NERA
ECONOMIC CONSULTING

NERA Economic Consulting
Marble Arch House
66 Seymour Street
London, UK W1H 5BT
+44 207 659 8500
www.nera.com

MEMO

TO: Gasunie Transport Services

DATE: 8 June 2020

FROM: George Anstey and Soren Christian (NERA)

SUBJECT: **Review of Sumicsid Age Corrections**

1. Introduction

Sumicsid performed the benchmarking for the Council of European Energy Regulators (CEER)in the TSO Benchmarks 2018 process (TCB18). It relied on a Data Envelopment Analysis (DEA) approach, based on cost and output data from a set of gas (and electricity) TSOs around Europe. Its final DEA models did not include any variables to control for the effects of age on costs (or cost efficiency). GTS has raised concerns that the final DEA methodology penalises gas TSO with high asset ages, as it does not capture the extent to which those companies' costs are higher as a result of their aging networks. GTS also argued that the effects of age are discontinuous[1] and that the data is therefore insufficient to address the problem statistically.

In response to GTS's concerns, Sumicsid has carried out further analysis of its sample. In particular, it seeks to test two hypotheses:

- H1: "high asset age leads to higher costs and, by consequence to lower cost efficiency score for the same output"; and
- H2: "age effects are discontinuous and […] the sample lacks data to statistically address the problem of GTS".[2]

With respect to H1, Sumicsid carries out a series of 12 Tobit regressions comparing the efficiency score for each year-company observation to a different variable measuring the network's age (with separate regressions for each age variable).[3] Out of these 12 regressions, it finds only that the relationship between efficiency scores and measures of asset age was statistically significant only for the share of compressors under 5 or 10 years in age, and that this relationship was negative (suggesting that companies' costs are *higher* when their networks are younger). All ten other relationships are statistically insignificant. Therefore Sumicsid concludes that hypothesis (H1) "cannot be supported" and, "if anything, […] GTS might have an advantage through older assets".[4]

---

[1] i.e. has intervals or gaps, such that asset age has no impact on efficiency scores until the age at which assets begin to break.

[2] Sumicsid (14 May 2020), A Note on Age Corrections in TCB18, p.1.

[3] (1) Average age of pipelines; (2) average age of compressors; (3) average age of all assets; (4-9) share of (pipelines/compressors/all assets) under (5/10) years; (10-12); share of (pipelines/compressors/all assets) over (60/30) years.

[4] Sumicsid (14 May 2020), A Note on Age Corrections in TCB18, p.4.

With respect to H2, Sumicsid argues that (1) GTS has the highest average asset age out of the 13 networks for which it has age data, but that the distribution of observations is relatively smooth; (2) GTS does not have the highest proportion of overage pipelines or compressors; and (3) “at least some of the direct peers for GTS have mean asset ages higher than average” and overage shares comparable to that of GTS.[5] It is therefore states that it cannot confirm the hypothesis that it is unable to benchmark GTS due to lack of sufficient age data in the sample.

We have reviewed Sumicsid’s arguments and have identified several shortcomings in its reasoning. We set these out as follows:

- Section 2 discusses the shortcomings of Sumicsid’s testing of H1;
- Section 3 discusses the shortcomings of Sumicsid’s testing of H2; and
- Section 4 concludes.

In short, we find that Sumicsid’s analysis is insufficient to identify the *absence* of age effects within the DEA models. The data is likely insufficient to definitively identify the *presence* of such effects using econometric methods. Therefore, analysis of a wider sample and/or detailed bottom-up or engineering approaches are likely to be necessary to quantify these effects. Relying on Sumicsid’s analysis would not provide a sound basis for setting GTS’s revenues in line with its efficient costs.

2. Sumicsid Analyses H1 over a Small Sample Using a Mis-specified Approach

2.1. Sumicsid Overstates the Value of its Evidence through Elementary Error

As described in Section 1, Sumicsid carries out a range of statistical tests to test the hypothesis that high asset age is correlated with lower efficiency scores in the DEA benchmarking. It reports two variants of its conclusions in different points of its note:

- At the end of Section 2 (the section in which it describes its testing of H1), it concludes that “the hypothesis H1 that older transmission assets would bias or decrease the cost efficiency score in TCB19 cannot be supported”;[6] and
- In Section 4 (the overall conclusions), it concludes that “the DEA-scores in TCB18 for gas transmission are not sensitive to high asset age. A higher share of newer compressors (5 or 10 years) may lead to a lower DEA-score”.[7]

At first glance, these seem to be identical conclusions replicated in different parts of Sumicsid’s note, but there is an important semantic difference: In the first iteration, Sumicsid

5 Sumicsid (14 May 2020), A Note on Age Corrections in TCB18, pp. 8-9.

6 Sumicsid (14 May 2020), A Note on Age Corrections in TCB18, p.4.

7 Sumicsid (14 May 2020), A Note on Age Corrections in TCB18, p. 8.

fails to reject the null hypothesis (i.e. that there is no relationship between asset age and efficiency score); In the second iteration, Sumicsid appears to *reject* the alternative hypothesis (H1).

This second iteration of its conclusions violates a core tennet of statistical hypothesis testing. Starting with a null hypothesis – usually that the hypothesised relationship (alternative hypothesis) does not exist – a researcher can either find evidence to *reject* the null hypothesis, and in so doing confirm the alternative hypothesis, or they can fail to find evidence to reject the null hypothesis. Critically, in this latter case, *it does not follow* that the alternative hypothesis is rejected. Similarly, it is never possible to *confirm* the null hypothesis – it merely remains unrejected in the absence of strong evidence to the contrary.

In the context of this particular hypothesis, it would be incorrect to read the evidence presented and conclude that there is no negative correlation between asset age and DEA efficiency score. Sumicsid's particular analysis does not support that assessment, only that it was unable to find sufficient evidence (and likely due to reasons set out subsequently in this memorandum). It does not follow that the relationship does not exist.

The first iteration of Sumicsid's conclusion places the burden of proof on rejecting or failing to reject the null hypothesis (in this case the latter), which is at least more philosophically accurate for the purpose of statistical inference. However, it is not strictly correct to say that H1 "cannot be supported". More accurately, H1 cannot be confirmed based on the evidence and analysis presented, but that there could be other approaches which do support it (such as alternative functional forms).

Moreover, as discussed throughout this memorandum, there are many reasons why we would not expect Sumicsid to be able to reject the null hypothesis based on the shortcomings of its analytical approach. These reasons have nothing to do with the absence of an underlying relationship between asset age and DEA efficiency score.

2.2. Sumicsid Excludes Over Half the Sample in an Already-Data Poor Environment

Sumicsid's principle method for assessing H1 is to regress efficiency score on asset age. It uses 12 different measures of asset age and regresses efficiency scores against those measures one by one. Sumicsid uses the "Tobit" method for censored data: Final efficiency scores cannot be in excess of 100 per cent (at least after the adjustment of super-efficient peers) and therefore simple OLS methods would tend to bias coefficients on age towards zero.

Sumicsid's analysis begins by discarding over half the sample of TSOs considered. Specifically, it excludes data from the 16 German TSOs, as they do not report asset ages. Sumicsid has developed a DEA model and attempted to verify that it is a sound model on the full sample. Exclusion of over half the sample increases the likelihood of a false negative by reducing the degrees of freedom and due to the risk that asset age and efficiency score may be correlated in the datapoints removed. In the absence of over half the sample Sumicsid's analysis cannot be regarded as reliable. We know of no previous statistical analysis used to support a benchmarking exercise where the regulator or its consultant excluded over half of such a limited sample set at the outset of that analysis, at least outside of the context of a CEER-benchmark.

Sumicsid could argue that removing half the sample would still shed statistical light on its hypotheses given strict statistical conditions, none of which apply. For instance, Sumicsid could argue that the exclusion of over half the sample would not risk drawing inaccurate conclusions if:

- the excluded data was random, which as a selection of TSOs exclusively from Germany it is not;
- it could reliably assert that there would be no relationship between asset age and efficiency score in the excluded data, which is unproven and *a priori* unlikely given the impact that age is likely to have on costs; and
- Sumicsid had access to such a superfluity of data and an enormous sample size with thousands of degrees of freedom that excluding a randomly-selected group of peers would be unlikely to have an impact on efficiency scores: In practice, Sumicsid reports sample sizes of just 40 (for variables measuring age of compressors) and 54 (for variables measuring age of pipelines and total assets).

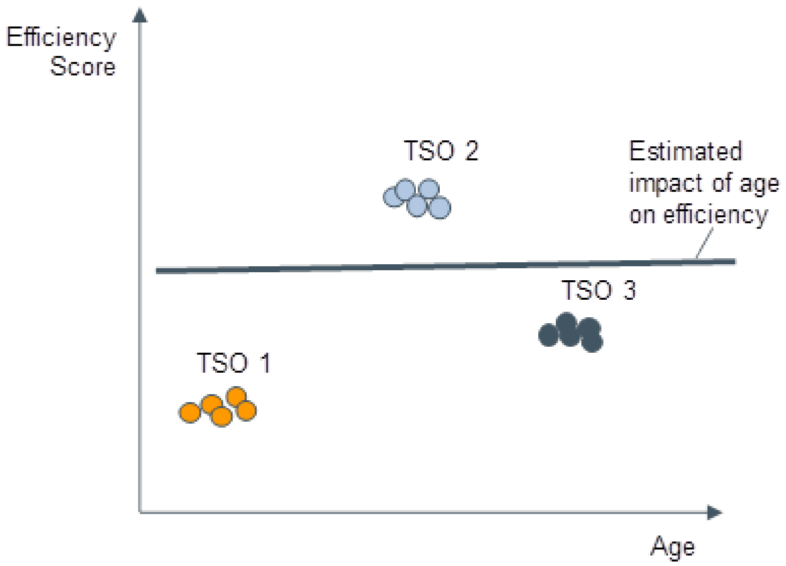
2.3. Sumicsid's Dataset Contains Insufficient Observations and Variation to Develop Sound Conclusions

Sumicsid does not provide full details of the data it includes in its analysis. It reports that it has data from 13 companies, and over five years. This would imply that it has 65 observations. It is not clear whether the 11 missing observations (65-54) for pipelines and total assets or 25 (65-40) for compressors arise from missing years or missing companies. It appears from figures presented with respect to H2 that it has 13 unique companies with at least some data on pipelines (and therefore 11 company-year pairs are missing) and 9 unique companies with data on compressors (and therefore 5 company-year pairs are missing).

There are no hard-and-fast rules in statistical inference for the minimum sample size because much depends on the data being examined and the hypotheses the researcher is investigating. However, by any reasonable standard this is not a large sample and therefore one prone to misleading results – particularly false negatives, which tend to increase as the sample size falls. In any case, the number of datapoints used is not the only consideration for assessing the sufficiency of a sample. In order to draw robust statistical conclusions, one needs sufficient variation within the sample. In other words one needs both a large number of datapoints and dispersed datapoints around the variables of interest.

Sumicsid's sample lacks the data variation that would allow it to perform reliable statistical analysis, because each company's age variable will not change much year-on-year (indeed, the maximum change is 1 year on average). And if it is to be regarded as a reliable measure of a company's underlying efficiency over its historical capital inputs, the DEA score will also not change much year-on-year. As a result, Sumicsid's datapoints are likely to be closely clustered together for each TSO. Though we have not seen the underlying data, we provide an illustrative example of this sort of clustering in Figure 1 below. As a result of this clustering, Sumicsid's analyses collapse to regressions on 9-13 observations (or tight clusters of data), rather than a distribution of 40-54 independent observations. In such circumstances, the apparent degrees of freedom Sumicsid has in its analysis are largely illusory.

Figure 1: Illustration of Data Clustering



Source: NERA Analysis

In a regression on closely clustered data, the impact of age on efficiency scores will be hidden by other factors which vary between TSOs. There are likely many other omitted factors which influence a company's efficiency score, many of which are also time invariant (or close to it), such as topography, differences in service levels and differences in regulatory standards between TSOs. These can be described as company-specific "fixed effects", and, assuming the size of these effects are not negligible, it becomes difficult to identify any other relationship (such as age) in a sample of up to 13 companies when they are not accounted for.

Fixed effects are a well-established challenge in analysing panel data (i.e. data that covers multiple entities over multiple years). There are several techniques Sumicsid could have used to account for fixed effects. For instance, Sumicsid could have taken first differences: If it were to assess the year-on-year change in efficiency score relative to the year-on-year change in asset age, the fixed effects would disappear as they do not change year-on-year.

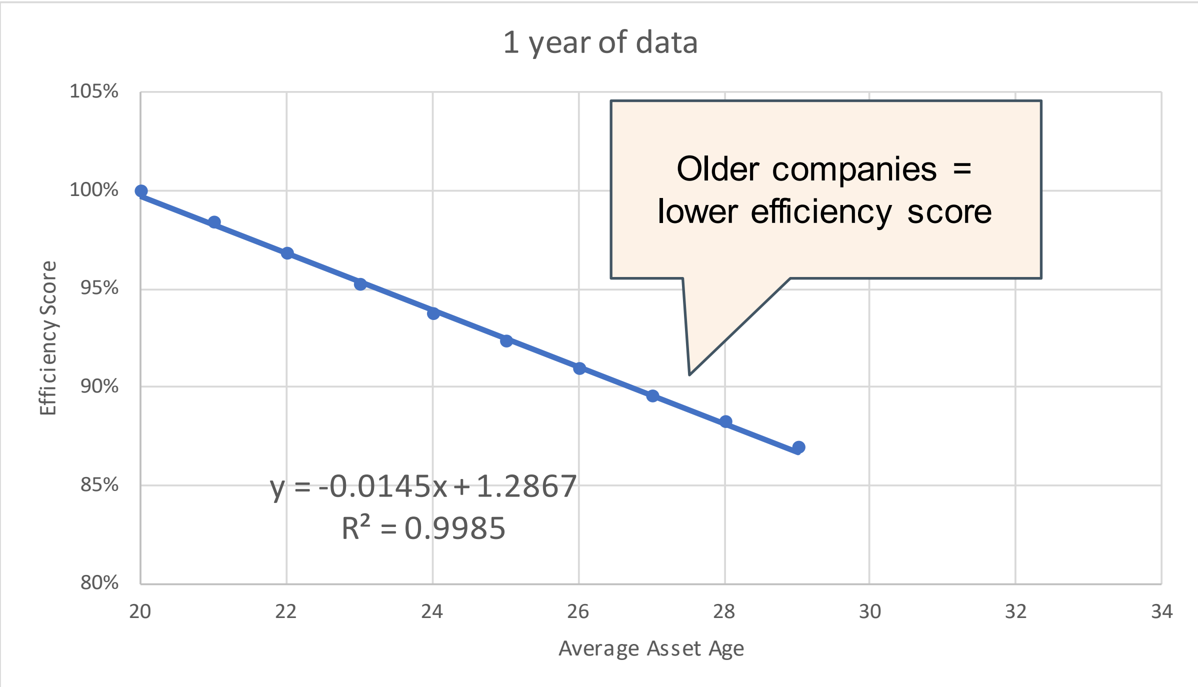
However, in order to for this analysis to identify a statistically significant result, there would need to be significant year-on-year variation in a company's efficiency score and its age. In the absence of such variation, age variables would be statistically indistinguishable from the other fixed effects.

2.4. Sumicsid's Measures of Age Are All Misspecified

Sumicsid's analysis uses data on a company's absolute age in each of the five years it assesses. This approach does not adequately measure the hypothesised relationship between efficiency and age. In particular, because the DEA efficiency scores are calculated separately for each year, the effect of age should be thought of in comparison only with the age of other networks *in that year*. For example, if every company ages at the same rate, then next year's efficiency scores will be similar to last year's, ceteris paribus, but resulting from an older set of networks. As a result the impact of age will suffer from "attenuation bias" and appear not to affect efficiency scores.

To illustrate this effect, assume a hypothetical world where a company's efficiency score is a perfect linear function of the network's age relative to the "youngest" network in that year. If we assess these 10 companies' efficiency scores against average age in a single year, we see the clear "perfect" negative relationship, shown in Figure 2 below.

Figure 2: Efficiency Score vs Asset Age in a Single Year

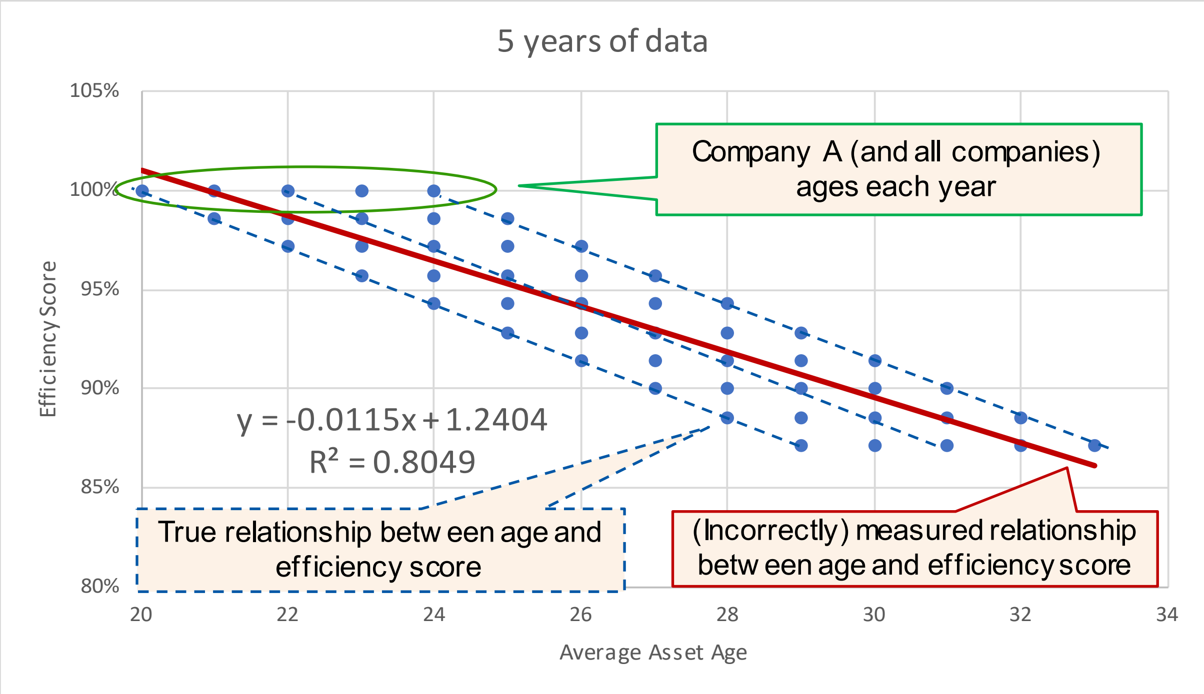


Source: NERA Analysis

In this hypothetical example, we also assume that there is no asset replacement, so each company's average asset age increases by one year every year. Because the DEA scores are calibrated annually, each company's efficiency score does not change annually, even if every company's actual *costs* do increase as a result of their ageing networks.

If we combine all of this data into a single regression, without accounting for the fact that every company's assets age at the same rate, we would see that each company ages but does not worsen in its efficiency score. We illustrate this effect in Figure 3.

Figure 3: Regression is Based on Absolute Age but Efficiency is Based on Relative Age



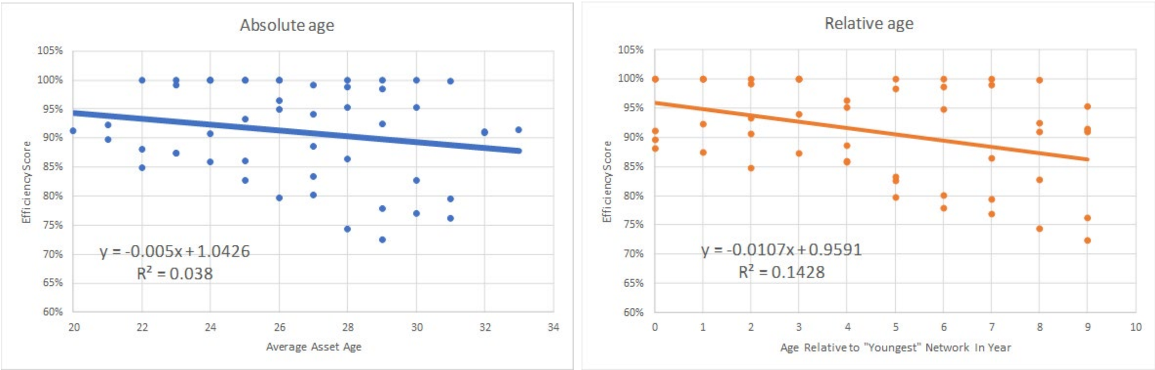
Source: NERA Analysis

In the figure, the true (and known) relationship is measured by the blue dashed lines – each company ages each year, but they do not age relative to each other. The incorrectly measured relationship is shown by the red line, which treats each data point as an independent relationship, and is therefore biased by the fact that all companies age but do not change in efficiency score.

Of course, as discussed in Section 2.3, there are other unobserved factors which drive efficiency scores, so we would not expect a perfectly clear relationship in the data, even if such a relationship does exist. Where there is "noise" in the data, econometricians rely on confidence intervals to determine whether a relationship does exist. If the independent variable is measured with error relative to its true value (i.e. measured in absolute rather than relative terms), this attenuating effect could be the difference between a statistically-significant relationship and the absence of one.

We illustrate the impact of using absolute versus relative age in with random variation in Figure 4 below, where the true relationship between relative age and efficiency score is the same as in the figures above. In the left figure, age is incorrectly measured in absolute terms. The observed relationship is not statistically-significant at the five per cent level (p-value is 8.7 per cent). In right figure, age is correctly measured in relative terms. Even though there are clearly other factors at hand, the relationship is statistically significant (p-value is 0.3 per cent), the coefficient is larger, and so is the goodness of fit (measured by R^2).

Figure 4: Efficiency vs Age plus Randomness



Source: NERA Analysis

In these illustrative examples, we have used average asset age as the independent variable, though this is only representative of three of Sumicsid's regressions. These examples could be extended to the other nine independent variables, relating to shares of certain assets above or below certain ages. In that case, the relevant factor driving efficiency scores would be the share relative to other companies, but Sumicsid's analysis would see all companies' share of overage assets increasing, or all companies' share of new assets decreasing.

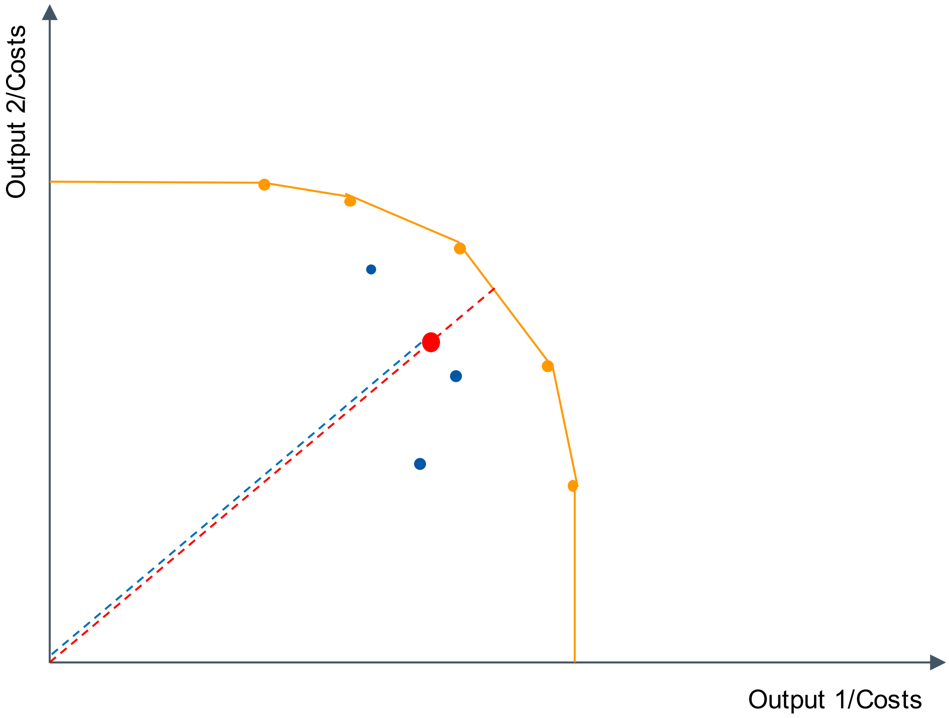
In conclusion, this mismeasurement of age in absolute terms rather than in relative terms attenuates the measured relationship, making it difficult to identify even though (in the case of these illustrative examples) we know that the relationship exists. Applying this logic to Sumicsid's actual analysis, there *may be* a true negative relationship between age and efficiency score that it is failing to capture because of this attenuating effect. In other words, it should be of no surprise that there is no relationship in Sumicsid's analysis because it has used the wrong explanatory variable.

2.5. Sumicsid's Assumptions in its DEA Mask the Impact of Age on Efficiency Scores Using Regression Analysis

In Section 2.4, we explain that the age variables that Sumicsid tests are a poor proxy for the ways in which network age affects efficiency in general. The question before Sumicsid is not, however, whether efficiency scores are affected on average by age, but the specific question of whether GTS's score is downwardly-biased by age. Sumicsid analyses this question by conducting a series of regressions, which are parametric and statistical measures of central tendency. DEA is a measure of the relationship of a firm to a non-parametric frontier defined without reference to inferential statistics. As a result, the average DEA efficiency scores of the sample *on average* are a poor proxy for the age-related penalty suffered *by GTS*: the average age-related penalty need not be related to whatever penalty GTS suffers.

We illustrate a simplified two-output example of a DEA model in Figure 5 below. In a standard DEA framework, companies are assessed on the level of outputs (e.g. network length or number of connection points) divided by cost. A company is deemed to be efficient if it sits on the concave frontier. For example, all of the companies in amber in the figure are determined to be efficient, whilst the blue and red companies sit inside the concave frontier and are therefore inefficient. For a company that is inefficient (e.g. the red company in the figure), its efficiency score is calculated by calculating the ratio between the blue dashed line and the red dashed line.

Figure 5: Illustration of Simplified DEA Approach (assuming CRS, two outputs)



Source: NERA Analysis

There are several ways in which the construction of the DEA efficiency score masks the age-related impact on costs of the sample as a whole and therefore the impact on GTS's efficiency score.

First, a company's efficiency score is only determined with reference to its two peers, i.e. the efficient companies that determine the slope of the efficient frontier for that particular company. Accordingly, the question that Sumicsid should be asking is how age *relative to peers* affects efficiency scores, not absolute age, or even relative to the sample on average. Only firms who are being compared to those same peers will be subject to the same age-related penalty.

Second, a company could have a relatively old network (and face higher costs as a result) but still be on the frontier because it provides a uniquely large level of outputs in a single dimension. For example, if the two outputs are network length and number of connection points, a very rural network will have high levels of network length but a low number of connection points. Even if it faces higher costs due to its age, its network length per unit of cost may still be best in class if its network length is enough of an outlier, or if the other companies with long networks also suffer from age effects. It would perform poorly in terms of connection points per unit of cost, but this is irrelevant to its efficiency score. If GTS is not uniquely large in any given dimension, age would impact its score if it were older than its peer firms.

Third, Sumicsid's age analysis only relies on data from 9-13 companies of the 29 companies participating in TCB18. We have not seen the underlying DEA analysis, but it is possible that a large proportion of these companies assessed are frontier companies with an efficiency score of 1.00, even if there are real differences in age-related cost pressures between the companies. In other words, GTS would have higher costs as a result of its more advanced age and therefore a lower efficiency relative to its peers, but the impact of age on costs would not be picked up for firms that are already at the frontier.

Fourth, in order to account for increasing returns to scale, the DEA models adopted by Sumicsid does not compare companies to those which provide higher levels of outputs *in absolute terms*. Therefore, the smallest companies are, by definition, efficient in the DEA framework, even if they are relatively old and more expensive as a result. For example, assume that the smallest company in a particular dimension has a relatively old network and faces higher costs as a result, while the second smallest company in that dimension is younger and thus does not face such cost pressures. The DEA model would ascribe the higher unit costs of the small company to the effects of increasing returns to scale, when in reality, a proportion relates to the effects of having an older network. The impact of age on the efficiency scores of the smallest companies would therefore provide no guide to the impact of age on GTS's efficiency score.

In very large samples, Sumicsid could potentially argue that the biases identified above were small: A handful of frontier firms whose efficiency scores could not or would not move with age would not affect the overall robustness of its results except in very marginal cases. However, in this case, the very small sample of just 9-13 firms means that the impact of the above-mentioned reasons for differences between GTS's age-related penalty and the sample as a whole could be very material.

3. Sumicsid Analyses H2 Using Non-Statistical Methods, Non-Specified Standards and Based on False Premises

3.1. Hypothesis Testing of H2

Sumicsid also reviews its second hypothesis (H2) that age effects are discontinuous and that there is insufficient data to address GTS's problem statistically. It does so without reference to statistical analysis and instead relies on naïve comparisons of the sample data. Specifically, Sumicsid presents three pieces of evidence to analyse this hypothesis:

- Out of 13 companies, GTS has the highest average asset age, but that the spread of the distribution is "relatively smooth and linear", although the standards by which Sumicsid judges the smoothness and linearity of the data is not specified;[8]
- Out of 13 companies, GTS has the second highest proportion of overage (>60 years) pipelines, although only three companies have *any* overage pipelines; and
- Out of 9 companies, GTS has the second highest proportion of overage (>30 years) compressors), although only five companies have *any* overage pipelines.

From this analysis, Sumicsid concludes that "the hypothesis H2 that the sample lacks comparison data for high age observations or that GTS would be compared to operators with radically younger networks cannot be confirmed".[9]

The wording of Sumicsid's conclusion is convoluted, and changes H2 from a broad hypothesis about the discontinuity of costs and the lack of data to assess GTS's score to a narrower hypothesis about high-age observations and/or "radically" younger networks. Nonetheless, much of the evidence Sumicsid presents tends to support H2 according to either expression.

3.2. Sumicsid Argues that the Distribution is Sufficiently Comparable to GTS Without Supporting Evidence

A core component of H2 as presented at the beginning of Sumicsid's note is that age effects are discontinuous. Sumicsid notes that, at 46.7 years, GTS's average asset age is the highest amongst the sample of 13 companies, and therefore, that "the hypothesis that the age effect is discontinuous, occurring exactly between GTS and the next highest observation cannot be rejected".[10]

It also concludes, however, that the distribution of age is "relatively smooth", and that GTS's DEA peers also have asset lives of "32.5 and 34 years, well above the mean" of 27.1 years. Therefore, according to Sumicsid, "it cannot be concluded that the sample lacks observations with high age or that GTS would be compared to a subsample with radically different age profile for the assets".[11]

Although it rightly does not reject this component of GTS's hypothesis outright, Sumicsid seems to present it with a high degree of scepticism which is unwarranted by the data it actually presents:

- First, GTS *does* have the oldest average asset life in the sample at 46.7 years, with the next oldest at around 39 years. This appears to be the largest gap in age within Sumicsid's sample, so it is unclear what Sumicsid means when it describes the distribution as "relatively smooth". Sumicsid provides no standard for smoothness or

8 Sumicsid (14 May 2020), A Note on Age Corrections in TCB18, p. 6.

9 Sumicsid (14 May 2020), A Note on Age Corrections in TCB18, p. 6.

10 Sumicsid (14 May 2020), A Note on Age Corrections in TCB18, p. 6.

11 Sumicsid (14 May 2020), A Note on Age Corrections in TCB18, p. 6.

linearity nor conducts any statistical testing to assess whether that standard has been met. Its assertion of "relative" smoothness and linearity is meaningless: the reader is simply left to ponder relative to what Sumicsid believes the data is smooth and linear.

- Second, Sumicsid implies that GTS's DEA peers are sufficiently comparable to GTS because their average ages are also "well above the mean". GTS's average age is 19.6 years older than the mean, while its peers are 5.1 and 6.9 years older than the mean. GTS's average age is thus nearly three times further from the mean than its nearest peer, and nearly four times further from the mean than its other peer. Sumicsid again provides no explanation or justification for why this apparently-material difference is in fact immaterial.
- Third, comparison to the mean or to the whole sample are irrelevant statistics, as DEA scores are determined only relative to the peer companies. GTS's average asset age is 37 to 44 per cent higher than that of its two peers.

Without further analysis of the relationship between age and costs, it is difficult to say whether a c. 40 per cent difference in average age is too large for GTS's peers to be considered comparable. However, Sumicsid is also incorrect to use this evidence to suggest that the sample includes adequately comparable firms.

3.3. In Practice, Discontinuity in Age-Related Costs May Indeed Occur between the Age of GTS and its Peers

Sumicsid acknowledges that the discontinuity in age-related costs could occur between GTS's average age and that of the next oldest firm, as it is impossible to statistically distinguish this hypothesis from other characteristics specific to GTS which may drive costs. However, as mentioned above, the relevant comparison in this case is not to age of the next oldest company, but rather to the ages of its DEA peers (32.5 and 34 years). In order to properly assess whether a discontinuity happens between 33 years (the approximate age of GTS's peers) and 47 years (the approximate age of GTS), we must consider the techno-economic evidence as well.

Sumicsid describes an asset's "techno-economic life time […] as the expected age of replacement for each asset, up until which the asset will provide full functionality in the power system compared to a new asset. Passed [sic.] the life time, the increase in expected maintenance cost is higher than the combined cost of capex and opex for a replacement. The potentially discontinuous part of the curve is therefore situated beyond the stipulated asset lifetimes", agreed by participants in the TCB18 process to be 60 years for pipelines and 30 years for compressors.[12]

Sumicsid's description implies that, for *all* pipelines under 60 years in age (and all compressors under 30 years in age), the cost of replacement is less than the cost of maintenance. This is an incorrect inference. While the *average* pipeline may have an asset

12 Sumicsid (14 May 2020), A Note on Age Corrections in TCB18, p. 7.

life of 60 years, this is simply an average out of a distribution. Some will last longer, and some must be replaced sooner.

This effect is explicitly accounted for in the "Survivor" model developed by engineering consultancy DNV-GL and used in British regulation. The Survivor model will set an average asset age for a particular asset class and assume that assets of that type must be replaced according to a normal distribution with a standard deviation of around 25 per cent of the mean.[13] In other words, the effects of age on replacement likelihood are not linear, with young assets having almost no chance of replacement up to a certain age, when the probability begins to increase more rapidly.

If the assumed asset life of a pipeline is 60 years with a standard deviation of 15 years, this means that about 3.6 per cent of pipelines of 33 years in age must be replaced, while 19.3 per cent of pipelines of 47 years in age must be replaced. According to this model, therefore, GTS would have to carry out substantially more replacement work than its peers, which are only 32.5 and 34 years old on average.[14]

The techno-economic evidence presented by Sumicsid speaks only of asset replacement rather than increased maintenance costs. As assets age, the costs of maintaining them tends to increase, but the operator should only replace them once the cost of maintenance exceeds the annuitised cost of replacement. Therefore, even if all pipelines will survive exactly 60 years, a company with *ageing* but not *overage* assets will still face increased maintenance costs, even if the maintenance costs do not exceed replacement costs. Sumicsid's presentation of techno-economic evidence says nothing of this effect, except that expected maintenance cost exceeds replacement cost only after the expected life of the asset.

3.4. Analysis of Over-age Shares

Based on its assumption that costs will only increase when an asset exceeds its techno-economic life (itself an incorrect assumption according to the survivor model discussed in section 3.3 above), Sumicsid presents two charts. The first shows that 13 per cent of GTS's pipelines are overage, the second-highest in the sample. 10 out of 13 companies, including both of GTS's DEA peers, have no overage pipelines. The second shows that 70 per cent of its compressors are over-age, also the second highest in the sample. 4 of 9 companies, including one of GTS's peers, have no overage compressors.

From this evidence, Sumicsid concludes that:

(1) GTS is not the maximum observation; and

(2) "at least some of the direct peers for GTS have […] a share of overage compressors that is comparable to that of GTS".[15]

---

13 As advised by GTS

14 In this example, we assume that GTS's pipelines are all 47 years in age whilst its peers' pipelines are all 33 years in age. In reality, there will be distribution around these ages which is not linearly related to replacement probability. The overall tendency of this effect will be the same as in this simplified example, however.

15 Sumicsid (14 May 2020), A Note on Age Corrections in TCB18, p. 9.

Conclusion (1) above is not relevant. As discussed throughout this memo, the only comparison that is relevant in determining a DEA score is to the peer companies. One company may have a higher over-age share in each of these asset types, but it may also be a frontier company with an efficiency score of 1.00 if, for example, it is smaller than its would-be peers. We do not know which company this is as Sumicsid has not identified any specific companies other than GTS.

Sumicsid overstates the importance of conclusion (2) above: only one peer company has an allegedly-comparable proportion of overage compressors. The other peer company does not have any, and neither peer company has any overage pipelines. Even if the hypothesis that age-related costs relate only to the overage share, GTS's high share of old-age assets will only be partially controlled for with respect to its compressors, and not at all with respect to its pipelines. Assuming that pipelines are a more important driver of a gas network's total costs than compressors, then we would conclude that the distribution of relevant data is not particularly comparable to GTS with respect to asset age.

Additionally, these charts demonstrate the lack of data variation of Sumicsid's sample with respect to over-age assets, particularly pipelines, which themselves make up over 75 per cent of gas networks' assets.[16] With only three companies with a non-zero volume of over-age pipelines, it is unsurprising that Sumicsid was unable to identify a correlation between that variable and DEA score.

4. Our Conclusions: Sumicsid's Analysis is Unreliable

Across the hypotheses that Sumicsid has tested, Sumicsid has generally not found conclusive evidence in support of GTS's hypotheses. However, any attempt to read that there is no impact of age on GTS's efficiency score from the evidence and analysis presented would be misguided.

Our core conclusions are the following:

- The absence of evidence relating asset age to DEA efficiency does not imply evidence of the absence of such a relationship. Instead, the absence of evidence could be the result of other analytical shortcomings, as appears to the case in Sumicsid's note.
- Sumicsid discards over half the sample in conducting its analysis which leads it with perilously few data points. The claimed 40-54 data points it includes in its analysis are likely to be illusory because the data points are not independent and age and efficiency score are likely to be clustered. With only 9-13 unique companies, Sumicsid does not have sufficient variation in its data to robustly test the relationship between age variables and DEA efficiency score. The presence of unidentified company-specific "fixed effects" which Sumicsid has not controlled for may outweigh the importance of age in determining DEA efficiency scores.
- The tested relationship between age and DEA efficiency score does not match the *hypothesised* relationship between age and DEA score, in which the DEA score is a

[16] As advised by GTS

function (in part) of a company's age relative to its peers *in that particular year*. Given the mismeasurement of the age-related variables in its regression analysis, it is unsurprising that it finds no correlation between age and efficiency score.

- The DEA efficiency score compared across the sample of companies is a poor proxy for how GTS's efficiency score would change if its assets were older or younger or the age-related penalty to GTS's efficiency score.
- GTS's average asset age is apparently-materially higher than that of the next oldest company, and, in particular, that of its peers. This gap, as Sumicsid acknowledges, means that any a discontinuity in the effects of age on efficiency in between the age of GTS's peers and GTS itself could affect efficiency scores. Moreover, evidence from the "Survivor" models implemented in British gas regulation suggests that any discontinuity is actually likely to fall within this range.
- Sumicsid's discussion of techno-economic factors does not account for cost pressures resulting from *ageing* but not over-age assets.
- With respect to its share of over-age assets, Sumicsid presents compelling evidence that GTS is an outlier relative to its peers, and that, across the whole sample, few networks have any overage assets which would allow for a robust econometric estimation. In spite of the evidence in this direction, it draws the opposite conclusion.

In short, Sumicsid's analysis does not serve as a useful piece of evidence on the relationship (or lack thereof) between ageing assets and GTS's efficiency score.

CONFIDENTIALITY

Qualifications, assumptions, and limiting conditions

This report is for the exclusive use of the NERA Economic Consulting client named herein. This report is not intended for general circulation or publication, nor is it to be reproduced, quoted, or distributed for any purpose without the prior written permission of NERA Economic Consulting. There are no third-party beneficiaries with respect to this report, and NERA Economic Consulting does not accept any liability to any third party.

Information furnished by others, upon which all or portions of this report are based, is believed to be reliable but has not been independently verified, unless otherwise expressly indicated. Public information and industry and statistical data are from sources we deem to be reliable; however, we make no representation as to the accuracy or completeness of such information. The findings contained in this report may contain predictions based on current data and historical trends. Any such predictions are subject to inherent risks and uncertainties. NERA Economic Consulting accepts no responsibility for actual results or future events.

The opinions expressed in this report are valid only for the purpose stated herein and as of the date of this report. No obligation is assumed to revise this report to reflect changes, events, or conditions, which occur subsequent to the date hereof.

All decisions in connection with the implementation or use of advice or recommendations contained in this report are the sole responsibility of the client. This report does not represent investment advice nor does it provide an opinion regarding the fairness of any transaction to any and all parties. In addition, this report does not represent legal, medical, accounting, safety, or other specialized advice. For any such advice, NERA Economic Consulting recommends seeking and obtaining advice from a qualified professional.