Opex Efficiency Modelling Strategy

January 2019

Northern Ireland Utility Regulator

FINAL

IMPORTANT NOTICE

This report was prepared by Cambridge Economic Policy Associates Limited (CEPA) for the exclusive use of the Northern Ireland Utility Regulator (UR).

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

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

CEPA does not accept or assume any responsibility in respect of the report to any readers of the report (third parties), other than the client(s). To the fullest extent permitted by law, CEPA will accept no liability in respect of the report to any third parties. Should any third parties choose to rely on the report, then they do so at their own risk.

CONTENTS

1. EXECUTIVE SUMMARY

Cambridge Economic Policy Associates Limited (CEPA) has been engaged to support the Northern Ireland Utility Regulator (UR) in assessing the relative efficiency of Northern Ireland Water's (NI Water) operating expenditure (opex) compared with water and wastewater companies in England and Wales. The objective of this short paper is to present the strategy that CEPA expects to employ when developing water and wastewater econometric opex models for PC21.

To develop our modelling strategy, we have followed best practice, learned from Ofwat's work at PR19, and taken a proportionate approach.

The table below presents a summary of the key points to our strategy for developing cost assessment models for PC21.

Table 1.1: Summary of model development strategy

Category	Approach
Target modelling suite	We will focus on developing top-down water and wastewater models. We will develop sewerage models that exclude bioresources to control for differences in sludge treatment/disposal between NI Water and England and Wales companies.
Data adjustments	We will exclude a number of costs from the models, including business rates, pension deficit repair costs, TMA costs and atypical costs. We do not expect to make any pre-modelling adjustments to the data to cover regional price differentials. Instead they will be dealt with through the special factor process.
Functional form	We will aim to develop simple models. If the data suggests that more complex relationships exist, we will consider whether these can be captured by other explanatory variables and whether higher order terms (i.e. quadratic terms) add much explanatory power to the models.
Estimation method and assumptions on efficiency	There are several different estimation methods available, each with different implications for how model residuals and company efficiency are calculated. According to UR transparency is a key priority for PC21. Therefore, we propose to focus on COLS as it is easy to replicate and understand compared with other modelling approaches.
Explanatory variables	Our work for Ofwat for PR19 identified a number of explanatory variables that could be used in the modelling and categorised them into five cost driver groups.[1] The models we develop will be based on a subset of these variables. We also note that some of the variables we eventually use may be transformations or combinations of the variables we have set out in this report.

[1] CEPA, March 2018. CEPA cost assessment report.

2. INTRODUCTION

Cambridge Economic Policy Associates Limited (CEPA) has been engaged to support the Northern Ireland Utility Regulator (UR) in assessing the relative efficiency of Northern Ireland Water's (NI Water) operating expenditure (opex) for PC21.[2] As we are considering relative efficiency, we need to compare NI Water to other companies and we do this using data that Ofwat has collected for the England and Wales water and sewerage companies. The objective of this short paper is to present the strategy that CEPA expects to employ when developing water and wastewater econometric opex models for PC21. It is designed to ensure that the models we develop are in line with UR's expectations and maximise transparency between UR and NI Water.

In the first part of this note we discuss the main characteristics, or target modelling suite, we are aiming to achieve. Secondly, we present the strategy CEPA will use to ensure a robust model development process. This is our initial approach, which could change as the models are developed. We also discuss how models could be adjusted during the analysis.

To ensure the process of model development and selection is objective and transparent, we developed a separate paper describing our a priori assumptions for explanatory variables (based on economic and technical rationale) and provided this to UR before model development commenced[3]. This demonstrates objectivity in modelling as opposed to 'picking and choosing', which is often raised as an issue by regulated companies.

3. TARGET MODELLING SUITE

The first step in developing our strategy for cost assessment is to ensure that there is a clear understanding between CEPA and UR of the modelling suite that UR would like to achieve as well the use or uses intended for these models.

The focus of this paper is the development of opex models. However, a similar methodology would apply for the development of maintenance (and potentially botex) models.

When assessing opex, the first stage is to identify what activities will be included. Ofwat developed models at a high level of cost disaggregation at PR19, as illustrated below.

Table 3.1: Ofwat PR19 model disaggregation

Activities being included		
High degree of aggregation	**Medium degree of aggregation**	**Disaggregated model**
• Wholesale water	• Water resources	• Water resources • Water licencing costs
	• Water network	• Raw water distribution • Water treatment • Treated water distribution
	• Wastewater network	• Sewage collection

[2] The price control for NI Water from 2021.

[3] Source: CEPA, March 2018. CEPA cost assessment report. Available in http://www.cepa.co.uk/userfiles/CEPA-cost-assessment-report.pdf

• Wholesale wastewater		• Sewage treatment
	• Bioresources	• Transport • Treatment • Disposal

After evaluating the data collected by UR and considering proportionality, we consider it is only possible to develop models at a high degree of aggregation. It would be very challenging to obtain (robust) more granular data and it is not clear to us that a proportionate approach warrants such an intensive data collection strategy. As a result, we will focus on developing wholesale water and wastewater models.

For wastewater models, we propose to consider sewage models (excluding bioresources). This approach reflects the presence of a public-private partnership (PPP) in Northern Ireland for the treatment of bioresources by incineration. As a result, NI Water operates under different conditions to England and Wales companies as:

- It does not use the same sludge treatment technology as companies in England and Wales. The sludge generated by NI Water is incinerated instead of using anaerobic digestors.
- It does not operate the sludge treatment activities. The sludge is treated by a third party through a PPP contract which limits NI Water's decision making.

We will develop wholesale water and sewage models. When developing these models, CEPA will try to balance robustness with simplicity / transparency to enable the various stakeholders to understand and challenge UR's findings. Model robustness will be ensured by assessing model performance against out model selection criteria.[4] Simplicity and transparency will be assured by presenting modelling results to stakeholders within the cost assessment working group and clearly describing our results in our final report.

4. MODELLING STRATEGY

When developing models, CEPA will need to use a number of assumptions. The application of these assumptions will need to be flexible to ensure that the models are robust and supported by the data. Proposed starting assumptions and the potential alternatives are discussed below.

4.1. Unmodelled costs

Certain costs may be excluded from opex efficiency analysis because:

- UR has decided that a separate treatment should be applied (e.g. business rates[5]).
- The costs are lumpy and not repeated over time. Therefore, their inclusion would distort the modelling. For example, any atypical costs incurred in PC15 that are unlikely to reoccur in PC21 (e.g. the costs of dealing with a flood or some other exogenous factor).

[4] As detailed in a separate paper that describes CEPA's approach to assessing model robustness, "Opex model assessment criteria".

[5] The decision to conduct a review of treatment of business rates was taken around the time of the RP6 price control of NIE Networks.

- The costs are only significant / incurred by a small number of companies. For example, traffic management act (TMA) costs are significant for some companies in England and Wales but zero for others (including NI Water).
- Certain costs do not follow the same cost drivers as the costs being modelled. For example, certain costs may be driven by statutory / regulatory requirements (e.g. environmental standards) rather than the scale / density of the company.

The initial list of costs that CEPA expects to exclude is presented in the table below:

Table 4.1: Proposed cost exclusions

Cost	Rationale
Business rates	Separate treatment by UR
Pension deficit repair costs	
Traffic Management Act (TMA) costs	Vary significantly across companies.
Abstraction charges / discharge consents	Abstraction charges are set by the Environment Agency, which reduces their controllability at a company level. This affects only water activities.
Statutory water softening	Atypical in nature as only one company in England and Wales incurs costs associated with statutory water softening.
Costs associated with the industrial emission directive	Costs associated with the Industrial Emission Directive were only incurred by a small number of companies in England and Wales over the historical period.

Source: CEPA

4.2. DATA ADJUSTMENTS

Adjustments are introduced to facilitate the homogeneity across the different companies in the sample. The table below presents a provisional set of data adjustments we expect to make to ensure we are using a comparable set of cost data when developing the benchmarking models.

Table 4.2: Proposed adjustments to costs

Adjustment	Rationale
• Regional wages • Energy (price and volume)	There could be regional pressures on the prices of some inputs, which would result in higher costs for companies simply due the location of their network. These may include differences in regional wages or energy prices. Regional price differences can be controlled for using one of three approaches: **(1) Ex-ante adjustment:** costs are adjusted for regional price differentials before modelling. This approach was taken by the UR at RP6 and GD17 when considering differentials in labour costs. **(2) Within the model:** price index variables are included within the econometric cost model. This approach was taken by Ofwat at PR14. **(3) Special cost factor:** no ex-ante or within model adjustments are made. Instead companies are required to supply a claim for a special cost factor adjustment where such costs are considered material. This approach was taken by UR from PC10 through PC15 and by Ofwat at PR19 for any potential differential in labour costs. Our initial view is that option 3 – the special cost factor approach - is preferred for PC21. This approach avoids using up degrees of freedom and also avoids placing restrictions on the underlying relationship between costs and regional price

Adjustment	Rationale
	differentials. This approach is also in line with the approach taken by UR at PC15 and therefore is well understood by NI Water and other stakeholders.

Source: CEPA

In addition to the above, it may also be necessary to make further adjustments for other data issues raised in the course of the data review process. We will assess these on a one-off basis, but these could cover:

- errors in reporting;
- differences in the scope of activities;
- differences in cost-allocation; and
- differences in operating environments.

We will record all data adjustments made and their rationale and will include them in our final report. We will also discuss them with relevant stakeholders at the cost assessment working group.

4.3. Capturing economies of scale

Economies of scale are an important consideration for efficiency benchmarking given the significant variation in the size of companies that we include in our sample. This issue is exacerbated with the inclusion of NI Water, which is relatively small compared to most England and Wales companies.

There are different approaches that could be used to account for economies of scales. The two approaches we propose to combine when developing models for PC21 are discussed below:

4.3.1. Using different cost drivers to capture economies of scale:

In this approach the models include variables that allow them to account for the factors that would affect the capacity of the company to realise economies of scale, for example, a larger than average size of water treatment plant would result in a company to profit from economies of scale. Using variables within the models that account for economies of scale is an approach that was favoured by Ofwat at PR19 as it is arguably more transparent than using complex functional forms to capture differences in economies of scale. For example, the percentage of load treated in small treatment works to capture economies of scale in sewage treatment.

4.3.2. Functional form

Different assumptions about the functional form can be used to capture economies of scale.[6] The main functional forms used to capture economies of scale are:

- Cobb-Douglas; and

[6] We note that the functional form cannot be considered independently from statistical performance of the variables in the models.

- Translog.

Cobb-Douglas is a standard functional form used in cost assessment literature that places weights on the input factors (i.e. cost drivers). When in a log-linear form, Cobb-Douglas models allow coefficients to be interpreted as cost elasticities. Cobb-Douglas models are relatively easy to replicate and interpret and therefore do not create opacity. However, they do suffer from the imposition of constant returns to scale (i.e. all companies are assumed to have the same level of economies of scale). Therefore, to control for differences in economies of scale one would need to use explanatory drivers that account for these differences.

Translog models relax the assumption of single value for economies of scale across the industry by allowing economies of scale to vary by company. However, the use of these models makes it more difficult to identify the specific effect of each variable on costs. Ofwat's PR14 models included translog terms but these were criticised by the CMA as being ambitious given the sample size. As a result, Ofwat for PR19, did not propose the inclusion of translog terms in their initial models.[7]

When developing PC21 models, we will start the analysis using a Cobb-Douglas functional form. This provides an intuitive result that is consistent with the challenge from the CMA and Ofwat's approach to cost modelling at PR19. However, we will test for evidence that other functional forms would better fit the data (i.e. whether there appear to be non-linear components).

If there is evidence to suggest that there are varying degrees of economies of scale we will consider:

- the benefit of including non-linear scale variables (e.g. length of mains squared) that allow for varying degrees of economies of scale across companies; and
- whether additional cost drivers could be used to control for the effect (e.g. percentage of load received at small sewage treatment works).

4.4. UNDERPINNING EFFICIENCY ASSUMPTIONS

There are different estimation methods available to us when estimating a given model specification. Each method introduces different assumptions about the composition of the error terms, and the associated assumptions about company (in)efficiency. Each method has pros and cons, such as the extent to which it recognises the panel structure of the data.

In this respect, methodologies can be classified into two main categories: pooled and panel data. Pooled data methodologies disregard the time component implicit in the observations for each one of the companies and assume these are uncorrelated observations. Panel data methodologies introduce assumptions about the sources of this potential correlation such that they can account for the time component in the data.

The pros and cons of each approach are summarised in the table below for the more common econometric techniques used in cost / efficiency benchmarking.

[7] Source: Ofwat, March 2018. Cost Assessment for PR19 – a consultation on econometric cost modelling.

Table 4.3: Econometric estimation techniques

Estimation method	Description	Strengths	Weaknesses
Pool models (e.g. Corrected Ordinary Least Squares (COLS))	• Pool models can be run on a cross-section (one year) or on multiple years. Each data point is treated as a unique firm. • Pool models place equal weight on the 'between variation' (differences between companies) and 'within variation' (differences between years for the same company).	• Use of pooled data increases the sample size and few distributional assumptions are required. • Estimates of efficiency are variable over time. • Relatively easy to understand and replicate.	• Pool models do not include assumptions that allow separation between the white noise, company heterogeneity, and inefficiency in the errors term. • No structure to estimates of efficiency across time.
Panel data models with Random Effects (RE)	• Uses Generalised Least Squares (GLS), which places more weight on the 'within' variation than OLS when calculating parameter estimates. • Requires that firm-specific effects be uncorrelated with cost drivers.	• Panel methods in general have the advantage that estimation takes into account the structure of the data. • The structure imposed on the error term allows efficiency to be differentiated from white noise.	• Efficiency is assumed to be constant over time. • Relatively complicated compared to pool models.
Panel data models with Fixed Effects (FE)	• Allows company specific effects to be correlated with cost drivers by estimating the company effect.	• Takes into account the panel structure of the data. • Produces unbiased and consistent parameter estimates in the presence of correlation between company effects and cost drivers.	• Efficiency is assumed to be constant over time. • Difficult to distinguish between inefficiency and company heterogeneity. • Data requirements are relatively high.
Stochastic Frontier Analysis (SFA)	• A maximum likelihood estimation method approach, requiring distributional assumptions on both the error term and efficiency.	• Allows for white noise to be separated from inefficiency and imposes a structure on the progression of inefficiency over time.	• Requires distributional assumptions and is data intensive. • SFA models have proven to be hard to implement and are rarely pursued.

Source: CEPA

For PC21, UR prioritises transparency and replicability. Therefore, we propose to prioritise COLS models as they are relatively easy to replicate and understand compared with other modelling approaches. In addition, COLS models have the benefit (over random effects) that efficiency is allowed to vary over time, which is arguably more important in the PC21 context given the focus on assessing relative efficiency rather than forecasting costs. Therefore, random effects will only be used if COLS models do not perform well against our model selection criteria.

While the way in which 'noise' is separated from inefficiency in SFA models is appealing, given the limited sample size available and difficulty Ofwat experienced in developing robust SFA models for PR14, we do not propose to pursue any SFA models for PC21. We would propose to use FE models only if COLS and RE models are unworkable, since FE models are both data intensive and make it relatively difficult to distinguish between heterogeneity and inefficiency.

4.5. Potential explanatory variables

The identification of the relevant explanatory variables that will be used within the econometric models as a proxy for cost drivers is a vital part of the model development process. Cost drivers can be split into five key areas, as presented in the figure below.

Figure 4.1: Cost drivers to be considered in model development

Scale of the activities	• The scale of the activities being undertaken by a company is expected to be a strong driver of its total costs. • It allows to undertake an initial evaluation of whether there are economies of scale in the activities and costs being modelled.
Density	• The density of population could affect the total cost of some of the activities of the companies (e.g. networks). The effect on the costs of the company, however, could be ambiguous.
System characteristics	• The characteristics of the assets and systems operated by the company could also affect its total cost for the provision of the services. For example, the topology of the network the company operates in.
Quality	• Increasing the quality of the service delivered by the companies can have an ambiguous effect. • On the one hand, additional quality could be coming at an additional cost as the company invests to provide this higher quality. On the other hand, higher quality could lead to lower leakage and in turn lower opex.
Level of activity	• Part of the differences in costs could reflect a higher (but efficient) amount of activity being undertaken.

Source: CEPA

When developing models, CEPA will use the 'specific' to 'general' approach, i.e. start running models with only a scale -related variable and then expand the number of variables to evaluate the potential effects of adding each one of these additional variables.

When deciding the variables to be added, more relevant cost drivers with independent explanatory variables will be added first while cost drivers where companies have a certain degree of control will be added later. In principle, the cost drivers will be added using the order presented above.

The CMA criticised the number of explanatory variables included in Ofwat's PR14 models and recommended three variables (plus a constant) be included. We do not consider this 'rule of thumb' used by the CMA to be determinative[8]. We consider there is not a single number of cost drivers that is theoretically correct. For PC21 we will refine the set of cost drivers included in the models based on CEPA's model selection criteria (discussed in CEPA's short paper titled 'Assessing model robustness'). Alongside this, it may be appropriate to use multiple models with different sets of cost drivers which could then be triangulated to arrive at a final efficiency estimate.

[8] The Law of Parsimony as applied to econometrics is however accepted as standard practice.

In terms of the cost drivers themselves, we have identified the potential explanatory variables that could be used in the models based on Ofwat's proposed PR19 models, published England and Wales company data and NI Water data. We do not present an exhaustive list of explanatory variables here but instead refer you to CEPA's March 2018 report for Ofwat.[9]

[9] Source: CEPA, March 2018. CEPA cost assessment report. Available in http://www.cepa.co.uk/userfiles/CEPA-cost-assessment-report.pdf

CEPA

Queens House
55-56 Lincoln's Inn Fields
London WC2A 3LJ
United Kingdom

Level 20, Tower 2
Darling Park, 201 Sussex St
Sydney NSW 2000
Australia

CEPA Ltd
@CepaLtd