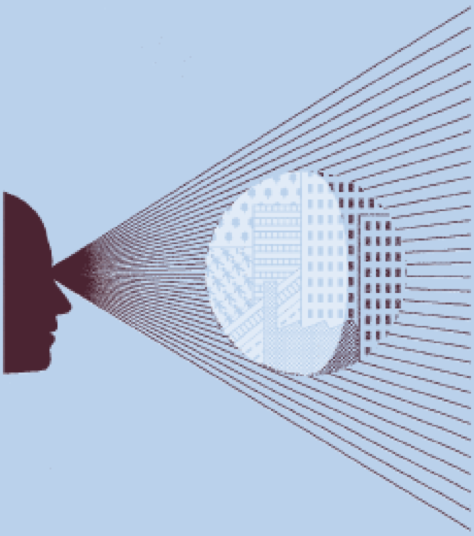
Oxera

How can the NMa assess the efficiency of GTS?

Prepared for the NMa

June 2012

Dossier number 104078



Executive summary

As part of its regulatory duties, the NMa regulates the tariffs of the national gas transmission company, GTS. When preparing for a new price control period, the NMa takes a ‘method decision’, setting out how it intends to regulate the tariffs of GTS, and including an assessment of the company’s cost efficiency. As the NMa’s current regulatory method bases tariffs on GTS’s actual costs, it wishes to understand whether the current costs are efficiently incurred and how they will evolve over time. With this objective in mind, the NMa has commissioned Oxera to consider alternative ways to assess GTS’s cost efficiency.

The aim of this report is twofold: to provide the NMa with an overview on approaches that could be employed to assess the relative efficiency of GTS, and to evaluate each of these approaches against a set of assessment criteria.

In theory, the potential for total efficiency improvement is made up of two components:

– **catch-up**, which measures whether the assessed company’s present cost level differs from current best practice, and, if so, by how much. Catch-up can be based on estimates of ‘relative’ or ‘static’ efficiency;

– **frontier shift**, which provides an estimate of the likely productivity improvements that the assessed company can make in the future, above and beyond any cost reductions owing to the company improving its static efficiency, usually by adopting new technologies and working practices. Frontier shift can be based on estimates of ‘dynamic’ efficiency and is set for every company in the industry. The purpose of the frontier-shift target is to encourage companies in the industry to improve their efficiency in accordance with technological improvements.

Regulators tend to be interested in both elements: catch-up efficiency estimates are generally used to inform the extent to which the assessed company’s costs need to be reduced in order to bring the company into line with current best practice, while frontier-shift estimates represent the savings that could become available in the time between regulatory reviews due to general productivity improvements.

Assessment criteria

To assess the different approaches, this report uses general criteria related to the method itself, and specific criteria related to the applicability of the method to GTS.

The general criteria can be summarised as follows.

– **Complexity/transparency**: the approach adopted by the regulator would need to be clear from the outset, and should enable a transparent monitoring framework to be established.

– **Reliability**: the output of the performance assessment must be regarded as reliable and robust by both the regulated company and any relevant third parties.

– **Suitability for catch-up and/or frontier-shift efficiency**: both catch-up and frontier-shift efficiency need to be assessed for the regulation of GTS. As a result, it is important to ascertain whether the methods assessed can allow a distinction to be made between catch-up and frontier-shift efficiency estimates.

The GTS applicability criteria can be summarised as follows.

– **Data availability**: the adopted approach needs to be able to be implemented using mature benchmarking data[1] that is either already available or can feasibly be gathered by the regulator—for example, by adjusting data on GTS such that it can be included within an existing dataset.

– **Integration**: different approaches produce different types of outcome. This criterion assesses whether it would be possible to integrate the outcome of the approach into the regulation of GTS. For instance, by seeking to understand whether the approach produces a static and dynamic efficiency estimate that can be directly applied to GTS's costs.

– **Implementation time**: this focuses on how long it might take to set up the method for assessing efficiency. A critical factor for assessing this criterion is the availability and maturity of benchmarking data.

– **Impact on other aspects of the regulation of GTS**: this relates to how the application of the various approaches is likely to affect GTS in the context of the current regulatory regime which is based on total expenditure (TOTEX).

The approaches discussed in this report adopt a high-level **top-down** or a **bottom-up** perspective

Top-down approaches

Top-down comparative efficiency modelling involves company- or functional-level comparisons between companies, business units or other economic aggregates. Where there are few companies or only one regulated company, top-down approaches might be less feasible, although international comparisons might be possible. However, in such circumstances, consistency of comparators could become more problematic; for example, data and operational differences may affect the ability to make like-for-like comparisons.

Several approaches are classified as top-down, and this report examines the following ones.

– **Frontier-based approaches**, which attempt to estimate a minimum cost frontier for the industry. These approaches could use econometric analysis, as in the case of corrected ordinary least squares and stochastic frontier analysis; or mathematical optimisation, as in the case of data envelopment analysis. Within this general category, there are many approaches, and both regulators and academics have used these when information on direct comparators has been available. Frontier-based approaches can be used to assess operating expenditure (OPEX) or TOTEX, and can derive estimates for both catch-up and frontier shift separately, provided that comparable data of sufficient quality is available.

– **Unit cost and real unit operating expenditure analysis**—unit cost or single factor productivity comparisons can be used to assess the regulated company's efficiency. Depending on the data available, such top-down unit cost comparisons can in general be used to analyse unit cost levels to estimate either catch-up or unit cost trends to analyse the total scope for efficiency saving which includes both catch-up and frontier-shift. The main difference between unit costs and RUOE approaches is whether they employ direct comparators in the form of top-down unit costs, or indirect comparators, referred to in this report as 'RUOE analysis'. Top-down unit costs are usually employed when the data does not allow for a more thorough frontier-based analysis, owing to issues of data comparability or simply the lack of enough comparators. RUOE analysis also relies on simple top-down unit costs, although the set of comparators is usually

[1] Mature data entails data that is already well-established and collated on a consistent basis, preferably audited, and has been used for comparative purposes such that there is already some confidence in its comparability.

broader and includes companies that are in similar industries (usually other regulated network utilities), rather than limiting the comparator set to companies in the same industry (for example, other gas transmission companies).

– **Growth accounting-based total factor productivity (TFP) analysis**, which provides a benchmark based on the overall productivity performance of a number of sectors of the economy that undertake activities deemed to be comparable with those undertaken by the assessed company. As such, this approach provides an estimate of the potential for productivity growth, which can be applied to TOTEX. The majority of regulators surveyed for this report have used this approach to inform their view on the likely frontier shift.

Bottom-up approaches

While top-down efficiency assessments use high-level comparisons, bottom-up assessments tend to be based on detailed information from the assessed company, including business plans and management accounting information. The assessments are built up by examining individual cost elements on a case-by-case basis. All the relevant cost reductions are then aggregated to provide an overall cost-reduction target. While top-down approaches attempt to make comparisons more like-for-like by including various cost drivers at the modelling stage, bottom-up approaches do so by undertaking comparisons at the business process level. This is because individual processes are likely to be similar across a wider range of companies—eg, the human resources (HR) processes in one company are likely to resemble the HR processes in another.

The bottom-up approaches examined in this report are as follows.

– **Process benchmarking**—this involves disaggregating the company into processes, where a process is defined as a collection of activities with identifiable inputs and outputs. These processes are then compared with other similar processes using internal or external benchmarks. Comparisons are undertaken based on unit costs, key performance indicators and simple productivity measures at a detailed cost line or functional level; for example, comparing HR, IT, finance or property functions within overheads. This approach provides an estimate of catch-up efficiency for the assessed functions.

– **Long-run incremental cost models** are based on the notion that a fully efficient company would price its products according to the long-run incremental cost (LRIC) of those products. With a view to estimating this cost, LRIC models have been used extensively to calculate wholesale access charges and assess cost-reflective pricing. Although there do not appear to be any examples of such models being used for efficiency assessment, the regulator could nevertheless adopt this approach as the basis for setting future prices or revenue since, according to theory, LRIC models aim to reflect the costs that a company would have incurred if it were operating in a competitive environment. The outputs from the LRIC model can be used to assess TOTEX. LRIC models cannot directly estimate the scope for frontier shift; rather, the rate of frontier shift is a required *input* to the model, so that it can properly estimate the long-run incremental costs.

– **Reference models**—this benchmarking approach is based on comparisons with a hypothetical efficient company ‘created’ through the use of a reference model. To create this hypothetical company, the model uses mathematical optimisation and externally sourced capital expenditure (CAPEX) unit costs either to redesign the network or to suggest improvements to the current structure. The model could be extended to assess TOTEX, but at the cost of increased complexity and potentially reduced accuracy. As with the LRIC model, a reference model can include an element of frontier shift, but this would need to be derived using a different method.

– **Comparing the unit CAPEX** of discrete, well-specified capital projects. Although similar in nature to the RUOE analysis, rather than taking a top-down view of the company, this approach relies on more disaggregated information: the unit costs of assets and a standardised set of activities relating to the maintenance and/or replacement of such assets. To evaluate these activities and/or assets, the analysis may involve several professional disciplines as varied as quantity surveying, contract design, engineering, and econometrics. CAPEX unit costs can be used to assess the catch-up efficiency in CAPEX.

Regulators use bottom-up approaches in particular when there are relatively few organisations against which the performance of the company in question can be compared. These approaches can be used where the regulated company is unique, either because it is the only company of its kind in its sector, or because its characteristics, such as topography or customers, are atypical.

Relative performance of the approaches against the criteria

It is difficult to rank with any great precision the approaches according to the selected criteria, mainly because of uncertainties about the required underlying data and issues relating to the implementation time. With regard to the latter, the availability and maturity of data are likely to be key constraints; from the literature review, it appears that the benchmarking of transmission companies focuses more on electricity transmission companies, predominantly using US data from the Federal Energy Regulatory Commission.[2] In terms of benchmarking of European gas transmission companies, there are relatively few comprehensive benchmarking studies (eg, the 2006 report by the Electricity Policy Research Group for the Council of European Energy Regulators,[3] which included data on only four European gas transmission service operators).

It is, however, possible to rank the approaches using a simpler, relative grading system with the inclusion of the necessary caveats. Such a ranking is provided in the table below, with each approach ranked against the criteria, from A, the highest ranking, to D, the lowest ranking. It should be stressed that these grades are relative; an approach ranked A for cost requirements does not mean that it is three times less costly to implement than an approach ranked C. Each approach has advantages and disadvantages, and the decision on which is likely to be most relevant to a particular assessment depends to a large extent on the circumstances and the type and availability of relevant information.

[2] For a review of data available from FERC on the US transmission companies, see, for example, Jamasb, T., Pollitt, M.G. and Triebs, T.P. (2008), 'Productivity and Efficiency of US Gas Transmission Companies: A European Regulatory Perspective', working paper. For data and annual reports on the electricity and gas transmission companies, see also www.ferc.gov.
[3] Electricity Policy Research Group (2006), 'International Benchmarking and Regulation of European Gas Transmission Utilities', Prepared for The Council of European Energy Regulators (CEER), Final report.

Relative rankings of the assessed approaches

	Top-down			Bottom-up			
	Frontier-based approaches	**TFP analysis**	**RUOE analysis**	**Process benchmarking**	**LRIC model**	**Reference model**	**CAPEX unit cost analysis**
General criteria							
Complexity and transparency	**A–C**, depending on the chosen approach	**A**	**A**, assuming that there is no issue about confidential information	**A–C**, depending on the overall transparency of external benchmarks	**D**	**C–D**, depending on the level of complexity	**A–B**, assuming that there is no issue about confidential information
Reliability	**A–C**, depending on the chosen approach and the type of comparator (for example, European only or European and US comparators)	**D**	**C**	**A–B**, for support functions. For gas transmission-specific functions, there are likely to be issues with the availability of reliable external benchmarks	**Unclear**, owing to the required assumptions and approach adopted	**B–C**, depending on the level of complexity	**B–C**, depending on the level of complexity and the quality of expert advice sought
Suitability for static and/or dynamic efficiency	**A**, if consistent data over time is available, it is possible to estimate catch-up and dynamic efficiency separately	**C**, measures total scope for efficiency saving although secondary sources could be used to decompose this measure into catch-up and frontier-shift estimates if required	**A–B**, if consistent data over time is available, unit cost *trends* measure overall productivity growth, thereby including catch-up and frontier shift, while unit cost *levels* could be used to provide an estimate of relative efficiency and thus catch-up efficiency	**C**, just catch-up; secondary source is needed to estimate frontier shift	**C**, just catch-up; secondary source is needed to estimate frontier shift	**C**, just catch-up; secondary source is needed to estimate frontier shift	**C**, just catch-up estimate on the current CAPEX of GTS; secondary source is needed to estimate catch-up on OPEX and frontier shift

	Top-down			Bottom-up			
	Frontier-based approaches	TFP analysis	RUOE analysis	Process benchmarking	LRIC model	Reference model	CAPEX unit cost analysis
GTS applicability criteria							
Data availability	**C–D**, data availability from other European gas transmission companies and data consistency are key constraints, although a high-level comparison may be possible with non-European gas (eg, US) transmission companies and transmission companies in Germany	**A**	**A–D**, comparison using direct comparators (other European gas transmission companies) could be limited owing to data constraints, although comparison with indirect comparators could be possible	**C–D**, for gas transmission-specific functions as there are likely to be issues of confidentiality and with the availability of reliable external benchmarks **A**, for support functions	**A–B**, highly dependent on the type of LRIC model used, accessibility of the gas transmission company's data and availability of information from direct comparators, which would be required to inform aspects of the LRIC that require judgement	**A–B**, highly dependent on accessibility of the gas transmission company's data	**A–B**, highly dependent on accessibility of the gas transmission company's data
Integration	**A–B**, depending on the chosen approach, the type of comparator (eg, European only, or European and US comparators), and the quality and quantity of the dataset (eg, data over time required to estimate frontier shift)	**B–C**, the tariffs could be adjusted for the overall productivity growth estimated by the approach	**A–C**, by definition RUOE focuses on OPEX, but it could be extended to cover TOTEX. Otherwise CAPEX may need to be assessed and reimbursed separately. While RUOE unit cost trends measure overall productivity growth, thereby including catch-up and frontier shift, RUOE unit cost levels could be used to provide an estimate of relative efficiency and thus catch-up efficiency	**B–C**, dependent on the scope of the cost base and the availability of external benchmarks. Would require separate assessment of the frontier-shift adjustment to adjust the tariffs	**Unclear**, owing to absence of precedent	**C–D**, can provide a catch-up estimate on the current CAPEX of GTS, but it could be extended to cover TOTEX. If limited to CAPEX, a separate assessment would be required of OPEX catch-up efficiency and the frontier-shift adjustment to adjust the tariffs	**C–D**, can provide a catch-up estimate on the current CAPEX of GTS. Would require a separate assessment of OPEX catch-up efficiency and the frontier-shift adjustment to adjust the tariffs
Implementation time	**B–D**, depending on the approach. Intrinsically linked to data availability (see above).	**A**	**A**	**B–C**	**C–D**, depending on the level of complexity	**C–D**, depending on the level of complexity	**C–D**, depending on the level of complexity

	Top-down			Bottom-up			
	Frontier-based approaches	TFP analysis	RUOE analysis	Process benchmarking	LRIC model	Reference model	CAPEX unit cost analysis
GTS applicability criteria							
Impact on other aspects of the regulation for GTS	**A–C**, the NMa's current method of GTS regulation bases tariffs on actual costs. The tariffs could be adjusted to remove the estimated static inefficiency of GTS identified by the method(s) and applying a frontier-shift adjustment over the period	**A–B**, the tariffs could be adjusted for the overall productivity growth estimated by the approach. However, limitations remain in terms of decomposing the overall measure into catch-up efficiency and frontier-shift improvements, although secondary sources could be used to facilitate this, if required	**A–C**, by definition RUOE focuses on OPEX, but it could be extended to cover TOTEX. Otherwise CAPEX may need to be assessed and reimbursed separately. While RUOE unit cost trends measure overall productivity growth, thereby including catch-up and frontier shift, RUOE unit cost levels could be used to provide an estimate of catch-up efficiency	**C–D**, can provide only estimates of catch-up efficiency which can be used to adjust the tariffs. However, additional adjustment in the form of frontier shift may need to be estimated separately	**C–D**, the model cannot directly estimate the scope for frontier shift; as such, additional adjustment in the form of frontier shift may need to be estimated separately. The efficient costs estimated by the model can be used to inform the tariff adjustment required on the current cost of GTS	**C–D**, depending on the complexity of the model, the reference model is likely to provide estimates of the need for additional capital investment and the likely cost of such investment; however, it could be extended to cover TOTEX. If limited to CAPEX, a separate assessment is likely to be required of OPEX static efficiency and the frontier-shift adjustment to adjust the tariffs	**D**, depending on the level of complexity. Separate assessment would be required of OPEX catch-up efficiency and the frontier-shift adjustment to adjust the tariffs

Source: Oxera analysis, based on input from the NMa.

Contents

List of tables

List of figures

List of boxes

1 Determining the potential for efficiency improvements

As part of its regulatory duties, the NMa regulates the tariffs of the national gas transmission company, GTS. When preparing for a new price control period, the NMa takes a 'method decision', setting out how it intends to regulate the tariffs of GTS, and including an assessment of the company's cost efficiency. As the NMa's current regulatory method bases tariffs on GTS's actual costs, it wishes to understand whether the current costs are efficiently incurred and how they will evolve over time. With this objective in mind, the NMa has commissioned Oxera to consider alternative ways to assess GTS's cost efficiency.

This section provides the NMa with an overview on approaches that could be employed to assess the relative efficiency of GTS in a performance assessment exercise.

The overall aim of a performance assessment exercise is to establish the scope for efficiency improvements that a company can achieve going forward. Theoretically, two components make up the potential for total efficiency improvements:

– **catch-up or static efficiency improvements**, which provide an estimate of the likely rate of improvement in catching up to current best practice. Catch-up can be based on estimates of 'relative' or 'static' efficiency;

– **frontier shift or dynamic efficiency improvements**, which provide an estimate of the likely productivity improvements that the assessed company can make in the future, above and beyond any cost reductions owing to the company improving its static efficiency, usually by adopting new technologies and working practices. The frontier-shift target is set for every company in the industry and is applied to encourage companies in the industry to improve their efficiency in accordance with technological improvements.

Some assessment approaches allow for both catch-up and frontier shift to be estimated within the same methodological framework; alternatively, a mixture of approaches can be used to estimate the two components separately. Where the assessment aims to estimate static efficiency, cross-sectional data from only one year is necessary, although some approaches can greatly benefit from using panel data covering a longer time period. Panel data allows the scope of historical frontier shift to be estimated and overall productivity growth to be measured, which, in a regulatory setting, is usually defined as the sum of catch-up and frontier shift.

The main aim of an efficiency analysis geared towards estimating the static component is to understand how inefficient a company is relative to best practice, and thus its potential to reduce its cost base to a more efficient level by catching up to the current frontier.

The first consideration is how to identify the efficiency frontier against which the regulated company is to be compared. All approaches examined in this report rely on a set of comparators to estimate the efficiency frontier, such as discrete business units, functions, regions, companies and/or aggregate industries. This set of comparators would ideally be made up of independent companies or business units that consume similar inputs to achieve similar outcomes. However, as it is not always possible to construct such a comparator set, other similar, or not so similar, economic units have previously been used to understand what cost reductions might be possible for the assessed company. These include:

– internal comparisons of discrete business units belonging to the assessed company, and/or regional comparisons if the assessed company has a regional structure where different business units undertake similar activities in each of the regions;

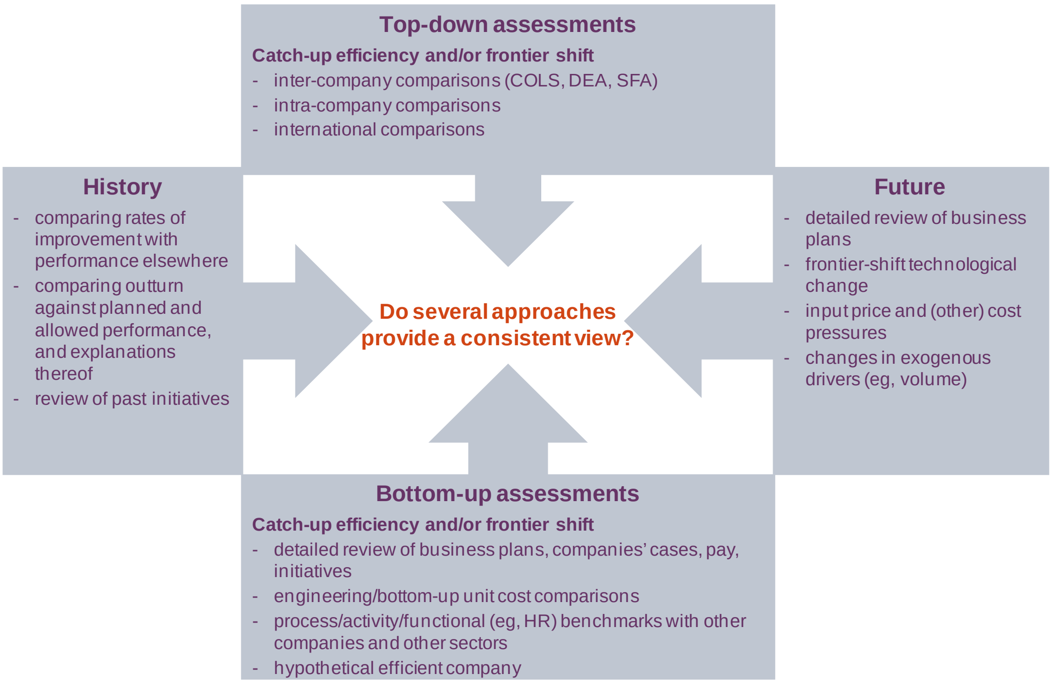
– discrete business units that undertake some similar functions belonging to other, sometimes dissimilar, companies—for example, bottom-up benchmarking of support functions such as accounting or HR;
– more broad economic aggregates, such as whole sectors of the economy that undertake broadly similar activities.

In some cases, the comparisons could be based on hypothetical businesses, whereby regulators have employed economic or engineering models to simulate the activities of the assessed company in order to form a view of the general level of efficiency displayed. In general, the decision regarding the appropriate set of comparators is dependent on a wide range of factors, including the regulatory regime, the industry structure and data availability.

In order to assess relative or static efficiency, direct comparators using a consistent dataset would be necessary, which in turn provide information to establish a required rate of catch-up efficiency. However, some approaches can provide an estimate of historical overall efficiency improvements as a benchmark rate of efficiency improvement for the regulated company of interest without recourse to a set of direct comparators. These approaches include 'growth accounting' TFP and trends in real unit costs.

The extent of possible efficiency improvements can be established from a more high-level **top-down** perspective, or a detailed **bottom-up** perspective.[4] In either case, a number of approaches can be considered. In addition, elements can be examined based on **historical** information or by looking at **future** forecasts. The regulatory approaches to determining potential efficiency improvements are summarised in Figure 1.1.

Figure 1.1 Determining potential for efficiency improvements



Note: COLS, corrected ordinary least squares; DEA, data envelopment analysis; SFA, stochastic frontier analysis.
Source: Oxera.

A number of top-down and bottom-up approaches are discussed in more detail in sections 3 and 4, respectively.

[4] In distinguishing between top-down and bottom-up approaches, there are also differences in principle. A top-down approach models a 'decision-making unit' (ie, a self-contained unit that has some degree of management autonomy for which inputs and outputs can be readily identified and ascribed). It analyses the inputs and outputs, and any external factors, in order to estimate the efficiency of transforming the inputs into outputs, without seeking to understand the details of the processes. In contrast, a bottom-up approach considers the workings of a process, including its cause-and-effect relationships.

2 Criteria for selecting an appropriate assessment methodology

Undertaking a performance assessment can be complex and includes numerous elements, a central one being the method adopted for the efficiency assessment exercise. Selecting the most appropriate assessment method or combination of methods is critical to the success of the whole exercise. This section examines some criteria that can be used to facilitate this selection process.

Owing to the overall complexity of a performance assessment, many criteria can be put forward to assess each method, ranging from technical considerations to relatively minor qualitative differences. However, the use of a large set of criteria can be counterproductive if the most important criteria are not given enough weight and the whole selection process becomes cumbersome. To limit their number to manageable levels and focus on those that best describe the NMa's objectives in benchmarking GTS, this report concentrates on the general criteria related to the method itself, and on the specific criteria related to the applicability of the method to GTS.

2.1 General criteria

In brief, the principles of 'good regulation'—as highlighted, for example, in the guidance from the UK Better Regulation Executive on Regulatory Impact Assessments—are that regulators should act in a manner that is targeted, accountable, transparent, consistent and proportionate.[5] With these principles in mind, general criteria can be devised to assess the overall suitability of the possible performance assessment approaches, as follows.

2.1.1 Complexity/transparency

In principle, the approach adopted by the regulator would need to be clear from the outset and enable a transparent monitoring framework to be established. Ideally, all relevant parties should be able to understand at least the principles underlying the adopted methodology and, if required, be able to replicate the analysis, assuming that they have the specialist knowledge and skills. In turn, a clear methodology:

– makes it easier to the regulator to explain its chosen approach if required. It also greatly assists in the discussion surrounding the factors that the regulator has chosen to include in the analysis;

– strengthens incentives since there is a clear link between the company's performance and its targets;

– allows the relevant parties to verify the regulator's findings if needed, unless the analysis relies on commercially sensitive data.

On February 22nd 2012, the Dutch court of appeal, College van Beroep voor het Bedrijfsleven, published a decision following a number of appeals from TenneT regarding the regulatory approach taken by the NMa. In one of its judgments, the court ruled that the regulated company need not be granted full access to the data used in the NMa's analysis. This suggests that when commercially sensitive data is used, full dissemination might not be required for the approach to be deemed appropriate

[5] See UK Department for Business, Innovation and Skills (2011), 'Principles of Economic Regulation', April, Chapter 1.

2.1.2 Reliability

Both the regulated company and any interested third parties must regard the output of the performance assessment to be reliable and robust. A robust approach is defined here as one that provides an estimate of performance that is as accurate as necessary under the specific circumstances under which the assessment takes place. For the NMa, for example, the results of the assessment must be accurate enough to be defendable in a court of law.

Robustness itself as a concept can have many dimensions in this setting. An assessment approach is considered able to produce robust results when it:

– is adaptable and requires as few assumptions and/or arbitrary decisions on the part of the regulator as possible;
– can be used to model a wide variety of situations, ideally with little or no modification;
– produces results that are not too sensitive to the assumptions underpinning it;
– can deal with real-world issues that the regulator is likely to face, such as noisy and imperfect data.

In more practical terms, an efficiency estimate is usually considered robust when it can be verified by several different approaches—ie, the critical issue is that the resulting efficiency estimates should not be volatile or too sensitive to the assumptions underpinning it.

As no efficiency approach is guaranteed to be absolutely robust under all possible scenarios, the choice of one or a range of suitable approaches would involve making trade-offs. In other words, the concept of robustness is relative. In addition, the situation faced by the regulator might be such that no approach can produce estimates that are accurate within a desired range. Even in this case, however, undertaking a performance assessment would be valuable in deriving at least a general range of cost reductions, and, more importantly, in identifying the gaps in the data or in the methodology itself so that they can be addressed in a future review.

2.1.3 Suitability for catch-up and/or frontier-shift efficiency

As both catch-up and frontier-shift efficiency need to be assessed for the regulation of GTS, it is important to ascertain whether the methods examined as part of this report are able to provide estimates for both of these.

2.2 Specific criteria related to the applicability of the method to GTS

In addition to the general criteria outlined above, a number of aspects are deemed relevant to assess the efficiency methods. These relate mainly to how the methods can be applied in practice to measure GTS's efficiency, given that the next regulatory period starts in January 2014. As proposed by the NMa, the specific aspects are as follows.

2.2.1 Data availability

The aim of this criterion is to ensure that the adopted approach can be implemented using sufficient and mature benchmarking data that is either already available or can feasibly be gathered by the regulator or its consultants, for example by adjusting data on GTS such that it can be included within an existing dataset. Some methodologies are data-intensive, while others can be run with less data. Understanding the data requirements of each methodology is fundamental so that the NMa can ascertain whether the data required can be collated successfully.

2.2.2 Integration

Different methodologies produce different types of outcome. This criterion assesses whether the outcome of the method can be integrated into the regulation of GTS. For instance, by seeking to understand whether the approach produces a static and dynamic efficiency estimate that can be directly applied to GTS's costs.

2.2.3 Implementation time

This criterion focuses on how long it might take to set up the method for assessing efficiency. This is particularly important for the NMa as there is currently no static efficiency testing method for GTS. With the next regulatory period starting in January 2014, the NMa needs to produce its draft decision in the first quarter or start of the second quarter of 2013, and its final method decision in the third quarter of 2013. Thus, any proposed method needs to be implementable, at least in an intermediate mode, and to produce a preliminary cost-reduction target for GTS before April 2013 so that the method can be revised if warranted and fully applied by the start of the new regulatory period in January 2014.

Although being implementable within this timeframe is a short-term objective, Oxera understands that it is not an absolute requirement for the NMa. For example, there may be a method or methods that the NMa considers most appropriate but which might be implementable only in the longer term. The NMA may therefore choose to develop such a method in the longer term, with the current review possibly a step towards this longer-term objective. Hence, this report considers approaches that may not be implementable in the short term and, where this is the case, provides an indication of how long it might take to develop them.

2.2.4 Impact on other aspects of the regulation of GTS

This relates to how the application of the different methods is likely to affect GTS in the context of the current regulatory regime. If changes are required, a particular method is not necessarily ruled out, although it may affect the implementation time.

First, it is important to consider whether each approach can be implemented without making significant changes to the current method of TOTEX regulation. Second, the NMa has stated that it may no longer need to have an additional reimbursement during the regulatory period for all expansion investments, depending on how the benchmark is applied. This is because other companies in the benchmarking exercise may also have expansion investments and thus GTS may be reimbursed automatically by the application of the benchmark. Critically, the NMa will need to examine carefully what expansion investments are included in the benchmarking exercise and how these compare to GTS's future expansion investments.

3 Top-down approaches

This section presents an overview of the top-down approaches used to assess efficient expenditure by other regulators in different sectors and countries. The aim here is to provide a broad overview of the approaches, rather than examining the more detailed workings of each one. The section provides:

– a brief technical description of the approach;
– some examples of where the approach has been applied, in different countries and sectors with a focus on gas transmission;
– the relative advantages and disadvantages of the approach;
– how it performs according to the assessment criteria listed in the previous section.

The top-down approaches discussed here include:

– data envelopment analysis (DEA);
– corrected ordinary least squares (COLS);
– stochastic frontier analysis (SFA);
– total factor productivity (TFP) estimates;
– unit cost comparisons, using either direct or indirect comparators.

The first three are referred to as **frontier-based approaches**, owing to the way they measure relative efficiency. Such approaches require the existence of direct national/regional, international or internal comparators, whereas TFP analysis, as described in this report, relies on indirect comparators, while benchmarks based on top-down unit costs can be derived from either direct or indirect comparators.

Most of these approaches have been used by regulators to assess electricity or gas transmission companies. Particular examples are as follows.

Frontier-based approaches

– The Task Force on Benchmarking of Transmission Tariffs of the Council of European Energy Regulators (CEER) commissioned a study to develop a framework for benchmarking of European gas transmission companies.[6] Benchmarking was undertaken using three frontier-based techniques: DEA, SFA and COLS.

– E3Grid, a regulatory benchmarking of European electricity transmission companies on behalf of CEER Workstream Incentive-based Regulation and Efficiency benchmarking (WS EFB), was commissioned in 2008. Benchmarking was undertaken using unit cost comparisons and DEA.[7]

– Ofgem, the regulator of the energy sector in Great Britain, had proposed to use COLS and DEA with US-based comparators to assess the efficiency of the UK gas and electricity transmission companies for the current review. However, given data consistency issues with the comparators and following stakeholders' concerns about the robustness of international benchmarking, the regulator focused on disaggregated unit

[6] Electricity Policy Research Group (2006), 'International Benchmarking and Regulation of European Gas Transmission Utilities'.
[7] Sumicsid (2009), 'International Benchmarking of Electricity Transmission System Operators', March.

cost benchmarking.[8] In the case of electricity and gas distribution companies, Ofgem has used this approach to establish efficiency targets.

TFP and similar sectoral-based estimates

– The NMa has considered a similar method in the past to assess the scope for future cost reductions in GTS. Although the focus of this analysis was the productivity of labour and intermediate inputs rather than TFP, the resulting benchmark was based on indirect comparisons with sectors of the Dutch economy using national accounts data.

– As part of the current transmission price control review, Ofgem is considering using TFP analysis to analyse long-term efficiency trends based on the EU KLEMS database.[9] To complement the EU KLEMS data, it has also proposed to use alternative productivity data, for example from the Office of National Statistics (ONS) on sectoral productivity.

Unit cost comparisons

– The NMa has employed a form of top-down unit cost analysis in the past for TenneT.[10] Given concerns with international TOTEX benchmarking, Ofgem has used disaggregated unit cost comparisons alongside trend analysis to assess companies' efficiencies, as in previous reviews.[11] Elsewhere, top-down unit costs have not seen widespread use to date when assessing transmission companies; rather, the cost trends have been used to examine rates of changes in order to provide benchmarks for rates of efficiency improvement. However, bottom-up unit costs have been used extensively in the past to assess capital expenditure (CAPEX) (see section 4).

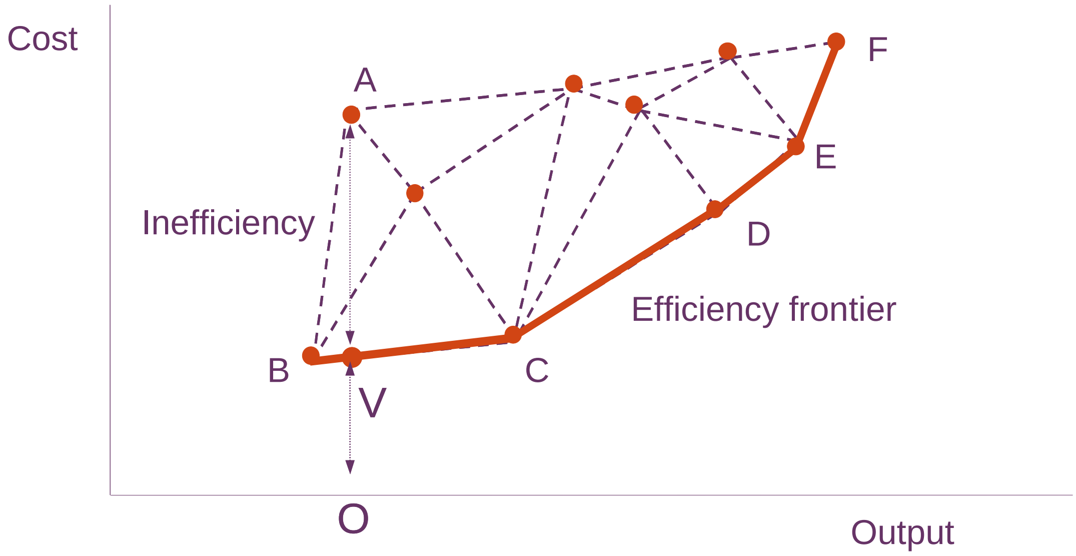
3.1 Data envelopment analysis

A mathematical, non-parametric approach, DEA is one of the most widely used approaches internationally when benchmarking regulated companies. As a frontier-based approach, it measures efficiency by reference to an efficiency frontier, which is constructed as linear combinations of efficient companies—ie, companies that produce the most output at the lowest cost.

In more detail, DEA assumes that two or more companies or decision-making units can be 'combined' to form a composite producer with composite costs and outputs—a 'virtual company'. The actual companies are then compared to these virtual and actual companies. If another actual or a virtual company or their combination achieves the same output as the actual company at a lower cost, the actual company is judged to be inefficient. DEA selects the efficient observations and constructs a frontier from them, ignoring those observations that turn out to be inefficient.

[8] Ofgem (2011), 'Decision on strategy for the next transmission price control - RIIO-T1 Tools for cost assessment', March, paras 4.7–4.8.
[9] Ibid, paras 3.2–3.3.
[10] Sumicsid (2010), 'Benchmarking TenneT EHV/HV', Project STENA, March.
[11] TPA Solutions (2006), 'Transmission Price Control Review 2007-2011—Efficiency Study and Forecast Opex', Final draft report, September.

Figure 3.1 Graphical example of data envelopment analysis



Source: Oxera.

In the example in Figure 3.1, the DEA frontier is given by the line joining points B, C, D, E and F. The efficiency of company A is given by the distance from A to point V. Point V is a 'virtual company', made up of a weighted average of frontier companies B and C, such that V has the same quality as A.[12] Companies B and C are referred to as A's 'peers', with B clearly being given a much higher weighting than C.

If the assessed company lies on the frontier, it would not have a catch-up efficiency target, although it may still have a frontier-shift target. However, if the company is at B or F then, although it can be considered as efficient with respect to the comparator set used, this may be due to B or F having no direct comparators. That is, if an observation is somehow unique—in this case small or large—then DEA may estimate the company to be efficient purely because it has no other comparator against which to compare it.

As with the other frontier-based approaches, DEA requires data on domestic, international or sub-company comparators, which are referred to as 'decision-making units'. The applicability of this approach is therefore dependent on the availability of comparators and data of sufficient quality. Oxera understands that GTS has no regional structure and that the dataset on comparators would therefore need to comprise gas transmission companies in other countries.

3.1.1 Applications in the regulatory setting

DEA has been widely used by regulators in Scandinavia; for example, it has been used to set efficiency targets for electricity distribution companies in Finland.[13] Here, overall costs to consumers in the form of TOTEX, comprising operating costs, depreciation and outage costs were benchmarked. Due to difficulties in applying an efficiency target to straight-line depreciation, and a lack of up-to-date data on outage costs, the efficiency target was applied to operating expenditure (OPEX) only for the 2008–11 price control period. An adjustment to the efficiency target was therefore applied using the ratio of OPEX to TOTEX for each DNO. The scope for industry-wide productivity improvements was estimated using a DEA model and data for the period 1999–2005.

[12] For a more detailed discussion on DEA, see Thanassoulis, E. (2001), *Introduction to the Theory and Application of Data Envelopment Analysis: A Foundation Text with Integrated Software*, Springer.
[13] See Energiemarkkinavirasto, 'Methods of determining the return DSOs during the regulatory period 2008-2011', available at http://www.energiamarkkinavirasto.fi/data.asp?articleid=1699&pgid=133&languageid=752.

The Independent Pricing and Regulatory Tribunal (IPART) of New South Wales has applied the DEA approach in the past, benchmarking eight Australian gas distribution companies against 50 US comparators,[14] using data sourced from local distributors, *The Pipeline & Gas Journal*, the Federal Energy Regulatory Commission (FERC) and Natural Gas LDC[15] databases. IPART's preferred model used operation and maintenance costs and length of mains as inputs, and energy delivered, residential and other customers as outputs. A number of environmental variables (ie, uncontrollable variables) including topography, age of mains, and degree of urbanity were considered in a second stage to test their influence on the DEA-based efficiency estimates. The sample of data was subjected to descriptive statistical summary and outlier analysis to identify any obvious errors and outliers in the data. The results were then cross-checked with alternative frontier-based approaches, SFA and COLS, for consistency. Australian distribution companies were found to be 73% efficient on average when compared against their US counterparts. The authors argued, however, that the results should not be used in a mechanical manner to set efficiency targets for the distributors. This was mainly due to the differences in the operating environments across competitors and observed variations in the quality of the data.

The final report of the study by EPRG, published in 2006,[16] used a sample of 43 US interstate gas transmission companies and four European companies with an initial panel dataset of 328 observations.[17] The cost measures considered in the study included operating and maintenance expenditure; a TOTEX measure comprising of operating and maintenance expenditure and depreciation; another TOTEX measure comprising operating and maintenance expenditure, depreciation and cost of capital, and revenue excluding fuel costs. All the cost measures were adjusted for inflation and purchasing power parity. The cost drivers considered include total annual throughput of gas transmitted (m^3/year), total length of pipelines (km), total amount of compressor horsepower on pipelines (HP), total number of compressor stations, total number of compressor units, capacity and load factor. Variables relating to quality and environmental factors were not considered in the study.[18] Benchmarking was undertaken using three frontier-based techniques: DEA, SFA and COLS. To account for possible outliers in the data, 10% of the most efficient companies were removed before estimating the efficiencies using DEA.

EPRG observed that, while Europe as a whole has sufficient numbers of comparators for benchmarking (78 national and regional gas transmission companies at the time), the lack of readily available and well-documented data was the reason for including US comparators in the study. Investigating the comparability of the data was a key issue raised in the study, along with data standardisation. Although US comparators are naturally expected to be substantially different to their European counterparts, differences within Europe were also quite large. However, it was stressed that achieving comparability among the European firms is easier than achieving comparability between US and European firms, as this most likely requires direct communication with FERC. As a result, the US data was not standardised in the EPRG study and the efficiency estimates were considered to be only indicative of the actual catch-up levels.

In Norway, TOTEX DEA has been used to benchmark domestic electricity distribution companies. Given the number of regional distribution companies (150 at the time), international comparisons were not necessary. The comparison was undertaken using both book and replacement values,[19] in order to account of the age profiles of different grids. The

[14] Independent Pricing and Regulatory Tribunal (1999), 'Benchmarking the Efficiency of Australian Gas Distributors', Research paper Gas99–9, December.
[15] Opri (1998), Natural Gas LDC Database, Boulder.
[16] EPRG (2006), 'International Benchmarking and Regulation of European Gas Transmission Utilities'.
[17] The European operators included in the EPRG study are confidential.
[18] EPRG (2006), op. cit., pp. 25–27.
[19] See Ajodhia, V., Kristiansen, T., Petrov, K. and Scarsi, G. (2005), 'Total cost efficiency analysis for regulatory purposes: statement of the problem and two European case studies', available at http://www.wip.tu-berlin.de/typo3/fileadmin/documents/infraday/2005/papers/petrov_scarsi_kristiansen_adjohia_Total_Costefficiency_analysis_for_regulatry_purposes.pdf

efficiency target was based on the most favourable result for each company. The efficiency estimate was restricted to 70% for the most inefficient companies, even if they were less than 70% efficient, and companies were expected to reduce at least 38% of their inefficiency gap over the four-year price control period, with any residual inefficiency carried into the next price control review. The efficiency target was applied to total costs in the previous year in order to obtain the allowed revenue.

In Germany, the regulator uses DEA along with SFA to assess the TOTEX efficiency of energy networks. Given the number of regional distribution companies (328 electricity and 488 gas), international comparisons are not used. The catch-up target is based on the efficiency score that is most favourable to the operator. Again, the results were not used in a mechanical manner to set efficiency targets for the distributors. A floor was set so that the maximum value that a company's inefficiency could take was 50%, and a company that did not provide data to allow an assessment of its efficiency was set a 50% catch-up target. To further ensure that a company's financial viability was not compromised, companies could submit evidence of any operational or structural factors that were not captured by the efficiency analysis and which might affect their costs. They could also argue for a longer period to achieve efficiency savings than might otherwise be the case.[20]

Recently, the Commission de Régulation de l'Electricité et du Gaz (CREG) commissioned a study to develop efficiency benchmarking models for the gas and electricity distribution companies in Belgium.[21] Given the number of regional distribution companies (25 electricity and 17 gas) and the availability of data over time, international comparisons were not necessary. The study proposed benchmarking the companies using a set of DEA models, with TOTEX as the single input, and total number of connections, total circuit length and total number of transformers as outputs in the case of the electricity distribution companies, and total number of connections, total weighted length of pipelines and total number of pressure stations as outputs in the case of the gas distribution companies.[22] The results and the analysis have not been published.

In 2010 the Finnish Energy Market Authority (EMV) applied a combined DEA and SFA approach (Stochastic Non-smooth Envelopment of Data, StoNED)[23] to the electricity sector in order to set company-specific efficiency targets for the regulatory period 2012–19. The benchmarking model used data collected over a six-year period (2005–10), with total cost (essentially, the sum of OPEX and half of outage costs) as the single input; energy transmitted, length of network and number of customers as outputs; and proportion of underground cables in the total network length as a contextual variable.[24]

3.1.2 Advantages and disadvantages of DEA relative to other frontier-based approaches

The major advantage of all frontier-based approaches is that the efficiency estimates are based on the realised performance observed in other, similar companies, rather than relying on expert knowledge of the industry or regulatory judgement. As such, the approaches are relatively robust and transparent, at least with respect to how the efficiency measures are derived.

The main advantages of DEA relative to the other frontier approaches are that it:

[20] For more on this, see Oxera (2007), 'Taming the Beast? Regulating German electricity networks', *Agenda*, May. Source: Bundeswirtschaftsministerium für Wirtschaft und Technologie (2007), 'Verordnung zum Erlass und zur Änderung von Rechtsvorschriften auf dem Gebiet der Energieregulierung', April 4th.
[21] Sumicsid (2011), 'Development of benchmarking models for distribution system operators in Belgium', Project NEREUS, Final Report, November 30th. See http://www.creg.be/pdf/Opinions/2011/P092011/Benchmarking_models_for_distribution_EN.pdf
[22] Ibid, p. 3.
[23] See Kuosmanen, T. and Kortelainen, M. (2010), 'Stochastic non-smooth envelopment of data: semi-parametric frontier estimation subject to shape constraints', *Journal of Productivity Analysis*, December.
[24] Kuosmanen, T. (2010), 'Cost efficiency analysis of electricity distribution networks: Application of the StoNED method in the Finnish regulatory model', working paper. Categorical, ordinal, interval or ratio scale data that characterise operating conditions and practices are commonly referred to as contextual variables in the productivity literature. See, for example, Banker, R.D. and Natarajan, R. (2008), 'Evaluating contextual variables affecting productivity using data envelopment analysis', *Operations Research*, 56:1, pp. 48–58.

– can accommodate multiple outputs and inputs, including measures of quality and factors outside the management of the company that might influence the overall efficiency of the assessed company, while standard COLS and SFA approaches generally use a single input;
– does not require a cost function to be specified, and therefore avoids many of the issues with cost function specification found in the other frontier-based approaches—ie, compounding the effects of mis-specification of the technology with inefficiency.

On the other hand, DEA is a deterministic approach, in that the deviations from the frontier are attributed to the inefficiency of a firm, and as such cannot directly account for the presence of noise or measurement error in the data. This has led to some applications making ad hoc adjustments to the efficiency estimates, or requiring regulatory judgement on them, in order to correct for the effects of noise. For example, in Finland the regulator applies a 0.84 adjustment to the efficiency target to account for known errors in the data.[25] Alternatively, some extensions to the standard DEA approach are possible.[26]

3.1.3 Assessment of DEA

Table 3.1 summarises how DEA measures against the general and GTS applicability criteria listed in section 2.

Table 3.1 Assessment of DEA against criteria

	Assessment
General criteria	
Complexity and transparency	DEA requires specialist software and could be considered a complex approach on the whole, even though its principles are relatively easy to explain. Extensions to DEA are significantly more complex and are likely to require specialist or bespoke software to implement. This approach is transparent, provided that the underlying models are made publicly available.
Reliability	Regulators and companies have made extensive use of DEA. Provided that the model is specified correctly and enough comparators are available, it is considered highly robust. The situation is somewhat more complicated when the analysis relies on international comparisons, since the normalisation to achieve like-for-like comparisons requires a degree of judgement, which could result in an additional source of measurement error or noise in the data. DEA is also a flexible approach; it can incorporate additional inputs and outputs if these are deemed to be important in the future, and can accommodate different assumptions about the nature of the underlying production or cost function, which forms the basis of the assessment, as the regulator's understanding and knowledge of the industry evolve over time. Although the standard DEA approach cannot distinguish between noise and inefficiency, extensions to it can address this weakness at the cost of greater complexity.
Suitability for catch-up and/or frontier-shift efficiency	DEA can be used to assess catch-up efficiency and estimate the frontier shift, provided that panel data is available.
GTS applicability criteria	
Data availability	The NMa would need to investigate which international comparators are most appropriate given GTS's structure, the regulatory framework and the characteristics of the Dutch market in general. While data availability from other European gas transmission companies and data consistency are likely to be key constraints, a high-level (ie, as per many studies examined, not mechanistic) comparison may be possible with US gas transmission companies following EPRG (2006) or with German

[25] Tahvanainen, K. (2010), 'Managing regulatory risks when outsourcing network-related services in the electricity distribution sector', Lappeenranta University of Technology, p. 79.
[26] See, for instance, the application of the bootstrapping technique on the DEA-based efficiency scores to enable statistical inference, in Simar, L. and Wilson, P.W. (2000), 'Statistical Inference in Nonparametric Frontier Models: The State of the Art', *Journal of Productivity Analysis*, **13**:1, pp. 49–78.

	Assessment
	gas transmission companies (where there are 15 national and 20 regional transmission companies). There are no set requirements regarding the number of required comparators; a simple analysis could involve as few as ten companies (provided that the number of variables included in the model is limited to three), although results are expected to be more robust when using larger samples. If the assessment relies on international comparators, the quality of the data might be impaired since the analysis may require some form of data normalisation. Sufficient data from comparators should also be available in the future to allow consistent assessment over time.
Integration	DEA is able to produce efficiency estimates that could then be applied to GTS's cost projections. If the DEA model includes both OPEX and CAPEX as inputs (separately or together as TOTEX), the estimated efficiency score for GTS would provide a percentage inefficiency in total cost that could be used to set a catch-up target for GTS. If data over time is available, DEA can also provide an estimate of frontier shift that can be applied to GTS's cost projections.
Implementation time	DEA is relatively easy to implement, and thus requires relatively few resources, as long as comparable data of reasonable quality is available. If data on international comparators is not readily available, the time required to collate this data could be significantly increased, as could the associated resource costs. The implementation time is likely to be affected by the quality of the international data, owing to the need to organise the data collation process and then make the necessary adjustments to ensure comparability. Given the possible dataset sizes, the time required to develop data for DEA could be similar to that required for COLS (although, in contrast to COLS, and other econometric approaches, DEA does not require a cost function to be specified and tested prior to estimation). This could be simplified if data on GTS were adjusted to slot into a dataset on comparable US or Germany gas transmission companies.
Impact on other aspects	DEA can be used to estimate both the current catch-up inefficiency and the additional frontier-shift adjustment to adjust the tariffs, assuming that consistent and sufficient data on comparable international comparators (European or US) is available on TOTEX and TOTEX drivers. If consistent and comparable data is limited to OPEX on the comparators then, while the catch-up and frontier-shift adjustments on OPEX can be provided by the DEA model, CAPEX would have to be assessed separately. If consistent and sufficient data on regional or European comparators is less likely to be collated given the time constraints, a high-level comparison may be possible with US gas transmission companies following the method set out in EPRG (2006) or using a dataset on German gas transmission companies.

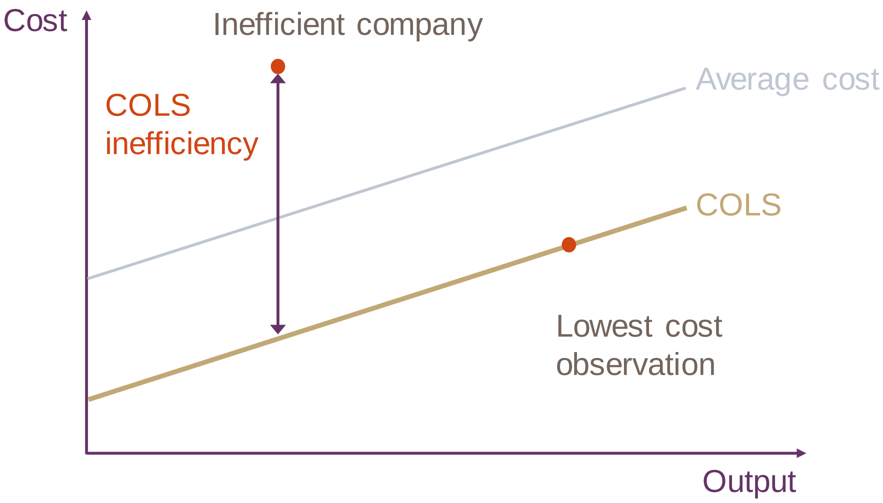
Source: Oxera.

3.2 Corrected ordinary least squares

COLS is an econometric approach based on the simple regression model known as ordinary least squares (OLS). Like DEA and SFA, it is a frontier-based approach; the frontier is derived by shifting the line of best fit from the estimated cost function[27] of the industry derived by OLS, so that instead of representing the average cost of the industry, the line represents the efficiency frontier. This shift is based on either the maximum negative residual of the regression model, resulting in the pure COLS frontier; or a function of the residual sum of errors of the regression models, resulting in the modified OLS (MOLS) frontier. Alternatively, an ad hoc adjustment of either the COLS or MOLS frontiers may be applied (see Figure 3.2 below).

[27] Or, indeed, the production, revenue or profit functions, depending on the goals of the analysis.

Figure 3.2 COLS frontier and efficiency



Source: Oxera.

As with DEA and SFA, COLS requires data on domestic, internal or international comparators. As such, its applicability is dependent on the availability of comparators and data of sufficient quality.

3.2.1 Applications in the regulatory setting

COLS, or similar regression-based approaches, have been extensively used by regulators in the UK to benchmark both regional and national monopolies. More specifically, Ofgem,[28] Postcomm, the postal services regulator,[29] and Ofwat, the water and sewerage services regulator, have used this approach to establish efficiency targets for electricity distribution companies, Royal Mail, the UK universal postal services operator, and the water and sewerage companies respectively. In addition, the EPRG (2006) considered COLS along with DEA and SFA to benchmark a sample of about 43 US interstate gas transmission companies as well as four European gas transmission companies.[30] While the EPRG's preferred models are as discussed in section 3.1.1, to account for possible outliers in the data (and possibly also for other misspecification errors), the residuals estimated by the COLS model were adjusted by 10%.

In the current reviews for the transmission and gas distribution companies and in the next review for the electricity distribution companies, Ofgem had proposed to use a toolkit of approaches to assess the companies' relative efficiency, to include aggregated and disaggregated econometric benchmarking, trend analysis, matrix analysis to assess cost trade-offs, expert review, and project-by-project review. The TOTEX benchmarking methodology that Ofgem had considered using in its econometric analysis was COLS.

Table 3.2 below indicates the cost drivers suggested by Ofgem for the current transmission reviews.

[28] Ofgem's OPEX cost drivers used in its review of electricity distribution companies are found in Ofgem (2009), 'Electricity distribution price control review final proposals - allowed revenue - cost assessment appendix', December 7th, p. 72.
[29] Used as a cross-check to its DEA and SFA results.
[30] EPRG (2006), op. cit.

Table 3.2 Ofgem's TOTEX drivers for the current transmission review

Explanatory variable	Comments
Length of pipeline	The length of gas transmission pipelines is an important driver in establishing the efficient level of costs. In such cost assessments pipeline lengths are normalised with their respective pressure tiers and diameters.
Peak demand	Decreasing the annual flow and increasing the peak demand will lead to significant increases in TOTEX as additional capacity would be required. Comparator data should be checked for compatibility; the definition and the units of measurement of peak demand should be consistent.
Compressor capacity	Compressors are required to transmit gas over longer distances. Installed compressor capacity is included as a cost driver.
Energy delivered	Equals the total annual energy delivered by the transmission network to offtake points. As with peak demand, for the sake of consistency it is important to convert the units across comparators. Although considered less important cost drivers than peak demand, large energy flows and variability warrant inclusion in the analysis.
Asset age	In general, older assets results in lower capital outlays, while operating costs associated with older assets tend to be higher. In the absence of asset age profiles, financial proxies such as depreciation to be used instead.
Other cost drivers	Other factors such as population density, type of terrain and number of road crossings can also be important cost drivers.

Source: Ofgem (2011), 'Decisions on strategy for the next transmission price control – RIIO-T1 Tools for cost assessment', March.

Box 3.1 Ofgem's experience with international benchmarking in RIIO-T1

The text below is based on public domain reports—the December 2010 consultation on strategy document and the March 2011 decision on strategy document—and correspondence with Ofgem in April 2012. In December 2010, when setting out its RIIO-T1 strategy, Ofgem stated its intention to use international benchmarking as one of its assessment approaches to determine the efficient cost. Note that international comparison is necessary in Great Britain because there are only three electricity transmission companies, each with a different scale of operation, and only one gas transmission operator. To facilitate international benchmarking, Ofgem commissioned a team from Cambridge University to collate international data from public domain sources in September 2010. A dataset with data from the USA, Europe, South America and Australia was compiled over a period of two months.

In its decision document in March 2011, Ofgem identified some issues with the US dataset, including but not limited to, the following:

– most transmission networks in the USA are integrated with distribution networks, unlike in Great Britain, which creates difficulties in apportioning costs to gas and electricity transmission alone;
– the measures of transmission network characteristics are different between the UK and the USA—for example, in the latter, pole miles are used instead of circuit miles to report network length.

In the case of data on European companies, it was identified that there were additional issues relating to differences in network characteristics, insufficient data on characteristics and activities across several years.[31] As a result, only the US data was considered to be complete and robust enough to facilitate comparison with GB data. Based on this dataset, Ofgem carried out intensive data analysis and statistic modelling to validate the data and models. With the completion of its international benchmarking report in summer 2011, Ofgem took around one year to conclude the work.

[31] Ofgem (2011), 'Decisions on strategy for the next transmission price control – RIIO-T1 Tools for cost assessment', March, p. 20.

In the end, Ofgem concluded that the international benchmarking was not robust enough to be relied upon without other cost assessment techniques. Nevertheless, it stated that the benchmarking exercise provided useful information and a foundation for future assessment. In particular, the benchmarking was found to be useful in that:

– its conclusion tested Ofgem’s prior judgement based on experience and the other international benchmarking results such as E^3Grid; and
– it provided a good foundation for Ofgem to develop ideas and approaches based on the discussion with transmission companies and academic advisers.

Ofgem also joined a pan-European benchmarking project in seeking a more collaborated benchmarking approach across European regulators, although legal and commercial sensitivity issues could limit the potential for a transparent dataset.

Source: Ofgem (2010), ‘Consultation on strategy for the next transmission price control – RIIO-T1 Tools for cost assessment’, December; Ofgem (2010), ‘Consultation on strategy for the next gas distribution price control - Supplementary Annex - RIIO-GD1 Tools for cost assessment’, December; Ofgem (2011), ‘Decisions on strategy for the next transmission price control – RIIO-T1 Tools for cost assessment’, March; and, correspondence with Ofgem in April 2012.

Based on its experience in RIIO-T1, Ofgem stated that the most difficult, challenging and time-consuming task of international benchmarking is to collect, analyse and normalise a comparable dataset, and that, without effective and efficient data validation and normalisation processes, it would be difficult to achieve convincing benchmarking results. While robustness issues compelled Ofgem to put less focus on international benchmarking during RIIO-T1, it has stated that there will be an ongoing and long-term development to establish a robust and mature benchmarking process during the RIIO-T1 period and beyond.

In October 2011, Ofgem published its initial assessment of the four transmission companies’ business plans. Its aim is to determine which of these plans are robust and well-justified, and therefore to allow a proportionate treatment of certain elements of the business plans or fast-tracking. Given the issues with developing an international comparable dataset, the initial assessment of the business plans was limited to unit cost analysis, for example, historical trend analysis, future volume and scope of activity justification.[32]

For the UK gas and electricity distribution companies, by using panel data covering several years, Ofgem has a large enough dataset to compare only GB distribution companies, and does not need to extend the comparator set internationally. There are eight gas distribution companies and 14 electricity distribution companies, and Ofgem uses a panel over a period of three years for the former and four years for the latter.[33]

Table 3.3 indicates the cost drivers suggested by Ofgem in the current gas distribution review.

[32] Ofgem (2011), ‘Initial assessment of RIIO-T1 business plans’, October 24th.
[33] In the case of benchmarking gas distribution companies, see Ofgem (2011), ‘Decision on strategy for the next gas distribution price control - RIIO-GD1 Tools for cost assessment’, March, para 4.23. In the case of benchmarking electricity distribution companies, see Ofgem (2009), ‘Electricity Distribution Price Control Review Final Proposals - Allowed revenue - Cost assessment appendix’, December, Appendix 8, p. 59.

Table 3.3 Ofgem's cost drivers for the gas distribution review

Cost area	Sub-cost category	Drivers
TOTEX (single model)		MEAV, CAPEX, CSV, REPEX CSV, total reports
TOTEX (aggregated CAPEX, REPEX, OPEX)	CAPEX	MEAV, CAPEX CSV
	OPEX	MEAV, PREs, total reports
	REPEX	Length of less than 7bar metallic network, REPEX CSV
Bottom-up regressions	Top-down OPEX	CSV of MEAV and PREs
	Work management	MEAV
	Emergency	PREs
	Repair	Total mains and service condition external reports
	Maintenance	MEAV
	Connections	Number of connections
	Mains reinforcement	CSV of length of main laid above and below 180mm
	REPEX	CSV of length of main by diameter band

Note: MEAV, modern equivalent asset value; CSV, composite scale variable; PRE, publicly reported gas escape; and REPEX, replacement expenditure.
Source: Ofgem (2011), 'Decision on strategy for the next gas distribution price control - RIIO-GD1 Tools for cost assessment: Supplementary Annex (RIIO-GD1 Overview paper)', March 31st.

In the England and Wales water industry, Ofwat estimates econometric cost models for different groups of costs, such as resource and treatment, or business support costs, for a single year. It then aggregates the results to obtain an overall efficiency target. With 21 companies providing water services, Ofwat compares costs at a company level. For sewerage services, with ten companies, it undertakes econometric modelling at a sub-company level (eg, treatment works), or compares unit costs at a company level.[34]

For postal services, since there are no domestic comparators for Royal Mail, the regulator has used sub-company, regional comparators to estimate efficiency using COLS cost models. Delivery offices and mail centres throughout the UK were benchmarked against each other to obtain an efficiency target. Postcomm also applied DEA and SFA to the same dataset in order to strengthen the robustness of its assessment.[35]

3.2.2 Advantages and disadvantages

As noted in section 3.1, one major advantage of COLS is that the resulting performance estimate of each company is based on the actual, realised performance of its direct peers assuming that these are available. In this case, these would be other gas transmission companies. As a result, although the assessment is high-level, it is also firmly grounded on the actual performance observed and achieved by other similar companies. As with DEA, the approach is relatively robust and transparent, at least with respect to how the efficiency measures are derived. Also, similar to DEA, the overall robustness of the approach greatly depends on the quality and the quantity of the available data, and an assessment based on international comparators could introduce a number of complications into the analysis.

As an econometric-based approach and in a similar way to DEA, COLS enables the efficiency assessment to take into account several different cost drivers. In addition, as it is grounded in econometric theory, a large number of statistical tests can easily be applied to examine the overall robustness of the results. COLS is also likely to be the easiest top-down

[34] Ofwat (2009), 'Relative efficiency assessment 2008-09 – supporting information', December.
[35] See Postcomm (2006), 'Royal Mail's Price and Service Quality Review 2006-10 – Licence Modifications Proposals', March.

approach to implement in an assessment exercise, assuming that the comparator dataset is already available or easy to collate.

The major disadvantage of COLS and similar approaches is that it is both deterministic and parametric. In other words, like DEA, it cannot directly control for the presence of noise in the data, but, unlike DEA, it requires the researcher to make an ex post assumption regarding the shape (functional form) of the estimated function. To deal with the first issue, in the past regulators have made subjective adjustments to the efficiency estimate to correct for the effect of noise, such as measurement error; this usually results in a decrease in the resulting efficiency targets.

3.2.3 Assessment of COLS

Table 3.4 summarises how DEA measures against the general and GTS applicability criteria listed in section 2.

Table 3.4 Assessment of COLS against criteria

	Assessment
General criteria	
Complexity and transparency	As noted, this approach is the easiest to implement of all the frontier-based approaches considered in this section. Like DEA and SFA, COLS is a transparent method, provided that the underlying models are made publicly available.
Reliability	COLS has been used extensively by regulators and can produce relatively accurate results, provided that the model is specified correctly. However, since COLS shares the most significant disadvantages of both DEA and SFA by being both deterministic and parametric, most performance measurement applications in the academic literature prefer to use either DEA when sample sizes are small or SFA when sample sizes are sufficiently large. To deal with the deterministic nature of COLS, in the past regulators have made subjective adjustments to the efficiency estimate to correct for the effect of noise, such as measurement error. Additionally, although COLS allows statistical tests to verify the reliability, owing to its parametric nature, it is not as flexible as DEA; the impact of this drawback on the overall performance assessment is likely to be small.
Suitability for catch-up and/or frontier-shift efficiency	COLS can be used to derive estimates of catch-up efficiency and frontier shift, provided that panel data is available.
GTS applicability criteria	
Data availability	As with DEA, the NMa would need to investigate which international comparators are most appropriate given GTS's structure, the regulatory framework and the characteristics of the Dutch market in general. While data availability from other European gas transmission companies and data consistency are likely to be key constraints, a high-level comparison may be possible with US gas transmission companies following EPRG (2006) or with German gas transmission companies. COLS is not particularly demanding in terms of the number of comparators; as an example, Ofgem has applied COLS using only eight comparators, although this was over a period of three years, providing a total of 24 observations, and Ofwat has applied COLS using data over a single year on 21 water companies, although results are expected to be more robust when larger samples are used. Sufficient data from comparators should also be available in the future to allow consistent assessment over time.
Integration	As for DEA, it should be possible to integrate the output of COLS in the regulation of GTS. This is because the methodology produces clear efficiency targets that can be applied to cost projections. If the COLS model includes TOTEX as the input, the estimated efficiency score for GTS will provide a percentage inefficiency in total cost that could be used to set a catch-up target for GTS. If data over time is available, COLS can also provide an estimate of frontier shift that can be applied to GTS's cost projections.

	Assessment
Implementation time	Of all the frontier-based approaches examined in this section, COLS is less complex to implement. Although differences in implementation time in terms of analysis are marginal between the frontier-based approaches, the time needed to implement the approach depends on the amount of data that needs to be collated and processed
Impact on other aspects	As with DEA, COLS can be used to estimate both the current catch-up inefficiency and the additional frontier-shift adjustment to adjust the tariffs assuming that consistent and sufficient data is available on TOTEX and TOTEX drivers on regional comparators, or that comparable European comparators are available. If consistent and comparable data is limited to OPEX on the comparators then, while the catch-up and frontier-shift adjustments on OPEX can be provided by the COLS model, CAPEX would have to be assessed separately.

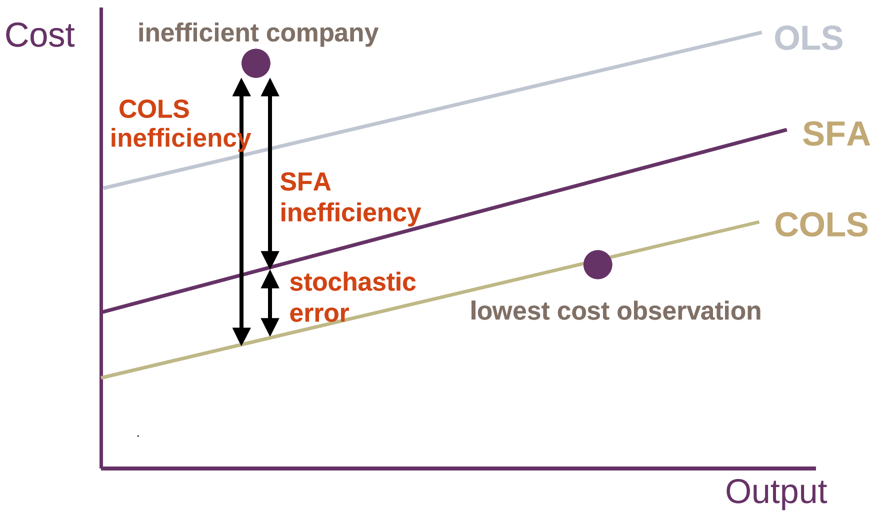
Source: Oxera.

3.3 Stochastic frontier analysis

One of the main weaknesses of the COLS (and the standard DEA) approach is that it assumes that any difference between a company's observed costs and the regression line (ie, the residual) represents inefficiency. It does not account for any stochastic error or noise in the model, such as measurement error, which affects the size of the residual. COLS (and standard DEA) may therefore impose harsh targets when the stochastic error element of the residual is large.

SFA is an econometric technique that attempts to distinguish between random error and inefficiency in the model. By assuming a distribution for the inefficiency and the random error components of the residual,[36] SFA is able to decompose the residual term into inefficiency and noise, and thereby identify the relative inefficiency of each firm in the sample. SFA can also test for the presence of inefficiency.[37] Figure 3.3 illustrates the difference between the COLS and SFA techniques for estimating inefficiency.

Figure 3.3 Estimating inefficiency using COLS and SFA



Note: This figure is intended only as an illustration of the difference in frontiers between OLS, COLS and SFA.
Source: Oxera.

[36] The stochastic error component of the residual is assumed to be normally distributed. The inefficiency component may be half-normal, truncated or exponentially distributed.
[37] For a more detailed discussion on SFA, see Kumbhakar, S.C. and Lovell, C.A.K. (2003), *Stochastic frontier analysis*, Cambridge University Press.

3.3.1 Applications in the regulatory setting

There are quite a few examples of SFA being applied in the regulatory setting by regulators and academics. As noted above, EPRG (2006) considered SFA, along with DEA and COLS, to benchmark a sample of about 40 US interstate gas transmission companies and four European companies.[38]

Between 2008 and 2011, the Finnish energy regulator used SFA to assess TOTEX incurred by electricity distribution companies in the Finnish market. Based on several years of company data, the SFA analysis was combined with DEA to determine the scope for efficiency catch-up, with 50% weighting attached to SFA.

In the UK, Ofcom, the UK telecoms regulator, has also used SFA to assess the TOTEX of BT's Openreach activities. Using publicly available information, BT was compared against the US local exchange carriers (LECs) by normalising BT's data to be comparable to that of the US companies.[39] This approach reduced the data burden, since it required an adjustment to BT's dataset only, given that the US LEC dataset was already well-established and collated on a consistent basis.

SFA has also been applied by Postcomm to assess the performance of Royal Mail; this relied on internal comparisons of the mail centres and delivery offices, which, although collectively managed by Royal Mail, operate with some degree of autonomy.[40]

The Office of Rail Regulation (ORR) used SFA to assess the efficiency of the rail network operator, Network Rail, using data from international comparators. It identified a significant efficiency gap between Network Rail and its comparators. For the price control review for the period 2013–18, the ORR is proposing to use advanced SFA models to assess the efficiency of Network Rail,[41] based on recommendations made by Oxera (2009).[42] SFA has also been explored briefly by Ofgem at previous gas and electricity distribution price control reviews, but, given the data limitations, it did not consider the approach appropriate.[43]

3.3.2 Advantages and disadvantages

As noted in the sections on DEA and COLS, frontier-based approaches offer the advantage that they all assess performance according to what has already been achieved by a group of direct comparators—ie, companies that undertake almost the same functions as the assessed company—assuming that data on direct comparators is available. This has the potential to increase the robustness of frontier-based estimates relative to other approaches that require more objectivity on the part of the regulator or the consultant undertaking the analysis. However, as with DEA and COLS, the overall robustness of the approach greatly depends on the quality and quantity of available data and comparators, and an assessment based on international comparators can introduce a number of complications in the analysis. For a more detailed discussion, see section 3.1.

SFA shares some of the main advantages of COLS. For example, it can take into account a number of different cost drivers, and several statistical tests are available for hypothesis testing both before and after the estimation phase. However, it also shares one of its disadvantages, in that it is also a parametric approach and is therefore sensitive to the assumptions on the overall shape of the estimated function, otherwise referred to as 'functional form mis-specification'.

That said, as SFA is stochastic, where noise is prevalent in the data it is likely to produce more robust results than DEA and COLS, although even SFA estimates can become quite

[38] EPRG (2006), op. cit.
[39] For more information, see NERA (2008), 'The comparative efficiency of BT Openreach', a report for Ofcom, March.
[40] See Postcomm (2006), 'Royal Mail's Price and Service Quality Review 2006-10 – Licence Modifications Proposals', March.
[41] ORR (2011), 'Establishing Network Rail's efficient expenditure', July, pp. 28–30.
[42] Oxera (2009), 'Recommendations on how to model efficiency for future price reviews', November.
[43] See, for example, Ofgem (2009), 'Electricity Distribution Price Control Review Final Proposals - Allowed revenue - Cost assessment appendix', December, p. 96.

inaccurate when measurement error is significant. In addition, SFA's ability to distinguish between noise and inefficiency comes at a cost. First, in order for SFA to produce robust results, the dataset needs to be quite large; second, it requires an assumption on the distribution of the inefficiency term, although this factor is usually not a major weakness of the approach on the whole since various academic studies have shown that the amount of bias introduced to the results owing to mis-specifying the inefficiency distribution tends to be negligible in most cases.[44]

3.3.3 Assessment of SFA

Table 3.5 summarises how SFA measures against the general and GTS applicability criteria listed in section 2.

Table 3.5 Assessment of SFA against criteria

	Assessment
General criteria	
Complexity and transparency	SFA requires specialist software, and on the whole could be considered a complex approach, even though its principles are relatively easy to explain. Extensions to SFA are even more complex, and are likely to require either specialist or bespoke software to implement. Like DEA and COLS, SFA is a transparent method, provided that the underlying models are made publicly available
Reliability	SFA is one of the most common approaches found in the academic literature for assessing efficiency in the regulatory setting. Owing to its stochastic nature, estimates derived from SFA have the potential to be more accurate than those produced by either DEA or COLS, assuming that the SFA model does not suffer from either severe functional form or inefficiency distribution mis-specification. The impact of this potential mis-specification on the overall performance assessment could be quite small, as several academic studies have demonstrated. In addition, the ability to distinguish between noise and inefficiency in the data makes SFA more flexible if data quality deteriorates for whatever reason in a future regulatory review. This gives SFA an additional advantage over the other frontier approaches.
Suitability for catch-up and/or frontier-shift efficiency	SFA can be used to derive estimates of catch-up efficiency and frontier shift, provided that panel data is available.
GTS applicability criteria	
Data availability	As with the other two frontier-based approaches, the NMa would need to investigate which international comparators are most appropriate given GTS's structure, the regulatory framework and the characteristics of the Dutch market in general. In addition, SFA has more stringent requirements than COLS and DEA, in terms of the number of data points (observations) required in the analysis. While data availability from other European gas transmission companies and data consistency are likely to be key constraints, a high-level comparison may be possible with US gas transmission companies following EPRG (2006) or with Germany gas transmission companies. There are no set minimum data requirements, but, in most cases, SFA models with fewer than 30–40 observations can fail to produce results. These data requirements are probably the main reason why SFA has not been as widely adopted by regulators as COLS and DEA. Sufficient data from comparators should also be available in the future to allow consistent assessment over time.

[44] See, for example, Kumbhakar, S.C. and Lovell, C.A.K. (2003), *Stochastic frontier analysis*, Cambridge University Press.

	Assessment
Integration	As with COLS and DEA, it should be possible to integrate the output of SFA in the regulation of GTS, since the methodology produces clear efficiency targets that can be applied to cost projections. If the SFA model includes TOTEX as the input, the estimated efficiency score for GTS would provide a percentage inefficiency in total cost that could be used to set a catch-up target for GTS. If data over time is available, SFA could also provide an estimate of frontier shift that can be applied to GTS's cost projections.
Implementation time	SFA is more complex than COLS and possibly DEA, and requires a larger dataset than the other frontier-based approaches to produce results. If the required data is available, comparatively little additional time is likely to be needed to implement DEA, COLS or SFA.
Impact on other aspects	As with DEA and COLS, SFA can be used to estimate both the current catch-up inefficiency and the additional frontier shift adjustment to adjust the tariffs, assuming that consistent and sufficient data is available on TOTEX and TOTEX drivers on regional comparators or comparable European comparators are available If consistent and comparable data is limited to OPEX on the comparators then, while the catch-up and frontier-shift adjustments on OPEX can be provided by the SFA model, CAPEX would have to be assessed separately.

Source: Oxera.

3.4 Growth accounting-based total factor productivity

Growth accounting-based TFP (in this report referred to simply as 'TFP') is arguably the most widely adopted method by regulators and academics for measuring productivity growth in economic aggregates—ie, the whole economy or sectors (whole industries) of the economy. A major factor in the widespread adoption of TFP in this context is that estimates can be (relatively) easily produced using country- or sector-specific national accounts data, without recourse to information from outside the country or the sector examined. On the other hand, TFP does not use direct comparators and requires the adoption of a number of simplistic assumptions, most notably that markets are perfectly competitive, which can lead to unreliable estimates.

A typical TFP analysis examines the productivity growth of a number of sectors of the economy that are deemed comparable to the assessed company, and, using the underlying cost structure of the assessed company, constructs a 'composite' benchmark. In essence, this composite benchmark is a weighted average of the historical productivity performance of the selected sectors of the economy. The sectors in question form the comparator group used to benchmark the regulated company.

Since the TFP analysis is based on productivity growth estimates, the resulting benchmark is also a productivity growth benchmark—ie, it includes both catch-up and frontier-shift elements. These cannot be primarily decomposed, but regulators have frequently used secondary evidence to translate the productivity benchmark into an estimate of frontier shift. These frontier-shift estimates are likely to be less robust than those derived from approaches that rely on direct comparators such as frontier-based approaches, but, as noted below, this type of TFP analysis is usually employed in conjunction with other approaches.

3.4.1 Applications in the regulatory setting

As noted above, regulators have frequently used TFP to produce either an overall cost-reduction or a frontier-shift estimate for regulated companies. However, in most cases, regulators tend to use other assessment approaches in conjunction with TFP analysis when considering the scope for overall cost reductions. In the UK, TFP analysis has been used by the ORR to estimate the frontier shift for Network Rail.[45] Other UK regulators, including

[45] For more information, see Oxera (2008), 'Network Rail's scope for efficiency gains in CP4', prepared for Office of Rail Regulation, April.

Ofgem, Ofwat, and the Office of the PPP Arbiter and the Civil Aviation Authority (CAA),[46] have adopted similar approaches, using the results as primary evidence for their assumption on frontier shift.

In the past, the NMa has used a similar approach to determine the productivity gains to be taken into account when setting a price control for the transportation business of GTS,[47] although its focus was on intermediate input and labour costs (labour productivity) rather than TFP.

3.4.2 Advantages and disadvantages

The major advantage of a TFP analysis that relies on composite benchmarks is that it can be implemented where there are no direct comparators, or where it is deemed that the data is of insufficient quality to rely on direct comparisons. Although the TFP approach described above requires consistent data on inputs, outputs and their relative prices for the sectors of the economy that form the comparator group, this information could be easily sourced either from pan-European productivity databases such as EU KLEMS or national statistical agencies (in this case, Statistics Netherlands).

The main disadvantage of TFP analysis is that the approach does not measure relative or static efficiency, but rather overall productivity growth, without making it clear what proportion of productivity gains are attributable to catch-up efficiency improvements and what to frontier shift. Evidence from other sources could be used to decompose the aggregate TFP productivity growth estimate into its constituent parts. Thus, it can be used to help determine a benchmark for future rates of improvement, but it cannot determine GTS's static efficiency.

The other disadvantage of TFP analysis that relies on composite benchmarks is that the companies in the comparator group do not necessarily undertake the same activities as the assessed company, but rather are from sectors of the economy that are deemed by the regulator to carry out similar activities. Owing to the nature of these indirect comparisons, the robustness of this approach is likely to be significantly reduced compared with the frontier-based approaches discussed above, assuming that such approaches are feasible in the first place. Also, the regulator needs to provide robust justification for why and how the comparator group of sectors is appropriate, since choosing a different group might have a significant impact on the results—although this latter issue has been partly addressed in previous assessments through the use of extensive sensitivity analysis.

3.4.3 Assessment of TFP

Table 3.6 summarises how TFP measures against the general and GTS applicability criteria listed in section 2.

Table 3.6 Assessment of TFP against criteria

	Assessment
General criteria	
Complexity and transparency	TFP analysis, as described in this section, is not particularly complex. It can also be transparent, since the underlying data is publicly available and the approach itself is well-documented.
Reliability	This analysis relies on indirect comparators, the selection of which requires a degree of judgement in undertaking the analysis. As such, the resulting estimates are likely to be less robust than those produced by the frontier-based approaches described earlier. In terms of stability, the approach is relatively flexible, in that both the comparator set and the relative weightings required to generate the composite benchmark can change as the activities of the assessed company evolve over time.

[46] Respectively, the regulators of the infrastructure companies that operate the London Underground and of the major UK airports.
[47] Reckon (2008), 'The productivity growth of GTS', July 15th; Reckon (2011), 'The productivity growth of GTS', March 23rd.

	Assessment
Suitability for catch-up and/or frontier-shift efficiency	TFP analysis measures overall productivity growth, which combines the estimates of catch-up efficiency and frontier shift. The approach cannot directly decompose the two elements from the overall productivity measure and does not measure relative or static efficiency. If this decomposition is required, external evidence from other sources would be needed, which could impair the robustness of the resulting estimates. In most cases regulators tend to use other assessment approaches in conjunction with TFP analysis.
GTS applicability criteria	
Data availability	TFP analysis using sectors of the economy as comparators relies on productivity growth estimates that are either readily available or relatively straightforward to estimate using Dutch national accounts data. Of all the top-down approaches considered in this section, TFP analysis is probably the most feasible to implement for the current review.
Integration	The analysis cannot provide an estimate of GTS's current levels of relative efficiency; rather, the focus is on estimating the scope for future cost reductions based on the performance of the composite benchmark.
Implementation time	The TFP analysis as described in this section is likely to be the least resource-intensive approach covered by this report, including the various bottom-up approaches. The implementation is relatively straightforward.
Impact on other aspects	The tariffs could be adjusted for the total scope of future cost reductions estimated by the approach. However, limitations remain in terms of decomposing the overall measure into catch-up efficiency and frontier-shift improvements although secondary sources could be used to facilitate this if required.

Source: Oxera.

3.5 Unit cost and single factor productivity comparisons with other regulated companies

Unit cost or single factor productivity comparisons can be used to assess the regulated company's efficiency. There are generally two ways such top-down unit costs comparisons could be used, depending on the available data: analysis of unit cost levels to estimate catch-up; or analysis of unit cost trends to estimate the total scope for efficiency saving, as it includes both catch-up and frontier-shift. Unit costs and single factor productivity measures are also extensively used in bottom-up approaches, as discussed in section 4.

3.5.1 Comparisons of unit cost or single factor productivity levels

When data on direct comparators is available,[48] a simple performance metric can be derived by comparing between companies the costs per unit of output (eg, OPEX per m^3 of gas transported or per customer served) or the output per unit of input (eg, customers per employee). Such comparisons provide an indication of static efficiency. While such unit costs or productivity measures are relatively simple to calculate, they do not take into account inherent differences between companies, such as size—larger companies would normally be expected to have lower unit costs than smaller companies. As such, when data on direct comparators is available, most regulators usually adopt frontier-based approaches, such as those discussed in sections 3.1 to 3.4.

Unit cost comparisons are usually undertaken as a cross-check or if the preliminary analysis finds that the data is not suitable for undertaking a more complex analysis owing to comparability concerns or because there are too few comparators. For example, Ofgas used unit cost comparisons such as supply costs per customer from UK public electricity suppliers to sense-check its analysis when reviewing British Gas Trading's price regulation.[49] Along with DEA, the NMa used similar unit cost comparisons, sourced from the 2009 E^3 Grid

[48] For GTS this will require international comparators.
[49] Ofgas (1999), 'Review of British Gas Trading's Price Regulation', Initial proposals, November.

report,[50] to set the most recent cost-reduction targets for TenneT. These unit costs were calculated using TOTEX, OPEX, CAPEX and adjusted TOTEX,[51] respectively, with the grid size as the output measure of choice. The comparisons were undertaken using both the whole sample of 22 transmission companies and a sub-sample derived after removing outlier companies.[52] The report considered that the most relevant unit cost measures were those based on TOTEX and adjusted TOTEX, after the removal of outliers.

3.5.2 Comparisons of unit cost and single factor productivity trends

Historical reductions in unit costs or increases in single factor productivity achieved by other network utility companies could also be used to provide a benchmark range of possible future cost reductions. As with the TFP analysis described above, this approach relies on indirect comparisons of performance over time to help determine a benchmark for future rates of improvement and does not, therefore, provide a measure of static efficiency. However, in this case, the comparator group is usually made up of other regulated network companies, rather than whole sectors of the economy. In addition, since the information required to undertake a total cost assessment of the companies in this comparator group is usually not available or difficult to normalise,[53] the performance measure used in such studies has generally been RUOE or output per employee, such as number of customers per employee, although TOTEX unit cost comparisons are also possible.[54]

RUOE is calculated by dividing a measure of OPEX by a measure of output and putting in constant price terms, while a single factor productivity measures is calculated by dividing the output measure by the input measure—often, the number of employees. Reductions in unit costs or increases in productivity are calculated for each year for which data is available for each of the comparator companies. The productivity increases of the selected companies are then used as a benchmark for productivity increases achievable by the assessed companies. A critical issue in this analysis is to ensure the comparability of the historical performances in the comparator group; to this end, the analysis usually makes adjustments to the cost data to control for issues such as inflation, the impact of economies of scale, changes in accounting standards, and changes in the nature of outputs.

3.5.3 Applications in the regulatory setting

Ofgem commissioned a report to assess the costs and efficiency of operations of Transco, the UK's gas transmission company at the time, as part of its gas price control review for the period 2002–07.[55] The RUOE analysis was based on regulated companies in other sectors of the UK economy, such as electricity and rail. The evidence suggested that Transco had scope to reduce RUOE by about 2–4% a year over 2002–07, or 1–3% once capital substitution had been taken into account. For the current price control reviews of electricity and gas transmission companies in the UK, Ofgem has compared the unit costs of similar activities, along with the volume of planned work and the impact on the output measures, between all the transmission companies, and with the gas and electricity distribution companies, where appropriate.[56]

Owing to a lack of domestic comparators for Network Rail, the ORR has also reviewed historical improvements in RUOE in regulated companies in other sectors for the 2008 price control review.[57] The results contributed to the establishment of an efficiency target for the

[50] Sumicsid (2009), 'International Benchmarking of Electricity Transmission System Operators, e^3 Grid Project Final Report', March.
[51] Adjusted TOTEX is basically an OPEX measurement that takes into account the capital input since any possible inefficiency in investments has been eliminated.
[52] A transmission company was considered to be an outlier if it outperformed the others to an extreme degree or had a considerable impact on the evaluation of a large number of transmission companies.
[53] See discussion in section 4.4.
[54] Sumicsid (2009), 'International Benchmarking of Electricity Transmission System Operators, e^3 Grid Project Final Report', March.
[55] Mazars Neville Russell (2001), 'Transco price control review 2002–2007', report to Ofgem, September 7th.
[56] Ofgem (2011), 'Decision on strategy for the next transmission price control - RIIO-T1 Tools for cost assessment', March.
[57] For more information, see Oxera (2008), 'Network Rail's scope for efficiency gains in CP4', prepared for Office of Rail Regulation, April.

company. Since this analysis used data from regulated utility companies, much of it was collected by other regulators and was therefore publicly available.

3.5.4 Advantages and disadvantages

The main advantage of using unit cost levels based on direct comparators or unit cost trend analysis that relies on direct comparators is that it avoids the more stringent requirement of the frontier-based approaches in relation to the number of available comparators. As such, a comparison of unit costs is still possible, even when a preliminary analysis of the data reveals that the use of frontier-based approaches is not practical. The disadvantages of this approach are numerous, however: trend analysis results in an average measure of performance and measures productivity growth that cannot directly distinguish between catch-up efficiency and frontier shift; both unit cost level and trend analysis can take only single output into consideration; and cannot deal with the existence of economies of scale or scope, and other issues.

RUOE analysis shares many of the advantages and disadvantages of the TFP analysis described in this section. Its advantage—but also main disadvantage—is that it relies on indirect comparators. While this avoids the need to collate data on international comparators, the resulting cost-reduction estimates are less robust relative to frontier-based approaches since they rely on comparators that do not undertake the same activities as the assessed company, and they take only one output into consideration.

It could be argued that RUOE trend analysis has one advantage over TFP analysis in that, since the comparator set is other regulated network companies, the resulting performance measure is likely to be closer to what can be achieved by the assessed company than the performance of a composite made up of different sectors of the economy. This comes at a cost, however, as the RUOE measure focuses on OPEX only, rather than TOTEX; although TOTEX unit cost comparisons are also possible.[58] As such, additional analysis could be required to assess the performance of the regulated company in relation to its CAPEX.

A further issue with RUOE analysis relates to the choice of the output measure used to construct the unit cost measure. Although this decision can be informed by using some simple statistics, it would always involve some degree of judgement by the regulator or consultant undertaking the analysis. Since network utilities are complex organisations with a large number of outputs and other factors that affect costs the ultimate choice is likely to be open to criticism.

3.5.5 Assessment of RUOE

Table 3.7 summarises how RUOE measures against the general and GTS applicability criteria listed in section 2.

Table 3.7 Assessment of RUOE against criteria

	Assessment
General criteria	
Complexity and transparency	Both unit cost comparisons that rely on direct comparisons and RUOE analysis (of trends and levels) as described in this section are considered relatively simple approaches. Additionally, they can be transparent, provided that the data is publicly available and the adjustments made to the data are well-documented.

[58] Sumicsid (2009), 'International Benchmarking of Electricity Transmission System Operators, e^3 Grid Project Final Report', March.

	Assessment
Reliability	Comparisons of unit cost levels or RUOE analysis that rely on direct comparisons are quite simplistic and cannot take into account a number of factors that could affect costs. If the data on a sufficient number of comparators is available, the frontier-based approaches described earlier are quite likely to provide more robust estimates. On the other hand, if this is not possible, direct unit costs can provide a view on the performance achieved by other companies that undertake the same activities as the assessed company. The selection process for the comparator set for RUOE trend analysis that relies on indirect comparators, the necessary adjustments to the cost data to ensure comparability between the comparators, and the selection of the most appropriate measure of output to construct the unit costs all require a degree of judgement when undertaking the analysis. In comparison, frontier-based approaches that rely on international comparators do not require such subjective decisions on the comparator set, and, while the data normalisation process can require a degree of judgement, the modelling process, including the selection of factors to include in the assessment, is informed by objective statistical tests in almost all cases. In addition, as RUOE focuses on OPEX alone, the regulator would need to undertake a separate exercise to assess CAPEX—although TOTEX unit cost comparisons are also possible. For all these reasons, the resulting estimates are likely to be less robust than those produced by the frontier-based approaches described earlier. Unit cost comparisons that rely on direct comparisons and the RUOE approach are both quite straightforward, in that the major decisions required relate to the comparator set, the length of the time examined, and the output measure used to create the unit costs. As such, the flexibility inherent in such approaches is limited.
Suitability for catch-up and/or frontier-shift efficiency	RUOE trend analysis, like TFP analysis, measures the overall level of improvement, which includes a measure of catch-up efficiency and frontier shift. The approach cannot directly decompose the two elements from the overall productivity; rather, the assessment would need to rely on external evidence, such as academic studies that have examined decomposition using different approaches. The unit cost approach in levels allows the current inefficiency level of GTS to be measured.
GTS applicability criteria	
Data availability	RUOE trend analysis that relies on indirect comparators requires data on other regulated network utilities. In the UK, such data is readily available through the publicly available regulatory accounts published by the companies. However, if the NMa chooses to use such an approach and data is not available for the Netherlands, data from the UK-based utilities could be collated.
Integration	As with the TFP analysis, RUOE trends cannot provide an estimate of GTS's current levels of efficiency; rather, the focus is on estimating the past efficiency gap based on the performance of the comparator set. The unit cost approach in levels measures the current degree of inefficiency of the assessed company.
Implementation time	The actual unit cost or RUOE analysis should not be resource-intensive in general. Given that data availability for either direct or indirect comparators is unknown at this time, it is difficult to judge what the overall resource burden would be.
Impact on other aspects	RUOE focuses on OPEX only rather than TOTEX, although it can be used to assess TOTEX. Otherwise, CAPEX may need to assessed and reimbursed separately. Also, similar to TFP analysis, RUOE trend measures overall productivity growth and cannot directly distinguish between catch-up efficiency and frontier shift, although secondary sources could be used to decompose the components if required. The unit cost approach in levels (OPEX and CAPEX separately or together as in TOTEX) allows the current inefficiency level of GTS to be measured.

Source: Oxera.

4 Bottom-up approaches

Bottom-up approaches look at the costs of individual activities to build up a picture of the overall cost of providing a service. This may be done through benchmarking of individual activities, processes and/or assets, using external benchmarks from within the industry or outside. For example, HR costs might be benchmarked using data from companies in different industries. It may also be done by constructing engineering models of the costs for a hypothetical efficient operator.

This section provides a broad overview of the bottom-up approaches. In general, these tend to be less standardised than their top-down counterparts, and their application depends to a large extent on the circumstances of the assessment—ie, the industry to which the assessed company belongs, its functions, and the general environment in which it operates. As such, the discussion provided here focuses on providing some relevant case studies rather than delving into the more detailed workings of each approach.

The bottom-up approaches examined here include:

– process benchmarking;
– long-run incremental cost (LRIC) models;
– benchmarking based on a reference model; and
– unit cost benchmarking for CAPEX.

Most of these approaches have been used by some regulators to assess gas or electricity transmission companies, as detailed below:

– **Process benchmarking**—Ofgem used this approach to assess the efficiency of support or 'headquarter' functions undertaken by TSOs, including HR & Scheme Trainees; Procurement & Logistics; Business Services Finance; Transmission Finance; Communications; Legal; Safety, Health, Environment (SHE); Regulation; and Audit. Ofgem is considering extending the use of this approach in the current and future price control reviews.[59]

– **LRIC models**—although used in the past to evaluate certain functions and 'products' offered by transmission companies, as far as Oxera understands they are not usually used to estimate efficiency in a regulatory context. As such, the use of this approach in an efficiency assessment setting could be considered to be original.

– **Benchmarking based on a reference model**—BnetzA, the German energy regulator, considered this as an alternative approach in case the SFA and DEA assessments were deemed infeasible or not robust.[60]

– **Bottom-up unit cost benchmarking**—Ofgem has used this form of benchmarking to assess the electricity and gas companies' CAPEX. It is also considering extending the use of this approach in the current and future price control reviews, given data and robustness issues with TOTEX econometric benchmarking.[61]

[59] Ofgem (2011), 'Decision on strategy for the next transmission price control - RIIO-T1 Tools for cost assessment', March; and Ofgem (2011), 'Decision on strategy for the next distribution price control - RIIO-GD1 Tools for cost assessment', March.
[60] BnetzA (2007), 'Vorstellung der Referenznetzanalyse'.
[61] Ofgem (2011), 'Decision on strategy for the next transmission price control - RIIO-T1 Tools for cost assessment', March; and Ofgem (2011), 'Decision on strategy for the next distribution price control - RIIO-GD1 Tools for cost assessment', March.

4.1 Process benchmarking

Process benchmarking compares the cost of individual activities against benchmarks from both within the sector and, where relevant, other sectors. In many cases, it is used to evaluate corporate costs, where functions are comparable across businesses and sectors and data is publicly available. For sector-specific functions, external benchmarks are harder to source. Benchmarks are typically based on key performance indicators (KPIs) and simple metrics, including ratios of function costs to total OPEX, cost per full-time equivalent (FTE), cost as a proportion of revenue, and number of FTEs.[62]

4.1.1 Applications in the regulatory setting

Gas and electricity transmission

In 2001 Ofgem used a bottom-up analysis of gas transmission direct and indirect costs as part of Transco's price control review.[63] Specific cost categories, such as employment costs and bad and doubtful debt, were examined. Also, specific activity costs, such as account management within the shipper service, and aggregate corporate overhead costs were considered. With respect to the latter, Ofgem analysed Transco's corporate cost data against third-party evidence from other UK utilities and international comparators. Using data from a survey commissioned by the US Council of Public Relations Firms, a sample of UK companies of comparable size and scale of legal activity, and US and Australian companies for administration and corporate overheads, the following costs were assessed against UK and international companies with similar business features:

– corporate communications spending;
– HR department costs;
– legal expenditure;
– regulatory costs, such as compliance.

In determining the electricity transmission price controls for 2007 to 2012, Ofgem reviewed the level of costs in business support functions of the NGC Group, comprising National Grid Electricity Transmission and National Grid Gas. The benchmarking was at a functional level on controllable operating costs, which were normalised to allow comparisons. These normalisations included functional costs per FTE, number of functional FTEs per total company FTE, and functional costs as a percentage of revenue. In total, £120m of NGC's £313m controllable costs were benchmarked. The functions benchmarked were the Corporate Centre; Operational Telecoms; HR and Scheme Trainees; Business Services Finance; Procurement and Logistics; Legal; Safety, Health, Environment and Security; Regulation; and Internal Audit.

High and low scenarios were provided for the potential savings. The process to derive the cost-reduction targets for support services started with a review of the Historic Business Plan Questionnaire (HBPQ), to identify trends and shocks in the costs for the four-year period. The model used by NGC to allocate costs between the functions was reviewed by the regulator. The review of the HBPQ led to a questionnaire, meetings and workshops with NGC to ensure an adequate understanding of the data in the HBPQ and to gain insights into the operational processes. The benchmarks came from external reports published by organisations that specialise in providing benchmarking information for particular functions, benchmarking reports commissioned by National Grid, and electricity distribution companies. This provided high-level comparison for certain functions. Data from other transmission companies was not available at the time.

[62] For example, see Oxera (2003), 'Benchmarking of Operating Expenditure', Report prepared for the Office of the Rail Regulator.
[63] Arthur Andersen (2001), 'Report on Transco's operating costs for the 2002-03 to 2006-07 Price Control Period', prepared for Ofgem, September 7th.

For the price control review covering the period 2013 to 2021, Ofgem has proposed to use a combination of top-down (limited to disaggregated unit cost benchmarking) and bottom-up analyses within their toolkit approach to assess the efficiency of the transmission companies' expenditure projections.[64]

Gas and electricity distribution

Ofgem has made extensive use of process benchmarking during price control reviews. For the last electricity distribution review, it commissioned external experts to undertake bottom-up comparisons of IT and property management costs incurred by electricity DNOs. It also benchmarked back-office costs for the gas distribution companies as part of the 2008–13 gas distribution price control review. This included comparisons of property management, HR, finance and regulation and legal costs across the companies, and against external comparators. The analysis was based on data collected by Ofgem in the Business Plan Questionnaire. Meetings with companies helped to ensure an adequate understanding of the data and the activities undertaken by each function. To make the costs of each company comparable with external benchmarks, the support service costs were normalised using relevant cost drivers, such as the number of FTEs. Third-party benchmarks included consultants' reports, electricity distribution companies and companies of a similar size. The gas distribution companies were benchmarked based on OPEX, number of FTEs and revenue. However, Ofgem placed less weight on the OPEX and FTE-based metrics, since these could be distorted by business strategies such as outsourcing or renting property, and by accounting policies.

For the current price review in terms of gas and forthcoming review in the case of electricity distribution companies, Ofgem has proposed to use a combination of top-down and bottom-up analyses within its toolkit approach to assess the efficiency of the companies' expenditure projections, including process benchmarking similar to previous reviews.[65]

Transport

In the transport sector, OPEX at Dublin Airport was assessed in 2009 by benchmarking individual activities using process-level data from other international airports to come up with a conservative and challenging case.[66] A bottom-up approach was adopted following consultation with stakeholders, who favoured this methodology over a top-down approach. As a result of this analysis, the regulator proposed a reduction in FTEs in several areas at the airport, as well as other non-payroll cost savings. External experts were commissioned to assess the expected operational costs of a new terminal. To this end, they examined expected volumes and other drivers of operating costs, combined with assumptions and pay data from external benchmarks.

For the price review of BAA, the UK CAA commissioned benchmarking of the operator's corporate functional costs, which included an assessment of individual processes. BAA was benchmarked against external comparators in Europe and the USA.[67] Qualitative assessments of the efficiency of particular processes including payroll were also undertaken.

As part of its 2003 interim review of Network Rail's access charges, the ORR examined controllable OPEX of £641m incurred at National Rail's headquarters, and some zonal/regional administrative and support-type expenditure.[68] A number of benchmarks from outside the rail sector were used, including utility companies, previous benchmarking surveys, and international audit office data. Where external process benchmarking was not

[64] Ofgem (2011), 'Decision on strategy for the next transmission price control - RIIO-T1 Tools for cost assessment', March.
[65] Ofgem (2011), 'Decision on strategy for the next distribution price control - RIIO-GD1 Tools for cost assessment', March.
[66] CAR (2009), 'Final Determination – Dublin airport charges 2010 -14', p. 49 and Table 6.2, p. 51; and Indecon-Jacobs (2009), 'Bottom-Up Efficiency Assessment of DAA/Dublin Airport OPEX: Final Report', p. 40.
[67] Booz Allen Hamilton (2006), 'Supporting Paper III: Airport efficiency assessment—overview and summary', December; KPMG (2006), 'Scrutiny of BAA Plc's IT costs', December.
[68] Oxera (2003), 'Interim review of Network Rail's track access charges', June, available at http://www.oxera.com/main.aspx?id=1429&&user=9204.

possible because some of the activities were specific to the rail industry, internal benchmarking of performance was used.

For the price control review covering 2013–18, the ORR has proposed to benchmark Network Rail using econometric and bottom-up approaches. Under bottom-up approaches, the ORR has stated that it will review Network Rail's bottom-up calculations of how it has justified its expenditure in its business plan—for example, its planned volumes of work—and proposes to benchmark Network Rail's unit costs for specific activities against other companies in Britain or overseas.[69]

4.1.2 Advantages and disadvantages

Widely used across the regulated industries, process benchmarking can provide useful insight into the performance of a regulated company on specific activities. Furthermore, many of the benchmarks for the more common support functions are publicly available, although identifying like-for-like companies and collating the required data can be time-consuming. Process benchmarking typically uses a number of metrics to assess each category, which provides a cross-check and improves the reliability of the results.

Process benchmarking using external comparators is often limited to corporate functions, however, and cannot easily be applied to assessing activities specific to a company, or where the costs depend largely on the company's operating circumstances. In such instances, international benchmarks would be required, which could be difficult to source. Another disadvantage of process benchmarking is that sometimes the external benchmarks can be distorted by accounting policies, and by business decisions such as outsourcing (data quality issues); in such circumstances the overall robustness of the approach is likely to be greatly diminished.

4.1.3 Assessment of process benchmarking

Table 4.1 summarises how process benchmarking measures against the general and GTS applicability criteria listed in section 2.

Table 4.1 Assessment of process benchmarking against criteria

	Assessment
General criteria	
Complexity and transparency	It is conceptually simple to calculate benchmarks based on ratios and unit costs, although the application of this approach can be difficult since the costs of the assessed company would first need to be allocated in a comparable process/activity format. Process benchmarking can be transparent, provided that the relevant benchmarks are publicly available and the cost allocation process is explained and justified. However, if the processes adopted by specialist consultancies when calculating such benchmarks are relatively opaque, this could in turn be a source of criticism from the assessed company. A possible risk could be that the assessed company could dispute the relevance of the external benchmarks, although this could be pre-emptively managed if the NMa ensures that the benchmarks come from reputable sources only.

[69] ORR (2011), 'Establishing Network Rail's efficient expenditure', July.

	Assessment
Reliability	Process benchmarking requires some degree of judgement both when choosing the relevant external benchmarks and in designing the cost allocation framework for the function to be benchmarked. It is difficult to assess whether the most appropriate benchmarks and cost allocation processes have been selected (since one cannot rely on an impartial statistical framework), but sensitivity analysis can increase overall robustness by examining how each decision affects the overall results. The use of a wide range of benchmarks and metrics could make this approach reliable and relatively flexible for benchmarking corporate functions, which is one of the reasons why it has seen widespread use. Comparison difficulties might hinder the use of this approach when benchmarking other activities, unless relevant external benchmarks are available or can be estimated based on expert technical advice.
Suitability for catch-up and/or frontier-shift efficiency	Process benchmarking can provide a catch-up efficiency estimate for the functions assessed. It is less suitable for estimating frontier shift.
GTS applicability criteria	
Data availability	Process benchmarking requires more effort to ensure comparability than many other benchmarking approaches. Benchmarking support or administrative functions should be relatively straightforward, since external comparators are likely to be readily available. However, such external benchmarks are expected to be more difficult to find for the functions that are unique to gas transmission, where the costs are dependent on operating circumstances or specific assets.
Integration	Process benchmarking can be used to derive cost-reduction targets for specific operations of GTS.
Implementation time	The implementation depends in the first instance on access to the costs of the various functions of the assessed company and the ability to express these in process/activity form. If such a distinction is not possible using GTS's accounts, the NMa would need to establish a new accounting/cost allocation framework for this purpose, which would add to the implementation time. The regulator would also need to solicit specific operational information from GTS, such as employee numbers or certain KPIs, to enable comparisons to be made. Process benchmarking is usually employed to assess the efficiency of support functions, since such external comparators are readily available. However, these external benchmarks are likely to be more difficult to find for the functions that are unique to the assessed company, where the costs are dependent on operating circumstances or specific assets. Identifying and collating such information would require additional resources by the NMa. Transmission-specific benchmarks could be available from consortia of transmission companies that undertake transmission specific benchmarking (see section 5.2). The overall implementation time of this exercise is difficult to estimate since it will depend on the effort required to put the necessary accounting/cost allocation framework in place and to source the necessary external benchmarks. The feasibility of the approach is likely to depend on data availability, the scope of the cost base and the availability of external benchmarks.
Impact on other aspects	Process benchmarking can provide only estimates of catch-up efficiency which can be used to adjust the tariffs. An additional adjustment in the form of frontier shift may need to be estimated separately. Also, it is generally applied on an OPEX rather than a TOTEX basis.

Source: Oxera.

4.2 LRIC modelling

LRIC are the total costs that can be avoided by stopping the production of a product, service or network element given that all other products, services and network elements are still being provided. Alternatively, LRIC can be defined as the additional costs that would be incurred in providing a product, service or network element given that all other products, services and network elements are currently being provided.

In economic theory, a competitive market would ensure that prices for products and services are bid down to the marginal costs of production,[70] leading to efficient market outcomes—ie, allocative, productive and dynamic efficiency. A regulator setting prices for the outputs of network companies may thus be able to mimic the outcomes of a competitive market by setting prices equal to marginal costs.

In network industries, however, the presence of significant fixed costs means that marginal cost is very small (potentially close to zero for small output increments). Hence, pricing down to this level would mean that many efficiently incurred costs would not be recovered. The concept of LRIC could be interpreted as a proxy for marginal costs in network industries. LRIC costs include all volume-sensitive costs associated with an increment, the fixed costs directly attributable to the production of the increment, and the increase in the common costs (overheads) that can be attributed to the increment.[71]

At a high level, there are two types of LRIC models.

– **Top-down models** take accounting information as the primary data source and calculate the costs of relevant increments (products, services or network elements) by applying cost–volume relationships linking these increments with predefined cost categories from the accounts. These cost–volume relationships in essence define the levels of efficient costs for the processes and assets included in the model and are sourced externally, mainly using engineering expertise or economic/econometric studies or interviews and field work. [72]

– **Bottom-up models** adopt an economic-engineering approach starting from the demand for the product, service or network element defined as the increment and then building an efficient network that can address this demand. The costs that correspond to the components of the model are also sourced externally, similar to the top-down LRIC models.

Another key dimension along which both types of model can vary is the assumption around the network topology. This is a question about whether to take, as a starting point, the existing network topology with the existing pipelines, compression stations, and other relevant assets of a gas transmission network—**the 'brownfield' approach**—or whether to model a fully efficient ideal network topology from scratch—**the 'greenfield' approach**. The question of which approach to choose is part of a more general point relating to the treatment of efficiency in the existing network configuration.

LRIC models have been used in the past mainly to calculate wholesale access charges and to assess cost-reflective pricing. A regulator could use these costs as the basis for setting future prices or revenue since, according to theory, LRIC aim to reflect the costs that a company would have incurred if it were operating in a competitive environment. However, as far as Oxera understands, LRIC models have not been used in an efficiency assessment context, the adoption of this approach could be considered an original application of LRIC models.

4.2.1 Applications in the regulatory setting

Telecoms

In the telecoms sector, a number of regulators have used LRIC-style models to estimate termination rates. For example, the French, Hungarian, Israeli and Dutch telecoms regulators

[70] Marginal cost is the cost of producing a single unit of output.
[71] When common costs that are not directly attributable to an activity are added to the estimates of LRIC, it forms a LRIC+ (LRIC plus) estimate.
[72] In modelling the cost–volume relationships, consideration needs to be given to the treatment of inefficiency. For example, if some inefficiencies were removed from the current network, the shape of a cost–volume relationship might change from that currently observed.

use a LRIC+ model[73] to assess mobile termination rates, while the Norwegian, Danish and Lithuanian regulators use a long-run average incremental cost (LRAIC) model.[74] Many of these models have been constructed by the same technical experts, indicating that standard approaches and assumptions are used.

In 1997, the predecessor of Ofcom, Oftel, constructed a bottom-up model to calculate the costs of an efficient operator. BT also did its own top-down work based on the actual interconnection costs. Work was undertaken to reconcile the results from both models and to establish whether factors other than efficiency could explain the gap. This included differences in utilisation rates, the cost of capital, and investment costs for a given piece of capital equipment.[75]

Energy

In the energy sector, electricity distribution companies in the UK have developed costing models to estimate cost-reflective distribution use-of-system charges. The approach used is not consistent across the companies, although a number of them estimate the incremental cost of increasing maximum demand by 500MW. Use-of-system charges for electricity transmission in the UK have also been established based on the marginal cost of investment in the transmission system required to meet the increased demand or generation for each node of that system. This model reflects the impact of users of the transmission network on costs at different locations, and takes into account a locational security factor.[76] Work is ongoing to agree a common methodology across the distribution companies.

Aviation

The UK CAA considered the use of a LRIC model to calculate the cost of constructing Terminal 5 at Heathrow Airport. This would have taken into account the OPEX and CAPEX of the new terminal and the expected increase in passengers. However, this approach was abandoned since there were concerns that it would increase prices in the short run if the incremental costs of Terminal 5 were higher than for the current facilities, while it might not allow full recovery of investment.

The CAA also considered using LRAIC to set price caps at Stansted Airport. A number of issues meant that this approach was not adopted, however. These included uncertainty about what the increment should be, difficulty in forecasting the long-term costs of assets with lifespans of 40–50 years, difficulty in identifying comparators for the increment, and concerns that regulatory commitment would be questioned if there were a move away from a regulatory asset base (RAB) approach to LRAIC.[77]

4.2.2 Advantages and disadvantages

There is significant regulatory precedent on the use of LRIC-based models for establishing wholesale access charges and assessing prices.[78] However, Oxera has not been able to find examples where LRIC models have been used to set allowed revenue, although the UK Competition Commission did consider this option for the regulation of Stansted Airport.

The main advantages of constructing a LRIC model is that it provides a detailed view of the operations undertaken by the assessed company—indeed, the level of detail afforded by a LRIC model can be matched only by a very detailed reference model, although LRIC models are likely to be even more ‘complete’ because they seek to take into account all long-run

[73] The difference between LRIC and LRIC+ is that, whereas the former includes no common costs, the latter does include an allocation for common costs. LRIC is thus a measure of the long-run incremental cost of a service, whereas LRIC+ includes all of these costs and, additionally, makes allowance for the recovery of common costs.
[74] LRAIC models avoid price fluctuations by using average incremental costs.
[75] Oftel (1997), ‘Pricing of telecommunications services from 1997: Oftel’s Proposals for Price Control and Fair Trading’, available at http://www.ofcom.org.uk/static/archive/oftel/publications/1995_98/pricing/pri1997b/chap4.htm.
[76] For more information, see National Grid (2011), ‘CUSC – Section 14: Charging Methodologies’, January.
[77] Competition Commission (2008), ‘Stansted Airport Ltd, Q5 price control review’, Appendix C.
[78] As well as the example sectors listed in this section, LRMC (long-run marginal cost) modelling has been adopted by Ofwat to determine prices for access to the water supply for large users. See http://www.ofwat.gov.uk/regulating/reporting/pap_tec_lrmca.pdf.

costs currently being incurred by the assessed company. Another advantage of LRIC modelling is that it can reduce the reliance on information supplied by the assessed company; this is more relevant with the bottom-up LRIC model, which calculates LRIC based on expert economic/engineering advice (ie, an external benchmark), rather than on the basis of the assessed company's costs and investment plans (which is the focus of the top-down LRIC models).

The major disadvantage of adopting a LRIC modelling approach is that the implementation can be quite complex and take time. LRIC models also require significant regulatory judgement when selecting the features and inputs of the model. This includes elements such as:

– the definition of the increment;
– the definition of 'long run'—over what period incremental costs should be calculated;
– how the required forecasts that determine future costs should be estimated;
– what constitutes as efficient costs for the process and assets that are included in the model.

These questions all rely in part on the judgement of the regulator and can thus be open to criticism from the assessed company. These issues could become more pronounced if the bottom-up LRIC model is adopted and the analysis does not take into account the current costs of the assessed company when designing and calibrating the model.

4.2.3 Assessment of LRIC

Table 4.2 summarises how LRIC modelling measures against the general and GTS applicability criteria listed in section 2.

Table 4.2 Assessment of LRIC modelling against criteria

	Assessment
General criteria	
Complexity and transparency	LRIC models are probably the most complex of any of the models discussed in this report. Provided that the models themselves and the underlying data are made available, the approach can be transparent. That said, the models, mainly under the top-down approach, may require information that the regulated company would like to keep confidential from third parties, making this approach less transparent to any interested third party. In addition, transparency could be reduced if some inputs for the model are derived from external sources that do not provide sufficiently detailed explanations on how such inputs were calculated.
Reliability	The extensive use of this approach in the regulatory setting provides an indication of its perceived reliability, at least when assessing prices. However, it may be significantly more difficult to construct a reliable model for establishing allowed revenue. In addition, the models require quite a few assumptions on future costs and how these should be discounted. Also, the models need to use external benchmarks to derive efficient costs for the processes and assets included in the model. Despite these shortcomings, LRIC models probably provide the most in-depth view of the activities undertaken by the assessed company, as well as their likely interactions, relative to all other approaches examined in this report. Also, given their modular nature, LRIC models are quite flexible and in principle could be modified to include any additional parameters that might be deemed important in a future assessment.
Suitability for catch-up and/or frontier-shift efficiency	The model cannot directly estimate the scope for the frontier shift; rather, the rate of future efficiency is an *input* to the model, so that it can properly estimate the LRIC. As such, if the NMa were to adopt this approach, additional analysis that uses a different approach would be required to estimate the frontier shift.

	Assessment
GTS applicability criteria	
Data availability	Under a top-down approach, most of the data required for this approach is related to the regulated company and should be made available to the NMa. Information from direct comparators would be required for benchmarking and informing aspects of the LRIC that require judgement. Availability of such data can vary.
Integration	LRIC models are used mostly for assessing prices. As a result, it is not clear how these models, and their outcomes, could be easily integrated into GTS's regulatory framework. LRIC models can be designed to assess OPEX and CAPEX separately or together, as in TOTEX.
Implementation time	Given the overall complexity of the approach, designing a LRIC model can require a significant amount of time and resources, making this arguably the most costly/time-intensive option considered in this report. In addition, a number of issues need to be resolved at all stages of the process. LRIC models could also require frequent updating to take into account changes in input prices, operating circumstances or obligations, and technology. This updating process is also likely to increase the implementation cost of the approach. The overall cost could be reduced (but still remain considerable) if the NMa could identify a standardised model for gas transmission or distribution that could be modified for the purpose of benchmarking GTS.
Impact on other aspects	The model cannot directly estimate the scope for the frontier shift; rather, the rate of future efficiency is an *input* to the model, so that it can properly estimate the LRIC. As such, if the NMa were to adopt this approach, additional analysis that uses a different approach would be required to estimate the frontier shift. The estimated efficient costs by the model can be used to inform the tariff adjustment required on the current cost of GTS. LRIC models can be designed to assess OPEX and CAPEX separately or together, as in TOTEX.

Source: Oxera.

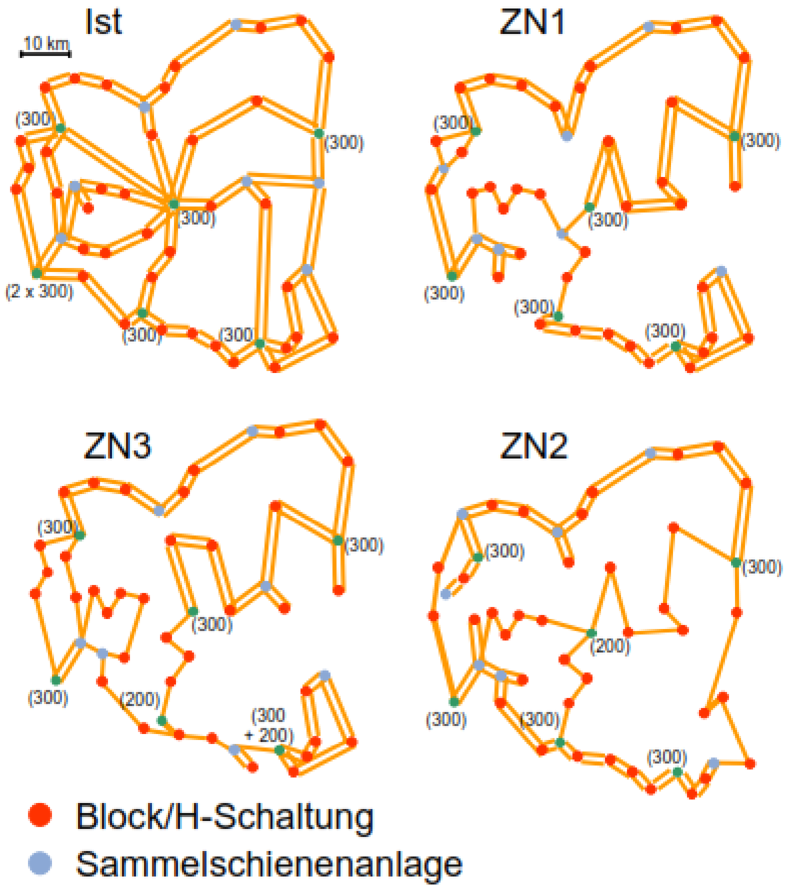
4.3 Reference model

Reference models are similar in their principles to LRIC models—one could even consider them as a sub-category of the more general LRIC modelling approach. However, since these two approaches focus on different areas, they are presented separately in this study. The LRIC model provides a more detailed and 'complete' view of the assessed company, while the reference model approach described below deals primarily with capital investment in the network, although such a model could be expanded in order to assess OPEX as well, as discussed in 4.3.1.

A reference model is an abstracted simulation of a transmission (or distribution) network. It starts with placing the points of supply and demand in space and attempts to optimise the process of transmission from supply points to demand points. This optimisation process can be completely unconstrained or constrained by the structure of the existing network. The first instance, the **greenfield approach**, assumes that the assessed company is building a network from new, based on the latest technology and optimal design, but with the same capacity, volumes and quality as the existing network. It does not consider the existing topology of the network. The second option, the **brownfield approach**, assumes that the assessed company uses the existing lines and substations, but when the network needs to expand or be modified for any reason, the most optimal network design derived from the reference model is chosen. Since the option of replacing the existing network and starting anew is uneconomical in most cases, pure greenfield models are not often seen in practice, at least not without developing a brownfield model as a sense-check.

In designing a reference model, the starting point is to design the reference network by linking supply and demand. This is often done using a mathematical simulation model that searches for optimal network designs that would reduce overall costs. In other cases, the reference network can be based on a more high-level assessment, which takes into account technical and legal requirements. Figure 4.1 below provides an example of the first approach, taken from the reference model adopted by BnetzA.

Figure 4.1 Optimisation examples offered by the reference model approach



Source: BnetzA (2007), 'Vorstellung der Referenznetzanalyse'.

In the above figure, red dots represent circuit switches, blue dots represent busbars and green dots represent transformers of differing capacities. The yellow lines are transmission power lines. The diagram denoted '1st' is the initial solution of the model, which would require approximately €16m to implement. The other three network structures represent optimised alternatives, and would each cost approximately 24% less to build. The overall build costs were estimated based on the network structure (as per Figure 4.1) and asset unit costs that were sourced externally.

The model described here focuses on capital investment. As the following case studies demonstrate, reference models could also be designed with the aim of assessing OPEX or TOTEX.

4.3.1 Applications in the regulatory setting

Energy

One of the first modern applications of this approach was by the Chilean National Energy Regulator, which used a reference model to establish efficient TOTEX for electricity distribution companies.[79] Based on the typical characteristics of an average operator, the model company was assumed to face the same restrictions and environment as the actual

[79] See Pollitt, M. (2004), 'Electricity Reform in Chile: Lessons for Developing Countries', MIT Center for Energy and Environmental Policy Research', September, Working paper.

companies. Information from the assessed company was collected and validated. Next, the structure and dimensions of the hypothetical efficient company were defined. The different categories of *efficient* costs were then allocated to high-voltage, low-voltage and customers. These efficient (benchmark) costs were estimated externally based on engineering or economic expert knowledge. Finally, the optimal costs were extracted from the reference model, including operational and maintenance costs, fixed costs, and distribution losses, and any special circumstances were identified.

Although the above reference model is quite complex and takes several factors into consideration, there have been a number of criticisms of this approach:

– there was disagreement about the level of assumed efficient costs, which were based on engineering assumptions rather than actual cost data;
– the model cannot be directly applied to heterogeneous companies;
– as the approach was based on five zone types, including rural and high density, companies operating in similar zones were set the same efficiency target regardless of their actual level of efficiency.[80]

In Sweden, the energy regulator constructed a greenfield reference network for each electricity distribution area based on engineering and economic analysis, which was used to assess the efficient levels of both OPEX and CAPEX required to service that area. A quality of service indicator was included in the model, and allowed revenue was reduced if actual interruptions exceed expected interruption costs. The overall assessment is based on whether the assessed companies' actual costs exceed by a certain margin those derived from the reference model. One issue identified in relation to this approach was that the model does not take into account future changes in demand, and may therefore underestimate the investment needs of the network.[81]

In Spain, a reference model has been constructed to assess the total costs of the distribution companies. This model was designed such that it could accommodate load growth, quality of service incentives and efficiency assumptions. It considers geographical and topographical constraints as well as peak loads. Data on the number of customers, their location and their annual consumption is required to populate the model. The model then minimises the costs of operating the specified network and constraints, based on external benchmarks. The model can design an optimal distribution network, or the replacement cost of the existing network.[82]

The reference model designed by the German regulator is in some ways simpler than the reference models detailed above, since its aim was to assess only the capital investment of electricity and gas transmission companies.[83] Its high-level principles have already been discussed above. However, just because the BnetzA reference model is simpler than the others detailed in this section, this does not mean that it has less merit. Indeed, it would be possible that the resulting estimates of the BnetzA model are more robust than the overall assessments derived from the more complex reference models, mainly because the BnetzA model focuses on the area where this approach excels—ie, the design and planning of the network. For the more extensive reference models to provide an estimate of efficient operating and support costs, they need to rely on external benchmarks and assumptions on how different factors (eg, customer density) affect costs; as such, the overall robustness of their results depends in large part on the robustness of these benchmarks and assumptions.

[80] For more information, see Rudnick, H. and Donoso, J. (2000), 'Integration of Price Cap and Yardstick Competition Schemes in Electrical Distribution Regulation', *IEEE Transaction on Power Systems*, **15**:4, November.
[81] See Viljainen, S., Tahvanainen, K., Lassila, J., Honkapuro, S. and Partanen, P. (2004), 'Regulation of electricity distribution business', Proceedings of the 6th Nordic conference on Nordic Distribution and Asset Management (NORDAC), Esbo, Finland.
[82] See Martinez, A.C. (2007), 'From reference network models to economics ones: first approach to network business total cost function', July, working paper.
[83] Although the BnetzA reference model is simpler in terms of its focus, it is quite complex in how the reference network is designed—ie, in the adopted optimisation methodology.

4.3.2 Advantages and disadvantages

Some of the advantages and disadvantages of reference models have already been discussed in the case studies above. In summary, possibly their main advantage is that they can be used to benchmark a natural monopoly without relying on direct, external comparators, while providing a detailed view of the functions and assets of the assessed company, similar to the LRIC model. A reference model is also likely to provide the most robust estimates of the need for additional capital investment and the likely cost of such investment, assuming that the underlying CAPEX unit costs are also accurate.

The main disadvantage of the reference models is that they are not easy to implement. They require significant resources to construct, are likely to involve several iterations to agree assumptions and methodologies between the relevant stakeholders; and require regular modifications to account for new technologies and changes in industry structure. In addition, to derive efficient underlying costs, they have to rely on external benchmarks, be they CAPEX unit costs, process benchmarks or simpler cost–volume relationships, which may not be fully transparent.

Concerns were also expressed that the reference model approach does not allow for innovation in the regulated companies. Since the companies know that they will be judged according to the reference model, they are likely to base all their future decisions for expanding and maintaining the network on how favourable or unfavourable these would be reflected in the model itself. As a result, the reference model may penalise a company that is transitioning from an older specification that is incorporated in the model to a more advanced/experimental specification that the model does not support.

4.3.3 Assessment of reference models

Table 4.3 summarises how reference models measure against the general and GTS applicability criteria listed in section 2.

Table 4.3 Assessment of the reference model approach against criteria

	Assessment
General criteria	
Complexity and transparency	The degree of complexity depends on the number and complexity of the assets and processes to be modelled. As with LRIC models, provided that the models themselves and the underlying data are made available, the approach can be transparent. However, transparency may be impaired if the model relies on assumptions that are themselves derived in an opaque fashion (eg, when the process of generating the required CAPEX unit costs is itself opaque). Again, as with LRIC models, constructing a reference model may require information that GTS would consider to be confidential, making this approach less transparent to any relevant third party.

	Assessment
Reliability	A well-specified reference model will be robust if it accurately reflects the processes and assets of the company. As with the LRIC model, a degree of judgement on the part of the regulator/consultant is required when constructing the model. This would depend on how complex the model is designed to be. Another factor that has a large impact on overall robustness is the quality of the external benchmarks that are required to estimate the overall cost of the reference network. The robustness of the resulting cost-reduction targets is also dependent on whether a green- or brownfield model is adopted. The latter is considered more equitable, but would not provide strong incentives for redesigning the network in a more efficient manner. Again, as with LRIC, a reference model can provide a detailed view of the operations of the regulated companies and their likely future costs. Changes that are more common, such as extending the network or installing extra capacity to accommodate greater demand or supply, should be easily accommodated by the reference model, by construction. More far-reaching changes, such as significant changes in the assessed companies' required outputs owing to new environmental legislation, might be more difficult to incorporate. However, this will depend on how the reference model is designed; a more modular approach is likely to be more useful, allowing the current model to take such changes into account.
Suitability for catch-up and/or frontier-shift efficiency	The reference model provides an estimate of the costs that an efficient network should be incurring currently, or the costs of implementing an efficient capital investment, although the approach can be extended to cover TOTEX. As such, the outputs of the model can be used to measure catch-up efficiency or OPEX or TOTEX. The model can incorporate an element of frontier shift in its evaluation, but, as with some of the LRIC models, this is an *input* to the model and therefore has to be estimated externally.
GTS applicability criteria	
Data availability	The main advantage of this approach is that it can be used to benchmark a natural monopoly without relying on significant data from external comparators, while providing a detailed view of the functions and the assets of the assessed company. On the other hand, the reference model requires considerable resources from the assessed company, which may not be available promptly enough when there are time constraints.
Integration	In the past regulators have used a hypothetical efficient company to measure the efficiency of the assessed operator. So, in principle, the catch-up estimates obtained through a reference model could be used to regulate GTS.
Implementation time	The overall implementation cost of constructing a reference model would depend on the complexity of the model in question. If it is designed to assess the network at a fine level of detail, the time required for this approach could be considerable. Some assistance might be required from external experts with engineering and economic background and with some prior experience in such modelling. The resource cost of this approach is likely to be substantial. It would involve collating the necessary inputs for the simulation model, which includes mapping the supply and demand points and creating a database of standardised unit costs in order to evaluate the asset options; designing the optimisation model and allocating the necessary cost for operating; and maintaining the network to its constituent parts. This last part can be omitted if the focus of the model is to assess capital investment, as is the case with the BnetzA model. Nevertheless, the implementation of this approach is likely to be costly and time-consuming.
Impact on other aspects	Depending on the complexity of the model, the reference model is likely to provide estimates of the need for additional capital investment and the likely cost of such investment, although the approach can be extended to cover TOTEX. If limited to CAPEX, a separate assessment of OPEX static efficiency and the frontier-shift adjustment may be required.

Source: Oxera.

4.4 Assessing unit costs in capital expenditure

This approach, usually referred to as 'CAPEX unit costs', has seen widespread adoption in the UK, especially when assessing the proposed future capital investment of regulated companies. It can also be applied ex post, to assess a company's performance in delivering the investment, and to ascertain whether the actual investment was justified. In more detail, the approach involves assessing whether:

– the proposed or realised investment is justified—ie, whether the company has made a correct (or at least justified) decision on, for example, whether it is cost beneficial to replace 100km minimum of pipes per annum or 50km (ie, what is the efficient level of activity?);

– the investment was efficiently delivered, by looking at the unit cost of the activities undertaken to deliver the investment.

On the first aspect, a number of methods are available, usually revolving around models of future demand and models for assessing the performance of current assets (asset replacement models). When the proposed investment is large and/or complex, additional analysis can be undertaken to ascertain the need for the investment; this usually includes a detailed cost–benefit analysis (CBA) and the consideration of possible alternatives.

The second aspect can involve a number of professional disciplines as varied as quantity surveying, contract design, and econometrics (for commodity price forecasting).[84] The approach is very similar in nature to the RUOE analysis described in section 3.6, except that instead of undertaking the analysis based on a top-down view of the company—ie, assessing OPEX as a whole—CAPEX unit costs rely on more disaggregated information; namely, the unit costs of assets and a standardised set of activities relating to the maintenance or replacement of such assets.

As presented in sections 4.2 and 4.3, CAPEX unit costs can also be an important input in the LRIC and reference models.

Rather than considering the varied theoretical basis for unit cost capital benchmarking, this section summarises its use in the water, rail and energy sectors. For the last of these, Oxera has approached a global engineering consultancy and obtained examples of how it has employed these techniques in the regulatory field.

4.4.1 Applications in the regulatory setting

Energy

CAPEX benchmarking has wide application in the energy sector. In the most recent periodic review of the regional electricity distribution companies in the UK, Ofgem undertook a number of projects to assess the efficiency of the companies' future investment, including:

– **load-related modelling**, to understand the impact of increasing demand on the network and how this affects future investment;
– **non-load-related modelling**, to look at the impact of factors not related to demand, such as asset age and condition, or the effects of distributed generation, and how these might affect future investment;
– **unit cost assessment**, which included developing a set of unit costs for all assets being installed on the electricity networks. This used information from existing unit cost data supplied by the distribution companies, cost data obtained from current or recent projects undertaken overseas, suppliers' published price lists, and industry indexes.

[84] For example, see Ofwat (2008), 'Cost base feedback report: August 2008', and Competition Commission (2000), 'Mid Kent Water Plc: A report on the references under sections 12 and 14 of the Water Industry Act 1991', chapter 6, paras 6.129–6.139 and Appendix 6.6, paras 22–33.

– **Review of specific large schemes**, detailing investment need; appraisal of schemes including consideration of alternatives and CBA; and requirements and measures for investment triggers.

– **Development of output measures**, which involved the development of suitable measures to quantify the outputs of network investment, and of appropriate tables for the collection of this data.

Ofgem has undertaken similar assessments in the past for both gas and electricity transmission, and is adopting similar approaches in the current gas and electricity transmission price control review. More specifically, it is looking to use disaggregated benchmarking approaches between the companies (mainly unit cost analysis), unit cost analysis based on engineering expertise, variance analysis, age-based modelling and spot checks on selected schemes.[85]

Water

Unit cost benchmarking is well-established in the water sector and, in England and Wales, has underpinned the regulatory process for decades. It is overseen by 'Reporters', who essentially provide an engineering auditor service. CAPEX unit costs can range from the general, such as the cost of replacing a 10m pipe, to the very detailed; an example of such a detailed definition, as used by Ofwat, is given in Box 4.1.

Box 4.1 Unit cost definition and guidance provided by Ofwat

Standard cost estimate: mains laying nominal bore 450mm
Estimated cost per metre of a 450mm nominal bore pipe laid, assuming that depth of cover to the mains is 900mm to the crown of the pipe, in the specified locations—grassland, rural/suburban highway and urban highway.

Standard costs shall represent the average unit costs for work that complies with the standard cost specification.

– Assume no unusual ground conditions. Omit de-watering, soil stabilisation, deep foundations, rafts, piling, special ground support, ground anchors and excavation in rock.
– Assume an average non-complex amount of trench water pumping and typical requirements for diversion of services. Include for non-complex geotechnical investigations.
– Assume excavated materials are not contaminated and can be used to refill trenches in field/verge.
– Assume trenches are refilled with granular material in highways and excavated material disposed of to a landfill tip 1km distant.
– Assume all necessary working space and areas for storage of materials are readily available at nominal cost.
– Assume open trench laying consistent with the AMP2 specification.
– Allow for the costs associated with the implication of NRSWA 1991 but exclude the lane rental costs as defined by Street Works (Charges for Occupation of the Highway) (England) Regulations 2001.
– Include costs required to comply with relevant Health and Safety regulations.
– Project management costs to be included in standard cost estimates are shown in the accompanying checklist. Note that a general adjustment to the standard costs can be made to cover the management costs above items. Adjustments for tender-outturn margin and contingencies should be the same as normal levels assumed in companies' investment projections.'

Source: Ofwat (2007), 'PR09 Business Plan Information Requirements: Cost Base'.

[85] Ofgem (2011), 'Decision on strategy for the next transmission price control - RIIO-T1 Tools for cost assessment', March.

However, the difficulties in undertaking like-for-like comparisons in capital unit cost benchmarking have been highlighted in some reports.[86]

Rail

In the UK, the ORR also employs independent Reporters to examine the stewardship of rail assets by Network Rail:

> The role of the Reporter is to provide ORR with professional advice on the quality of Network Rail's provision, as specified in their licence. This provides ORR with an independent review of Network Rail's performance and stewardship of the network. ORR uses this review to assure Network Rail's performance on an ongoing basis, in our capacity as the industry regulator.[87]

Motivated by the safety issues that surfaced post-privatisation, the remit of the Reporters extends far beyond unit cost benchmarking, and there are three strands to rail reporting:

– review Network Rail's data to validate its accuracy and reliability;
– review and assess Network Rail's business planning and asset management;
– support the ORR's monitoring of Network Rail's enhancement projects.

Notwithstanding the broad remit, unit cost benchmarking is an important part of the task, as can be seen in a 2011 report by Network Rail.[88]

4.4.2 Advantages and disadvantages

As evidenced above, unit cost benchmarking has broad application in a number of sectors.[89] As with the majority of the bottom-up approaches discussed in this section, the main advantage of CAPEX unit cost assessment is that it may not require data on direct comparators, although if CAPEX unit cost data is available from a number of comparators, it can be used to increase the confidence of the benchmarks. Ofwat compares the regulated water companies' CAPEX unit costs against each other to derive the benchmark value that it includes in its database. Also similar to the other bottom-up approaches, it can provide a highly detailed view of the functions and asset base of the assessed company. CAPEX unit costs are also quite modular, in that they can be as detailed or as high-level as the regulator wishes or deems appropriate; this can have a significant impact on the overall cost of implementing this type of benchmarking approach.

The main disadvantage of relying on CAPEX unit costs is that their overall reliability is quite dependent on the level of detail. Box 4.1 above provided an indication of what Ofwat considers to be a sufficient level of detail, but, even here, there are still concerns that this level of detail is insufficient to provide like-for-like comparisons. Another feature of CAPEX unit costs is that the relevant unit cost benchmarks have to be provided or at least validated by an external source, usually by engineering experts, increasing the implementation time and cost.

4.4.3 Assessment of unit CAPEX approaches

Table 4.4 summarises how the unit CAPEX approach measures against the general and GTS applicability criteria listed in section 2.

[86] See Jacobs (2009), 'Findings from the Cost Base Audits (March-May 2009)', a report for Ofwat, and Jacobs (2007), 'Advice on capital works unit costs (Cost Base)', a report for Ofwat.
[87] From the ORR's website on Independent Reports, http://www.rail-reg.gov.uk/server/show/nav.147.
[88] Network Rail's response to the Nichol's report, sourced from http://www.rail-reg.gov.uk/upload/pdf/nr_response_nichols_report.pdf.
[89] In addition to the sectors detailed above, this approach has been proposed to assess efficiency in the provision of road infrastructure and large, publicly procured civil infrastructure projects. See http://www.hm-treasury.gov.uk/d/nationalinfrastructureplan251010.pdf and http://www.hm-treasury.gov.uk/iuk_cost_review_interim_report.htm.

Table 4.4 Assessment of the unit cost CAPEX approach against criteria

	Assessment
General criteria	
Complexity and transparency	The degree of complexity depends on the level of scrutiny that the regulator adopts in its assessment. Greater scrutiny might require more complicated underlying models, but would also be likely to increase the robustness of the results. The transparency of this approach depends on how involved GTS would be in constructing the framework necessary for this type of assessment. Usually, assessment includes many rounds of consultation with the assessed company to ensure that all matters relevant to the assessment are discussed and taken into consideration.
Reliability	The approach has been widely adopted by regulators for assessing not only efficiency of, but also the need for, capital investments. Overall, the robustness of this approach would depend on the quality of the technical input and the availability of comparable data. As with both the LRIC and the reference model approaches examined above, expert engineering assistance is critical, and the robustness of the approach would be significantly affected by the quality of this assistance. Owing to its open-ended nature, this approach is quite flexible. The unit cost database would need to be updated frequently to provide an accurate reflection of the actual costs that the regulated companies are likely to face. Large projects are usually assessed according to a CBA, which can be modified to accommodate any likely changes in the regulated companies' environment or required outputs.
Suitability for catch-up and/or frontier-shift efficiency	CAPEX unit costs represent the 'standardised'/efficient costs required to deliver a unit of capital investment with the current technology; as such, they can be used to derive an estimate of overall catch-up efficiency for CAPEX. As with all bottom-up approaches, forward-looking cost reductions in unit costs can be included in the valuation process, but would need to be sourced externally.
GTS applicability criteria	
Data availability	The data required for the creation of the relevant unit costs relies on engineering practice, and is likely to be available from the assessed company.
Integration	CAPEX unit costs approach can provide a catch-up estimate on the current CAPEX of GTS, but would require separate assessment of OPEX catch-up efficiency and the frontier-shift adjustment to adjust the tariffs.
Implementation time	The implementation time of this approach would depend to a large extent on the level of scrutiny that the NMa chooses to adopt. The cost of an assessment that examines all proposed capital investments would probably be substantial. However, this level of detail might not be necessary; instead, greater emphasis could be placed on those projects that are sufficiently large and/or deemed to be of importance. Additionally, separate assessment of OPEX catch-up efficiency and the frontier-shift adjustment may need to be carried out.
Impact on other aspects	To adjust the tariffs would require a separate assessment of OPEX catch-up efficiency and the frontier-shift adjustment.

Source: Oxera.

5 Assessment of approaches

A variety of approaches, both top-down and bottom-up, have been reviewed in this report, and these can be used to estimate GTS's efficiency. This section summarises how these approaches perform relative to the assessment criteria discussed in section 2. In addition, a high-level comparison of these methods is provided below.

5.1 General criteria

5.1.1 Complexity and transparency

All top-down approaches are similarly transparent, which means that a knowledgeable outside party can verify the findings using the same dataset, assuming that full disclosure is possible of both the underlying data and the assessment methodology used to derive the results. However, full disclosure of underlying data for assessments that rely on direct comparators (namely, the frontier-based approaches) might not be possible if the analysis uses confidential information limiting the transparency of the approach.[90] The TFP approach does not suffer from similar transparency issues, given that data necessary for its implementation is publicly available. RUOE trends can be similarly transparent, assuming that the data on the network utilities that serve as comparators can be made publicly available.

In the past, the NMa has used frontier-based approaches with international data for TenneT. These approaches rely on principles that should be fairly straightforward to implement; especially COLS. Extensions to DEA and SFA are likely to be quite complex and require specialist software to apply. On the contrary, TFP and RUOE are considered relatively simple approaches.

The bottom-up approaches reviewed here can be characterised as transparent on the whole, assuming that both the underlying data and the models used to estimate the potential for cost reductions are made available. That said, the LRIC (especially, under the top-down approach) and reference models may require information that GTS would like to keep confidential, which would impair the transparency of these approaches to an interested third party.

With regard to complexity, the process benchmarking is probably conceptually the simplest to implement of all bottom-up approaches. Nevertheless, allocating the costs of the assessed company in a comparable process/activity format can be difficult. As far as the CAPEX unit costs approach is concerned, the degree of complexity would depend on the level of scrutiny that the NMa chooses to adopt. Reference and LRIC models are the most complex of the approaches considered in this report, as they are likely to involve a significant amount of data and require several assumptions to be made.

5.1.2 Reliability

With regard to reliability, the top-down approaches that use direct comparators are considered more robust in general than those that rely on indirect comparisons such as TFP and RUOE trends, because they base their evaluation on information from companies that undertake the same, or very similar, functions as the assessed company. As such, fewer adjustments are necessary, both to the underlying data and in the modelling/assessment process. Another feature of these approaches is that, in general, they all base their assessments on actual, realised performance. These two features can result in an objective

[90] CBb (the Trade and Industry Appeals Tribunal) has recently ruled that the regulator is not obliged to provide full access of third-party confidential information to a transmission company.

assessment, which does not rely on expert knowledge of the industry or regulatory judgement.

The above discussion does not mean that a top-down assessment can be undertaken in a mechanistic matter, without any direct inputs from the regulator or other industry experts. The frontier approaches offer much flexibility and can be modified to address issues specific to the assessment at hand, such as the need to undertake total cost modelling, owing to the significant variation in capital asset bases in the comparator set, or to incorporate a measure of quality of service in the analysis. Both of these can be controlled for directly at the modelling stage, provided that consistent information regarding them is available for all comparators and that the number of comparators is sufficient to include all these factors in the analysis.

Having enough comparators for the purposes of the analysis is a major factor that influences the reliability of frontier-based approaches in general (see discussions in section 3.1–3.3). Data quality is also important for the overall reliability of all approaches that use direct or indirect comparators in the assessment, be they top-down or bottom-up.

As for TFP and RUOE trends, both should be considered less reliable in general relative to their frontier-based counterparts because they both rely on indirect comparators. This is not to say that such approaches will always be less robust; there could be cases where the available data is of such poor quality, and/or the number of comparators so few, that even though frontier-based approaches are feasible in theory, the regulator would be better served by using the alternative top-down methods that may use more reliable data.

For the bottom-up approaches, process benchmarking requires some degree of judgement about the choice of external benchmarks and the design of the necessary cost allocation framework. However, extensive sensitivity analysis can mitigate this weakness and increase robustness by examining how these decisions affect the overall results. In addition, the overall reliability of the approach can be strengthened through the use of a wide range of metrics to inform the potential for overall cost reductions.

It is difficult to assess at this time the expected reliability of a LRIC model, both because a successful application of this approach would need to resolve a number of issues (as discussed in section 4.2), and because, to Oxera's knowledge, there are no examples of such an approach being applied in this context (the latter suggests a risk in its adoption). Nevertheless, a LRIC model has the potential to provide a detailed view of the activities undertaken by the assessed company, as well as their likely interactions, and to greatly increase the NMa's understanding of the industry.

A reference model, such as a LRIC model, could provide a detailed description of the NMa's activities and related costs, although this would depend on how detailed the model is. Reference models also require some degree of judgement and could thus be open to criticism, especially in view of the fact that the benchmarks required for the valuation of the reference network need to be sourced externally, for example based on expert engineering advice. Although such models can be used to assess total costs, they could also be designed with a more limited scope in mind—namely, to assess future capital investment. Such more limited reference models might actually be more robust, given that they are less complex and require fewer assumptions. Lastly, the reliability of cost-reduction estimates derived from a reference model would depend on whether a green- or brownfield model were adopted. The latter is considered more equitable, but would not provide strong incentives for redesigning the network in a more efficient manner.

CAPEX unit cost models have recently been applied by a number of regulators, and are likely to be both reliable and flexible. The overall robustness depends critically on the data availability and external benchmarks for assessing efficiency.

In order for the assessment approach to be reliable, it needs to be sufficiently flexible to accommodate any possible future changes to GTS's activities and its required outputs. Potentially, the most flexible top-down approaches are the frontier-based ones that can readily incorporate both additional inputs and external factors in the analysis. Furthermore, all bottom-down approaches can be modified to include additional parameters that might be deemed important in a future assessment.

To summarise, assessing the likely relative robustness of the bottom-up approaches is quite difficult at this stage, especially given the issues regarding the construction of the LRIC model and the overall complexity inherent in designing a reference model. Process benchmarking for support activities and unit cost CAPEX analysis could be considered 'safer', given the wider adoption of these approaches in the regulatory setting. However, as noted above, the other two bottom-up models could be quite flexible and could provide significant insights into GTS's operations and resulting costs.

5.1.3 Suitability for catch-up and/or frontier-shift efficiency

It is important that the adopted approach is suitable for assessing the GTS's static or relative efficiency. Furthermore, if an alternative approach is used in conjunction to evaluate the frontier-shift, the two methods must be compatible.

The frontier-based approaches examined in this report are capable of assessing relative efficiency. They can also estimate the frontier shift, provided that panel data is available.

One of the main disadvantages of the other two top-down approaches (TFP and RUOE trend) is that they provide an estimate of the overall productivity growth, without making clear what proportion of productivity gains is attributable to relative efficiency and what to frontier-shift improvements. Decomposing the two elements would need to rely on external evidence, such as academic studies that have examined the issue of decomposition (or recourse to frontier-based or bottom-up approaches). As this approach requires assumptions that might be open to third-party criticism, it is advised that sensitivity analyses be carried out to verify these results.

All the bottom-up approaches reviewed in this report are unsuitable for estimating a frontier-shift. Process benchmarking can provide a catch-up efficiency estimate for the assessed functions. The LRIC and the reference modelling approaches are also unsuitable for directly estimating the scope for future improvement; rather, the rate of frontier shift is an *input* to the models. As a result, an additional approach needs to be used in combination with either of the two methods to estimate the frontier shift separately. As with all the other bottom-up approaches, CAPEX unit cost analysis can produce only an estimate of the catch-up efficiency and by definition is limited to CAPEX.

5.2 GTS applicability criteria

5.2.1 Data availability

One of most important criteria is data availability—whatever the relative advantages of an approach, if the appropriate data is not accessible in the given timeframe, it will be of little use to the NMa. There appear to be a few international benchmarking studies of transmission companies involving possibly mature datasets. This section provides a selective review of international benchmarking studies on transmission companies by the Sumicsid Group in project ECOM+ and e^{3}GRID, the International Transmission Operation and Maintenance Study (ITOMS), the International Comparison of Transmission System Operation (ICTSO), the Gas Transmission Benchmarking Initiative (GTBI) and the EPRG benchmarking study on European gas transmission companies for CEER in 2006.

A group of European energy regulators (in Austria, Denmark, Finland, the Netherlands, Norway and Sweden) joined forces to develop international benchmarking of electricity transmission companies by Sumicsid in 2004–05. The benchmarking was based on a unit cost model including OPEX, normalised CAPEX and size of grid as the driver, although extension of the model to accommodate different returns to scale using DEA was also discussed.[91] ECOM+ has evolved since into the e^{3}GRID benchmarking model. E^{3}GRID is a regulatory benchmarking of European electricity TSOs on behalf of CEER Workstream Incentive-based Regulation and Efficiency benchmarking (WS EFB) in 2008 by the Sumicsid Group, van Dijk Management Consultants and Tractebel Engineering SA. The benchmarking involved participation of 19 national regulatory authorities in Europe and 22 TSOs.[92] The e^{3}GRID project involved more electricity TSOs, new and improved data definitions and scope of activities. It also relaxed the limitations with unit cost benchmarking by allowing for both increasing and constant returns to scale in the cost function and included cost drivers such as connection density and capacity for renewable energy, apart from grid size.

ITOMS is a consortium of international transmission companies that work together with the UMS Group to compare performance and practices and identify transmission industry best practice worldwide.[93] Since 1994, the UMS Group has provided programme management and analysis for the consortium of approximately 40 TSOs with comparators drawn from North America, Europe, Scandinavia and Asia Pacific.[94] The biannual programme collects data on the transmission companies and analyses maintenance and operation costs and service levels. It appears that data is collected on some quality and safety measures, and OPEX and CAPEX on the TSOs, and that benchmarking is undertaken based on KPIs.[95] However, it is not clear whether the comparator set was limited to electricity transmission companies or included both gas and electricity transmission companies.

The ICTSO programme was initiated by the National Grid Company in the UK in 1994. Currently, the group is managed by a steering committee with six selected members and is supported by KEMA Consulting. It includes 20 transmission companies spanning Asia, Europe, Africa, the South Pacific, and North and South America.[96] A database with transmission companies' system data and performance measures since 2000 appears to be available for members of the group. Based on this database, benchmarking studies are available for the members. These studies compare number of staff and the costs of the different transmission companies, considering their most important cost drivers, such as size and activity level.[97]

GTBI is a consortium of eight gas transmission companies in Europe initiated in 2004 and supported by the Juran Institute.[98] The benchmarking study appears to cover the main operations and maintenance aspects of gas transmission operation, along with non-commercial system operation elements, including network spending and staffing; compressor spending and staffing; dispatching spending and staffing; support costs and staffing; and emissions, safety and reliability.[99] As with other studies discussed in this section (ITOMS, ICTSO, ECOM+ and E^{3}Grid), the results and membership of the GTBI study are protected by confidentiality agreements, and cannot be shared outside of the membership. As part of the 2007–11 transmission review, Ofgem's consultants carried out international benchmarking of National Grid against four European gas transmission companies using data from ITOMS and GTBI.[100]

The EPRG was commissioned by CEER to develop a framework for benchmarking of European gas transmission companies.[101] The final report of the study, published in 2006,

[91] Sumicsid (2005), 'ECOM+ project - benchmarking of transmission system operators', October.
[92] Sumicsid (2009), 'International Benchmarking of Electricity Transmission System Operators', March.
[93] UMS Group (2010), 'Measuring the Benefits of Asset Management', December.
[94] See http://www.umsgroup.com/partnerforums/itoms.asp.
[95] UMS Group (2010), 'Measuring the Benefits of Asset Management', December, pp. 2–3.
[96] Dicaprio, A., Garrote, J.S., Carter, A., Russell, D., Franken, B. and Scarci, G.C. (2004), '"Real Time Operation" Multidimensional benchmarking of 20 TSOs', August.
[97] Franken, B., Rogge, M., Mak, C.M., Albassam, L.A. and Thijssen, G. (2005), 'Benchmarking System Operation Processes for 20 international Transmission System Operators', August.
[98] See http://www.juran.com/industries_oil_and_gas_descriptions.html, accessed April 23rd 2012.
[99] National Grid Gas Transmission (2011), 'Innovation, efficiency and value for money' July, pp. 39–40.
[100] TPA Solutions (2006), ' Transmission Price Control Review 2007-2011 - Efficiency Study and Forecast Opex', Final draft report, September, pp. 54–55.
[101] Electricity Policy Research Group (2006), op. cit.

used a sample of about 40 US interstate gas transmission companies as well as four European operators. Benchmarking was undertaken using three frontier-based techniques: DEA, SFA and COLS. The report observed that, while Europe as a whole has sufficient numbers of comparators for benchmarking, the lack of readily available and well-documented data was the reason for including US comparators in the study. The report also observed that comparability of data was a key issue, along with data standardisation, not only because the European companies operated differently from the US companies, but also because there were substantial differences within the European companies. To overcome these issues, the report suggested employing stochastic techniques such as SFA, as measurement errors cannot be avoided entirely.

The US data on 43 companies was taken from annual regulatory reports filed by the energy utilities to FERC; and the European data on the four companies was obtained from the relevant companies or estimated by the regulator. Based on the analysis, the report made the following observations with respect to data and analysis:[102]

- long-term commitment to build appropriate databases. This would involve agreeing on functional/cost reporting boundaries for the transmission companies and standardising all data (cost measures for exchange rate and inflation);
- in the case of insufficient numbers of European comparators, US data from FERC could be an option, with direct communication with the regulator to reduce comparability issues;
- it might be possible to produce a relative ranking of the European companies even without high comparability between US and European firms;
- under incentive regulation, a few high-quality strategic variables might be sufficient to obtain robust X factors;
- revenue is highly correlated with the expenditure measures and produces similar efficiency measures across firms;
- measures of capacity and pipeline network length could be sufficient drivers of gas transmission companies' costs and revenues;
- gas transmission operations can be relatively easily modelled using econometric approaches, the specific dataset would need to be fine-tuned before processing to produce robust results;
- in terms of DEA, the use of weight restrictions should be explored to incorporate the relative importance of particular inputs or outputs;
- revenue should be an important regulatory focus area next to cost measures, given that ratepayers fund the utilities' activities in a regulated industry.

Other gas transmission-specific benchmarking reports and academic articles include the following.

- Analyses of the productivity of the US transmission industry following the deregulation of well-head prices in the USA, by Sickles and Streitwieser (1991). The input variables considered included fuel volume, total labour expenditure, transmission pipeline capital, capacity utilisation, compressor station capital, and throughout. Delivery volume multiplied by average miles transported was used as the single output variable. Both DEA and SFA were used to assess the productivity of the transmission companies.[103]

- An assessment of an open-access transportation programme of US gas transmission utilities, in Granderson (2000).[104] The input and output variables considered in the study were the same as in Sickles and Streitwieser (1991). An SFA model was used to estimate the efficiency levels of the companies.

[102] Ibid, pp. 66–67.
[103] Sickles, R.C. and Streitwieser, M.L. (1991). 'Technical inefficiency and productive decline in the U.S. interstate natural gas industry under the natural gas policy act', Working paper.
[104] Granderson, G. (2000). 'Regulation, Open-Access Transportation, and Productive Efficiency', *Review of Industrial Organization*, **16**, pp. 368–85.

– Benchmarking of Australian and New Zealand gas transmission and distribution utilities, by Meyrick & Associates in 2004.[105] The cost measures considered included operation and maintenance expenditure and CAPEX, and the cost drivers included throughput and capacity. The benchmarking technique used was DEA.

DEA, COLS and SFA all require data on direct comparators. These comparators can be external, such as other international gas transmission companies or internal, such as the regional distribution companies. The overall robustness of the estimates from these approaches is dependent on the underlying data. Ideally, the data should be free of measurement error (noise),[106] have enough comparators, and include the factors that significantly influence costs. Each of the frontier approaches (DEA, COLS and SFA) requires the same data in the form of costs and cost drivers. Tables 3.2 and 3.3 give examples of the sort of data required; however, the approaches do differ in terms of the amount of data required for estimation purposes. Indicative rules of thumb are as follows.

– For DEA, the literature has suggested a variety of rules of thumb to relate the number of companies and the number of inputs and outputs included in the model, as the discriminatory power of DEA reduces as more variables are included in the model compared with the number of companies. For example, it is suggested that the number of companies needs to be at least twice the product of the number of inputs and the number of outputs, or at least three times the sum of the inputs and outputs. Moreover, it has been suggested that if more than one-third of the companies are deemed efficient by the DEA model, its discriminatory power is considered to be poor.[107]

– For COLS, models with as few as eight observations over three years and 21 observations over a single period have been used in regulation.[108]

– For SFA, models with fewer than 30–40 observations can fail to produce results.

As stated above, Oxera understands that the dataset on comparators for GTS will need to consist of gas transmission companies in other countries. An assessment that relies on international comparators would need to ensure comparability of the benchmarked companies and require some form of data standardisation in order to ensure robust, like-for-like comparisons. The NMa would therefore need to assess which type of comparator, US- or European-based, is similar enough to allow it to separate efficiency from structural differences before collating the necessary data. This process is crucial because international comparators might use different definitions for costs and cost drivers, have different corporate structures (eg, they might undertake both transmission and distribution activities), or have different boundaries between the transmission and distribution function. It is possible that the process of data normalisation could introduce an additional element of measurement error into the data, which could reduce the overall accuracy of the resulting estimates.

The issues of data quality and availability, however, cannot be determined prior to the analysis. A judgement can be formed only once appropriate comparators have been identified and the data has been analysed. From the above review, in terms of benchmarking of transmission companies, the focus appears to be more on electricity transmission companies, with US data from the FERC predominately being used. In terms of benchmarking gas transmission companies in Europe, the availability and maturity of data are likely to be key constraints given that there are relatively few comprehensive

[105] Meyrick & Associates (2004). 'Comparative Benchmarking of Gas Networks in Australia and New Zealand', New Zealand Commerce Commission.
[106] Although DEA has been shown to provide accurate estimates under conditions of moderate measurement error.
[107] For example, see Cooper, W.W., Seiford, L.M. and Zhu, J. (2011), *Handbook on Data Envelopment Analysis*', Second Edition, Springer Science.
[108] COLS has been used by Ofgem based on data for eight distribution companies over a two-year period. See Ofgem (2011), 'Decision on strategy for the next gas distribution price control - RIIO-GD1 Tools for cost assessment', March, para 4.23. In its 2008–09 assessment of water companies, Ofwat applied COLS to cross-sectional data on 21 water companies. See Ofwat (2009) 'Relative efficiency assessment 2008-09 – supporting information', December.

benchmarking studies (eg, the 2006 report by the EPRG for the CEER, which included data on only four gas transmission companies in Europe). The remaining top-down approaches discussed in this report are TFP and RUOE analysis. TFP analysis requires data on aggregate sectors of the economy, which is readily available from a variety of sources. While simple unit comparisons of levels require data on direct comparators, RUOE trend analysis requires data on other network utilities that in theory would not be hard to acquire, although it is not clear at this stage whether such data is available for the various Dutch utilities. Nevertheless, such data can be gathered from other countries' utility companies, such as those in the UK.

Data availability is unlikely to be an issue for process benchmarking to assess the efficiency of support functions (ie, overheads or headquarter costs). However, at this time it is not clear whether TSO-specific functions could also be assessed, since many studies of TSO-specific benchmarking are confidential, although GTS may be able to provide evidence to the NMa on its efficiency levels if it has taken part in such studies. The LRIC modelling (in particular, the bottom-up method), CAPEX unit cost and reference model approaches do not have to rely on data from external comparators, but require considerable resources from the assessed operator, which may not be made available to the regulator, as GTS may consider some of them confidential.

5.2.2 Integration

It should be possible to integrate the output from all the frontier-based approaches and the unit cost comparisons of levels in the regulation of GTS, as the approaches produce clear efficiency targets that can be applied to cost projections. That said, the remaining top-down approaches, TFP and RUOE trend analysis, are unable to provide an estimate of GTS's current levels of relative efficiency; rather, their aim is to provide an estimate of the scope for future cost reductions based on the performance of a composite benchmark.

Bottom-up approaches could be expected to satisfy this criterion more effectively, although this is likely to be dependent on data availability, the scope of the cost base in the benchmarking exercise and the availability of external benchmarks. Process benchmarking, which tends to focus on OPEX, together with a unit CAPEX approach can be used to derive cost-reduction targets for the OPEX- and CAPEX-specific operations of GTS.

Regulators have used reference models with hypothetical efficient companies in the past to measure the efficiency of the assessed operator. So, in principle, a reference model could be used to regulate GTS. LRIC models, as far as Oxera understands, have not been used to assess efficiency. As a result, it is not clear whether their outcomes could be integrated into the GTS regulatory framework.

5.2.3 Implementation time

Assuming that data on comparable European or US comparators is available, implementing a *high-level* catch-up efficiency method before the start of the next regulatory period in 2014 is expected to be feasible for many of the approaches reviewed in this report. Exceptions include *comprehensive* frontier-based approaches where data collation and validation is required, and LRIC and reference models which require very detailed data from the regulated company. In general, the burden to the regulator is likely to be greater if the analysis is based on international comparisons, owing to the need to organise the data collation process and make the necessary adjustments to ensure comparability between the international comparators and GTS. Any proposed method therefore needs to be implementable, at least in intermediate mode, and produce a preliminary cost-reduction target for GTS before April 2013, so that it can be fully applied by the start of the new regulatory period in January 2014. Given Ofgem's experience in RIIO-T1, it does not appear that one year is a long enough to collate and analyse a full international dataset. It may be possible to undertake a high-level (ie, non mechanistic) analysis using data on US transmission companies from FERC following EPRG (2006), or to use data on German gas transmission companies, as this would only require adjusting GTS data to be consistent with the US or German dataset.

With regard to implementation time, the frontier-based approaches are relatively easy to apply, at least in their 'standard' form, with no extensions or refinements. Although the implementation time is likely to be modest provided that the necessary know-how is available, the time needed to collate the data and standardise the process could be substantial, especially if regional data is unavailable. SFA, by requiring a much larger dataset than DEA and COLS, is generally considered to be the hardest to implement in a given timeframe. However, once the data becomes available, the additional time needed to implement either DEA or COLS is likely to be small.

TFP analysis, as described in this report, is likely to be the least time-intensive, and thus the least costly, of all the approaches discussed here, including the various bottom-up ones. It is relatively easy to implement, and the required data is publicly available. RUOE analysis is also likely to be quite straightforward to implement, but a significant part of the overall cost associated with this approach relates to data collation and standardisation. Since it is unclear at this stage whether such data is actually available from the Dutch network utilities, and in what form, it is difficult to assess the likely resource requirements of this approach. Also, RUOE focuses on OPEX only rather than TOTEX, although it can be used to assess TOTEX; otherwise, CAPEX may need to assessed and reimbursed separately. Also, similar to TFP analysis, RUOE trend analysis measures overall productivity growth and cannot directly distinguish between catch-up efficiency and frontier shift.

All the bottom-up approaches are likely to require significant implementation time since they all have substantial set-up costs. Possibly the least costly approach would be process benchmarking of support activities, although designing the cost allocation framework required to make the comparisons, and testing it to ensure that comparisons are like-for-like, could be a lengthy and involved process. As discussed earlier, for gas transmission-specific functions, there are likely to be issues of confidentiality and with the availability of reliable external benchmarks.

It is expected that the most time-intensive approaches would be those that employ either a LRIC or a reference model for the assessment; a less detailed reference model could provide some time savings, but might also reduce the robustness of the resulting estimates. The time required to adopt a LRIC modelling approach is not easy to determine for the two reasons already noted: designing the model is not straightforward; and, to Oxera's knowledge, LRIC models have not yet been applied by a regulator in an efficiency assessment context. It is expected that research would be required to determine how to construct a LRIC model for a gas transmission company with the specific goal of assessing its costs.

Lastly, the time needed to implement a CAPEX unit cost analysis depends on the level of scrutiny that the NMa wishes to adopt and the availability of data and external benchmarks. The time needed to examine all proposed capital investments would be likely to be substantial. Such level of detail might not be necessary, however; instead, greater emphasis could be placed on those projects that are sufficiently large and/or deemed to be of importance. Also, the model can estimate a catch-up efficiency only on the current CAPEX of GTS, and additional analysis would be required to assess OPEX and estimate the frontier-shift adjustment.

5.2.4 Impact on other aspects of the regulation of GTS

It is important to consider whether each approach can be implemented without making significant changes to the current method of TOTEX regulation. In addition, the NMa has stated that it may no longer need to have an additional reimbursement during the regulatory period for all expansion investments, depending on how the benchmark is applied. Critically, the NMa will need to examine carefully what expansion investments are included in the benchmarking exercise and how these compare to GTS's future expansion investments.

Ideally, any performance assessment exercise of infrastructure companies needs to take into consideration both OPEX and CAPEX. When both types of expenditure are included in the same modelling process, the resulting process is usually referred to as total cost modelling.

In total cost modelling, the target measure could be a measure of TOTEX, or OPEX and a measure of capital, which could be CAPEX. These can be analysed separately, with any interactions between them accounted for during a second stage; or they can be incorporated separately into the same model if the selected approach can accommodate multiple inputs and outputs, as is the case with DEA for example. The main advantage of total cost modelling is that it avoids biases arising from the misallocation of costs between CAPEX and OPEX. However, it is often difficult to create a single, unified measure of the CAPEX used by the assessed companies.[109]

There are typically two approaches to incorporating capital costs in the TOTEX benchmarking: the capital cost (TOTEX) approach and the capital stock (total cost) approach. Both approaches were considered in the benchmarking work in EPRG (2006). Under the former, the total cost is simply the sum of the outturn OPEX and CAPEX. This has the advantage that it is simple to compute and the resultant measure of TOTEX is easy to understand. However, it does not consider the long-lived nature of assets; rather, it takes a snapshot of the typically lumpy capital costs during a single year. Also, as the investment timing and accounting policies of companies can vary, even if they invest at similar levels, the costs can fluctuate and the resulting benchmark figures can be misleading. To offset this, a rolling average of TOTEX can be considered or the sum of operating costs and average network CAPEX. Under the total cost approach, the long-lived nature of assets is recognised and an estimate of the annual capital consumption is arrived at by converting a stream of annual investment figures into a stock of capital on which a return is paid and a stream of annual depreciation figures—ie, annual capital consumption is equal to an annual depreciation charge plus a return on capital. While this could be a realistic approach to measuring the CAPEX part of TOTEX, it requires a range of assumptions, adding to the regulatory burden in terms of collecting additional data. Both TOTEX and total cost approaches have been applied in the UK and Europe in different sectors, as can be seen in Table 5.1 below.

[109] Oxera (2003), 'Total Cost Modelling', *The Utilities Journal*, pp. 42–3.

Table 5.1 Applications of total cost modelling

Sector	Application
Water	In a 1996 report for Ofwat, Bosworth, Stoneman and Thanassoulis considered the TOTEX modelling approach for measuring comparative total efficiency in the sewerage and water industry.[1] The Norwegian Water and Energy Directorate has developed a TOTEX benchmarking approach using DEA
Energy	Ofgem's original total cost modelling (in DPCR3) involved modelling OPEX plus a ten-year average CAPEX.[2] Similarly, in DPCR4 Ofgem considered several versions of total cost analysis, differing mainly in how the CAPEX part of TOTEX was accounted for.[3] In DPCR5, Ofgem also acknowledged the trade-off between OPEX and CAPEX and partly accounted for this by including in its OPEX benchmarking non-load-related CAPEX (the average non-load-related CAPEX for the period 2005/06 to 2014/15, as submitted by the companies), considering that this approach best captured the trade-off.[4] In its October 2010 report, Ofgem announced that, in future price control reviews (RIIO-GD1, RIIO-T1 and RIIO-ED), the base revenue requirements of the gas and electricity transmission and distribution companies would be assessed according to an efficient level of TOTEX (OPEX + normalised CAPEX) under the RIIO framework.[5] Regulators such as the Norwegian Water and Energy Directorate and the DTe in the Netherlands have also moved towards TOTEX benchmarking. Both NVE and DTe use DEA as the main benchmarking technique. Other examples include energy regulation in Austria, Belgium and Finland.[6]
Transport	Railtrack's 2000 periodic review of track access charges was based on total operating, maintenance and renewal expenditure. In the econometric modelling, two total cost functions were employed: one in which the dependent variables were total way-and-structures expenses, and the other in which depreciation was discounted from the total expenses (way + structures). In PR08, the ORR used a number of approaches to examine Network Rail's efficiency. Two of particular relevance here are the total cost econometric model and the TFPapproach.[7] Oxera reviewed the approaches employed by ITS Leeds and provided some alternatives together with recommendations.[8]
Communications	In 2001, Ofcom examined BT's efficiency in terms of total costs (operating costs, depreciation and returns).[9] A subsequent study for Ofcom carried out an efficiency assessment of BT Openreach's activities following a total cost approach—ie, the analysis focused exclusively on total cost computed as OPEX + depreciation + cost of capital.[10]

Source: [1] Bosworth, D., Stoneman P., and Thanassoulis E. (1996), 'The Measurement of Comparative Total Efficiency in the Sewerage and Water Industry: An Exploratory Study', October. Ofwat (1999), 'Review of Public Electricity Suppliers 1998–2000, Distribution Price Control Review', Consultation Paper, May. [2] Ofgem (1999), 'Review of Public Electricity Suppliers 1998-2000, Distribution Price Control Review', Consultation Paper, May, pp. 36–7. [3] Ofgem (2004), 'Electricity distribution Price Control Review Final Proposals', November, section 7.23. [4] Ofgem (2009), 'Electricity Distribution Price Control Review: Final Proposals - Allowed revenue - Cost assessment appendix', December. [5] Ofgem (2010), 'Handbook for implementing the RIIO model', October, chapter 8. [6] Haney, A.B. and Pollitt, M. (2009), 'Efficiency Analysis of Energy Networks: An international survey of regulators, CWPE & EPRG'; DTe (2000), 'Choice of model and availability of data for the efficiency analysis of Dutch network and supply business in the electricity sector', background report accompanying 'Guidelines for price cap regulation in the Dutch electricity sector', a report by Frontier Economics, February; SUMICSID (2009), 'International Benchmarking of Electricity Transmission System Operators', e3GRID project, final report for CEER. [7] NERA (2000), 'Review of overseas railway efficiency', a report for the Office of the Rail Regulator, July. [8] ITS Leeds (2008), 'International benchmarking of Network Rail's maintenance and renewal costs: an econometric study based on the LICB dataset', October, section 5.3, p. 23. [9] Oxera (2009), 'Recommendations on how to model efficiency for future price reviews', November.[10] NERA (2001), 'The profitability and efficiency of the UK mobile network operators', a report for Oftel, August; NERA (2008), 'The comparative efficiency of BT Openreach', a report for Ofcom, March.

From the above discussion and applications by different regulators in different sectors, it appears that both the total cost and TOTEX approaches have their particular merits and limitations. In addition, there are data-consistency issues with either approach—for example, a change in capitalisation policy by a company could result in a distortion of benchmarking results; similarly, the number of years chosen to normalise capital costs. Oxera notes that many of the issues are empirical in nature and necessitate exploration of different modelling approaches to develop robust solutions.

Under the frontier-based approaches, if the focus of assessment is TOTEX, it should be possible to integrate the output from them and the unit cost comparisons of TOTEX levels in the regulation of GTS, as the approaches produce clear efficiency targets that can be applied to the total cost projections. The remaining top-down approaches, TFP and RUOE trend analysis (the latter extended to cover TOTEX), are unable to provide an estimate of GTS's current levels of TOTEX relative efficiency; rather, their aim is to provide an estimate of the scope for future total cost reductions based on the performance of a composite benchmark.

Among the bottom-up approaches, process benchmarking tends to focus on OPEX, but together with a unit CAPEX approach can be used to derive cost-reduction targets for the OPEX- and CAPEX-specific operations of GTS. As discussed earlier, regulators have used reference models with hypothetical efficient companies in the past to measure the total cost efficiency of the assessed operator. So, in principle, a reference model could be used to regulate GTS. LRIC models, as far as Oxera understands, have not been used to assess efficiency. As a result, it is not clear whether their outcomes could be integrated into the GTS regulatory framework.

5.3 Summary and overall ranking of the approaches

It is difficult to assess with any great precision how the approaches reviewed in this report rank relative to each other, according to the criteria set by NMa. This is mainly due to a number of uncertainties about the required underlying data, and issues relating to implementation time. However, ranking the approaches based on a simpler, relative grading system with the inclusion of the necessary caveats is possible, and likely to be helpful to the NMa when shortlisting the approaches. Such a ranking is provided in Table 5.2, which ranks each approach from A to D (A denoting the highest ranking and D the lowest) against the general and the applicability criteria for GTS. It should be stressed that these grades are relative; an approach ranked A for cost requirements does not mean that is three times less costly to implement than an approach ranked C.

It should also be noted again that the given rankings are meant solely as an approximate indicator of relative performance; each approach has advantages and disadvantages, and the decision on which approach is likely to be more suited in an assessment greatly depends on the specific details of the assessment being undertaken.

Table 5.2 Relative rankings of the reviewed approaches based on the assessment criteria

	Top-down			Bottom-up			
	Frontier-based approaches	**TFP analysis**	**RUOE analysis**	**Process benchmarking**	**LRIC model**	**Reference model**	**CAPEX unit cost analysis**
General criteria							
Complexity and transparency	**A–C**, depending on the chosen approach	**A**	**A**, assuming that there is no issue about confidential information	**A–C**, depending on the overall transparency of external benchmarks	**D**	**C–D**, depending on the level of complexity	**A–B**, assuming that there is no issue about confidential information
Reliability	**A–C**, depending on the chosen approach and the type of comparator (for example, European only or European and US comparators)	**D**	**C**	**A–B**, for support functions. For gas transmission-specific functions, there are likely to be issues with the availability of reliable external benchmarks	**Unclear**, owing to the required assumptions and approach adopted	**B–C**, depending on the level of complexity	**B–C**, depending on the level of complexity and the quality of expert advice sought
Suitability for static and/or dynamic efficiency	**A**, if consistent data over time is available, it is possible to estimate catch-up and dynamic efficiency separately	**C**, measures total scope for efficiency saving although secondary sources could be used to decompose this measure into catch-up and frontier-shift estimates if required	**A–B**, if consistent data over time is available, unit cost *trends* measure overall productivity growth, thereby including catch-up and frontier shift, while unit cost *levels* could be used to provide an estimate of relative efficiency and thus catch-up efficiency	**C**, just catch-up; secondary source is needed to estimate frontier shift	**C**, just catch-up; secondary source is needed to estimate frontier shift	**C**, just catch-up; secondary source is needed to estimate frontier shift	**C**, just catch-up estimate on the current CAPEX of GTS; secondary source is needed to estimate catch-up on OPEX and frontier shift

	Top-down			Bottom-up			
	Frontier-based approaches	**TFP analysis**	**RUOE analysis**	**Process benchmarking**	**LRIC model**	**Reference model**	**CAPEX unit cost analysis**
GTS applicability criteria							
Data availability	**C–D**, data availability from other European gas transmission companies and data consistency are key constraints, although a high-level comparison may be possible with non-European gas (eg, US) transmission companies and transmission companies in Germany	**A**	**A–D**, comparison using direct comparators (other European gas transmission companies) could be limited owing to data constraints, although comparison with indirect comparators could be possible	**C–D**, for gas transmission-specific functions as there are likely to be issues of confidentiality and with the availability of reliable external benchmarks **A**, for support functions	**A–B**, highly dependent on the type of LRIC model used, accessibility of the gas transmission company's data and availability of information from direct comparators, which would be required to inform aspects of the LRIC that require judgement	**A–B**, highly dependent on accessibility of the gas transmission company's data	**A–B**, highly dependent on accessibility of the gas transmission company's data
Integration	**A–B**, depending on the chosen approach, the type of comparator (eg, European only, or European and US comparators), and the quality and quantity of the dataset (eg, data over time required to estimate frontier shift)	**B–C**, the tariffs could be adjusted for the overall productivity growth estimated by the approach	**A–C**, by definition RUOE focuses on OPEX, but it could be extended to cover TOTEX. Otherwise CAPEX may need to be assessed and reimbursed separately. While RUOE unit cost trends measure overall productivity growth, thereby including catch-up and frontier shift, RUOE unit cost levels could be used to provide an estimate of relative	**B–C**, dependent on the scope of the cost base and the availability of external benchmarks. Would require separate assessment of the frontier-shift adjustment to adjust the tariffs	**Unclear**, owing to absence of precedent	**C–D**, can provide a catch-up estimate on the current CAPEX of GTS, but it could be extended to cover TOTEX. If limited to CAPEX, a separate assessment would be required of OPEX catch-up efficiency and the frontier-shift adjustment to adjust the tariffs	**C–D**, can provide a catch-up estimate on the current CAPEX of GTS. Would require a separate assessment of OPEX catch-up efficiency and the frontier-shift adjustment to adjust the tariffs

	Top-down			Bottom-up			
	Frontier-based approaches	TFP analysis	RUOE analysis	Process benchmarking	LRIC model	Reference model	CAPEX unit cost analysis
			efficiency and thus catch-up efficiency				
Implementation time	**B–D**, depending on the approach. Intrinsically linked to data availability (see above).	**A**	**A**	**B–C**	**C–D**, depending on the level of complexity	**C–D**, depending on the level of complexity	**C–D**, depending on the level of complexity
GTS applicability criteria							
Impact on other aspects of the regulation for GTS	**A–C**, the NMa's current method of GTS regulation bases tariffs on actual costs. The tariffs could be adjusted to remove the estimated static inefficiency of GTS identified by the method(s) and applying a frontier-shift adjustment over the period	**A–B**, the tariffs could be adjusted for the overall productivity growth estimated by the approach. However, limitations remain in terms of decomposing the overall measure into catch-up efficiency and frontier-shift improvements, although secondary sources could be used to facilitate this, if required	**A–C**, by definition RUOE focuses on OPEX, but it could be extended to cover TOTEX. Otherwise CAPEX may need to be assessed and reimbursed separately. While RUOE unit cost trends measure overall productivity growth, thereby including catch-up and frontier shift, RUOE unit cost levels could be used to provide an estimate of catch-up efficiency	**C–D**, can provide only estimates of catch-up efficiency which can be used to adjust the tariffs. However, additional adjustment in the form of frontier shift may need to be estimated separately	**C–D**, the model cannot directly estimate the scope for frontier shift; as such, additional adjustment in the form of frontier shift may need to be estimated separately. The efficient costs estimated by the model can be used to inform the tariff adjustment required on the current cost of GTS	**C–D**, depending on the complexity of the model, the reference model is likely to provide estimates of the need for additional capital investment and the likely cost of such investment; however, it could be extended to cover TOTEX. If limited to CAPEX, a separate assessment is likely to be required of OPEX static efficiency and the frontier-shift adjustment to adjust the tariffs	**D**, depending on the level of complexity. Separate assessment would be required of OPEX catch-up efficiency and the frontier-shift adjustment to adjust the tariffs

Source: Oxera analysis, based on input from the NMa.

Park Central
40/41 Park End Street
Oxford OX1 1JD
United Kingdom

Tel: +44 (0) 1865 253 000
Fax: +44 (0) 1865 251 172

Stephanie Square Centre
Avenue Louise 65, Box 11
1050 Brussels
Belgium

Tel: +32 (0) 2 535 7878
Fax: +32 (0) 2 535 7770

200 Aldersgate
14th Floor
London EC1A 4HD
United Kingdom

Tel: +44 (0) 20 7776 6600
Fax: +44 (0) 20 7776 6601

www.oxera.com