Commentary on the report by Frontier Economics to Ofgem on the feasibility of econometric benchmarking in DNO cost regulation

Final version

RJ Gibbens

S Zachary

May 2013

1. Introduction

In April 2013 Frontier Economics (FE) submitted a report [1, 2] to Ofgem on the feasibility of using an econometric approach to estimate the relative efficiency, and hence assist in the cost regulation, of the 14 GB Distribution Network Operators (DNOs). In particular the report examined the possibility of using **total expenditure benchmarking** (*totex*) as the primary measure of observed company costs.

We have a number of serious reservations both about the fundamental econometric approach in this setting—of which the single most important is, we believe, the logical inability of such an approach to distinguish between the *heterogeneity* of DNO operating environments and the *efficiencies* of individual DNOs, and with regard to the conclusions of the FE report itself, in which we believe various aspects of the statistical analysis to be seriously flawed.

2. The econometric approach to DNO benchmarking

Useful background reading on the econometric approach to DNO benchmarking is given by Greene [3], by Stern [4], by Haney and Pollitt [5], and in the short and very valuable guide by Schmidheiny [6], the latter dealing in particular with fixed and random effects in panel data.

The underlying approach for DNO benchmarking is that of the construction of a statistical model in which the costs of the 14 DNOs (e.g. *total expenditure*) are appropriately regressed against a number of explanatory variables (*number of customer connections*, *peak load*, *network density*, *wage costs*, etc). This typically leaves much residual variation which is further decomposed as a *company effect*—constant over observations for each company—plus a smaller amount of residual variation corresponding to differences in company performances between the successive years for which data is available. Thus the mathematical form of the model is

$$y_{it} = \sum_{j=1}^{k} \beta_j x_{jit} + \sum_{i=1}^{14} \alpha_i + \epsilon_{it} \tag{1}$$

where

- y_{it} is an appropriate transformation of the chosen expenditure measure for company i in year t,
- x_{jit} is an appropriate transformation of the observed value of the explanatory variable j for company i in year t,
- β_j is the corresponding regression coefficient for the variable x_j,

- α_i is the company effect for company i which is assumed to be the same for all years t,
- ϵ_{it} is the residual for company i in year t.

Additionally the model (1) may include an intercept, corresponding to normalising the company effects α_i so that they sum to zero. However, whether or not such an intercept is included makes no difference to either the analysis or its conclusions—it is simply a matter of presentation.

The chosen form of the model (1) in the FE study is that of the *Cobb-Douglas* production model, in which in general each x_{jit} is the logarithm of the corresponding observed value of the explanatory variable. (For more on the appropriateness of this model in the current study, see Section 3.)

The *company effects* α_i are ascribed to some mixture of cost contributions from unidentified exogenous variables—e.g. difficulty of operating terrain—and *inefficiencies* associated with each company. (In the present FE study the company effects appear to be identified totally with inefficiencies.) These effects are used as the basis of company rankings and in some sense incorporated in the cost regulation process.

We have three very serious reservations with regard to the use of such an approach in the present context.

1. *The identification of the explanatory variables to be used in the model.* In the above econometric approach, such variables must necessarily include readily determinable economic measures of the volume of those services delivered by each company; this therefore focuses attention on such measures as *number of customer connections* and *peak load*, and perhaps others such as population or network *density* within each company area. But in reality such variables are typically little more than *proxy* measures for those costs which are necessarily incurred by the companies concerned. These output measures do at least have the virtue that they are difficult to distort; nevertheless it remains the case that they are merely typically *correlated* with such costs, and *changes* in these variables do not necessarily result in anything which resembles the cost changes suggested by the fitted regression models. For example, additional customer connections cost very little and certainly very much less than would be suggested by the fitted models; the same is true of additional peak load. It follows, we believe, that the use of such variables is at best *highly* unreliable, and—as is always the case with the use of proxy variables in statistical analysis, even when their values are not artificially distorted for gaming purposes—introduces much residual variation which *cannot* be ascribed to such factors as inefficiency.

In our opinion it follows that the use of proxy measures is to be avoided if at all possible. Rather what is required is the additional effort—and indeed understanding of the industry cost structure, contrary to the *totex* philosophy—needed to identify those variables, such as network characteristics, which are *causally* connected to company costs—even if these variables are not readily identifiable by the application of purely statistical techniques (such as those considered in the present FE study).

2. *The identification of residual (company) variation with unexplained costs and/or inefficiency.* We believe that in reality it is almost impossible to decompose the residual variation in the model, i.e. that left over after removing the fitted contributions of the chosen explanatory variables, as a sum of contributions from costs due to unmeasured exogenous variables which are essentially constant over observations for each company, and costs due to the inefficiencies of individual companies. If the unmeasured exogenous variables are essentially constant over time then, from a deterministic point of view it is completely impossible: *any* unexplained variation which might be ascribed to company inefficiency might equally well be ascribed to exogenous variables whose values differ between companies—see the next point for a more detailed discussion of this case. The decompositions occasionally claimed by econometric models, such as stochastic frontier analysis, seem to rely entirely on probabilistic assumptions about differences between the shapes of the distributions of exogenous variables and "inefficiency" variables. For such a separation to be at all reliable it is necessary that (a) the distributional assumptions made should correspond very closely indeed to reality, and (b) extremely large quantities of data should be available. Typically neither of these assumptions is remotely satisfied.

3. *The inclusion of variables which do not change significantly over time.* The inclusion of variables which do not change over time introduces collinearities into the fit of the model (1) which means that the regression coefficients **cannot be determined**. (If the constancy over time is merely approximate, then the regression coefficients may in principle be estimated, but the associated confidence intervals will be extremely wide.) Such variables are typically those such as might be used to capture *heterogeneity* between companies—a measure of *difficulty of operating terrain* or *density of network* will typically not change significantly over time. To be precise, if the variable j is such that, for each company i, the measured values x_{jit} of the variable are constant over time t then the corresponding regression coefficient β_j may be adjusted to any value we like and, by correspondingly adjusting the *company effects* α_i, the overall fit of the

model may be retained exactly as before (so that the residuals ϵ_{it} are unchanged). The problem in the case of variables which represent heterogeneity, such as those mentioned above, is simply an extreme case of the difficulty noted earlier, namely the difficulty in models such as (1) being able to distinguish between *heterogeneity* and *efficiency*. Nor, of course, will it do to leave out of the model variables representing heterogeneity as this simply redefines all heterogeneity as efficiency.

It appears to be sometimes argued, as in the present FE study, that the use of a *random effects* model—in which the company effects α_i are treated as observations of random variables—enables such models to be satisfactorily fitted after all. In our opinion this is simply not the case. The difficulty is a **logical** one, implicit in the functional form (1) whenever this includes one or more variables which, in addition to the company effects, are constant (or almost so) over time. The use of a random effects model merely introduces certain implicit assumptions—in essence that some distributions of the *company effects* α_i are more plausible than others—independently of the data, and this additional information fed into the model enables some sort of estimates to be made in principle. However, the precision of such estimates is *extremely* poor—one is really getting out of the analysis nothing more than the additional assumptions fed in. Software for fitting random effects models typically does not attempt to produce confidence intervals for such effects, as they are not formally speaking parameters to be estimated.

A nice demonstration of the failure of the above approach is given by using a Bayesian analysis, in which the distinction between fixed and random effects disappears (formally everything is now random) and in which confidence intervals *are* obtainable. Again, in the presence of explanatory variables which are almost constant over time, these confidence intervals are impossibly wide. In the next section we report the results of such an analysis of the data of the present FE study.

3. The present modelling and analysis

The present Frontier Economics study is based on the philosophy outlined in the preceding section. The functional form of the preferred statistical model is that of the Cobb-Douglas production model. This is transformed to the model (1) by taking as the response y the logarithm of the chosen measure of *company costs* (for each company in each year), and as the explanatory variables x_j the logarithms of *customer numbers*, *peak load*, a measure of network *density*, and measures of *wages* and the *price of capital* (the latter two constrained to have regression coefficients summing to one—see below).

We have thus all the concerns with this approach discussed in the previous section. We further have more detailed concerns as follows.

1. *Heterogeneity and efficiency.* **Our overriding concern** is what we believe is the almost complete failure of the analysis to discriminate between *heterogeneity* and company *efficiency*—as discussed jointly in the points 2. and 3. of Section 2. *Heterogeneity* is represented by the *density* variable—there is much discussion in the FE report as to the chosen measure of this variable, and as to whether it sufficiently captures heterogeneity between companies and their operating environments; *efficiency* is represented by the company effects α_i in the model (1). The primary goal of the entire analysis is to estimate the latter, and yet the FE study seems to make little or no attempt to measure the *precision* of these estimates (as measured by standard errors or confidence intervals). However, we believe that the degeneracy in the model is such that this is essentially impossible.

   Figure 1 shows a plot of (the logarithm of) network *density* for each company and for each year. It will be noted that, for every company, the densities are almost constant over time. This in itself suggests the almost total impossibility of discriminating between *heterogeneity* and company *efficiency* (for the reason given in point 3. of the previous section), and in particular of estimating the efficiencies of the individual companies. This is confirmed in the *fixed effects* analysis, where the FE report itself points out that the collinearities in the data—resulting from the constancy over time of *density*, and also as it happens of *customer numbers*—mean that this model may not be reliably fitted. But, as we have pointed out in the previous section, in our view the *random effects* analysis merely ignores this problem—adding a little information-theoretic fuzz to the data so that fitting seems *in principle* possible, and then failing to report the extremely wide standard errors which should be associated with the estimates of the *company effects*.

   In order to obtain these standard errors, or confidence intervals, we have carried out a *Bayesian* analysis. The Bayesian versions of the fixed and random effects models *coincide*—all parameters are regarded as random having prior distributions from which posterior distributions may be obtained by the use of the data. We have used the usual "improper" prior distributions of parameters in the linear model, corresponding to the case where no significant prior information is available about their likely values—though we believe that our conclusions are unaffected by any remotely plausible choice of prior distributions. The results (obtained

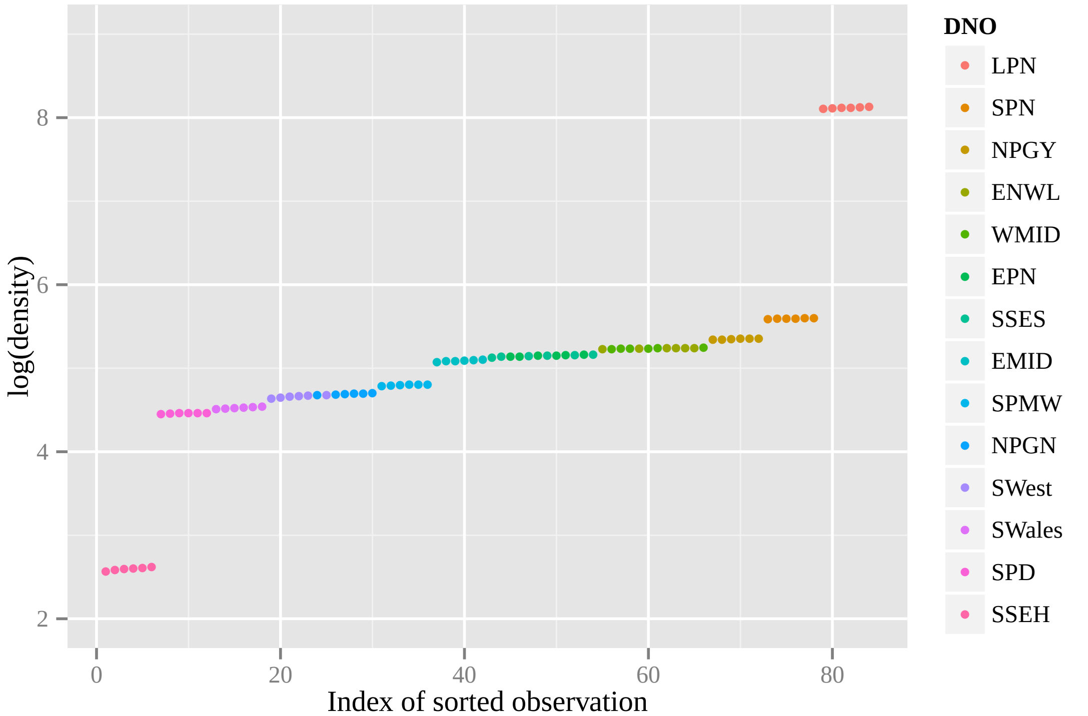

Figure 1: Figure showing that log(density) is essentially collinear with the company DNO effect.

with the use of the R function `bayesglm` in the package `arm` and showing parameter estimates, together with their standard errors, t-values and p-values) are shown in Table 1. We have here chosen not to normalise the *company effects* α_i to sum to zero, but rather to omit the intercept—as we have earlier remarked this is a purely presentational matter.

Figure 2 shows a dotplot of all the estimated parameters, together with the 95% *confidence interval* (in strict Bayesian terminology we should say *credible interval*) for each. It will be seen that all of these confidence intervals, other than that for *wage*, include the value zero, i.e. there is no evidence that anything else whatsoever enters into the model.[1] In particular the confidence intervals for the 14 company effects so overlap each other that none differs significantly from any other. Hence we believe that in this model **essentially no measure of company efficiency is possible**.

2. *The preferred statistical model.* The deterministic component of the model

[1]One might reasonably expect that either *customer numbers* or *peak load*, as measures of size of company business, *should* enter significantly into the model. Indeed this is the case for *peak load* if the *density* variable is omitted from the analysis. The collinearities introduced by the inclusion of this variable appears to have a destabilising effect on the entire analysis.

	Estimate	Std. Error	t-value	Pr(>|t|)
log(customers)	-0.0925	0.1237	-0.75	0.4569
log(peak)	0.2886	0.1451	1.99	0.0500
log(density)	-0.0011	0.2549	-0.00	0.9966
log(wage)	0.2967	0.1009	2.94	0.0043
EMID	0.1358	0.2938	0.46	0.6451
ENWL	0.1137	0.2930	0.39	0.6990
EPN	0.4548	0.3012	1.51	0.1348
LPN	-0.1067	0.8254	-0.13	0.8974
NPGN	-0.1482	0.2931	-0.51	0.6145
NPGY	-0.0352	0.2970	-0.12	0.9060
SPD	0.0205	0.3246	0.06	0.9497
SPMW	-0.0125	0.2844	-0.04	0.9651
SPN	0.0931	0.3178	0.29	0.7703
SSEH	-0.3599	0.6020	-0.60	0.5515
SSES	0.1555	0.2980	0.52	0.6033
SWales	-0.4256	0.2894	-1.47	0.1451
SWest	-0.1872	0.2936	-0.64	0.5255
WMID	0.2263	0.2930	0.77	0.4420

Table 1: Bayesian analysis: table of estimated coefficients and the associated standard errors.

used in the FE study is that of the use of the Cobb-Douglas production function. Thus, in the model (1), both the response variable y and explanatory variables x_j are transformed by taking logarithms of the original variables.

It is important that the variables in the model and the allowed values of the parameters are such that the model is economically realistic, and is capable of being reasonably reliably fitted from the available data. In this respect two weaknesses of the present analysis are as follows.

(a) *The scaling properties of the model.* Because company sizes vary considerably, and there is no allowance for this other than in the functional form of the model itself, it is important that the model scales correctly as company size is varied. For this to be so requires that the coefficients of those explanatory variables which grow in proportion to the size of the company should, in the multiplicative Cobb-Douglas model, sum (to a very good approximation) to one. In the present model, the two variables concerned are *number of customers* and *peak load*. In the reported FE analyses their coefficients appear to sum to somewhere around 0.81 or 0.82. This is considerably less

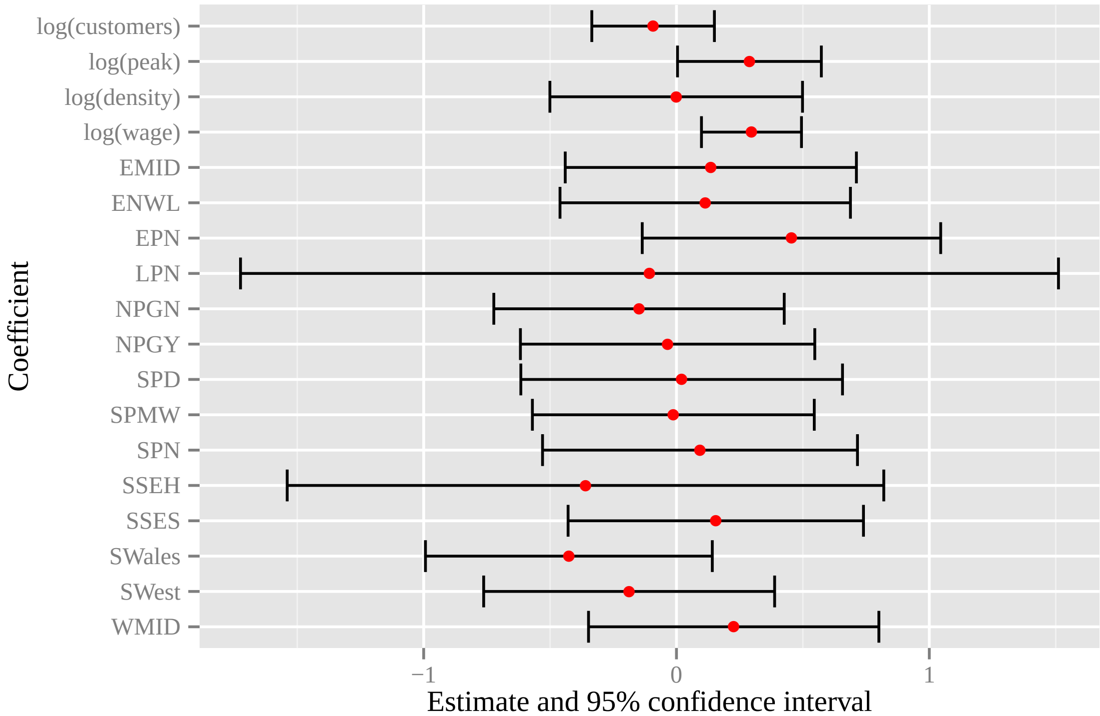

Figure 2: Bayesian analysis: figure showing the estimated coefficients and the associated 95% confidence intervals.

than one, and is explained, in the words of the report (p 39), as representing "modestly increasing returns to scale". But in reality it represents a major instability in the model, if company effects are to be interpreted as measures of efficiency. To see this clearly consider—as a thought experiment—amalgamating two or more of the companies; their estimated company effects, under any of the fitted models, will then increase dramatically, completely changing their efficiency rankings.

In summary, increasing *returns to scale* correspond to a size effect for which the model makes no allowance, and which therefore greatly distorts the efficiency scores.

(b) *The number of genuinely independent companies.* It might be argued that there are only 6 genuinely independent companies, instead of 14. Reanalysing the data on this basis may be expected to produce very different results. This is partly because of the *scaling* problems discussed in the previous paragraph, and partly because the data appropriate for a smaller number of genuinely independent companies obviously have a lower information content, resulting in larger standard errors and confidence intervals.

3. *A further major problem with the density variable.* Network densities, as measured by *meter density*, are broadly similar for all companies except two. These are LPN, for which meter density is extremely high, and SSEH for which meter density is (on the logarithmic scale used in the model) extremely low. (See Figure 1, which plots *density* both by year and by company.) The results of this is that the influence of the observations for these two companies is so high that the regression coefficient for *density* is almost entirely determined by the observations for these two companies (the other companies having a much smaller effect). The consequence is that it is impossible to make any estimate of the efficiency of either of these two companies: whatever their recorded costs the regression model will simply readjust its fit so that these companies each have a mid-ranking performance. This is well illustrated in Figure 2, which shows confidence intervals for the company effects α_i (of which the efficiency scores are a monotonic transformation, and so in particular have the same ranking). The confidence intervals for these two companies are huge, confirming that, even if one were to accept the identification of the *company effect* with *efficiency*, the entire analysis has almost nothing to say about either of these two companies. Nor will it do to leave out *density* since, *a priori*, it is clearly a variable which has an important influence on costs—as do a number of other not-included environmental variables.

4. Conclusions

We are of the opinion that, for all the reasons outlined above, an econometric approach to DNO benchmarking in the present GB context is *so* unreliable as to produce efficiency scores which might almost as well have been randomly generated. We therefore believe that the nature of the problem, **and of the available data**, is such that the proposed approach is simply not feasible for this purpose. In particular, the fact that *heterogeneity* measures change little over time means that this variable cannot be separated from the company *efficiency* which it is the purpose of the exercise to measure.

We have not discussed the further difficulties of the specifically *totex* approach, in which it is necessary to somehow further account for *capital expenditure*.

References

[1] Total cost benchmarking at RIIO-ED1, Phase 2 report. Vol 1. April 2013. Report prepared for Ofgem by Frontier Economics Ltd.

[2] Total cost benchmarking at RIIO-ED1, Phase 2 report. Vol 2. April 2013. Report prepared for Ofgem by Frontier Economics Ltd.

[3] Greene, William. *Reconsidering heterogeneity in panel data estimators of the stochastic frontier model.* Journal of Econometrics. Elsevier. 126 (2005) 269–303.

[4] Stern, Jon. *Econometric Benchmarking and its Uses by ORR: a Review.* Centre for Competition and Regulatory Policy. City University London. CCRP Working Paper No 21. February 2013.

[5] Haney, Aoife Brophy and Pollitt, Michael G. *International Benchmarking of Electricity Transmission by Regulators: Theory and Practice.* Electricity Policy Research Group. University of Cambridge. EPRG Working Paper 1226. November 2012. http://www.eprg.group.cam.ac.uk/wp-content/uploads/2012/12/1226-Final.pdf

[6] Schmidheiny, Kurt. Panel data: Fixed and Random Effects. Short Guides to Microeconometrics. Universität Basel. Fall 2012. http://www.schmidheiny.name/teaching/panel2up.pdf

Richard Gibbens is a Reader in Network Modelling at the University of Cambridge. He has a MA degree, a Diploma in Mathematical Statistics and a PhD from the University of Cambridge. He held a Royal Society University Research Fellowship in the Statistical Laboratory, University of Cambridge before joining the Computer Laboratory. He has research interests in mathematical modelling that span communication networks, transport networks and energy networks and collaborates with colleagues in academia, government and industry in these areas.

Stan Zachary is Reader in Statistics at Heriot-Watt University, Edinburgh. He has an MA degree in mathematics and an MMath degree from the University of Cambridge, and a PhD in probability and statistics from the University of Durham. He is also a Fellow of the Royal Statistical Society. He has an extensive research record in probability and statistics, particularly in the area of stochastic networks, heavy-tailed distributions and extreme events, and has written or edited books and organised many international research meetings focused on these areas.