Note for Ofgem on Alternative Methodologies: some preliminary analysis

Professor Andrew Smith, University of Leeds, March 2020

1. Introduction

In a previous report for Ofgem (Smith, 2019)[1], it was noted that it is good practice for economic regulators to consider and test alternative approaches as part of their cost benchmarking framework. Whilst some caveats were noted in respect of the use of DEA and SFA in the particular context of gas distribution networks in Great Britain, it was suggested that preliminary analysis be undertaken in order to obtain a clearer view as to the potential value of using these techniques either instead of, or alongside, existing approaches. This approach was also supported by the companies in their responses to Ofgem's consultation on RIIO-2 tools for cost assessment.

The purpose of this note is to report on the results of that preliminary analysis. It should be noted that the analysis shown is not intended to be a full, definitive model selection process with a view to selecting a particular model. It is intended simply to illustrate the possibilities and form a judgement as to whether alternative techniques – namely stochastic frontier analysis (SFA) and data envelopment analysis (DEA) – could form part of Ofgem's suite of models for use within the benchmarking framework for gas distribution. It is clear that there are a wide range of possible SFA and DEA approaches that may be applied, some of them quite complex in terms of their assumptions and estimation. Further work would be needed to select a model using the most up to date data and chosen variables for inclusion within the frontier; also taking account of a wide range of criteria, including transparency.

GDPCR and/or RIIO-GD1 data is used and the results shown focus on the sample including all GDNs observed over a period of 6 years (2013/14 to 2018/19, i.e. RIIO-GD1 actuals). Models using longer time periods were also estimated (2008/09 to 2018/19) as a cross-check. In taking this work forward, more detailed work will be needed to determine the appropriate sample for analysis, including also a decision on whether to include forecast data.

The models shown here have been implemented in the econometric software package, STATA, and run by Ofgem's cost analysis team. Only models for Totex have been included. Here it should be noted that there is a debate in economic regulation around whether to model costs at an aggregate (Totex) level, or

[1] Smith, A.S.J. (2019), Note for Ofgem on Alternative Methodologies, June 2019.

disaggregate level. Disaggregate modelling potentially allows a more tailored set of cost drivers for the individual cost category, but raises data quality issues and substitution effects between categories of cost. There is also the associated problem of cherry picking, whereby it is considered that an unrealistic efficiency target may be obtained by selecting the best performer for each cost category, since it could be argued that this is not realistic for one firm to achieve. This point can (and has been, for example in ED1), dealt with in standard regression analysis work, by first combining cost estimates from disaggregated models prior to conducting an efficiency benchmark (e.g. upper quartile) adjustment[2].

2. Stochastic frontier analysis: alternative model specifications

As discussed in the previous note (Smith, 2019)[3], although the academic literature would tend to support the use of SFA techniques over the COLS approach, this preference does not necessarily flow through into the sphere of economic regulation for a number of reasons. First of all, Ordinary least squares (OLS) may well be a reasonable means of obtaining estimates of the influence of the explanatory variables in the model and, particularly given the small sample size, SFA will not necessarily improve on OLS in this respect.

SFA also relies on distributional assumptions about the random noise and inefficiency components, which may be hard to justify – and the decomposition of the error term merely preserves the original firm rankings that result from comparing the overall residuals in the model. Therefore, when it comes to the step of decomposing the model residual in order to obtain an estimate of firm inefficiency, it is not clear that SFA is superior to the use of regulatory judgement (e.g. through the application of an upper quartile adjustment). There may also be concern about the use of distributional assumptions (which can be seen to be arbitrary) when estimating the parameters themselves. Further, there are numerous, relatively complex, SFA approaches to choose from. Transparency and simplicity could also provide increased support for COLS compared to SFA, though these arguments need to be balanced against the other model selection criteria.

However, SFA is a candidate approach to be considered. This then raises the question as to which SFA model to apply, as there are many, as noted. This

[2] Whilst in principle SFA could be adapted, for example, by combining the cost predictions for an averagely efficient firm (rather than the frontier) from the disaggregate analysis, and then making an adjustment at the overall cost level, this does then raise the question as to why SFA is being applied, in place of simpler methods. Potentially, complex SFA system cost models could also be estimated, recognising the correlation in errors between the disaggregate activities, though this approach is most likely not sensible given the sample size.

[3] Smith, A.S.J. (2019), Note for Ofgem on Alternative Methodologies, June 2019.

document is not intended to cover an exhaustive set, but three broad model types can be identified:

1. The simple, pooled SFA model. This model does not recognise the panel structure of the data (i.e. does not recognise that the data comprises observations on the same firms, repeated over time; but rather treats each data point as independent). The model permits efficiency to vary over time, but in a non-structured and random way;
2. Time invariant inefficiency panel models. These models recognise the panel structure of the data, but impose the assumption that inefficiency does not vary over time;
3. Time varying inefficiency panel models. These models recognise the panel structure of the data[4], but allow inefficiency to vary over time. Within this class of models, there is also a distinction between those models that attempt to decompose time invariant effects into inefficiency and unobserved heterogeneity.

The different models do have different properties, which may have advantages and disadvantages, but there is not necessarily a definitive a priori preferred model. It should also be highlighted that the different SFA models use various assumptions to obtain estimates of inefficiency, and their plausibility requires careful consideration prior to use in a regulatory context.

In the next section the results from the set of models listed below are presented and discussed; and a conclusion drawn on the potential usefulness of SFA (in general terms) as part of Ofgem's modelling approach going forward:

- Pooled model

- Time-invariant efficiency models:
    - Battese and Coelli (1988) – BC88:
    - Pitt and Lee (1981) – PL81

- Time-varying efficiency models:
    - Battese and Coelli (1992) – BC92[5]
    - True Random Effects - TRE
    - Kumbhakar, Lien and Hardaker (2014) – KLH4[6]

The above list is not intended to be exhaustive but to be indicative of the model

[4] Though, depending on the exact model, some inefficiency terms may be assumed to be independent across firms and over time.
[5] A more flexible model, Cuesta (2000), allows the direction of time path of inefficiency to vary by firm. Other models, such as Cornwell, Schmidt and Sickles (1990) could also be estimated.
[6] Kumbhakar et. al. (2014): Technical efficiency in competing panel data models: a study of Norwegian grain farming, Journal of Productivity Analysis, 2014, vol. 41, issue 2, 321-337.

types available. It is worth noting also the approach used by Farsi et. al. (2005)[7], which seeks to distinguish between unobserved heterogeneity and inefficiency using the Mundlak (1978)[8] approach, based on the assumption that inefficiency is uncorrelated with the regressors, whilst unobserved heterogeneity might be. The Mundlak approach can be combined with other models.

The results are also presented alongside the traditional OLS, random effects (RE) and fixed effects (FE) models for comparative purposes.

3. Stochastic frontier analysis (SFA): some initial results

The parameter estimates obtained from the different models are reported in Table 1 below.

Table 1: Parameter Estimates

	(OLS)	(RE)	(FE)	(SFA pooled)	(BC88)	(PL81)
Variables	log_Totex	log_Totex	log_Totex	log_Totex	log_Totex	log_Totex
log_Totex_csv	0.787***	0.727***	0.560***	0.790***	0.758***	0.766***
	(0.031)	(0.066)	(0.121)	(0.023)	(0.056)	(0.049)
Trend	-0.020***	-0.020***	-0.020***	-0.012***	-0.020***	-0.020***
	(0.006)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)
Constant	-0.144	0.268	1.405*	-0.278*	-0.026	-0.075
	(0.224)	(0.450)	(0.182)	(0.084)	(0.369)	(0.330)
Observations	48	48	48	48	48	48
Adj. R-squared	0.927	(0.930)	0.926			
Log-likelihood	64.585	80.376	96.580	65.936	80.849	80.770

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.10

[7] Farsi, M., Filippini, M. & Kuenzle, M., Unobserved heterogeneity in stochastic cost frontier models: an application to Swiss nursing homes, Applied Economics, 2005, v.37, iss.18, pp. 2127-2141.
[8] Mundlak, Y., On the Pooling of Time Series and Cross Section Data, Econometrica, 1978, 46, iss. 1, pp. 69-85.

Table 1 (continued)

Variables	(BC92) log_Totex	(TRE) log_Totex	(TFE) log_Totex
log_Totex_csv	0.815***	0.751***	0.579***
	(0.035)	(0.066)	(0.108)
Trend	-0.007	-0.020***	-0.021***
	(0.005)	(0.003)	(0.003)
Constant	-0.456*	0.074	
	(0.251)	(0.455)	
Observations	48	48	48
Adj. R-squared			
Log-likelihood	85.295	80.658	96.988

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.10

Table 1 shows that the parameter estimates on the included variables in the model are largely invariant to the choice of model specification, both between OLS and SFA, and within the class of SFA models. In assessing the statistical significance of the parameters in SFA models, we do have to remember the problems of drawing inference in finite samples for SFA models (here our sample size is only 48). It is worth noting that the Hausman test does not reject, thus indicating that it is reasonable to proceed with random effects (and indeed more generally with approaches, including OLS, which assume no correlation between the firm effects and the regressors).

In terms of the coefficients, the fixed effects model (and the true fixed effects counterpart) is perhaps the only model that deviates, with a coefficient on the CSV measure of 0.56 as compared to the other models, which put this coefficient broadly in the range 0.70-0.80. That said, the confidence interval for the fixed effects model extends from 0.31 to 0.81. In general, in economic regulation, fixed effects models have tended to produce less plausible results. Further consideration of the expected size of the coefficient would be useful to inform this debate further.

It is reassuring that the point estimates for the included variables are similar across models. The question then is which model or models might be preferred in respect of deriving efficiency scores for the companies, taking account of the reported test statistics and the underlying assumptions of the models. Here the aim is not to make definitive judgments with a view to selecting a particular model, but to offer general observations with the aim of forming a judgement on the appropriate use of this class of models going forward. Below we discuss each of the models in turn, and then draw conclusions about the general usefulness of SFA for Ofgem.

3.1 Pooled SFA

With respect to the pooled SFA model (half-normal), a likelihood ratio test indicates that the inefficiency effects are statistically significant at the 10% level, but not at the 5% level (although the test statistic is only just smaller than the required 5% critical value). These results suggest that the model struggles to some extent to disentangle inefficiency from random noise, which may reflect that the data does not support the particular assumptions of the model.

It must be remembered that the model treats the panel dataset as if it was a cross-section, which neglects an important feature of the data structure. That is, it assumes that all data points are independent. This means that inefficiency is assumed to be independently distributed over firms and over time. Thus if a firm is inefficient in one period, this has no bearing on whether it is likely to be efficient in the next period. This assumption is not very plausible in a regulatory context. It should of course be noted that the estimation of a (pooled) COLS model likewise ignores the panel structure of the data. That said, since the SFA approach is making specific assumptions about inefficiency, it is important to be clear about what those assumptions are and whether they are reasonable.

Since SFA can be sensitive to distributional assumptions, we also tested a truncated normal model; however this model failed to converge.

3.2 Time invariant models

Turning to the time invariant models, we estimate a half-normal time invariant (PL81) and truncated normal model (BC88). The test statistics indicate that the BC88 model does not add anything compared to the PL81 model and the results are almost identical (in respect of the parameter estimates and efficiency findings). Given that the panel is relatively short, the assumption of time invariant inefficiency may be considered plausible in many academic applications. However, in the context of a regulatory environment with strong incentives for efficiency gains within regulatory control periods, this is less clear.

The results indicate slightly lower average efficiency scores (greater inefficiency) for SFA than for COLS or RE (see Table 2) [9]; and in general, the cost allowances generated from the SFA models estimated are lower than those from OLS and RE. However, given the imposition of a time invariant efficiency assumption, it is not clear

[9] The cost allowances for OLS and RE (and FE) are computed after carrying out an upper quartile adjustment.

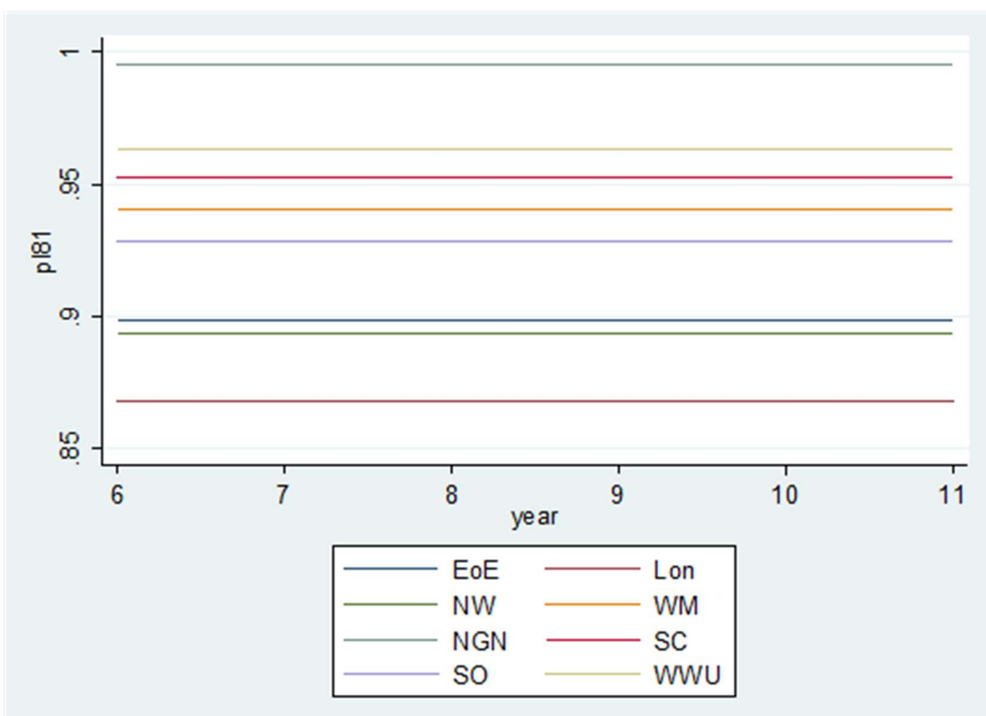
that the PL81 approach is beneficial as compared to RE, which also has the same assumption, but does not impose distributional assumptions. The efficiency scores for the PL81 model are shown graphically below in Figure 1 for each firm.

The obvious extension is to consider time varying inefficiency models. Particularly in respect of economic regulation, the selection of models that permit inefficiency to vary over time would seem to be preferable; ideally, with inefficiency varying in a structured way that recognises the panel structure of the data.

Table 2: Summary of Efficiency Scores

Variable	Obs	Mean	Std. Dev	Min	Max
Eff_pooledSF	48	.919	.054	.808	.992
Eff_pl81	48	.930	.040	.868	.996
Eff_bc88	48	.924	.041	.864	.994
Eff_bc92	48	.930	.053	.794	.993
Eff_tre	48	.986	.011	.955	.995
Eff_tfe	48	.978	.016	.914	.994
Eff_re	48	.947	.051	.852	1
Eff_fe	48	.931	.065	.803	1
Eff_cols	48	.941	.052	.832	1

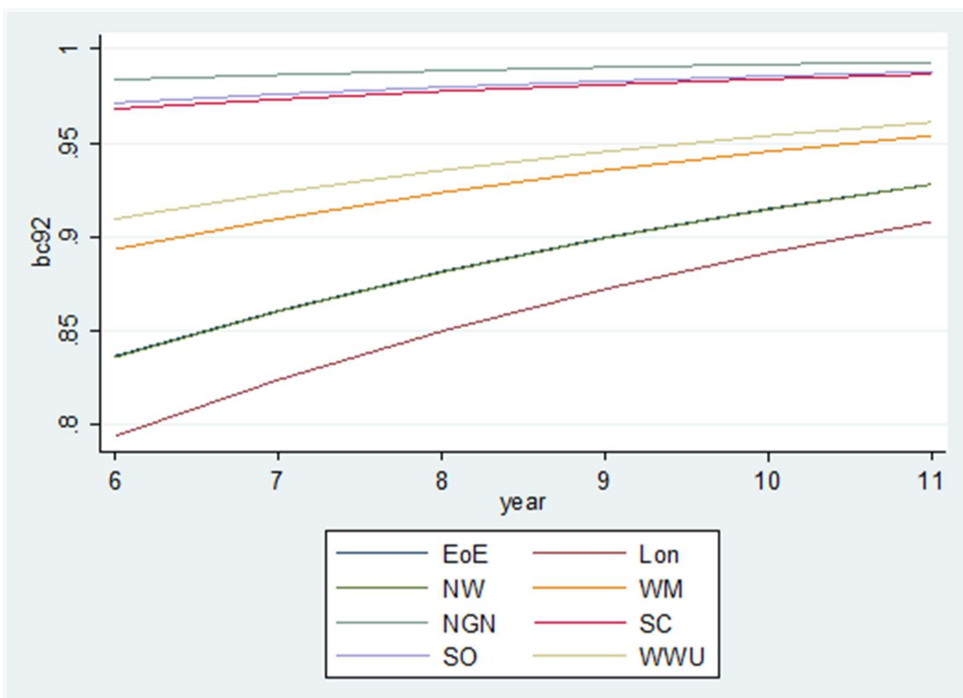
Figure 1: PL81 Efficiency Scores by Firm



3.3 Time varying efficiency models

The natural model to consider next is the BC92 model, as this contains the time invariant model (PL81) discussed in the previous sub-section as a special case. It is thus possible to test the restriction of time invariant efficiency. Further, the BC92 model explicitly recognises the panel structure of the data and thus permits inefficiency to vary over time in a structured way. One restriction of the model is that the direction of efficiency change must be the same for all firms – that is, the model should show either convergence of divergence. Thus, firm rankings stay the same over time, but firms either catch up or diverge from the frontier (but cannot over-take each other). A less restrictive version of the model (Cuesta, 2000) permits firms to have different directions of efficiency change over time, and this model should be tested in future work (this model was used by ORR in both PR08 and PR13)[10]. The model can also be made more general to allow turning points in the time path of inefficiency, which can be particularly useful for longer panels (and was used in the aforementioned work for ORR). The scores are shown graphically below in Figure 2.

Figure 2: BC92 Efficiency Scores by Firm



[10] See, e.g. Smith ASJ. 2012. The application of stochastic frontier panel models in economic regulation: Experience from the European rail sector. Transportation Research Part E: Logistics and Transportation Review. 48(2), pp. 503-515. The model could be tested alongside Cornwell, C., Schmidt, P. and Sickles, R.C., 1990. Production Frontiers with Cross-Sectional and Time-Series Variation in Efficiency Levels. Journal of Econometrics 46, 185-200. This model has a fixed and random effects version, and is a useful comparator as it does not impose distributional assumptions.

Statistical testing indicates that we can strongly reject the null hypothesis of time invariant inefficiency – and thus the change in inefficiency over time is statistically significant. The results are plausible and show some degree of convergence over time, as might be expected within a regulatory environment. The average efficiency is similar to that of the PL81 model (see Table 2), though the range of scores is slightly larger. This model has similar parameter estimates to the COLS model, except in respect of the time trend variable, where part of the time trend may have moved into the time varying inefficiency part of the model. It should be noted that the BC92 model does make a very specific assumption about the functional form of the time variation in inefficiency.

3.4 Models dealing with time invariant unobserved heterogeneity

One issue that must be faced when dealing with panel data is that there are likely to be fundamental differences between firms, referred to in the literature as unobserved heterogeneity. Such heterogeneity causes costs to vary between firms, for reasons that have nothing to do with inefficiency. The SFA approach in general seeks to decompose random noise from inefficiency, but this may not fully address the point that fundamental differences between firms – relating to, for example geography – are likely to be time invariant. As a result, the academic literature has sought to develop models that separate inefficiency from such factors.

The true fixed and random effects models assume that inefficiency is time varying, whilst unobserved heterogeneity is time invariant. Thus, the standard fixed and random effect terms are deemed to capture time invariant, unobserved heterogeneity, whilst time varying inefficiency is captured by decomposing the error term using the usual SFA assumptions. With the current data, the test statistics indicate that these models do not add much to the underlying standard fixed and random effects models. The parameter estimates are almost identical and the log-likelihood values change only marginally.

It is also the case that the model assumptions of true fixed and random effects may not be particularly plausible given that some inefficiency may be persistent (time invariant), so the random effect term may partly be hiding inefficiency as well as unobserved heterogeneity. Related to that, the notion that inefficiency is independently distributed over time is not very palatable in a regulatory environment as it implies that if a firm is inefficient in one period this has no impact on its probability of being inefficient in the next period. However, this latter assumption is common amongst most SFA models, with the exception of some of the models discussed in section 3.3.

One attempt to address the issue that inefficiency may comprise an element that is persistent, and a time varying element, is the “four component” model put forward by Kumbhakar, Lien and Hardaker (2014). This model is often referred to as the KLH model. In a regulatory context, it can be seen as a relatively complex model in which the error term is decomposed into four components: (1) time invariant unobserved heterogeneity; (2) time invariant inefficiency; (3) time varying inefficiency; and (4) time varying unobserved heterogeneity or random noise.

The multi-stage version of this model has been estimated with the gas distribution data. The first stage is simply a random effects model. The second and third stages then decompose the random effect and the error term respectively into time invariant and time varying effects using a standard SFA model.

It should be noted that, in respect of the time invariant component of inefficiency, this model will not change the rankings imposed by ordering companies according to the random effects from the standard random effects model (though the scores could vary). Further, the decomposition of the time invariant effects into inefficiency and unobserved heterogeneity is based on a SFA model with just 8 data points (this model only uses the cross-section variation), which makes the results highly questionable. Not surprisingly therefore it was not possible to get statistically significant inefficiency effects from this model (either for the persistent of time varying component).

3.5 Conclusions on SFA

The above results indicate that the parameter estimates in the models are in general robust to the estimation method used. Whilst SFA could have a role to play in the regulatory framework, it has to be remembered that the models impose numerous assumptions. These include assumptions about the distribution of the inefficiency and random noise terms (e.g. half-normal, exponential, truncated normal etc.), which may be challenged. The assumption that inefficiency is distributed randomly over time (applied in many SFA models) with the panel data at hand is also problematic.

Likewise, the assumption that inefficiency is invariant over time may not be plausible in a high-powered incentive regulation framework. At the same time, other models utilise the assumption that inefficiency is time varying, whereas unobserved heterogeneity is not, which may not be plausible. Very complex models are then needed to relax this assumption, but their application to a situation with just eight cross-sections is highly questionable, particularly given the implications for transparency. Further, in respect of the time invariant component of inefficiency, such models do not in any case overturn the company rankings that emerge from a standard random effects framework (though the scores could vary).

Perhaps the most appropriate SFA models come from the class of approaches that recognise the panel nature of the data and permit inefficiency to vary over time in a structured manner. Here we estimated one such model (BC92) to illustrate the possibilities, though recognising that there are other models of a similar type that can be tried. These models enable the unrealistic assumption of independence in inefficiency over time (a problem that plagues many comparator models) to be relaxed, which is important in a regulatory context[11].

The weakness of the above models is that they do not explicitly decompose time invariant inefficiency from time invariant unobserved heterogeneity. However, our initial results indicated that the so called "true" and KLH models did not add much to the alternatives. It is also the case that unobserved heterogeneity could, in part, be dealt with via regulatory judgement – which in essence is what is implied by choosing an adjustment to the COLS framework, for example choosing the upper quartile. More widely, given the transparency benefits of OLS, and its well-known properties in small samples, there is much to be said for retaining the existing COLS framework, comparing also against standard random effects.

That said, my view is that SFA should be tested against the COLS / RE models, but from the perspective of requiring good reasons to depart from the current COLS / RE framework, given the nature of the assumptions of many of the SFA models, and the small sample size. Of the SFA models, I would consider the class of models that permit structured time variation in inefficiency (e.g. BC92 and similar models), the most likely to deliver plausible results, and for which the assumptions of the model might be seen to be reasonable and relatively transparent in a regulatory environment. There is also regulatory precedent in the UK for such models. The remaining alternatives seem unlikely to add much to the current framework, given their assumptions, the sample size, and the preliminary analysis done.

Ultimately, only detailed analysis and comparison of the results will yield a final model selection decision, and Ofgem would need to balance the various model selection criteria, including taking into account that of transparency, which may weigh against the more complex SFA approaches. As noted in the introduction, if SFA models are to be used, it seems most sensible to restrict their application to Totex models.

As a final caveat, it should be noted that we have only estimated models using a single, composite cost driver in the model. If it was deemed necessary to depart from that approach, then the different SFA approaches would need to be tried for the alternative possible model specifications (in terms of included variables in the cost function).

[11] See, for example, Alvarez et. al. (2006). Alvarez, A., Amsler, C., Orea, L. and Schmidt, P., 2006. Interpreting and testing the scaling property in models where inefficiency depends on firm characteristics. Journal of Productivity Analysis 25, 201-212.

4. Data envelopment analysis (DEA): some initial results

In the previous note (Smith, 2019), a number of issues were highlighted concerning the likely value of using DEA in this context. As discussed further below, some of these weaknesses can partly be overcome by relying on advances in the literature beyond traditional DEA implementations, but many of these models might be considered highly complex and not necessarily applicable in a regulated environment. The challenges in using DEA are as follows.

Firstly, in a model with just a single output (e.g. CSV), constant returns to scale (CRS) DEA is simply a ranking of data points based on unit costs which is not particularly informative. Further, we would not want to impose the assumption of CRS, given the findings from econometric work. Second, one of the advantages of the econometric framework is that is produces estimates of the elasticity of costs with respect to scale (and other drivers, where relevant), also enabling statistical testing of the scale properties. This permits discussion of the results and whether they are deemed plausible. Whilst a variable returns to scale (VRS) variant of DEA is available, most DEA studies do not report scale elasticities (though the literature does contain methods, for deriving scale measures from DEA, and an additional computation is needed). Any elasticities computed from a DEA model would need to be subject to the same degree of scrutiny that has applied in the UK regulated context in respect of the econometric approach; in order to demonstrate their plausibility at all points in the sample.

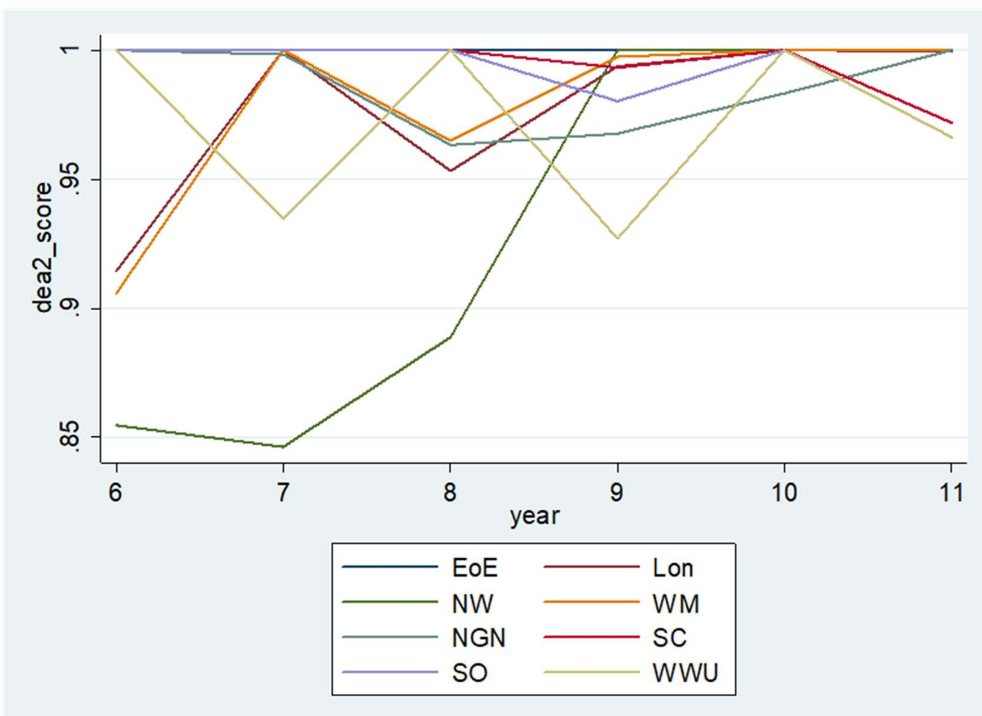
Third, in its traditional form, the model does not deal with the problem of decomposition of cost gaps into random noise and inefficiency (though this could be dealt with via an upper-quartile-type adjustment as in COLS). Although more advanced approaches have been developed in the literature, these approaches start to blur the distinction between DEA and statistical methods, and would appear, without further justification, to be too complex for application in a regulated environment with the small sample sizes available. It is further not clear any benefits would warrant the increased complexity, particularly taking into account the potential implications for transparency.

The final weakness of DEA is that it can struggle to differentiate between firms in the case of many inputs and outputs (which could occur if the individual elements of a composite scale variable are included separately in the model); particularly in small samples. This situation can be made worse if combined with the VRS assumption. In the case of multiple outputs, DEA may give zero weights to some outputs, which can be considered a problem in a regulatory framework. This would indicate firms with a particularly high ratio of one output to one input are efficient, which may not be credible in a regulatory context. Whilst restrictions can be placed on DEA (such as minimum weights to be applied to outputs), it is not clear on what basis these restrictions should be set.

Overall, the previous note was sceptical of the value of DEA. Some of the issues raised could perhaps be dealt with by moving to much more complex versions of DEA, which then start to blur the distinction between econometric and DEA approaches. Below some standard DEA models have been estimated in the first instance, to indicate the potential for this approach to be utilised or developed further.

The first model estimated was a CRS Totex model, with a single output (CSV). As noted, this model is not very informative as it is simply a comparison of unit costs. Further, the findings of the econometric models indicate cost elasticities of less than one, so we would not want to impose the assumptions of CRS. The models shown below therefore assume VRS. The first model has multiple outputs based on the components of CSV (Figure 3); the second has a single output (CSV), as in Figure 4.

Figure 3: VRS DEA Multiple Outputs (components of CSV)

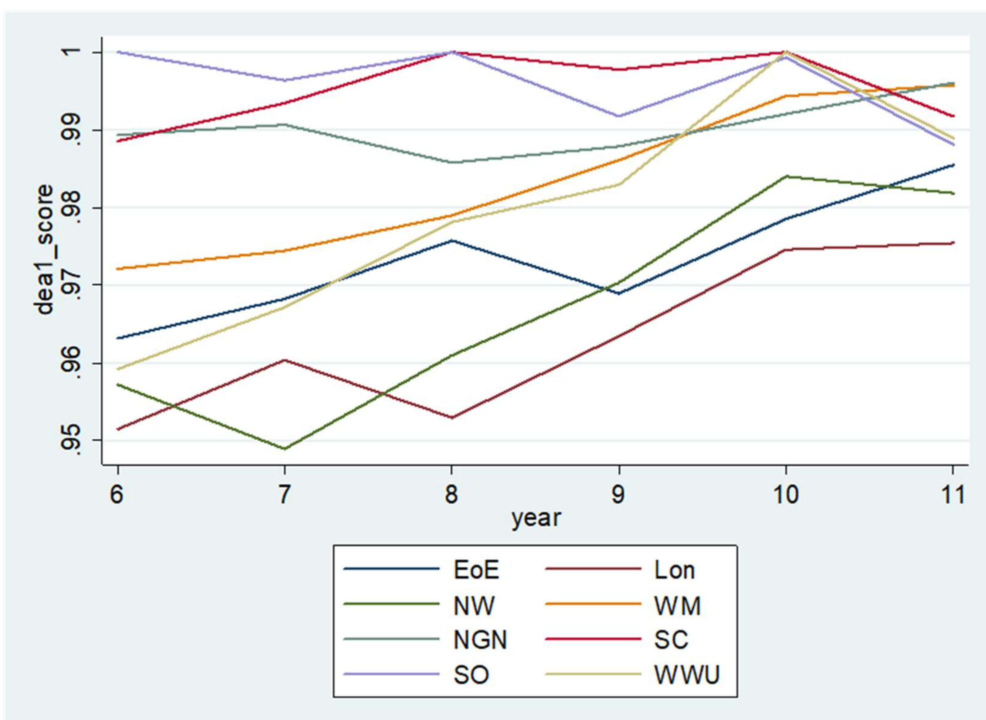


The results in Figure 3 largely confirm the prior expectations set out above. For most firms the efficiency scores fluctuate within a narrow range close to one, and each firm hits the frontier in at least one of the years. One firms goes from having an efficiency gap of 15% to being on the frontier within two years. Overall, these results do not seem plausible and could not realistically form part of the regulator's cost assessment methodology.

Figure 4 shows the results for a single output (CSV), VRS model. This model likewise puts most firms either on or very close to the frontier, with scores between 0.95 and unity, or between 0.97 and unity in the final year. The model therefore is not able to differentiate firm performance very well, and I do not consider this to be a plausible approach.

As set out in the previous note (Smith, 2019), there are other possible approaches that could be used, including two stage approaches, and also Malmquist DEA approaches. Some attempt could be used to restrict the weights possibly, as discussed above, and further analysis of peers and scale findings would enhance transparency. However, it seems unlikely that such approaches will produce plausible results, though they could be tried for completeness. With the above caveats, my view is that at best DEA could only be used as a cross-check against other approaches.

Figure 4: VRS DEA Single Output (CSV)



5. Summary and conclusions

This short note has set out some indicative analysis of alternative SFA and DEA approaches with a view to forming a view as to the value of using such approaches within Ofgem's suite of models. The overall conclusion is that DEA does not seem to offer great promise in the current context, and at best could be used as a cross-check against other approaches, and even then further analysis would be warranted. In terms of priorities, I would therefore argue that developing the other approaches is a better use of resources, but that some further limited investigation of DEA could be carried out as well for completeness. It should be noted that some of the advances in the literature in the DEA space could lead to modelling becoming increasingly complex, creating approaches that lie between parametric (econometric) and non-parametric (e.g. DEA) methods. Based on the initial results such approaches do not seem warranted.

I do not consider that SFA is necessarily a panacea to solve all the perceived problems of the COLS/RE approaches. OLS is a familiar and transparent method, with well-known properties. However, I do consider that SFA models should be tested alongside COLS. Ofgem can then select its preferred approach based on its model selection criteria, exercising its judgement in the standard way. Of the available SFA models, I consider those that explicitly recognise the panel structure of the data and permit efficiency to vary over time in a structured way to be the most likely to add value (the results of one such model, BC92 is presented in this note, but there are others with more flexible properties). These models also make reasonable assumptions and are relatively transparent for a regulatory case. There is also regulatory precedent in the UK for such models. The remaining alternatives seem unlikely to add much to the current framework, given their assumptions, the sample size, and the preliminary analysis done. However, they remain relevant for comparative purposes.

Ultimately, only detailed analysis and comparison of the results will yield a final model selection decision, and Ofgem would need to balance the various model selection criteria, including taking into account that of transparency, which may weigh against the more complex SFA approaches. However, transparency / simplicity should not be prioritised at the expense of statistical robustness or economic plausibility, although perhaps a strong preference in favour of SFA would be needed in order to deviate from the current approach.