

Selecting A Benchmarking Model Creates Need For An Error Margin, Says Dutch Court



By Tomas Haug

Energy Regulation Insights

From the Editor

Attempts to adapt the managerial technique of statistical benchmarking to regulatory purposes have frequently run into problems. A court decision from the Netherlands provides further indication that international benchmarking models are of little use in a regulatory context. In its decision, the Dutch appeals tribunal repeatedly rejected the regulator's attempt to adopt benchmarking results from one particular model, insisting that the regulator's decision should make due allowance for the margin of error inherent in picking one model rather than another. The other model gave the target company a high score, and as a result, adding this margin drastically reduced the ability of the regulator to demand cost savings, which rather undermines the rationale for this extremely time-consuming technique. Although the court decision applies only in the Netherlands, it is sufficiently general to affect the practice of benchmarking in other regulatory jurisdictions. It certainly gives regulators everywhere reason to reconsider the value of regulatory benchmarking and to look at other, more objective ways of setting targets for cost reduction.

Tomas Haug, Director

Background

Like their counterparts in many countries, Dutch energy regulators have adopted the managerial technique of statistical benchmarking as the basis for setting regulated network tariffs for the last 20 years or so. The technique has remained a source of dispute due to its inherent subjectivity, however, leading to many appeals. A recent attempt to benchmark the Dutch electricity transmission system operator, TenneT TSO (TenneT), has culminated in a court decision that explicitly recognises the wide variation in possible benchmarking scores. While the court decision is applicable only within the Netherlands, it may influence benchmarking practices in other parts of the world.

The Dutch energy regulator, Autoriteit Consument & Markt (ACM), and its various predecessors have been benchmarking TenneT since 2005. Given the lack of Dutch comparators for a national monopoly transmission system operator (TSO), the regulators used international benchmarking as the basis for setting TenneT's allowed revenue and tariffs.

In its most recent benchmarking attempt, ACM commissioned a consortium of consultants (the Consortium) to conduct a specifically Dutch run of the international TSO efficiency benchmarking model known as e3GRID2012, which used 2012 data from TSOs in many different countries. (The e3GRID process has been running for some time, using data from previous years.) This Dutch run of the model was given the title "Special TENnet Assessment", or STENA. ACM used the efficiency score emerging from STENA to determine what TenneT's costs would (or perhaps "should") be at the end of the next regulatory period. Using this figure, ACM was able to calculate the annual rate of change (the "X-factor") required to reduce TenneT's revenues from the current level to the forecast level.



In STENA, ACM ordered the use of specific assumptions about the Weighted Average Cost of Capital (WACC), asset depreciation lives, and the consumer price index, rather than using the assumptions adopted for e3GRID2012. STENA used parameters that corresponded to ACM's regulatory decisions for the regulatory period from 2011 to 2013. As a result of changing these assumptions, TenneT's efficiency score dropped from 100% (in e3GRID2012) to 85% (in STENA).

According to the Consortium, the main reason for the drop in TenneT's efficiency score was the change in the WACC. In the words of the Consortium:

In STENA 2012, the result of [85]% can be interpreted simply as 'if TenneT would invest and operate as efficiently as the peer units subject to the parameters of TenneT, then [85]% of the total expenditure for construction, maintenance and share of support would be enough to provide all current services'. Inversely, '[15]% of the current total expenditure for construction, maintenance and share of support could have been avoided if TenneT had applied the best practices of TSO peers operating under the same financial constraints as TenneT'.¹

In fact, it is mere presumption to assume that TenneT could have reduced its costs by 15% merely by adopting best practices; the cost difference may also be due to other factors not taken into account in the model. Regardless, that was the manner in which ACM chose to interpret the benchmarking results.

Key Driver of TenneT's "Efficiency"

In the e3GRID2012 model, the Consortium assumed a common WACC of 4.36% (post-tax). For STENA, ACM changed the WACC to 6% (pre-tax), which is the WACC that applied in the Netherlands for the period 2011 to 2013. The corresponding post-tax WACC is 4.81%, and therefore close to the e3GRID2012 assumption on a like-for-like basis. Hence, the main difference between e3GRID2012 and STENA arose over the treatment of corporate taxes. The former excluded such taxes (as irrelevant to a comparison of efficiency), whilst the ACM included such taxes within its specifically Dutch model, as a component of the formula for a pre-tax WACC.

To understand how including a (mostly uncontrollable) cost item like tax in a benchmarking model can significantly affect the resulting scores, it is worth considering in more detail how benchmarking works in general and how it was conducted by the ACM's consultants.

STENA (like e3GRID) is based on the method of Data Envelopment Analysis (DEA)—a benchmarking approach used in a number of regulatory contexts. Normally described as "non-parametric", DEA is a deterministic calculation of relative performance.² In DEA, the relative importance of different cost drivers is found by a linear optimisation procedure. The optimisation procedure defines a "frontier" of best-performing firms, which are classed as "100% efficient". For each firm, the model then constructs a composite made up by weighting the outputs of the frontier firms that are closest to the frontier. DEA then awards each firm an "efficiency score" defined by its performance relative to this weighted composite firm.

Given a relatively small sample, consisting of 21 highly varied European electricity TSOs, a DEA model will inevitably place many of the participating TSOs on the frontier, simply because no other TSO is comparable. The Consortium expressed concern with this outcome, stating that a "...disproportionate[ly] large share of the TSOs appear fully efficient by default simply because within the small sample there are not sufficiently many similar entities to allow comparison".³ Rather than accepting that the small sample invalidates any proper comparisons, the Consortiums claimed that there were "*methodologically sound ways to alleviate these problems*", for example, by constraining the weights that are placed on each output (so-called "weight restrictions"). The Consortium acknowledged that there was no single objective way to determine "reasonable values for the restrictions on the output weights". Thus, lacking any objective way to determine the constraints, the Consortium set a range of weight restrictions (+/-50%) around the regression coefficients estimated in their analysis of cost drivers. The assumed range was chosen arbitrarily, but affected the outcome of the process (a common problem with most benchmarking exercises). Box 1 illustrates how this adjustment can affect the scores emerging from DEA.

In the Dutch national run of the model, the weight restrictions were re-calculated for the national parameters, most notably the pre-tax WACC of 6%. The Consortium explained that the change in WACC had a large impact on the regression coefficients (and, therefore, on the weight restrictions), as the coefficients were unstable. In particular, for the DEA output variable known as "density",⁴ the weight restriction fell to a much lower (i.e. more restrictive) level. This tightening of the restriction had a significant impact on TenneT, for which density was a key explanatory cost driver. TenneT operates in an area with relatively high population density, which drives up construction costs.

Tightening the weight restriction on "density" had the effect of reducing the extent to which density could explain TenneT's



costs. Thus, in this model, changing to a higher WACC (by including tax—a non-controllable cost item) reduces the impact of density on total costs. From an engineering perspective, this outcome has no plausible explanation.

Therefore, the STENA2012 study should be regarded as unreliable and defying common sense. The Consortium does not provide any intuitive explanation for the fall in TenneT's efficiency score stating only that:

TenneT is obtaining a lower score in the STENA2012 than the e3GRID2012 study (...) primarily driven by the WACC at 6% via impact on the regression coefficients, the effect of other changes being marginal. Above 6% and below 5%, the OLS coefficients are stable.⁵

The reference to the regression ("OLS") coefficients being stable "[a]bove 6% and below 5%" is irrelevant, given that the change in the WACC lies specifically *within* this range. Unstable statistical regressions do not provide any justification for the observed result and suggests that any efficiency targets emerging from the Consortium's benchmarking must be unreliable.

NERA's Role

In parallel to the e3GRID2012 benchmarking project, a sub-group of the TSOs performed a "Shadow Benchmarking" project. Its purpose was to follow and replicate the analysis undertaken by the Consortium. NERA experts carried out the shadow benchmarking on behalf of these TSOs.

"The judgement of the tribunal has a number of implications, both for the practice of benchmarking, and for the higher level choice as to whether this kind of benchmarking is worthwhile in a regulatory context."

Using the data provided by these TSOs, we were able to replicate the regulators' results in our shadow runs of the Consortium's benchmarking analysis. We could then analyse the sensitivity of the ACM's results with respect to variations in the STENA assumptions, most notably by replacing the pre-tax STENA WACC of 6% with the corresponding e3GRID2012 post-tax WACC of 4.81%.

Our shadow calculations showed that in the post-tax WACC scenario, TenneT's efficiency score was 100% (and that TenneT was not an unusual case or "outlier"). NERA economists, therefore, demonstrated that the STENA result provides an unreliable basis for setting efficiency targets.

Court Ruling

In 2015, the tribunal for regulatory appeals in the Netherlands (the CBB⁶) rejected the use of a single benchmarking model, ordering that the regulator should include a margin for error, because other models gave different answers (i.e. higher scores).⁷ In a decision on a second appeal published on 8 December 2016,⁸ the same tribunal rejected the regulator's proposal to add a 5% error margin and ordered the regulator to add 10%, to reflect the difference between the benchmarking models concerned. This decision raised the benchmarking score of the appellant from 85% to 95%.

The judgement of the tribunal has a number of implications, both for the practice of benchmarking, and for the higher level choice as to whether this kind of benchmarking is worthwhile in a regulatory context.

The first implication of this judgement (and of the interim judgement of 2015) is that regulators may not be free to choose a model and a single set of inputs, even if their selection is based on some statistical criteria (such as closeness-of-fit in regression analysis). The tribunal recognised that the selection of particular input data created a margin of error in itself, because other input data might give different results. In this context, the tribunal noted particularly that TenneT had little or no influence over the WACC, the main source of variation in results.⁹

The second implication is that the regulatory use of benchmarking may represent an expensive and time-consuming technique, liable to provoke multiple disputes, but with no discernible benefit for consumers. The addition of the 10% margin for error gave TenneT an "efficiency score" of 95%. This result would require TenneT to reduce its costs by 5% over a period of three to five years, i.e. by 1.0-1.7% per year, depending on the length of the next regulatory period.¹⁰ For most regulated businesses, such putative cost savings would be offset by a rise in the WACC of less than 1% (which might easily emerge from the increased regulatory risk caused by such subjective methods).¹¹ Having to allow for equally valid benchmarking models with results closer to 100%, by adding a margin for error, would effectively negate the value of all of the effort put into benchmarking.



Of course, every utility regulator wants to ensure that imprudently incurred costs are not passed through to customers via higher prices. The Dutch tribunal's judgement suggests that benchmarking is not worth the effort to achieve such aims. It does not give regulators a shortcut that avoids detailed scrutiny of costs. Consumers would be better served by simpler, more objective methods of setting cost targets, such as long-term trends in Total Factor Productivity (to set annual rates of efficiency growth), and detailed, like-for-like comparisons of individual expenditures (to identify items that merit further investigation).

Box 1: DEA and Weight Restrictions

Data Envelopment Analysis (DEA) uses a linear program to compare the outputs (or, in a different formulation, the costs) of a set of firms. Figure 1.1 shows broadly how it works for a purely illustrative example of oil refineries with two outputs.

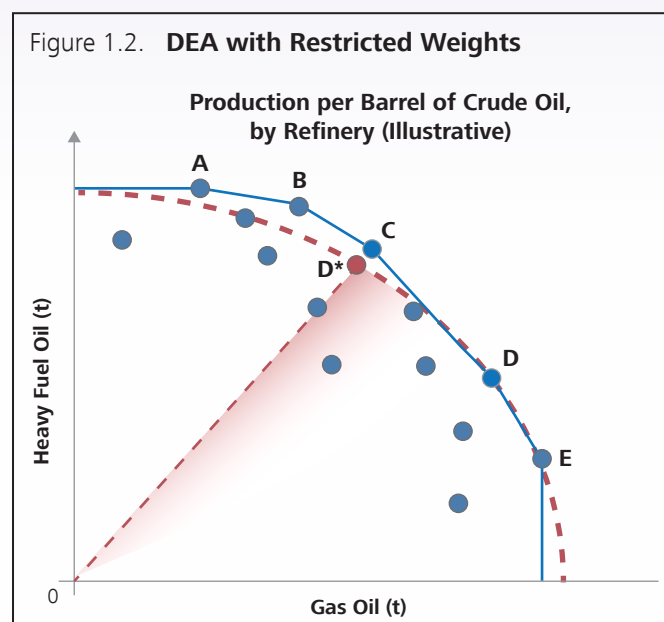
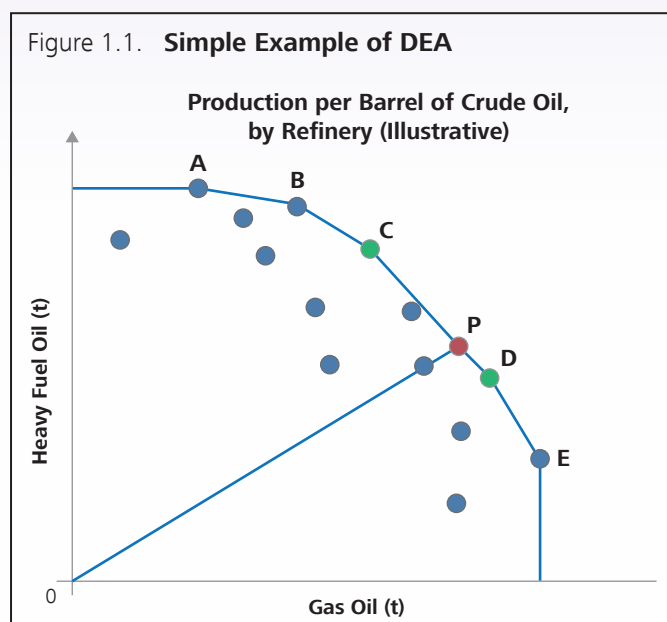
Each of the dots represents a single oil refinery and shows how much it is able to produce from a single barrel of crude oil, assuming that it can split the barrel into only two products: heavy fuel oil and gas oil. The "frontier" of maximum outputs is defined by refineries A, B, C, D, and E, and the solid blue lines drawn between them. The efficiency of the "target" refinery, T, is calculated by constructing a "peer", P, as an average of refineries C and D. The peer produces outputs in the same proportions as refinery T, and so lies on a line drawn from the origin (O) through point T to the frontier. The efficiency score of refinery T is defined by its closeness to this peer, i.e. as the ratio of distance OT to distance OP—in this case about 85%.

The Cbb's latest judgement applies Dutch law to a specific decision by a national regulator, but the underlying economic concerns affect all benchmarking exercises and may raise similar issues under other national laws. The decision gives all regulators a reason to consider whether it is worth embarking on a benchmarking exercise, if it will not be reliable enough to define allowed costs without a substantial margin for error.

Figure 1.2 shows the effect of restricting the weights that can be assigned to different outputs. In this case, refinery C represents the maximum allowed ratio of gas oil to heavy fuel oil. Refineries to the right and below the line OC produce a higher ratio of gas oil to heavy fuel oil, but positions below the line OC (in the shaded area) are discounted. Instead, refineries in this area are treated as if they lie on the line OC.

Thus, refinery D lies on the frontier and would have a score of 100%. However, to abide by the weight restriction, it must be swung around to the position D*, along the quarter circle marked by the dashed red line. At this point, it lies within the frontier defined by refinery C, and, therefore, has a score below 100%—in this case, about 93%.

Imposing the weight restriction, therefore, affects the scores of some of the refineries (but not all), and removes some from the frontier, by shifting them into another part of the graph.





Endnotes

- 1 Consortium (2016), *STENA2012–Benchmarking TenneT TSO 2007-2011*, Frontier Economics/Sumicsid/Consentec, p29. The previous score of 100% is taken from the e3GRID2012 results for TenneT NL, September 2013, p4. Note that the Consortium’s report actually refers to an efficiency score of 83%, which is based on the “STENA2012 base case” scenario. The score of 85% used by ACM and inserted here for the sake of clarity is consistent with the Consortium’s “STENA2012 excluding NorNed” scenario.
- 2 Unlike regression, DEA offers no statistics on the reliability of the results. For that reason alone, regulators often prefer DEA to regression, when analysing small samples of data (e.g. less than 50 observations). In a small sample, however, the absence of reliability statistics does not make the results of DEA any more reliable than those of regression. The absence of such statistics merely hides the underlying problem caused by the small sample size.
- 3 e3GRID2012 – European TSO Benchmarking Study, July 2013, p43.
- 4 Density mainly affects a TSO’s capex (and hence its costs of depreciation and return) by raising construction costs.
- 5 Frontier Economics/Sumicsid, *STENA2012 – Note on reports by Polynomics/NERA for TenneT*, April 2014, p. 19.
- 6 College van Beroep voor het bedrijfsleven (in English, Industry Appeals Tribunal).
- 7 CBB (2015), *Interim Judgement (Tussenuitspraak)* on matters 13/855 and 13/865, Industry Appeals Tribunal (*College van Beroep voor het bedrijfsleven*), 11 August 2015, reference ECLI:NL:CBB:2015:272, available at Rechtspraak.nl.
- 8 CBB (2016), *Judgement (Uitspraak)* on matters 13/855-862 and 13/865-868, Industry Appeals Tribunal (*College van Beroep voor het bedrijfsleven*), 8 December 2016, reference ECLI:NL:CBB:2016:374, available at Rechtspraak.nl.
- 9 CBB (2016), paragraph 2.4.
- 10 The Dutch Electricity Law of 1998 obliges ACM to adopt a regulatory period of three to five years, and to set an X-factor that brings a regulated business’s allowed revenue into line with its forecast of the business’s costs by the end of that period. The X-factor applies to the regulated business’s allowed revenue at the start of that period, rather than to its costs, and so may differ from the required rate of change in its costs.
- 11 This statement relies on a few basic assumptions about the cost structure of a regulated business, namely that half of total costs are made up of operating expenses, whilst the other half comprises a return on capital of 5% and depreciation based on an average asset life of around 30 years.

Contributor

Tomas Haug

Director

Berlin | +49 30 700 1506 10

tomas.haug@nera.com



NERA Contacts

EUROPE

Dr. Richard Hern

Managing Director
London | +44 20 7659 8582
Berlin | +49 30 700 1506 01
richard.hern@nera.com

NORTH AMERICA

Dr. Jeff D. Makhholm

Managing Director
Boston | +1 617 927 4540
jeff.makhholm@nera.com

AUSTRALIA/NEW ZEALAND

James Mellsop

Managing Director
Auckland | +64 9 928 3290
james.mellsop@nera.com

About NERA

NERA Economic Consulting (www.nera.com) is a global firm of experts dedicated to applying economic, finance, and quantitative principles to complex business and legal challenges. For over half a century, NERA's economists have been creating strategies, studies, reports, expert testimony, and policy recommendations for government authorities and the world's leading law firms and corporations. NERA serves clients from more than 25 offices across North America, Europe, and Asia Pacific.

Our Services

Energy Regulation Insights showcases insightful quantitative analysis from NERA's top energy experts on critical issues affecting electricity and natural gas markets around the globe. Other examples of recent energy market studies we have completed include:

- Creating prototype commodity and asset valuation models for a large European utility (jointly with our sister company Oliver Wyman).
- A review of wholesale electricity markets and renewable investment incentives in Eastern Europe.
- Assessment of the impact of vesting contracts in the Singaporean power market, including game theoretical modelling of competitive dynamics.

For more information on our capabilities in these and other related areas, please visit our website at www.nera.com.

The opinions expressed herein do not necessarily represent the views of NERA Economic Consulting or any other NERA consultant.