This is a repository copy of *Benchmarking water retail cost efficiency in England and Wales*.

**Article:**

# Benchmarking Water Retail Cost Efficiency: Evidence for Companies in England and Wales

Sam Williams[*], Christopher Pickard[#], Karli Glass[§] and Anthony Glass[¶]

## Abstract

Privatised water companies in England and Wales are subject to economic regulation by the industry regulator (Ofwat). Ofwat sets 5-year caps on the prices companies can charge their customers. These caps are in part based on the benchmarking of companies' costs. Ofwat has not previously used econometrics to benchmark domestic retail costs, but it undertook such modelling in its 2019 price review analysis. This is the first journal article to present an efficiency analysis of domestic water retail costs in England and Wales. Our approach is different from Ofwat's, as we propose two new ways of accounting for differences in the effect on cost of the number of single and dual service customers (water / sewerage-only and water and sewerage). Some companies' cost efficiencies vary greatly between the two ways. Depending on the approach to price capping, this could possibly have non-negligible implications for companies' caps.

**JEL Classification:** D24; L95; L51

**Key words:** Periodic price caps; Frontier models; Single service (water / sewerage-only) customers; Dual service (water and sewerage) customers; Bad debt costs

## 1. Introduction

Water and sewerage services in England and Wales are supplied by privatised companies that are monopolists in their supply areas. To provide incentives to improve their performance and prevent the companies from taking advantage of their monopoly power, they are subject to economic regulation by the Water Services Regulatory Authority (Ofwat). Among other things, this regulation involves Ofwat imposing caps on the prices that water companies can charge their customers at price reviews that occur every five years, the most recent being in 2019 (PR19). These caps are set, in part, based on an assessment of the relative cost efficiency of a company vis-à-vis the other water companies in the industry. For wholesale water costs and sewerage costs at the 2014 Price Review (PR14), but not domestic retail costs, Ofwat used econometric modelling to benchmark companies' cost efficiencies (CEPA, 2014).[1]

---

[*] Corresponding author. Economic Insight, 125 Old Broad Street, London, EC2N 1AR. Email: sam.williams@economic-insight.com. The authors acknowledge comments on an earlier version of this paper from colleagues at Economic Insight, and participants in a special session on 'Regulatory Cost Assessment in the Water Industry' at the 16th European Workshop on Efficiency and Productivity Analysis, June 2019, London. The authors also acknowledge the constructive comments on an earlier version of the paper from an editor and two anonymous reviewers.

[#] Economic Insight, 125 Old Broad Street, London, EC2N 1AR. Email: christopher.pickard@economic-insight.com

[§] School of Business and Economics and Centre for Productivity and Performance, Loughborough University, Leicestershire, LE11 3TU, UK. Email: k.glass@lboro.ac.uk

[¶] Sheffield University Management School, The University of Sheffield, Conduit Road, Sheffield, S10 1FL, UK. Email: a.j.glass@sheffield.ac.uk

In the water industry, retail activities include: customer services (billing, payment handling, enquires and complaints, etc.); debt management; meter reading; decisions and administration of disconnections and reconnections; demand-side water efficiency initiatives; customer-side leaks; general support; and providing developer information (Economic Insight, 2017, p. 18). As part of Ofwat's PR19 analysis, separate econometric modelling of companies' domestic retail cost efficiencies was undertaken (Ofwat, 2018).[2] In line with the modelling in Ofwat (2018), we estimate the relative retail cost efficiencies of English and Welsh water companies, which is timely because of the importance that is attached to the quality of customer service in the industry.[3] Our approach to benchmark companies' retail cost efficiencies involves using two well-established methodologies for econometric frontier models. In doing so, our paper makes four contributions, which we set out in the remainder of this introductory section and relate to the relevant literature in subsection 2.2.

Given the parallels between Ofwat's (2018) modelling of domestic retail costs and our modelling, our first and headline contribution is to suggest how to address practical issues that arise when implementing a general efficiency framework, in order to arrive at an empirical model that can inform public utility policy on benchmarking. A key aspect of our practical suggestions is to ensure that our empirical models reflect the CMA's recommendations about Ofwat's cost models at PR14 (CEPA, 2014). The CMA set out its recommendations in its redetermination of Bristol Water's price controls in response to the company's appeal against Ofwat's PR14 determination (CMA, 2015, pp. 72 - 73).

The CMA's recommendations concerning Ofwat's PR14 models that we take account of in our modelling approach relate to: (i) counterintuitive signs and magnitudes of coefficients on some explanatory variables; (ii) a large number of explanatory variables relative to the small sample size – the CMA note that Ofwat's full total expenditure (totex) model has 27 explanatory variables and is estimated using 90 observations; (iii) the use of the translog functional form for the models; (iv) questions concerning the relationships between cost and some explanatory variables (e.g., taking logs of variables that are expressed as proportions).

Each practical step we take to ensure that our empirical models reflect the corresponding above recommendation is as follows.

(a) For explanatory variables to be retained in our preferred set of models the fitted coefficients must have the expected signs (based on microeconomic theory or economic intuition) and have plausible magnitudes. The latter is important when having to form a preference between models because we may expect a fitted coefficient to lie within a narrow range of values (e.g., when measuring scale economies). The decision to retain an explanatory variable will, of course, also depend on whether it has a statistically significant impact. As we will discuss in due course, we take a rather liberal

view of what constitutes statistical significance, which can involve retaining an explanatory variable if it does not have a significant impact at the typical nominal levels (5% or lower). This situation would arise when based on these nominal levels an explanatory variable does not have a significant impact, but because we deem it to be economically important it is retained in the model, which will have implications for the companies' efficiencies.

(b)    The sample size we use is small, which is very similar to those used for PR19 across the industry (i.e., by Ofwat and water companies), so there is a need for parsimonious model specifications to conserve degrees of freedom. We obtain parsimonious fitted models through the functional form for the models, which we discuss in (c), and by using a general-to-specific modelling approach, which we describe in due course.

(c)    As Ofwat's PR14 cost models have a translog functional form some of the regressors are quadratic and interaction variables. The presence of these variables, which account for up to six regressors in a model, is one of the factors that contributes to Ofwat's PR14 models having a large number of explanatory variables relative to the small sample size. Also, with reference to (a) above, it is not clear from microeconomic theory or economic intuition what the expected signs are of the coefficients on the interaction variables. By using the simpler Cobb-Douglas functional form for our retail cost models the quadratic and interaction variables are omitted by construction, which is one of the factors that contributes to our models being parsimonious and hence conservers of degrees of freedom. Moreover, by omitting the interaction variables, we avoid having explanatory variables where is not clear what the expected signs of the coefficients on these variables should be.

(d)    We provide a justification of the relationship between domestic retail cost and each explanatory variable by motivating the inclusion and form of the cost driver (e.g., how we account for companies' single and dual customer numbers).

While CEPA (2014) use a translog functional form and Ofwat (2018) and the models we report employ a Cobb-Douglas form, the two methods we use to estimate efficiency are among the approaches that CEPA (2014) adopt and, as Welsh Water (2019, p. 7) confirm and we discuss in due course, are also the approaches that Ofwat (2018) use. In particular, the two methods we use are the approaches that Ofwat (based on advice from CEPA, 2014) ultimately used to set the companies' base wholesale water and sewerage cost allowances at PR14. CEPA (2014) also consider a range of other efficiency models. Among these models are a number of stochastic frontier specifications that assume a composed error. This error is split into inefficiency and noise using assumptions about the distributions of these error components. We do not, however, consider stochastic frontier efficiency models here for the following two reasons.

First, since our headline contribution is to suggest how to address practical issues that arise when developing models to inform public utility policy on benchmarking, at the time when our work was undertaken, it made sense to employ the approaches that Ofwat (in light of the CMA's (2015) comments on Ofwat's PR14 modelling) was likely to adopt at PR19. As it turned out and in line with our expectations, Ofwat did not consider stochastic frontier efficiency models at PR19. Second, when the CMA (2015, pp. 73 - 74) queried the complexity of the modelling CEPA (2014) carried out for Ofwat because the results can be difficult to interpret, although the CMA did not explicitly refer to stochastic frontier models, to avoid our modelling being viewed as similarly complex, we take the view that the best course of action is not to consider such models. We discuss this in more detail in subsection 3.2, but, in short, this is because stochastic frontier models are the most complex model specifications that CEPA consider.

While there are similarities between our approach to the econometric benchmarking of domestic retail costs and Ofwat's (2018) approach, there are also some important differences.[4] The second contribution of this study therefore is to compare our approach with Ofwat's. This study is also the first journal article to model the domestic retail costs of water companies in England and Wales. As a result, our third contribution is to complement the studies that estimate cost models for: (i) the water distribution of English and Welsh companies (e.g., Cubbin and Tzanidakis, 1998, Thanassoulis, 2000, Ashton, 2003, and Molinos-Senante and Maziotis, 2018; 2019a; 2019b); (ii) the sewerage operations of companies in England and Wales (e.g., Thanassoulis, 2002); (iii) water industries in other countries (e.g., 19 Asia-Pacific countries (Estache and Rossi, 2002), Peru (Corton, 2003), Portugal (Correia and Marques, 2011), Italy (Guerrini *et al.*, 2013) and Denmark (Guerrini *et al.*, 2015)); and (iv) retail activities in Portugal (e.g., Correia and Marques, 2011, Marques and De Witte, 2011, and Carvalho, and Marques, 2014). More generally, as our study is the first journal article to benchmark the domestic retail costs of water companies in England and Wales, our fourth contribution is in the form of a new application of efficiency analysis in the academic literature. Our study adds to a large number of interesting efficiency applications to, for example, UK banks (Tanna *et al.*, 2011), European life and non-life insurance companies (Vencappa *et al.*, 2013), quoted commercial banks in 46 countries (Gaganis *et al.*, 2013) and banks in 10 EU accession countries (Mamatzakis *et al.*, 2013), to name but a few.

The remainder of this paper is organised as follows. In section 2 we put the paper into context by providing some industry background and discussing how it fits into the extant literature. The methodology is presented in section 3, which has two parts. In the first part we set out the general forms of the two costs models we estimate. In the second part we explain how these general forms lead to two approaches to estimate the industry best practice frontier and thus the companies' efficiencies. As well as the simpler Cobb-Douglas functional form we discussed above, in section 4 we discuss four further practical considerations when implementing our methodology. In section 5 we motivate our choice of independent

variables to explain domestic water retail costs. Section 6 describes the data and in section 7 we present the fitted models and estimates of the cost efficiencies. Section 8 discusses the empirical findings and their natural policy implications. The latter focus on the regulatory mechanism, whereby companies' domestic retail cost efficiencies influence the companies' domestic retail price controls that Ofwat set. In section 9 we conclude by summarising the salient features of the study.

## 2. Industry Background and the Relevant Literature

### 2.1 *Industry Background*

Ofwat (2016, pp. 12 - 15) sets out the broad range of activities included within domestic retail in the English and Welsh water and sewerage industry. Using this categorisation, we present in figure 1 the constituent parts of the total expenditure (£816 million) in the industry on domestic retail in 2016/17. £359.2 million (44%) is categorised as bad debt, which is the sum of the bad debt charges reported in the companies' profit and loss accounts (i.e., the sum of the companies' adjustments for write-offs). Bad debt collectively refers to doubtful debt (£277.8 million) and expenditure on debt management (£81.4 million).[5] A further £228.7 million (28%) is customer services expenditure and depreciation of retail assets amounts to £34.5 million (4%).[6] Meter reading costs account for £42.4 million (5%) and relate to cyclical meter reads, ad hoc read requests, scheduling, transport, processing and the management of meter data and meta readers. The remaining £151.2 million (19%) comprises costs associated with office provision, insurance premiums, demand-side efficiency initiatives, local authority rates and all other costs.

[Insert figure 1 about here]

17 companies in the English and Welsh water and sewerage industry currently provide retail services to domestic customers and are subject to detailed price cap regulation. All 17 companies also provide wholesale services to domestic customers. 10 of the 17 are water and sewerage companies (WASCs) and the remaining 7 are water-only companies (WOCs). In many cases the same WASC will supply both water and sewerage services to a customer, a situation referred to as 'dual service', though these may be supplied by different WASCs. Where a WOC supplies a customer's water, the customer's sewerage services will be supplied by a WASC. Unlike business customers, domestic customers cannot choose their retail supplier.

At PR14 Ofwat used a rather different approach to econometric benchmarking of domestic retail costs to set the binding price controls for companies' domestic retail operations. To enable companies to recover their efficient costs, at PR14 Ofwat set the companies' allowed revenues for domestic retail based on the companies' average costs to serve their customers. These unit costs were weighted to reflect differences in, among other things, average costs between water-only, sewerage-only and dual service customers, and

between metered and unmetered customers. Adjustments were also applied to the average costs to serve to reflect differences in the levels of bad debt between companies.

## 2.2    *Fit of the Study within the Relevant Literature*

We discuss how the paper fits within the extant literature by focussing on two issues. First, how our study fits within the academic literature on the efficiency of water operations, and, second, key similarities and differences between our study and Ofwat (2018). Interestingly, a number of the academic efficiency studies of water operations we cited above use different frontier methods. For instance, Estache and Rossi (2002) use stochastic frontier analysis (SFA), which, as we noted above and will expand on in subsection 3.2, we do not use. This is to prevent our modelling from being subject to the same concern that the CMA (2015) had about CEPA (2014) using complex models that yield results that can be difficult to interpret.

One of the methods that Cubbin and Tzanidakis (1998) use (Corrected Ordinary Least Squares, COLS), which is also one of the methods we use to estimate efficiency, and, although parametric like SFA, does not involve splitting the error term using distributional assumptions. COLS therefore, like the other method we use to compute efficiency (Schmidt and Sickles, 1984, denoted SS from hereon), is referred to as deterministic frontier analysis (DFA). As we noted above, we use these two approaches to estimate efficiency because the headline contribution of this paper is to resolve practical considerations when developing models to inform public utility policy on benchmarking. It made sense therefore to use these two approaches because both were used in a PR14 report for Ofwat by CEPA (2014) and in analysis Ofwat undertook for PR19 (Ofwat, 2018), where Welsh Water (2019) confirm the use of COLS in the latter.[7] The other method that Cubbin and Tzanidakis use, and also Thanassoulis (2000; 2002) and Guerrini *et al*. (2013; 2015), is data envelopment analysis (DEA), which is a non-parametric linear programming technique. Ofwat did not use DEA at PR19, and never appeared to be considering doing so, so we do not discuss it any further here.[8]

There are a number of similarities between Ofwat's (2018) approach to the modelling of domestic retail costs and that which we adopt. Among other things, this involves taking account of the broad range of activities within domestic retail, as set out in figure 1 above. Ofwat's approach to this, which is also the approach we take for reasons we will explain in due course, is to group retail costs into two categories, bad debt costs and non-bad debt costs. A separate model is then estimated for each category.

There are also various differences between Ofwat's (2018) domestic retail cost modelling and how we go about it, such as the debate about what data is used for some of the explanatory variables (e.g., the data and explanatory variables that are used to capture income deprivation). We acknowledge this debate, but instead focus on two more involved issues. The first concerns the type of standard errors that are used. Ofwat use clustered standard errors, while for reasons we elaborate on further in the paper, we favour

robust standard errors and the usual unadjusted standard errors. The type of standard error is very important because in line with the CMA's (2015) recommendation that Ofwat's cost modelling at PR14 was too complex, and reflecting our small sample, which is essentially an industry standard sample, we use a general-to-specific modelling approach to arrive at simple parsimonious model specifications. This basically involves beginning with a relatively large number of explanatory variables and omitting variables that are deemed to be statistically / economically unimportant. The type of standard errors will influence the variables that are deemed to be statistically important and are thus retained in the model, which will affect the estimates of the companies' efficiencies.

The second difference we focus on between Ofwat's approach and that which we adopt, concerns how the potential retail cost sharing between single and dual service customers is accounted for. Ofwat account for this effect by including the ratio of single to dual service customers as an explanatory variable. We, on the other hand, use two different approaches that involves including separate explanatory variables for the number of single and dual service customers, or, alternatively, including the total number of customers and the number of single service customers as regressors. A priori, we do not have strong reasons to favour Ofwat's approach or either of our approaches to capture this effect. A key finding from our empirical analysis, however, is that for a number of companies there is quite a large variation in their efficiencies depending on which approach we use to capture this effect. This suggests that it is important to consider different ways of accounting for this effect if econometric models are used to benchmark companies' domestic retail costs. This is because how this effect is captured could potentially have a non-negligible implication for a company's price cap. It is not though necessarily a matter of having a preferred way of accounting for this effect. This is because we estimate a suite of frontier models, which is appealing as quite different efficiencies can be averaged across models to take account of the effects of different viable model specifications.

## 3. Methodology

### 3.1 *General Presentation of the Retail Cost Models*

Eqs. 1 and 2 represent the general forms of the pooled OLS and random effects models that are used to calculate the efficiencies using COLS and the SS method, respectively.

$$c_{it} = \alpha + c(y_{it}, w_{it}, t) + \gamma' z_{it} + \varepsilon_{it}, \tag{1}$$

$$c_{it} = \alpha + c(y_{it}, w_{it}, t) + \gamma' z_{it} + \theta_i + \varepsilon_{it}. \tag{2}$$

In Eqs. 1 and 2 the variables are logged and the panel data comprises observations for $N$ firms and $T$ periods, which are indexed $i=1,\ldots,N$ and $t=1,\ldots,T$. $c_{it}$ is an observation for total retail operating costs or one of its two cost categories that we consider - bad debt costs or non-bad debt costs. $\alpha$ is the common intercept

across the firms; $y_{it}$ is either an observation for a single output or vector of observations for multiple outputs; $w_{it}$ is either an observation for a single input price or vector of observations for multiple input prices; $t$ is the time trend in the dependent variable; and $\varepsilon_{it}$ is the idiosyncratic error, where all the other variables and parameters in Eqs. 1 and 2 are discussed in due course.

$c(y_{it}, w_{it}, t)$ is a general representation of the Cobb-Douglas functional form that we use for the cost models. Note that estimates of Eqs. 1 and 2 do not represent an approximation of the relevant logged cost frontier. This is because having estimated Eqs. 1 and 2 the frontiers are specified by shifting the fitted models to the location of the best performing firm, which is taken to lie on the frontier and is thus 100% efficient. For the estimates of Eqs. 1 and 2 we use a different approach to shift the fitted model to specify the frontier and estimate efficiency, which we discuss in the next subsection.

Since we use a Cobb-Douglas functional form we do not include $t^2$ or interactions with $t$. Due to the omission of these interactions $t$ captures Hicks-neutral technical change.[9] $t$ therefore shifts the cost frontier and does not influence its slope. In addition, $z_{it}$ in Eqs. 1 and 2 is a vector of observations for further variables that shift the cost frontier and $\gamma'$ is the associated vector of coefficients to be estimated. In the empirical analysis the explanatory variables are chosen based on the premise that changes in $c(y_{it}, w_{it}, t)$ and $z_{it}$ explain changes in domestic retail cost that are outside company control. As a direct consequence, the cost efficiencies are within company control. This underlines the importance of justifying that the explanatory variables are, at least for the most part, not at the discretion of the companies, which is the focus of section 5.

As a final point on the general form of our domestic retail cost models, we note that $\theta_i$ in Eq. 2 denotes the random effect for the *ith* firm. In contrast to the time-varying idiosyncratic error, $\varepsilon_{it}$, as is standard, the random effects are modelled as time-invariant random errors. In other types of empirical analysis random effects are used to account for unobserved heterogeneity. In our analysis, however, we apply the SS approach to efficiency estimation, which, as discussed in detail in the next subsection, involves using the random effects to obtain consistent estimates of the time-invariant efficiencies of the firms.[10]

Instead of using random effects we could also apply the SS approach using fixed effects. We do not, however, estimate models with fixed effects as, unlike random effects, fixed effects can be correlated with the independent variables. As a result, when there is not a lot of variation over time in some of the explanatory variables, which is the case in our analysis (e.g., number of customers in a company's fixed supply area), the fixed effects may explain most of the variation in cost, leaving few (if any) statistically significant explanatory variables.

## 3.2    *Approaches to the Estimation of Efficiency*

From the OLS estimate of Eq. 1, we use the estimates of the time-varying residuals, $\hat{\varepsilon}_{it}$, to calculate relative time-varying efficiency estimates for the firms. For a cost function, as illustrated in figure 2 in the simple two-dimensional cost and single output space, in each period this involves shifting the estimate of Eq. 1 down to the position that represents the frontier. Hence why this efficiency estimator is referred to as Corrected OLS. In each period we specify the efficiency frontier by shifting the estimate of Eq. 1 down to the lowest (most negative) residual, which represents the best performing firm.

[Insert figure 2 about here]

Formally, the COLS time-varying efficiency estimate for the *ith* firm in period *t* is calculated as follows.

$$\widehat{Eff}_{it} = \exp\left[\min_{i}(\hat{\varepsilon}_{it}) - \hat{\varepsilon}_{it}\right], \tag{3}$$

where the estimate of inefficiency from the fitted Eq. 1 using logged variables is in square brackets, which we transform into an efficiency estimate by taking the exponential. From Eq. 3 we can see that the firm that represents industry best practice in period *t*, which need not of course be the same firm in each period, has an efficiency score of 100%. The efficiencies of all the other firms are then calculated relative to the performance of this frontier firm.

The SS efficiency estimator is based on the premise that all the heterogeneity is observed and is therefore accounted for by the variables on the right-hand side of Eq. 2, which is supported by this equation having a theoretical functional form. The random effects are therefore interpreted as time-invariant inefficiencies. SS efficiencies are calculated in similar fashion to the COLS time-varying efficiencies in figure 2. This involves specifying the frontier by shifting the Generalised Least Squares (GLS) estimate of Eq. 2 to the point of the lowest (most negative) random effect, which represents the best performing firm. Formally, the SS time-invariant efficiency estimate for the *ith* firm is calculated as follows.

$$\widehat{Eff}_{i} = \exp\left[\min_{i}(\hat{\theta}_{i}) - \hat{\theta}_{i}\right], \tag{4}$$

where because SS efficiencies are time-invariant the same firm represents industry best practice over the entire sample.

The COLS and SS deterministic frontier approaches are appealing because they support our headline contribution, which is to recommend how to address practical issues when implementing a general efficiency framework, in order to arrive at an empirical model that can inform public utility policy on benchmarking. This appeal is for three reasons. First, Ofwat (2018) also use the COLS and SS methods to

benchmark companies' domestic retail cost efficiencies, where, as we noted above, Welsh Water (2019, p. 7) confirm Ofwat's use of the former. Second, using the COLS and SS efficiency estimators is complementary because the former yields time-varying efficiency (i.e., transient performance), while the latter yields time-invariant efficiency (i.e., persistent performance). There would still be this complementarity if we were able to use the Cornwell *et al.* (1990) (denoted CSS from hereon) method to estimate the firms' efficiencies instead of COLS, where the CSS efficiency estimator is the time-varying counterpart of the SS estimator. It is not possible to use the CSS efficiency estimator for our empirical application because it is a big *T* estimator and *T* is small in our empirical analysis as we use five years of data.[11] Third, when the CMA (2015, pp. 73 -74) queried the complexity of the econometric modelling CEPA (2014) carried out for Ofwat at PR14 on the grounds that the results can be difficult to interpret, although the CMA did not explicitly refer to stochastic frontier models, to avoid our modelling being viewed as similarly complex, we take the view that the best course of action is not to use such models and to use the COLS and SS methods.

To provide an insight into the added complexity of stochastic frontier models, lets transform Eqs. 1 and 2 into such models. To do so, replace $\varepsilon_{it}$ in Eq. 1 with the following composed error terms. (i) $v_{it} + u_{it}$, where $v_{it}$ is our revised notation for the normally distributed idiosyncratic error and $u_{it}$ is half-normally distributed time-varying inefficiency, which yields the pooled Aigner *et al.* (1977) stochastic frontier model.[12] As an alternative to (i), (ii) $v_{it} + u_i$, where $v_{it}$ is as in (i) and $u_i$ is half-normally distributed time-invariant inefficiency, which yields the Pitt and Lee (1981) stochastic frontier model that accounts for the panel structure of the data. By replacing $\varepsilon_{it}$ in Eq. 2 with the composed error term in (i) above, we transform the model into the true random effects stochastic frontier model (Greene, 2005). In contrast to the assumption that heterogeneity is observed, which underlies the SS method, with this true stochastic frontier specification it is assumed that there is some unobserved heterogeneity, which is captured by the random effects.

In the above transformations of Eqs. 1 and 2 into stochastic frontier models, we assumed for convenience that (time-varying / time-invariant) inefficiency is half-normally distributed. Although this is the most common assumption for the inefficiency distribution in the literature, it is one of a number that can be made. Parmeter and Kumbhakar (2014) survey the literature and note that nine assumptions have been made about the inefficiency distribution (e.g., gamma, truncated normal and exponential). The added complexity of stochastic frontier models from our perspective is due to the number of potential assumptions about the inefficiency distribution, which can lead to the estimation of a large number of models. The COLS and SS deterministic frontier methods avoid such complexity because no assumption needs to be made about the inefficiency distribution. This is particularly appealing for our empirical analysis as it involves estimating a suite of models with different dependent and independent variables.

By not pursuing stochastic frontier models, we choose not to consider models that include determinants of the mean of inefficiency (e.g., Battese and Coelli, 1995), which is for two reasons. First, in addition to the technical requirement of employing a distributional assumption for inefficiency, this type of model adds a further layer of complexity. This is because calculating the elasticities of the determinants of the mean of inefficiency would not involve using the standard approach to calculate the elasticities of the drivers of the dependent variable in a regression equation (e.g., the approach we use to estimate the vectors of coefficients denoted $\gamma'$ in Eqs. 1 and 2). Instead, it would involve using the method that was specifically developed by Kumbhakar and Sun (2013) for stochastic frontier models.

Second, since the companies' cost efficiencies from our models are within company control, this suggests that including determinants of the mean of inefficiency in our models, which are similar to the type of models Ofwat use, is akin to an imperfectly informed regulator specifying the drivers of a regulated firm's cost efficiency. This is at odds with the classic asymmetry in information between the regulator and regulated firms because a water company knows far more about the drivers of its cost efficiency than Ofwat. Accordingly, by omitting determinants of the mean of inefficiency, the models are giving water companies the flexibility to identify the most effective strategies to improve their cost efficiencies.

## 4.     Further Practical Considerations when Implementing the Methodology

In light of the practical issues involved in our water industry application, it is necessary to adopt an approach to the practical implementation of the general methodology in the previous section that takes such issues into account. Accordingly, our approach to implement the general methodology highlights some transferable practical steps that often need to be adopted when applying frontier methodologies to public utility benchmarking.

To summarise our approach to the practical implementation of the general methodology, we estimate a suite of 16 empirical cost models using a general-to-specific approach to the modelling. The use of a suite of models is intended to address several trade-offs between different modelling choices. Such trade-offs across a suite of models should not be seen as a drawback, vis-à-vis the practice of having a single preferred model in some other types of applied microeconometric analysis. These trade-offs can instead be viewed as an advantage, as they will feature in the calculation of average efficiency across different model specifications. Equal weights have thus far been attached to company efficiencies from different model specifications when benchmarking companies' costs in the English and Welsh water industry. This is not unlike model averaging in the efficiency literature to obtain an average set of firms' efficiencies from a set of stochastic frontier models that make a range of distributional assumptions about inefficiency (Parmeter, *et al.*, 2019). The methods for model averaging though will typically not yield equal weights.

The suite of models we estimate is presented schematically in figure 3. From this figure we can see that the models differ according to:

- The two modelling approaches we adopt to distinguish between differences in the effects of single and dual service customers (i.e., water / sewerage-only and water and sewerage).

- The choice of dependent variable, which determines if the model represents a 'top-down' total retail cost / 'bottom-up' cost category approach (i.e., disaggregating total retail cost into bad debt and non-bad debt cost categories).

- Whether the estimation procedure accounts for the panel structure of the dataset by estimating time-invariant efficiency (i.e., the random effects models) or overlooks this feature of the data (i.e., the pooled OLS models).

[Insert figure 3 about here]

In addition to the practical concerns we have previously discussed that motivate us to adopt the Cobb-Douglas functional form, the implementation of the general methodology also involves adopting a practical approach to the following: (a) the determination of the efficiency frontier; (b) the choice of dependent variable; (c) model selection and statistical inference; and (d) heteroscedasticity and serial correlation. We now turn to discuss (a) - (d) in turn.

*(a)    Determination of the Efficiency Frontier*

With COLS, having estimated the cost model using OLS, the unexplained variation in costs, contained in the residuals, is taken to be a measure of inefficiency. The predicted values of any cost model therefore provide an estimate of what each company's costs would be at average levels of efficiency. The most efficient firms are those with the lowest actual costs relative to predicted costs – in other words, the lowest (most negative) residual. Efficiency estimates are then calculated as the difference in performance between all firms and the best performers.

Translating this general principle in a practical approach is, however, complicated by the fact that model residuals do not only comprise inefficiency. Residuals are also driven by statistical noise, measurement error and errors in the regression specification (Greene, 2003; Street, 2003). As such, calculating cost efficiency scores based on the firm with the lowest residual (i.e., the firm with the lowest costs relative to predicted costs) would overstate other firms' inefficiencies. This is analogous in a random effects model to random effects comprising not only inefficiency but also unobserved heterogeneity. Applying the SS approach therefore to calculate cost efficiency scores based on the firm with the lowest random effect would also overstate other firms' inefficiencies.

With COLS and the SS approach, we address the problem of overstating other firms' inefficiencies, by using a firm other than the one with the lowest residual / random effect as the basis for the calculation of the cost efficiencies.[13] We follow the approach adopted over recent years in public utility regulation of using the lower quartile residual / random effect and therefore use upper quartile performance as the benchmark.[14,15] Rather than directly address the problem that the inefficiency component of residuals / random effects is unobservable, this approach instead addresses the implication that using the lowest residual / random effect would generate excessively high estimates of firms' inefficiencies. It also avoids the need to make an arbitrary assumption about the split between inefficiency and other residual components, while retaining the feature that efficiency scores are calculated based on the performance of an actual firm.

In addition to the above adjustment to the efficiency frontier, we note that in practice regulators often make further adjustments to estimated efficiency scores. These include feasibility adjustments to the total efficiency challenge for each company and glide paths, which is where the implied efficiency challenge is imposed incrementally across the years of the price control period. The purpose of these adjustments is to ensure that the effective efficiency challenge imposed on each company *within the control period* is feasible. As such, they are not directly relevant to the question of what the actual gap to the frontier is. Rather, their purpose is to ensure that efficiency challenges are not so demanding that firms have short-term incentives to cut their variables costs in an inefficient manner.

Regulators have applied various combinations of these adjustments in practice. For the water industry in England and Wales, in table 1 we compare the adjustments at PR04 with those at PR14. At PR04 downward adjustments of 10% and 20% were applied to firms' residuals. In addition, rules were applied to ensure that efficiency gaps were not calculated with respect to atypical firms. This included a 'pragmatic turnover rule' that excluded very small firms from acting as the benchmark, and outlier treatment. At PR04 feasibility adjustments and glide paths were also applied. In contrast, at PR14 the only adjustment that was applied was to calculate the efficiency scores relative to the upper quartile company.

[Insert table 1 about here]

*(b)    Choice of Dependent Variable*

Since we use a theoretical functional form for our cost models, from production theory the dependent variable is a measure of the total cost associated with domestic retail activities. The dependent variable is not therefore a unit cost such as retail cost per connected property. The measure of the total cost that is used as the dependent variable will, of course, depend on factors such as the number of connected properties. Such factors should therefore be captured by the cost drivers (see section 5 for discussion of these drivers).

In an efficiency benchmarking context, there is flexibility as to the level of aggregation of costs included in the models. One could include the totality of household retail costs as the dependent variable or, alternatively, one could estimate separate models for different types of retail cost, such as bad debt, debt management, metering, etc. This is often referred to as a distinction between the use of bottom-up cost models, which require a number of sub-models to be combined to estimate efficiency, and top-down models of operations as a whole.

There is a strong prima facie case for estimating costs at the highest possible level of aggregation. If there are trade-offs between different cost categories, for example between bad debt and the amount spent on debt management, this will only be captured in an aggregated model. On the other hand, if there are no such trade-offs between cost categories, there is no loss of generality in estimating an aggregated model, provided all relevant cost drivers are included. In this case, an estimated coefficient will simply reflect the sum of the impacts of the individual explanatory variable on each cost category.

In practice, however, with a limited sample size it is advantageous to also estimate cost models at more disaggregated levels, as this tends to reduce the number of variables included in the model and therefore increases the number of degrees of freedom. In addition, in small samples it may be preferable to avoid introducing additional noise to the dependent variable by summing across a large number of cost categories, which can lead to erroneous coefficient estimates and/or a lack of statistical significance. Coefficients are also more likely to be statistically significant, and the coefficients are likely to be more intuitive as they reflect the impact of the relevant explanatory variable on a single cost category. Further, the coefficient estimates may be more intuitive in disaggregated models as their interpretation is likely to be more obvious.

Reflecting the trade-offs between different levels of aggregation, we use a combination of top-down and bottom-up models. The top-down models simply include the totality of household retail operating costs as the dependent variable. In deciding on the bottom-up models, the question was whether to create separate cost models for the most disaggregated cost categories (these categories being: doubtful debts, debt management, customer service, meter reading, depreciation and other operating expenditure), or to combine these into models of intermediate levels of costs.

We note that the disaggregated cost categories form two distinct groups of costs: debt-related costs, comprising doubtful debt and debt management, and the other cost categories which mostly relate to customer service functions. We consider that there are significant advantages of bottom-up models for these two groups of costs. This is because the main trade-offs between incurring different types of cost occur within these two groups, rather than between them. For instance, there are clearly trade-offs between debt management and doubtful debt costs, as achieving a reduction in defaults will tend to involve greater

expenditure on debt management activities. The trade-offs between bad debt related costs and other costs are, however, less obvious.

*(c)      Model Selection and Statistical Inference*

There are many cost drivers that potentially need to be controlled for in our models. While theory can provide a guide, the lack of precedent in the use of an econometric approach to household retail cost assessment means that the question of which of the many potential explanatory variables should be included within a model is essentially an empirical one. This is complicated by the small sample size, which also introduces a trade-off between including more variables in a model to minimise the risk of omitted variable bias and the robustness of the statistical inference. A further issue, encountered in wholesale cost modelling and likely also to apply to retail, is that the dominance of scale as a cost driver means that few other variables included in any model are statistically significant at conventional levels, even though, as a matter of principle, they clearly drive costs.

To address these concerns, we have combined a general-to-specific modelling approach to determine which explanatory variables are included within the models with a 'liberal' approach to statistical significance. Specifically, our general-to-specific approach retained variables that were correctly signed at significance levels of up to 15%. In addition, we also estimated alternative models that retained any explanatory variables that were correctly signed, but which had significance levels in excess of 15%. Retaining correctly signed explanatory variables that are not significant at conventional marginal levels is important, as their inclusion will certainly affect the residuals and one would also suspect the random effects. This will in turn affect the company efficiencies that are used to inform the caps that are placed on companies' prices for the duration of the 5-year control period (1 April 2020 - 31 March 2025).

*(d)      Heteroscedasticity and Serial Correlation*

The random effects models deal with concerns regarding heteroscedasticity and serial correlation in the residuals due to the presence of the random effects and via the GLS estimator. These issues, however, are a potential concern for the pooled OLS models. Options to deal with these issues include using robust standard errors or clustered standard errors. For the water retail cost case we consider, we do not see any disadvantages from using robust standard errors.

There are, however, several limitations from using clustered standard errors for the case in hand. A priori, it is unclear as to what the standard errors should be clustered on. Clustering by firm is a possibility, but it may be that clustering among water-only companies and water and sewerage companies would be more appropriate. Moreover, as part of Ofwat's consultation on its approach to the econometric cost modelling at PR19 (Ofwat, 2018), Glass and Glass (2018) question the use of clustered standard errors because of the

small sample size. They refer to a recent study by Ibragimov and Muller (2016), who conduct a Monte Carlo experiment to analyse this issue. Based on under rejection of the null hypothesis in simulations for some scenarios, they conclude that when clustered standard errors are used, a variable is more likely to be not significant when the sample size is small, than is actually the case.

The upshot is that the practical effect of using clustered standard errors would be similar to using random effects, in that the number of variables in the model after using the general-to-specific approach would be reduced. As such, clustered standard errors would reduce the advantages of estimating a suite of models that include both pooled OLS and random effects specifications. Our approach therefore is to use robust standard errors for the pooled models and the usual unadjusted standard errors for the random effects models.

## 5.    Motivation of the Independent Variables to Explain Water Retail Costs

In view of the novelty of econometric modelling of the domestic retail costs of English and Welsh water companies, and to ensure that we do not exclude valid explanatory variables, we considered the full range of potential cost drivers from first principles. In this respect, one should be mindful in efficiency benchmarking of the purpose of an econometric frontier model. The role of the model is to distinguish between variation in costs that occurs due to inefficiency – captured within the model by the residuals / random effects – and variation that occurs due to other factors that would cause one company to have higher efficient costs than another. This means that only drivers of differences in *efficient costs* should be included within the econometric model – as opposed to drivers of differences in costs that occur due to inefficiency.

In most cases, this is equivalent to saying that only explanatory variables that are fully outside of the companies' control should be included within the models. If a cost driver is within company control, then it follows that the company could reduce its costs by changing the level of the cost driver. For example, it would be inappropriate to include company-specific labour cost indices as an explanatory variable, as these are to some extent within each company's control. On the other hand, low traffic speed will tend to make meter reading more time-consuming, and therefore costlier. This is outside of company control, and therefore other things equal will tend to increase a firm's efficient costs.

Cost drivers that are within company control will tend to be correlated with company inefficiency. Including drivers that are within company control will therefore tend to reduce the model residuals and one would suspect also the random effects, and hence lower the implied gap between inefficient companies and the efficiency frontier. At the same time, including such variables would appear to increase a model's explanatory power, as they would account for more of the observed variation in costs – even though in reality they were *suppressing* inefficiency. As a consequence, it is important to ensure that only drivers of

differences in efficient costs are included *at the start of the modelling process,* i.e., when constructing the generalised model that represents the starting point for the general-to-specific modelling. This is because there is no way, ex post, to determine whether a variable should have been excluded on the grounds that it is within company control – and indeed, the inclusion of such variables may give the appearance of superior model fit.

The exception to the requirement that explanatory variables are outside of company control relates to cost drivers that explain differences in firms' efficient costs *at the efficiency frontier*. These drivers may be within company control and tend to relate to non-price aspects of the service that firms provide to customers, such as service quality. For companies at the frontier, where one company's higher costs are due to better service quality, these costs cannot be lowered without also reducing quality. In other words, a cost driver that is within company control may only be included within the generalised model to the extent that it measures trade-offs that firms face at the efficiency frontier.

In practice, of course, the distinction between variables that are within and outside company control is likely to be a matter of degree. Although it is domestic water retail that we analyse, to illustrate consider the case of wholesale water, where bills are determined by both the regulatory process, which is outside of company control, and the efficiency of the company's wholesale business. As such, a pragmatic approach to variable selection is required, that acknowledges that the 'ideal' variables relate to variation in efficient costs, while allowing for the fact that the lines between variables that are within / outside company control are blurred in practice. The practical criterion that we apply is therefore that the variables included are substantially outside of company control, and any connection between company actions and retail efficiency is 'distant'.

Having considered potential cost drivers from first principles – and excluded drivers that did not relate to efficient costs – we determined that the set of cost drivers summarised in table 2 were appropriate to include within the generalised cost models. In the remainder of this section we discuss the motivation for including each of these cost drivers.

[Insert table 2 about here]

*(i)      Scale and Scope*

The number of domestic retail customers that companies serve is the main driver of their total retail costs. For most of the retail functions that companies carry out, costs increase with the number of customers, across bad debt costs and other costs. For example, each additional customer that companies serve requires them to issue additional bills, respond to additional customer contacts, and increases the likelihood of a

default and the need to incur debt management costs. Serving additional metered customers also generates the need to undertake more meter readings.

A complication arises due to the distinction in domestic water retail between single service customers, who are served by one company for retail services relating to water and a different company for retail services relating to sewerage, and dual service customers, who are served by the same company for both types of retail service. This raises the possibility of retail cost savings for a company by serving the water and sewerage requirements of the same customer. For example, there may be savings involved in a single company dealing with customer service for both water and sewerage, relative to two companies dealing separately with customer service requirements for each.

As outlined in subsection 2.2, we use two approaches to account for these potential retail cost savings from providing a dual service to a customer. The first approach is to include the number of single and dual service customers as explanatory variables in the retail operating cost regression, which of course yields separable impacts of these variables. We denote the above variables in levels using the upper-case notation $S$, $D$ and $C$, respectively, and in natural logs using the corresponding lower-case notation. The coefficient on $d$ from the fitted regression equation is of course an elasticity, and represents the average percentage impact on $C$ of a 1% increase in $D$. Underlying this interpretation is the theoretical assumption that there is a 1% increase in dual customers for all the companies, which is not the case as some companies have no such customers.

In terms of the practical interpretation of the coefficient on $d$, this coefficient can be interpreted as an elasticity pertaining to only those companies that have dual service customers. This is because the model recognises that the 1% increase in dual customers is not coming from any variation in the dual customer observations for companies with only single service customers, as these observations remain constant at zero over the sample.

To reconcile the theoretical assumption that underlies elasticities and the practical interpretation, the second approach we use avoids the situation where there is, in theory, an increase in dual customers for single service companies. This involves including the total number of customers served and the number of single service customers as explanatory variables, where we denote the total number of customers in levels as $TS$ and in natural logs using the corresponding lower-case notation.[16] The coefficient on $ts$ from the fitted regression equation using the second approach is an elasticity, and represents the average percentage impact on $C$ of a 1% increase in $TS$. As $S$ is part of $TS$, the coefficient on $s$ is interpreted relative to that on $ts$. In other words, the coefficient on $s$ is capturing if there is any difference between the impacts of $S$ and $TS$ on $C$.

*(ii)    Population Transience*

Population transience – the movement of customers within and between company supply areas – has the potential to affect domestic retail costs in several respects. Any move will require companies to undertake additional account management activities, due to the need to record new customer details and open and close accounts. Higher population transience may also make debt recovery more difficult, as customers owing money become more difficult to trace, especially if they move outside company supply areas. Other things equal, this will make customer debt management more expensive, for a given level of customer debt (or vice versa).[17] This will lead to companies incurring higher debt management costs and / or higher doubtful debts. There may also be a relationship between customers' propensity to move and their probability of default or arrears. This would further increase companies' debt management / bad debt costs in totality.

To control for differences in population transience across companies, we use Office for National Statistics (ONS) data on population movements between local authorities. Ideally the data would measure all customer movements, whether they occur within or between local authorities, but this was the most granular level available. Data were available on both inflows into local authorities and outflows. As both are, in theory, relevant for retail costs, we therefore included the sum of inflows and outflows as a percentage of the total population.

*(iii)    Wage Levels*

As many areas of domestic water retail are labour intensive, there is a strong prima facie case for the inclusion of wage levels in the cost models. Many of the jobs in this sector, such as call centre staff, can, however, be outsourced. In fact, it is only jobs that relate to meter reading and debt collection that are required to take place within a company's supply area. As such, there is likely to be limited cross-company variation in the efficient level of wage-related costs. Related to this, in its draft PR19 methodology Ofwat states: *'With the exception of metering, which is around 5% of total retail costs, we consider that the impact of regional labour costs can be substantially mitigated or removed in retail activities'* (Ofwat, 2017b, p.184)*.*

Reflecting this, we consider it appropriate to focus on measures of wage levels for only those retail staff categories that cannot be outsourced. These categories primarily relate to roles that must be undertaken within company supply areas. Within water companies' retail businesses such roles primarily relate to meter reading and debt collection. As such, we use data on average weekly wages from the ONS Annual Survey of Hours and Earnings (ASHE) for the Standard Occupational Classification (SOC) "Sales Related Occupations", which includes debt collectors and meter readers.

*(iv)* *Service Quality*

As set out above, the cost drivers in the models should drive differences in *efficient cost* only. This is generally equivalent to a cost driver being outside of the company's control, though this is not the case for service quality. At the efficiency frontier, there is a cost-quality trade-off, meaning that service quality cannot improve without costs increasing, and costs cannot decrease without service quality declining. As such, without controlling for service quality, one would risk concluding that a company was more efficient than it really was, whereas in reality it was providing a lower quality service.

Several measures of service quality for water companies are available at the company-level, including percentage customer satisfaction and 'net promoter score'. For our purposes, however, a retail-specific measure of service quality is clearly preferable. Ofwat has historically used a measure called the 'Service Incentive Mechanism' (SIM) to provide companies with incentives to improve service quality. The calculation of SIM scores includes a measure of customer satisfaction (out of 5) with respect to customers' experiences of interactions with firms' billing functions. We note that in practice it may be difficult for customers to think of their experiences of firms' retail functions independently from their experiences of firms' wholesale functions. We nevertheless consider the SIM billing score to be the most appropriate of the available measures of service quality.

*(v)* *Customer Ability to Pay and Deprivation*

The extent to which a company's customers fall into arrears and default on their bills depends on the economic circumstances of the company's customer base. As such, economic circumstances are the primary driver of companies' bad debt related costs. This may be in relation to income deprivation, with customers on low income being more exposed to the financial shocks that would lead to bill defaults. It may also reflect other aspects of customers' financial circumstances, such as personal debt levels, which also leave customers more exposed to financial shocks. In addition, factors such as a high cost of living in particular areas could lead to a greater tendency for customers to default on their bills.

In principle, these economic circumstances have the potential to affect both doubtful debts and debt management costs. In fact, companies face a trade-off between these two types of expenditure – and customer deprivation and ability to pay affect this trade-off. Other things equal, a firm seeking to reduce its doubtful debt costs will need to increase its expenditure on debt management activities, such as generating repayment plans for customers in financial distress. On the other hand, a firm is unlikely to be able to reduce its (efficient) debt management costs without increasing its doubtful debt costs, for example due to employing fewer debt collectors or drawing up fewer repayment plans. In the face of, for example, an increase in deprivation, a firm would therefore be faced with a choice of 'taking the hit' of higher doubtful debt costs, or trading this off against higher debt management expenditure.

As we set out above, the explanatory variables in the models should ideally be outside of company control. This has important implications for how deprivation and customer ability to pay are captured in the models. Data are available, for instance, from credit reference agencies on default rates by location. These default rates will, however, be affected by debt management activities – meaning that a local area could appear to have a low default rate, but this could be the result of higher debt management activities. As such, we consider that it is preferable to include variables that directly measure deprivation and customer ability to pay in a company's supply area.

The issue is further complicated by both the 'breadth' of the concepts of deprivation and ability to pay and associated difficulties in measuring them. This is most easily demonstrated with reference to the 'official' measure of deprivation in neighbourhoods in England, the Index of Multiple Deprivation (IMD). This is a composite index across several 'domains', each representing a different aspect of deprivation. These domains are: income; employment; education, skills and training; health deprivation and disability; crime; barriers to housing and services; and living environment. In principle the income deprivation domain would appear to be of most relevance to bad debt related costs, but the other domains are also potentially of relevance.

While there would clearly be advantages in principle from using an average overall IMD score across each company's supply area as an explanatory variable, there are practical difficulties with this. The overall IMD score is not available for Welsh neighbourhoods. A separate IMD for Welsh neighbourhoods is available, but it is not calculated on a comparable basis. The IMD income domain, which measures the proportion of the population entitled to income-related benefits in a neighbourhood, is, however, broadly comparable between the English and Welsh indices. As such, we include an explanatory variable for deprivation and ability to pay comprising the average IMD income domain across neighbourhoods in each company's supply area.

There is, however, the additional complication that the IMD income domain scores are only produced every five years. As such, we also include broader measures of deprivation and customer ability to pay that vary from year-to-year. To measure broader financial cost pressures with respct to housing, we include the ratio of the median house price to average income in a company's supply area. In addition, to measure the extent of financial distress, we include a measure of property repossessions in the supply area.

*(vi)      Wholesale Bill Size*
In relation to debt related costs, higher water bill amounts will lead to greater amounts of doubtful debt, for a given propensity of customers to default. Water bills relate both to retail and wholesale activities. The former is clearly, to some extent, within management control (noting we are considering

the 'retailer' here), whereas the latter will be fixed, at least from the perspective of the retail business. As such, it is the wholesale bill amount that is relevant in this context.

*(vii)   Metering*

Other things equal, the retail costs companies incur will be higher for customers that have a water meter, who pay water bills in proportion to their use, than for unmetered customers, who pay a flat bill based on their property's rateable value, regardless of their water usage. This is because companies incur additional cost due to the need to send operatives to physically read these customers' meters. As such, the non-bad debt component of retail costs will increase with the proportion of a company's customers that are metered. We therefore include the proportion of customers that are metered as an explanatory variable.

Several factors specific to company supply areas also have the potential to make it more expensive to serve metered customers. We therefore include two explanatory variables to capture traffic speed in the company's supply area and the proximity of metered properties, as these factors will affect the time required to read meters. The latter is a metered household density variable, calculated as the ratio of the number of metered households to water mains length. We use mains length as the denominator (rather than, say, supply area) as we consider this a better indicator of distances between properties on each company's supply network.

Finally, the nature of housing stock has the potential to help or hinder operatives when they read meters. For instance, it may be more difficult to access meters in large detached houses than in a block of flats where numerous meters are co-located – though the latter may be more difficult where an operative has to call a concierge to gain access. Reflecting these considerations, we also include a variable that measures the proportion of housing stock in each company's supply area that are flats and maisonettes.

## 6.     Data

To estimate the models, we use unbalanced panel data for 18 companies in the English and Welsh water industry for the period 2012/13 - 2016/17. These companies comprise the 17 WASCs and WOCs that are currently in the industry, which we have data on for all five years of the sample, plus one company that is not in the dataset for all five years.[18] We explored including 2011/12 in our dataset, as this is also part of the period that Ofwat considers relevant to PR19, but we ultimately omitted this year. This is because there was a tendency for the cost data to appear materially different to other years, which raised concerns that this may be due to accounting separation issues.

The dataset was assembled by adding to data that is available from industry sources, such as the *water company data share*, data for further variables that we collected from a range of other sources. See table 3 for descriptions of the variables and the data sources and table 4 for the summary statistics.

## 7.        Empirical Cost Models and Estimated Efficiencies

Using the data described above for the 18 companies, in tables 5 and 6 we present the 16 estimated cost models after the general-to-specific modelling. Specifically, tables 5 and 6 present models when the potential retail cost savings from providing a dual service is accounted for using the first and second approaches we propose.

For the 17 companies in the last year of the sample, in table 7 we present the efficiency estimates (and their rankings) from the 16 models based on a benchmark of upper quartile performance.[19] For companies that outperform this benchmark the efficiencies in table 7 are negative. In these cases, we interpret the companies as being at the frontier, rather than having inefficiently low costs. In table 8, for each approach to account for the potential retail cost savings from providing a dual service, we report the average efficiency gaps (and their rankings) to the upper quartile performance benchmark across subsets of models (top-down models and bottom-up models). Recall that top-down models are those with total retail operating costs as the dependent variable, and the bottom-down models have bad debt costs and non-bad debt costs as the dependent variables. To obtain the results in table 8 we calculated equally weighted average efficiencies across the relevant subset of models.

[Insert tables 5-8 about here]

Ahead of a more detailed discussion of the empirical results in the next section, inspection of the results in tables 5 and 6 indicates that some variables in the dataset were dropped from all the reported models during the general-to-specific modelling. We can also see from a comparison of the models that there were a number of cases in the general-to-specific modelling where a variable is dropped from one model but retained in another. This highlights nicely the impact of estimating a suite of models with three different dependent variables using estimation procedures that either account for or overlook the panel structure of the dataset. The upshot of the general-to-specific modelling is that all the reported models are parsimonious, which is appropriate for our small sample and is also in line with a recommendation of the CMA in its redetermination of Bristol Water's price control at PR14 (CMA, 2015).

Turning to the diagnostic analysis of the fitted models. The results of the cross-sectional Ramsey RESET test on each pooled OLS model (models 1 - 4 and 9 - 12 in tables 5 and 6) accepts (at the 5% level or lower), for all but two of these models (9 and 12), the null hypothesis that there is no general functional form misspecification. The panel data version of the RESET test on each random effects model (models 5 - 8 and 13 - 16 in tables 5 and 6), accepts the above null (at the 5% level or lower) for all eight fitted models.

Since the RESET test is a test of omitted non-linear combinations of the variables in a fitted model (such as squared and interactions terms), but is not a test for other omitted variables, the results of the RESET test for models 9 and 12 illustrates the potential trade-off between the Cobb-Douglas and translog functional forms. This is because the presence of squared and interaction terms in a translog model can be the reason for accepting the null of the RESET test. We do, however, think it is important to take heed of the CMA's reservations about the translog functional form at PR14, and failing to accept the null of the RESET test for only two of the 16 models would seem quite a small price to pay for doing so.

We use the variance inflation factor (VIF) as the diagnostic check for multi-collinearity in the OLS models. A rule of thumb is that multi-collinearity is a concern if the VIF is greater than 10 (Curto and Pinto, 2011). Multicollinearity is not therefore a concern in the OLS models because the VIF ranges from 1.33 – 5.96.[20]

# 8. Discussion of the Empirical Results and Policy Implications

## 8.1 *Discussion*

The coefficients on all the variables in the reported models in tables 5 and 6 have the correct signs and there are no cases where the magnitude of a coefficient appears implausible. This is in line with the CMA's recommendation when redetermining Bristol Water's price control at PR14 (CMA, 2015, pp. 72 - 73). This recommendation was a response to CEPA (2014) (acting on behalf of Ofwat) reporting models where the signs and / or magnitudes of some coefficients where counter to expectations. For instance, as the CMA note in the above report, CEPA report base expenditure models that imply diseconomies of scale with respect to the size of a company's customer base, when one would expect to find constant or increasing returns. On the other hand, we find that the coefficient on the total number of customers (models 9 - 16) is generally around one (i.e., constant returns to scale), which lies within the expected range of values.

In terms of the goodness-of-fit of the models, the overriding pattern is that the adjusted $R^2$ is always higher for a model that uses our second approach to account for cost savings from providing a dual service, vis-à-vis the corresponding model that uses our first approach to account for such savings (e.g., model 9 versus model 1). We also observe that the sum of the coefficients on the number of single and dual service customers (models 1 - 8) is always less than one. Comparing this to the coefficient on the total number of customers, suggests that there are some retail costs that are shared between single and dual service customers that are not being picked up by models 1 - 8 (i.e., by the separate single and dual customer variables). In models 1 - 8 we can see that the magnitude of the coefficient on the number of dual customers is far less than the coefficient on the number of single service customers. This finding is intuitive as it is consistent with retail cost sharing between service lines at dual supply companies.

As we previously noted, the coefficient on the number of single service customers in models 9 - 16 must be interpreted relative to the coefficient on the total number of customers. This is because single service customers are, of course, part of the total customer base. Suppose based on models 9-16, there is a marginal increase in the total number of customers that corresponds to a marginal increase in both single and dual service customers. Such an increase will lead to an increase in total retail operating / bad debt / non-bad debt costs. In models 9 - 16, however, the coefficient on the number of single service customers is negative, significant and small compared to the coefficient on the total customers served. This suggests that there is a relatively small retail cost saving associated with single service customers vis-à-vis the total number of customers, where the difference between the two is, of course, the number of dual service customers. Such a relatively small cost saving points to there being retail cost sharing between the service lines of dual supply companies.

We dropped sales-related wages from all the reported models because even at the 15% level it was generally not significant, which we considered to reflect the low proportion of retail-related jobs that cannot be outsourced. Billing SIM (i.e., service quality) and the house price to income ratio were also dropped from all the models as the coefficients generally had the wrong sign. In the case of the billing SIM variable, this is conceivably because for companies that are not on the efficient frontier cost efficiency is correlated with service quality.

Regarding explanatory variables that are dropped from some models and retained in others, variables associated with bad debt (average wholesale bill and deprivation measures) are often retained, while the metering variables feature less. We suggest this reflects the fact that bad debt related costs make up a larger part of total retail operating costs than metering costs. In line with this, even in models that relate specifically to non-bad debt costs, the coefficients on the metering variables are always small and sometimes not significant (albeit correctly signed).

In contrast to the cases where we retain a metering variable when its coefficient is small and insignificant at conventional levels (1%, 5% or even 10%), there are a few insignificant variables that we retain with non-negligible coefficients, such as average wholesale bill in model 5 and peak traffic speed in model 15. Including such variables on the grounds they are economically important, as opposed to statistically important, is a key feature of the reported models, which would not be as important in some other types of applied microeconomic analysis. This is because it is the sign and magnitude of the coefficient on a variable, and not its statistical significance, that affects the residuals / random effects, and consequently and importantly the efficiency estimates.

To aid the discussion of the efficiency results, we summarise the three methodological differences between the models in tables 5 and 6 as follows. (i) Some models in these tables have different dependent and

independent variables. (ii) The pooled OLS models overlook the panel structure of the data, while the random effects GLS models take account of it. (iii) From a pooled OLS model we estimate time-varying efficiency (i.e., transient performance), while from a random effects model we estimate time-invariant efficiency (i.e., persistent performance). In light of such methodological differences, it is not surprising that there is a lot of company-level evidence in table 7 of non-negligible variations in the efficiencies from different models. Such variations, however, should not be seen as a drawback. Rather these variations highlight the benefits of estimating a suite of frontier models. These benefits are felt in table 8 because these variations in efficiencies are used in the calculation of the average efficiency gaps across subsets of models. Even after this averaging there is company-level evidence in table 8 of non-negligible differences in the average efficiency gaps across different subsets of models. That said, we can see for individual companies in table 7 that there is a robustness between the efficiency rankings from some pairs of models. Furthermore, such robustness for individual companies is particularly evident in table 8 between the rankings of the average efficiency gaps across top-down and bottom-up models that use the same approach to capture cost savings from dual supply.

We now turn to explore why there is a such a marked difference for some companies between the average efficiency gaps from top-down and bottom-up models that use the same approach to account for cost savings from dual supply (see table 8). These differences could conceivably be because of trade-offs between bad debt costs and non-bad debt costs, which would not be captured when these costs are modelled separately. For instance, it could be that there is a trade-off between minimising bad-debt costs and minimising non-bad debt costs, if pursuing bad debt leads to increased customer contacts.

### 8.2 *Policy Implications*

It follows from the nature of our analysis that there some important policy implications, which we begin discussing by revisiting the marked differences for some companies between the average efficiency gaps from top-down and bottom-up models. If there are no trade-offs between bad debt costs and non-bad debt costs, in theory, there should be no difference between the average efficiency gaps from top-down and bottom-up models. This is because a coefficient on a variable in a top-down model should equal the sum of the coefficients on the variable in the corresponding bottom-up models. There are, however, likely to be companies with non-negligible trade-offs between bad debt costs and non-bad debt costs and, even if there are not, in practice, the coefficient on a variable in a top-down model may not be the sum of the corresponding coefficients in the bottom-up models. The latter can be because a variable is a driver of a disaggregated retail cost component, but not sufficiently so for the variable to also explain variations in total retail cost (or vice versa). The implication is that top-down and bottom-up models would have different sets of explanatory variables, which there is a lot of evidence of in the models we report. For example, in the pooled OLS model for bad debt costs in table 5 population flow is retained and has a

significant positive non-negligible impact, whereas it was omitted from the pooled OLS models for total retail operating costs in the same table.

To avoid using for some companies markedly different average efficiency gaps from top-down and bottom-up models to set the retail price controls, we therefore suggest using only the average gaps from the top-down models. This is not to say that estimating the bottom-up models was not useful because without estimating these models we would not have been aware of such marked differences in these gaps. Moreover, using the average efficiency gaps from the top-down models to set the retail price controls is in line with the previously discussed prima facie case for such models.

The efficiency measurement literature was for some time of the view that a choice needed to be made between estimating time-invariant / time-varying efficiency. More recently, estimators have been developed for models that include both time-invariant and time-varying efficiencies (Tsionas and Kumbhakar, 2014; Filippini and Greene, 2016; Glass and Kenjegalieva, 2019; Douch *et al*., 2020; Glass *et al*., 2020). Now although these complex models are outside the scope of this paper because, as we will expand on in the concluding section, our focus is on the simpler frontier models that are applied in industry, these models highlight the importance of considering both time-invariant and time-varying efficiencies. In particular, these models recognise that performance measurement should not viewed in the way the efficiency literature evolved by moving from calculating time-invariant efficiency to computing its time-varying counterpart because the latter is the more appropriate measure of performance in most situations. Rather these models recognise the benefits of both types of efficiency because they measure different aspects of performance – time-varying efficiency measures transient performance, while time-invariant efficiency measures persistent performance. For this reason, we favour considering both types of efficiency, but because we place the emphasis on particularly tractable frontier models that are appealing to industry, we measure the two types of efficiencies using different models (COLS and random effects models).

The reasoning for our interest in water companies' time-varying efficiency is obvious, but our practical motives for considering companies' time-invariant efficiencies is less so. Time-invariant inefficiency can be the result of rigidities in fixed assets and technology, and inertia in processes and internal organisation. When econometric efficiency benchmarking is used to set the water companies' retail price controls, we posit that time-invariant efficiency should be included in the calculation of the equally weighted average efficiency across the top-down models (as in table 8). This is because it would send a clear message to companies that they should not just focus on improving their time-varying efficiencies, as this is where they view the 'quick wins' to be, and should also concentrate on addressing their persistent underperformance by improving their time-invariant efficiencies.

Having made the case for top-down models that include random effects models, we explain the important role that econometric efficiency benchmarking plays in Ofwat's determination of the companies' retail price controls. While Ofwat conducts its econometric retail cost efficiency benchmarking, companies independently conduct such benchmarking to inform discussions with Ofwat during the PR19 consultation period. Using a preferred suite of retail cost models, which will comprise well-specified models from different justifiable perspectives, for each company Ofwat calculates unweighted average annual forecasts across the models for the 5-year price control. To set the companies' retail cost allowances for the price control, Ofwat uses an unweighted average of the efficiencies from its suite of preferred models to scale up the forecasts of each company's efficient costs to the level of upper quartile performance in the industry. Based on a company's cost allowance, Ofwat determines the remuneration the company should be allowed for its retail activities, which represents the retail price control.

Although the above represents a summary of the process Ofwat intended to follow to set the companies' retail price controls at PR19, a slightly different approach was adopted. This was because the companies effectively followed the above process. Based on their own efficiency benchmarking, companies proposed retail price controls in the final versions of their business plans that were acceptable to Ofwat. The importance therefore of well-specified retail cost frontier models to Ofwat and water companies is clear. This is because if the models are misspecified the efficiency estimates will be biased, which will lead to retail cost allowances and thus retail price controls that are lower / higher than should be the case. If the remuneration to a company for its retail activities is lower than it should be, the company may not be able to adequately fulfil some of its public retail responsibilities (e.g., customer services such as dealing with enquiries and complaints).

## 9.    Concluding Remarks

The overriding contribution of this paper is to provide an in-depth study of the use of frontier models in industry. To this end, this paper is the first journal article to analyse the econometric benchmarking of the retail cost efficiencies of English and Welsh water companies at PR19. Notwithstanding the academic contribution to the field of efficiency measurement of technical state-of-the-art stochastic frontier models, such models are largely not used by UK public utility companies and their regulators for efficiency benchmarking purposes. Rather, as we set out in this paper, there is a tendency for more tractable frontier models to be used in UK public utility industries.

To provide the best possible demonstration of the use of frontier models in the English and Welsh water industry to benchmark companies' retail cost efficiencies at PR19, when choosing our econometric approach, we rather successfully predict the approach that Ofwat (2018) adopts. This was important because this paper emerged out of work for water companies (Economic Insight, 2018), who independently

of Ofwat's retail cost efficiency benchmarking, conduct their own efficiency modelling to inform discussions with Ofwat during the PR19 consultation period. At the same time, however, there are some subtle differences between Ofwat's approach and our approach, such as how we capture the retail savings from dual supply.

With regard to the policy implications from our empirical analysis, we made the case for a top-down modelling approach that includes estimating random effects models. We then highlighted the important role of the models in the determination of the companies' retail price controls, and in doing so emphasised the implications of the model specifications for the controls. Ultimately, Ofwat did not draw on its econometric benchmarking of the companies' retail cost efficiencies to set the retail price controls at PR19. This is because Ofwat set these controls at the levels the companies proposed in the final versions of their business plans. Similar econometric benchmarking of companies' retail cost efficiencies to that reported here would have been undertaken by individual companies to arrive at the proposed levels of their retail price controls. Ofwat choosing to accept the companies' proposed retail price controls was a good outcome all round, as this indicated that Ofwat and the companies thought the controls represented the right balance between company remuneration and value for money for customers. That said, at the time of writing, four companies (Anglian, Bristol, Northumbrian and Yorkshire) were appealing to the CMA against Ofwat's final determination of their PR19 price controls. Since their proposed retail price controls were accepted by Ofwat, the focus of their appeals will be on other elements of Ofwat's determination. An appeal though involves a complete redetermination, so the CMA is not prohibited from adjusting the four companies' retail price controls that Ofwat approved.

## References

Aigner, D., C. A. K. Lovell and P. Schmidt (1977): 'Formulation and estimation of stochastic frontier production function models'. *Journal of Econometrics*, vol. 6, pp. 21 - 37.

Ashton, J. (2003): 'Capital utilisation and scale in the English and Welsh water industry'. *Services Industries Journal*, vol. 23, pp. 137 - 149.

Battese, G. E. and T. J. Coelli (1995): 'A model for technical inefficiency effects in a stochastic frontier production function for panel data'. *Empirical Economics*, vol. 20, pp. 325 - 332.

Cambridge Economic Policy Associates (CEPA) (2014): *Cost assessment – advanced econometric models.* Final Report to Ofwat.

Carvalho, P. and R. C. Marques (2014): 'Computing economies of vertical integration, economies of scope and economies of scale using partial frontier nonparametric methods'. *European Journal of Operational Research*, vol. 234, pp. 292 - 307.

Chung, W. (2011): 'Review of energy-use performance benchmarking methodologies'. *Applied Energy*, vol. 88, pp. 1470 - 1479.

Competition and Markets Authority (CMA) (2015): *Bristol Water plc: A reference under section 12(3)(a) of the Water Industry Act 1991- Report.*

Cornwell, C., P. Schmidt and R. C. Sickles (1990): 'Production frontiers with cross-sectional and time-series variation in efficiency levels'. *Journal of Econometrics*, vol. 46, pp. 185 - 200.

Correia, T. and R. C. Marques (2011): 'Performance of Portuguese water utilities: How do ownership, size, diversification and vertical integration relate to efficiency?' *Water Policy*, vol. 13, pp. 343 - 361.

Corton, M. (2003): 'Benchmarking in the Latin American water sector: The case of Peru'. *Utilities Policy*, vol. 11, pp. 133 - 142.

Cubbin, J. (2004): *Assessing Ofwat's efficiency econometrics: A report for Water UK*.

Cubbin, J. and G. Tzanidakis (1998): 'Regression versus data envelopment analysis for efficiency measurement: An application to the England and Wales regulated water industry'. *Utilities Policy*, vol. 7, pp. 75 - 85.

Curto, J. D. and J. C. Pinto (2011): 'The corrected vif (cvif)'. *Journal of Applied Statistics*, vol. 38, pp. 1499 - 1507.

Douch, M., A. J. Glass and K. Kenjegalieva (2020): 'Uncovering spatial productivity centers using asymmetric bidirectional spillovers'. *European Journal of Operational Research*, vol. 285, pp. 767 - 788.

Economic Insight (2018): *Household retail cost assessment for PR19- Final report for Bristol Water and Wessex Water*. Available at:
https://www.ofwat.gov.uk/wp-content/uploads/2018/03/BRL-WSX_Pelican-Household-Retail-Cost-Assessment-Final-STC-16-03-18.pdf

Economic Insight (2017): *Household retail margins at PR19- A report for Bristol Water and Wessex Water*. Appendix 8.3.E of Wessex Water's PR19 Business Plan. Available at:
https://www.wessexwater.co.uk/-/media/files/wessexwater/corporate/strategy-and-reports/business-plan/0803e--economic-insight-report-on-household-retail-margins-at-pr19.pdf

Estache, A. and M. Rossi (2002): 'How different is the efficiency of public and private water companies in Asia?' *World Bank Economic Review*, vol. 16, pp. 139 - 148.

Filippini, M. and W. Greene (2016): 'Persistent and transient productive inefficiency: A maximum simulated likelihood approach'. *Journal of Productivity Analysis*, vol. 45, pp. 187 - 196.

Gaganis, C., F. Pasiouras and A. Tsaklanganos (2013): 'Taxation and bank efficiency: Cross-country evidence'. *International Journal of the Economics of Business*, vol. 20., pp. 229 - 244.

Glass, K. and A. Glass (2018): 'Comments on 'Cost assessment for PR19: A consultation on econometric cost modelling'.' Available at:
https://www.ofwat.gov.uk/wp-content/uploads/2018/03/KA-Glass-consultation-comments.pdf

Glass, A. J. and K. Kenjegalieva (2019): 'A spatial productivity index in the presence of efficiency spillovers: Evidence for U.S. banks, 1992-2015'. *European Journal of Operational Research*, vol. 273, pp. 1165 - 1179.

Glass, A., K. Kenjegalieva and R. C. Sickles (2016): 'A spatial autoregressive stochastic frontier model for panel data with asymmetric efficiency spillovers'. *Journal of Econometrics*, vol. 190, pp. 289 - 300.

Glass, A., K. Kenjegalieva and R. C. Sickles (2014): 'Estimating efficiency spillovers with state level evidence for manufacturing in the U.S.' *Economics Letters*, vol. 123, pp. 154 - 159.

Glass, A. J., K. Kenjegalieva and T. Weyman-Jones (2020): 'The effect of monetary policy on bank competition using the Boone index'. *European Journal of Operational Research*, vol. 282, pp. 1070 - 1087.

Greene, W. H. (2003): *Econometric Analysis*. New Jersey: Pearson Education.

Greene, W. H. (2005): 'Reconsidering heterogeneity in panel data estimators of the stochastic frontier model'. *Journal of Econometrics*, vol. 126, pp. 269 - 303.

Greene, W H. (2008): *The econometric approach to efficiency analysis*. In *The Measurement of Productive Efficiency and Productivity Growth*, Fried H. O., C. A. K. Lovell and S. S. Schmidt (Eds.). Oxford: Oxford University Press.

Guerrini, A., G. Romano and B. Campedelli (2013): 'Economies of scale, scope, and density in the Italian water sector: A two-stage data envelopment analysis approach'. *Water Resources Management*, vol. 27, pp. 4559 - 4578.

Guerrini, A., G. Romano, C. Leardini and M. Martini (2015): 'The effects of operational and environmental variables on efficiency of Danish water and wastewater utilities'. *Water*, vol. 7, pp.3263 - 3282.

Hjalmarsson L., S. C. Kumbhakar and A. Heshmati (1996): 'DEA, DFA and SFA: A comparison'. *Journal of Productivity Analysis*, vol. 7, pp. 303 - 327.

Ibragimov, R. and U. K. Muller (2016): 'Inference with few heterogeneous clusters'. *Review of Economics and Statistics*, vol. 98, pp. 83 - 96.

Kumbhakar, S. C. and K. Sun (2013): 'Derivation of marginal effects of determinants of technical inefficiency'. *Economics Letters*, vol. 120, pp. 249 - 253.

Kumbhakar, S. C. and E. G. Tsionas (2014): 'Firm heterogeneity, persistent and transient technical efficiency: A generalized true random-effects model'. *Journal of Applied Econometrics*, vol. 29, pp. 110 - 132.

Mamatzakis, E., A. N. Kalyvas and J. Piesse (2013): 'Does regulation in credit, labour and business matter for bank performance in the EU-10 economies'. *International Journal of the Economics of Business*, vol. 20., pp. 341 - 385.

Marques, R. C. and K. De Witte (2011): 'Is big better? On scale and scope economies in the Portuguese water sector'. *Economic Modelling*, vol. 28, pp. 1009 - 1016.

Molinos-Senante, M. and A. Maziotis (2019a): 'Estimation and evaluation of productivity change and its drivers in the English and Welsh water sector: A stochastic cost frontier approach'. *Urban Water Journal*, vol. 16, pp. 625 - 633.

Molinos-Senante, M. and A. Maziotis (2019b): 'Cost efficiency of English and Welsh water companies: A meta-stochastic frontier analysis'. *Water Resources Management*, vol. 33, pp. 3041 - 3055.

Molinos-Senante, M. and A. Maziotis (2018): 'Assessing the influence of exogenous and quality of service variables on water companies' performance using a true-fixed stochastic frontier approach'. *Urban Water Journal*, vol. 15, pp. 682 - 691.

Ofwat (2008): *Setting price limits for 2010-15: Framework and approach.*

Ofwat. (2016): *RAG 4.06- Guideline for the table definitions in the annual performance report.*

Ofwat. (2017a): *Delivering Water 2020: Our final methodology for the 2019 price review, Appendix 11: Securing cost efficiency.*

Ofwat. (2017b): *Delivering Water 2020: Consulting on our methodology for the 2019 price review.*

Ofwat. (2018): *Cost assessment for PR19: A consultation of econometric cost modelling.*

Ofwat. (2019): *Technical Appendix 2 - Securing cost efficiency.*

Office of Rail Regulation (ORR) (2013): *PR13 efficiency benchmarking of Network Rail using LICB.*

Parmeter, C. F. and S. C. Kumbhakar (2014): 'Efficiency analysis: A primer on recent advances'. *Foundations and Trends in Econometrics*, vol. 7, pp. 191 - 385.

Parmeter, C. F., A. T. Wan and X. Zhang (2019): 'Model averaging estimators for the stochastic frontier model'. *Journal of Productivity Analysis*, vol. 51., pp. 91 - 103.

Pitt, M. M. and L-F. Lee (1981): 'Measurement and sources of technical inefficiency in the Indonesian weaving industry'. *Journal of Development Economics*, vol. 9, pp. 43 – 64.

Ruggiero, J. (2007): 'A comparison of DEA and the stochastic frontier model using panel data'. *International Transactions in Operational Research*, vol. 14, pp. 259 - 266.

Schmidt, P. and R. C. Sickles (1984): 'Production frontiers and panel data'. *Journal of Business and Economic Statistics,* vol. 2, pp. 367 - 374.

Street, A. (2003): 'How much confidence should we place in efficiency estimates?' *Health Economics*, vol. 12, pp. 895 - 907.

Tanna, S., F. Pasiouras and M. Nnadi (2011): 'The effect of board size and composition on the efficiency of UK banks'. *International Journal of the Economics of Business*, vol. 18., pp. 441 - 462.

Thanassoulis, E. (2000): 'The use of data envelopment analysis in the regulation of UK water utilities: Water distribution'. *European Journal of Operational Research*, vol. 126, pp. 436 - 453.

Thanassoulis, E. (2002): 'Comparative performance measurement in regulation: The case of English and Welsh sewerage services'. *Journal of the Operational Research Society*, vol. 53, pp. 292 - 302.

United Utilities (2018): *PR19 Business plan, Chapter 7 – Securing efficient cost.* Available at:
https://www.unitedutilities.com/corporate/about-us/our-future-plans/our-proposed-business-plan/

Vencappa, D., P. Fenn and S. Diacon (2013): 'Productivity growth in the European insurance industry: Evidence from life and non-life companies'. *International Journal of the Economics of Business*, vol. 20., pp. 281 - 305.

Welsh Water (2019): *PR19 Business plan supporting information, IAP response (Ref B2.8.WSH.CE.A1) - Cost assessment methodology.* Available at:
https://www.welshwater.com/cy-GB/Library/-/media/F5993E40FE444536982B1A426AA4C519.ashx

Wessex Water (2018): *PR19 Business plan appendices, Section 8 – Securing cost efficiency, Appendix 08.03.B – Peer review of residential retail cost assessment.* Available at:
https://www.wessexwater.co.uk/corporate/strategy-and-reports/business-plan-2020/appendices
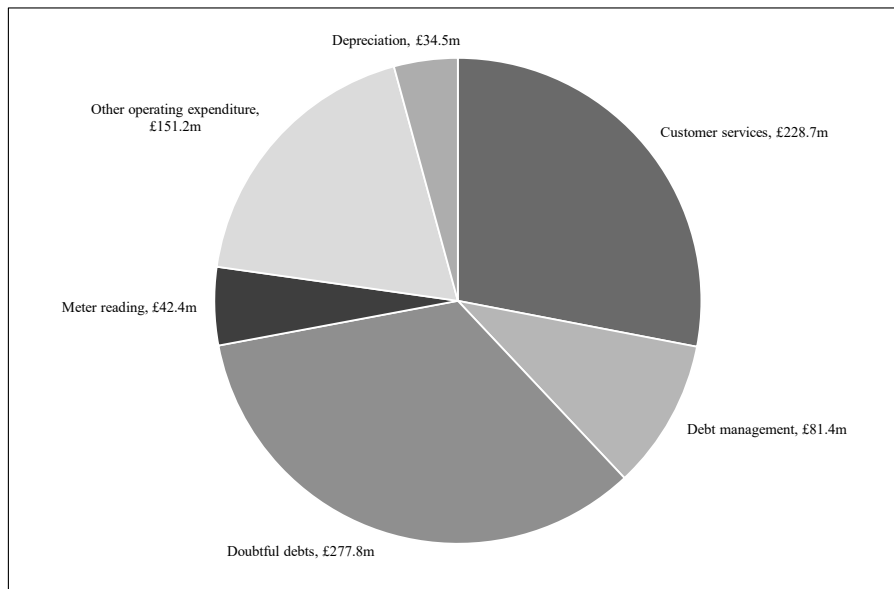


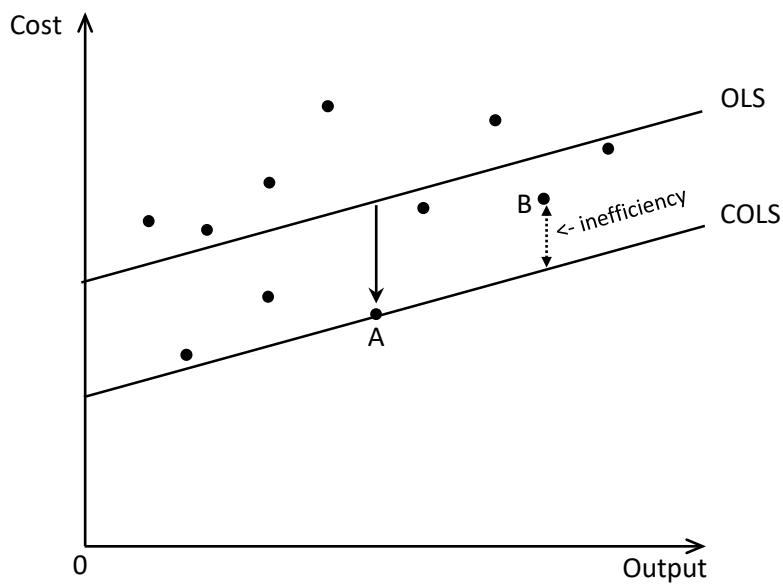Figure 1: Domestic retail cost categories and 2016/17 industry expenditure



Figure 2:  Illustration of inefficiency using Corrected Ordinary Least Squares (COLS)
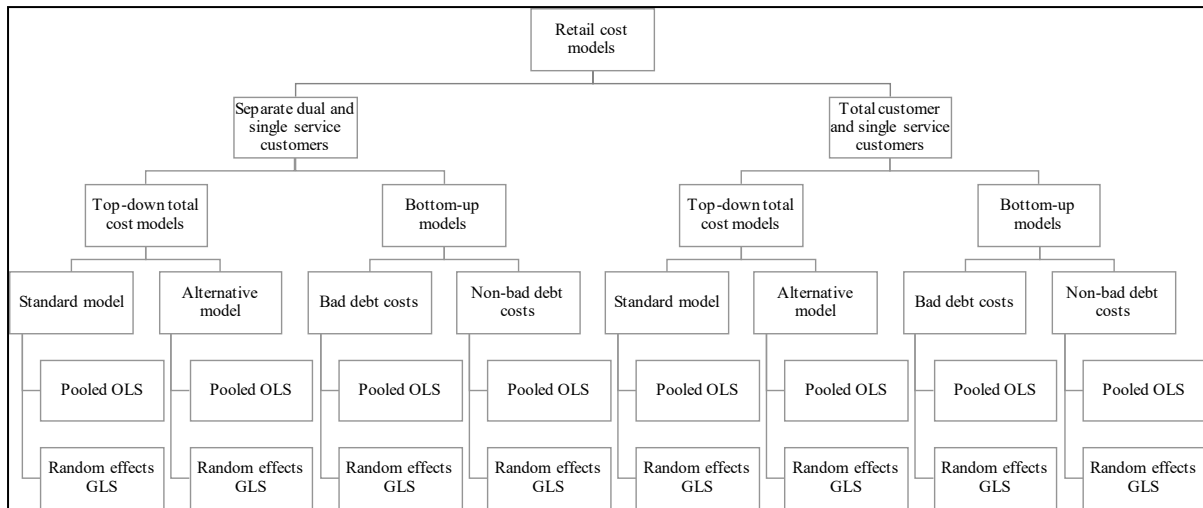
Figure 3: Schematic presentation of the suite of models

Table 1*:* Ofwat's adjustments of efficiency scores at PR04 and PR14

| Policy tool | PR04 | PR14 |
|---|---|---|
| Residual adjustment | 10% water, 20% wastewater. | None |
| Frontier selection | Pragmatic turnover rule, outlier treatment. | Upper quartile |
| Feasibility adjustment | 60% opex, 40% capital maintenance. | None |
| Glide path | 5-year opex, 3-year capital maintenance. | None |

Table 2*:* Drivers of domestic water retail costs

| Cost category | Drivers |
|---|---|
| All | Scale (number of customers) |
| | Scope (dual versus single service customers) |
| | Population transience |
| | Wages |
| | Service quality |
| Bad debt costs | Income deprivation |
| | Exposure to financial shocks |
| | Wholesale bill |
| Non-bad debt costs | Metered customers |
| | Density of metered customers |
| | Traffic speed |
| | Housing stock |

## Table 3: Description of the variables

| Variable | Description | Source |
|---|---|---|
| Total retail costs | Total operating costs for domestic water retail | Water company data share |
| Bad debt costs | Doubtful debt and debt management costs for domestic water retail | Water company data share |
| Non-bad debt costs | Total retail costs less bad debt costs | Water company data share |
| Single service customers | Total number of water-only or sewerage-only customers served by each company | Water company data share |
| Dual service customers | Total number of water and sewerage customers served by each company | Water company data share |
| Total customers | Total number of customers served by each company | Water company data share |
| Population flow (%) | Sum of population inflows and population outflows between UK local authorities, divided by population | Office for National Statistics, Local Area Migration Indicators |
| Sales-related wages | Average weekly wages for Standard Occupation Classification 721 (sales-related occupations) | Office for National Statistics Annual Survey of Hours and Earnings |
| Billing SIM | Service Incentive Mechanism quality rating for companies' billing functions | Ofwat |
| IMD income score (%) | Index of Multiple Deprivation income domain (proportion of population eligible for income-related benefits) | Ministry of Housing, Communities and Local Government, English IMD 2015; Welsh Government, Welsh IMD 2014 |
| Property repossessions (%) | Number of mortgage and landlord repossessions divided by population | Ministry of Justice, Mortgage and landlord possession statistical tables |
| House price to income ratio | Ratio of median house price to mean gross annual earnings | Office for National Statistics, Housing summary measures analysis |
| Wholesale bill | Average wholesale bill per customer | Water company data share |
| Metered customers (%) | Proportion of total customers that are metered | Water company data share |
| Metered household density | Number of metered households divided by water mains length | Water company data share |
| Traffic speed | Average A road peak hour traffic speed | Department for Transport, Congestion and reliability statistics |
| Flats (%) | Number of flats, apartments and maisonettes, divided by total number of dwellings | 2011 Census |

## Table 4: Summary statistics

| Variable | Mean | Std. Dev. | Minimum | Maximum |
|---|---|---|---|---|
| Retail operating costs | 48.517 | 46.976 | 2.745 | 184.600 |
| Bad debt costs | 21.998 | 24.364 | 0.651 | 100.900 |
| Non-debt costs | 26.518 | 23.960 | 2.074 | 88.400 |
| Single service customers | 615.649 | 491.389 | 75.811 | 1,946.537 |
| Dual service customers | 938.038 | 1,112.385 | 0.000 | 3,388.495 |
| Total customers | 1,553.687 | 1,373.792 | 111.415 | 5,335.032 |
| Population flows | 0.100 | 0.019 | 0.065 | 0.145 |
| Sales-related wages | 397.373 | 21.784 | 349.350 | 438.277 |
| Billing SIM | 4.480 | 0.166 | 3.860 | 4.700 |
| IMD income score | 0.138 | 0.024 | 0.099 | 0.174 |
| Property repossessions | 0.024 | 0.007 | 0.012 | 0.044 |
| House price to income ratio | 9.040 | 1.987 | 6.362 | 14.391 |
| Average wholesale bill | 239.234 | 106.107 | 78.324 | 487.518 |
| Metered households | 728.538 | 601.674 | 59.739 | 2,158.572 |
| Metered household density | 38.528 | 16.493 | 18.041 | 99.393 |
| Flats | 0.143 | 0.037 | 0.092 | 0.223 |
| Peak traffic speed | 26.129 | 4.026 | 17.353 | 32.255 |

Table 5: Estimated models using the first approach to account for any retail cost savings from a dual service

| Variable | Pooled OLS | | | | Random effects GLS | | | |
|---|---|---|---|---|---|---|---|---|
| | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 | Model 7 | Model 8 |
| | ln(total retail operating costs) | ln(bad debt costs) | ln(non-bad debt costs) | ln(total retail operating costs) | ln(total retail operating costs) | ln(bad debt costs) | ln(non-bad debt costs) | ln(total retail operating costs) |
| ln(single service customers) | 0.536*** (0.000) | 0.535*** (0.000) | 0.498*** (0.000) | 0.563*** (0.000) | 0.349*** (0.001) | 0.532*** (0.000) | 0.268** (0.025) | 0.318*** (0.003) |
| ln(dual service customers) | 0.122*** (0.000) | 0.121*** (0.000) | 0.263*** (0.000) | 0.159*** (0.000) | 0.226*** (0.000) | 0.184*** (0.003) | 0.250*** (0.000) | 0.246*** (0.000) |
| Metered customers (%) | | | 0.014*** (0.0002) | 0.007* (0.062) | | | 0.002 (0.610) | 0.002 (0.500) |
| Metered household density (per km mains) | | | -0.016*** (0.001) | -0.007** (0.041) | | | | |
| Flats (%) | 0.057*** (0.000) | | | 0.060*** (0.001) | | | | 0.053 (0.144) |
| ln(peak traffic speed) | | | -1.830*** (0.000) | -0.364 (0.290) | | | -1.217** (0.047) | |
| IMD income (%) | 0.164*** (0.000) | 0.189*** (0.000) | | 0.155*** (0.000) | 0.066 (0.167) | 0.136*** (0.008) | | 0.105* (0.056) |
| Property repossessions (%) | | | | | 0.107*** (0.000) | | | 0.119*** (0.002) |
| ln(average wholesale bill) | 1.206*** (0.000) | 1.744*** (0.000) | | 0.999*** (0.000) | 0.341 (0.152) | 1.235*** (0.002) | | 0.301 (0.213) |
| Internal population total flow (%) | | 0.0909*** (0.001) | | | | | | |
| Time trend | | | | | | | -0.037** (0.014) | |
| Constant | -10.020*** (0.000) | -14.372*** (0.000) | 4.539*** (0.000) | -8.063*** (0.000) | -2.741 (0.103) | -10.254*** (0.000) | 4.104* (0.067) | -3.836** (0.039) |
| $R^2$ adj. | 0.9284 | 0.9333 | 0.8743 | 0.9283 | 0.8957 | 0.9260 | 0.8539 | 0.9060 |
| No. of Observations | 89 | 89 | 89 | 89 | 89 | 89 | 89 | 89 |

Notes: ***, ** and * denote statistical significance at the 1%, 5% and 10% levels, respectively. Probability levels are in parentheses.

Table 6: Estimated models using the second approach to account for any retail cost savings from a dual service

| Variable | Pooled OLS | | | | Random effects GLS | | | |
|---|---|---|---|---|---|---|---|---|
| | Model 9 | Model 10 | Model 11 | Model 12 | Model 13 | Model 14 | Model 15 | Model 16 |
| | ln(total retail operating costs) | ln(bad debt costs) | ln(non-bad debt costs) | ln(total retail operating costs) | ln(total retail operating costs) | ln(bad debt costs) | ln(non-bad debt costs) | ln(total retail operating costs) |
| ln(total customers) | 0.877*** (0.000) | 0.979*** (0.000) | 1.061*** (0.000) | 0.966*** (0.000) | 1.043*** (0.000) | 0.933*** (0.000) | 1.069*** (0.000) | 1.065*** (0.000) |
| ln(single service customers) | | | -0.120*** (0.000) | -0.069* (0.087) | -0.134** (0.041) | | -0.138** (0.021) | -0.150** (0.030) |
| Metered customers (%) | | | 0.005*** (0.004) | 0.005*** (0.005) | | | 0.005 (0.114) | 0.002 (0.400) |
| ln(peak traffic speed) | | | -0.257* (0.062) | | | | -0.327 (0.286) | |
| IMD income (%) | 0.027*** (0.001) | 0.067*** (0.000) | | 0.027*** (0.003) | | 0.055* (0.071) | | |
| Property repossessions (%) | 0.121*** (0.000) | | | 0.147*** (0.000) | 0.113*** (0.000) | 0.147** (0.015) | | 0.130*** (0.000) |
| ln(average wholesale bill) | 0.659*** (0.000) | 1.091*** (0.000) | | 0.480*** (0.000) | 0.400*** (0.004) | 1.165*** (0.000) | | 0.351** (0.019) |
| Internal population total flow (%) | | | | | | | | |
| Time trend | | | | | | | -0.035*** (0.002) | |
| Constant | -6.974*** (0.000) | -11.307*** (0.000) | -3.200*** (0.000) | -6.502*** (0.0000) | -5.519*** (0.000) | -11.572*** (0.000) | -2.820** (0.011) | -5.446*** (0.000) |
| $R^2$ adj. | 0.9821 | 0.9616 | 0.9676 | 0.9835 | 0.9815 | 0.9639 | 0.9709 | 0.9824 |
| No. of observations | 89 | 89 | 89 | 89 | 89 | 89 | 89 | 89 |

Notes: ***, ** and * denote statistical significance at the 1%, 5% and 10% levels, respectively. Probability levels are in parentheses.

Table 7: Efficiency estimates from the suite of models

| Company | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 | Model 7 | Model 8 | Model 9 | Model 10 | Model 11 | Model 12 | Model 13 | Model 14 | Model 15 | Model 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Affinity | 33.4% | 56.7% | 58.7% | 32.2% | 59.3% | 59.2% | 68.1% | 53.5% | 16.2% | 44.3% | 21.3% | 13.6% | 10.1% | 13.9% | 19.9% | 7.3% |
|  | (15) | (16) | (17) | (14) | (17) | (16) | (17) | (15) | (13) | (16) | (11) | (13) | (9) | (6) | (13) | (9) |
| Anglian | 21.8% | 29.2% | 0.0% | 15.4% | 4.5% | 22.7% | 17.1% | 28.7% | 4.5% | 33.1% | -9.3% | -3.1% | 4.4% | 23.6% | -8.2% | 0.0% |
|  | (14) | (13) | (5) | (13) | (6) | (8) | (8) | (12) | (6) | (11) | (2) | (3) | (7) | (10) | (3) | (5) |
| Bristol | 0.0% | 35.3% | -12.9% | -1.7% | 17.1% | 38.3% | 1.9% | 27.2% | -3.1% | 41.2% | -23.5% | 0.0% | -2.2% | 29.6% | -19.1% | -1.0% |
|  | (5) | (14) | (2) | (4) | (10) | (14) | (6) | (11) | (3) | (15) | (1) | (5) | (3) | (15) | (1) | (4) |
| Dee Valley | -29.6% | -53.5% | 14.7% | -22.7% | -66.7% | -19.4% | -14.6% | -36.6% | 5.7% | 0.0% | 24.2% | 1.7% | 19.1% | 25.1% | 24.2% | 16.6% |
|  | (2) | (2) | (6) | (2) | (1) | (2) | (2) | (2) | (8) | (5) | (12) | (6) | (15) | (11) | (15) | (14) |
| Northumbrian | -95.9% | -79.7% | -18.4% | -98.1% | -52.1% | -49.0% | 0.0% | -40.0% | -18.2% | -2.5% | 12.0% | -11.2% | 0.8% | -2.8% | 10.4% | 1.4% |
|  | (1) | (1) | (1) | (1) | (2) | (1) | (5) | (1) | (1) | (4) | (8) | (2) | (6) | (4) | (7) | (7) |
| Portsmouth | 7.3% | 20.0% | 15.2% | 8.6% | -11.6% | 30.2% | -3.5% | -13.7% | 5.3% | 9.4% | 0.0% | 8.2% | 0.0% | 27.8% | 0.0% | 1.3% |
|  | (8) | (9) | (7) | (8) | (4) | (11) | (4) | (4) | (7) | (6) | (5) | (9) | (5) | (13) | (5) | (6) |
| Sutton & East Surrey | -12.9% | 1.9% | -6.9% | -20.7% | 0.0% | -4.9% | -10.7% | -20.0% | 10.0% | 20.5% | 25.6% | 11.9% | -1.3% | -26.0% | 17.8% | -1.9% |
|  | (3) | (6) | (4) | (3) | (5) | (3) | (3) | (3) | (9) | (7) | (14) | (12) | (4) | (2) | (11) | (2) |
| South East | 42.0% | -29.2% | 55.7% | 37.2% | 55.7% | 41.1% | 64.2% | 62.7% | 10.4% | -94.4% | 18.6% | 2.3% | 10.6% | 0.0% | 16.5% | 7.9% |
|  | (16) | (3) | (16) | (16) | (15) | (15) | (16) | (17) | (10) | (1) | (10) | (7) | (10) | (5) | (10) | (10) |
| Southern | 20.6% | 54.0% | 55.5% | 32.5% | 29.4% | 36.6% | 29.3% | 22.4% | 47.0% | 65.6% | 35.9% | 44.3% | 40.0% | 45.9% | 36.4% | 38.6% |
|  | (13) | (15) | (15) | (15) | (14) | (12) | (11) | (9) | (17) | (17) | (17) | (17) | (17) | (17) | (17) | (17) |
| South Staffordshire | 11.8% | 22.6% | 34.4% | 13.5% | 25.0% | 21.4% | 47.2% | 38.4% | 12.3% | 31.3% | 11.0% | 11.4% | 18.7% | 18.7% | 18.8% | 17.8% |
|  | (10) | (11) | (11) | (12) | (13) | (7) | (14) | (14) | (11) | (9) | (7) | (11) | (14) | (7) | (12) | (15) |
| Severn Trent | 18.0% | 0.0% | 22.1% | 12.2% | 15.3% | -0.8% | 44.0% | 29.8% | -1.8% | -13.3% | 12.9% | -0.6% | 4.7% | -9.5% | 23.5% | 4.7% |
|  | (12) | (5) | (10) | (11) | (8) | (4) | (13) | (13) | (4) | (2) | (9) | (4) | (8) | (3) | (14) | (8) |
| South West | 13.7% | 8.9% | 38.2% | 9.0% | 19.9% | 26.3% | 20.1% | 20.8% | 17.1% | 22.6% | 25.5% | 11.1% | 15.1% | 19.4% | 12.7% | 12.3% |
|  | (11) | (7) | (12) | (9) | (11) | (10) | (9) | (8) | (14) | (8) | (13) | (10) | (13) | (8) | (8) | (13) |
| Thames | 0.8% | -4.1% | 20.8% | 0.0% | 19.9% | 36.6% | 13.1% | 4.2% | 16.0% | 32.5% | 26.7% | 16.2% | 12.7% | 21.1% | 14.4% | 10.9% |
|  | (6) | (4) | (8) | (5) | (11) | (12) | (7) | (6) | (12) | (10) | (15) | (15) | (12) | (9) | (9) | (12) |
| United Utilities | 48.5% | 61.4% | 51.9% | 47.2% | 56.7% | 68.6% | 52.6% | 54.6% | 18.9% | 36.1% | 4.8% | 15.7% | 11.2% | 30.1% | 3.7% | 9.8% |
|  | (17) | (17) | (14) | (17) | (16) | (17) | (15) | (16) | (15) | (13) | (6) | (14) | (11) | (16) | (6) | (11) |
| Welsh | 2.8% | 25.4% | 45.6% | 9.5% | 16.4% | 26.2% | 40.4% | 23.4% | 22.7% | 40.6% | 27.7% | 27.7% | 23.4% | 29.2% | 26.4% | 24.7% |
|  | (7) | (12) | (13) | (10) | (9) | (9) | (12) | (10) | (16) | (14) | (16) | (16) | (16) | (14) | (16) | (16) |
| Wessex | -6.0% | 22.1% | -7.9% | 1.7% | -28.1% | 10.5% | -39.0% | 0.0% | 0.0% | 34.4% | -3.4% | 6.5% | -2.7% | 27.4% | -11.5% | -1.4% |
|  | (4) | (10) | (3) | (6) | (3) | (6) | (1) | (5) | (5) | (12) | (4) | (8) | (2) | (12) | (2) | (3) |
| Yorkshire | 7.3% | 17.3% | 21.2% | 1.9% | 6.2% | 0.0% | 24.2% | 9.0% | -8.8% | -8.2% | -5.8% | -12.1% | -19.4% | -27.5% | -3.1% | -20.9% |
|  | (8) | (8) | (9) | (7) | (7) | (5) | (10) | (7) | (2) | (3) | (3) | (1) | (1) | (1) | (4) | (1) |

Note: Rankings of the efficiency estimates are in parentheses.

Table 8: Estimates of the average efficiency gaps across subsets of models

| Company | Single and dual service customer variables (1st approach to model any retail cost savings from a dual service) | | | Total and single service customer variables (2nd approach to model any retail cost savings from a dual service) | | |
|---|---|---|---|---|---|---|
| | Top-down | Bottom-up | Average | Top-down | Bottom-up | Average |
| Affinity | 44.6% (15) | 61.6% (17) | 53.1% (16) | 11.8% (11) | 23.3% (15) | 17.6% (14) |
| Anglian | 17.6% (11) | 17.4% (9) | 17.5% (9) | 1.5% (5) | 10.1% (6) | 5.8% (7) |
| Bristol | 10.6% (8) | 11.0% (6) | 10.8% (6) | -1.6% (3) | 0.8% (2) | -0.4% (3) |
| Dee Valley | -38.9% (2) | -6.2% (3) | -22.6% (2) | 10.8% (10) | 22.2% (13) | 16.5% (11) |
| Northumbrian | -71.5% (1) | -32.4% (1) | -51.9% (1) | -6.8% (2) | 5.4% (4) | -0.7% (2) |
| Portsmouth | -2.4% (5) | 9.1% (5) | 3.4% (5) | 3.7% (7) | 3.2% (3) | 3.4% (4) |
| Sutton & East Surrey | -13.4% (3) | -7.3% (2) | -10.4% (3) | 4.7% (8) | 16.8% (9) | 10.7% (9) |
| South East | 49.4% (16) | 54.1% (15) | 51.8% (15) | 7.8% (9) | 10.6% (7) | 9.2% (8) |
| Southern | 26.2% (14) | 43.9% (14) | 35.0% (14) | 42.5% (17) | 46.2% (17) | 44.3% (17) |
| South Staffordshire | 22.2% (13) | 35.3% (13) | 28.7% (13) | 15.0% (15) | 17.9% (10) | 16.4% (10) |
| Severn Trent | 18.8% (12) | 22.0% (10) | 20.4% (11) | 1.8% (6) | 8.4% (5) | 5.1% (5) |
| South West | 15.9% (10) | 24.1% (11) | 20.0% (10) | 13.9% (12) | 19.9% (11) | 16.9% (12) |
| Thames | 6.2% (7) | 16.7% (7) | 11.4% (7) | 13.9% (12) | 23.1% (14) | 18.5% (15) |
| United Utilities | 51.8% (17) | 59.3% (16) | 55.5% (17) | 13.9% (12) | 20.3% (12) | 17.1% (13) |
| Welsh | 13.0% (9) | 33.6% (12) | 23.3% (12) | 24.6% (16) | 31.4% (16) | 28.0% (16) |
| Wessex | -8.1% (4) | -4.8% (4) | -6.4% (4) | 0.6% (4) | 10.6% (7) | 5.6% (6) |
| Yorkshire | 6.1% (6) | 16.9% (8) | 11.5% (8) | -15.3% (1) | -10.0% (1) | -12.7% (1) |

Note: Rankings of the average efficiency gaps are in parentheses.

---

[1] Cambridge Economic Policy Associates (CEPA) advised Ofwat on the econometric modelling for PR14.

[2] Ultimately econometric benchmarking *by Ofwat* of companies' domestic retail cost efficiencies did not form the basis of the final determinations of the companies' PR19 domestic retail price controls. This was because Ofwat set these controls for the companies for 1 April 2020 - 31 March 2025 at the levels companies proposed in their business plans (Ofwat, 2019). In doing so, Ofwat acknowledged the performance improvements inherent in the levels the companies proposed. On the other hand, it is evident that econometric benchmarking of domestic retail cost efficiencies was conducted *by companies* as part of their analysis for PR19, as they reported this benchmarking in their PR19 business plans (e.g., Welsh Water, 2019, United Utilities, 2018, and Wessex Water, 2018).

[3] This academic paper emerged from earlier work by Economic Insight (2018).

[4] It should be noted that Ofwat's modelling of retail costs was undertaken completely independently of the modelling in this paper.

[5] Doubtful debt relates to debts that are unlikely to be repaid and represents the movement in bad debt provision (and the bad debt charge). Expenditure on debt management relates to monitoring, issuing reminders, follow-up calls, and field recovery which includes customer visits and costs relating to third party debt collectors and legal fees.

[6] Customer services expenditure relates to billing, payment handling, vulnerable customer schemes, and customer enquiries and complaints.

[7] Specifically, Welsh Water (2019, p. 7) note that 'Ofwat's approach to residential retail modelling uses Corrected Ordinary Least Squares (COLS) to determine the 'efficient' level of expenditure.'

[8] SFA, DFA and DEA all have their advantages. For discussions of the relative merits of different frontier methods see Hjalmarsson *et al.* (1996), Ruggiero (2007) and Greene (2008).

[9] Technical change is Hicks-neutral if the change does not influence the productivity of the inputs. Relating this to the specifications of our cost models, where the dependent variable reflects the input cost to produce the output, technical change is Hicks-neutral as $t$ is not interacted with output, and so the change in the input cost when the output level changes is independent of technical change. This contrasts with the translog cost function, where technical change is non-neutral because $t$ is interacted with output, and so how the input cost changes when the output level changes depends on, among other things, the magnitude of technical change.

[10] There has recently been a resurgence of methodological interest in the SS efficiency estimator, as it is the basis for a method to calculate efficiency performance spillovers (e.g., Glass *et al*., 2014; 2016). This draws parallels with the well-established literature on productivity performance spillovers.

[11] The common way of using the random effects from an estimate of a model such as Eq. 2 to calculate time-varying efficiency estimates using the CSS approach is a three-step process. First, using a single model regress the disturbances from the fitted random effects model on, for each firm, a time trend and its square. For example, if there are 17 firms in the sample, there would be 34 time and time squared regressors in total. We cannot apply the CSS approach in our empirical analysis because in the first step this would involve estimating, for each firm, the coefficients on the time trend and its square using only five observations. To guard against this type of problem, the CSS efficiency estimator should only be applied to samples where $T$ is big. In practice, researchers do apply the CSS estimator when $T$ is not strictly big, but even in these cases $T$ is much bigger than five. Second, for each firm add its random effect to the sum of the time trend and its square pre-multiplied by their firm specific coefficients. Third, using the values for the firms from the second step, calculate the time-varying efficiency estimates in the same fashion as COLS (see Eq. 3).

[12] Note that the sign of inefficiency in a cost frontier model specification is positive. This is because firms aim to minimise costs so inefficient firms will lie above the cost frontier (see figure 2).

[13] Another approach that has widely been applied to COLS by regulators, in spite of it being conceptually problematic, is to split the residuals between inefficiency and other components (Ofwat, 2008; Office of Rail Regulation (ORR), 2013). This typically involves applying a fixed percentage reduction to all firms' residuals (Cubbin, 2004; Chung, 2011) and effectively pulls all the observations closer to the regression line, so that all firms' distances to the lowest residual are reduced. While this approach directly addresses the problem at hand, the choice of percentage reduction is arbitrary and assumes that the split between inefficiency and other residual components is the same across all firms. We do not therefore use this approach here.

[14] We note that, at earlier price controls, regulators including Ofwat tended to address this issue by selecting the 'frontier' firm as the benchmark (i.e., not the upper quartile) and then applying downwards adjustments to the residuals, to reflect the fact that they do not entirely relate to inefficiency. Note, this is separate from other downwards adjustments to residuals made by regulators for reasons other than the efficiency gap – which we discuss subsequently.

[15] We do not use our fitted models for future prediction, so our benchmarking of companies' domestic retail cost performance is based on historical upper quartile performance. At PR19 Ofwat (2017a, p. 18) planned to use such an approach as part of its benchmarking of companies' domestic retail cost performance, by using an average of historical and predicted future upper quartile performance. Ultimately, such benchmarking *by Ofwat* did not form the basis of the final determinations of the PR19 retail price controls. See the policy implications in section 8 for more details on this.

[16] The Pearson correlation coefficient between $ts$ and $s$ is only 0.53. There are evidently a sufficient number of companies in the sample with sufficient dual service customers to not be unduly concerned about collinearity between $ts$ and $s$ in the models.

[17] Noting that companies typically will face a trade-off between debt management and bad debt costs. Hence, in the face of higher transience a company could: (i) invest in more debt management activities in order to attempt to recover a higher proportion of debt than it otherwise would; or (ii) not incur the higher debt management costs, but accept that its debt costs will increase as a result of the transience. The 'optimal' balance will vary from company-to-company, depending on their circumstances. The point, however, is that transience must logically increase these costs, when considered as a whole.

[18] This additional company is Bournemouth Water, which was purchased by the parent company of South West Water in 2015, and, as a result, is not in the dataset in the last year of the sample.

[19] We report average efficiencies for these 17 companies, as these are the companies Ofwat set retail price controls for at PR19.

[20] Although the VIF pertains to cross-sectional models, since based on the VIF we have no concerns about multi-collinearity in the OLS models, we can make the case that there are also no multi-collinearity issues in our random effects models. This is because the same explanatory variables are used to estimate both sets of models.