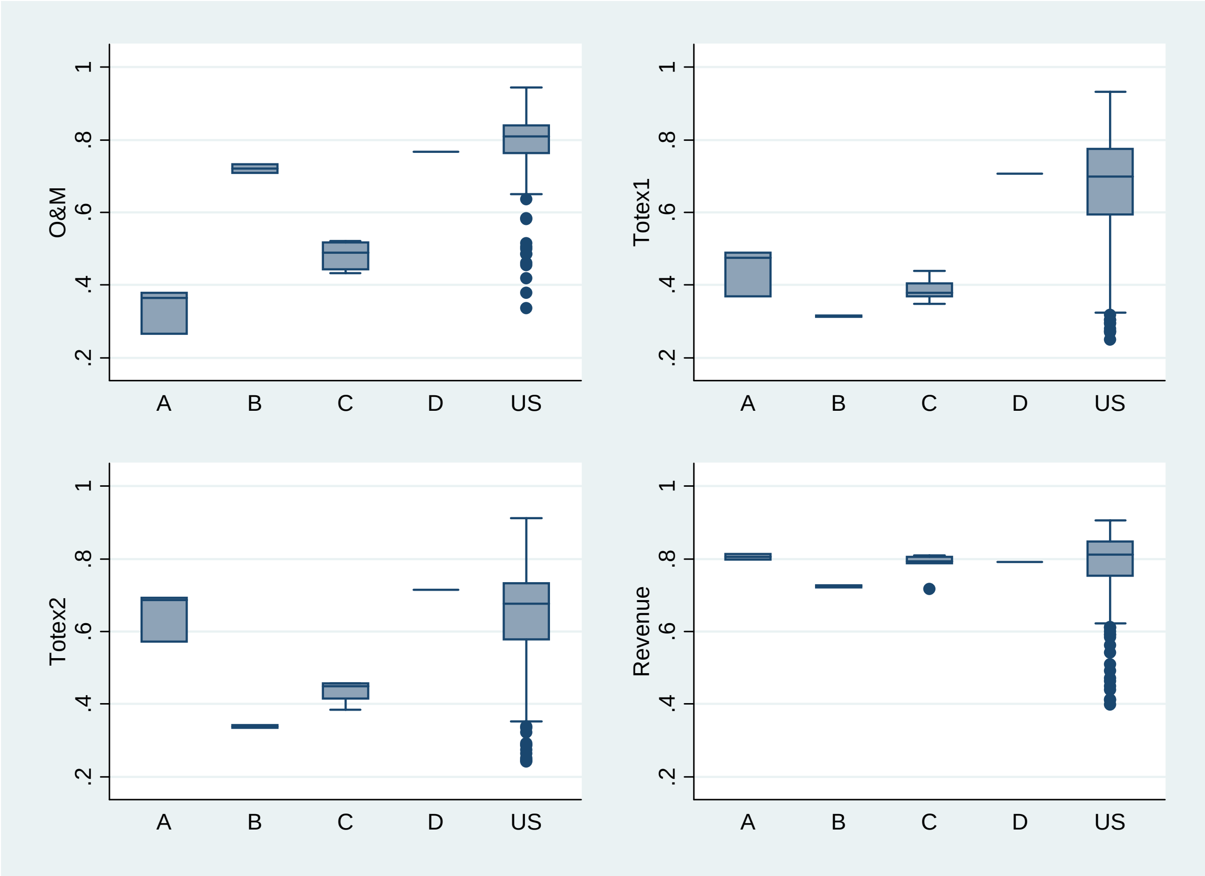

Figure 13: Box plots of SFA efficiency scores by cost measure and country

Figure 14: Scatter plots of SFA efficiency scores by cost measure and year

Figure 14 gives individual efficiency scores by country and year. This highlights that there might be differences in performance across time. But again, the relative ranking of the European firms does not change significantly across time.

Note that to facilitate comparison across techniques we reproduce the same tables and graphs in the same order for COLS and DEA below.

8.2. Corrected Ordinary Least Squares

COLS mean efficiency scores are given in Table 20. The high ranking for A for Totex2 is explained again by its depreciation practice. Remember that we "lowered" the frontier to the 90^{th} percentile.

Table 20: COLS mean efficiency scores by cost measure and country

Country	O&M	Totex1	Totex2	Revenue
A	.15	.35	.58	.50
B	.45	.23	.21	.33
C	.22	.28	.28	.44
D	.53	.64	.66	.44
US	.67	.63	.59	.55

The histograms in Figure 15 give the distribution of efficiency scores for COLS. Unlike for the SFA distributions above O&M has not the narrowest distribution. Revenue, however, has a low variance as above. Generally, the difference in the distribution between SFA and COLS is caused by the differences in the assumptions on the distribution of inefficiency component. In particular, SFA distributions have a left skew whereas COLS distributions do not.

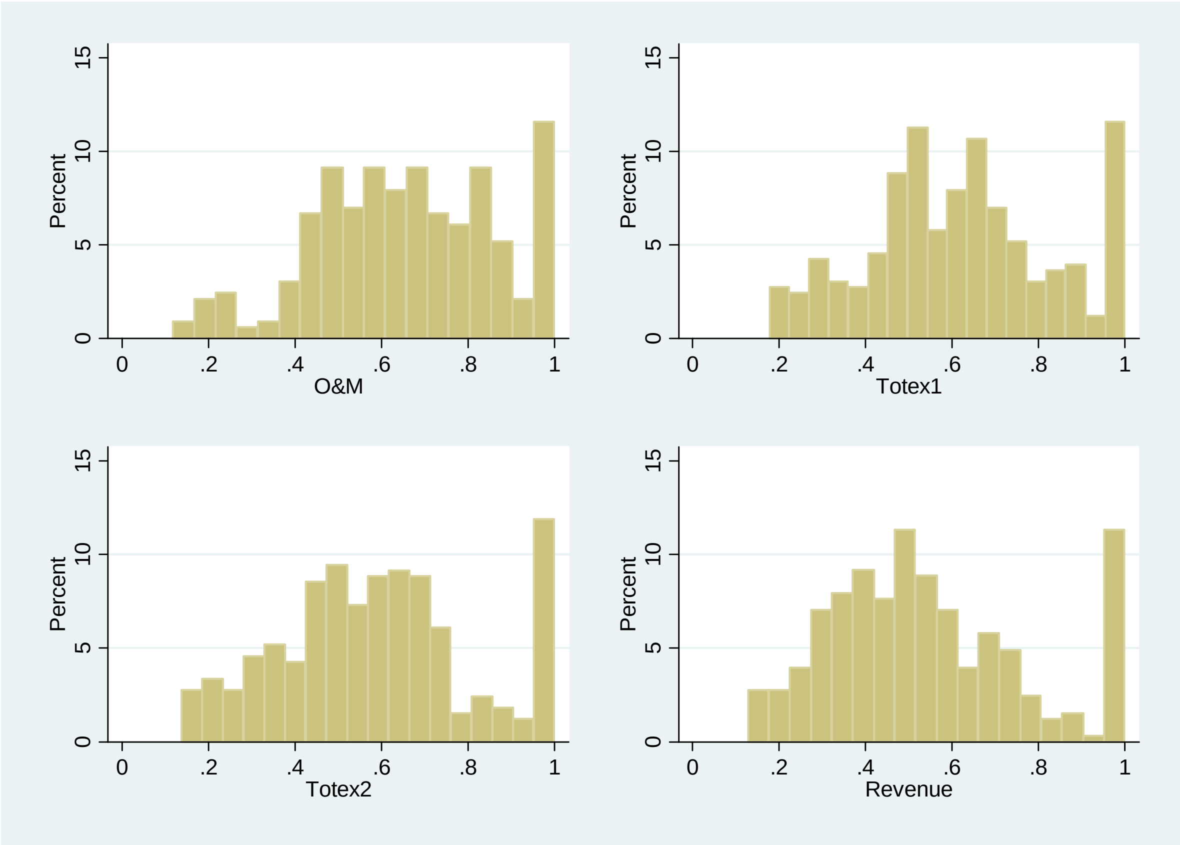

Figure 15: Histogram of COLS efficiency scores by cost measure

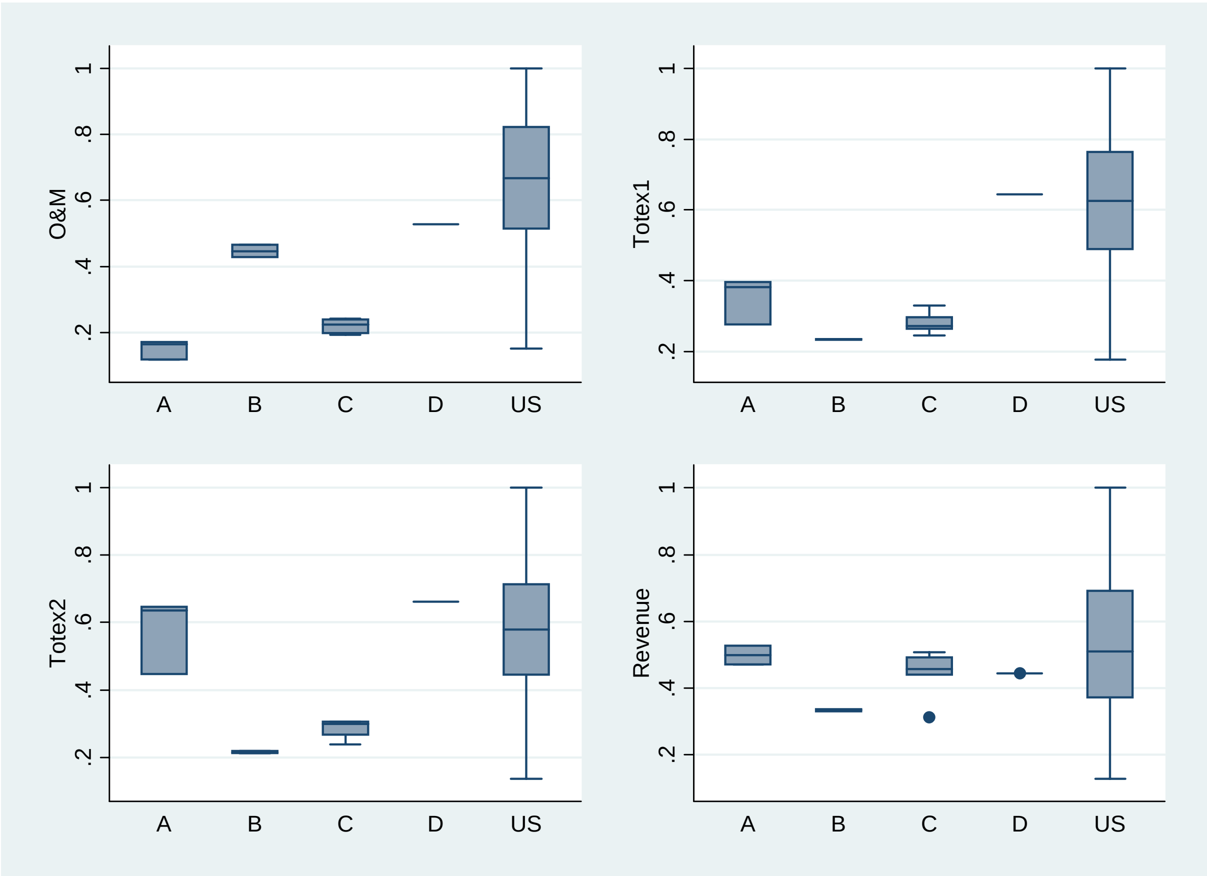

Figure 16: Box plots of COLS efficiency scores by cost measure and country

Figure 16 gives box plots for the COLS scores. Generally, COLS scores confirm the relative ranking of the European countries among each other and vis-à-vis the US companies. The rankings are identical to the rankings produced by SFA above.

Figure 17 gives the COLS scores across years. Again, the relative rankings do not vary much across years.

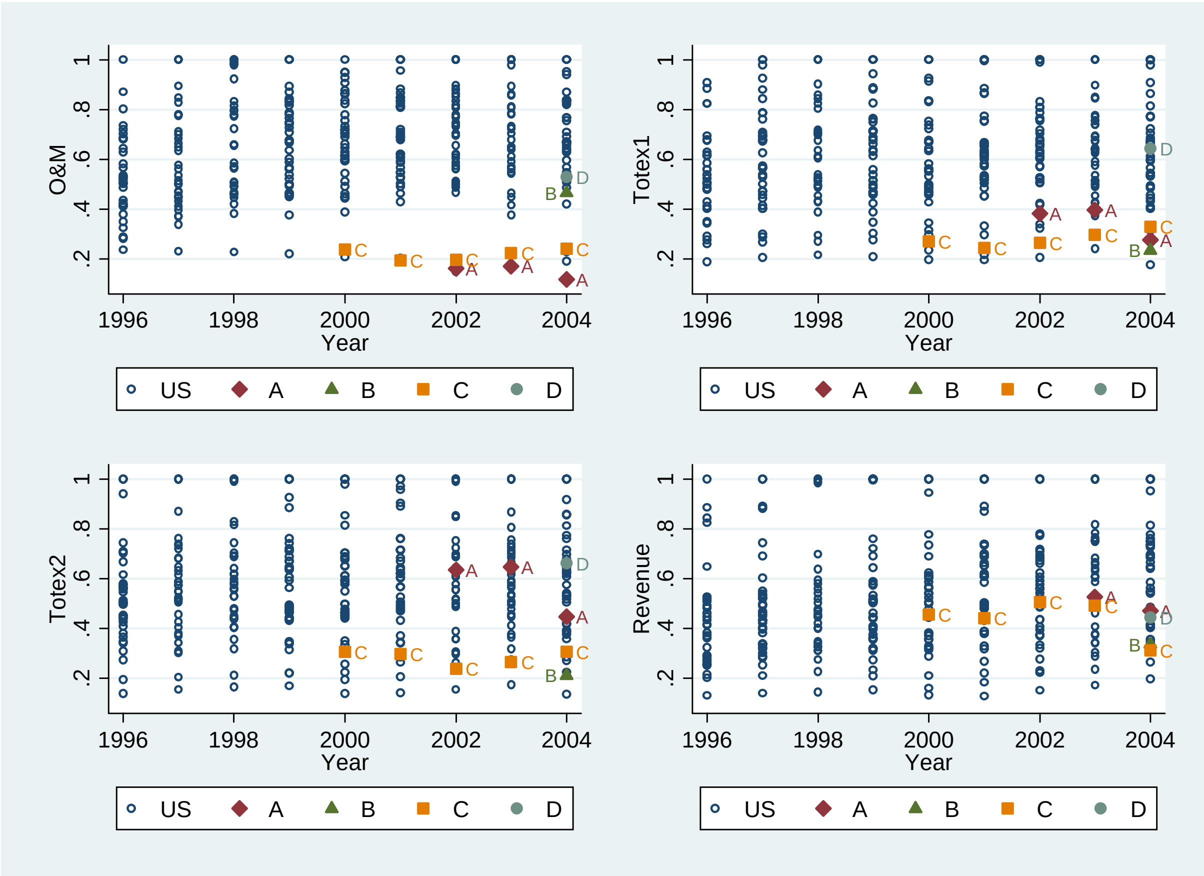

Figure 17: COLS efficiency scores by cost measure and year

8.3. *Data Envelopment Analysis*

Table 21 and Table 22 give the DEA mean efficiency scores for constant returns to scale and variable returns to scale models respectively. The SFA and COLS score above should be compared to the VRS scores. The intuition for VRS is that gas transmission has decreasing average production costs and therefore increasing returns to scale. In DEA CRS implies a more stringent frontier. Thus, if firms actually exhibit VRS one would expect the VRS scores to be higher than the CRS scores.

Except for countries B and C, all scores increase when moving from CRS to VRS. Country D moves from a below average score for CRS to the frontier for VRS. As VRS tries to compare firms of the same absolute size it seems that DEA cannot identify any comparator for country D. Thus, by construction D moves to the frontier. The result that B (and to a lesser extend C) perform at their optimal levels of scale seems odd and requires further investigation.

Table 21: DEA (CRS) mean efficiency scores by cost measure and country

Country	O&M	Totex1	Totex2	Revenue
A	.12	.28	.59	.42
B	.43	.19	.21	.27
C	.17	.24	.24	.39
D	.32	.52	.70	.38
US	.55	.58	.57	.52

Table 22: DEA (VRS) mean efficiency scores by cost measure and country

Country	O&M	Totex1	Totex2	Revenue
A	.17	.45	.71	.65
B	.44	.19	.21	.29
C	.17	.25	.27	.42
D	1	1	1	1
US	.60	.63	.61	.57

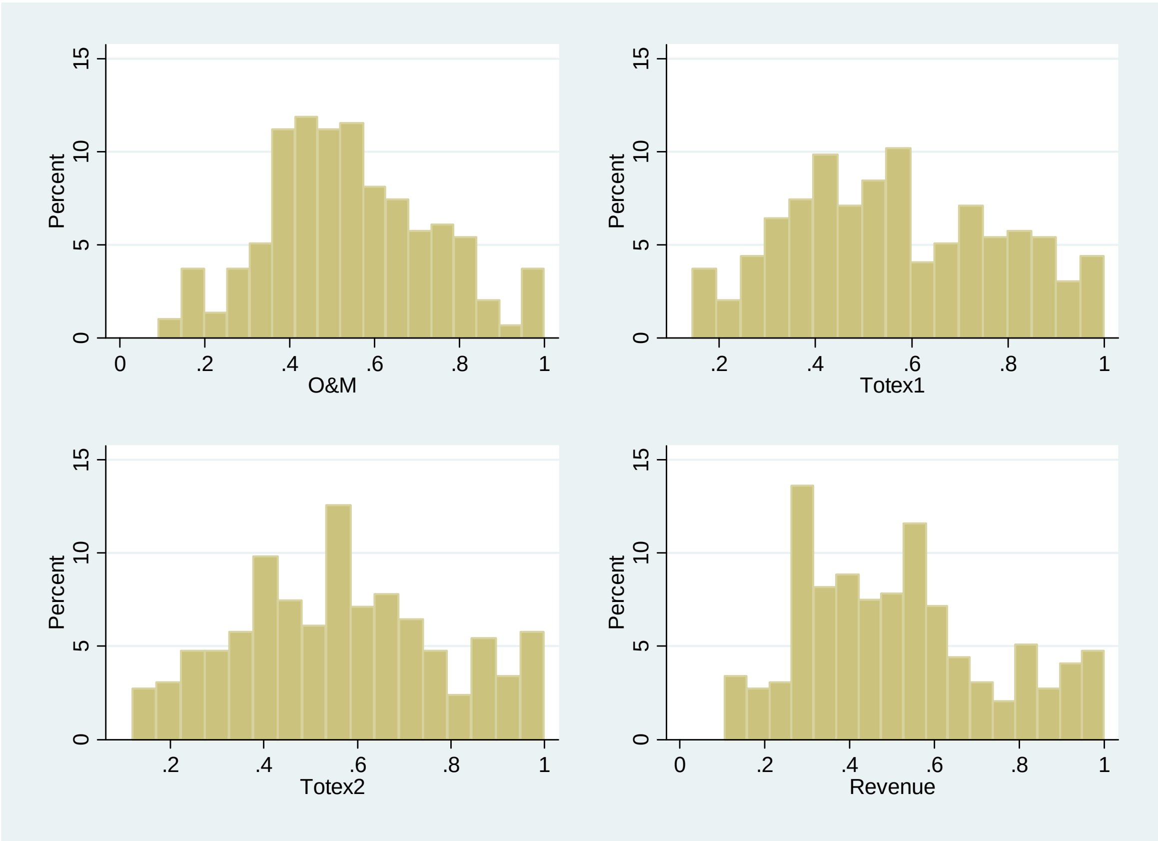

Figure 18: Histogram of DEA (CRS) scores

Figure 18 gives histograms for the DEA (CRS) scores (Figure 22 in the chapter's appendix does the same for VRS). When comparing the distribution of the DEA scores with both the

SFA and COLS scores the DEA distribution is less normal. This illustrates that both SFA and COLS produce specific distributions for the scores that reflect the assumptions being made on the distribution of the variables or inefficiencies. No such assumptions are required for DEA.

Figure 19 and Figure 20 give box plots of the CRS and VRS DEA scores by country. Both give support to the patterns of relative rankings identified above.

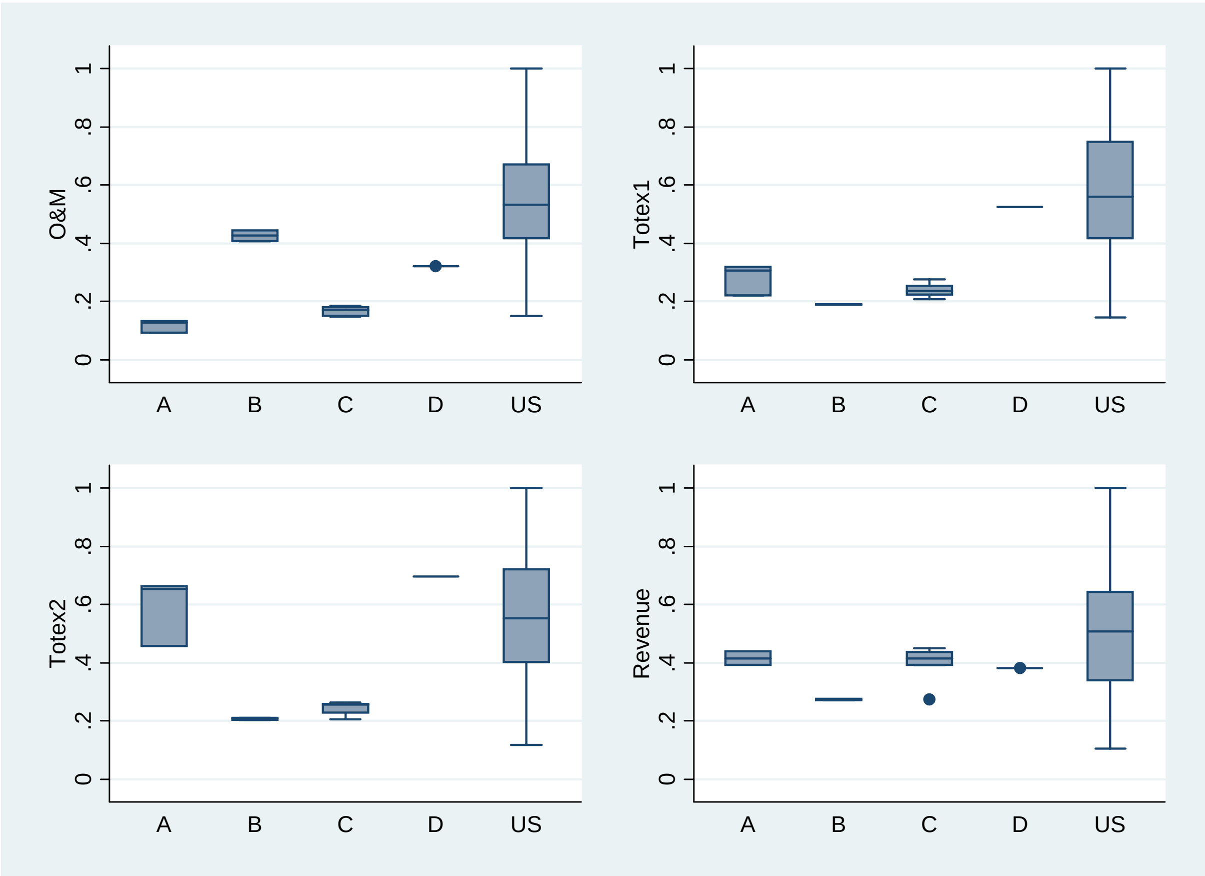

Figure 19: Box plots of DEA (CRS) efficiency scores by cost measure and country

Figure 21 gives the DEA (CRS) scores across time (Figure 23 in the chapter's appendix does the same for VRS). Again, the pattern hardly changes across time.

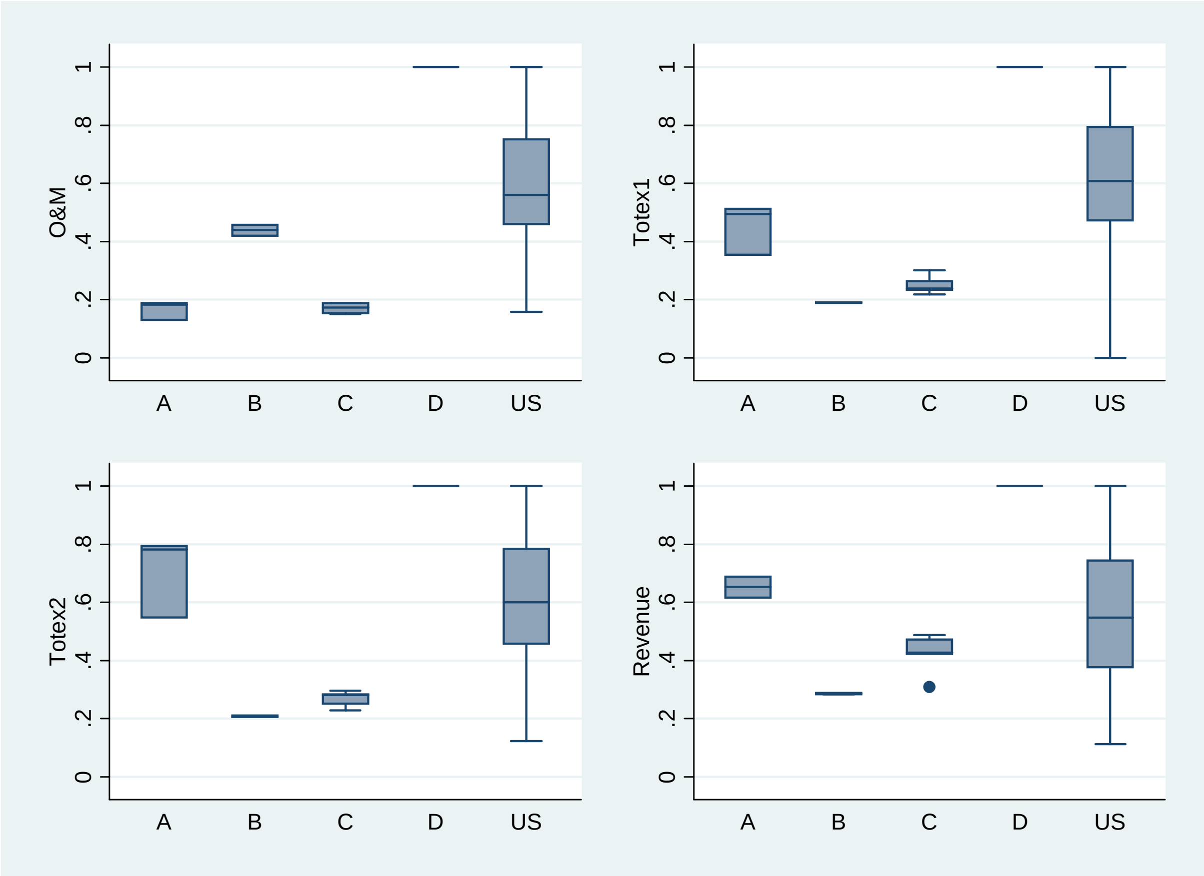

Figure 20: Box plots of DEA (VRS) efficiency scores by cost measure and country

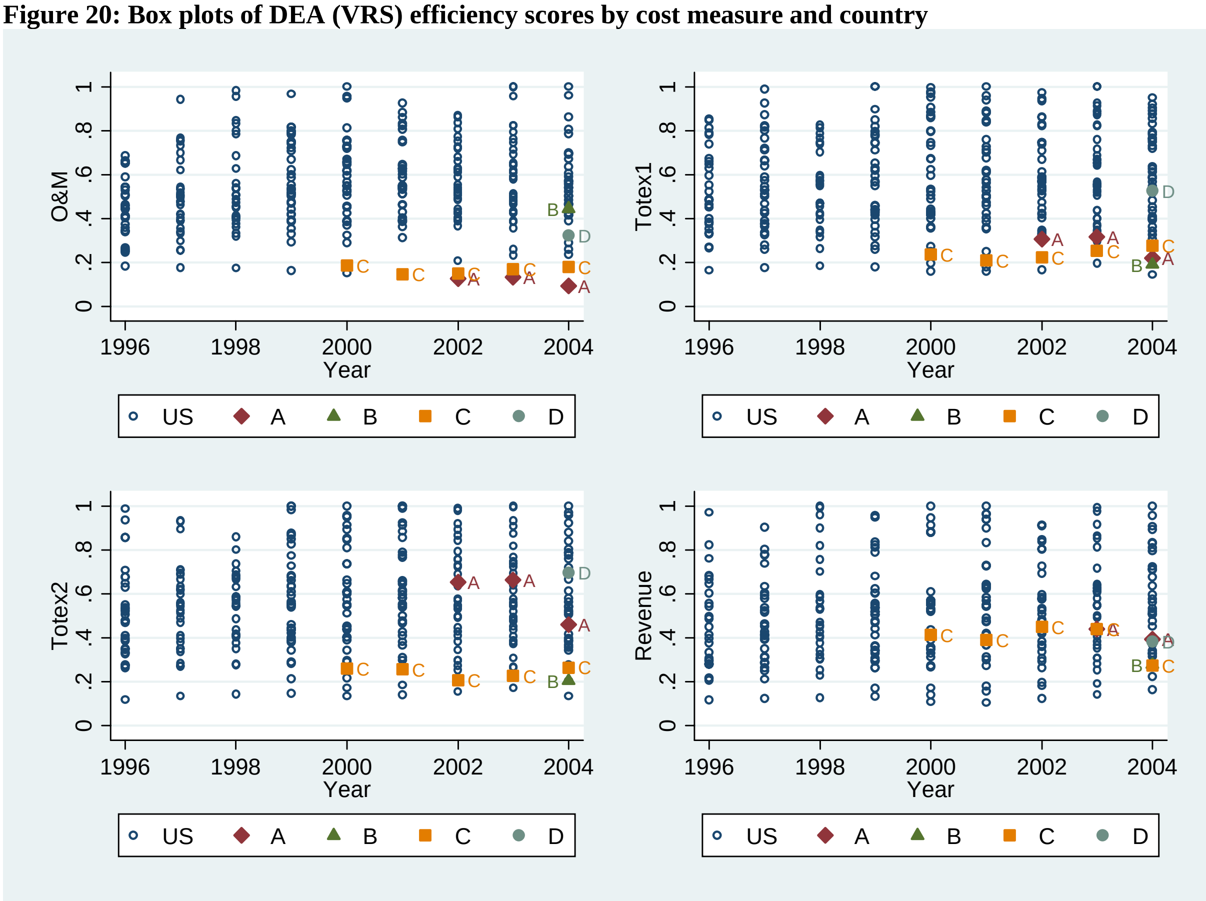

Figure 21: DEA (CRS) scores across time

8.4. Consistency

8.4.1. Correlations

Visual inspection of the efficiency scores for SFA, DEA, and COLS already led to conclusions on the relative rankings of the participating countries. Here we formally test the (rank) correlations of all individual companies.[23] The correlation coefficients range from -1 to 1. Missing values indicate that a given correlation is statistically insignificant. Roughly speaking a coefficient of 1 indicates that any two pairs of firms are ranked exactly the same by the two methods tested. One would expect correlations to be higher for a given technique or a given model and lower across techniques and/or models. Table 23 shows that across the SFA models all correlations are significantly positive.

Table 23: Rank correlations for SFA scores

	O&M	Totex1	Totex2	Revenue
O&M	1.00			
Totex1	0.54	1.00		
Totex2	0.36	0.93	1.00	
Revenue	0.23	0.71	0.79	1.00

The correlations for the COLS and DEA scores are reported in Table 24 and Table 25 respectively. Note that all three techniques produce virtually the same rank correlation scores.

Table 24: Rank correlations for COLS scores

	O&M	Totex1	Totex2	Revenue
O&M	1.00			
Totex1	0.53	1.00		
Totex2	0.35	0.92	1.00	
Revenue	0.23	0.71	0.79	1.00

Table 25: Rank correlations for DEA (CRS) scores

	O&M	Totex1	Totex2	Revenue
O&M	1.00			
Totex1	0.55	1.00		
Totex2	0.42	0.93	1.00	
Revenue	0.23	0.73	0.78	1.00

[23] We measure the Spearman rank correlation as opposed to the Pearson product-moment correlation we used in Chapter 7. The latter is more stringent as it requires joint-normality and linearity.

Table 26: Rank correlations for DEA (VRS) scores

	O&M	Totex1	Totex2	Revenue
O&M	1.00			
Totex1	0.56	1.00		
Totex2	0.41	0.89	1.00	
Revenue	0.29	0.72	0.82	1.00

To test whether our visual impression that the correlation among European firm is quite strong we tested for the correlation across European firms only. We would expect that the rank correlations for the European firms are somewhat higher then for the total sample. Due to the low number of European observations most correlations are insignificant. For the significant ones the evidence is mixed as illustrated in Table 27 for DEA.

Table 27: Rank correlations for DEA (VRS) scores, European only (missing implies insignificance)

	O&M	Totex1	Totex2	Revenue
O&M	1.00			
Totex1		1.00		
Totex2		0.90	1.00	
Revenue		0.84	0.75	1.00

Table 28 gives the correlations between the DEA scores for VRS and CRS. The highlighted cells on the diagonal are the correlations across identical cost measures. Unsurprisingly, these are higher than the ones off-diagonal.

Table 28: Rank correlation for DEA VRS to CRS

CRS/VRS	O&M	Totex1	Totex2	Revenue
O&M	0.91	0.52	0.34	0.18
Totex1	0.51	0.90	0.86	0.66
Totex2	0.39	0.85	0.93	0.73
Revenue	0.21	0.70	0.79	0.93

Ultimately, we are interested in the correlations across techniques shown in Table 29 and Table 30. Whereas the former gives the standard (Pearson) correlations the latter gives the rank correlations. Again, correlations across cost measures, which are highlighted on the diagonals, rather than across cost measures *and* techniques, are fairly high. In particular the correlation between SFA and COLS is very high. As mentioned above SFA and COLS produce rank correlations equal to one by construction. Once we move off-diagonal correlations drop considerably.

Table 29: Correlation coefficients for SFA, COLS and DEA (VRS) scores

		SFA				COLS			
		O&M	Totex1	Totex2	Revenue	O&M	Totex1	Totex2	Revenue
COLS	O&M	0.87	0.57	0.33	0.15				
	Totex1	0.53	0.96	0.88	0.66				
	Totex2	0.34	0.89	0.94	0.73				
	Revenue	0.13	0.66	0.74	0.85				
DEA (VRS)	O&M	0.76	0.48	0.27	0.11	0.91	0.51	0.31	0.17
	Totex1	0.41	0.86	0.80	0.64	0.51	0.88	0.81	0.63
	Totex2	0.29	0.84	0.90	0.73	0.36	0.82	0.89	0.72
	Revenue	0.15	0.63	0.69	0.80	0.24	0.58	0.64	0.89

Table 30: Rank correlation for SFA, COLS, and DEA (VRS) scores

		SFA				COLS			
		O&M	Totex1	Totex2	Revenue	O&M	Totex1	Totex2	Revenue
COLS	O&M	0.99	0.58	0.36	0.21				
	Totex1	0.57	0.99	0.91	0.68				
	Totex2	0.35	0.91	0.99	0.77				
	Revenue	0.21	0.68	0.76	0.99				
DEA (VRS)	O&M	0.91	0.52	0.32	0.19	0.92	0.52	0.31	0.19
	Totex1	0.54	0.89	0.81	0.66	0.53	0.88	0.81	0.66
	Totex2	0.38	0.84	0.91	0.75	0.37	0.82	0.89	0.74
	Revenue	0.27	0.65	0.72	0.92	0.25	0.63	0.70	0.92

Generally, the high correlations confirm that our results are consistent. Bauer (1998) found an average rank correlation across parametric and non-parametric techniques for his sample of less than 0.1.

However this is only one of several consistency tests that the literature suggests. Bauer (1998), for instance suggests six. He suggests stability across time which we do not test formally due to a lack of observations. But various figures giving scores by year above suggest that there is some stability across time. He further suggests comparison with non-frontier standard performance measures. The one measure that is readily available is the unit costs we presented in Chapter 5. Table 31 gives the rank correlations between various unit cost measures and the efficiency scores for SFA (Table 35, and Table 36 in the chapter's appendix do the same for COLS and DEA). All significant correlations are negative as expected. We expect negative correlations because a high unit cost implies low efficiency. However, we do not expect perfect negative rank correlations (i.e. coefficients equal to -1) because the unit costs take into account only one cost driver at a time and not multiple cost drivers as the benchmarking techniques do. The correlations for the same cost measures (highlighted) are stronger than the ones across cost measures.

Table 31: Rank correlations between SFA efficiency scores and various unit cost measures

	O&M	Totex1	Totex2	Revenue
O&M/Cap.	-0.47	-0.28	-0.19	-0.14
O&M/Del.	-0.44	-0.14		
O&M/Mains	-0.84	-0.53	-0.38	-0.22
O&M/HP	-0.43	-0.16		
Totex1/Cap.	-0.41	-0.75	-0.69	-0.58
Totex1/Del.	-0.42	-0.54	-0.44	-0.35
Totex1/Mains	-0.40	-0.73	-0.67	-0.47
Totex1/HP	-0.24	-0.41	-0.32	-0.19
Totex2/Cap.	-0.28	-0.81	-0.88	-0.74
Totex2/Del.	-0.30	-0.60	-0.62	-0.51
Totex2/Mains	-0.31	-0.68	-0.70	-0.51
Totex2/HP	-0.14	-0.45	-0.44	-0.31
Revenue/Cap.	-0.17	-0.59	-0.66	-0.91
Revenue/Del.	-0.23	-0.44	-0.46	-0.69
Revenue/Mains	-0.20	-0.56	-0.60	-0.66
Revenue/HP		-0.35	-0.36	-0.46

Another test proposed is the test for equal distributions of the scores across models and techniques. Above we already depicted several distributions of scores and mentioned we would *not* expect them to be identical across techniques. For completion we give the numerical summary statistics for all 16 distributions in Table 32.

Table 32: Summary statistics of the distributions of the efficiency scores

Technique	Variable	Obs.	Mean	Std. Dev.	Min	Max
SFA	O&M	328	.78	.10	.27	.94
	Totex1	328	.66	.15	.25	.93
	Totex2	328	.64	.14	.24	.91
	Revenue	327	.79	.09	.40	.90
COLS	O&M	328	.66	.21	.12	1
	Totex1	328	.62	.22	.18	1
	Totex2	328	.59	.23	.14	1
	Revenue	327	.54	.23	.13	1
DEA (CRS)	O&M	295	.54	.19	.09	1
	Totex1	295	.57	.22	.14	1
	Totex2	295	.56	.22	.12	1
	Revenue	294	.52	.22	.11	1
DEA (VRS)	O&M	295	.59	.21	.13	1
	Totex1	295	.62	.22	0	1
	Totex2	295	.61	.23	.12	1
	Revenue	294	.56	.23	.11	1

Across models for a given technique mean and standard deviations are rather stable. Note that the similar means *across* techniques are partly a product of our approach towards outliers.

Note also that the SFA results have the lowest variance, which might be due to the explicit accounting for measurement error.

Last, we list the ten best and worst performers by technique in Table 33. This allows us to assess whether the frontier is stable across techniques (here we only show the results for O&M). We observe that SFA and COLS are more likely to produce the same frontier whereas the DEA frontier seems to be somewhat different.

Table 33: Highest and lowest scores for O&M

	SFA		COLS		DEA (CRS)		DEA (VRS)	
	score	name	score	name	score	name	score	name
Highest	0.94	Trailblazer Pipeline Company	1.00	Gulf South Pipeline LP	1	Natural Gas Pipeline of America	1.00	TransColorado Gas Transmission
	0.94	Trailblazer Pipeline Company	1.00	Northern Border Pipeline	1	Wyoming Interstate Company,	1.00	Natural Gas Pipeline of America
	0.93	Trailblazer Pipeline Company	1.00	Great Lakes Gas Company	1	TransColorado Gas Transmission	1.00	Tennessee Gas Pipeline
	0.91	Trailblazer Pipeline Company	1.00	Wyoming Interstate Company,	1.00	Natural Gas Pipeline of America	1.00	Tennessee Gas Pipeline
	0.90	Trailblazer Pipeline Company	1.00	Trailblazer Pipeline Company	0.98	Gulf South Pipeline LP	1.00	Northern Natural Gas
	0.89	Trailblazer Pipeline Company	1.00	Trailblazer Pipeline Company	0.97	Gulf South Pipeline LP	1.00	East Tennessee Natural Company
	0.89	Natural Gas Pipeline of America	1.00	Trailblazer Pipeline Company	0.96	Wyoming Interstate Company,	1.00	**D**
	0.89	Wyoming Interstate Company,	1.00	Wyoming Interstate Company,	0.96	Wyoming Interstate Company,	1.00	Natural Gas Pipeline of America
	0.89	Natural Gas Pipeline of America	1.00	Gulf South Pipeline LP	0.96	Kern River Gas Company	1.00	Northern Natural Gas
	0.89	Natural Gas Pipeline of America	1.00	Natural Gas Pipeline of America	0.95	East Tennessee Natural Company	1.00	Tennessee Gas Pipeline
Lowest	0.46	Dominion Transmission	0.21	Dominion Transmission	0.17	Dominion Transmission, Inc.	0.18	**A**
	0.46	Dominion Transmission	0.20	Dominion Transmission	0.17	**C**	0.18	Dominion Transmission, Inc.
	0.44	**C**	0.20	**C**	0.16	Dominion Transmission, Inc.	0.18	Dominion Transmission, Inc.
	0.43	**C**	0.19	**C**	0.15	Dominion Transmission, Inc.	0.17	**C**
	0.42	MIGC	0.19	MIGC	0.15	Dominion Transmission, Inc.	0.17	Dominion Transmission, Inc.
	0.38	**A**	0.17	MIGC	0.15	**C**	0.16	Dominion Transmission, Inc.
	0.38	MIGC	0.17	**A**	0.15	**C**	0.16	Dominion Transmission, Inc.
	0.36	**A**	0.16	**A**	0.13	**A**	0.15	**C**
	0.34	MIGC	0.15	MIGC	0.13	**A**	0.15	**C**
	0.27	**A**	0.12	**A**	0.09	**A**	0.13	**A**

Overall, our results are rather consistent across cost measures and/or models with a somewhat less favourable picture for the consistency of the actual frontier. In the following section we shortly test for the robustness of our results with respect to the chosen rate for cost of capital.

8.4.2. Cost of Capital

Note that, as mentioned, we applied a 7 percent rate for the cost of capital for the construction of the Totex2 measure. Here we test whether lowering or increasing the rate has an effect on the SFA efficiency scores. The alternative rates are 5 percent and 9 percent. Table 34 shows that the correlation coefficients are virtually 1. This is not surprising as rates are equal for all companies. Last, note that all else equal the absolute score for a given company decreases when the allowed return increases as total costs increase.

Table 34: Correlations among SFA scores with different rates for cost of capital

	Totex2 (7%)	**Totex2 (9%)**	**Totex2 (5%)**
Totex2 (7%)	1.00		
Totex2 (9%)	0.99	1.00	
Totex2 (5%)	0.99	0.99	1.00

8.5. Appendix

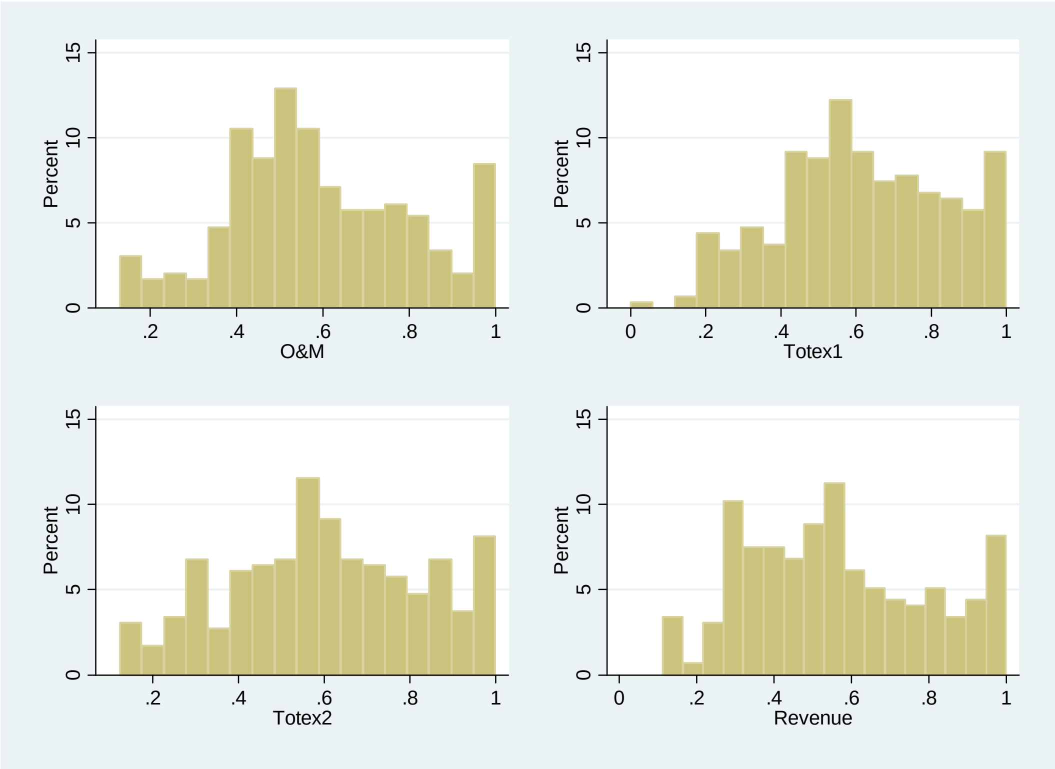

Figure 22: Histograms of DEA (VRS) efficiency scores by cost measure

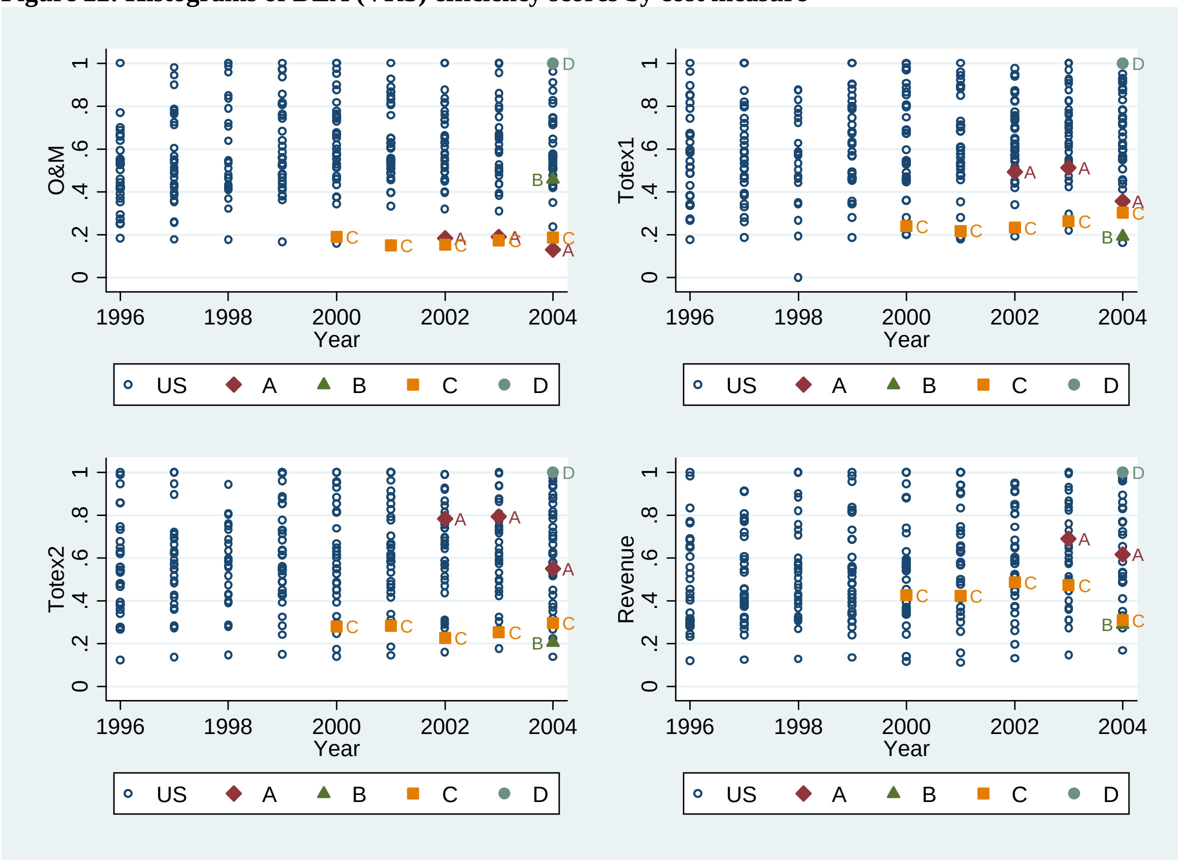

Figure 23: DEA (VRS) efficiency scores by cost measure and year

Table 35: Rank correlations between COLS efficiency scores and various unit cost measures

	O&M	Totex1	Totex2	Revenue
O&M/Cap.	-0.51	-0.36	-0.25	-0.15
O&M/Del.	-0.47	-0.20	-0.06	0.02
O&M/Mains	-0.83	-0.50	-0.37	-0.22
O&M/HP	-0.44	-0.17	-0.00	0.06
Totex1/Cap.	-0.43	-0.81	-0.74	-0.59
Totex1/Del.	-0.44	-0.57	-0.46	-0.35
Totex1/Mains	-0.38	-0.66	-0.62	-0.45
Totex1/HP	-0.24	-0.38	-0.27	-0.18
Totex2/Cap.	-0.30	-0.86	-0.92	-0.75
Totex2/Del.	-0.30	-0.62	-0.63	-0.51
Totex2/Mains	-0.28	-0.60	-0.64	-0.49
Totex2/HP	-0.13	-0.39	-0.39	-0.29
Revenue/Cap.	-0.19	-0.63	-0.70	-0.92
Revenue/Del.	-0.24	-0.46	-0.48	-0.69
Revenue/Mains	-0.18	-0.50	-0.55	-0.65
Revenue/HP	-0.07	-0.31	-0.31	-0.45

Table 36: Rank correlations between DEA (VRS) efficiency scores and various unit cost measures

	O&M	Totex1	Totex2	Revenue
O&M/Cap.	-0.61	-0.18	-0.01	0.09
O&M/Del.	-0.52	-0.04	0.15	0.23
O&M/Mains	-0.82	-0.56	-0.42	-0.25
O&M/HP	-0.42	-0.17	0.03	0.07
Totex1/Cap.	-0.49	-0.67	-0.57	-0.43
Totex1/Del.	-0.41	-0.46	-0.31	-0.20
Totex1/Mains	-0.35	-0.77	-0.75	-0.58
Totex1/HP	-0.24	-0.41	-0.28	-0.21
Totex2/Cap.	-0.29	-0.71	-0.78	-0.61
Totex2/Del.	-0.21	-0.51	-0.49	-0.37
Totex2/Mains	-0.23	-0.70	-0.78	-0.61
Totex2/HP	-0.10	-0.42	-0.39	-0.31
Revenue/Cap.	-0.18	-0.56	-0.62	-0.82
Revenue/Del.	-0.17	-0.43	-0.41	-0.62
Revenue/Mains	-0.14	-0.62	-0.71	-0.79
Revenue/HP	-0.07	-0.35	-0.30	-0.45

9. Discussion

This report outlined a general approach for the benchmarking of individual European gas transmission companies. Though there is room for improvement our approach seems to be capable of producing relevant results, especially for the ranking of firms. The emergence of a rather consistent pattern of relative efficiencies across countries is evidence in support of the validity of our approach.

Whereas country D performs close to the US average A, B and C perform worse. Clearly insufficient standardization puts a question mark behind these results.

Next we try to explain the ranking of countries we observe. A candidate for explaining the higher US average performance might be the higher load factor. The European average load factor is about 20 percent lower than the US. When abstracting from the possible problem of transit flow measurement for the European firms, this might be an indication that the developing pipe-to-pipe competition in the US with strong third-party access rights make US companies more efficient. For one of the TSO's, a prime bias could be a relatively low wage level. However any remaining differences in this area this only put this TSO at an advantage in the cost benchmark.

Last, when comparing the various cost measures some interesting conclusions emerge. Even though the cost measures are highly correlated they show differences in terms of the efficiency scores they produce. The distribution of the scores for revenue stands out for its low variance compared to the other totex measures. Additionally, the similarity in mean and median scores across the countries is strong. As already stated above one reason for this might be that revenue is the most standardized proxy for the cost at hand. Revenue compared to the other measures includes *all* cost components. However, revenue should be adjusted for a minimum rate of return to exclude loss-making firms from the sample and prevent them from dominating the frontier. This might imply that when a regulator has little trust in the quality of data, revenue might be a cost measure that merits particular attention.

10. Recommendations

This report uses a limited number of variables, models and techniques. Undoubtedly, the results obtained here can be further probed and improved upon by using standardized data including more variables, models and specifications.

Standardization of data should be a priority area for the regulators. Inspiration for data collection can be taken from the data requested for this report. Moreover, the US regulator FERC is certainly the main case study for regulatory data collection. However, the benefits of standardization are limited where operators are genuinely different as seems to be the case for country B.[24] Moreover, as we saw, DEA has difficulty producing useful results when there is only one firm of a particular size (this was the case for country D in the DEA VRS model above). This implies that even when there is a large number of European observations, inclusion of non-European data might be required to produce useful results for all techniques.

In this report we made particular choices on which variables to include, which techniques to use, and which models to specify. However, different model specifications might be used in the future. These include:

- using more flexible functional forms for SFA and COLS,
- CRS models for SFA and COLS,
- use of weight restrictions for DEA.

Moreover, it may be possible to improve on the comparability of transit/export volumes and related costs as well as labour costs. The sensitivity to specific assumptions such as the use of PPP as exchange rate can be tested. More variables such as diameter, number of shippers, and tariff system might be added to exclude the possibility that there are better specifications for the cost function.

Last, we would like to summarize our recommendations for the application of benchmarking to gas transmission in Europe. For data and comparability:

[24] The more detailed the cost break-down the better the data standardization can be. In practice, though, it might be difficult for instance, to separate the labour expense related to a particular operation from the total expense. In the case of country B a separate sample of similar companies may need to be prepared.

- robust benchmarking results cannot be obtained without long-term commitment to build appropriate databases,
- this implies agreeing on functional boundaries for transmission operators and standardizing all data,
- in case of insufficient European data, benchmarking with US companies is a viable strategy, surely direct communication with the US regulator FERC can reduce issues of comparability,
- but even without high comparability between US and European firms, it is possible to produce a *relative* ranking of European firms,
- last, under incentive regulation few high quality strategic variables may be sufficient to obtain robust X-factors.[25]

For the variables:

- high correlations among cost-drivers make it difficult to include several of them in standard cost functions for the purpose of econometric analysis,
- on the other hand, this means that some variables can be feasible substitutes for others,
- measures of capacity and network length might be sufficient to capture the cost function of gas transmission operators,
- revenue is highly correlated with the cost measures and produces very similar efficiency scores across firms.

For the benchmarking techniques:

- gas transmission seems to be a business that can easily be modeled and measured econometrically (using OLS). With few strategic output variables we were able to produce rather consistent results,
- Stochastic Frontier Analysis, a theoretically very appealing technique, may not be sufficiently robust to be readily employed without fine-tuning the specific dataset at hand,
- finally, in a regulated industry where rate-payers fund utilities' activities, revenue should be an important regulatory focus area next to cost measures.

[25] This report seems to confirm the point made for instance by FERC (1996) that a move from rate-of-return regulation to incentive regulation shifts the emphasis from *quantity* to *quality* for data collection.

References

Bauer, P. W., Berger, A.N., Ferrier, G.D., Humphrey, D.B. (1998). "Consistency Conditions for Regulatory Analysis of Financial Institutions: A Comparison of Frontier Efficiency Methods." Journal of Economics and Business **50**: 85-114.

CEPA (2003). Background to Work on Assessing Efficiency for the 2005 Distribution Price Control Review, Report produced for Ofgem.

Coelli, T. J., D. S. P. Rao, et al. (2005). An Introduction to Efficiency and Productivity Analysis, Springer.

Comnes, G. A., S. Stoft, et al. (1995). Performance-Based Ratemaking for Electric Utilities: Review of Plans and Analysis of Economic and Resource Planning Issues. Berkeley, Oak Ridge National Laboratory and University of California.

DTe (1999). Price Cap Regulation in the Electricity Sector. The Hague, Netherlands Electricity Regulatory Services (DTe).

Estache, A., M. A. Rossi, et al. (2004). "The Case for International Coordination of Electricity Regulation: Evidence from the Measurement of Efficiency in South America." Journal of Regulatory Economics **25**(3): 271-295.

European Commission (2005). Technical Annexes to the Report from the Commission on the Implementation of the Gas and Electricity Internal Market.

FERC (1996). "The Governance of Energy Displacement Network Oligopolies." Discussion Paper **96-08**.

Gabriel, S. A. and D. Minehart (2004). "Vertical Structure in the Natural Gas Market." Presentation to the East Coast Energy Group.

Granderson, G. (2000). "Regulation, Open-Access Transportation, and Productive Efficiency." Review of Industrial Organization **16**: 368-385.

Hall, G. L. (2000). "Electricity Pricing and Regulatory Practice in a Competitive Environment - Workshop Paper No. 2: Analysis of Alternative Ratemaking Methodologies. Manila. Prepared by Hagler bailly Services, Inc. for Asian Development Bank and the Philippines Energy Regulatory Board and Department of Energy."

Hill, L. J. (1995). "A Primer on Incentive Regulation for Electric Utilities." Oak Ridge National Laboratory.

IPART (1999). Benchmarking the Efficiency of Australian Gas Distributors. New South Wales.

Jamasb, T. and M. Pollitt (2002). International Benchmarking and Regulation of European Electricity Distribution Utilities., Council of European Energy Regulators.

Jamasb, T. and M. Pollitt (2003). "International benchmarking and regulation: an application to European electricity distribution utilities." Energy Policy **31**(15): 1609-1622.

Juskow, P. L. and R. Schmalensee (1986). "Incentive Regulation for electric utilities." Yale Journal on Regulation **4**(1): 1-49.

Kim, T. Y., J. D. Lee, et al. (1999). "International comparison of productivity and its determinants in the natural gas industry." Energy Economics **21**: 273-293.

Kittelsen, S. A. C. (1993). Stepwise DEA: Choosing variables measuring technical efficiency in Norwegian electricity distribution. Oslo, SNF.

Meyrick & Associates (2004). Comparative Benchmarking of Gas Networks in Australia and New Zealand, New Zealand Commerce Commission.

Ofgem (2004). Electricity Distribution Price Control Review: Final Proposal **265/04**.

Pacific Economics Group (2004). New Zealand Gas Distribution Cost Performance: Results from International Benchmarking, commissioned by Vector Networks.

Pollitt, M. (1995). Ownership and Performance in Electric Utilities. Oxford-New York, Oxford University Press.

Sickles, R. C. and M. L. Streitwieser (1991). Technical inefficiency and productive decline in the U.S. interstate natural gas industry under the natural gas policy act.

Weyman-Jones, T., J. C. G. Godinho, et al. (2005). Efficiency and Productivity Analysis for Incentive Regulation.

11. Appendix A: Detailed Literature review

11.1. Meyrick & Associates (2004)

The New Zealand Commerce Commission commissioned this benchmarking of Australian and New Zealand gas transmission and distribution utilities. The data for both functions is separated and here only the transmission sample is discussed. The unbalanced transmission sample contains 8 companies for the years 1997-2003, which results in 31 observations.

The variables used are summarized in Table 37 below. Note that the study includes both capacity as output and capital stock (length of mains) as input. The benchmarking technique used is a multilateral Malmquist index where inputs and outputs are weighted according to cost and production functions respectively. The weights are taken from a different sample as the sample used here is too small to produce robust econometric results.

Table 37: Variables used in Meyrick and Associates (2004)

Outputs		Inputs	
Description	**Measurement**	**Description**	**Measurement**
throughput (1,2)	vol.	O&M (1,2)	including labour
capacity (1)	daily maximum feasible capacity (or peak demand) * pipeline length	capex (1,2)	length of mains
capacity (2)	asset value (optimized deprival value, depreciated optimized replacement cost)	capex (alternative measure)	optimized deprival value, depreciated optimized replacement cost (fixed assets)

Note: the numbers in parenthesis indicate models, in which the respective variable is included. Corresponding numbers in parenthesis are found in the text below.

The authors argue that generally, physical measures should be preferred due to the "one hoss shay" depreciation character of pipelines (which would imply ordinary accounting depreciation might overstate real costs).

It is argued that no other environmental factors than the ones automatically included in the model (for instance density as throughput over length) need to be included as New Zealand and Australia are sufficiently similar.

As there are measurement problems with the direct measure for capacity (1) an alternate model (2) is specified where asset values measure capacity.

Results indicate that Australian companies consistently outperform New Zealand companies**.** Moreover, an informal inspection of the results shows that "straight line" networks performed better than "radial systems". This might warrant the inclusion of a proxy for the "complexity" of the network.

The study might be criticized on the following grounds. First, results might be less robust due to the small sample size, which also prohibits the use of alternative techniques like DEA and SFA to check robustness. Secondly, the use of a Malmquist index relies on input and output variable weights, which cannot be measured directly here (i.e. as prices or revenue shares). Last, though multilateral Malmquist indices allow cross-section comparison the method is not used widely.

11.2. Granderson (2000)

This paper is an assessment of open-access transportation program for US gas transmission utilities.[26] The program, which provides for non-discriminatory access to pipelines, intends to increase competition. By analyzing firm-level cost efficiency some comments on the relative success of the program are made.

The sample consist of 20 US interstate pipeline companies for the years 1977 to 1987 (75 percent of total output of major interstate companies). Table 38 below lists the input and output variables used.

A frontier cost function (SFA) is estimated and yearly average as well as individual efficiency scores are calculated. By restricting the original (regulated) cost function an unregulated cost function is derived as well.

[26] Historically US gas transmission utilities sold gas they owned to distribution companies and had to hold a license for each customer. In the mid 1980's utilities could obtain blanket licenses for a class of customers on the condition that they would not discriminate. This led to an increase in the amount of gas purely transported for other. In 1992, the regulator mandated the actual unbundling of sales, transportation and storage services. Thus today in the US, gas transmission services are provided on an unbundled, non-discriminatory basis.

Table 38: Variables used in Granderson (2000)[27]

Outputs		Inputs	
Description	**Measurement**	**Description**	**Measurement**
compressor station fuel[28] (a proxy for capacity)	vol.	labour*	# employees for transmission plant
		labour*	total expenditure
		fuel	vol.
		fuel*	total expenditure (price = exp./vol.)
		compressor station capital*	horsepower
		pipeline capital*	pipe steel given by a function of the pipe diameter and length.

Besides calculating implied labour and fuel prices, physical capital prices are calculated by dividing total expenditure by the number of physical units. Capital unit user prices are calculated for mains and compressors[29].

The results indicate that for the sample period industry transmission costs decreased by 0.3 percent, whereas at the same time individual efficiency scores diverged. Last, Granderson finds that rate-of-return regulation leads to higher production costs compared to no regulation.

11.3. Kim et al. (1999)

This study benchmarks gas utilities from several countries including the US, Europe, and Korea. The unbalanced sample contains 68 observations for 9 transmission companies and the years 1987-1995.[30]

Table 39 below lists the variables used. The only output is the volume of gas supplied. Note that the capital expense input contains maintenance expenditure.

[27] * indicates that (unit) factor prices/cost are also used.
[28] According to Granderson an ideal output measure would be the sum product of shipment volumes and the distance shipped.
[29] Here the user price of capital is calculated from information on the depreciation rate, O&M expense, overhead and the tax rate. To compute the O&M contribution, transmission labour and compressor station fuel expenses are subtracted from total transmission O&M. The remainder is allocated to compressor and pipeline capital according to some ratio. The resulting amounts are divided by the total value of the respective capital stocks to obtain the user price. The overhead component is based on an allocation of non-labour overhead to transmission plant. Alternatively, Sickles and Streitwieser (1991) start from total transmission revenue (therefore disregarding depreciation and tax) to apply a similar break-down.
[30] For the DEA the sample is restricted to 45 observation in order to obtain a balanced sample.

The benchmarking methods used among others are Tornqvist index and DEA (separating allocative, scale and managerial efficiency). The use of several methods allows the testing robustness.

Table 39: Variables used in Kim et al 1999[31]

Outputs		Inputs	
Description	**Measurement**	**Description**	**Measurement**
gas supplied	vol.	labour*	average full-time equivalent numbers
		administrative costs*	total cost less labour and capital expense[32]
		capex*	Cost of acquiring and maintaining capital stock (includes maintenance cost, depreciation, non-income tax, interest, insurance)

The results show that for both methods Ruhrgas (Germany) and Transwestern Pipeline Co. (US) are the best performers in terms of average productivity across years. Thus, indicating that US companies are potentially a good benchmark for European companies. Another finding is that the sources of inefficiency differ across companies (for instance scale vs. allocative inefficiency).

11.4. Sickles and Streitwieser (1991)

This study analyses how the deregulation of wellhead prices in the US affected the productivity of the US transmission industry. As the deregulation of wellhead prices led to higher prices and lower demand the hypotheses is that efficiency of transmission companies deteriorated.

The sample consists of a panel of 14 US interstate pipeline companies for the years 1977 to 1985 resulting in 126 observations. Table 40 summarizes the input and output variables used.

The firm level benchmarking techniques used are SFA and DEA.

[31] * indicates that (unit) factor prices/cost are also used.
[32] Administration cost is obtained by subtracting labor and capital cost from total cost. To calculate unit costs. administration cost is regressed against pipeline length and labor input. Last, the number of administration "units" is calculated by dividing total cost by unit cost.

Table 40: Variables used in Sickles and Streitweiser (1991)[33]

Outputs		Inputs	
Description	**Measurement**	**Description**	**Measurement**
deliveries * miles transported	delivery vol. * average miles transported	labour*	# employees for transmission plant
		labour*	total expenditure
		fuel*	vol.
		fuel*	total expenditure (price = exp./vol.)
		compressor station capital*	horsepower
		transmission pipeline capital*	amount of pipe steel given by a function of the pipe diameter and length.[34]
		capacity utilization	average daily to peak day deliveries

Confirming the above hypotheses firm efficiency fell during the sample period. The estimated coefficients for compressors and capacity are all significant at standard levels.

Cross-price elasticities indicate that fuel and compressor services are complements. The same is true for labour in respect to fuel and pipeline services. All other input pairs are substitutes.

The respective minimum and maximum yearly average SFA scores are 69.7 percent and 77.7 percent. For DEA these are 78.1 percent and 86 percent respectively. The authors report that the DEA scores become more similar to the SFA scores when the averages are weighted by the firm's output shares.

11.5. IPART (1999)

IPART uses a sample of US and Australian gas distribution companies to benchmark one Australian company (AGLGN). The study enlarges a small domestic sample using US data. The sample contains 59 observations for 1999.

The benchmarking techniques used are DEA, SFA and COLS. Table 41 lists the variables of the preferred models. As a proxy for the quality of service IPART measures "**U**n**A**ccounted for **G**as", which is the difference between gas supply and the sum of deliveries.

[33] * indicates that (unit) factor prices/costs are also used. For the actual calculation see footnote 29.
[34] Where the amount of steel equals 0.382*(diameter)^2*length.

Prior to the actual benchmarking a regression analysis is performed for the identification of cost drivers (outputs). Potential cost drivers include: customer numbers, deliveries, length of mains, climate (heating degree days), age of network (percentage of asset life expired). The statistical results indicate that length of mains and climate are not significant cost drivers. All other variable are. The authors stress however, that these statistical relationships are not intended to determine input-output relationships. Rather they determine the relative weight given to certain cost drivers (note that length of mains is included in the preferred models in spite of its insignificant in the regression analysis).

The company to be assessed (AGLGN) proposes the following environmental variables: climate, soil type, topography, mains material, age of mains, degree of urbanisation, location of industrial loads within the network. Due to data limitations only climate and age of network were taken into account. These environmental factors are accounted for in two ways. First, they are included into the DEA model in the same way as scale restrictions (thus as non-controllable variables). Secondly DEA scores are regressed against the environmental variables in a second stage Tobit regression.

Table 41: Variables used in IPART (1999)

Outputs		Inputs		Non-discretionary	
Description	**Measurement**	**Description**	**Measurement**	**Description**	**Measurement**
residential customers (1)	#	length of mains (1,2)	km	climate	HDD
other customers (1)	#	O&M (1,2)	$	age of network	% of asset life expired
total customer (2)	#				
delivery (1,2)	volume				
reciprocal of UAG (2)	% of total throughput				

Note: the numbers in parenthesis indicate models in which the respective variable is included. Corresponding numbers in parenthesis are found in the text below. HDD = Heating Degree Days.

The preferred models use the DEA technique and one exclude **(1)** and one includes unaccounted for gas **(2)**. The sensitivity of the preferred models is tested by applying several variations. These include the use of a different measure for O&M costs for AGLGN (the company claimed that its actual costs deviated from its standardized costs), removing the original peers, treating length of mains as an environmental variable, and using different benchmarking methods (COLS, SFA).

These sensitivity tests show that the choice of techniques does not "unduly" influence the results. Nor does the inclusion of the two environmental variables in these various models change the respective results significantly. Generally, though absolute results differ between the various models "general patterns of efficiency" and rankings are broadly consistent.

11.6. Pacific Economics Group (2004)

In this study two New Zealand gas distribution companies are benchmarked against a set of US companies. The sample contains 240 observations (40 firms, 1997-2002).

Table 42 lists the variables used in this study. Note that as a cost function is estimated the sum of O&M expenses and administrative and general expenses (on the input side) constitute the dependent variable of the regression.

Various adjustments are made for both the US and the NZ data to align both. For instance New Zealand capital stock measures are based on the optimized deprival value (ODV) whereas US data is based on historical cost.

Table 42: Variables used in PEG (2004)[35]

Outputs		Inputs	
Description	**Measurement**	**Description**	**Measurement**
retail customers	#	O&M expense	$ (corrected for sales activities)
deliveries	vol.	share of administrative & general expenses	$
		labour*	#
		capital stock*	net plant values in base year plus additions
		length of pipe	km
		pipe material	percentage of main not made of cast iron

A Translog cost function is estimated, which implies that companies are benchmarked against average and not best performance. Unlike in IPART (1999) the regression controls for wages. The authors defend the use of their method on three grounds. First, they state that the method does not require choosing peers explicitly where there are not sufficient observations in one country. Secondly, as the production function is based on economic theory the method cannot be accused of a "black box" approach. Last, like all parametric approaches it allows the specification of a confidence interval around a certain efficiency score.

New Zealand companies compare favourably as their *actual* costs are lower than *predicted* cost.

35 *indicates that (unit) factor prices/costs are also used.

A sensitivity test is performed for changes in the opportunity cost of capital and the construction price index used for the calculation of the capital price. The results are not sensitive to these changes.

11.7. Kittelsen (1993)

The author shows how with the help of econometric techniques the decision to include or exclude variables in non-parametric models can be "aided by the data". A sample of 172 Norwegian electricity distribution utilities for 1989 is used.

The approach allows choosing a set of significant variables for non-parametric benchmarking techniques. The intuitive basis for this approach is that in a DEA model efficiency scores will never decrease when a variable is added or disaggregated. This implies that the inclusion of more variables in itself decreases the risk of erring to the regulated companies' disadvantage.

Table 43 list the variables used. In addition the following environmental variables are included: a distance index, which includes both distance and travelling difficulties in one variable, a corrosion index, derived from depreciation and multiplied by the distance index to get an index that covers the necessary length of lines that are exposed to corrosion, and a climatic index, multiplied by the number of customers to get an index of the need for peak capacity compared to average capacity.

Table 43: Variables used in Kittelsen (1993)

Outputs		Inputs	
Description	**Measurement**	**Description**	**Measurement**
maximum power	kW	labour	hours
energy delivered to other electricity utilities and energy intensive industry	MWh	energy loss	MWh (truncated 5-year moving average, i.e. for 1989: (loss87+2*loss88+3*loss89)/6)
energy delivered to other industry and commerce	MWh	transformers	number and capacity of local transformers ($)
energy delivered to others (households and agriculture)	MWh	lines	voltage level and length ($).
customers	#	goods and services	$

The starting model contains the following variables: energy delivered, labour, energy loss, capital and goods and services.

After the stepwise inclusion of number of customers and the distance index the remaining variables show insignificant test statistics and are therefore not included.

The approach allows to chose significant variables to include in a non-parametric benchmarking model when the sample size is rather large (the authors suggest >100) and the models are nested.

11.8. Estache et al. (2004)

This paper benchmarks the efficiency of 84 South American electricity distributors between 1994 and 2001.

The techniques used are SFA and DEA. The focus of the paper is on tests of consistency. Table 44 lists the variables used.

Table 44: Variables used in Estache et al (2004)

Outputs		Inputs		Non-discretionary	
Description	**Measurement**	**Description**	**Measurement**	**Description**	**Measurement**
final customers	#	labour	# employees	residential sales share	res. sales/total
energy supplied	GWh	transformer capacity	MVA	GDP	per capita
service area	square kilometres	length of network	kilometres of distribution		

For each technique two functions are proposed: an input distance and an input requirement function. The latter takes only one (endogenous) input into account. In total there are six specifications: a SFA model with an input distance function (1), a SFA model with an input requirement function (2), DEAs with input distance functions for VRS (3) and CRS (4) respectively, DEAs with input requirements functions for VRS (5) and CRS (6) respectively. All runs use the same set of variables.

Next, the authors test the consistency of the various approaches. In particular the results should be consistent "in their efficiency levels, rankings, identification of the best and worst result and they should be consistent over time”. Consistency over time establishes *external* consistency (or consistency with “reality”), the other scores establish *internal* consistency.

Both a statistical test and a correlation test are used to check the consistency of the 6 models, where the null is that the approaches generate the same distribution of efficiency scores.

The only pairs that are consistent are formed by the DEA models for a given scale assumption.

Though according to the statistical test the distributions of efficiency scores are not identical the evidence from the correlation test is mixed. Consistency of upper and lower quartiles is rather strong for the various DEA approaches but weak for parametric vs. non-parametric approaches.

Correlations across years for each model, however, are high and positive for all available lags.