Total cost benchmarking at RIIO-ED1 – Phase 2 report – Volume 1

A REPORT PREPARED FOR OFGEM

April 2013

Total cost benchmarking at RIIO-ED1 – Phase 2 report – Volume 1

1 Introduction 7

1.1 What is total cost (totex) benchmarking? ... 7

1.2 Objectives of the study ... 8

1.3 Phase 1 findings ... 8

1.4 Phase 2 objectives ... 9

1.5 Structure of this report ... 9

2 Methodology 11

2.1 Overview ... 11

2.2 Costs ... 12

2.3 Cost drivers ... 14

2.4 Sample ... 17

2.5 Technique ... 18

3 Further investigation into input prices 23

3.1 Phase 1 ... 23

3.2 Phase 2 approach ... 23

3.3 Capital price series ... 24

3.4 Labour price series ... 29

3.5 General inflation indices ... 36

3.6 Results ... 36

3.7 Conclusion ... 48

4 Alternative measures of density 49

4.1 Phase 1 ... 49

4.2 Phase 2 approach ... 49

4.3 Developing sub-DNO measures of density ... 50

4.4 Results ... 61

4.5 Conclusion ... 75

5 Accounting for quality of supply 77

5.1 Phase 1 77

5.2 Phase 2 approach 77

5.3 Results and conclusions 81

6 Summary of key results 83

6.1 Confirmation of model coverage 83

6.2 Results of our preferred model 85

6.3 Comparison of efficiency scores under RE and POLS 87

Annexe 1: Density data 89

Annexe 2: Meter density histograms 93

Annexe 3: Testing different density measures 101

Annexe 4: Density outlier analysis 107

Annexe 5: Regression results excluding outliers 111

Total cost benchmarking at RIIO-ED1 – Phase 2 report – Volume 1

Figure 1. The four main components of a regulatory benchmark 12

Figure 2. Generic cost function 15

Figure 3. Indices considered as potential proxies for capital input prices 27

Figure 4. Correlations between capital series 28

Figure 5. Evolution of labour input price series considered, using national averages 31

Figure 6. Correlation between labour input price series using national averages 32

Figure 7. Regional mapping of ASHE regional averages 34

Figure 8. SIC-35, Regional wages 36

Figure 9. Scatter of efficiency scores under Specification 1 and Specification 2 48

Figure 10. Example of histogram of meter density; Yorkshire 52

Figure 11. Example of histogram of meter density; South Wales 52

Figure 12. Example of histogram of meter density; Scottish Hydro 53

Figure 13. Example of histogram of meter density; London 53

Figure 14. Scatter plot of basic Phase 1 density measure against mean meter density (2010) 58

Figure 15. Scatter plot of basic Phase 1 density measure against mean meter density (2010); excluding LPN 59

Figure 16. Rates used to monetise the quality of supply delivered by DNOs during DPCR4 (first two tables) and DPCR5 (last table) 78

Figure 17. Quality of supply indicators. CIs and CMLs (IIS weighted) 80

Figure 18. Monetised values derived from quality of supply performance 81

Figure 19. Scatter of efficiency scores under Specification 1 and Specification 2 86

Figure 20. Comparison of meter numbers from Frontier's analysis with customer numbers provided by Ofgem 92

Figure 21. Histogram of meter density; EMID 93

Figure 22. Histogram of meter density; ENW 94

Figure 23. Histogram of meter density; EPN 94

Figure 24. Histogram of meter density; LPN 95

Figure 25. Histogram of meter density; NEDL 95

Figure 26. Histogram of meter density; SP 96

Figure 27. Histogram of meter density; SP Manweb 96

Figure 28. Histogram of meter density; SPN 97

Figure 29. Histogram of meter density; SSE Hydro 97

Figure 30. Histogram of meter density; SSE Southern 98

Figure 31. Histogram of meter density; WMID 98

Figure 32. Histogram of meter density; WPD SWales 99

Figure 33. Histogram of meter density; WPD SWest 99

Figure 34. Histogram of meter density; YEDL 100

Figure 35. Regression replacing basic density with mean meter density; Specification 1 101

Figure 36. Regression replacing basic density with mean meter density; Specification 2 102

Figure 37. Regression replacing basic density with standard deviation of meter density; Specification 1 102

Figure 38. Regression replacing basic density with standard deviation of meter density; Specification 2 103

Figure 39. Regression replacing basic density with skewness of meter density; Specification 1 103

Figure 40. Regression replacing basic density with skewness of meter density; Specification 2 104

Figure 41. Regression replacing basic density with kurtosis of meter density; Specification 1 104

Figure 42. Regression replacing basic density with kurtosis of meter density; Specification 2 105

Figure 43. Scatter plots of the density and total cost residuals; Specification 1 108

Figure 44. Scatter plots of the density and total cost residuals; Specification 2 109

Figure 45. Regression excluding SSEH; Specification 1 111

Figure 46. Regression excluding LPN; Specification 1 112

Figure 47. Regression excluding SSEH and LPN; Specification 1 112

Figure 48. Regression excluding SSEH; Specification 2 113

Figure 49. Regression excluding LPN; Specification 2 113

Figure 50. Regression excluding SSEH and LPN; Specification 2 114

Table 1. Example of efficiency score calculation. 21

Table 2. Capital price series considered 25

Table 3. Labour input price series and cost indices considered during Phase 2 30

Table 4 Input price series for categories capital, labour and the general inflation 37

Table 5. Specification 1 – Regional wages; Random Effects and Pooled OLS 39

Table 6. Efficiency score and rankings of Specification 1 40

Table 7. Specification 2 – National wages; Random Effects 42

Table 8. Specification 2 – National wages; Pooled OLS 43

Table 9. Efficiency score and rankings of Specification 2; Random Effects 44

Table 10. Efficiency score and rankings of Specification 2; Pooled OLS 45

Table 11. Comparison of Specification 1 (SIC_35, regional) and Specification 2 (SIC_35, national); Random Effects 47

Table 12. Measures used in the density analysis 55

Table 13. Meter density measures (2010) 57

Table 14. Percentage of DNO surface area with meter density in each bracket 60

Table 15. Regression results using mean density vs. basic density; Random Effects 62

Table 16. Regression results using standard deviation of density; Random Effects 63

Table 17. Regression results using skewness of density; Random Effects 65

Table 18. Regression results using kurtosis of density; Random Effects 66

Table 19. Regression results excluding density outliers, regional wage specification; Random Effects 69

Table 20. Regression results excluding density outliers, national wage specification; Random Effects 70

Table 21. Efficiency scores and rankings for the regional wage specification 72

Table 22. Efficiency scores and rankings for the national wage specification 73

Table 23. Scoring of alternatives to account for quality of supply performance using the statistic associated with the Wald goodness of fit test 82

Table 24. Comparison of Specification 1 (SIC35, regional) and Specification 2 (SIC-35, national); Random Effects 85

Table 25. Efficiency estimates for Specification 1 under RE and POLS 87

Table 26. Efficiency estimates for Specification 2 under RE and POLS 88

Table 27. Summary of density data following cleaning 91

1 Introduction

Ofgem has traditionally used a variety of benchmarking techniques during price control reviews and has signalled its intent to continue to do so. Ofgem has also indicated that it wishes to consider the role that total cost (totex) benchmarking might play at future price controls, including the forthcoming RIIO-ED1 review.

In the summer of 2012 Frontier Economics was commissioned by a group of DNOs, led by UKPN, to undertake an assignment to demonstrate the feasibility of totex benchmarking for the electricity distribution companies regulated by Ofgem. Since the conclusion of that first study, Frontier has worked with Ofgem and the DNOs to take forward our work on totex benchmarking.

Given the leading role that Ofgem has played in supporting the work, and the potential importance of totex benchmarking in the RIIO-ED1 review, Ofgem has now formally taken control of the Frontier totex study, with input from the DNOs through regular meetings of the Cost Assessment Working Group (CAWG). Ofgem has commissioned a second phase of work to address a range of issues left outstanding by the initial work.

This report provides a comprehensive summary of our findings and sets out our recommendations in respect of totex benchmarking for RIIO-ED1.

Throughout our work, during both Phase 1 and Phase 2, the commissioning party has stressed to us the need for our research to be conducted independently. Our Phase 2 report should therefore be understood to represent the views of Frontier Economics[1]. It does not necessarily reflect the views of Ofgem or any of the DNOs.

1.1 What is total cost (totex) benchmarking?

The phrase totex benchmarking is not associated with a single, well specified approach to benchmarking. Instead it can be understood to cover a potentially very wide range of methodologies. The common feature of all of these methodologies is that they seek to include all relevant costs in a single, summary regression, but otherwise there are many approaches that could be explored.

By its nature, totex benchmarking is high level since it does not seek to delve into the detail of the cost structure of any business, and therefore avoids the need to define the cost boundaries within the cost base, which might otherwise give rise to measurement error or create perverse reporting and operating incentives.

1 We are grateful for the advice and guidance of our academic advisors, Professor Tom Weyman-Jones and Professor Ron Smith.

1.2 Objectives of the study

Our original study, commissioned by UKPN and its project partners, was intended to assess the extent to which totex benchmarking could be implemented with sufficient robustness to be informative in the context of the 14 Electricity DNOs and the forthcoming RIIO-ED1 review. It is our view that total cost benchmarking can always be done in principle, but the successful application of a totex technique is only possible if there is sufficient, and sufficiently reliable, data available. Our Phase 1 work was therefore focused on practical application and the identification of an explicit technique.

1.3 Phase 1 findings

Following the conclusion of our Phase 1 work Frontier concluded that we had been successful in identifying a model specification that, based on the available data, appears to describe reasonably well the relationship between totex and key cost drivers such as outputs, input prices and environmental variables. Our proposed totex regression model explained totex (adjusted for quality of supply) as a function of:

- customer numbers;
- peak demand;
- average connection density;
- labour prices;
- general inflation; and
- a time trend.

We therefore concluded that developing a benchmarking methodology that considers totex was feasible. The model specification we presented relied solely on data that DNOs already collect, integrated with data on prices which are publicly and readily available.

However, Phase 1 also identified a range of areas where Frontier considered that further research might be needed. Our report outlined a number of areas where it was possible that some modifications might be made to the model to take account of factors that were otherwise not accounted for and this set of next steps became the agenda for our Phase 2 work.

1.4 Phase 2 objectives

The objective of Phase 2 has been to investigate as fully and thoroughly as possible the next steps identified at the conclusion of Phase 1. The areas where we considered that further investigation might be needed were:

- the potential impact of investment cycle on totex;
- the potential impact of asset condition on totex;
- whether, and if so how, to take account of assets in our modelling (e.g. network length);
- further and richer investigation of the potential impact of connection density on totex;
- assessment of alternative methods of accounting for quality in our benchmarking;
- further investigation of capital price series;
- further investigation of alternative labour cost data;
- assessment of whether and how to take account of differences in voltage structure.

We also note that our Phase 2 work has benefited from the availability of a further year of data, with our sample now covering the six years ending 2011/12.

We have sought to identify in which of these areas it might be necessary to modify our proposed Phase 1 model.

1.4.1 Excel Model

During Phase 1, in addition to advice on totex benchmarking as summarised in our report, Frontier also developed an Excel model to facilitate totex benchmarking. This model allowed Ofgem and the DNOs to replicate our results and also to test alternative totex benchmarking approaches. The model included all the data assessed in the course of our Phase 1 work, whether that data played a role in our model or otherwise.

At the conclusion of Phase 2 we will update our model to take account of certain requests from Ofgem and the DNOs. This will provide all stakeholders with the capacity to review directly our work and also to consider any alternative modelling approaches that they consider have merit.

1.5 Structure of this report

Due to the scope and breadth of our research during Phase 2 and the resulting extent of the material that we wish to report, we have split our final report into two volumes. In Volume 1 we provide a comprehensive review of the factors that have ultimately been included in our recommended model. In Volume 2, we describe the research we have undertaken in a number of areas that ultimately has not led to a change in the specification of our recommended model.

The remainder of Volume 1 of our report is comprised of the following sections.

- Section 2 provides an outline of the **methodology** we have adopted for this study. We review the choices made with respect to costs, cost drivers, sample and technique. For the avoidance of doubt, we have continued to adopt the approach that was adopted during Phase 1 of our report.
- Section 3 provides a discussion of our investigation into whether and how we might capture **input prices** in our benchmarking model.
- Section 4 covers our work investigating a wider and richer set of measures of **density**.
- Section 5 summarises our findings in respect of how **quality of service** might be accounted for in our model.
- Finally, in section 6 we bring together our analysis and present the conclusions that we draw as a result of our Phase 2 investigation.

In a series of annexes we also provide the following.

- The data used in the analysis of connections density
- The meter density histograms
- The results of testing alternative density measures
- An analysis of outliers in the connections density exercise
- The regression results excluding outliers in the connections density exercise

2 Methodology

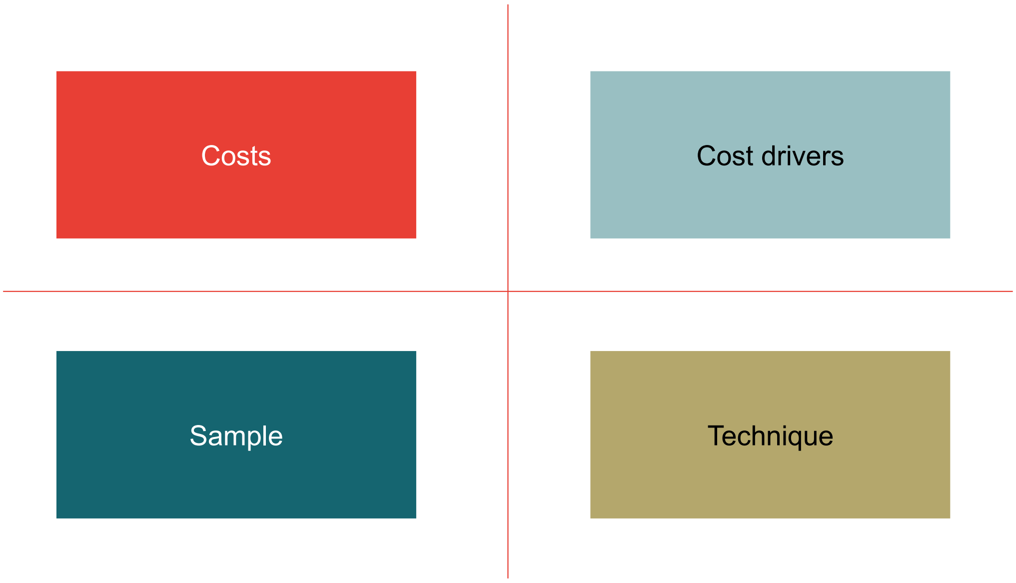
In this section we provide a description of the methodology we have adopted to investigate the feasibility of total cost benchmarking. We follow the structure we developed for our report on future benchmarking for Ofgem, which requires us to assess our choices with respect to cost, cost drivers, sample and technique.

2.1 Overview

In our report for Ofgem we set out a framework for specifying the elements that together comprise any benchmarking analysis. **Figure 1** lists the 4 components of a regulatory benchmarking exercise:

- **Costs:** One of the key challenges associated with this work is to identify how the capital-related part of the DNOs cost base could be incorporated into a totex measure. We considered this question carefully during Phase 1. We also set out any excluded costs (e.g. non-controllable costs, pension deficit funding etc.).
- **Cost Drivers:** We provide a discussion of the types of cost driver that we consider appropriate for a totex benchmarking study, including a review of the cost drivers that have been used in previous benchmarking studies in other places. In particular we discuss the philosophy that underpins our modelling approach and how this philosophy supports cost driver selection.
- **Sample:** We discuss briefly the sample that has been used to support our Phase 2 work.
- **Technique:** The predominant benchmarking technique used by Ofgem in previous regulatory reviews is the OLS/COLS approach. We have also made use of parametric statistical methods (rather than non-parametric methods such as DEA) in this assignment. We provide a discussion of the technique used, including a discussion of the merits of using a Random Effects (RE) approach.

Figure 1. The four main components of a regulatory benchmark



In the following subsections we review the key issues that arise in each of these four areas.

2.2 Costs

Accurate, reliable and comparable cost data is necessary for any robust benchmarking. In this section we discuss the cost data that we have used.

2.2.1 Coverage

The cost data we have used was originally drawn from the regulatory reporting processes undertaken by the companies. Our main sources have been the data that Ofgem has placed in the public domain through the DPCR5 Financial Model[2] and the 2010-11 Cost Reporting document.[3] We have also been provided with access to data for 2011-12 now that it has been collected by Ofgem. We anticipate that in due course this data will be made publically available through a future Cost Reporting document. As a consequence, we have cost data available for the years 2006-07 to 2011-12.

2 Ofgem, Electricity Distribution Price Control Review Final Proposals - Allowed revenue financial issues – Financial Model, 07/12/2009.

3 Ofgem, Electricity Distribution Annual Report for 2010-11, 30/03/2012.

During our work on Phase 1, the DNOs undertook a thorough audit of the cost data we were using. This audit identified a number of discrepancies in the older data that had been provided to populate Ofgem's DPCR5 financial model. The DNOs have therefore updated this data to ensure a greater degree of comparability and all of our Phase 2 work has been conducted on this revised data.

The measure of totex we have used in this study, as agreed at the outset of Phase 2 with the Ofgem and the DNOs, is comprised of:

- network investment;
- network operating costs;
- closely associated indirects;
- business support costs;
- non-operational capex;
- ES4 RAV adjustment;
- pensions[4];
- logged up costs; and
- connections costs, where included within the scope of the price control.

For the avoidance of doubt, the following costs are excluded:

- atypicals;
- non-price control activities;
- non-activity based costs;
- standalone funding through, for example, IFI and LCNF mechanisms; and
- sole use connections for 2010/11 and 2011/12.

We are grateful to the DNOs for their efforts in ensuring the highest degree of comparability in our input cost data.

In addition to these costs, we have also considered including the impact of customer interruptions directly into our totex measure. A full description of our approach and the results can be found in Section 5.

4 Excluding established deficit funding costs.

2.2.2 Incorporating capex

During Phase 1 we considered carefully how capex might be incorporated in a totex benchmarking study. We discussed and investigated two high-level approaches.

- **Use capital expenditure:** Simply using actual capex in a certain period (e.g. annually or possibly averaged over a number of years) has the advantages of simplicity and also that the flow of expenditure is under the direct control of the current management. However, since capex can be lumpy, companies could look efficient or inefficient depending on factors related to specific network needs, e.g. as a consequence of where they are in their investment cycle. There will be a need to control for this effect when using an expenditure flow measure of capital, as we discuss later when evaluating alternative modelling techniques.

- **Use a measure of capital consumption:** This means constructing a measure of capital consumption comprised of returns on and depreciation of some defined asset base. The strengths and weaknesses of this approach mirror the ones of direct capital expenditure. In principle the investment cycle is better considered through the treatment of historic capex flows in the asset base, but current management has limited control over a measure of capital consumption in the short run as much of it relates to historic expenditure that is now sunk. Given the need for a large volume of historic data, there are also limits to the extent to which measures of capital consumption can be reconstructed. Additionally, there is less variation in measures of this kind over time, which can give rise to challenges in estimation.

Following Phase 1, we discussed these two approaches in detail with the DNOs and with Ofgem. It was agreed that limitations on the availability of historic data on a sufficiently consistent basis would greatly reduce the viability of the capital consumption approach. Consequently, our Phase 2 analysis has focused solely on totex measures based on capital flows.

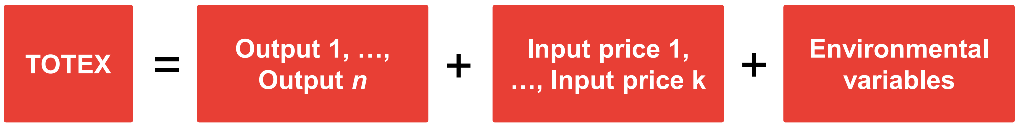
As highlighted above, this gives rise to a concern over the underlying volatility of the data and the extent to which it is therefore possible to estimate efficiency reliably. We return to this question when we discuss modelling techniques in Section 2.5 below.

2.3 Cost drivers

The selection of cost drivers is central to any benchmarking study. By changing the set of cost drivers that are used to "explain" cost, the underlying nature of the exercise can be changed fundamentally. It is therefore necessary for cost driver choice to be guided by a clear underpinning economic rationale in addition to the resulting econometric properties of the model. Similarly, the benchmarking model should fit naturally with the wider regulatory framework and provide consistent incentive properties. Where proxy variables are included that are inputs that are within the control of the operator (e.g. network length as a possible proxy for geographic factors that could drive cost), the potential perverse incentives that may be created should be considered carefully. Finally, it is also necessary to be aware of the practical constraints that exist over the number of cost drivers that can be included, given the available sample size and multicollinearity between drivers.

All of our analysis during Phase 2 has been focused on defining a totex function for DNOs. Economic theory suggests that such function should take the form shown in **Figure 2** below.

Figure 2. Generic cost function



Therefore, there are three groups of variables that contribute to explain a DNO's total cost:

- **Outputs:** this group contains the firm's main outputs, such as for example electricity distributed and peak demand served in each year.
- **Input prices:** this group contains the prices of the inputs used by the firm, such as labour and capital, in order to capture changes in totex that arise as a result of changes in the prices of inputs.
- **Environmental variables:** these are the variables that describe the operating environment of the firm; these variables are outside the firm's control (e.g. service area) but may affect its observed costs.

We discuss each category in turn.

2.3.1 Core outputs

The set of core outputs that might be included are the amount of electricity distributed, the number of customers served and the peak load. In practice, it may not be possible to include all these variables in the final model specification because:

- output variables tend to be highly correlated with each other, which raises issues of multicollinearity; and

  - even considering all six years, the sample size is relatively small: this limits the number of explanatory variables that can be used.

We consider this set of variables to be the ideal candidates to include in a study of this kind since they are outputs valued by customers and which capture well the scale of the supply task and, to a large degree, determine the size, number and layout of the assets required to serve.

Over the course of the study it has been suggested that we should also consider and include other variables, such as network length, as additional core outputs. Our in principle view on this is that it would be inappropriate for a number of reasons.

- Asset related variables are **not core outputs** and are **not directly valued by customers**.

- Similarly, such variables are **inputs** and their inclusion in a benchmarking model may create **perverse incentives**, in addition to giving rise to technical estimation problems arising from endogeneity.

- Finally, including network length in the benchmarking model **eliminates the ability to test for optimal network design**. Any line installed on the network will be regarded in such a model as a "good", whereas it is entirely possible that a more efficient company might be able to serve the same set of customer outputs with a smaller network. A model that includes network length will "explain away" potentially excess network length and fail to reward companies that excel in network planning design.

Notwithstanding these concerns, we have analysed whether asset related outputs should play a role in our final model. A full discussion of our investigation can be found in Section 4 of Volume 2.

2.3.2 Input prices

The role of input prices is clear. Where prices change either over time or between regions, it is reasonable to anticipate that this will lead to changes in totex. In order to ensure a robust estimation (and specifically to avoid the risk of a missing variable bias) it is necessary to capture these effects through the inclusion of appropriate input prices in the model. We provide an exhaustive review of our investigation into input prices in Section 3.

The coefficient on each price can be interpreted as the estimated budget share of the input in question in the cost base. Consequently, it must follow that the sum total of all budget shares of all included inputs must equal 1. This condition ensures the resulting cost function has the necessary properties to be considered well specified, specifically that it should be homogeneous of degree 1. Loosely speaking, this ensures that if all input prices double, then totex should also double.

This condition is typically imposed by restricting the form of the cost function. Consider an example in which the totex cost function includes one output, and two input prices, labour and capital:

$$\ln(totex) = \beta_0 + \beta_1 \ln(output) + \beta_2 \ln(wages) + \beta_3 \ln(capital\ prices) + \varepsilon \qquad (1)$$

Homogeneity of degree one effectively restricts the parameters β_2 and β_3 to sum to one. If we impose such restriction, we can rewrite the equation (1) as

$$\ln(totex) = \beta_0 + \beta_1 \ln(output) + \beta_2 \ln(wages) + (1 - \beta_2) \ln(capital\ prices) + \varepsilon \qquad (2)$$

which can be also expressed as

$$\ln\left(\frac{totex}{capital\ prices}\right) = \beta_0 + \beta_1 \ln(output) + \beta_2 \ln\left(\frac{wages}{capital\ prices}\right) + \varepsilon \qquad (3)$$

Equation (3) can be understood to be equivalent to equation (1) under the restriction of homogeneity of degree 1 in input prices. We have also tested the validity of this assumption directly (by estimating equation (1) and testing the restriction on the relevant parameters), and found no evidence to reject the restriction on the model.

2.3.3 Controlling for operating environment and other factors

It is necessary to ensure that the operating environment of each DNO, in so far as it might impact on its costs, is captured in the model. Otherwise any estimates of efficiency may be biased. For example, there is clear evidence to suggest that a DNO's costs to serve will be driven by the population density of the service region. We provide a thorough investigation of density in Section 4 below.

We have also considered a range of other variables that could capture some important aspect of each DNO's operations that could justify a difference in costs for reasons other than differences in efficiency. A full discussion of our conclusion in each of these areas can be found in Volume 2 of our report.

2.4 Sample

In agreement with Ofgem and the DNOs, in this study we have regarded each licensee as a separate entity, i.e. we have benchmarked the 14 GB DNOs, rather than the ownership groups. This is consistent with Ofgem's approach at previous price controls. It also ensures a larger sample size, containing greater heterogeneity, allowing more ambitious/robust econometric analysis.

We have not included companies from other jurisdictions given the difficulties this would create in terms of data collection/standardisation.

As noted above, totex data is available for 6 years, providing us with a panel of 84 observations. For certain candidate cost drivers data is only available for a single year. Our approach in these cases is detailed in the relevant chapter.

Since such data is not yet available, we have not included within our study any forecast totex data.

2.5 Technique

Following regulatory precedent in GB, and after discussion with Ofgem and the DNOs, we have restricted our attention to statistical methods and have not investigated non-parametric techniques such as Data Envelopment Analysis (DEA). However, within this general approach there exist a relatively wide variety of models that could be used to estimate a totex function and a set of efficiency scores.

2.5.1 Assumed frontier model

Several methods are available to estimate our proposed cost function. These include:

- OLS on a cross-section of the mean value for each variable for the 14 DNOs, this is also known as the "between" regression;
- Pooled OLS (POLS) on the panel 84 observations;
- Random Effects (RE) on the panel; and
- Fixed Effects (FE) on the panel. This is equivalent to OLS on the deviations from the mean for each variable and is also known as the "within" estimator.

Ofgem has more recently tended to adopt a POLS approach, e.g. at its recent RIIO-GD1 review, although given the "shifting" of the regression line to some frontier point (e.g. the upper quartile level of performance) perhaps it might be better understood to be an application of a Pooled Corrected OLS (P-COLS) approach. This approach does not take direct account of the panel structure of the sample (i.e. the fact that all observations are not independent, but instead we have repeated observations on the same entity over time) explicitly in the model specification. However, estimation can be made robust to this by taking account of clustering in the data when calculating so-called robust standard errors[5]. Ofgem has then estimated efficiency for each company through the use of adjusted residuals for a given year.

Random and Fixed Effects[6] models recognise that there are potentially systematic differences between the 14 DNOs in the sample, even after controlling for all the cost drivers in the regression and estimate these systematic differences, separating them from the idiosyncratic error using a two-component error form. The OLS and POLS approaches do not.

In this benchmarking exercise, we therefore advocate the use of the Random Effects approach, for the following reasons.

- Despite being widely used in regulatory benchmarking, the statistical properties of the POLS estimator rely on the assumption that there are no systematic differences amongst the various DNOs in the sample. This is in contrast with the main objective of benchmarking itself, which tries to identify such systematic differences between DNOs.

- A possible treatment to control for the existence of systematic differences between the various DNOs in the sample would be to use robust standard errors in the pooled OLS regression. This treatment could empirically work to restore the statistical properties of the OLS estimator. However, as the structure of the data in the sample is known to be a panel of 14 different DNOs it seems more appropriate to use an estimation method that explicitly takes this structure into account, a panel data method.

- The Fixed Effects estimator is inappropriate in our case. First, our set of cost drivers includes variables that change very slowly (or are essentially fixed over time), e.g. density, number of customers, and these are estimated with very poor precision using Fixed Effects models. After performing the Hausman test over the various specifications tried so far, we have concluded that the Random Effects (RE) estimator is consistent and consequently, more efficient that the FE estimator.[7]

---

5 It is worth noting that the choice of whether to use robust standard errors or not will have no impact on the parameter estimates and hence on estimated efficiency scores. It could, however, have an impact on the assessment of the significance of parameter estimates and the regression model as a whole. In estimating our preferred Random Effects specification we have checked to ensure that all our models remain significant irrespective of ones choice of standard error estimation.

6 See for example Greene, W (2005) 'Fixed and random effects in stochastic frontier models', Journal of Productivity Analysis.

7 We have performed a Hausman test, both in its standard version and a robust version of it in case there was any intracluster correlation left in the Random Effects model. Both tests deliver the same answer that Random Effects is appropriate in this case.

- Panel data methods determine an estimate of each DNO's inefficiency averaged over periods, based on the assumption that efficiency is fixed over time. This is helpful in the current context where our totex variable might be "lumpy" as a consequence of the inherent lumpiness of capex programs. The RE estimator provides a method through which any noise in the data can be isolated through the decomposition of the error term. This feature is not available in pooled OLS, where the averaging of each DNO's inefficiency depends on *ad hoc* assumptions, for example the inefficiency for each DNO might be calculated as the mean of the residuals for this DNO across the periods in the sample.

2.5.2 Underlying production function

We have chosen a Cobb-Douglas cost function. This functional form has been widely used in applied cost benchmarking studies as it is simple to understand and analyse. We use a log linear specification, taking logarithms of the dependent variable and the cost drivers except for the time trends and any dummy variable. The translog specification is a popular alternative functional form. A preliminary assessment of its suitability in the context of the 14 GB DNOs did not produce econometrically robust results (e.g. we found many square and interaction terms that were not statistically significant). Consequently, we have not considered the translog functional form further. The dataset is a panel comprising observations for all GB DNOs (N=14) over six years (T=6), resulting in a sample of 84 observations.

2.5.3 Derivation of efficiency scores from a Random Effects model

The "efficiency score" for each DNO is calculated as the relative distance of that DNO with respect to the most efficient DNO. In the Random Effects regression the "efficiency score" is based on the estimated systematic, time invariant component of the error term for each DNO, i.e. the lower the time invariant component of the error term the higher the estimated efficiency for a DNO. In Pooled OLS the "efficiency score" is based on the full residual.

Residuals can have positive and negative values. For this reason we have re-scaled the original residuals by assigning a value of 0 to the DNO with the lowest original residual (the most efficient DNO in the sample). For the other DNOs, their re-scaled residuals will be equal to their original residuals minus the lowest original residual. These re-scaled residuals are always positive numbers. The efficiency scores for each DNO are then calculated taking the exponent of the negative of the previous re-scaled numbers. With this approach, the most efficient DNO (the one with the lowest residual) has a 100% efficiency score, while the other DNOs have an efficiency score below 100%.

Our approach is entirely consistent with the relevant literature on efficiency benchmarking.

Table 1 shows an example of how efficiency scores have been calculated both under Random Effects and Pooled OLS.

Table 1. Example of efficiency score calculation.

DNOs	Original residual (time-invariant under Random Effects)	Re-scaled values	Efficiency score
	A	B	C
1	-0.050	0.054	95%
2	-0.020	0.084	92%
3	-0.019	0.085	92%
4	0.038	0.142	87%
5	0.065	0.169	84%
6	0.123	0.227	80%
7	0.080	0.184	83%
8	-0.047	0.057	94%
9	0.060	0.164	85%
10	-0.104	0.000	100%
11	0.057	0.161	85%
12	-0.059	0.045	96%
13	-0.059	0.045	96%
14	-0.066	0.038	96%
		A – min(A)	exp (-B)

3 Further investigation into input prices

As described in our methodology section, the prices of relevant inputs are a key determinant of costs and a total cost function estimated without taking account of input prices is likely to be poorly specified.

In this section we describe the work we have undertaken to investigate which input prices should be included in our preferred model specification and which input price series deliver the best results.

3.1 Phase 1

Our Phase 1 candidate econometric model included a labour input price and the UK GDP deflator. We used the GDP deflator as a proxy for the price of general inputs used by DNOs that broadly track economy-wide prices[8]. The price of capital goods purchased by DNOs was not included as we were unable to find a suitable series in the time available. Regarding the labour price, we used both a national wage index and a regional wage index and found that using regional averages of the representative wage in the utilities sector delivered slightly better results.

3.2 Phase 2 approach

In Phase 2 we have further investigated which input prices should be used as cost drivers in the estimation of the totex cost function. These input prices should relate to the categories of inputs that are most relevant for electricity DNOs.

Our starting point in Phase 2 has been to consider a totex cost function using three categories of inputs: labour, capital and a general category of costs that move in line with economy-wide prices. In this section we:

- describe the capital price data series we have investigated;
- describe the labour price data series we have investigated;
- describe the economy-wide price data series we have investigated; and
- present the results of combining different input prices in our base model and identify which combination of input prices results in a well specified model, both in terms of economic interpretation and statistical properties.

---

8 As explained in more detail in Section 2, we deflated costs and the labour price series by the GDP deflator in order to impose homogeneity on the estimated cost function.

3.3 Capital price series

Because the totex measure that we utilise to benchmark electricity distributors is based on capital expenditures, the total cost function should control for the price changes of materials bought by DNOs over the sample period.

To this end we have considered a range of data series from different sources[9], and in **Table 2** we list those that we believe could be closely related to the prices faced by DNOs. Some of these indices have been previously considered by Ofgem[10].

9 Office of National Statistics, British Electrotechnical and Allied Manufacturers Association (BEAMA), Eurostat, Department for Business, Innovation & Skills (BIS), The National Institute of Economic and Social Research (NIESR) and Department for Communities and Local Government.

10 RIIO – T1: Initial proposals – Real price effects and on-going efficiency, 27 July 2012, Ofgem

Table 2. Capital price series considered

Index Name	Index acronym	Description	Considered by Ofgem?	Industry, activity or product covered	Geographical coverage	Source
Price index of Materials used in the Basic Electrical Equipment Industry	**BEAMA**	Producer price index (PPI) which measures the evolution of input prices faced by manufacturers of electrical equipment, division 27 in UK Standard Industrial Classification 2007. It is measured monthly.	Yes	Division 27 in SIC-2007	United Kingdom	BEAMA
PPI – Electricity Production and Distribution	**PPI-pro&dis**	PPI in input prices for the electricity production and distribution industries. It is a base weighted index of the materials and fuel purchased. It is measured monthly.	No	Electricity production and distribution (I/O group 88)	United Kingdom	ONS (PPI:7167880000)
PPI - Electric motors, generators and transformers, EU Imports	**PPI-MGT(EU)**	PPI in output prices that EU manufacturers charge for electric motors, generators and transformers destined for the UK. It is measured monthly.	No	Electric motors, generators and transformers	Imports from the EU	ONS (PPI:8271100100)
PPI - Electric motors, generators and transformers, UK	**PPI-MGT(UK)**	PPI in output prices that UK manufacturers charge for electric motors, generators and transformers. It is measured monthly.	No	Product: electric motors, generators and transformers	United Kingdom	ONS (PPI:2711000000)
PPI - Electricity distribution and control apparatus, EU Imports	**PPI-appa(EU)**	PPI in output prices that EU manufacturers charge for electricity distribution and control apparatus. It is measured monthly.	Yes	Electricity distribution and control apparatus	Imports from the EU	ONS (PPI:8271200100)
PPI - Electricity distribution and control apparatus, UK	**PPI-appa(UK)**	PPI in output prices that UK manufacturers charge for electricity distribution and control apparatus. It is measured monthly.	Yes	Electricity distribution and control apparatus	United Kingdom	ONS (PPI:2712000000).

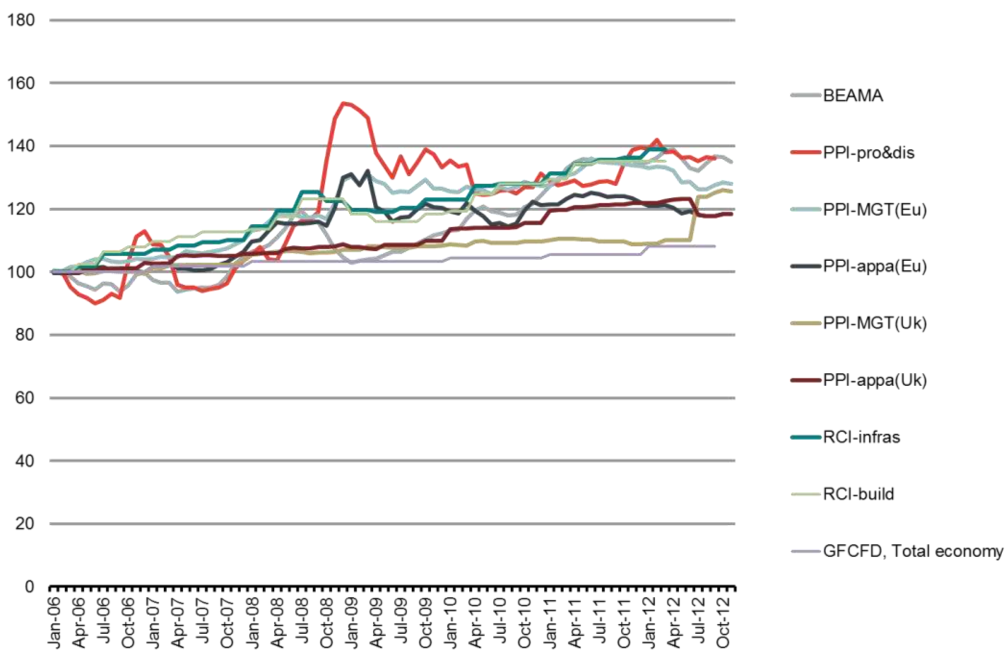
Resource Cost index of Infrastructure	**RCI-infras**	The notional trend of labour, materials and plant costs faced by a contractor derived by applying the price adjustment formulae for civil engineering works to a cost model for an infrastructure project. A disaggregated version of this series captures the effect of materials only and excludes plant and labour price trends is also available. It is measured quarterly.	Yes	Infrastructure project	United Kingdom	BIS
Resource Cost Index of Non-Housing Building	**RCI-build**	Same concept as above. However, the price adjustment formula is applied to a cost model for a non-housing building project. A disaggregated version of this series captures the effect of materials only and excludes plant and labour price trends is also available. It is measured quarterly.	Yes	Non – housing building	United Kingdom	BIS
Gross Fixed Capital Formation Deflator	**GFCFD**	The change in the value of a basket of fixed assets[11], We have considered two measures of the GFCFD. One for the whole economy and one for the non-residential construction and civil engineering activities.	No	UK Economy and non-residential construction and civil engineering activities (code AN112, as classified by ESA-95)	United Kingdom	AMECO

Source: Frontier Economics

11 Fixed assets are defined by ESA as tangible or intangible assets produced as outputs from processes of production that are themselves used repeatedly, or continuously, in processes of production for more than one year.

Data for each of the indices in the table are only available on a national basis. We do not consider that this gives rise to any concern as it seems reasonable to assume that capital goods are purchased in markets where there is limited, if any, regional price variation. Hence, we expect the most important components of a DNO's asset base (such as transformers or conductors), to show the same price trend regardless of where they are bought within the domestic market. **Figure 3** shows the evolution of these price indices for the period 2006-2012.

Figure 3. Indices considered as potential proxies for capital input prices



Source: Frontier. *Note that RCI series are available quarterly, PPIs are available monthly and GFCFD are available annually.*

All indices show similar ascending patterns, with PPI – Electricity Production and Distribution being the more volatile, probably because the index is highly affected by fuel prices paid by electricity generators. We can also observe that the two RCI indices follow a similar ascending pattern. More variation is observed among the remaining five PPI indices. **Figure 4** below shows the correlations among the price series.

Figure 4. Correlations between capital series

	BEAMA	PPI-pro&dis	PPI-MGT(Eu)	PPI-appa(Eu)	PPI-MGT(Uk)	PPI-appa(Uk)	RCI-infras	RCI-build	GFCFD, Total economy
BEAMA	1								
PPI-pro&dis	0.5979	1							
PPI-MGT(Eu)	0.8327	0.8769	1						
PPI-appa(Eu)	0.736	0.9087	0.9566	1					
PPI-MGT(Uk)	0.8412	0.7964	0.9592	0.9099	1				
PPI-appa(Uk)	0.949	0.6588	0.8843	0.7657	0.8795	1			
RCI-infras	0.9409	0.7551	0.9351	0.8553	0.9328	0.9558	1		
RCI-build	0.933	0.692	0.8869	0.7989	0.8842	0.9602	0.9823	1	
GFCFD, Total economy	0.9003	0.7091	0.8889	0.7985	0.909	0.9411	0.9548	0.9354	1

Source: Frontier. *Note that RCI series are available quarterly, PPIs are available monthly and GFCFDs are available annually. Therefore, any two correlation values are only directly comparable if the two pairs have been measured under the same periodicity.*

We note that, in general, there is a high level of correlation between certain price series, which suggests they are capturing similar price pressures. This implies that we might expect to find a number of series for which the statistical fit with the available data is similar.

Among all the series listed in **Table 2** we do not have any *a priori* reason to discard them as candidate series to include in the model. The only exception is the series PPI – Electricity Production and Distribution which includes fuel prices paid by electricity generators, and given that fuel costs are very volatile and not a major share of DNOs' input prices we have rejected it as a potential candidate.

Our approach with the other series is to include them one at a time as well as in pairs in the cost function. We did not try combinations of three or more capital indices, due to the high correlations among them as well as to our limited sample size.

On the basis of this testing process, evaluation of the coverage of each index and its previous use by Ofgem in regulatory proceedings, we have concluded that the BEAMA price index of materials used in the basic electrical equipment industry is the most appropriate to include in the cost function.

3.4 Labour price series

In Phase 1 we concluded that including a labour input price in the set of cost drivers was required for the estimated cost function to be well specified. We also found that a price series based on regional averages of wages from the sector "electricity, gas, steam and air conditioning supply" (using the UK Standard Industrial Classification - SIC) delivered somewhat better results than a price series based on the national average wage for all employees in the UK.[12]

This section presents the candidates we have further considered as the labour input price. Most of the data has been sourced from the ASHE database provided by the ONS [13], and an additional labour cost series has been sourced from BEAMA.

A criterion we have followed when selecting candidates for the labour input price is that the underlying wages and labour costs should be related to the electricity distribution sector. Some of the series we have identified as candidates have been previously considered by Ofgem.

There are two specific issues that we have explored that are worth emphasizing:

- Whether average wages based on professional occupations (using the UK Standard Occupational Classification – SOC) are better cost drivers in our cost model than wages based on industrial classifications (SIC). For this reason we have considered prices series based on average wages for occupational categories and series and based on average wages for sectors in the UK. The ASHE database provides average wages for both occupational categories and sectors.

- Whether the labour input price is better represented by a national price or regional price series. To the extent it is possible we have matched regional wages to the geographical areas where DNOs operate. The ASHE database provides both national and regional average wages (based on UK regions) for most of the wage series considered. Subject to availability, we have included price series based on national and on regional averages in our set of selected candidates for the labour input price.

---

12 This result was reported in the Appendix of the Phase 1 report

13 The ASHE database reports yearly average wages across different dimensions like sectors, occupational categories, UK regions, earnings definitions and subsets of employees. http://www.ons.gov.uk/ons/guide-method/surveys/list-of-surveys/survey.html?survey=%27Annual+Survey+of+Hours+and+Earnings+%28ASHE%29%27

Table 3 lists all the labour input price and cost series considered in Phase 2. The table distinguishes between prices series based on average wages from specific sectors or based on professional occupations. The table also reports whether the price series is available on a regional and national basis. The full set of labour input price candidates includes 17 series, 10 based on national wages and 7 based on regional wages.

Table 3. Labour input price series and cost indices considered during Phase 2

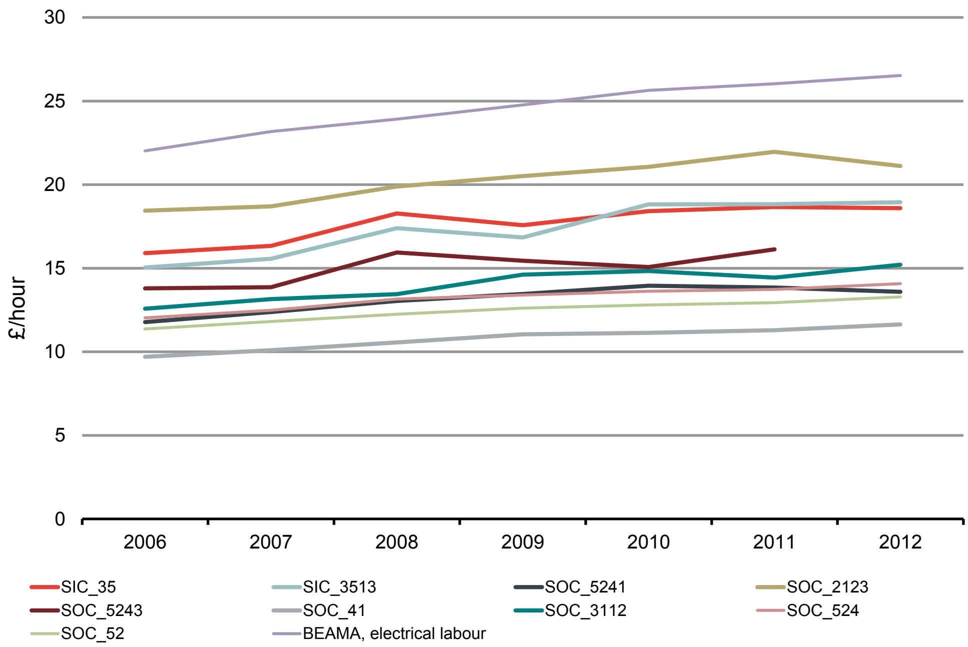
Name	Sector / Occupation	Definition	National / Regional	Other	Source
SIC_35[14]	Sector	Electricity, gas, steam and air conditioning supply	Both		ONS
SIC_3513	Sector	Distribution of electricity	National		ONS
SOC_2123	Occupation	Electrical engineers	Both		ONS
SOC_3112	Occupation	Electrical/ electronic technicians	Both		ONS
SOC_41	Occupation	Administrative occupations	Both		ONS
SOC_52	Occupation	Skilled metal and electrical trade	Both		ONS
SOC_524	Occupation	Electrical trade	Both		ONS
SOC_5241	Occupation	Electricians, electrical fitters.	Both		ONS
SOC_5243	Occupation	Lines repairers and cable jointers	National	2012 not availabl e	ONS
BEAMA_electrical -labour	Not specified	Labour cost index (Electrical) - CPA/4 Electrical Engineering	National		BEAMA

Source: Frontier Economics using information from ONS (ASHE) and BEAMA

Figure 5 shows the evolution of the labour input price series in **Table 3**. The figure only shows the price series based on national wages.

14 In the report, we use the SIC_35 to refer to the sector "Electricity, gas, steam and air conditioning supply" using the UK Standard Industrial Classification from 2007 (SIC2007). Wage data for this sector has been available in the ASHE database for the sample period considered 2006 to 2012. However, in the initial years of 2006 and 2007 the precise description of the series was "Electricity, gas, steam and hot water supply". We do not think this affects the possibility of building a wage series based on this code over the sample period

Figure 5. Evolution of labour input price series considered, using national averages



Source: Frontier Economics using data from ONS (ASHE) and BEAMA

Note that all SIC and SOC series have a common measurement unit, gross hourly pay, whereas the BEAMA series is an index. In order to be able to plot it in the same graph, we have re-scaled the BEAMA series dividing it by a factor of 4. Hence, only the trend between BEAMA and ASHE series can be compared, not the absolute values.

The wage series depicted in **Figure 5** above grew in the period 2006-2012 at an annual average rate between 2.3% and 4.1%. The actual trend followed by each series also varied substantially, with some occupations or industries showing nominal wage decreases in certain years. Nevertheless, as **Figure 6** below shows, the correlation between series is generally high.

Figure 6. Correlation between labour input price series using national averages

	SIC-35	SIC-3513	SOC -5241	SOC-2123	SOC-5243	SOC-41	SOC-3112	SOC-524	SOC-52	BEAMA, electrical labour
SIC-35	1									
SIC-3513	0.9624	1								
SOC -5241	0.9228	0.9467	1							
SOC-2123	0.9094	0.9443	0.952	1						
SOC-5243	0.9195	0.8032	0.7911	0.8331	1					
SOC-41	0.899	0.9113	0.9883	0.9655	0.817	1				
SOC-3112	0.7896	0.8353	0.9601	0.8874	0.6575	0.9625	1			
SOC-524	0.948	0.9419	0.9917	0.9585	0.86	0.9902	0.9339	1		
SOC-52	0.9172	0.933	0.9933	0.9693	0.8201	0.9974	0.9515	0.9945	1	
BEAMA, electrical labour	0.8988	0.945	0.9859	0.975	0.771	0.9839	0.9387	0.9776	0.9917	1

Source: Frontier Economics using data from ONS (ASHE) and BEAMA

In section 3.4.1 we discuss the construction of the data and in section 3.4.2 we highlight some initial considerations relating to the data.

3.4.1 Construction of the data

Choice of wage definition and average in ASHE

A very wide variety of measures are presented in the ASHE database. These include:

- both mean and median;
- hourly, weekly and annual pay;
- full time, part time and all employee splits;
- male, female and all employee; and
- use of gross pay (including all bonuses, overtime etc.).

Of these measures, we have chosen to make use of the mean of the hourly gross pay for all employees.

We believe that hourly data, in comparison with annual data, is better suited for our exercise since hourly pay will take account of variations in terms and conditions that govern the length of the working week and when/whether overtime is paid. Nevertheless, we also tested our model specifications using annual measures and results showed poorer statistical properties in comparison with hourly measures.

We have chosen to use mean wages in our modelling, although again we have tested both. Compared to the mean, the median is not always reported in the ASHE database for some wage series and regions in some years. We have nonetheless used both measures and the results obtained using the mean showed superior statistical properties.

Mapping the ASHE data to financial years and to DNOs

The ASHE data provides averages corresponding to calendar years. Data specific to DNOs, totex and cost drivers are given in regulatory years (April to March). In order to construct wage series that are matched to the specific DNO data, we have calculated a weighted average using the observations from two consecutive ASHE databases. That is, a wage figure for year 2006-2007 is calculated as ¾*2006 ASHE wage + ¼*2007ASHE wage.

As well as a temporal match a spatial match is also needed. While there is no direct correspondence between GB regions and the DNO's service areas, we have mapped regional wages based on geographical proximity. **Figure 7** below shows the mapping employed to this end.

Figure 7. Regional mapping of ASHE regional averages

DNO	Region
NPGN	North East
SPMW	50% North West + 50% Wales
ENWL	North West
NPGY	Yorkshire and The Humber
EMID	East Midlands
WMID	West Midlands
SWest	South West
SSES	South East
EPN	East
LPN	London
SPN	South East
SWales	Wales
SSEH	Scotland
SPD	Scotland

Source: Frontier Economics

We acknowledge that this mapping is not perfect, but believe that it is broadly reasonable and, in any event, is the only approach that can be taken to consider regional data in the absence of requesting that the ONS re-states its data using different regional boundaries. This mapping was applied to all regional wage series.

During the course of Phase 2, UKPN provided a recut of the ASHE database based on the geographical areas covered by each DNO. This data includes average wages for some occupational categories relevant in the electricity distribution sector. However, this data does not cover all the years in the sample and it is only based on SOC codes (wage averages for SIC codes are not available). We have therefore decided not to use this recut of the ASHE database.

3.4.2 Initial considerations of the data

Choice of SOC or SIC series

As we set out below, we have tested price series based on each SOC, separately and in combination with various SOC-based price series, and assess whether they fit the data well. However, given data limitations, it is not possible to use more than two labour prices in the set of input prices. This restricts the use of SOC codes as we would expect that workforce of DNOs is composed of several professional categories. Alternatively, we have considered constructing a blended labour price series from several SOC codes. We regard this as arbitrary because it requires knowledge on the composition of the DNOs' workforce. Even if this composition was known, a blended labour price using SOC codes would largely replicate the labour price based on the SIC code corresponding to the electricity distribution sector. Using wage data based on SIC codes rather than SOC codes will take account of the mix of labour actually used in the relevant industry.

We have identified two SIC codes that may be representative of the labour employed in electricity distribution sector. One covers the utilities sector to a wide extent, code SIC_35 "electricity, gas, steam and air conditioning supply", and the other focuses on electricity distribution only, code SIC_3513. While national and regional average wages are available for SIC_35 only national wages are reported for SIC_3513 in the ASHE database.

Figure 5 above shows that the labour input price series based on the SIC_35 and SIC_3513 codes are quite similar, both in terms of absolute values and the general trend over time.[15] Based on this we consider that a price series based on the SIC_35 might be a better alternative over the SIC_3513 code because it is possible to compare the national and the regional wage specification using the same underlying data source. Despite this consideration, we have kept the labour price series using the SIC_3513 code in the list of alternative labour prices.

Given these considerations, we would expect that using a labour price series based on SIC codes will result in a better specified model than using price series based on SOC codes.

Choice of national and regional wages

In our set of candidate labour input prices we have included series using national and regional wage averages.[16]

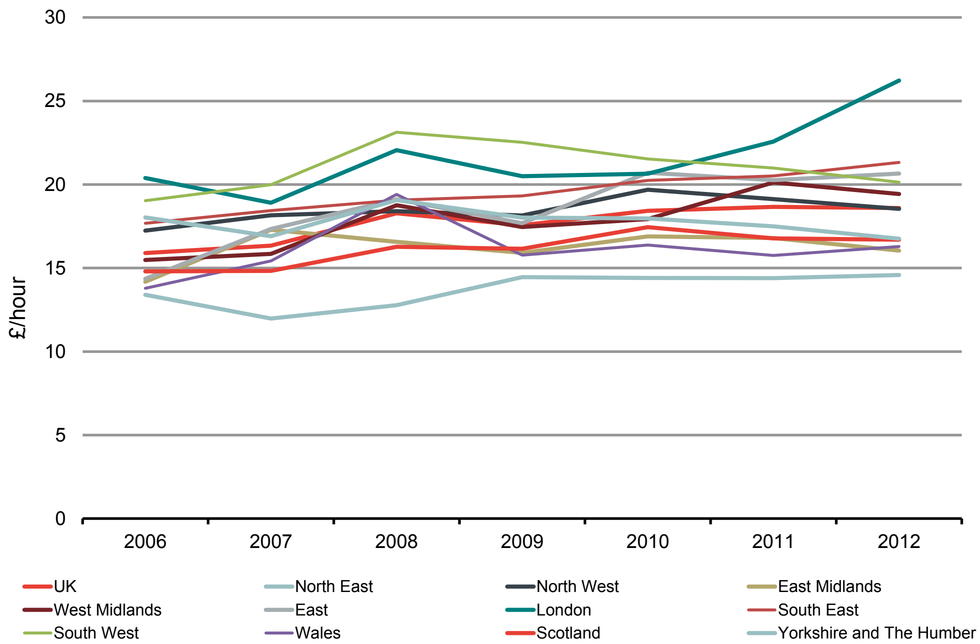
Using the SIC_35 series, **Figure 8** shows the evolution of labour input prices across regions. The figure reveals that there is considerable variation between the labour prices across UK regions. This variation prevails in all series constructed using regional wage averages.

---

15 For the SIC_35 series we have used the series based on national average wages.

16 The use of regional wages should not be understood to suggest that all labour is purchased regionally or that the DNO market for labour is regional. It merely reflects regional variation in wages that might be relevant in explaining the cost levels of the DNOs.

Figure 8. SIC-35, Regional wages



Source: Frontier Economics using ASHE data

3.5 General inflation indices

In the econometric model of Phase I, the dependent totex variable and the independent wage variable were expressed in real terms (2010-11 prices) using the UK GDP deflator. As described in the methodology section, this is equivalent to including the price of a general input in the regression and imposing homogeneity of degree one in prices.

In Phase 2 we have again considered using a measure of general inflation in the totex cost function. To do this we have selected the two measures of general inflation considered in Phase 1, the RPI index and the UK GDP deflator, and included them in the set of possible input prices.

3.6 Results

In this section we present the results of including combinations of input prices in our base model. The objective of this exercise is to identify which combinations of input prices, covering all or some of the previous input categories, work best when included in the model. **Table 4** shows the three lists of alternative input prices for capital, labour and the general inflation.

Table 4 Input price series for categories capital, labour and the general inflation

Capital	Labour	General Inflation
BEAMA	SIC_35 (regional)	UK RPI
PPI-MGT(Eu)	SIC_35 (national)	UK GDP deflator
PPI-MGT(Uk)	SIC_3513 (national)	
PPI-appa(Eu)	SOC_2123 (regional)	
PPI-appa(Uk)	SOC_2123 (national)	
RCI-Infras	SOC_3112 (regional)	
RCI-build	SOC_3112 (national)	
GFCFD	SOC_41 (regional)	
	SOC_41 (national)	
	SOC_52 (regional)	
	SOC_52 (national)	
	SOC_524 (regional)	
	SOC_524 (national)	
	SOC_5241 (regional)	
	SOC_5241 (national)	
	SOC_5243 (national)	
	BEAMA_electrical_labour (national)	

Source: Frontier Economics

We have imposed some restrictions to the set of considered input price combinations. The rationale for these restrictions is twofold: first, we require our recommended cost specification to have a sound economic interpretation; and second, given data limitations and significant degree of collinearity between input price variables, the model can only accommodate a limited number of input prices as cost drivers. The restrictions we have imposed are:

- input price combinations should include one or two labour input prices and one or two capital prices; and
- input price combinations cannot include more than one measure of general inflation.

Although not an input price, we have also considered, in line with our model specification in Phase 1, combinations of input prices with and without a linear time trend.

Among all the possible input price combinations fulfilling the above restrictions, we have found four specifications with particularly robust econometric properties and which are consistent with economic theory. All four specifications include a labour price and a capital price. None of them includes a general measure of inflation or the time trend. The reason for the model rejecting the use of a general measure of inflation and the time trend might be due to a high degree of collinearity between these two variables and the capital and labour price series used.

The four specifications identified use the same capital price but different labour input prices. The capital price used is the BEAMA index for Basic Electrical Equipment. In terms of the labour input prices, one specification uses a price series based on regional wages while the other three use each a labour price based on national wages. The labour prices used by each of the four specifications are[17]:

- SIC_35 (regional)
- SIC_35 (national)
- SIC_3513 (national)
- BEAMA_electrical_labour (national)

In the process of exploring all possible input prices, we have found other combinations of input prices that result in reasonably well specified cost functions. However, in no case have these alternative specifications shown a better performance, in terms of econometric properties and economic interpretation, than the four candidates above.

In order to simplify the presentation of the results, we refer to the specification using a regional wage as specification 1. The results for the three specifications using a national wage are being referred to as specification 2.

3.6.1 Specification 1 – Regional wages

Table 5 below presents the main regression output, under both random effects and pooled OLS for this specification. In the Pooled OLS regression, we use clustered standard errors based on the 14 DNOs.

[17] We imposed homogeneity on the cost function by deflating totex and the labour input price by the price of capital.

Table 5. Specification 1 – Regional wages; Random Effects and Pooled OLS

	Random Effects	Pooled OLS
Customers	0.469***	0.396*
Peak	0.351***	0.434**
Density	-0.078***	-0.082***
Wages (regional SIC-35)	0.326***	0.337***
Price of capital[18] (BEAMA)	0.674	0.663
Constant	-8.21***	-7.78***
R^2 [19]	0.887***	0.887***

The table reports the estimated coefficient for each variable and the confidence intervals using a 95% probability.[20]

*** Significant at 1% ** Significant at 5% *Significant at 10%

Source: Frontier Economics

The sum of the coefficients for the output cost drivers, customers and peak demand, equals 0.81 under Random Effects and 0.82 under Pooled OLS which suggests modestly increasing returns to scale.[21] The sign and value of this overall elasticity is broadly consistent with findings in other studies of the GB DNOs, including previous work undertaken by members of the Frontier team[22].

---

18 Due to the imposition of homogeneity of degree +1 in input prices, we can infer the coefficient for the capital price as 1-coefficient on wages.

19 The table reports the overall R^2 in the Random Effects and the Pooled OLS regression. We use ***, **, * to indicate the overall goodness of fit using the p-value of the Chi-square test for Random Effects and the p-value of the F test for Pooled OLS.

20 The intervals report that with 95% probability the estimated coefficient will be within the confidence interval. The intervals are calculated using the variances of the estimated coefficient, the higher the variance the less precise are the estimates of the coefficients and the wider the confidence intervals.

21 We have tested the hypothesis of increasing returns to scale and confirm this with 98% (random effects) and 96% (Pooled OLS) probability

22 Burns, Philip and Tom Weyman-Jones, Cost Drivers and Cost Efficiency in Electricity Distribution: A Stochastic Frontier Approach, *Bulletin of Economic Research*, Vol. 48, No.1, pp. 41-64, 1996

Under both Random Effects and Pooled OLS models, the estimated coefficients for the input prices show a relative contribution to total costs of around 40% for labour and 60% for capital. We do not regard these values as unreasonable.[23]

We have found that the Random Effects model performs better than the Pooled OLS model for three reasons:

- the significance of the individual parameters is higher under Random Effects than under Pooled OLS;
- the Breusch Pagan test rejects the hypothesis that there are not significant differences across DNOs, which indicates that the Random Effects estimator is more appropriate than Pooled OLS.
- the Random Effects model requires the explanatory variables to be uncorrelated with the individual specific error term, and we do not find any correlation between the cost drivers and the estimated time invariant residuals.

Table 6 presents the rankings and efficiency scores obtained for this specification under both Random Effects and Pooled OLS. The Pooled OLS regression does not provide an aggregate efficiency score for each DNO for the whole period. Instead, we have calculated the average efficiency score taking the mean of each DNO scores across the six years in the sample. We have then normalised the resulting average efficiency scores under Pooled OLS giving a value of one to the most efficient DNO in the sample.

Table 6. Efficiency score and rankings of Specification 1

DNO	Random Effects		Pooled OLS	
	Ranking	Efficiency Score	Ranking	Efficiency Score
WMID	13	0.840	13	0.809
EMID	5	0.947	5	0.933
ENWL	8	0.900	9	0.869
NPgN	7	0.938	7	0.909
NPgY	1	1.000	2	0.988
SWales	2	0.996	3	0.98

23 Remember that the coefficient of the input prices can be interpreted as the budget share of each input in the costs.

SWest	4	0.967	4	0.947
LPN	9	0.896	8	0.877
SPN	10	0.874	10	0.847
EPN	12	0.842	12	0.816
SPD	6	0.941	6	0.923
SPMW	14	0.820	14	0.798
SSEH	11	0.865	11	0.839
SSES	3	0.996	1	1

Source: Frontier Economics

Both models show very similar efficiency rankings.

3.6.2 Specification 2 – National wages

The three specifications presented under specification 2 use a labour price based on national wages.

The results are very similar across all three models, in terms of both estimated coefficients and efficiency scores. Despite these similarities, at the end of this section we argue that using the series based on SIC_35 wages could be marginally preferred over the series based on SIC_3513 wages and the BEAMA cost index. **Table 7** and **Table 8** present the results under Random Effects and Pooled OLS (with clustered standard errors) for the three specifications using national wages.

Table 7. Specification 2 – National wages; Random Effects

	SIC_35 (national)	SIC_3513 (national)	BEAMA_electrical-labour
Customers	0.585***	0.536***	0.566***
Peak	0.239*	0.288**	0.258*
Density	-0.056*	-0.058*	-0.057*
Wages	0.542***	0.744***	0.542***
Price of capital [24] (BEAMA)	0.458	0.256	0.458
Constant	-8.64***	-7.91***	-9.44***
R^2 [25]	0.875***	0.876***	0.875***

The table reports the estimated coefficient for each variable and the confidence intervals using a 95% probability.[26]

*** Significant at 1% ** Significant at 5% *Significant at 10%

Source: Frontier Economics

[24] See footnote 18.

[25] See footnote 19.

[26] See footnote 20.

Table 8. Specification 2 – National wages; Pooled OLS

	SIC_35 (national)	SIC_3513 (national)	BEAMA_electrical-labour
Customers	0.422*	0.410*	0.417*
Peak	0.415*	0.427*	0.420*
Density	-0.064***	-0.064***	-0.064***
Wages	0.466***	0.668***	0.470***
Price of capital [27] (BEAMA)	0.534	0.332	0.530
Constant	-7.840***	-7.355***	-8.608***
R^2 [28]	0.878***	0.878***	0.878***

The table reports the estimated coefficient for each variable and the confidence intervals using a 95% probability.[29]

*** Significant at 1% ** Significant at 5% *Significant at 10%

Source: Frontier Economics

It is clear from these tables that the results are very similar. The sum of the coefficients of the output cost drivers, customers and peak, is virtually the same under the three wage series, and across both estimation techniques and is very close to what we found for Specification 1. However, peak demand has now lost its statistical significance, especially when using SIC_35 or BEAMA as labour prices. Density is also found to be marginally less statistically significant compared to the specification using regional wages.

---

27 See footnote 18.

28 See footnote 19.

29 See footnote 20.

The estimated coefficients for the labour and capital prices are very similar under SIC_35 and BEAMA. They both show labour cost shares of around 60% - 70% and consequently capital cost shares of around 30% - 40%, which do not seem unreasonable. On the other hand, series SIC_3513 yields a labour cost share of around 90%, a rather large value, which is implausible (but we note that the confidence interval around this estimate would support values at more reasonable, lower levels).

We have again assessed which of the two estimation techniques performs better, and conclude that the Random Effects model performs better than Pooled OLS on the basis of the Breusch-Pagan test and correlation analysis of the residuals and explanatory factors.

Table 9 and **Table 10** show the rankings and efficiency scores obtained under Random Effects and Pooled OLS.

Table 9. Efficiency score and rankings of Specification 2; Random Effects

DNO	SIC_35 (national)		SIC_3513 (national)		BEAMA_electrical-labour	
	Ranking	Efficiency Score	Ranking	Efficiency Score	Ranking	Efficiency Score
WMID	12	0.828	12	0.828	12	0.829
EMID	4	0.952	4	0.955	4	0.953
ENWL	8	0.891	8	0.886	8	0.890
NPgN	2	0.996	2	0.994	2	0.995
NPgY	3	0.989	3	0.989	3	0.989
SWales	1	1.000	1	1.000	1	1.000
SWest	7	0.904	7	0.902	7	0.904
LPN	9	0.872	9	0.876	9	0.874
SPN	10	0.855	10	0.851	10	0.854
EPN	13	0.822	13	0.822	13	0.822
SPD	6	0.945	6	0.945	6	0.945
SPMW	14	0.789	14	0.797	14	0.792
SSEH	11	0.829	11	0.834	11	0.831

SSES	5	0.947	5	0.953	5	0.950

Source: Frontier Economics

In **Table 10** we have again normalised the resulting average efficiency scores under Pooled OLS giving a value of one to the most efficient DNO in the sample.

Table 10. Efficiency score and rankings of Specification 2; Pooled OLS

DNO	SIC_35 (national)		SIC_3513 (national)		BEAMA_electrical-labour	
	Ranking	Efficiency Score	Ranking	Efficiency Score	Ranking	Efficiency Score
WMID	12	0.816	12	0.816	12	0.816
EMID	5	0.965	5	0.965	5	0.965
ENWL	9	0.866	9	0.863	9	0.865
NPgN	3	0.99	3	0.989	3	0.99
NPgY	2	0.996	2	0.996	2	0.996
SWales	1	1	1	1	1	1
SWest	7	0.887	7	0.886	7	0.887
LPN	8	0.877	8	0.877	8	0.877
SPN	10	0.837	10	0.836	10	0.836
EPN	13	0.816	13	0.816	13	0.816
SPD	6	0.947	6	0.947	6	0.947
SPMW	14	0.799	14	0.801	14	0.8
SSEH	11	0.825	11	0.826	11	0.825
SSES	4	0.977	4	0.979	4	0.978

Source: Frontier Economics

The rankings are very stable across the three national wage series using both Random Effects and Pooled OLS. As before Random Effects and Pooled OLS deliver similar efficiency rankings for each of the three national wage specifications.

In summary, all three national wage series show similar and broadly reasonable econometric and economic properties. The major differences between the three specifications are:

- the series based on SIC_3513 delivers marginally more significant coefficients;
- the labour price coefficient using the series based on SIC_3513 is higher than we would expect for the budget share of labour costs.

Given these results, we have a preference for the specifications with a labour price based on either the SIC_35 code or the BEAMA_electrical-labour index. In order to have some consistency with Phase 1, we recommend the specification with a labour price based on the SIC_35 code. We note however, that the other two specifications using a national wage could both be used without leading to any discernible difference in the conclusions one would draw from the results.

3.6.3 Comparison of specifications 1 and 2

Table 11 compares the two preferred specifications with regional and national labour prices. Whilst the results are very similar across the two specifications there are some differences. First, we observe that some explanatory power is transferred from peak to customers when regional wages are replaced by a national average. This is not a particular concern as the sum of the two coefficients for the output cost drivers is almost identical in both specifications. This effect can be justified by the high correlation that exists between *customers* and *peak*, which is equal to 0.9573 and is the highest among any two cost drivers.

Regarding density, the estimated coefficient is different under regional and national wages, though this difference is small and not significant once confidence intervals are taken into account.

The specification with regional wages allocates 40% of the costs to labour inputs, whereas the specification using national wages estimates that this share is just below 70%. Despite the difference, the two parameters are not statistically different from each other, confirming that the two models are not necessarily contradictory.

In terms of statistical indicators, the specification using regional wages has a slightly higher R^2. Based on this, we do not find strong evidence that one specification performs better than the other and therefore recommend that both are considered by Ofgem.

Table 11. Comparison of Specification 1 (SIC_35, regional) and Specification 2 (SIC_35, national); Random Effects

	Specification 1 (regional wage)	Specification 2 (national wage)
Customers	0.469***	0.585***
Peak	0.351***	0.239*
Density	-0.078***	-0.056*
Wages	0.326***	0.542***
Price of capital[30] (BEAMA)	0.674	0.458
Constant	-8.21***	-8.64***
R^2 [31]	0.887***	0.875***

The table reports the estimated coefficient for each variable and the confidence intervals using a 95% probability.[32]

*** Significant at 1% ** Significant at 5% *Significant at 10%

Source: Frontier Economics

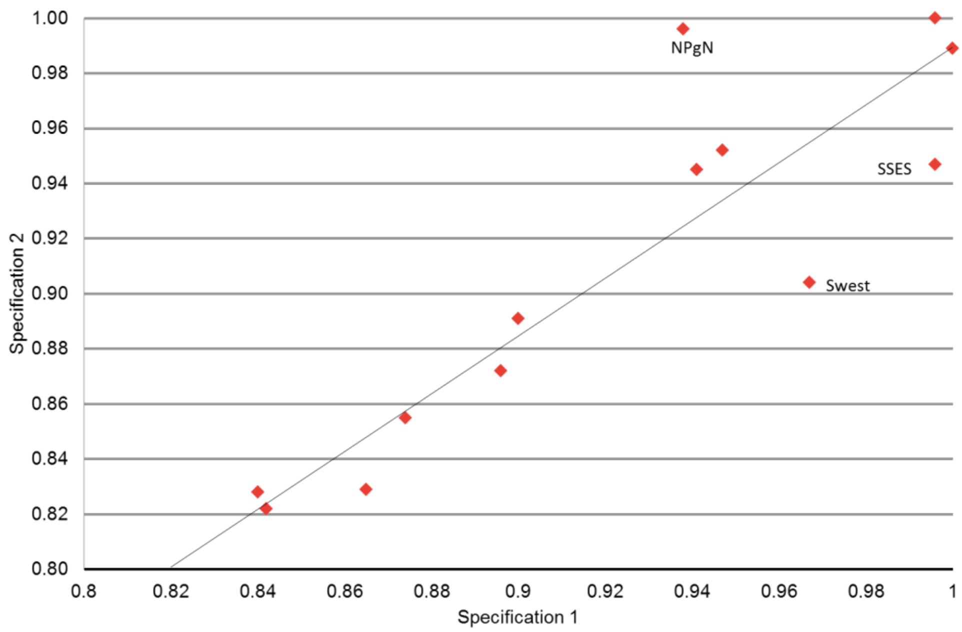
Figure 9 below shows a comparison of efficiency scores under the two specifications. We observe that except for a few exceptions, namely NPgN, SWest and perhaps SSES, both specifications deliver very similar results in terms of efficiency scores.

30 See footnote 18.

31 See footnote 19.

32 See footnote 20.

Figure 9. Scatter of efficiency scores under Specification 1 and Specification 2



Source: Frontier Economics

3.7 Conclusion

In our analysis of input prices, we have found two different specifications with sound econometric and economic properties.

The two specifications use identical output and environmental cost drivers, customers, and peak load, and the average density of customers in the service area of each DNO. Both specifications also use the same series for the input price of capital, the BEAMA series of Basic Electrical Equipment is used.

The two models differ in the series used for the labour input price. Specification 1 uses a price series based on regional wages while Specification 2 uses a national average wage. Given the results obtained, we recommend using SIC-35 to obtain the average wages underlying the two labour price series, national and regional.

Both specifications are economically sensible and can be supported by the data. Consequently we recommend that Ofgem considers the results of both of these specifications.

4 Alternative measures of density

In this section we investigate whether the simple average measure of density used in our Phase 1 report can be improved on by using measures that capture the variation of density not only across DNOs, but also within each DNO.

4.1 Phase 1

As we described in our Phase 1 report, we expect connection density to be an important cost driver. This is because of two effects:[33]

- **Geometric effect** – Fewer assets are needed to serve customers as they become closer together, reducing costs as density increases. This implies a downward sloping relationship between density and total costs.

- **Urbanisation effect** – At some point the geometric effect could be, at least partly, offset by increased costs associated with serving high density areas. For example, this could be the result of safety requirements resulting in more distribution assets being located underground in urban areas, increased traffic congestion, more difficulty accessing infrastructure, and associated higher installation and maintenance costs.

In principle both low and high density could lead to higher costs, implying a U-shaped relationship between connection density and total costs. This would be the case if the geometric effect dominates at low density levels, while the urbanisation effect dominates at higher densities.

During Phase 1 we analysed two simple measures of density, i.e. customers per unit of service area and customers per network length. Using these very simple measures, the Phase 1 analysis found evidence of increased density decreasing costs only (i.e. evidence that the geometric effect outlined above dominates in the GB sample). However we noted that the simple measures used might be insufficiently detailed to allow the identification of both density effects. Consequently we have revisited the issue of connection density in more detail during Phase 2.

4.2 Phase 2 approach

The simple measure of density used during Phase 1 allow the variations in average density across DNOs to be captured in the econometric model, but they

---

33 The characterisation of these two effects follow the notation used in Frontier Economics and Consentec, 2009, "Impact of connection density on regional cost differences for network operators in the Netherlands", A report prepared for Energiekamer.

do not enable an investigation of the impact of the variation in density within each DNO's service area on costs incurred. Yet most DNOs serve a wide variety of different types of terrain, including relatively sparsely populated rural regions, moderately dense suburban regions and (possibly highly dense) urban regions. Our Phase 2 work is therefore focused on assessing whether these within-DNO variations in density are an important driver of costs.

To do this we have adopted the following steps:

- Developing sub-DNO measures of density:
  - gather data from public sources on the density of sub areas within each DNO's operating region;
  - use this data to prepare histograms that describe the underlying density composition of each DNO's operating area;
  - use these underlying histograms as a basis from which to develop a wide range of alternative measures of density that describe more fully the underlying distribution.
- Test empirically whether these measures are able to better describe the data than the simple measures used during Phase 1.

The remainder of this section describes these steps in more detail.

4.3 Developing sub-DNO measures of density

4.3.1 Gathering source data

Our approach for Phase 2 begins with the collection of detailed underlying data from the ONS, collated at various levels of geographic granularity.

For England and Wales, we used information available on the Medium Layer Super Output Area (MSOA)-level. An MSOA is an area that is defined by the Office of National Statistics (ONS) and covers a population between 5,000 and 15,000 inhabitants. The exact boundaries of MSOAs are determined in cooperation with local authorities such that they match boundaries of other areas whenever possible (e.g. to follow borders of local authorities, wards or postal code areas). For Scotland we used data at the Intermediate Geographic Zone (IGZ) level, the Scottish equivalent of the MSOA.

Great Britain consists of 7,193 MSOAs and 1,234 IGZs. On average therefore there are approximately 600 subdivisions within each DNO region. We consider that this data is sufficiently granular to allow the remaining steps of our methodology to be undertaken robustly.

We used underlying information to derive analysis for two types of density:

- **meter density** – the number of meters (domestic, Economy 7 and non-domestic) in an area divided by its surface area; and
- **demand density** – the total electricity demand (domestic, Economy 7 and non-domestic) in an area divided by its surface area.

Each of these measures was calculated for the sub-areas served by GB DNOs, taking into account the different surface areas of these sub-areas.

In working with the underlying ONS data it was necessary to make a number of judgements while processing the data. In Annexe at the end of this Volume we provide more detail on the data sources and the steps we have taken to match the data to our sample, including the cross checks we have undertaken to verify the robustness of the data.

4.3.2 Preparation of underlying density histograms

At the conclusion of our first stage, we were able to map the MSOA and IGZ data for each sub-area to the DNO areas. Using this data we then derived meter density and demand density for each sub-area. The sub-areas have different sizes, and in order to create comparable statistics (i.e. per hectare and not per MSOA or IGZ), sub-area surface area was used as a weight in the analysis.

Using weighted density measures we were then able to develop the full distribution of density for each GB DNO in 2010.

We illustrate the output of our analysis by showing the meter density histograms for Yorkshire, South Wales, Scottish Hydro and London license areas.

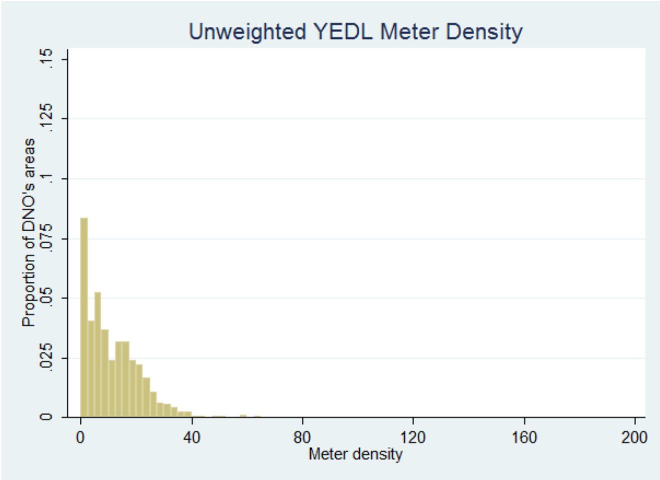
These histograms report the proportion of areas (MSOAs or IGZs) for each DNO that fall in each of the meter density classes considered, and defined by intervals of width 2.5 meters (customers) per hectare. For example, **Figure 10** shows that in the Yorkshire region 8.5% of the areas have a meter density between 0 and 2.5 customers per hectare, and 4% of the areas have a meter density between 2.5 and 5 customers per hectare. The height of the bars sums to one for each histogram.[34]

These histograms illustrate that there are material differences between DNOs, with SSE-Hydro and LPN appearing markedly different in respect of the underlying distribution of their density.

---

[34] In order to show more clearly the differences between regions, we show unweighted histograms in this section, i.e. without taking account of the differing size of the sub areas. When weighted histograms are drawn, the density accounted for by sparse sub areas increases markedly. We present the weighted histograms for all DNOs in the Annexes at the end of this volume.

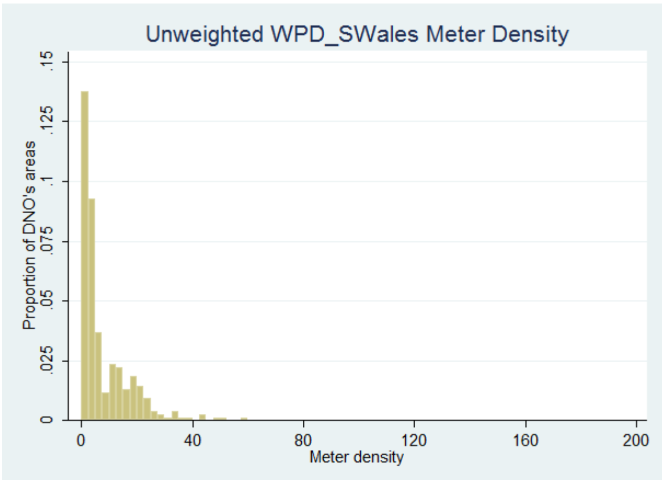
Figure 10. Example of histogram of meter density; Yorkshire



Source: Frontier Economics analysis of ONS data

Note: The x-axis shows the underlying meter density bands we have considered. For each we then calculate the proportion of the histogram's density that falls within, as shown on the y-axis.

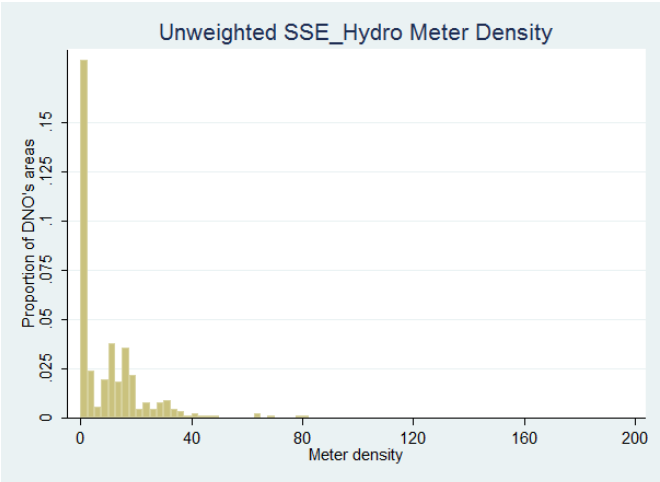
Figure 11. Example of histogram of meter density; South Wales



Source: Frontier Economics analysis of ONS data

Note: The x-axis shows the underlying meter density bands we have considered. For each we then calculate the proportion of the histogram's density that falls within, as shown on the y-axis.

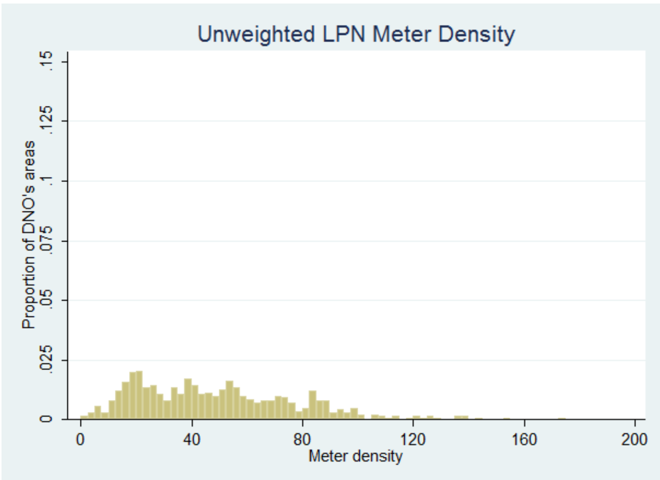
Figure 12. Example of histogram of meter density; Scottish Hydro



Source: Frontier Economics analysis of ONS data

Note: The x-axis shows the underlying meter density bands we have considered. For each we then calculate the proportion of the histogram's density that falls within, as shown on the y-axis.

Figure 13. Example of histogram of meter density; London



Source: Frontier Economics analysis of ONS data

Note: The x-axis shows the underlying meter density bands we have considered. For each we then calculate the proportion of the histogram's density that falls within, as shown on the y-axis.

The histograms reveal that – with the exception of London – sparsely populated sub-areas make a large proportion of all DNOs' service areas. This is particularly marked for SSEH. Each of these DNOs then has a "tail" of its service area which is made up of denser sub areas.

As one would expect, the histogram for London is visibly different, with sparsely populated regions making up a much smaller proportion of its total service area. Instead a significant proportion of London's service region is made up of relatively dense service areas.

4.3.3 Developing alternative measures of density

Using the data we have derived, we have considered a very wide range of ways in which it is possible to summarise the resulting distributions of density. We have calculated a range of statistics to capture the underlying heterogeneity of connection density and these are summarised in **Table 12** below.

Each measure was estimated at the DNO level both from the meter density data and the demand density data (with the exceptions of the Gini coefficient, which was estimated using meter statistics only). We also estimated squared terms for some of the measures in order to test for the possible existence of a quadratic relationship (i.e. to allow direct testing of the existence of a U-shaped relationship between density and cost). The analysis looked both at including single and multiple density measures in the model.

Table 12. Measures used in the density analysis

Measure	Description
Mean	Mean density, weighted by sub-area surface area. Conceptually identical to the density variable used in our Phase 1 work, but derived from the detailed ONS data.
Standard deviation	Standard deviation of the distribution of density, weighted by sub-area surface area.
Skewness	Skewness of density, weighted by sub-area surface area, summarising the extent to which the tail on one side of the distribution is longer than the other (equivalently, whether the bulk of the distribution lies to below or above the mean).
Kurtosis	Kurtosis of density, summarising how "peaked" the distribution is.
Gini coefficient	A measure of inequality between zero and one where zero would imply that density is equal across the DNO's surface area and 1 would imply that customers are concentrated in one unit of the DNO's surface area, with the remaining surface area empty. .
Share of surface area below a given density threshold	The proportion of the DNO's surface area below a given density level.
Total surface area below a given density threshold	The DNO's total surface area below a given density level.
Share of surface area above a given density threshold	The proportion of the DNO's surface area above a given density level
Total surface area above a given density threshold	The DNO's total surface area above a given density level

Note that the density data used in the analysis was from 2010. We assumed that density did not vary significantly over time and therefore applied the 2010 data over the whole panel period. We view this as a reasonable assumption for the meter density variables given that surface area is fixed and the number of meters is unlikely to have changed substantially year on year at the MSOA or IGZ level during the sample period.

In the tables and figures below we provide a brief descriptive analysis of each of the measures set out in Table 1. We focus on meter density, but can confirm that qualitatively similar results have been derived for demand density.

Table 13 below shows a range of summary statistics, including the first four moments of the histogram distributions (mean, standard deviation, skewness and kurtosis) and also the Gini coefficient. The information in **Table 13** allows us to confirm, as one would anticipate, that London (LPN) and Northern Scotland (SSEH) are both clear outliers, representing opposite extremes in the sample. On all measures, with the exception of the Gini coefficient, both LPN and SSEH are either far above or far below the typical range for the other DNOs.

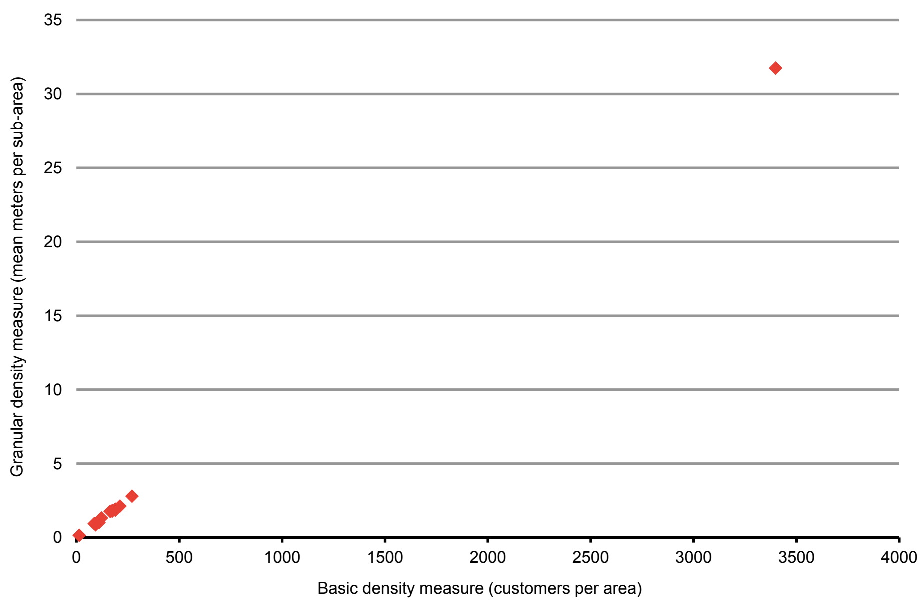
Table 13. Meter density measures (2010)

DNO	Mean density	Standard deviation	Skewness	Kurtosis	Gini coefficient
WMID	1.85	4.47	3.76	18.98	0.43
EMID	1.77	4.05	4.67	32.55	0.42
ENWL	1.91	4.72	4.00	22.14	0.45
NPGN	1.00	3.22	5.95	46.36	0.45
NPGY	2.12	4.45	3.96	23.42	0.42
SWales	0.86	2.50	8.06	98.99	0.43
SWest	1.03	3.27	7.61	78.35	0.42
LPN	31.75	26.13	1.29	4.71	0.30
SPN	2.80	6.22	5.63	57.09	0.42
EPN	1.77	4.56	5.74	50.14	0.42
SPD	0.94	4.04	8.83	122.71	0.47
SPMW	1.32	3.76	5.42	40.91	0.44
SSEH	0.15	1.29	24.33	849.83	0.45
SSES	1.82	4.49	5.15	40.31	0.42
GB	1.25	4.32	8.51	124.44	0.50

Source: Frontier Economics

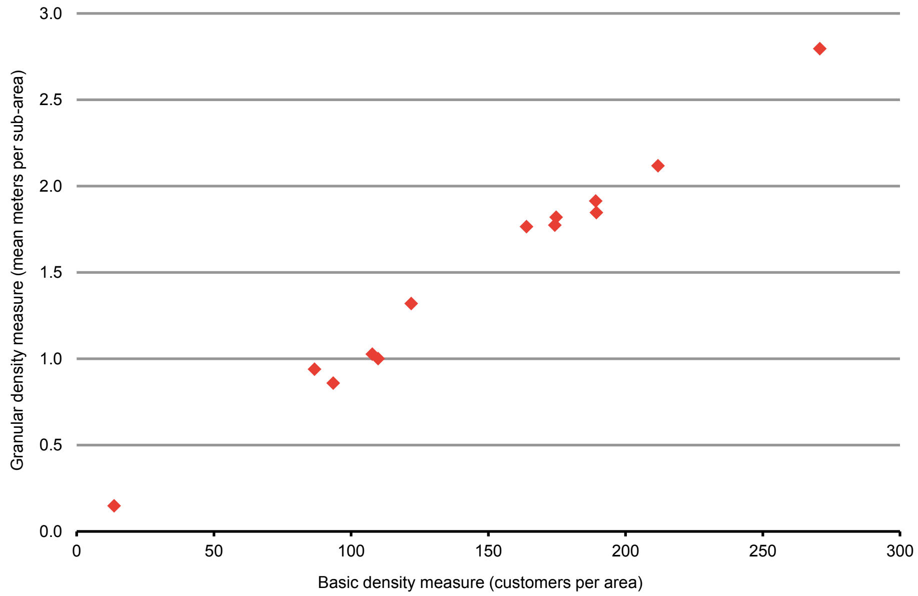
Figure 14 and **Figure 15** below compare the basic density measure from Phase 1 (customers per service area) with mean meter density for all DNOs. Since the inclusion of LPN in the chart limits the ability of the reader to view the dispersion of the remaining DNOs, we show the same figure excluding LPN. As one would expect we find that these measures are highly correlated – the correlation coefficient between the two measures for all DNOs is 0.99. This confirms that the "simple" Phase 1 measure of density can be understood to be capturing essentially the same information as the "sophisticated" mean variable we developed as described above for Phase 2, and therefore confirms and validates that data drawn from the sources used to construct the alternative measure is consistent with the data that is used to derive the simple measure used in Phase 1.

Figure 14. Scatter plot of basic Phase 1 density measure against mean meter density (2010)



Source: Frontier Economics

Figure 15. Scatter plot of basic Phase 1 density measure against mean meter density (2010); excluding LPN



Source: Frontier Economics

As described in **Table 12**, we have also calculated a wide range of other variables for econometric testing. These variables are based on calculating the proportion (or absolute amount) of each DNO's service region that lies above/below some given threshold. **Table 14** shows these proportions for each DNO for meter density. The motivation for constructing these variables was to allow the direct testing of whether density below or above some given level could be shown to result in costs higher than some baseline level (i.e. to investigate directly the support for a U-shaped curve).

Taken together with the first four moments of the distribution and the Gini coefficient described in **Table 13**, these variables provide a rich basis from which to investigate the relationship between density and cost. This is the final stage of our investigation and the results of this are reported in the following section.

Table 14. Percentage of DNO surface area with meter density in each bracket

DNO	< 0.25	0.25 ≤ x < 0.5	0.5 ≤ x < 0.75	0.75 ≤ x < 1	1 ≤ x < 2	2 ≤ x < 5	5 ≤ x < 10	10 ≤ x < 25	25 ≤ x < 50	50 ≤ x < 75	75 ≤ x < 100	100 ≤ x < 125	125 ≤ x
EMID	16%	44%	10%	5%	9%	7%	5%	4%	0%	0%	0%	0%	0%
ENW	54%	11%	7%	3%	7%	6%	5%	5%	1%	0%	0%	0%	0%
EPN	14%	38%	18%	9%	9%	5%	4%	4%	1%	0%	0%	0%	0%
LPN	0%	0%	0%	0%	3%	7%	8%	35%	25%	13%	6%	1%	1%
NEDL	68%	13%	4%	2%	4%	5%	2%	2%	0%	0%	0%	0%	0%
SP	76%	8%	5%	2%	3%	2%	1%	2%	1%	0%	0%	0%	0%
SPN	0%	35%	19%	8%	14%	8%	7%	6%	1%	0%	0%	0%	0%
SP_Manweb	52%	22%	7%	3%	4%	5%	3%	3%	0%	0%	0%	0%	0%
SSE_Hydro	95%	3%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
SSE_Southern	18%	41%	12%	5%	9%	6%	3%	5%	1%	0%	0%	0%	0%
WMID	26%	39%	12%	4%	6%	4%	3%	6%	1%	0%	0%	0%	0%
WPD_SWales	62%	13%	5%	3%	7%	7%	2%	1%	0%	0%	0%	0%	0%
WPD_SWest	41%	29%	13%	5%	4%	3%	2%	2%	0%	0%	0%	0%	0%
YEDL	27%	26%	12%	6%	9%	9%	7%	5%	1%	0%	0%	0%	0%
GB	54%	20%	8%	3%	5%	4%	2%	3%	1%	0%	0%	0%	0%

Source: Frontier Economics

4.4 Results

We have undertaken what we consider to be an exhaustive review of all candidate measures, including looking at certain measures in combination. This section presents results for the density measures that worked best when added to our base model, both in terms of their economic interpretation and statistical properties.

We discuss results for:

- mean density (both the Phase 1 and Phase 2 measures);
- standard deviation;
- skewness;
- kurtosis;
- Gini coefficient;
- threshold variables (i.e. density above and density below some cut off);
- a combination of thresholds and mean; and
- squared terms.

Finally, we provide a discussion of the two outliers in our sample of 14, i.e. LPN and SSEH. We show how the resulting estimates of efficiency for those two DNOs are influenced by the inclusion or otherwise of the other.

In each case we have tested these density measures as a possible addition to our base model, i.e. the preferred capital and labour price series (as set out in Section 3) and our preferred approach to capturing quality of service (as set out in Section 5). We note, however, that we have also tested these density measures against a variety of other input price combinations and quality of service treatments in order to ensure that we were not missing an alternative approach, making use of an entirely different set of drivers, which was better in aggregate. No superior alternative was found.

4.4.1 Analysis of mean density

We have estimated our base models replacing the Phase 1 density measure (i.e. customers per service area) with the mean density per hectare for each DNO derived from our detailed analysis of ONS data. Results are reported in **Table 15** below.

Table 15. Regression results using mean density vs. basic density; Random Effects

Variable	Regional wage specification		National wage specification	
	Using Phase 1 density measure[35]	Using mean meter density	Using Phase 1 density measure	Using mean meter density
Customers	0.469***	0.467***	0.585***	0.585***
Peak	0.351***	0.354***	0.239*	0.238*
Regional wage	0.326***	0.330***		
National wage			0.541***	0.548***
Phase 1 density measure	-0.0777***		-0.0564*	
Mean meter density		-0.0780**		-0.0552*
Constant	-8.207***	-8.540***	-8.635***	-8.868***
R^2 [36]	0.887***	0.884***	0.975***	0.872***

The table reports the estimated coefficient for each variable and the confidence intervals using a 95% probability.[37]

*** Significant at 1% ** Significant at 5% *Significant at 10%

Source: Frontier Economics

As expected, the coefficient on the mean meter density is very similar to the coefficient on the Phase 1 density measure. There is also little change in the coefficients and significance of the other variables in the model, and the overall performance of the models (whether using the regional or the national wage).

This confirms and validates that data drawn from the sources used to construct the alternative measure is consistent with the data that is used to derive the simple measure used in Phase 1. The two measures can be understood to contain essentially the same information. Consequently, if mean density is to be our preferred measure, our preference is to use the simpler, more transparent and more readily updated Phase 1 density measure.

35 Customers per service area

36 See footnote 19.

37 See footnote 20.

4.4.2 Analysis of the standard deviation of density

We tested using the standard deviation of meter density in the model. In theory, higher standard deviation of density could:

- raise costs if greater diversity of density means that DNOs are less able to benefit from economies of scale in serving different area types; or
- lower costs given that standard deviation of meter density and mean meter density are highly positively correlated;[38] and increased mean meter density is associated with lower costs.

The results supported a negative impact of standard deviation on total costs, as shown in **Table 16**. The coefficient on the standard deviation is statistically significant at 5% for the regional wage specification, and suggests a stronger effect of density on costs than when using the mean or basic measures. However, the standard deviation variable was insignificant at the 10% level in the national wage specification.

Table 16. Regression results using standard deviation of density; Random Effects

Variable	Regional wage specification	National wage specification
Customers	0.455***	0.579***
Peak	0.341**	0.221
Regional wage	0.326***	
National wage		0.553***
S.d. of meter density	-0.115**	-0.0762
Constant	-8.137***	-8.555***
R^2 [39]	0.875***	0.865***

The table reports the estimated coefficient for each variable and the confidence intervals using a 95% probability.[40]

*** Significant at 1% ** Significant at 5% *Significant at 10%

Source: Frontier Economics

The change in the efficiency scores and rankings was limited for these specifications. In the regional wage specification the correlation between

38 Their estimated correlation coefficient is 0.993.

39 See footnote 19.

40 See footnote 20.

efficiency scores between the model including mean density and that including the standard deviation was 98%, while for the national wage specification it was 99%. The absolute change in efficiency scores compared to the specification using mean density was limited, averaging at 2.5% for the regional wage specification and 1.7% for the national wage specification, with a maximum change of 5.2% in the regional wage specification and 3.8% in the national wage specification, both for SPD.

4.4.3 Analysis of the skewness of density

We similarly analysed the possible impact of the skewness of distribution of meter density on total costs for DNOs. Density is positively skewed for all the DNOs (most of the distribution located below the mean, long tail to the right), and markedly more so for SSEH.[41] As with standard deviation, higher skewness could theoretically raise or lower total costs:

- more positive skew could raise costs if it implies the DNO must serve a relatively smaller amount of urbanised areas which may for example be associated with higher costs of meeting safety requirements; or
- more positive skew could reduce costs if there is a scale economy effect from having a more highly heterogeneous service area in terms of connection density.

The results in **Table 17** support the hypothesis that costs increase with skewness. This could also be driven by the impact of mean density on costs, which is negatively correlated with skewness.

41 Skewness of meter density is 24.3 for SSEH, compared to an average skewness of 5.4 for the remaining DNOs.

Table 17. Regression results using skewness of density; Random Effects

Variable	Regional wage specification	National wage specification
Customers	0.473***	0.596***
Peak	0.352**	0.240*
Regional wage	0.314***	
National wage		0.548***
Skewness of meter density	0.133**	0.106*
Constant	-8.905***	-9.258***
R^2 [42]	0.880***	0.875***

The table reports the estimated coefficient for each variable and the confidence intervals using a 95% probability.[43]

*** Significant at 1% ** Significant at 5% *Significant at 10%

Source: Frontier Economics

As before, the change in the efficiency scores and rankings was limited for these specifications. For example, for the regional wage specification the correlation of efficiency scores between the model including mean density and that including the skewness was 96%, while for the national wage specification it was 98%. The absolute change in efficiency scores compared to the specification using mean density was limited, averaging at 1.6% for the regional wage specification and 1.3% for the national wage specification, with a maximum change of 3.3% in the regional wage specification and 2.9% in the national wage specification, both for SPN.

4.4.4 Analysis of the kurtosis of density

We also analysed the kurtosis of the distribution of density, which measures the "peakedness" of a distribution. Again, both SSEH and LPN stand out by this measure, with meter density kurtosis of 850 and 5 respectively compared to 53 on average for the remaining DNOs. The results suggest a small, positive coefficient on kurtosis, which suggests that the more peaked the density distribution, the higher the DNO's total costs. These results are summarised in **Table 18**. While the model including kurtosis performs reasonably well, the interpretation of the density coefficient is less clear than when using a simpler density measure.

---

42 See footnote 19.

43 See footnote 20.

Table 18. Regression results using kurtosis of density; Random Effects

Variable	Regional wage specification	National wage specification
Customers	0.482***	0.611***
Peak	0.340**	0.228
Regional wage	0.310***	
National wage		0.553***
Kurtosis of meter density	0.0687**	0.0571*
Constant	-8.988***	-9.402***
R^2 [44]	0.875***	0.873***

The table reports the estimated coefficient for each variable and the confidence intervals using a 95% probability.[45]

*** Significant at 1% ** Significant at 5% *Significant at 10%

Source: Frontier Economics

Again, the change in the efficiency scores and rankings was limited for these specifications. For example, for the regional wage specification the correlation between efficiency scores for the model including mean density and that including the kurtosis was 92%, while for the national wage specification it was 96%. The absolute change in efficiency scores compared to the specification using mean density is limited, averaging at 2.2% for the regional wage specification and 1.7% for the national wage specification. The maximum change is 5.1% for LPN in the regional wage specification and 3.5% in the national wage specification for SPN.

4.4.5 Gini coefficient

We have analysed the inclusion of the Gini coefficient (described above) in our regression models. The results showed a positive and statistically significant coefficient in the regional wage specification but not the national wage specification.

A positive coefficient is consistent with the results found for standard deviation, as the DNOs with higher standard deviation of density are typically those with a flatter, more equal distribution of density across their areas.

[44] See footnote 19.

[45] See footnote 20.

4.4.6 Testing the moments of the distribution and Gini coefficient in combination

We have analysed a wide range of combinations of the density variables. We found that multi-collinearity between the measures, coupled with our small sample size, meant that when more than one was included in the model, at least one of the density variables became statistically insignificant (often both). Consequently, we have not been able to test successfully these coefficients in combination.

Overall, the specifications including one of standard deviation, skewness, kurtosis and the Gini coefficient did not improve the model relative to the specifications using basic density or mean meter density and we saw limited variation in efficiency scores. From this we conclude that our initial Phase 1 finding with respect to the use of mean density appears to be stable. Our results are consistent with the dominance of the "geometric effect" outlined in section 4.1 in the GB sample.

4.4.7 Threshold variables

As described in Section 4.3.3, we have also generated a range of further density variables:

- variables measuring a DNO's total surface area with density above (or below) a given level; and
- variables measuring the share of a DNO's total surface area with density above (or below) a given level.

We used a wide variety of different thresholds in the analysis. To measure low density, we used thresholds of meter density equal to 0.25, 0.5 and 1, and to measure high density, we used thresholds of 2, 5, 10, 25, 50 and 75. The larger number of high density thresholds reflects the long tail of higher density areas shown in the histograms. Since the analysis of mean density identified above has found strong evidence in support of the geometric effect (i.e. the downward sloping portion of the U-curve) we have focused attention on testing for the possible existing of the upward sloping portion of the U-curve. The results from including these metrics in the regressions were mixed and in some cases seemingly contradictory.

In the specification using the regional wage, four out of the six "high" density surface area measures had positive coefficients when included in place of the conventional density measure. However, only one of these was statistically significant, and then only at the 10% level. For this threshold, the metric using the same threshold but the proportion of the DNO's surface area (rather than the actual surface area) had a coefficient that was negative and statistically significant at the 5% level.

For the specification using the national wage, none of the high density surface area measures had statistically significant coefficients, and two out of the six high density share measures had statistically significant coefficients, both of which were negative.

This analysis did not therefore reveal direct support for the existence of the upward sloping portion of the U-curve.

4.4.8 Combining mean density with threshold variables

The analysis also looked at combining measures capturing areas of high density with measures of overall density (e.g. the Phase 1 mean meter density).

For the specification using the national wage and high density threshold measures, only one of the twelve high density variables tested showed a statistically significant coefficient, and this was negative (i.e. reinforcing the geometric effect). For the regional wage specification, again only one of the twelve high density measures tested showed a statistically significant coefficient (at the 10% level), which was positive.

Again, this analysis did not therefore reveal direct support for the existence of the upward sloping portion of the U-curve.

4.4.9 Introducing squared terms

We also looked at including the basic density measure in both level and squared terms. This resulted in a negative coefficient on the level term and a small positive coefficient on the squared term. However both were statistically insignificant at the 10% level, in both the national and regional wage specifications.

4.4.10 The impact of outliers

As we have noted above, our analysis of the detailed composition of the density of each DNO’s service region has highlighted both LPN and SSEH as potential outliers, with characteristics markedly different from those of more typical GB DNOs. We wished to test the effect of these potentially pivotal observations in our sample by testing the effect of dropping LPN, SSEH and both of these outliers from the sample. The regression results for our two preferred specifications with a reduced sample are shown in **Table 19** and **Table 20**.

Across both specifications we observe that while the coefficient estimates for most variables are broadly stable, the coefficient on density is more sensitive to the sample. Specifically:

- with SSEH excluded, the estimate of the coefficient on density decreases in magnitude;

- with LPN excluded, the estimate of the coefficient on density increases in magnitude; and
- with both excluded, the absolute coefficient estimate reduces and is no longer statistically significant (i.e. absent the two outliers we would conclude that density should not be included in the model).

Table 19. Regression results excluding density outliers, regional wage specification; Random Effects

Variable	Full sample	Excluding SSEH	Excluding LPN	Excluding both outliers
Customers	0.469***	0.553***	0.546***	0.538***
Peak	0.351***	0.311**	0.325**	0.307**
Regional wage	0.326***	0.330***	0.292***	0.289***
Phase 1 density measure	-0.0777***	-0.0645**	-0.111*	-0.0301
Constant	-8.207***	-9.149***	-9.012***	-9.164***
R^2 [46]	0.887***	0.863***	0.893***	0.870***

The table reports the estimated coefficient for each variable and the confidence intervals using a 95% probability.[47]

*** Significant at 1% ** Significant at 5% *Significant at 10%

Source: Frontier Economics

46 See footnote 19.

47 See footnote 20.

Table 20. Regression results excluding density outliers, national wage specification; Random Effects

Variable	Full sample	Excluding SSEH	Excluding LPN	Excluding both outliers
Customers	0.585***	0.701***	0.662***	0.645***
Peak	0.239*	0.185	0.232	0.214
National wage	0.541***	0.578***	0.468***	0.482***
Phase 1 density measure	-0.0564*	-0.0372	-0.106	0.00302
Constant	-8.635***	-9.915***	-9.596***	-9.714***
R^2 [48]	0.875***	0.855***	0.882***	0.861***

The table reports the estimated coefficient for each variable and the confidence intervals using a 95% probability.[49]

*** Significant at 1% ** Significant at 5% *Significant at 10%

Source: Frontier Economics

Efficiency scores and rankings for these specifications are reported in **Table 21** and **Table 22** below. The resulting changes to efficiency scores are around 2-4% for LPN when SSEH is excluded and around 1-4% for SSEH when LPN is excluded. We observe that the changes for both outliers through this analysis are more pronounced under the national specification (where the efficiency score for LPN improves by 4.5% for example when SSEH is excluded) than under the regional specification (an improvement of 2.3% for LPN). We also observe that the exclusion of outliers has some effect on other DNOs, with SWales appearing the most sensitive, and particularly sensitive to the exclusion of SSEH.

Notwithstanding the impact on efficiency scores, we observe that this outlier analysis has a more limited effect on efficiency ranking. For example, under either the national or regional specifications we observe that LPN's ranking improves from 9th to 7th or 8th when SSEH is excluded. Similarly, when LPN is excluded SSEH's ranking improves from 11th to 9th.

Though the density results are affected by the two outliers, we believe that this does not undermine the model results estimated using all fourteen DNOs, as LPN and SSEH provide a richer set of information on how density affects costs.

48 See footnote 19.

49 See footnote 20.

In addition, it is preferable to keep the density measure in the model as it avoids an omitted variable problem.

Table 21. Efficiency scores and rankings for the regional wage specification

DNO	All DNOs		SSEH excluded		LPN excluded		LPN and SSEH excluded	
	Efficiency score	Ranking	Efficiency score	Ranking	Efficiency score	Ranking	Efficiency score	Ranking
WMID	83.97%	13	84.15%	12	83.63%	12	83.52%	11
EMID	94.74%	5	94.81%	5	95.39%	6	94.06%	4
ENWL	90.01%	8	90.60%	9	89.83%	8	90.00%	8
NPGN	93.76%	7	92.00%	7	94.14%	7	91.49%	6
NPGY	100.00%	1	100.00%	1	99.22%	2	100.00%	1
SWales	99.64%	2	96.19%	3	98.42%	3	95.20%	3
SWest	96.69%	4	94.92%	4	95.82%	4	92.90%	5
LPN	89.62%	9	91.95%	8	-	-	-	-
SPN	87.43%	10	88.01%	10	85.85%	10	88.24%	9
EPN	84.20%	12	85.49%	11	85.37%	11	84.06%	10
SPD	94.10%	6	92.83%	6	95.58%	5	90.78%	7
SPMW	82.00%	14	79.69%	13	80.53%	13	78.60%	12
SSEH	86.53%	11	-	-	88.48%	9	-	-
SSES	99.61%	3	99.88%	2	100.00%	1	98.55%	2

Source: Frontier Economics

Table 22. Efficiency scores and rankings for the national wage specification

DNO	All DNOs		SSEH excluded		LPN excluded		LPN and SSEH excluded	
	Efficiency score	Ranking	Efficiency score	Ranking	Efficiency score	Ranking	Efficiency score	Ranking
WMID	82.78%	12	84.13%	12	83.20%	12	83.47%	10
EMID	95.20%	4	96.48%	3	96.81%	5	95.31%	3
ENWL	89.15%	8	91.03%	8	89.57%	8	90.12%	7
NPGN	99.56%	2	98.00%	2	100.00%	1	96.40%	2
NPGY	98.87%	3	100.00%	1	98.59%	3	100.00%	1
SWales	100.00%	1	96.05%	5	98.90%	2	94.87%	5
SWest	90.42%	7	88.96%	9	90.67%	7	87.38%	9
LPN	87.22%	9	91.71%	7	-	-	-	-
SPN	85.48%	10	87.35%	10	84.19%	11	87.79%	8
EPN	82.15%	13	85.08%	11	84.76%	10	83.39%	11
SPD	94.46%	6	93.70%	6	97.18%	4	90.93%	6
SPMW	78.85%	14	76.68%	13	78.17%	13	76.19%	12
SSEH	82.87%	11	-	-	86.83%	9	-	-
SSES	94.66%	5	96.18%	4	96.62%	6	95.18%	4

Source: Frontier Economics

4.5 Conclusion

We have undertaken further analysis to allow a richer testing of the potential impact of density on cost. We found that replacing the Phase 1 density measure with measures based on the new, more detailed dataset offered little improvement in how density is accounted for in the model. Consequently, our preference is to use the basic density measure (customers per service area) as it is more transparent, easy to collect and more readily available than density measures based on detailed underlying density distributions for each DNO.

We have sought to test explicitly for an urbanisation effect, whereby costs start to rise as density increases above a certain level. We did not find that this effect was supported by the data. However, it should be noted that this could also be the result of the limited sample size and in particular the relative similarity of the majority of the DNOs to one another. It is clear from our analysis that SSEH and LPN are outliers in the dataset.

We therefore recommend that the final specifications should include our Phase 1 measure of density. We also conclude that there is a modest range of uncertainty around the efficiency scores of SSEH and LPN for any given treatment of density, and quantify this in section 4.4.10.

5 Accounting for quality of supply

Delivering electricity with fewer and shorter interruptions is costly for DNOs, and they consequently face a trade-off between quality and cost. This means that it is necessary to take account of quality of supply in our totex benchmarking to ensure that the model is not biased, and the estimates of DNOs' efficiency reflect the quality of supply delivered.

5.1 Phase 1

During Phase 1 we took account of quality of supply by adjusting the totex variable by the relative number of interruptions (CIs) and the average length of interruptions (CMLs). We calculated this adjustment by monetising the delivered quality of supply, using the DNOs' specific rates set by Ofgem in the IIS during DPCR4 and DPCR5 (our sample covers both regulatory periods), relative to a benchmark level of zero interruptions and minutes lost (i.e. DNOs were "charged" for every interruption and every minute lost). The estimated parameters in this adjusted totex model were similar to the unadjusted totex model, and the significance of the estimated parameters generally improved.

5.2 Phase 2 approach

In Phase 2 we have considered two issues:

- whether we include the actual number and length of interruptions in the set of drivers of the totex cost function; and
- whether the monetisation of the quality of supply should be against a zero benchmark or the benchmarks set by Ofgem.

As far as the first of these issues is concerned, the modelling approach would need to recognise that totex and quality of service can both be considered choice variables of the company - quality performance will be driven by the costs the DNO incurs to invest and maintain its system, and equivalently, costs will be driven by the realised level of quality performance. This necessitates the estimation of a two-equation system of regressions, one each for quality and totex, estimated simultaneously. We have rejected this approach because it is complex and "data hungry".

In contrast, the approach of monetizing the actual quality of supply delivered by DNOs adjusting the dependent totex variable accordingly reduces the number of variables included in the model (and preserves degrees of freedom), and also addresses the issue of endogeneity. Consequently we prefer the approach of modifying totex, which is simple, clear and captures the relevant cost-quality trade off.

During Phase 1 we converted the delivered level of performance using the DNOs' specific rates set by Ofgem relative to a benchmark level of zero interruptions or minutes lost. In response to this work the DNOs requested that we also considered making adjustment relative to the target levels embodied in the IIS scheme described above. We have therefore estimated two models:

- The first model consists of monetising all the actual CIs and CMLs, and treats this monetary value as additional costs to be added to the totex variable. In effect, this monetizes all outages against a benchmark of zero, and represents the Phase 1 approach.

- The second model reflects the current incentive mechanism, and consists of monetising the difference between the actual CIs and CMLs and the targets set by Ofgem for every DNO and year, where these targets are different for each DNO. When this difference is positive, implying that the specific DNO has underperformed the target, this results in additional costs to be added to the totex. Conversely, when the DNO outperforms the target, the monetary value of the difference between actual performance and the target results in a cost deduction to be applied to the totex. In this second model we have limited the adjustment of the totex by the revenue exposure faced by DNOs in the Interruption Incentive Scheme during DPCR4 and DPCR5.

5.2.1 Data

For both models, we have used the rates in Ofgem's Interruption Incentive Scheme to monetise the actual performance of DNOs. **Figure 16** shows the rates we have used to monetise the quality of service performance indicators.

Figure 16. Rates used to monetise the quality of supply delivered by DNOs during DPCR4 (first two tables) and DPCR5 (last table)

Table 4.5 CI incentive rates

Incentive rates for the number of customers interrupted per 100 customers (£m/CI – 02/03 prices)						
DNO	2005/6	2006/7	2007/8	2008/9	2009/10	2004/5 IIP incentive rate
CN - Midlands	0.10	0.11	0.11	0.11	0.11	0.06
CN - East Midlands	0.15	0.15	0.15	0.15	0.16	0.09
United Utilities	0.18	0.18	0.18	0.19	0.19	0.13
CE – NEDL	0.10	0.10	0.10	0.10	0.10	0.06
CE – YEDL	0.13	0.14	0.14	0.14	0.14	0.08
WPD - South West	0.10	0.10	0.10	0.10	0.11	0.07
WPD - South Wales	0.07	0.07	0.07	0.08	0.08	0.03
EDF – LPN	0.29	0.30	0.30	0.31	0.31	0.24
EDF – SPN	0.09	0.09	0.09	0.10	0.10	0.05
EDF – EPN	0.15	0.15	0.16	0.16	0.17	0.10
SP Distribution	0.23	0.23	0.23	0.23	0.23	0.13
SP Manweb	0.18	0.18	0.18	0.18	0.18	0.11
SSE - Hydro	0.08	0.08	0.08	0.09	0.09	0.04
SSE - Southern	0.18	0.18	0.18	0.19	0.19	0.11
Average	0.15	0.15	0.15	0.15	0.15	0.10

Table 4.6 CML incentive rates

Incentive rate for the number of customer minutes lost per customer (£m/CML)						
DNO	2005/6	2006/7	2007/8	2008/9	2009/10	2004/5 IIP incentive rate
CN - Midlands	0.14	0.15	0.15	0.16	0.17	0.10
CN - East Midlands	0.18	0.19	0.20	0.21	0.23	0.17
United Utilities	0.22	0.23	0.23	0.24	0.25	0.16
CE – NEDL	0.13	0.13	0.14	0.14	0.14	0.08
CE – YEDL	0.17	0.18	0.18	0.19	0.20	0.16
WPD - South West	0.17	0.17	0.17	0.18	0.18	0.13
WPD - South Wales	0.12	0.12	0.12	0.12	0.13	0.05
EDF – LPN	0.33	0.33	0.34	0.35	0.35	0.25
EDF – SPN	0.12	0.13	0.14	0.15	0.16	0.09
EDF – EPN	0.23	0.24	0.25	0.25	0.26	0.17
SP Distribution	0.27	0.28	0.30	0.33	0.35	0.14
SP Manweb	0.20	0.21	0.22	0.23	0.24	0.12
SSE – Hydro	0.10	0.11	0.11	0.11	0.11	0.04
SSE - Southern	0.24	0.25	0.26	0.27	0.28	0.15
Average	0.19	0.19	0.20	0.21	0.22	0.13

Table 16.5 - Annual CI and CML incentive rates for DPCR5

DNO	CI incentive rate £m	CML incentive rate £m
CN West	0.11	0.40
CN East	0.12	0.42
ENW	0.11	0.56
CE NEDL	0.07	0.26
CE YEDL	0.10	0.37
WPD S Wales	0.05	0.18
WPD S West	0.07	0.25
EDFE LPN	0.30	0.34
EDFE SPN	0.10	0.36
EDFE EPN	0.16	0.57
SP Distribution	0.09	0.33
SP Manweb	0.07	0.21
SSE Hydro	0.03	0.15
SSE Southern	0.13	0.47

Source: Ofgem, Final Proposal documents during DPCR4 and DPCR5

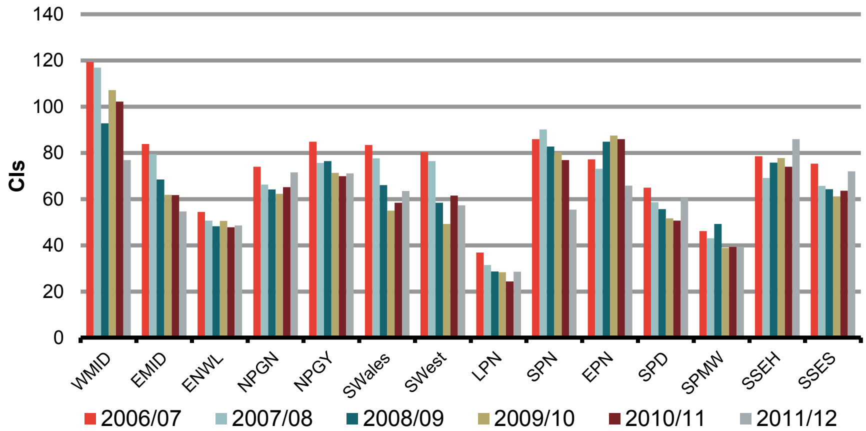
The indicators used to monetise the delivered quality of service are the CIs (number of customer interruptions by 100 customers) and CMLs (minutes of interruption per customer) reported by Ofgem and used in the IIS. We use the CIs and CMLs without storms and weighted by the type of interruption according to the IIS in each price control period.[50]

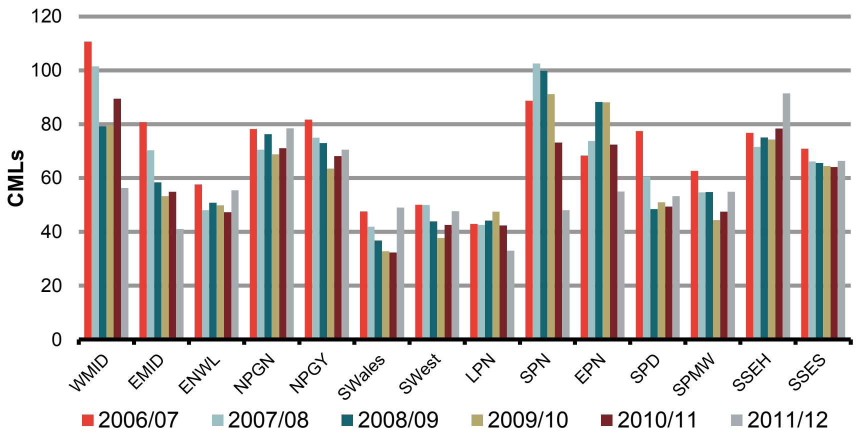
Figure 17 shows the underlying data used to monetise the quality of supply performance by DNOs.

---

50 Ofgem, Electricity Distribution Annual Report for 2010-11 and attached Excel file, March 2012. For the year 2011-12, interruptions performance data has been provided directly by Ofgem. Ofgem could only provide us with un-weighted data for year 2011-12.

Figure 17. Quality of supply indicators. CIs and CMLs (IIS weighted)

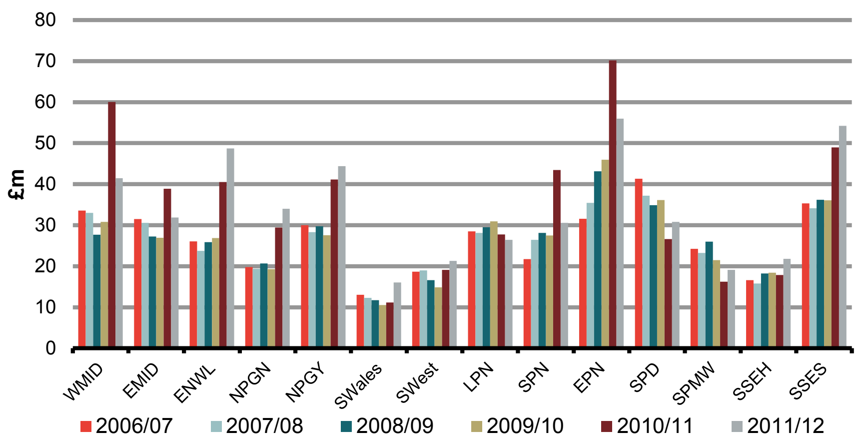


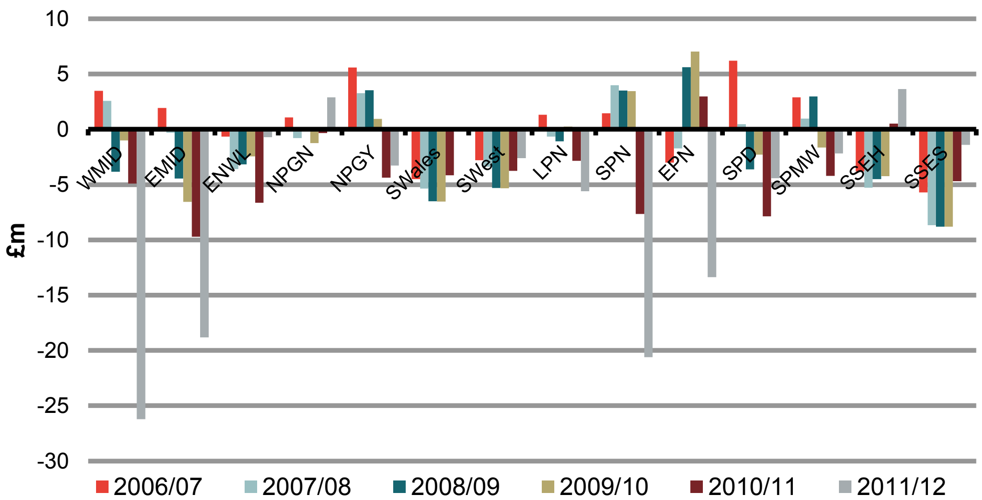


Source: Frontier Economics based on Ofgem data

Figure 18 shows the monetised values of the quality of supply delivered by DNOs under the two approaches. We have used these values to adjust the totex variable. Positive values increase the totex amount while negative values decrease the totex.

Figure 18. Monetised values derived from quality of supply performance





Source: Frontier Economics

5.3 Results and conclusions

We have evaluated the three alternatives of the model (totex unadjusted, totex adjusted using a common baseline of zero and totex adjusted using the targets set by Ofgem) using the Wald goodness of fit test, and reported in **Table 23**.

The results show that adjusting the totex variable by the delivered quality of supply results in a model with stronger econometric properties. This result is generally consistent across all the specifications we have considered. It is also clear that monetizing from a baseline of zero delivers the most statistically robust results.

Table 23. Scoring of alternatives to account for quality of supply performance using the statistic associated with the Wald goodness of fit test

	Specification 1 (regional wages)	Specification 2 (national wages)
Totex not adjusted	143	132
Totex adjusted (using Ofgem targets)	141	118
Totex adjusted (zero benchmark)	**161**	**152**

Source: Frontier Economics

The Appendix reproduces the regressions results, where it can be clearly seen that the adjusted totex (zero baseline) model delivers more significant coefficients.

For these reasons we recommend using the zero baseline measure as the adjusted totex variable in our preferred specifications.

6 Summary of key results

In this section we draw together the different elements of our review to present a summary of the results that have emerged from our Phase 2 research.

- We confirm the elements of our final model, including confirming those areas of research where we have concluded that no adjustment is necessary.
- We present final results for our preferred approach, including the estimated efficiency rankings that emerge.
- Finally, we compare and contrast results for the Random Effects specification with the results that are derived under a Pooled OLS approach, including a comparison of efficiency scores and rankings.

6.1 Confirmation of model coverage

Our research during Phase 2 has allowed us to draw firm conclusions of the elements that should be contained in the model and those that should not.

Our preferred model specification includes the following cost drivers.

- Core outputs.
  - Customer numbers.
  - Peak demand.
- Input prices[51].
  - Labour prices (we consider two variants, one in which regional prices are used, and a second in which national prices are used. For both variants we have used the series SIC_35 "electricity, gas, steam and air conditioning supply", which is available at both regional and national level.)
  - Capital prices (based on the BEAMA producer price index for Basic Electrical Equipment).
- Environment.

---

51 Our assessment of the results of the Phase 2 analysis suggests that we should not include a measure of general inflation in our specification. We find more econometrically robust results when general inflation is excluded and the model is estimated using two input prices (for labour and capital). Note that by imposing homogeneity of degree one in input prices, we are implicitly transforming all monetary variables in real prices.

- Density (measured as customer numbers divided by surface area).

We have also undertaken a thorough investigation of a wide range of additional topics. The results of this work are presented in detail in Volume 2 of this report. In each case we have concluded that no adjustment should be made to our preferred model. We provide a high level summary of our rationale for not including each element below.

- **Asset condition**: there is little data available and no econometric support for the inclusion of the variables that do exist. Additionally, there are concerns over the endogeneity of asset condition and which would make its inclusion in a totex model technically challenging, and could also give rise to confused regulatory incentives.

- **Investment cycle**: based on long run historic data, and expert technical assessment at DPCR4, we find no evidence to suggest any material difference in the investment cycle between companies. Consequently we see no requirement for controlling for this in the totex model.

- **Asset related outputs**: as would be expected, we find evidence to show that asset related outputs such as network length and MEAV are correlated with totex. Models that include network length and to an extent MEAV, essentially as a replacement for a density measure, are econometrically viable. Nevertheless, we consider these models inappropriate for use in a regulatory context. Re-specifying the model to depend on either network length or MEAV would eliminate the ability of the model to judge and provide incentives for optimal network design. It would also create perverse incentives to, at the margin, favour operational solutions that are asset heavy. We consider that there is strong evidence to show that our density measure captures well the relevant environmental effects and given that it is beyond the control of the companies, should be preferred for econometric and regulatory purposes.

- **Voltage structure**: we have found no evidence to suggest that voltage structure is a significant driver of cost for the GB DNOs. We anticipate that this result, which contrasts with findings of studies undertaken on DNOs operating in other countries, arises as a consequence of the scale of the GB DNOs and the resulting averaging of the customers served. Similarly find no evidence to suggest that the absence of 132 kV assets in Scotland should be accounted for in our totex model. It is likely that any residual effect that might be captured by a voltage structure variable is likely to be addressed by our density measure.

6.2 Results of our preferred model

As set out in Section 3 (Input prices) we have identified two candidate models that vary in just one respect, i.e. whether the variable that captures the price of labour should be regional or national. We reproduce the core regression results for the Random Effects specification in **Table 24** below.

Table 24. Comparison of Specification 1 (SIC35, regional) and Specification 2 (SIC-35, national); Random Effects

	SIC 35, Regional	SIC 35, National
Customers	0.469***	0.585***
Peak	0.351***	0.239*
Density	-0.078***	-0.056*
Wages	0.326***	0.542***
Price of capital[52] (BEAMA)	0.674	0.458
Constant	-8.21***	-8.64***
R² [53]	0.887***	0.875***

The table reports the estimated coefficient for each variable and the confidence intervals using a 95% probability.[54]

*** Significant at 1% ** Significant at 5% *Significant at 10%

Source: Frontier Economics

A comparison between the two models reveals the following features:

- The sum of coefficients on peak and customers are very similar, equal to 0.812 in the regional wage model and 0.817 in the national specification.
- The estimates of budget shares for both labour and capital broadly are plausible and not statistically different from each other, once confidence intervals are taken into account.

---

52 Due to the imposition of homogeneity of degree +1 in input prices, we can infer the coefficient for the capital price as 1-coefficient on wages.

53 See footnote 19.

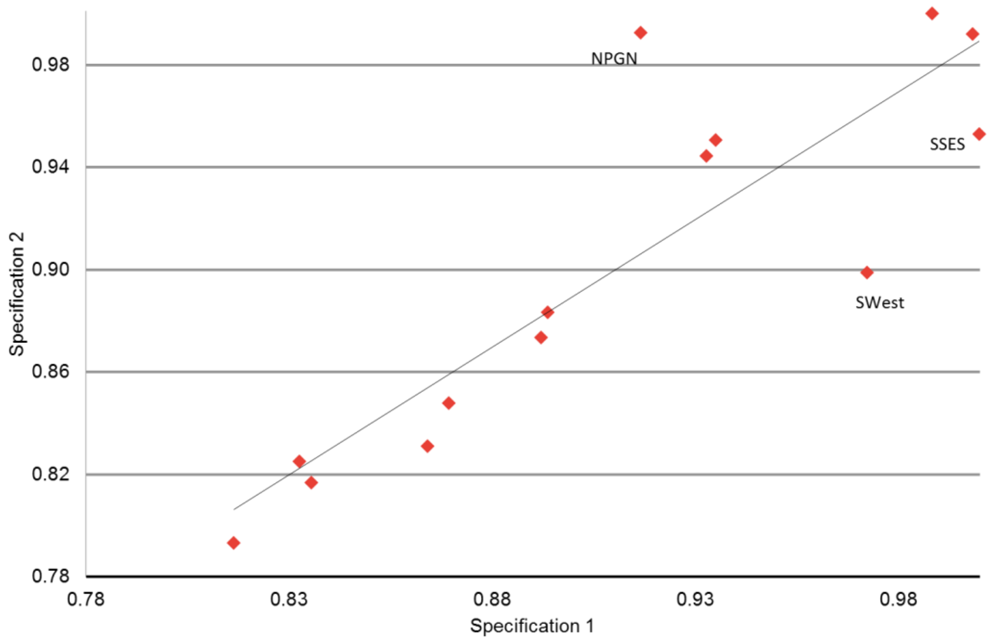
54 See footnote 20.

- The R^2 coefficient is slightly higher for the regional wage specification. However, differences are small and we do not consider that there is enough evidence to suggest that that one specification performs better than the other.

We consider that both models are econometrically sound, have robust incentive properties and that both have a reasonable economic interpretation. However, we note that the two specifications are to a degree supported by competing underlying hypotheses over the nature of the labour markets from which DNOs source input, i.e. whether they are more national or more regional in nature. On balance, we do not consider that it is helpful to reject one of these models at this stage and we recommend that Ofgem considers carefully the results of both models.

As we show in **Figure 19**, the estimates of efficiency that are derived from these two specifications are similar for the majority of the DNOs, e.g. the correlation between them is 88%. However, there are certain DNOs for which the choice of labour input price gives rise to sizeable changes in estimated efficiency.

Figure 19. Scatter of efficiency scores under Specification 1 and Specification 2



Source: Frontier Economics

6.3 Comparison of efficiency scores under RE and POLS

It is helpful to consider the extent to which our assumption on the structure of the data (i.e. the assumptions that underpin the Random Effects specification as described in **Section 2** (Methodology) might give rise to different results from those that would be derived from the application of Pooled OLS. We present an analysis of this in **Table 25** for specification 1 and **Table 26** for specification 2.

Table 25. Efficiency estimates for Specification 1 under RE and POLS

DNO	RE	POLS (average sample period)	POLS (Last year estimate)
WMID	0.840	0.809	0.805
EMID	0.947	0.933	0.939
ENWL	0.900	0.869	0.683
NPgN	0.938	0.909	0.893
NPgY	1.000	0.988	0.928
SWales	0.996	0.98	0.879
SWest	0.967	0.947	0.852
LPN	0.896	0.877	0.996
SPN	0.874	0.847	0.809
EPN	0.842	0.816	0.854
SPD	0.941	0.923	0.881
SPMW	0.820	0.798	0.736
SSEH	0.865	0.839	0.767
SSES	0.996	1.000	1.000

Source: Frontier Economics

Table 26. Efficiency estimates for Specification 2 under RE and POLS

DNO	RE	POLS (average sample period)	POLS (Last year estimate)
WMID	0.828	0.816	0.809
EMID	0.952	0.965	1.000
ENWL	0.891	0.866	0.701
NPgN	0.996	0.991	0.990
NPgY	0.989	0.996	0.978
SWales	1.000	1.000	0.935
SWest	0.904	0.887	0.834
LPN	0.872	0.877	0.994
SPN	0.855	0.837	0.810
EPN	0.822	0.816	0.854
SPD	0.945	0.947	0.924
SPMW	0.789	0.799	0.760
SSEH	0.829	0.825	0.771
SSES	0.947	0.977	0.987

Source: Frontier Economics

Efficiency scores and rankings under both models, Random Effects and Pooled OLS over the whole sample period, are very similar. Efficiency scores under Pooled OLS are significantly different from Random Effects when the first are calculated using a single year in the sample.

Annexe 1: Density data

This annexe describes the density data sources, cleaning, matching and quality control.

Data sources

The following provides more details on the measurements and sources used in this analysis. Please read MSOA as MSOA and IGZ if not specified otherwise.

Meters and demand per MSOA

- **Source 1 –** The Department of Energy and Climate Change (DECC) estimates the number of meters (domestic, Economy 7 and non-domestic meters) and energy consumption per MSOA. We use estimates from 2010. The meter variable is a direct measurement of our variable of interest, and therefore we prefer it over indirect approximations like population and the number of households. We also use the energy consumption variable to construct a comparator density measure. The dataset includes the local authority which each MSOA corresponds to. This is our primary way to match MSOAs to the service area of DNOs.

Surface area per MSOA

- **Source 2 –** The ONS reports the surface area in hectares, population and population density per MSOA in England and Wales. We use the surface area to calculate meter density. The population figures are used to check plausibility.

- **Source 3 –** We were unable to find direct information about the surface area of the Scottish IGZs. However, Scottish Neighbourhood Statistics (SNS) report the dwelling density and the number of dwellings in each IGZ. From this we calculate the surface area in hectares.

- **Source 4 –** UKPN provided a list of areas served by each DNO.

Supporting data and cross checks

- **Source 5 –** The number of customers per DNO, as reported in Ofgem's annual report, is used to cross-check the bottom up approach we take by aggregating all meters per MSOA.

For some MSOAs, information on the local authority was not sufficient to assign the area to a DNO. Therefore we use sources that provide town names of MSOAs. An appendix from the NHS is used to provide more details on the

Cheshire area and an appendix from the Association of Research Observatories is used to provide more details on the Shropshire area.

Data cleaning

Some meter and energy consumption statistics are not assigned to a particular MSOA. There were three different entries we found instead of one MSOA number:

- **"Unallocated"** meter and energy consumption statistics are assigned to a local authority, but could not be assigned to a specific MSOA.
- Some statistics cover **multiple areas**, making it unclear how to assign values to the individual areas.
- Sometimes we also had meter and consumption data classified as **"Half hourly consumption"** instead of an MSOA. We assume these data entries refer to large commercial customers.

In all of the three cases, we could not assign the meter and energy consumption data to an MSOA, so they were excluded from the analysis at the MSOA level.

Some MSOAs are in the list as multiple entries. We added these together after checking that the entries are part of the same local authority.

The Scottish dwelling density and numbers of dwellings per IGZ is given for several years. We calculate the density per year and take the average. There are small rounding differences between the years that are mitigated by averaging.

Matching

In most cases, the name of the area provided by UKPN exactly matched the name of the local authority in the DECC data. We assigned all MSOAs to a DNO that served a matching local authority.

However, not all matches were perfect, so we adapted some entries, mostly by changing minor details.[55]

In other cases the DECC description did not easily match the UKPN data. For a small number of MSOAs we changed the local authority name to match the area name in the UKPN dataset.

Table 27 below summarises the number of meters, total demand, the total surface area and the number of sub-areas for each DNO following data cleaning.

55 For example, by changing symbols: "Dumfries & Galloway" becomes "Dumfries and Galloway".

Table 27. Summary of density data following cleaning

DNO	Number of meters	Total electricity demand (MWh)	Surface area (hectares)	Number of sub-areas (MSOAs or IGZs)
EMID	2,657,280	14,590,041	1,505,462	733
ENW	2,330,772	12,730,189	1,218,227	634
EPN	3,495,300	20,428,177	1,970,873	953
LPN	2,275,634	12,966,410	71,671	565
NEDL	1,629,585	8,355,890	1,628,335	433
SP	2,021,408	10,655,845	2,151,348	886
SP Manweb	1,496,512	8,305,451	1,134,206	419
SPN	2,189,859	12,560,989	783,457	596
SSE Hydro	892,915	5,616,414	6,031,039	348
SSE Southern	3,069,695	18,225,005	1,687,124	837
WMID	2,389,505	13,221,912	1,294,536	690
WPD SWales	1,098,798	5,528,904	1,279,383	306
WPD SWest	1,506,316	8,820,755	1,467,103	395
YEDL	2,247,501	11,567,771	1,061,332	632

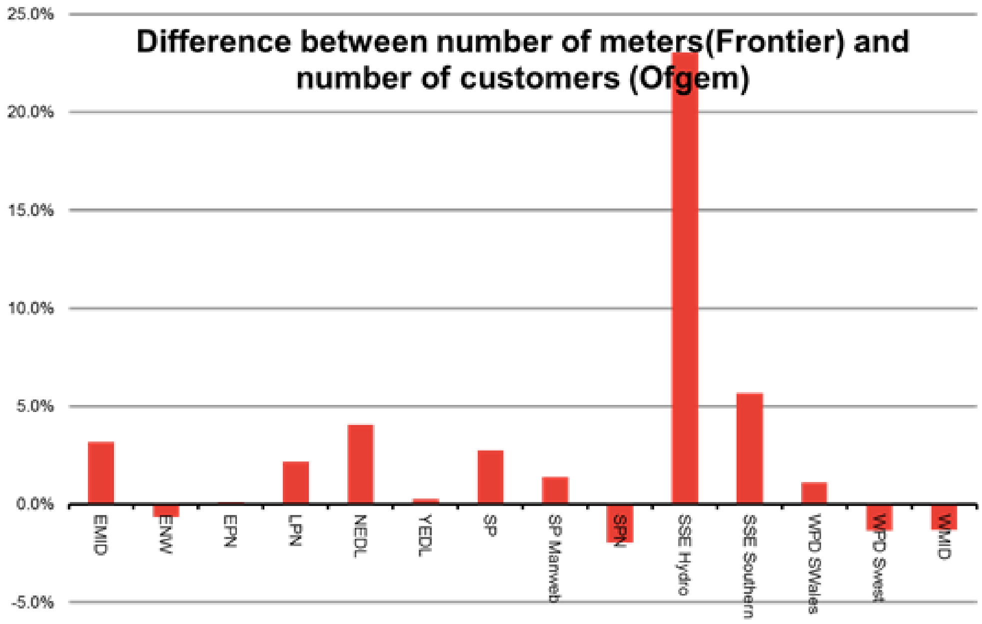
Sources: DECC, ONS, SNS, UKPN, Ofgem

Quality

We performed two quality checks. First, we compared the total number of meters from our bottom-up approach to the number of customers reported by Ofgem. On aggregate we have 2% more meters than customers in our dataset, which we consider to reveal a reasonable degree of consistency.

Figure 1 shows the deviation between sources per DNO. For most DNOs the deviation is within a 3% margin. The deviations for NEDL and SSE Southern are 4% and 6% respectively. For Scottish Hydro the two numbers deviate by 23%. This result indicates that our aggregation and matching are fairly robust, with the exception of Scottish Hydro. We speculate that this is likely to arise as a consequence of the challenges in taking account of the Scottish islands, but have been unable to confirm that this is the case.

Figure 20. Comparison of meter numbers from Frontier's analysis with customer numbers provided by Ofgem



Source: Frontier Economics

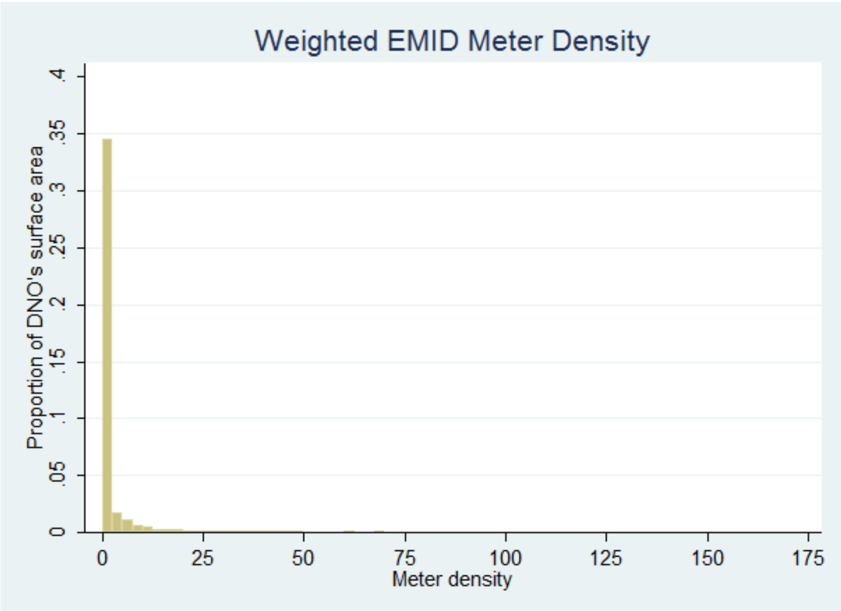
As a second quality check, we checked if meter density has a strong correlation with population density. As this correlation is 95%, we are confident that our data are reasonably robust.

Annexe 2: Meter density histograms

This annexe contains histograms of meter density for each DNO in 2010. The histograms use the same scale for each DNO for comparability. All density observations are weighted by the surface area of the corresponding MSOA (or IGZ) to correct for the different sizes of the sub-areas.

These histograms report the proportion of surface area (in hectares) for each DNO that falls in each of the meter density classes considered, with of width 2.5 meters per hectare. The height of the bars sums to one for each histogram.

Figure 21. Histogram of meter density; EMID



Source: Frontier Economics

Figure 22. Histogram of meter density; ENW

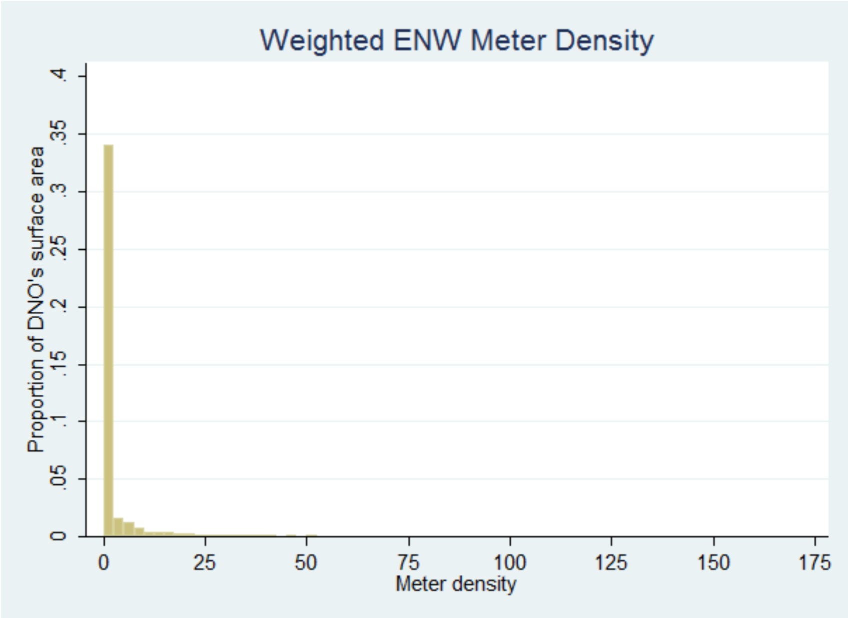


Figure 23. Histogram of meter density; EPN

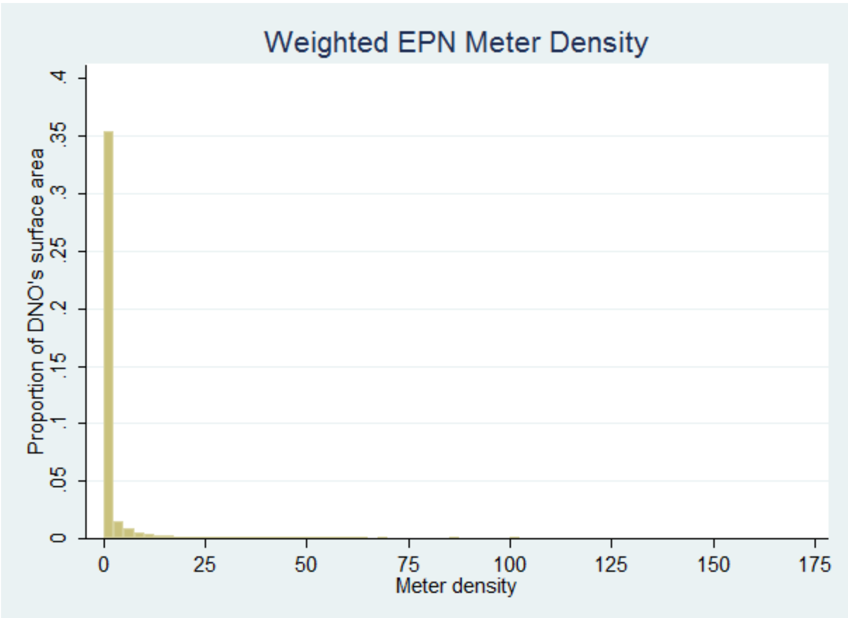


Figure 24. Histogram of meter density; LPN

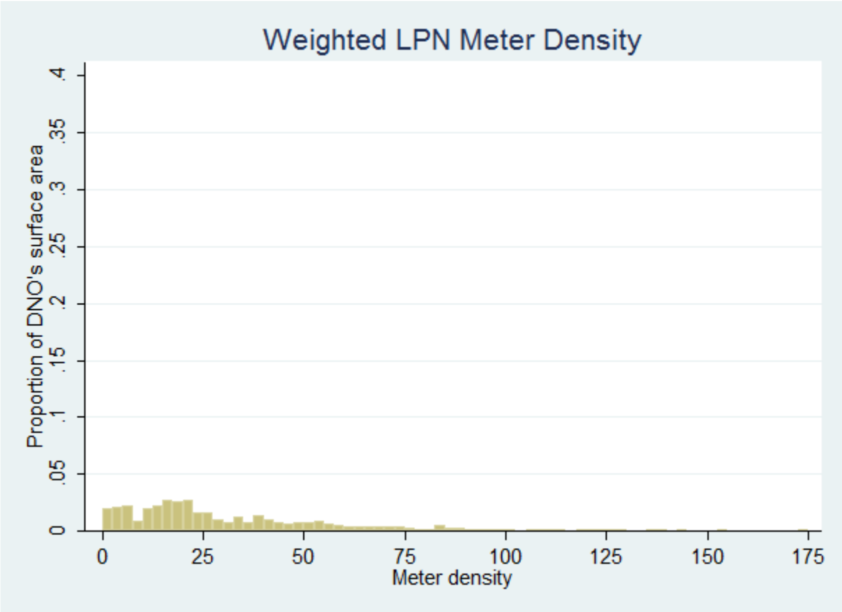
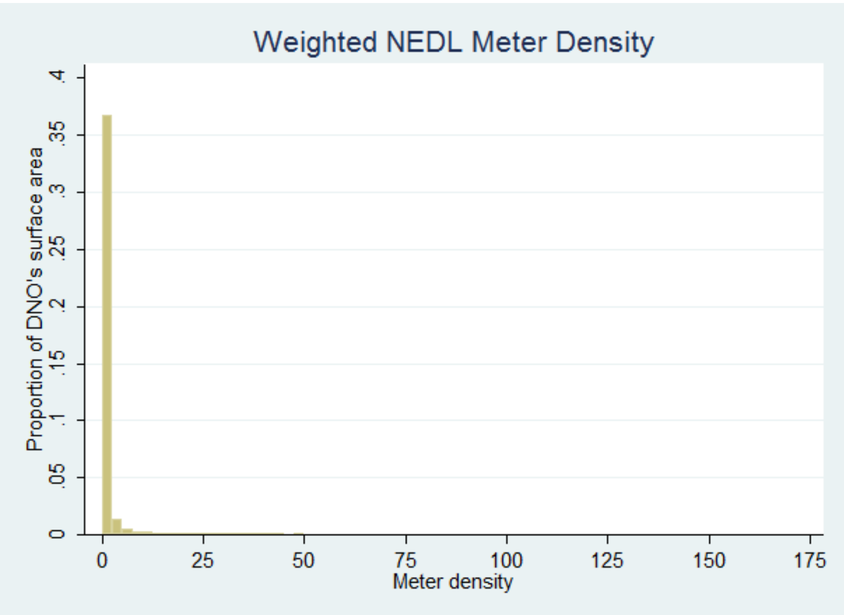


Figure 25. Histogram of meter density; NEDL



Source: Frontier Economics

Figure 26. Histogram of meter density; SP

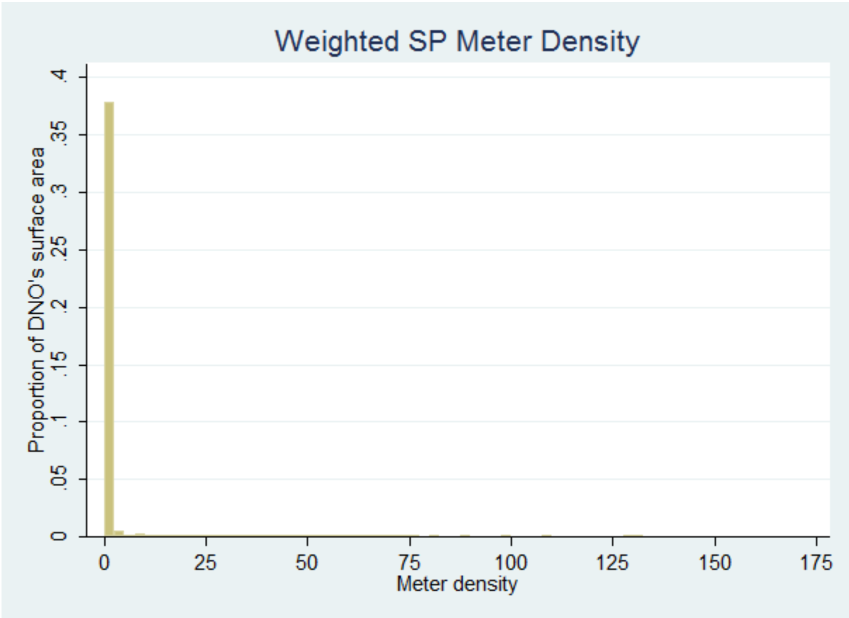


Figure 27. Histogram of meter density; SP Manweb

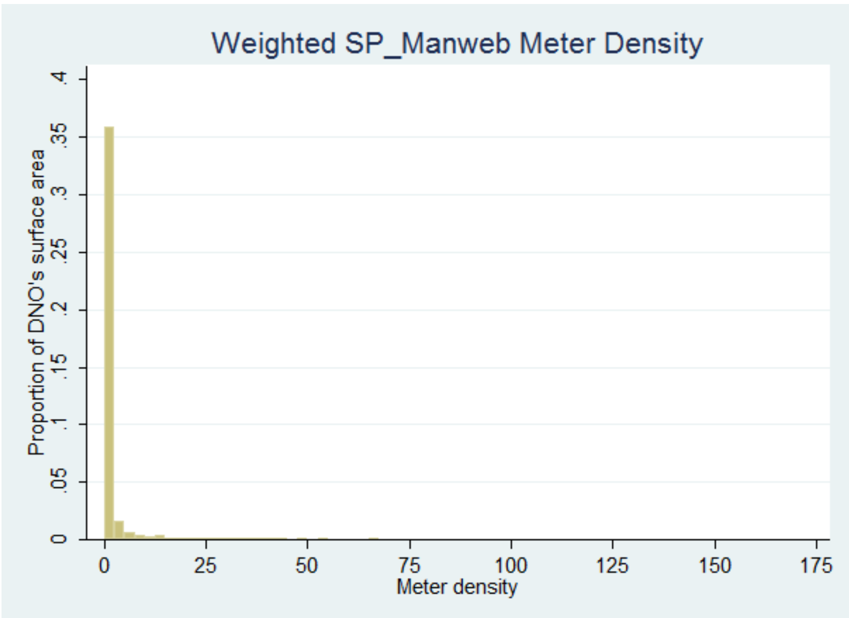


Figure 28. Histogram of meter density; SPN

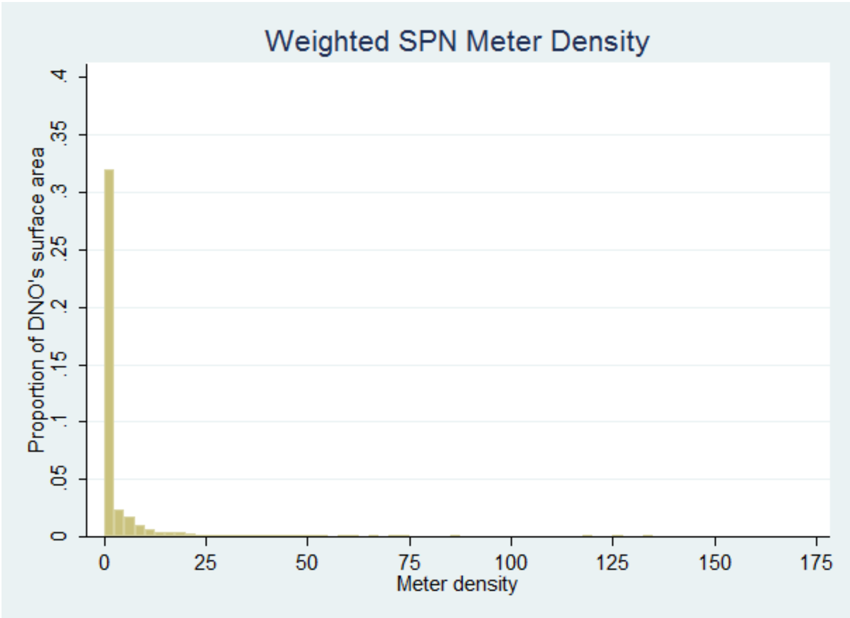


Figure 29. Histogram of meter density; SSE Hydro

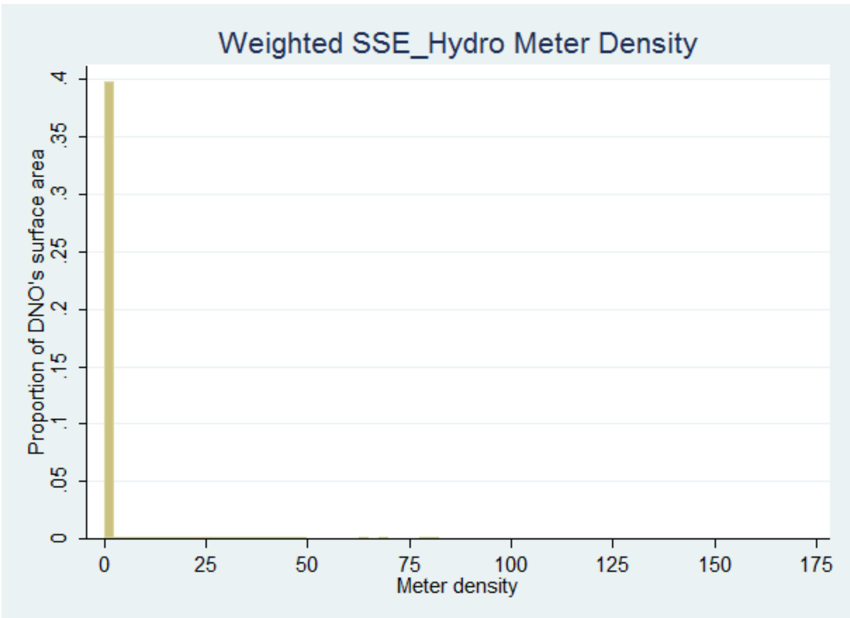


Figure 30. Histogram of meter density; SSE Southern

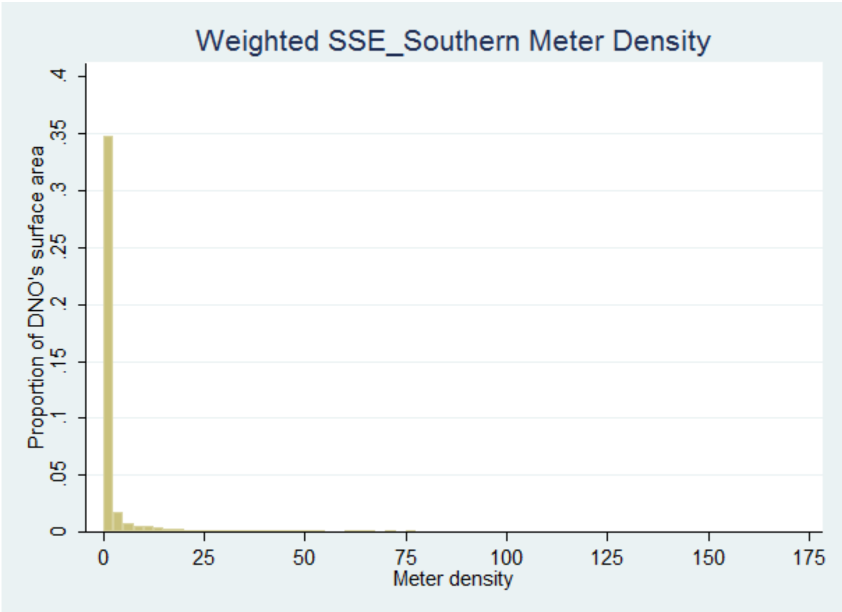


Figure 31. Histogram of meter density; WMID

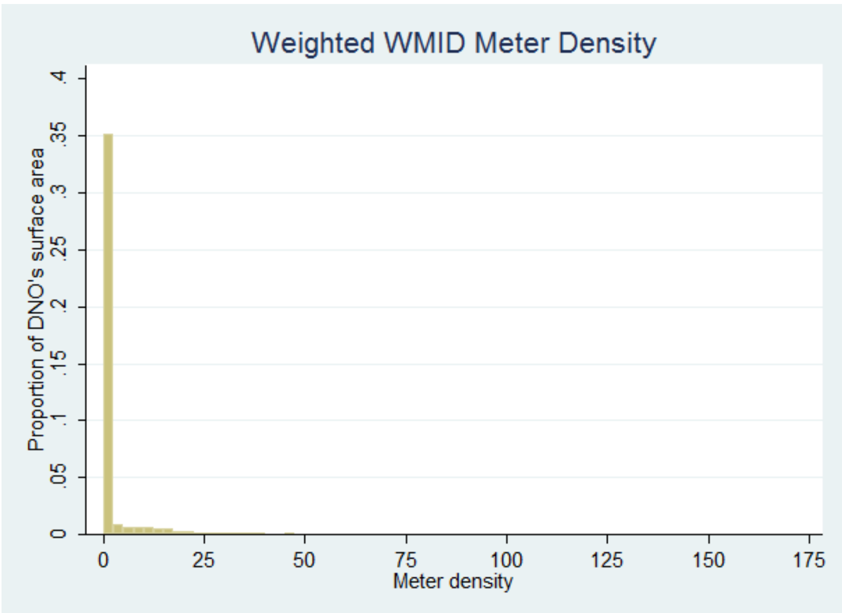


Figure 32. Histogram of meter density; WPD SWales

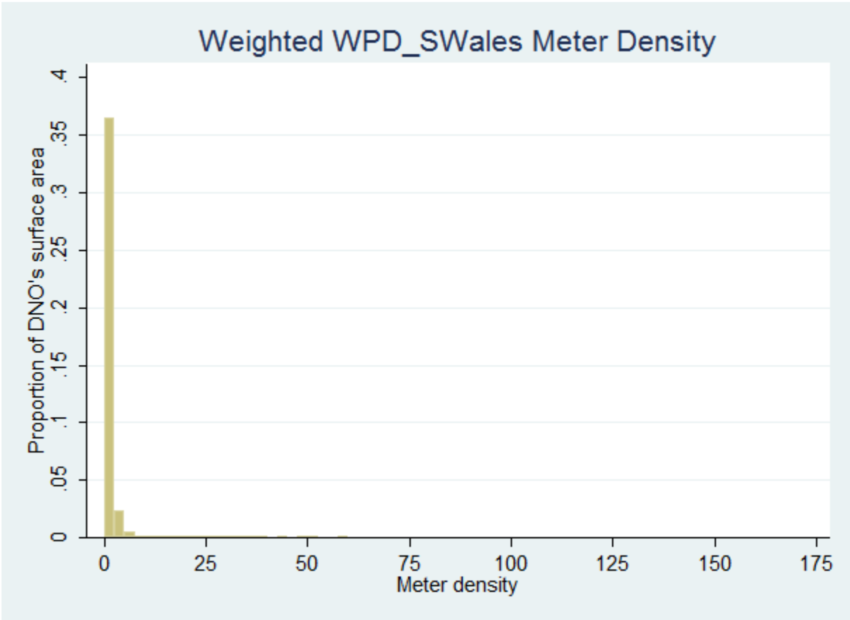


Figure 33. Histogram of meter density; WPD SWest

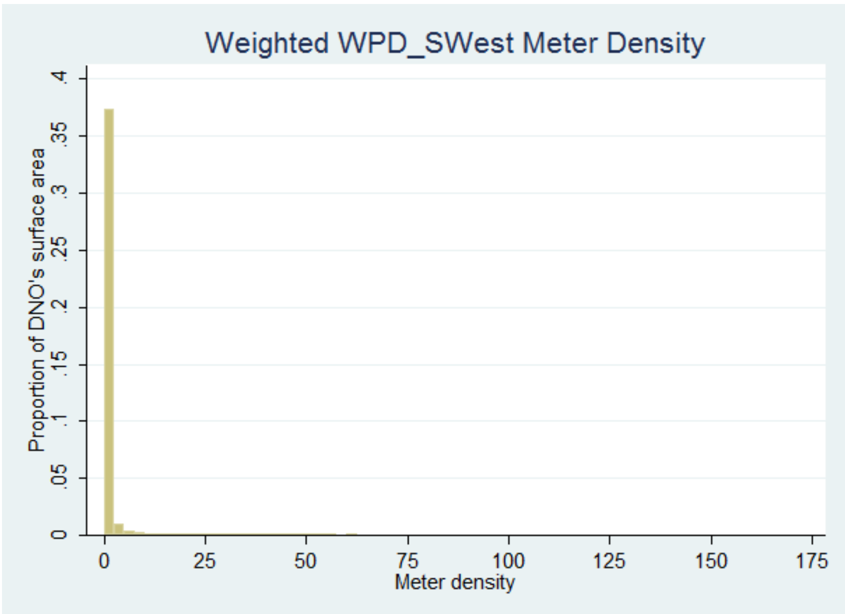
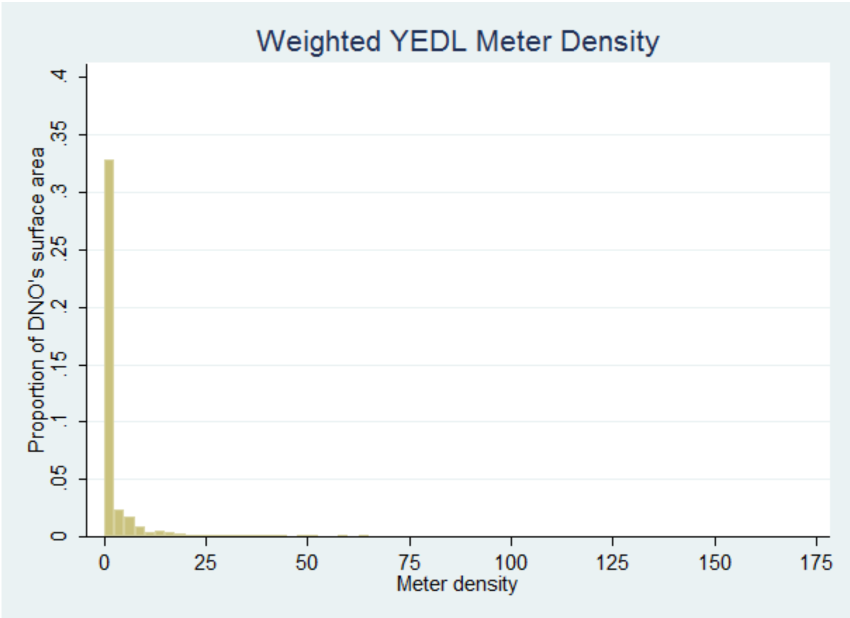


Figure 34. Histogram of meter density; YEDL



Annexe 3: Testing different density measures

Figure 35. Regression replacing basic density with mean meter density; Specification 1

```
Random-effects GLS regression                   Number of obs      =        84
Group variable: num_id                          Number of groups   =        14

R-sq:  within  = 0.1627                         Obs per group: min =         6
       between = 0.9374                                        avg =       6.0
       overall = 0.8844                                        max =         6

                                                Wald chi2(4)       =    154.16
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
l~t_capital1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 l_customers |   .4665153   .1491693     3.13   0.002     .1741489    .7588818
      l_peak |   .3542143    .137172     2.58   0.010     .0853621    .6230664
  l_den_mean |  -.0779753   .0307697    -2.53   0.011    -.1382828   -.0176677
l_wage_reg~1 |   .3298067   .1015054     3.25   0.001     .1308597    .5287536
       _cons |  -8.540449   1.324366    -6.45   0.000    -11.13616   -5.944739
-------------+----------------------------------------------------------------
     sigma_u |  .08815598
     sigma_e |  .08438377
         rho |  .52185241   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Source: Frontier Economics

Figure 36. Regression replacing basic density with mean meter density; Specification 2

```
Random-effects GLS regression                   Number of obs      =        84
Group variable: num_id                          Number of groups   =        14

R-sq:  within  = 0.2357                         Obs per group: min =         6
       between = 0.9193                                        avg =       6.0
       overall = 0.8727                                        max =         6

                                                Wald chi2(4)       =    146.63
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
l~t_capital1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 l_customers |   .5845967   .1576007     3.71   0.000     .2757051    .8934884
      l_peak |   .2375895   .1436486     1.65   0.098    -.0439566    .5191357
  l_den_mean |   -.055191     .03186    -1.73   0.083    -.1176355    .0072534
l~1_capital1 |   .5475041   .1403404     3.90   0.000      .272442    .8225663
       _cons |  -8.867573   1.370444    -6.47   0.000    -11.55359   -6.181553
-------------+----------------------------------------------------------------
     sigma_u |  .09347219
     sigma_e |  .08149205
         rho |  .56815242   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Source: Frontier Economics

Figure 37. Regression replacing basic density with standard deviation of meter density; Specification 1

```
Random-effects GLS regression                   Number of obs      =        84
Group variable: num_id                          Number of groups   =        14

R-sq:  within  = 0.1631                         Obs per group: min =         6
       between = 0.9266                                        avg =       6.0
       overall = 0.8745                                        max =         6

                                                Wald chi2(4)       =    130.88
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
l~t_capital1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 l_customers |   .4548464   .1546044     2.94   0.003     .1518275    .7578654
      l_peak |   .3409358   .1401237     2.43   0.015     .0662984    .6155731
    l_den_sd |  -.1152441    .056787    -2.03   0.042    -.2265446   -.0039435
l_wage_reg~1 |   .3264336    .102802     3.18   0.001     .1249455    .5279218
       _cons |  -8.137174    1.37078    -5.94   0.000    -10.82385   -5.450495
-------------+----------------------------------------------------------------
     sigma_u |  .09731401
     sigma_e |  .08438377
         rho |  .57080487   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Source: Frontier Economics

Figure 38. Regression replacing basic density with standard deviation of meter density; Specification 2

```
Random-effects GLS regression                   Number of obs      =        84
Group variable: num_id                          Number of groups   =        14

R-sq:  within  = 0.2369                         Obs per group: min =         6
       between = 0.9111                                        avg =       6.0
       overall = 0.8652                                        max =         6

                                                Wald chi2(4)       =    132.51
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
l~t_capital1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 l_customers |   .5793504   .1611652     3.59   0.000     .2634725    .8952283
      l_peak |   .2205223   .1455176     1.52   0.130    -.0646869    .5057315
    l_den_sd |  -.0761587   .0572053    -1.33   0.183    -.1882791    .0359617
l~1_capital1 |   .5527625   .1404901     3.93   0.000     .2774071     .828118
       _cons |   -8.55523   1.386253    -6.17   0.000    -11.27224   -5.838225
-------------+----------------------------------------------------------------
     sigma_u |  .09976663
     sigma_e |  .08149205
         rho |  .59980595   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Source: Frontier Economics

Figure 39. Regression replacing basic density with skewness of meter density; Specification 1

```
Random-effects GLS regression                   Number of obs      =        84
Group variable: num_id                          Number of groups   =        14

R-sq:  within  = 0.1616                         Obs per group: min =         6
       between = 0.9326                                        avg =       6.0
       overall = 0.8800                                        max =         6

                                                Wald chi2(4)       =    142.88
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
l~t_capital1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 l_customers |   .4727479   .1527471     3.09   0.002      .173369    .7721267
      l_peak |   .3516646   .1388753     2.53   0.011     .0794741    .6238551
  l_den_skew |   .1331784   .0580666     2.29   0.022     .0193699    .2469868
l_wage_reg~1 |   .3135907   .1014111     3.09   0.002     .1148285    .5123529
       _cons |  -8.905014   1.456844    -6.11   0.000    -11.76038   -6.049651
-------------+----------------------------------------------------------------
     sigma_u |  .09225144
     sigma_e |  .08438377
         rho |   .5444537   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Source: Frontier Economics

Figure 40. Regression replacing basic density with skewness of meter density; Specification 2

```
Random-effects GLS regression                   Number of obs      =        84
Group variable: num_id                          Number of groups   =        14

R-sq:  within  = 0.2355                         Obs per group: min =         6
       between = 0.9216                                        avg =       6.0
       overall = 0.8749                                        max =         6

                                                Wald chi2(4)       =    150.56
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000
```

l~t_capital1	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
l_customers	.5961485	.157546	3.78	0.000	.2873641	.904933
l_peak	.239636	.1428109	1.68	0.093	-.0402682	.5195403
l_den_skew	.1060046	.05751	1.84	0.065	-.0067129	.2187222
l~l_capital1	.5479109	.1401905	3.91	0.000	.2731425	.8226792
_cons	-9.257514	1.445038	-6.41	0.000	-12.08974	-6.425292
sigma_u	.09196538					
sigma_e	.08149205					
rho	.56016051	(fraction of variance due to u_i)				

Source: Frontier Economics

Figure 41. Regression replacing basic density with kurtosis of meter density; Specification 1

```
Random-effects GLS regression                   Number of obs      =        84
Group variable: num_id                          Number of groups   =        14

R-sq:  within  = 0.1614                         Obs per group: min =         6
       between = 0.9271                                        avg =       6.0
       overall = 0.8749                                        max =         6

                                                Wald chi2(4)       =    132.52
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000
```

l~t_capital1	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
l_customers	.4821434	.1573314	3.06	0.002	.1737795	.7905073
l_peak	.3400818	.1398819	2.43	0.015	.0659184	.6142452
l_den_kurt	.068703	.0336249	2.04	0.041	.0027994	.1346065
l_wage_reg~1	.3103528	.1019346	3.04	0.002	.1105647	.5101409
_cons	-8.988483	1.562307	-5.75	0.000	-12.05055	-5.926417
sigma_u	.09643619					
sigma_e	.08438377					
rho	.5663594	(fraction of variance due to u_i)				

Source: Frontier Economics

Figure 42. Regression replacing basic density with kurtosis of meter density; Specification 2

```
Random-effects GLS regression                   Number of obs      =        84
Group variable: num_id                          Number of groups   =        14

R-sq:  within  = 0.2362                         Obs per group: min =         6
       between = 0.9195                                        avg =       6.0
       overall = 0.8730                                        max =         6

                                                Wald chi2(4)       =    145.22
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000
```

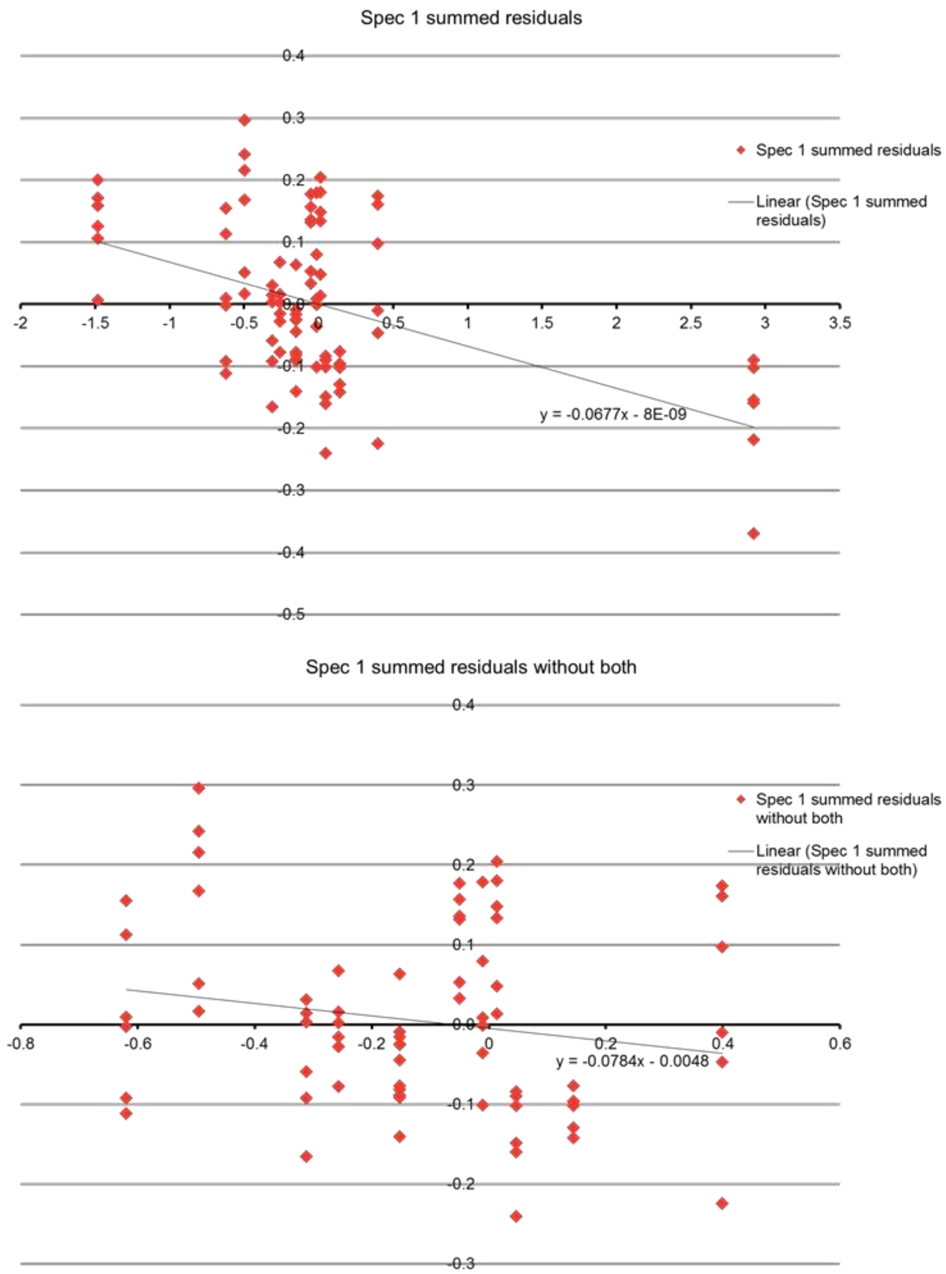
l~t_capital1	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
l_customers	.6109142	.1606583	3.80	0.000	.2960297	.9257987
l_peak	.2278904	.1427172	1.60	0.110	-.0518303	.507611
l_den_kurt	.0571128	.0327325	1.74	0.081	-.0070417	.1212674
l~1_capital1	.5532747	.1400319	3.95	0.000	.2788172	.8277323
_cons	-9.402293	1.526373	-6.16	0.000	-12.39393	-6.410658
sigma_u	.09430737					
sigma_e	.08149205					
rho	.57251212	(fraction of variance due to u_i)				

Source: Frontier Economics

Annexe 4: Density outlier analysis

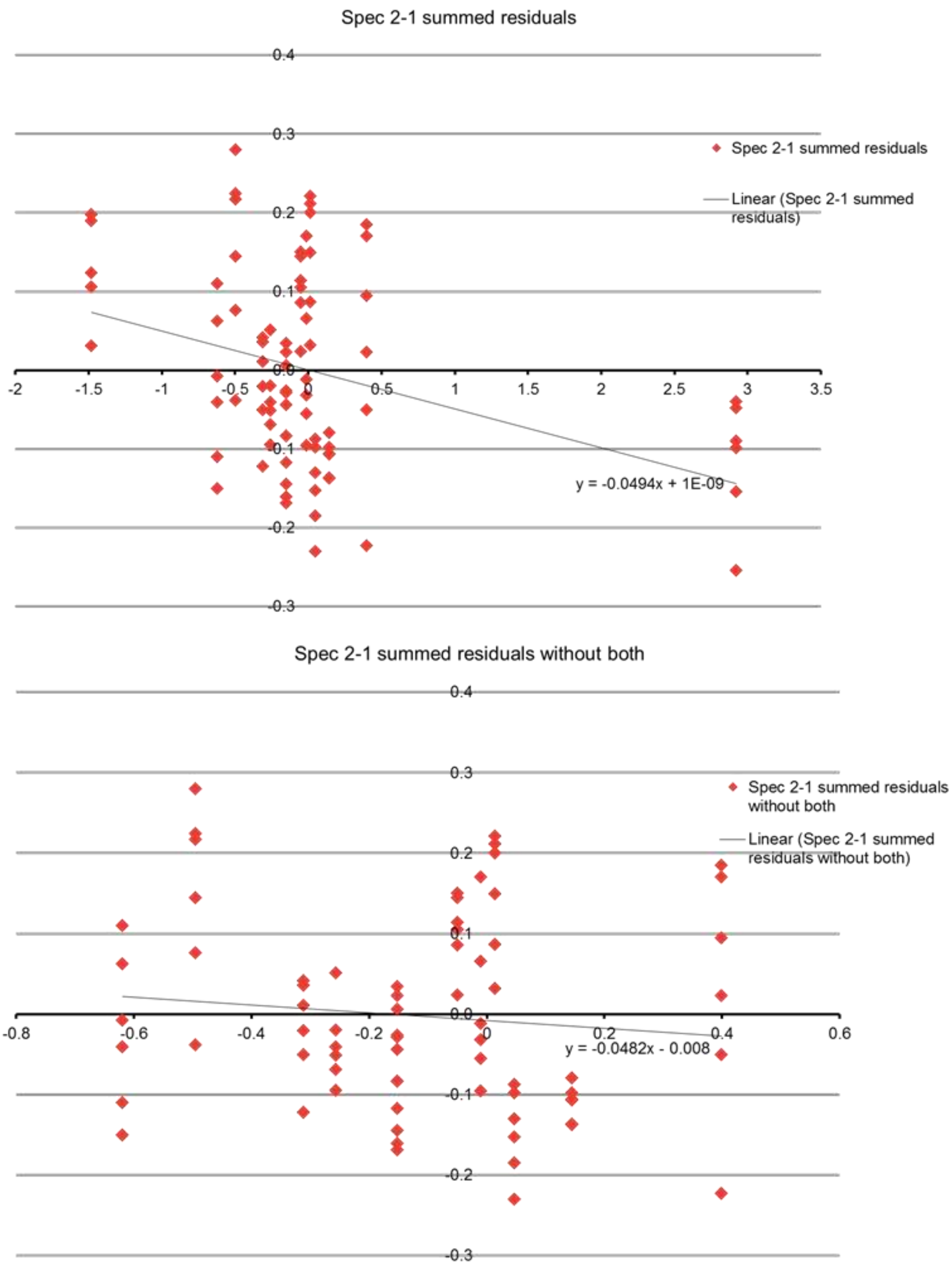
We first looked at the impact of outliers by running the totex regression excluding density, and separately regressing density on the other explanatory variables in the model (customers, peak and the regional or national wage). This allows us to look at the relationship between density and total costs once the other explanatory variables have been controlled for. We plotted the total residual term (both the idiosyncratic and time/DNO varying components) for density against the summed residual for total costs. The scatter plots are presented in **Figure 43** and **Figure 44** below, showing the line of best fit for all the residuals, and the line of best fit excluding the LPN and SSEH residuals.

Figure 43. Scatter plots of the density and total cost residuals; Specification 1



Source: Frontier Economics

Figure 44. Scatter plots of the density and total cost residuals; Specification 2



Source: Frontier Economics

Comparing the gradients of a linear line of best fit for the residuals indicates the direction of the impact the two density outliers have on the density coefficient in the model. A comparison of these gradients shows that, overall, LPN and SSEH result in a more steeply sloped relationship between density and costs than would be estimated if both the outliers were excluded. For the national wage specification, taking each of the two outliers out of the residual plots in turn (and leaving the other outlier in) shows that both result in a more steep gradient on the density residuals. For the regional wage specification, the same comparison showed that excluding LPN makes the density gradient steeper than otherwise, while excluding SSEH makes the density gradient marginally less steep.

Annexe 5: Regression results excluding outliers

Figure 45. Regression excluding SSEH; Specification 1

```
Random-effects GLS regression                   Number of obs      =        78
Group variable: num_id                          Number of groups   =        13

R-sq:  within  = 0.1566                         Obs per group: min =         6
       between = 0.9280                                        avg =       6.0
       overall = 0.8625                                        max =         6

                                                Wald chi2(4)       =    118.89
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
l~t_capital1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 l_customers |   .5526899   .1613427     3.43   0.001      .236464    .8689158
      l_peak |   .3108114   .1401661     2.22   0.027     .0360909    .5855318
   l_density |  -.0644905   .0314825    -2.05   0.041    -.1261951   -.0027858
l_wage_reg~1 |   .3302548   .1041859     3.17   0.002     .1260543    .5344554
       _cons |  -9.149028   1.491395    -6.13   0.000    -12.07211   -6.225947
-------------+----------------------------------------------------------------
     sigma_u |  .08698713
     sigma_e |  .08574492
         rho |   .5071912   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Source: Frontier Economics

Figure 46. Regression excluding LPN; Specification 1

Random-effects GLS regression		Number of obs	= 78
Group variable: num_id		Number of groups	= 13
R-sq: within	= 0.1500	Obs per group: min	= 6
between	= 0.9409	avg	= 6.0
overall	= 0.8932	max	= 6
		Wald chi2(4)	= 145.56
corr(u_i, X)	= 0 (assumed)	Prob > chi2	= 0.0000

l~t_capital1	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
l_customers	.5460198	.1858286	2.94	0.003	.1818024	.9102372
l_peak	.3253916	.1394159	2.33	0.020	.0521414	.5986418
l_density	-.1110918	.0669495	-1.66	0.097	-.2423103	.0201268
l_wage_reg~1	.2918486	.1019023	2.86	0.004	.0921238	.4915735
_cons	-9.012302	1.677289	-5.37	0.000	-12.29973	-5.724877
sigma_u	.09191151					
sigma_e	.08342296					
rho	.54830024	(fraction of variance due to u_i)				

Source: Frontier Economics

Figure 47. Regression excluding SSEH and LPN; Specification 1

Random-effects GLS regression		Number of obs	= 72
Group variable: num_id		Number of groups	= 12
R-sq: within	= 0.1407	Obs per group: min	= 6
between	= 0.9289	avg	= 6.0
overall	= 0.8700	max	= 6
		Wald chi2(4)	= 107.92
corr(u_i, X)	= 0 (assumed)	Prob > chi2	= 0.0000

l~t_capital1	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
l_customers	.5383634	.193143	2.79	0.005	.1598101	.9169167
l_peak	.307313	.1435725	2.14	0.032	.0259161	.5887099
l_density	-.0301063	.1122284	-0.27	0.788	-.2500699	.1898573
l_wage_reg~1	.2889491	.1059807	2.73	0.006	.0812308	.4966674
_cons	-9.164127	1.695705	-5.40	0.000	-12.48765	-5.840607
sigma_u	.09256291					
sigma_e	.085089					
rho	.54199623	(fraction of variance due to u_i)				

Source: Frontier Economics

Figure 48. Regression excluding SSEH; Specification 2

```
Random-effects GLS regression                   Number of obs      =        78
Group variable: num_id                          Number of groups   =        13

R-sq:  within  = 0.2343                         Obs per group: min =         6
       between = 0.9122                                        avg =       6.0
       overall = 0.8547                                        max =         6

                                                Wald chi2(4)       =    121.74
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000
```

l~t_capital1	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
l_customers	.7012417	.1692776	4.14	0.000	.3694636	1.03302
l_peak	.1851531	.1461228	1.27	0.205	-.1012423	.4715485
l_density	-.0372067	.0317148	-1.17	0.241	-.0993667	.0249532
l~1_capital1	.5781098	.1473917	3.92	0.000	.2892274	.8669922
_cons	-9.915169	1.502628	-6.60	0.000	-12.86027	-6.970072
sigma_u	.08903592					
sigma_e	.08289145					
rho	.53569315	(fraction of variance due to u_i)				

Source: Frontier Economics

Figure 49. Regression excluding LPN; Specification 2

```
Random-effects GLS regression                   Number of obs      =        78
Group variable: num_id                          Number of groups   =        13

R-sq:  within  = 0.1978                         Obs per group: min =         6
       between = 0.9263                                        avg =       6.0
       overall = 0.8824                                        max =         6

                                                Wald chi2(4)       =    136.86
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000
```

l~t_capital1	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
l_customers	.6617862	.1980193	3.34	0.001	.2736755	1.049897
l_peak	.2317371	.1478605	1.57	0.117	-.0580641	.5215384
l_density	-.1058631	.0696747	-1.52	0.129	-.242423	.0306967
l~1_capital1	.4683905	.1459331	3.21	0.001	.182367	.7544141
_cons	-9.596267	1.741174	-5.51	0.000	-13.00891	-6.183628
sigma_u	.09639446					
sigma_e	.08145319					
rho	.58342286	(fraction of variance due to u_i)				

Source: Frontier Economics

Figure 50. Regression excluding SSEH and LPN; Specification 2

```
Random-effects GLS regression                   Number of obs      =        72
Group variable: num_id                          Number of groups   =        12

R-sq:  within  = 0.1899                         Obs per group: min =         6
       between = 0.9151                                        avg =       6.0
       overall = 0.8610                                        max =         6

                                                Wald chi2(4)       =    109.69
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
l~t_capital1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 l_customers |   .6445099   .2005794     3.21   0.001     .2513815    1.037638
      l_peak |   .2135408   .1510097     1.41   0.157    -.0824327    .5095144
   l_density |    .003019   .1116583     0.03   0.978    -.2158271    .2218652
l~1_capital1 |   .4823039   .1540239     3.13   0.002     .1804226    .7841852
       _cons |  -9.714653   1.705705    -5.70   0.000    -13.05777   -6.371534
-------------+----------------------------------------------------------------
     sigma_u |  .09264285
     sigma_e |  .08308316
         rho |  .55424059   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Source: Frontier Economics

FRONTIER ECONOMICS EUROPE
BRUSSELS | COLOGNE | LONDON | MADRID

Frontier Economics Ltd 71 High Holborn London WC1V 6DA
Tel. +44 (0)20 7031 7000 Fax. +44 (0)20 7031 7001 www.frontier-economics.com