A Review of Ofwat's January 2019 Wholesale Water and Wastewater Botex Cost Assessment Modelling for PR19

March 2019

Professor David Saal and Dr Maria Nieswand

Centre for Productivity and Performance
School of Business and Economics
Loughborough University
Loughborough, LE11 3TU
UK

Contact Details

D.S.Saal@lboro.ac.uk
+44 (0) 1509 227 123

Table of Contents

Introduction, Terms of Reference and Expression of Interest

We have been engaged by Anglian Water to provide an independent review of the regulatory cost assessment models that Ofwat developed and published in January 2019 as part of its Initial Assessment of Company Business Plans (IAP). Given that we have worked closely with Anglian Water to first review its own cost modelling[1], and then engaged directly with it in its subsequent efforts to providing improved cost models[2], we fully acknowledge our relationship with Anglian Water. However, our understanding with Anglian Water has always prioritised the importance we place on maintaining academic integrity and independence. We, therefore, came to this review as independent academics with substantial experience and reputation in cost modelling who have engaged with Anglian Water on cost modelling, and not as consultants whose arguments have been driven exclusively by Anglian Water's agenda.[3]

We therefore emphasise that while Anglian Water asked us to review Ofwat's Wholesale Water and Wastewater modelling, it did not place any general or specific restrictions on the approach we took nor on the issues we raised, nor on the conclusions which we reached. Moreover, while we kept Anglian Water abreast of the content of our review as it developed, they have only contributed to this report via extremely minor comments that did not substantively influence the conclusions that we had already independently reached. In fact, the only binding restriction placed on us by Anglian Water was to provide the report within a timeline that would allow this review to inform its formal response to Ofwat's IAP.

While we initially conceived of the review as providing only expert commentary on the appropriateness of Ofwat's modelling, Ofwat's provision of the underlying data and Stata coding for its models on January 31st, but failure to provide a report detailing its econometric modelling approach until a week later, took us down an unexpected path: as we needed to run Ofwat's Stata codes to get timely access to Ofwat's econometric modelling, we unexpectedly started our review by running econometric models. This led us to begin econometrically testing for some of the issues that we believed might be present in Ofwat's models, subsequently resulting in the detailed empirical testing of Ofwat's models that our final report now draws many of its conclusions from. We stress that consistent with our understanding of Anglian's expectations to provide a review of Ofwat's modelling, as opposed to providing alternative models, the resulting models should not be seen as our suggestion of the models that Ofwat or Anglian should employ in this or future price reviews. Instead, they should be seen as empirical evidence indicating issues with Ofwat's models, and in some cases, as possible ways forward for at least making incremental improvements to Ofwat's PR19 modelling.

Our resulting report is therefore one of two parts. As presented, the second part includes detailed commentary on Ofwat's wholesale water and wastewater models, augmented with the presentation of additional models that allow us to empirically demonstrate the relevance of many of the concerns we raise with Ofwat's models rather than simply relying on verbal commentary. The first part

[1] http://www.anglianwater.co.uk/_assets/media/cost-modelling-report.pdf
[2] http://www.anglianwater.co.uk/_assets/media/AW_Cost_Modelling_Report_March_2018_MAIN_REPORT.pdf
[3] We note that the views expressed in this review are made in a personal capacity and are not those of Loughborough University or any of its constituent parts or affiliated entities.

remains more consistent with our originally intended approach of further elaborating on the concerns about Ofwat's modelling approach already raised by Professor Saal in his response to Ofwat's March 2018 Cost Assessment consultation. However, this section now also incorporates additional conclusions based on our empirical review, as well as empirical evidence that supports the comments that we would have otherwise made in the absence of our detailed empirical review.

Executive Summary

Our empirical review of Ofwat's PR19 cost assessment models has identified and empirically demonstrated significant underlying issues in the entire suite of models that Ofwat employed in its Initial Assessment of Business Plans (IAP). Moreover, we believe that our approach of explaining and then making small, reasonable changes to Ofwat's models provides direct incontrovertible evidence of significant limitations and omissions in its models. We are therefore confident that our empirical review alone provides sufficient evidence to demonstrate that Ofwat's models do not provide a level of confidence that should be met before models are applied to regulatory cost assessment.

However, we also strongly believe that the problems with Ofwat's PR19 cost modelling are by no means limited to the models it chose for use in the IAP. Instead, these problems result from, and are inextricably linked to Ofwat's regulatory cost modelling framework, which has resulted in excessively parsimonious cost assessment models. Moreover, we emphasise that we have already raised our concerns with the approach that Ofwat was developing on two occasions within the PR19 process. Thus, the Centre for Productivity and Performance's September 2017 review of Anglian Water's initial cost modelling, included a section entitled "Issues Raised When Reconciling our Academic Approach to Cost Assessment with Anglian Water's PR19 Determined Approach", that noted several concerns with the approach being developed by Ofwat, which Anglian had mirrored in its own preliminary cost modelling. Furthermore, Professor Saal's May 2018 response to Ofwat's cost assessment consultation elaborated significant concerns with the models and modelling framework provided by Ofwat's consultants, which in practice, Ofwat continued to follow extremely closely in its January 2019 models[4].

Thus, it is our professional opinion that Ofwat's rigid adherence to a modelling framework that only effectively always includes only a single scale variable, and an extremely limited number of noninteractive control variables, resulted in a set of excessively parsimonious models that cannot properly account for the complex cost determinants and interactions between multiple outputs and characteristics that drive water and wastewater costs. It was therefore not surprising, that our empirical assessment of Ofwat's models could demonstrate that very simple modifications of its models would undermine confidence in their use for regulatory cost assessment. Moreover, this empirical evidence validates that our concerns with regard Ofwat's modelling framework are not based on esoteric academic arguments but are instead consistent with our experience of applying academic knowledge to meet the real-world requirements of regulatory cost assessment.

Notwithstanding the strong concerns we express with regard to Ofwat's cost modelling approach at IAP, we have demonstrated objectively that in some cases Ofwat's existing models can be improved by small modifications or additional variables while largely maintaining Ofwat's existing cost modelling structures. We stress that we do not consider that the changes we made in this review are adequate, as they were made as a means of demonstrating underlying issues with Ofwat's modelling approach and models. However, they do suggest that pursuing such changes could incrementally

[4] While Ofwat added a few lines to its econometric modelling commentary suggesting that it would consider the use of interactive and squared variables, in practice its modelling maintained a rigid adherence to always modelling with a single output, and employing an extremely limited set of limited explanatory variables (Ofwat, 2019a p. 7). Thus, Ofwat's only use of squared interaction terms was its specifications for density, which we raise concerns with below.

improve the quality of Ofwat's modelling, by better taking into account economic, managerial, and economic factors influencing cost. As such, we hope that our observations and the approach we have followed may help Ofwat to improve the final botex models used for PR19.

Given this overall summary we now summarise our conclusions with regard to our overall assessment of Ofwat's modelling framework, and with regard to its Wholesale Water and Wastewater Modelling.

Our Assessment of Ofwat's Modelling Framework

Our assessment is that Ofwat's strict adherence to a particular cost modelling framework is the primary reason that its PR19 cost assessment modelling does not provide a suite of models that can be considered sufficiently robust for regulatory cost assessment. We therefore summarise, our key conclusions regarding Ofwat's modelling framework:

1. **Ofwat's modelling framework is excessively parsimonious and represents an overzealous response to the criticisms made by the Competition and Markets Authority (CMA) that Ofwat's model was excessively complex in its 2015 review of Bristol Water's PR14 Price Determination.**

   While Ofwat was correct to respect and strive to improve its modelling in response to the CMA, it is unlikely that the CMA intended to suggest the use of an arbitrarily simplistic modelling framework, and the resulting models that Ofwat chose to use in its IAP. Thus, Ofwat's PR19 models can now be characterised as being derived from a framework that was designed to be excessively parsimonious, and therefore incapable of allowing for sufficient complexity to adequately control for the underlying determinants of company costs.

   Moreover, Ofwat's PR19 models can not only be verbally and conceptually criticised for "excessive simplicity", as we have also provided clear empirical evidence that demonstrates that its models and modelling approach suffers from omitted variables bias.

   This is clearly demonstrated by Ofwat's assertions that it could not specify a model for integrated wholesale wastewater, while with only very limited modelling effort, we were able to demonstrate that such a model can indeed be created. We did this by simply adding a single control variable to the set of variables employed by Ofwat in its wastewater modelling and using the standard econometric technique of general to specific modelling to identify a model, which we believe meets minimum standards for use in regulatory cost assessment. Similarly, a very simple modification of how Ofwat controls for pumping stations, and the addition of network length to its Integrated Wholesale Water models also suggests that these models suffer from omitted variables bias and/or other issues, and are therefore "excessively simplistic".

2. **We observe that Ofwat is at times unclear about its applied modelling principles and also applies them inconsistently. In some cases, we found indication for Ofwat even violating**

its own principles or modifying them since the cost assessment consultation without explaining or indicating the change in its modelling approach.

While the IAP documents clarify some of those principles (e.g., the choice of Random Effects), others were emphasised in previous phases of the PR19 process (e.g. the VIF multicollinearity criterion) but violated and removed from discussion in the IAP modelling without even mentioning why these principles were changed or dismissed from the IAP set of modelling principles. Moreover, some principles are only stated vaguely which opens the doors for their inconsistent application, e.g., the accepted level of significance is vaguely defined but serves as a justification for excluding wholesale waste water models but, at the same time, variables in some models are included at a significance levels well below the normal minimum standard of 10 percent.

We are particularly concerned with this lack of transparency, as we believe Ofwat's modelling framework was largely designed to provide what we believe to be the excessive parsimony required to meet its inappropriately stringent multicollinearity standards. However, while the original modelling framework and its limitations remain in the January 2019 models, Ofwat appears to abandon key principles it once emphasised as important in the design of its modelling framework.

Moreover, while we cannot be certain, it appears that Ofwat's motivation in abandoning its strict multicollinearity standards is its modellers' choice to exclusively rely on weighted average local area density in its water models. Unfortunately, as we discuss further below, we do not believe this approach provides results consistent with managerial, engineering, or economic understanding of the J-shaped relationship between density and costs.

3. **Even if the models chosen by Ofwat were appropriate, Ofwat's approach to triangulation is not, because it fails to triangulate models that can be considered alternatives.**

Ofwat's triangulation amounts to changing control variables that, at best, provide only a small portion of overall explanatory power. Moreover, as it does not offer an Integrated Wholesale Wastewater model, it does not effectively triangulate wholesale wastewater. Given the importance of the single scale variable in each model (e.g., connected properties in its Integrated Wholesale Water, and Water Resource Plus models), a logical way of reducing specification biases would be to triangulate across model specifications that include alternative scale variables (e.g., length of mains) depending on the activity of interest. Instead, Ofwat even triangulates models that violate its own principles if academic standards were to be applied, e.g., BR1 collapses to a single scale variable model if it had treated the control variable as statistically insignificant according to normal practice.

4. **Given Ofwat's modelling framework's extremely limited ability to appropriately control for heterogeneity in the economic, engineering, spatial , and demographic characteristics that determine the cost of the complex activities , we have considerable concerns with regard**

to whether it is appropriate to employ random effects modelling, and/or interpret the resulting firm specific error terms as evidence of inefficiency.

To state this more strongly, Ofwat provides no evidence to suggest that its random effects models provide evidence of inefficiency. Instead we believe it is far more likely that Ofwat's random effects residuals should be interpreted as statistical allowance for its models' failure to adequately control for the complex factors influencing differences in firms' costs.

While this point is a relatively sophisticated econometric issue, its significance should not be underestimated by the lay reader, as there is strong reason to suspect that Ofwat's models may actually provide no evidence of inefficiency between firms. This is because its random effects residuals may be simply controlling for unmodelled heterogeneity that cause legitimate differences in the determinants of firms' costs. Furthermore, the fact that 9 out of 13 of Ofwat's models fail Ramsey reset specification tests when estimated with OLS, but 12 out of 13 of its otherwise unchanged models pass Ofwat's Ramsey reset when estimated random effects, provides suggestive evidence for this conclusion.

5. **Ofwat has not considered carefully how its approach to modelling at the econometric level in a panel data relates to its use of those models in forecasts.**

   In sum, because Ofwat uses a random effects model, and allows for no change in the underlying cost relationship over the seven year period that it models, its econometric modelling assumes that both firm efficiency and the underlying relationship determining costs CANNOT change. However, it then uses these models to project forward and assumes not only that firms can meet efficiency catch up targets, but also that the cost frontier will move to allow continuing efficiency improvements (technical change). Thus, Ofwat's backward looking modelling assumptions are logically inconsistent with its use of those estimated models to make forward looking cost projections. As a result of this inconsistency, there is likely to be bias in both Ofwat's backward- and forward-looking assessment of costs.

6. **Moreover, the forecast of costs is flawed due to changes in the nature of some cost data**.

Ofwat uses the paraamters of models that are estimated with past botex expenditure to assess forward looking business plan data that includes enhancement opex. While it is true that there has always been enhancement opex incurred as a result of enhancement capex schemes, and that this could potentially be handled through botex cost models, there are now opex solutions to what would previously have been purely capex schemes which are also included in enhancement opex. This is a new category of expenditure and as such cannot possibly be modelled using historic data which has no such cost included in it. Consequently, companies who have put forward innovative solutions of this nature will be deemed to be inefficient compared to companies which have taken an enhancement capex route. As rectifying what was seen as a capex bias was a key reason for Ofwat introducing totex cost assessment, it is surprising that Ofwat should be undermining its own approach. Thus, this provides another reason why Ofwat's forward projections are suspect.

Our Assessment of Ofwat's Wholesale Water Models

When compared to Ofwat's suite of Wholesale Wastewater models, its Wholesale Water models appear relatively strong at the superficial level. Thus, Ofwat specifies a total of five models, all of which pass the Ramsey model specification test, have appropriately signed coefficient estimates, and, with the exception of its density controls, always have statistically significant coefficients.

However, considerable weaknesses become apparent with closer inspection, and we emphasise that the below points are only a partial list with further enumeration limited by the time frame of this review. Moreover, our empirical review has been able to provide empirical evidence supporting several of these points.

- **Ofwat's models do not allow for enough controls to capture the complex factors influencing water supply costs.**

  When taken together as a whole, Ofwat's models use no more than a single scale variable, a density control (in logged levels and squared log levels), a single control for treatment complexity, and a control for booster pumping stations. Factors such as significant differences in water resource and treatment costs attributable to the type of water source employed, regional variation in water scarcity, and substantial variation in company efforts to reduce water demand and scarcity via leakage control and water metering are only a few examples that quickly come to mind. Thus, compared to both the CMA's models in the Bristol Water determination and the academic literature, Ofwat's models do not appear to have controlled sufficiently for the complexity of water supply.

- **Moreover, we further emphasize that Environment Agency concerns that within 25 years England will not have enough water to meet demand, are particularly inconsistent with Ofwat's failure to control for past differences in water scarcity, leakage levels, metering, and other factors that have already resulted in additional costs for water stressed companies in its models. E.g this omission suggest Ofwat's models do not adequately address past nor future implications of steadily tightening water demand balances on water supply costs in England and Wales.**

- **Ofwat's models are unlikely to accurately capture the impact of differences in population density on water supply costs**.

  In all its models, Ofwat relies exclusively on the specification of a single scale variable and its density controls to capture these effects. While statistically significant in all other models, the density variables are not individually significant in the WRP2 Water Resources model. Moreover, its discussion of its estimated elasticity of cost with respect to density and the figure it provides illustrating this (Ofwat, 2019a, p. 15) suggest that the elasticity of cost with respect to density is always increasing. This is contrary to the general understanding that in network industries, costs generally decrease as density increases before some critical level where congestion leads to increased costs. ***Moreover, there is an extremely well-established academic literature on modelling and controlling for economies of density in network industries that provides alternative approaches to that employed by Ofwat.***

We also note that, given the statistical insignificance of its density controls in the WRP2 Water Resource model, and our below discussions suggesting that Local Area District density measures may provide too coarse a measure of density to accurately capture how population settlement patterns influence both water and wastewater system design, we strongly suspect that alternative approaches to controlling for density would yield different cost assessment results. Moreover, we have been able to provide preliminary evidence supporting this assertion, as inclusion of Ofwat's "lnboosterperlength" variable, which is highly negatively correlated with Ofwat's weighted density variable, into both of its Water Resource Plus models, causes its density control variables to become jointly insignificant, thereby demonstrating not only that Ofwat's Water Resource Models are not only mis-specified, but also that the density controls it **exclusively** relies on may not be appropriate in other models.

- **Our extensions of Ofwat's models set out below empirically demonstrate that Ofwat's models are under-specified and therefore suffer from omitted variables bias, because they do not control for the multiple output characteristics of water supply.**

  Water distribution and Integrated Wholesale Water are characterized by multiple output production. These outputs include the provision of a network transportation output and outputs which are provided with that network. The outputs provided by the network can be variously modelled with inclusion of output proxies such as population served, volumetric water delivered proxies, and/or proxies for connections served. Moreover, the same well-established academic literature that provides viable alternatives to the density approach employed by Ofwat suggests that cost interactions between network transportation outputs, volumetric outputs, and connection outputs substantially influence costs. Our below empirical review demonstrates that additionally controlling for properties or water volumes in Ofwat's water distribution models is appropriate, and controlling for water networks in its integrated wholesale water models also shows that its models suffer from omitted variables bias.

- **A striking characteristic of all of Ofwat's Water Models, is that they do not include the key output delivered by water companies: water**.

  While we are sympathetic to Ofwat's justification that using the number of households instead of water volumes allows it to use an output proxy that will allow better controls and incentives for leakage reduction and water efficiency schemes (Ofwat, 2019a, p. 12), we are not convinced, as this measure is also the standard proxy related to controlling for the cost of serving connections on a network. Moreover, many other viable alternative scale variables could be employed, while also allowing for incentives for efficient water usage. Thus, water delivered (distribution input less distribution losses) and Effective Water (Distribution Input less Total Leakages) are all equally valid measures that would meet Ofwat's stated criterion of using an output proxy that incentivizes for leakage reduction and water efficiency schemes. Moreover, within the scope of the time allowed for this review, we have demonstrated via models replacing Ofwat's property-based scale variable with Effective Water, that such alternatives provide viable alternatives models to Ofwat's properties-based approach. ***Ofwat's choice of output should be further considered and, at a minimum, should be triangulated with alternative specifications that employ alternative and potentially more appropriate proxies for the key water output delivered by companies.***

- **While beyond the scope of this review to carefully develop, Ofwat should more carefully consider the incentive implications of its models with regard to leakage controls and water demand reducing activities such as metering.**

  These are both costly activities that will result in lower water volumes. Models which use simple scale proxies as their primary drivers may therefore provide perverse incentives that suggest companies that have higher outputs should have higher costs, when the relationship is much more complex when water scarcity, leakage and water demand management are considered properly. Moreover, there are tradeoffs between the increased cost of leakage reduction and water demand management which increase network costs, and the savings to be gained in water resource plus.

- **Even when it does provide alternative models for Water Resources Plus and Integrated Wholesale Water, Ofwat does not carry out appropriate triangulation at the model level**.

  Thus, even though 87.8 to 94.8 percent of the variance of costs is explained by the single scale variable included in each of its models, the triangulation carried out by Ofwat amounts to no more than testing two alternative controls for water treatment complexity in its Water Resource Plus and Wholesale Water Models. Thus, Ofwat's "triangulation" amounts to modifying control variables which provide little of the overall explanatory power in its models. Similarly, as all of Ofwat's models **exclusively** rely on the same variable and specification of density, and density contributes far more to the explanatory power of the models than treatment complexity controls, Ofwat should also provide triangulation of models which capture density through different modeling approaches.

- **Ofwat offers a single distribution model, and therefore does not triangulate its estimation of distribution costs.** This is even though the underlying model estimated by Ofwat treats the number of pumping stations as an output which is effectively estimated to be responsible for 46.5 percent of all water distribution costs. In contrast, the number of properties serviced, or the volume of water delivered on the network, are effectively assumed to have zero cost elasticity, i.e. no effect on botex. Ofwat should, therefore, at a minimum triangulate its water distribution models with alternative specifications.

In sum, while Ofwat has provided a complete set of models for wholesale water, we believe that there are still considerable conceptual issues with its modelling. Moreover, we have also provided empirical evidence demonstrating omitted variables bias and other issues in its chosen models. We therefore believe that Ofwat's models can and should be improved before Ofwat's draft and final determinations, and believe that the issues we raised above as well as our empirical review provide strong indicators of several changes that would result in such improvement.

Our Assessment of Ofwat's Wholesale Wastewater Models

In our opinion, the striking characteristic of Ofwat's wastewater modelling for PR19 is its lack of a clear and well-developed conceptual relationship with the complex engineering, demographic, spatial, and other characteristics that determine the costs of providing wastewater services. In contrast, Ofwat displays strong confidence in its conceptual understanding of the determinants of wastewater costs in its assertion that "…wastewater companies in our sample are relatively similar, whereas water companies are more heterogenous." (Ofwat, 2019a, p.23). However, despite this supposed similarity Ofwat was unable to model integrated wastewater costs, and its explanation for this failure (Ofwat, 2019a, p.19), can be interpreted as a statement that since the modelling framework and variables it chose to employ did not provide an appropriate model, it simply chose not to provide a model.

We find its choice not to provide an integrated wholesale model more than puzzling as we were able to rapidly develop several robust and parsimonious models of integrated wholesale wastewater suitable for regulatory application, while also deliberately limiting ourselves to Ofwat's modelling framework. Moreover, we would expect even more sound and robust models to be possible, once adherence to Ofwat's modelling framework was relaxed to allow for specifications better able to account for the complex heterogeneity that actually characterises wastewater service provision both within and between companies.

Furthermore, we cannot understand why Ofwat did not seek to adopt or further develop any of the many alternative models of integrated wholesale wastewater that companies (including Anglian Water) submitted to the 2018 Cost Assessment consultation, particularly given that all the models it presented to the consultation were statistically inadequate based on Ramsey Reset model specification tests. Moreover, we interpret this as suggesting the possibility that, when the limitations imposed by its of its own modelling framework prevented development of an appropriate model, Ofwat appears to have chosen simply not to model integrated wholesale wastewater costs rather than explore the alternative modelling approaches that were presented by companies.

As a result, this has left Ofwat to rely on the following disaggregated models in its Initial Assessment of Plans:

- Sewage Collection models that can be readily challenged based on a correct interpretation of their underlying parameters (wrong signs and magnitude of the length of mains variable), and which therefore empirically demonstrates that they are not suitable for regulatory cost assessment.

- Bioresources Models that cannot be seen to have adequately controlled for the engineering, managerial and topographic factors influencing the costs associated with transporting, treating, and disposing of sludge as Ofwat relies only on a measure of density, in addition to a scale variable: This has resulted in one bioresources model that fails basic model specification tests, and another that relies on a single marginally statistically insignificant explanatory factor in addition to output, which implies that both of Ofwat's models "self-demonstrate" their statistical inadequacy*.* Moreover, we emphasise that Ofwat provides no compelling explanation about why other factors were not considered in its bioresources modelling, despite its modelling framework suggesting the need for further controls for topography and complexity.

- Ofwat's Sewage Treatment and Bioresources Plus models are superficially robust but identically specified. As the Sewage Treatment and Bioresources Plus models are identical, by definition, the Bioresources models do not allow for appropriate controls to capture the implications of transporting, treating, and disposing of sludge and should therefore be suspect on these grounds alone.

  Moreover, we were able to easily demonstrate that Ofwat's Bioresource Plus models suffer from omitted variables bias, and that its conceptual approach fails to account for the fact that decisions with regard to local sewage collection system scale, sewage treatment plant scale, and sludge treatment and transportation are integrally linked, thereby causing companies to optimise wastewater costs by running over 6,300 sewage treatment collection and treatments systems in England and Wales. Thus, adding a control variable for nonindigenous treatment of sludge at treatment works, as well as Ofwat's own control variables for pumping capacity and weighed average density, and employing a standard general to specific modelling approach yielded a model that unequivocally demonstrates that Ofwat's bioresource models are mis-specified. This, in our opinion, suggests that Ofwat's models are not yet suitable for regulatory cost assessment.

In sum, Ofwat's failure to provide an Integrated Wholesale Waste water model, coupled with the empirically demonstrable failings of its disaggregated models, demonstrates our overall conclusion that the entire set of wholesale waste water models employed in the IAP can and must be improved before Ofwat's draft and final determinations. Moreover, we believe that while our empirical review demonstrates significant issues with Ofwat's models, it also demonstrates several directions of travel that should result in substantially improved models.

Part 1 - General Assessment of Ofwat's Modelling Approach

We acknowledge Ofwat's stated ambition to develop sensible and meaningful models for setting efficient cost allowances, which are consistent with an engineering, operational and economic understanding of cost drivers (Ofwat, 2019a, p. 5). Moreover, Ofwat explicitly stated in its 2018 cost assessment consultation document, that its approach to model development and assessment would:

> *"Use engineering, operational and economic understanding to specify an Econometric model, and form expectations about the relationship between cost and cost drivers in the models." (Ofwat, 2018, p. 8).)*

Our review of Ofwat's January 2019 regulatory cost assessment models, therefore judges if Ofwat's econometric models are robust with respect to statistical, engineering, managerial and economic cost modelling considerations, and therefore appropriate for regulatory cost assessment. Our conclusion is that they are not, and that Ofwat has not sufficiently improved its overall modelling approach relative to what was proposed in the March 2018 Cost Assessment consultation.

Instead, Ofwat has largely remained committed to a rigid and flawed modelling framework, thereby producing conceptually inadequate models that therefore suffer from omitted variables bias and various other model mis specifications, as readily demonstrated in our below analysis.

We contrast Ofwat's rigid approach to both the knowledge of water industry managers and engineers we have worked with as well as a wide academic literature (particularly in the case of water). That literature accepts and understands that water and wastewater systems are complex multiple output industries, where complex interactions, and a wide variety of operating characteristics, cost interactions and constraints influence operations and hence costs. Moreover, as became particularly clear in our wastewater modelling work for Anglian Water, complex cost interactions between network and treatment activities will vary both within and between companies, thereby requiring careful consideration of how to best specify econometric models capturing these complex engineering, operational, and economic interrelationships. This will of course also be true for water services where both managers and a wide academic literature accept that the type and location of water sources (which again vary within and between firms) relative to population location influence optimal water acquisition AND network system design.

As experts in both academic and applied modelling in regulated network industries, we understand the need to balance academic nicety against the parsimony and clarity required for regulatory cost modelling. However, regulatory cost modelling should still seek to specify models that account for the economic, operational, and engineering complexity of what is modelled and not abandon such modelling in favour of excessive parsimony. Thus, we strongly believe that Ofwat would have been better served by adopting a regulatory cost modelling approach that adapted academic understanding with regard to cost modelling in complex multiple output network industries, rather than employing its rigid modelling framework, a modelling framework which amounts to little more than establishing a correlation between a company's costs and its scale of operations, and subsequently including a non-exhaustive list of control variables to the specification.

Moreover, we, and particularly Professor Saal, have already made some of these points before as part of the PR2019 Cost Assessment consultation process. We therefore begin our review by considering the key concerns he raised to Ofwat in May 2018, while also providing comment on how Ofwat's January 2019 cost assessment modelling can be judged given those concerns.

Review of Concerns Raised by Professor Saal in his Cost Assessment Consultation Response

Professor Saal raised the three following key concerns with regard to the appropriateness of Ofwat's approach to modelling regulatory cost in his May 2018 cost assessment consultation response[5]:

- **Concerns with the appropriateness of both Botex and Totex modelling, which falsely aggregate operating costs with capital investment, and therefore do not accurately reflect the true economic cost of regulated activities**
- **Concerns with Disaggregated Cost Assessment and Cost Interactions**
- **Concerns with Developing Appropriate Cost Driver Based Cost Assessment Models**

The following section briefly summarises the reasons for each of those concerns, and draws further conclusions on n whether and how Ofwat's approach developed between the 2018 Cost Consultation and the models it provided in January 2019.

In sum, with regard to the first concern, we continue to note significant issues with regard to the use of Totex and Botex based costed assessment, before proceeding to accept this inappropriate cost definition so as to allow further assessment of Ofwat's modelling.

With regard to the issue of excessive disaggregation, we note and approve considerable movement by Ofwat, as they have largely abandoned disaggregated cost assessment. However, we note considerable continuing issues with their modelling approach and its inability to account for cost interactions between activities.

Finally, as Ofwat's January 2019 modelling approach remains largely consistent what the methodological framework provided by its consultant CEPA as well as the models it presented as part of the 2018 cost assessment consultation, there are considerable concerns which can be raised with regard to whether Ofwat has developed appropriate models for regulatory cost assessment.

Thus, as our below review strongly suggests, the concerns raised by Professor Saal as part of the 2018 cost assessment consultation remain relevant. Moreover, our review will systematically demonstrate that Ofwat has employed cost assessment models as part of the IAP, despite that fact that these models can be readily demonstrated to be of inadequate quality for regulatory cost assessment.

---

[5] https://www.ofwat.gov.uk/wp-content/uploads/2018/03/Professor-David-Saal-consultation-submission.pdf

Concerns with the appropriateness of both Botex and Totex modelling

As noted in Professor Saal's consultation response, Ofwat's approach since PR2014 of using TOTEX as its primary cost measure is entirely inconsistent with an economic definition of costs. However, Professor Saal also accepted the constraints imposed by Ofwat's already established regulatory framework for PR2019, and noted that Ofwat's movement to botex modelling was an improvement on using totex modelling,

We are therefore pleased to see that in its January 2019 econometric modelling, Ofwat has continued using only botex modelling. However, we emphasise that there is still potential bias within these models because of the continuing deviation from an appropriate definition of economic costs, Nevertheless, this issue will not be discussed further in this review, beyond noting here that it could result in substantial errors in Ofwat's assessment of appropriate regulatory cost allowances for companies.

We therefore emphasise that the implication of potential biases resulting from the inconsistency between botex and economically consistent definitions of costs should be further considered by Ofwat as it moves to draft and final determinations for PR19

However, having made this important point, the remainder of this review simply accepts botex as the "cost" modelled by Ofwat and goes forward without further considering the appropriateness of that assumption.

Concerns with Disaggregated Cost Assessment and Cost Interactions

Professor Saal's consultation response elaborated this point by emphasising the need for better consideration of the role of cost interactions when defining and modelling disaggregated units for cost assessment, and noting that both the main text and the academic annex A of CEPA's 2011 report for Ofwat (both of which he co-authored) highlighted that the presence of significant cost interactions between disaggregated units of assessment can result in considerable biases if not controlled for properly. Moreover, CEPA (2011) highlighted that as there is considerable evidence that such cost interactions may exist in the water industry, cost assessment and regulatory price determination at inappropriate levels of disaggregation may result in perverse incentives.

He therefore noted concern about Ofwat carrying out cost assessment done below the level of disaggregation that Ofwat has committed to setting price caps. However, he also welcomed that at the time of the cost assessment consultation, Ofwat had supported CEPA's preliminary modelling efforts in this direction, as evidenced by CEPA's provision of models at the water resource plus and bioresource plus level.

Moreover, as it was consistent with his understanding of the most common level of disaggregation of observed water and sewerage services and cost modelling internationally, as evidence by a literature review he had carried out (Saal, et al, 2013) he also indicated that he would be very supportive of an approach to modelling **and** price capping that involved the following maximum disaggregation of wholesale services , while also properly testing and controlling for cost interactions between them:

- water abstraction and treatment (water resources plus)
- water distribution
- wastewater collection
- sewage treatment, sludge treatment, and disposal (bioresources plus)

Consideration of Ofwat's January 2019 cost assessment modelling in light of Professor Saal's cost assessment consultation response therefore leads to a mixed assessment with regard Ofwat's January 2019 cost modelling.

Thus, except for its models for Bioresources, Ofwat relies on a level of disaggregated cost assessment which is consistent with Professor Saal's suggestion. ***We therefore broadly welcome Ofwat's clear choice to abandon highly disaggregated modelling***.

However, Ofwat does not provide models that are consistent with the level it is setting price caps, and we therefore continue to question whether it is appropriate to set price caps and hence incentives at a given level, if it is infeasible or inappropriate to assess costs at that level. ***Moreover, setting price caps at a different level than costs are assessed requires further assumptions and extrapolations which are potentially open to challenge for a variety of reasons***. Thus, while noting that exploring the implications of this particular issue is beyond the scope of this review, due consideration should of this issue is required as PR19 proceeds.

Furthermore, while Ofwat does provide a model of Wholesale Integrated Water Costs that can be used to triangulate/control/check for cost interactions between its disaggregated water models, it provides no such model for Wholesale Integrated Wastewater, which in our opinion is sufficient grounds alone to invalidate its entire assessment of wastewater costs. Moreover, as we will demonstrate below, several of its disaggregated and integrated wholesale models can be shown to have failed to control for statistically significant factors demonstrating evidence of cost interactions between activities. For example, this includes conceptual and statistical evidence suggesting that it is infeasible to assess bioresource costs in isolation form sewage treatment costs, as well as evidence that Ofwat's s integrated wholesale water models are biased due to a failure to control for network activities in its models***.***

Thus, while Ofwat has clearly moved to assessing costs at a more appropriate level of aggregation, its models and triangulation approach do not yet properly control for and/or triangulate appropriately given the cost interactions extant between disaggregated activities. Moreover. this makes it a virtual certainty that its cost assessments are biased.

Concerns with Developing Appropriate Cost Driver Based Cost Assessment Models

We choose not to elaborate Professor Saal's May 2018 concerns with regard to Ofwat's ability to develop appropriate cost assessment models within its modelling framework in great detail in this section. This is because our below review systematically demonstrates significant issues with Ofwat's models.

However, we emphasise to the reader that Professor Saal's commentary noted that as early as our September 2017 review of cost assessment modelling for Anglian Water (Anglian Water 2017), we had raised concerns about the approach being developed by Ofwat's Cost Assessment Working Group. Moreover, his consultation response highlighted that the report he co-authored with CEPA(2011) for Ofwat not only highlighted many of the issues that would impact Ofwat's intended PR2019 cost assessment methodology, but also included an academic annex that provided Ofwat with a basic review of the relevant issues that must be considered when modelling complex multiple output systems.

Thus, while Saal(2018) highlighted several potential issues with the modelling framework that Ofwat had presented in the cost assessment consultation, .it is for our purposes sufficient to focus on the following summary statement:

> *"In sum, it is the conceptual and explanatory quality of the model that matters, and not adopting an a priori modelling approach which assumes that a single output variable and other explanatory variables can and must be used to explain what are complex multiple output systems. Such a modelling approach incurs a high risk of excluding appropriate models that would better capture legitimate reasons for differences between firms' expenditures." (Saal, 2018)*

Unfortunately, as our below review will demonstrate, Ofwat's continued adherence to the modelling approach it had developed with its consultants at the time of the 2018 cost assessment consultants, appears to have resulted in a set of models that can be systematically demonstrated to exclude legitimate and statistically significant explanatory factors, and/or result in models that are not robust to its criterion with regard to sign and plausibility of the magnitude of estimated coefficients. Moreover, it also resulted in Ofwat's failure to provide a model for integrated wholesale wastewater services, despite our below demonstrated feasibility of providing such a model.

We therefore next turn our attention to the consideration of a non-exhaustive set of comments detailing concerns that we have identified with Ofwat's January 2019 cost assessment models and modelling approach.

A Non-Exhaustive Set of Comments on Ofwat's Modelling Approach

According to Ofwat (2019a, p. 5) its final PR19 modelling is based on assessment criteria outlined in their March 2018 consultation on cost modelling, i.e. Ofwat (2018). To develop their models for determining efficient cost allowances, Ofwat engaged with CEPA (and others) whose opinions are outlined in a separate report, i.e. CEPA (2018).

In the final published version of the "Supplementary technical appendix: econometric approach" not all decision rules are made explicit. In our evaluation, we therefore assume that decisions regarding PR19 models used in Ofwat's January 2019 Initial Assessment Plans, were made based on decision criteria previously proposed in CEPA (2018) and Ofwat (2018), where not explicitly mentioned by Ofwat (2019a,b).

The primary purpose of our review is to assess the suitability for regulatory costs assessment of the models estimated by Ofwat, and as stated above, this review accepts Ofwat's approach of treating botex as an appropriate cost measure, despite our concerns about this approach.

We also restrict our evaluation to wholesale water and wholesale wastewater models, i.e. we do not consider residential retail models.

While explicit modelling and estimation approaches are mentioned by Ofwat regarding cost drivers, functional form, estimation procedure, triangulation, and enhancement costs, we do not exhaustively review these in detail. E.g. we only address the detail of those and related issues where relevant in our evaluation.

Moreover, we do not consider issues with regard to the definition of historical cost that Ofwat has chosen to model, nor its treatment of inflation, and the lack of controls for regional variation in labour and other costs. Thus, as with our concerns about the appropriateness of Totex and Botex as definitions of costs that should be modelled, we have simply set these issues to the side for the purpose of this review.

We do however, note one considerable issue, which was highlighted in the 2017 Frontier Economics report for Water UK on productivity trends since privatisation, that Professor Saal co-authored[6]. E.g. the dramatic change in regulatory accounting standards and requirements that Ofwat implemented after 2015, resulted in a considerable inconsistency in data from that date.(Frontier Economics, 2017) While, some might argue that this inconsistency applied only to Frontier Economics' productivity modelling, our discussions with Anglian Water's cost accountants, as well as our own modelling with the PR19 databases strongly suggested evidence for such structural breaks in the data. As such structural breaks would need to be systematically tested for and controlled for, given the panel data approach employed by Ofwat, we simply note, that Ofwat's substantial change in regulatory accounting standards post 2015 coupled with limitations in the ability to accurately

[6] https://www.water.org.uk/news-water-uk/latest-news/water-uk-publish-frontier-economics-report-productivity

restate historic data to be fully compliant with the new standards, may be a significant potential source of bias in Ofwat's modelling.

We finally emphasize that our comments are non-exhaustive, with the time frame available for the review limiting the number of concerns as well as the level of detail that can be provided.

Comment #1:

Ofwat's modelling approach is fundamentally flawed and does not stand up to legitimate principled academic scrutiny. Ofwat neglects the academically well accepted nature of network industries by restricting itself to single-output models (and thereby contravenes its own modelling principles). So even though Ofwat seems to potentially allow for multiple scale (also called output or volume) measures in their models (Ofwat, 2018, p. 9), none of the proposed models for wholesale water and wholesale sewerage include more than one scale variable.

A fundamental characteristic of economic production and cost models, that are consistently applied in the academic literature to model technologies of regulated infrastructure networks, is its multi-output nature. Moreover, this characteristic is so fundamental, that it applies to a variety of networks, e.g., electricity transmission and distribution, gas distribution, water distribution, wastewater collection.

Particularly for business activities that involve physical networks, e.g., wastewater collection or water distribution, outputs such as connected properties or the volume delivered (abstracted), are incorporated jointly with the length of the respective network. By doing so, economic, operational and engineering rationale is acknowledged.

While an updated review of the relevant academic literature supporting these comments is beyond the scope of this report, we point the interested reader to the academic annex of Cambridge Economic Policy Associates (2011) report for Ofwat (CEPA,2011), which states:

> *"The production of both water and sewerage services involves a complex process that is not easily characterised by a single output. Thus, for example, the academic literature provides recent papers in which the aggregate provision of water services by a WoC is modelled as the joint production of water volumes, customer connections, and water transportation. Similarly, Ofwat's past modelling approaches, and particularly the separation of activities in the accounting separation guidelines that came into effect in 2009-2010, provide alternative characterisations of the multiple outputs produced along the water and sewage supply chains. Moreover, consideration of even a single output component of the accounting separation chain reveals that firm activities such as "water treatment" or "sewerage treatment" are in fact the horizontal summation of the multiple outputs produced within the firm's multiple treatment facilities, each with different output characteristics." (CEPA, 2011 p. 97)*

It must therefore be made very clear that restricting the model specification to only one scale (output) variable has fundamental implication and imposes very strong assumptions about the underlying technology (and its associated activities) that is supposed to be modelled. Thus, for example when Ofwat's model for treated water distribution (TWD1) only incorporates length of mains Ofwat imposes the very strong and unrealistic assumption that there are no costs associated with maintain connections to properties or the delivered volume of water delivered to those properties. Similarly, when it models water resource plus costs, after assuming no difference in the output characteristics and costs associated with borehole, river, impounding, and pumped reservoir abstraction, it assumes away, what are generally accepted to be important differences that drive the costs of water abstraction.

This is especially important when it comes to regulatory benchmarking exercises as the efficiency scores of the respective company depends on the choice of scale (output) variable. A simple example can illustrate the inherent issue: Suppose two companies, company 1 serves 20 properties and operates 50 miles of network while company 2 serves 40 properties and operates 80 miles of network. Further suppose the costs of company 1 are 50 and 80 for company 2. Both variables, properties and miles are highly correlated. However, choosing one or the other would yield a different performance score: when length of network is applied, both companies perform at 1£ per miles of network and are therefore equally efficient. This changes when properties are chosen as cost driver: company 1 spends £2.50 per property, company 2 only £2 per property and is therefore more efficient. Certainly, this example simplifies reality but demonstrates that different measures of output yield different performance scores even though both measures are highly correlated. This result would remain irrespective of the estimation method.

Econometrically there is no reason to restrict a multiple-output technology to a single-output technology. Collinearity of scale (output) variables does not sufficiently justify this restriction per se, as implied by Ofwat (2018, p. 9). This is readily demonstrated by referencing the academic annex of CEPA's (2011) report to Ofwat, which notes that a substantial, and subsequently expanded, academic literature regularly models water companies with multiple outputs. Moreover, we would emphasise that those results and models withstand the very demanding scrutiny levels required for publication in peer reviewed journals.,

Further, accounting for multiple scale (output) variables does not introduce complication nor does it compromise the practicability of regulatory benchmarking, one of the principles Ofwat follows (2019a, p. 5). Other European regulators, for example the Bundesnetzagentur (BNetzA) in Germany and the Norges vassdrags- og energidirektorat (NVE) in Norway, consistently rely on models of multi-input-multi-output technologies. Thereby the choice of variables respects economic and engineering considerations simultaneously as, for example, suggested in SUMICSID (2007, 5.156). Moreover, to our knowledge there is no reason to believe that scale (output) variables are notably less correlated in other countries, than they are in England and Wales.

Moreover, our empirical assessment of Ofwat's models demonstrates with simple examples that it is feasible and appropriate to augment Ofwat's models via simply adding an additional output to improve its specification. Thus, for example, augmenting its underspecified Integrated Wholesale Water models by adding network length to the connected properties variable it employs, demonstrates not only the feasibility of modelling with multiple outputs, but also the omitted variables bias caused by Ofwat's insistence on always using a single output specification

We will further address alternative scale (output) measures in our discussion on triangulation.

Comment #1a:

Ofwat's modelling approach is arbitrarily simplistic. A consequence of single-output models is that they do not allow modelling complex systems and cost interactions that determine costs within it.

Again, Ofwat's modelling principles would in principle allow for non-linear and cross-product terms (Ofwat, 2019a, p. 7). Their final models, however, only allow for non-linear terms for wholesale water when a squared term of the weighted average density measure is included.

> *As no other non-linear or cross-product terms, nor a second output variable, can be found in any (!) of the water and wastewater models and Ofwat provides, we can only conclude that Ofwat still heavily relies on the modelling principles of CEPA that suggested to use only one variable to be used for each "group of cost drivers", e.g., scale (CEPA, 2018, p. 40).*

Consequently, Ofwat's models are highly restrictive with regard to variation within firms due to (plant size) economies of scale, or variations in cost elasticities related to factors such water source, or indigenous/nonindigenous treatment of sludge, nor interactions between variables. However, all of these are important sources of complexity that potentially influence the underlying cost elasticity relationship between the scale of a company and its costs.

Moreover, as the academic appendix of CEPA (2011) highlights, scale economies in complex systems result from a complex relationship and cost interactions between multiple outputs and will vary depending on those relationships. A single non-interacted scale variable cannot capture that complexity.

Hence, even without evaluating Ofwat's actual models, it is possible to conclude that they are extremely unlikely to provide conceptually appropriate specifications and are therefore extremely likely to suffer from misspecification and omitted variables bias.

We emphasise that while not summarised again in the above section, this issue had already been raised in Professor Saal's our response to Ofwat's cost assessment PR19 consultation (Saal, 2018).

Comment #2:

Because Ofwat effectively limits itself to specifications that amount to specifying a scale variable that will by definition explain most of the observed intercompany variance in costs, and further limits itself to the use of an extremely limited number of simplistic control measures for complexity, a single measure for topography, and controls density, (while providing no credible reason to justify this modelling restriction), its models are highly likely to provide biased estimates of the determinants of company costs.

Our argument rests on the observation that Ofwat's modelling approach effectively specifies a relationship between cost and the single allowed scale variable and therefore allows no deviation in the estimated cost elasticity faced by firms with regard to that scale variable. Instead, deviation in the cost response of firms is only allowed via the noninteractive control variables, which amounts to an assumption that all the complex factors explaining differences in company costs, can be modelled by simply shifting the estimated relationship between scale and cost up or down.

However, our understanding based on over 30 years combined experience in modelling network industries, is that the factors being controlled for generally have a much more significant/complex impact on a firm's costs than what is captured by modelling them as shifters of a constant, that does not fundamentally alter the core relationship between scale and costs at the heart of Ofwat's modelling approach.

A brief discussion of Ofwat's wastewater modelling highlights this point. Ofwat appears to believe that simple controls for the intensity of treatment (proxied only by ammonia consent requirements) the share of load treated in different sized plants and controls for density, coupled with the use of disaggregated models can explain all the variation observed in intra and interplant wastewater costs. However, a more fully developed conceptual framework would reveal that the avoidance of excessive pumping and other network transportation costs is likely to be a significant reason for not consolidating sewage treatment plants, and choosing to accept the cost implications associated with smaller treatment plants and nonindigenous treatment of sludge. E.g. it is the cost interactions between activities and how a company optimizes its overall waste water costs that determines pumping and plant scale, and not pumping and plant scale that determines costs.

As Ofwat's modelling framework effectively precludes modelling such complexity appropriately, its resulting models are therefore highly likely to result in biased cost assessment models.

Comment #3:

Ofwat abandons previously adhered to model selection criterion with little or no discussion of the implications, and no acknowledgement of their change of practice is provided in n the final reports Ofwat (2019a,b) Moreover, in the case of both multicollinearity statistics and its use of Ramsey reset tests to test for potential model misspecification, Ofwat's appears to further disguise changes in its earlier practice, by simply failing to report the relevant test statistics in its reports. Furthermore, these test statistics are only recoverable via running the Stata do files that Ofwat provided, which is not consistent with a regulator seeking transparency.

Multicollinearity - As part of the consultation process, Ofwat and its consultant CEPA gave every indication that models with low multicollinearity were a priority, and suggested excluding models for which a test on multicollinearity results in a VIF value larger than 10.

Moreover, Saal(2018) took exception to this noting that:

> *"Unequivocally, this approach is unnecessarily restrictive, and thereby deliberately and arbitrarily excludes the potential inclusion of statistically significant variables that are legitimate and important potential determinants of differences in firm's costs, that may be necessary when modelling the complex cost determinants within a network industry. Moreover, simple reference to the Stata reference manual section on VIF tests (p. 2275) indicates that "some choose a more conservative threshold value of 30" i.e. the arbitrary use of an effective VIF criteria of less than 5 suggests that an appropriate assessment of the trade-off between improved and statistically significant model specifications, and the*

resulting statistical costs created by increased multicollinearity, did not form part of the modelling strategy employed by CEPA" Saal(2018)

Given the specifics of this comment, and the previous focus on a draconian approach to multicollinearity, we find it extremely interesting that every water model specified by Ofwat now has a VIF test statistic in excess of 200, but Ofwat has provided no explanation for this change in its modelling criterion. We cannot be certain, but we are quite confident that Ofwat has made this change because of its decision to rely on weighted average Local Area density in its models and also despite the fact that this violated its previous modelling principles with regard to multicollinearity.

Ramsey Reset Tests - Ramsey Reset misspecification tests have long been applied to regulatory cost benchmarking, and we found it fully acceptable and appropriate that Ofwat and CEPA's modelling framework appeared to acknowledge the importance of this test for potential model misspecification. Moreover, we note that Ofwat required companies to report the Ramsey reset test and VIF test statistics along with the models that they submitted to the 2018 cost assessment consultation exercise. However, Ofwat's cost assessment reports do not provide this statistic for any of its models, even though at least one of the random effects models it used in earnest for regulatory cost assessment (BR2) demonstrates evidence of model misspecification**.**

Comment #4:

Ofwat claims to apply Cobb-Douglas models and incorrectly pretends this would overcome issues of stability, multicollinearity and difficulty over interpretation associated with Translog models.

We respect the regulators choice of Cobb Douglas over Translog models, based on the CMA Bristol Water PR14 case. However, we do not believe that this respect of the CMA's opinion should translate to employing only arbitrarily simplistic model specifications.

The consequences of restricting the modelling to Cobb Douglas functions is that cost elasticities are assumed to be the same for each company, irrespective of their business and plant sizes. As we can see no reasonable economic, engineering, or regulatory reason to assume that all firms operate with the same economies of scale, this assumption is highly likely to result in biased estimation of costs.

Ofwat mentions that companies agreed with Cobb Douglas specifications (Ofwat, 2019a). However, the reasoning Ofwat gives seems to be a misconception regarding the issues of multicollinearity and interpretation of the models.[7]

Moreover, the potential issue of multicollinearity does not go away in Cobb Douglas specifications as such. It is an issue related to the underlying data (variables), which does not vanish when using Cobb Douglas instead of Translog. Moreover, as discussed above, Ofwat's own approach to modelling

[7] The fact that several of the models submitted to Ofwat by companies as part of the cost assessment consultation used squared or interactive scale variables to model costs is also not consistent with this assertion (https://www.ofwat.gov.uk/wp-content/uploads/2018/03/Appendix-1-Modelling-results_Final.pdf)

water costs now contradicts this assertion given high multicollinearity in its January 2019 cost assessment models.

Furthermore, dropping variables with the intention to reduce the issue of multicollinearity is an option but potentially introduces a strong bias due to omitted variables bias (Wooldridge, 2006, p. 104). Ofwat (2019a, p. 21) acknowledges that the coefficients could be potentially biased by the omitted variable problem but seems to care less about it when using very restricted number of explanatory variables per se. Favouring one bias over the other would either require a theoretical or an empirical justification that is based on the actual data and the obtained regression results. At least, for the sake of transparency, Ofwat should mention that this trade-off exists and that has consciously made this choice.

It seems that a modelling approach from “general”, i.e. more variables, to a “specific”, i.e. less variables, is a better approach instead of going from “specific” to “general” as applied (CEPA, 2018, p. 39). Academically, there is no reason to prefer one bias over the other as their impact on the coefficients if not clear. Besides that, omitting variables can lead to severe misspecifications (also see comment #1).

Regarding the difficulty to Ofwat’s assertions that non-linear and cross terms are difficult to interpret, which is then used as an argument against Translog specification, it seems implausible why a model including the weighted average density and its squared term (in models WPR1/2, TWD1, WW1/2) should be easier to interpret as a model when compared to, for example, included a squared scale (output) variable, or interactions between variables .

Moreover, Ofwat appears to ignore standard academic practices such as demeaning data when using interactive variables that can be readily applied to allow for easily interpretable coefficients, which we regularly demonstrate to our students .

Thus, rather than being based on any consistent criterion, Ofwat’s modelling selection criterion appear to be based solely on a low willingness to provide comprehensible explanations for including certain terms.

Comment #5:

Ofwat claims to model with only a single scale variable, and control variables for complexity, topography, and density, but in several cases the actual underlying models it specifies are really poorly specified multiple output “Cobb-Douglas models” which, not only violate its stated modelling framework, but also provide underlying parameter estimates that violate its criterion with regard to not using models with inappropriately signed or sized coefficient estimates.

As discussed below in our detail review of Ofwat’s models, we find that Ofwat’s only model for treated water distribution (TWD1) is estimated with a suspiciously high estimated cost elasticity of 0.465 for booster pumping stations, and a suspiciously low cost elasticity of 0.549 for mains length, which is not consistent with our expectations of the relative cost implications of these variables.

Similarly, Ofwat's first sewage collection model, (SWC1) can be shown to have an underlying estimated cost elasticity of -0.901, while its SWC2 model has implausibly similar underlying cost elasticity estimates for sewer length (0.369) and pumping capacity (0.346).

Thus, it is readily demonstrable that Ofwat does not carefully consider and/or understand the implications of the mathematical specifications it employs in at least 3 of the 13 models it specified, and that these models are in fact highly suspect if the parameters are appropriately interpreted.

We believe this occurs because Ofwat's modelling approach sometimes effectively treats what should legitimately be control variables as outputs.

Comment #6

Building from Comment #1 and in contrast to Comment #5, Ofwat fails to include additional scale variables when it is conceptually appropriate to do so due to the underlying multiple output characteristic of what is being modelled. Moreover, as it can be readily demonstrated that these additional scale variables are statistically significant when added to some of Ofwat's models, their exclusion, demonstrates that at least some of Ofwat's models suffer from omitted variables bias, due to the failure to control for multiple outputs properly.

For example, Ofwat includes only connected properties in its models of Integrated Wholesale Water but does not include the lengths of main variable that it uses a proxy for the sewage collection output in its treated water distribution models. However, as we readily demonstrate below, the simple inclusion of network length as an additional output scale variable, provides models in which this variable is not only statistically significant, but both the properties and length of mains variables have appropriate cost elasticity estimates.

Similar arguments therefore suggest, that not only Ofwat's integrated wholesale water models (WW1 & WW2) but also its water distribution model (TWD1) are subject to omitted variables bias due to its failure to include multiple outputs in its specification, in a manner consistent with the specifications it uses for having different outputs for the disaggregated models it specifies.

Comment #7

It can be further demonstrated that adding a single control variable to those employed by Ofwat in both its bioresources its water resource models and applying a conventional general to specific modelling framework to model selection, results in the choice of an alternative preferred specification to those chosen by Ofwat.

Moreover, in the case of Integrated Wholesale Wastewater, the addition of a single variable to the set of variables used by Ofwat to model Sewage collection and Bioresources Plus, and again applying conventional general to specific modelling demonstrates that contrary to Ofwat's claims, it is perfectly possible to work within its modelling framework and provide a model of Integrated Wholesale Wastewater costs

Thus, even if we adhere to Ofwat's single scale variable framework, it is readily demonstrable that Ofwat's model selections and conclusions with regard its chosen models are not robust, its models

and model selection process are subject to omitted variables bias, and its cost assessment results are therefore not appropriate for use in regulatory cost assessment.

As demonstrated below, the simple inclusion of a control variable for the share of treatment load taking place in plants where sludge is not treated indigenously, to the set of variables used by Ofwat, results in selection of an alternative model than that chosen by Ofwat, when general to specific modelling is employed.

Similarly, inclusion of the same variable to the set of variables employed by Ofwat for modelling sewage collection and bioresource plus modelling and employing general to specific modelling yields a model specification satisfying reasonable model selection criterion. Thus, Ofwat's claims that it cannot provide a model of integrated wholesale wastewater costs are readily refuted.

Finally, simply adding the control variable employed by Ofwat to control for booster pumping stations to its water resource plus models, on the grounds that water resources require pumping and should be controlled for, and applying standard general to specific modelling, reveals that the weighted average density variables that Ofwat employs are no longer statistically significant, thereby demonstrating that its specifications for water resources plus are not robust to a simple test for omitted variables bias.

Comment #8 – Ofwat's January 2019 cost assessment methods have (largely) moved in the right direction by focussing on the assessment of costs for sewage collection, bioresource plus, water distribution, and water resource plus.

This approach is consistent with comments made by Profess Saal in his March 2018 consultation response to Ofwat which emphasised that there are significant issues cost interactions between different parts of the value chain, and both the accuracy of accounting cost allocations as well as the ability to consistently model distinct cost output relationships at the level of highly fragmented cost assessment once proposed by Ofwat.

Moreover, this level of disaggregation is consistent with not only the academic water industry cost assessment literature, but also the level of disaggregation that is generally observed outside of the UK in real world companies.[8]

Comment 8a – But Ofwat does not provide appropriately specified models of wholesale water and wastewater costs that can be used to respectively cross check and control for cost interactions between water resource plus and water distribution, and bioresource plus and sewage collection.

Thus, as already discussed Ofwat does not even attempt to provide a model of Integrated Wholesale Wastewater costs, with which to cross check its disaggregated findings, and as per our above comments Ofwat's model of wholesale water costs can be shown to suffer from omitted variables bias because of its failure to properly account for the multiple outputs inherent in integrated water system.

[8] https://www.degruyter.com/view/j/rne.2013.12.issue-1/rne-2012-0004/rne-2012-0004.xml

Properly specified aggregate model are required to cross check the disaggregated results for at least the four following reasons

1. Incorrect Cost Allocation between activities in the dependent variables for separated functions

2. Inappropriate output specifications which cannot capture the distinct outputs produced in each separated activity, and/or the total activity (Ofwat sometimes uses same output for all levels!)

3. Imposing the condition of strict separability of costs when cost interactions actually exist, and the goal of the company should be to minimise overall costs, not the sum of the arbitrarily defined separated costs

4. Other measurement and prediction error compounded by the misspecifications created by the above factors and inadequate modelling at disaggregated level.

Comment #9:

Ofwat's triangulation fails to account for alternative models and triangulates models that do not meet minimum acceptable standards for model quality.

We acknowledge that Ofwat aims reduce imperfections that are inherent to statistical modelling in cost assessment for regulatory purposes by triangulation and in principle agree with the overall approach.

The purpose of triangulation is to combine outcomes of alternative models, which are alternative and credible representations of the determinants of costs. However, consideration of Ofwat's bioresource models reveals that they do not meet this criterion.

Thus, both the BR1 and BR2 models include sludge as a scale variable and only differ by the addition of a single alternative variable controlling for density. As discussed above, BR2 fails a Ramsey Reset test and does not meet acceptable standards for inclusion in regulatory cost assessment.

In contrast, BR1 passes a Ramsey reset test, and the regression results in Ofwat (2019a, p. 17) show that in this model only the scale (output) variable shows a significance level greater than 10 percent. However, the other included control variable, i.e. weighted average density, exhibits a level of statistical significance larger than 10 percent, i.e. 12 percent.

We acknowledge the regulator's opinion on not strictly following scientific thresholds (as an example Ofwat mentions a threshold of 5 (!) percent, i.e. accepting to fail predicting the coefficient correctly with probability of 5 percent in Ofwat, 2018, p. 9) in favour of the practicability of the models. However, a significance level that larger than 10 percent seems to be questionable, particularly when the model being considered only includes a single other variable.

Given academic standards, the density variable in model BR1 could not be considered as having an significant impact on costs (as the coefficient is not statistically different from zero at a 12 percent significance level). Further, the constant in this model is also statistically insignificant, at least at the 10 percent level. This has two important implications:

Thus, if BR1 effectively collapses to a model that only includes scale variable as the only explanatory variable, and BR2 provides a model that fails the Ramsey Reset specification test, Ofwat is effectively triangulating a model that uses a single scale variable to predict costs (BR1) with a model that fails to to meet acceptable standards for use in regulatory cost assessment .

Therefore, triangulating BR1 and BR2 is highly inappropriate. Moreover, we emphasise that this affects all higher aggregation levels, and therefore cost allowances, which these models contribute to.

Considering another case, we note that when taken at face value, Ofwat's triangulation of its sewage collection models seems appropriate as the reported SWC1 and SWC2 appear to be statistically robust and use alternative measures of density (Ofwat, 2019a, p. 22). However, as our discussion in Comment #4 has already revealed, both models, and particularly SWC1 have underlying cost elasticity estimates that are implausible. Thus, Ofwat has again triangulated models that do not meet acceptable standards for regulatory cost assessment.

Comment #10

Even when Ofwat's triangulation occurs over models that meet basic modelling criterion, its triangulation is suspect because its models never vary with regard to the key scale variable that explains most of the variance in its models. Instead, its models only vary in the specification of a single control variable which generally contributes no more than a few percent of the overall explanatory power of its models.

Our below review of Ofwat's modelling reveals that it is not only feasible to specify a model of integrated wholesale wastewater costs using Ofwat's chosen measure of scale, which is treatment load. Alternative models using either the properties variable Ofwat favoured as a proxy for both water resources plus, and integrated wholesale water, or alternatively total population served, were successfully specified and could be used to provide meaningful triangulation with variation in the key scale variable explain most of the model variance.   Similarly, our below modelling demonstrates a volumetric alternative to Ofwat's specification of properties as the output proxy for water resources plus, and while not reported below, such an alternative specification is also potentially feasible for outputs-based triangulation of its integrated wholesale water models.

A benefit of this approach is that it could partially address the multi-output nature of network-based business activities (discussed in comment #1) and reduces the bias that is imposed by selecting only one of the potential multiple outputs produced by the firm.

Comment #11

Ofwat's triangulation does not accurately control for cost interactions between cost assessment at the aggregated level, and its disaggregated modelling. Moreover, Ofwat does not even triangulate its wastewater modelling because it fails to provide an integrated wholesale wastewater model to triangulate its disaggregated models with.

CEPA(2011) argued that one of the most important reasons for triangulation is to insure that potential cost interactions and accounting cost allocation inconsistencies that may exist between activities that are assessed separately are accounted for in aggregated model that will more readily capture and control for such effects.

However, Ofwat's PR19 models do not meet these objectives on at least two grounds. Firstly, as discussed above, Ofwat fails to provide an integrated wholesale wastewater model. It therefore does not even attempt to triangulate its disaggregated wastewater models. Moreover, we emphasize that this is even though our review has demonstrated that such a model is feasible to generate within Ofwat's modelling framework.

Secondly, as Ofwat always chooses single-output models, its integrated wholesale water model, does not add up to the sum of its parts, by which we mean that if it is appropriate to model water distribution with length of mains as a proxy for water distribution output and properties as a proxy for water resources outputs, an integrated model including both these activities should in principle include both of these outputs. Moreover, the conceptual logic of this argument is demonstrated empirically below, as such two output specifications provide what we believe are robust alternatives to Ofwat's integrated wholesale water models.   In fact, as discussed above, these two output models demonstrate that Ofwat's integrated wholesale water models are not appropriate models for triangulation nor direct cos assessment as they can be demonstrated to suffer from omitted variables bias.

Comment #12

Ofwat's reliance on weighted density in all its water modelling suggests a further lack of appropriate modelling and triangulation

Given the implications of Comment #7 (Ofwat's density measures can be shown to be jointly insignificant in a water resources model variant derived from the set of variables it uses for water modelling) and Comment #3 (extremely high multicollinearity in its specifications) the reader should already be concerned about Ofwat's heavy reliance on of Local Area District (LAD) based weighted density measures. However, we wish to further emphasiSe that as Ofwat uses the same density formulation in all of its water modelling and does not triangulate alternative output specifications, the triangulation of its water models amounts to little more than testing the implications of the alternative measures for treatment complexity. E.g. triangulation on measures which add a very small proportion to the explanatory power of Ofwat's models.

Comment #13

Alternative approaches to triangulation could be employed by Ofwat

Additionally, diverse estimation methods could be triangulated: The German regulator BNetzA, for example, relies on the "Best of four" principle, which can be considered as a triangulation of model alternatives. The German regulation for electricity distribution companies relies currently on two DEA (Data Envelopment Analysis) and two SFA (Stochastic Frontier Analysis) models for efficient cost estimation. By law it is determined that the best efficiency score out of the four models would be used with the intension to overcome particularities and limitations of the underlying estimation procedures.

Comment #14

Ofwat's heavy reliance on the weighted average of Local Area District (LAD) population density is potentially suspect on conceptual and empirical grounds and may result in biased estimates within its water modelling. This suggest the need for alternative density data and/or alternative modelling approaches to controlling for density with which to confirm that Ofwat's water models are robust.

LAD density measures are based on the 398 LADs in the UK and therefore provide by far too course a measure of density to be suitable for modelling the determinants leading to the engineering and managerial choice to optimise wastewater system design.

Our below discussion of wastewater modelling clarifies that the over 6300 sewage treatment collection and treatment plant systems observed in England and Wales is consistent with managers making wastewater system optimisation decisions at a level of disaggregation far below that observed at the level of aggregation implied by the use of LAD based density measures. E.g. this reveals the existence of many distinct waste water collection and treatment systems operated within LAD boundaries. Moreover, the irrefutable evidence of clear differences in factors such as local wastewater system design, treatment plant scale, and decisions about the indigenous treatment of sludge, within LAD boundaries, demonstrates that decision making is made for population settlements much smaller than LADs.

E.g. the trade-off between the costs of transporting more sewage and the resulting benefits of increased treatment plant scale, are optimised at the very local level of population settlement, not the LAD level. Thus, it is not surprising that weighted average density is statistically insignificant in the only wastewater model Ofwat directly employed it in, and also drops out as statistically insignificant in the integrated wholesale wastewater models we developed as part of this review.

However, while further analysis would be required on this issue, these concerns can also potentially be raised about Ofwat's water modelling. E.g., as the population settlement patterns determining wastewater system design are the same as those influencing local water distribution systems as well as the transportation of treated water to population centres, there is reason to believe that Ofwat's LAD based weighted density measures may also be too course to rely on for modelling water costs.

Thus, even if LAD density measures are statistically significant, they may not in fact measure density at the appropriate level to capture many of the differences in water industry costs. This observation is consistent with the peculiar always increasing cost elasticity for density Ofwat describes for its water models, which is contrary to expectations that cost would first decrease before increasing as density increases. Thus, it is unclear why Ofwat fully abandoned its previously developed measures for sparsity and density, which were based on a much finer (more local) definition of population density.

Comment #15

While our below review has simply accepted Ofwat's underlying econometric approach including the use of random effects estimation, and the use of clustered standard errors, we do not believe that Ofwat has adequately justified these choices. We therefore offer a few brief observations about these econometric choices, the implications of which should be subject to further review and consideration as PR19 proceeds.

- While we cannot be sure of its actual modelling approach, Ofwat appears to have modelled by simply choosing its scale variables, selecting what it believed by *a priori* assessment to be an appropriate group of control variables, and then running the resulting models. As consideration of its Stata codes and modelling reveal, this resulted in the majority of its models failing Ramsey Reset tests when specified with OLS but passing this test when estimated with Random Effects. As the use of random effects estimation is a standard econometric approach for controlling for unspecified firm level heterogeneity (Greene,2012 pp. 385-7) this result is not surprising. However, the resulting random effects are, in fact, likely to be controlling for the many factors and issues with Ofwat's models that we have discussed above. E.g. there is no reason to believe in or favour Ofwat's interpretations of the residuals resulting from its random effects models as evidence of cost inefficiency, as opposed to the more standard interpretation of these firm specific residuals as controlling for unspecified homogeneity in the factors determining company costs.

  Stated differently, we highly suspect that Ofwat's interpretation of its random effects residuals as evidence of inefficiency as opposed to evidence of unmodelled factors influencing company costs, is incorrect.

- Ofwat's random effects models assume that measured "inefficiency" is constant over the 7 years modelled. Moreover, Ofwat assumes that there is no technical change, or stated differently that the coefficients of its models have not changed over time, change that would allow for deterioration or improvement in the underlying cost relationships faced by firms. Moreover, Ofwat, provides no justification for why this very strong assumption is appropriate, over the period modelled when companies were subject to price caps designed to incentivize both efficiency catch up and frontier shift by firms.

- Furthermore, the lack of for time, not only assumes that no technical changes has occurred, but also assumes that changing economic and regulatory conditions, including the profound change in regulatory cost accounting that occurred after 2015, has no effect on firms and/or the data. This is particularly important as our work with Anglian Water, as well as the recently published report for Water UK co-authored by Professor Saal, both encountered significant discontinuity in the data, after 2015 (Water UK, 2017). There is therefore also a lack of an adequate justification as to why time controls are not required , and how this omission influences the appropriateness of Ofwat's models and the appropriateness of how it applies these models in its subsequent assessment steps

- While Ofwat does have relatively small panel databases consisting of repeated observations of the same firm, its models can still be considered excessively parsimonious as it never has respectively less than 118 and 66 remaining degrees of freedom in its water and wastewater models. This supports the thrust of the above comments which suggest that Ofwat's modelling approach is excessively parsimonious, and favours reduced variable quantity over model quality.

- Ofwat relies on clustered standard errors as the basis of the statistical inference with which variable and model selection are driven. As our colleagues (Glass & Glass, 2018) pointed out in their own submission to Ofwat's cost assessment consultation, such clustered standard errors, can provide excessively robust measures of statistical evidence, particularly in relatively small samples. As a result, we have therefore accepted the legitimacy Ofwat's models including moderately statistically insignificant variables, but as this issue is not adequately discussed, we are unsure how Ofwat's approach to statistical inference has actually influence Ofwat's choice of appropriate models.

- Ofwat does not consistently employ a set of criterion with regard Ramsey Reset model specification tests, multicollinearity and decision rules with regard to statistical inference in its models, and has changed/or abandoned its emphasis on some of these measures since the cost assessment consultation, without explaining that it had done so,, let alone why it had done so.

- Ofwat should employ a much broader range of model specification tests in considering the quality of its models

Comment #16

Although related to the issues raised in Comment #15 we must further emphasise that Ofwat's modelling framework is also ill suited to be used to project costs forward into the PR19 regulatory control period

In addition to preventing adequate modelling of the complexity that determines water and wastewater costs, Ofwat's static modelling approach assume that inefficiency is constant across the 7 years of the sample, that no improvements in technology occur, and that the factors which influence changes such as population growth, increased leakage control, etc. do not vary significantly because of the local operating characteristics of the firm.

Moreover, many of its models simply do not include factors such connected properties or water volumes that would be required to project forward into the next review period accurately. Related to this, Ofwat, does not adequality explain why it is appropriate to subsequently violate the assumptions implicit in its random effects modelling (see comment #15), when it subsequently uses them to project forward in its assessment of required costs during the PR19 control period. Thus, in complete violation of the assumptions imposed in its econometric modelling, Ofwat's projections not only assume that at efficiency catch up can occur but also assumes that positive technical change will take place.

As such, this contradiction in its methodological approach suggests that it is highly likely that Ofwat's projections of future cost requirements during the PR19 period, are very likely to be inaccurate.

Moreover, as our above comments have highlighted, many of Ofwat's models can be demonstrated to be inadequate due to issues such as model misspecification, omitted variables bias, and relying on underlying estimated cost relationships that have implausible signs and/ or magnitude, and therefore cannot be considered appropriate for current cost assessment let alone future cost assessment.

We also note a contradiction in Ofwat's approach to employing evidence from studies that use an empirical methodology that allow for changes in productivity over time, Thus, In October 2018 Rachel Fletcher, Ofwat's Chief Executive Officer, cited evidence from Frontier Economics' recent report for Water UK (Frontier, 2017) on productivity trends to support her case for "why change is needed" in the industry. In contrast, as discussed above, Ofwat's modellers employ a static modelling framework where no change in efficiency or technical change is employed, despite the fact that the very report cited by Rachel Fletcher, included a section suggesting how panel data might be employed to allow better estimates of productivity change and its decomposition into efficiency change and technical change (continuing efficiency improvement). Moreover, the brief literature review in Frontier (2017) identifies several papers where such methodologies have been employed in the English and Welsh water industry.

We therefore emphasize that it is our opinion that Ofwat should better employ its panel data sets to provide evidence on past trends in productivity growth, efficiency change, and technical change so as to better inform its estimates of potential improvements in the future. Furthermore, we strongly emphasise that such an assertion is not inconsistent with a regulatory approach to cost assessment as Professor Saal was commissioned by Ofwat itself (Stone and Webster, 2004) during the 2004 price review to provide econometric evidence of past productivity trends to help support Ofwat's assumptions with regard to potential continuing efficiency improvement.

Comment #17:

Ofwat actually predicts a different future based on the past.

From our understanding of Ofwat's modelling of projected costs for the PR19 review period, its projection methods are also inappropriate for another reason. Thus, Ofwat models using botex, which is net of enhancement expenditures, but then uses these models to project forward in a manner that is used to evaluate company costs that include enhancement costs.   This is not appropriate, as the models employed were specified and estimated to fit a different cost relationship than that they are being used to project forward.

Comment #18:

Ofwat uses a simplistic forecast of cost drivers.

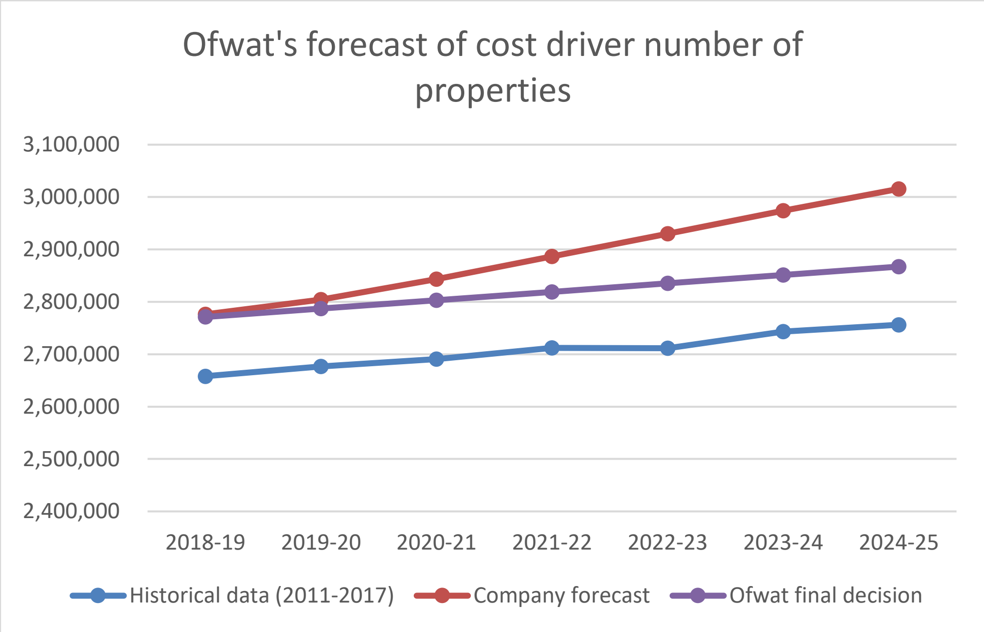
Cost drivers are forecasted based on a linear relationship (OLS) and historical data from 2011-2018 (Ofwat, FM_WWW3, forecast). This prevents taking non-linear population growth and structural changes within the population, e.g., changes in demographics, into account. not allow for preparing for, e.g. growth in population.

While Anglian Water predict an over-proportional growth in the number of properties (see red line in Figure 1), Ofwat predicts a much flatter increase in properties given historical data (see blue line in Figure 1) and therefore assumes the exact same growth rate of properties for the time horizon 2018-2025 (see purple line in Figure 1) as observed over the time period 2011/12-2017/18.

Assuming a straightforward correlation between population growth and the number of properties (more population means more houses) and a likewise potentially straightforward correlation between the structure of the population and the number of properties (more elderly need more housing/bungalows), Anglian Water's prediction seems to much better capture the population growth that has been predicted by the Office of National Statistics for the South East region (see Figure 2).

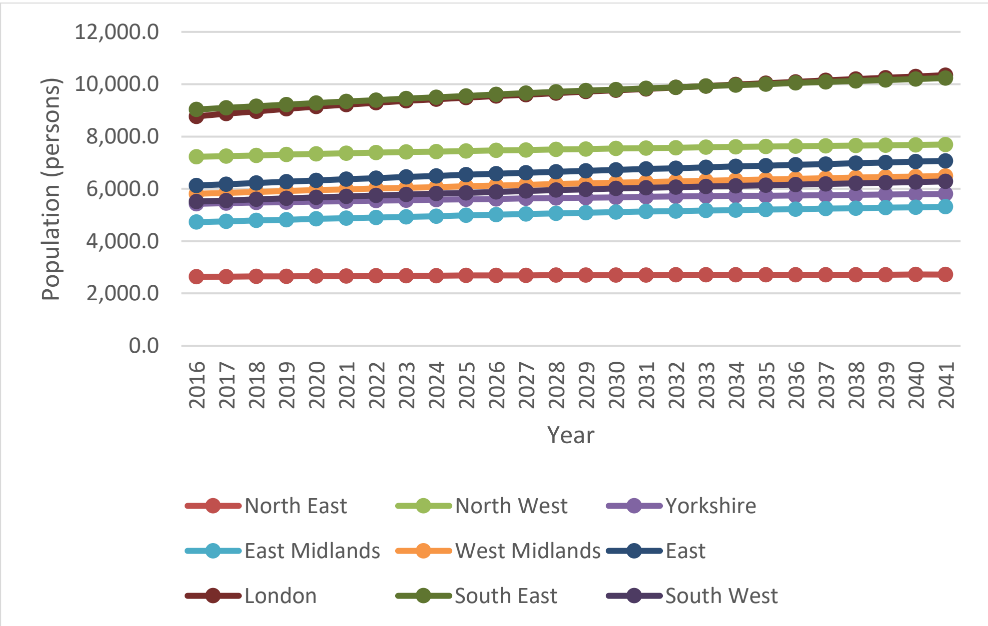
The predictions by the Office of National Statistics forecast an over-proportional population growth for the South East (see dark-grey line in Figure 2), comparable to the growth expected for London.

Figure 1: Ofwat's forecast of cost driver number of properties



Source: Ofwat, FM_WWW3, spreadsheet "Forecast".

Figure 2: 2016-based subnation population projections 2016 - 2041



Source: Office for National Statistics: Population projections for regions: Table 1 "2016-based subnational population projections for regions in England". Url: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationprojections/datasets/regionsinenglandtable1.

Part 2 - Empirical Assessment of Ofwat's Wholesale Water and Wastewater Models

Assessment of Ofwat's Wholesale Water Models

Overall Assessment

When compared to Ofwat's suite of Wholesale Wastewater models, its Wholesale Water models appear relatively strong at the superficial level. Thus, Ofwat specifies a total of five models, all of which pass the Ramsey model specification test, have appropriately signed coefficient estimates, and with the exception of its density controls, always have statistically significant coefficients. Moreover, the estimation of Water Resources Plus, Distribution, and Integrated Wholesale Water not only provides a set of triangulated models in which disaggregated costs can be checked against aggregated costs, but also does so at a level of disaggregation that is consistent with what is observed in the academic literature and allows modelling of clearly defined activities with identified outputs. Moreover, this level of modelling is consistent with Professor Saal's comments in his response to the 2018 cost assessment consultation (Saal, 2018), as well as the focus on more aggregated models taken in Anglian Water's 2018 cost modelling report (Anglian Water, 2018).

However, without even engaging in modelling to test Ofwat's models, a myriad of limitations and weaknesses are apparent. The below is a partial list of these issues, with a more exhaustive list not feasible given the time constraints we face in preparing this review.

- **Ofwat's models do not allow for sufficient controls to capture the complex factors influencing water supply costs.** When taken together as a whole, Ofwat's models use no more than a single scale variable, a density control (in logged levels and squared log levels), a single control for treatment complexity, and a control for booster pumping stations. Factors such as significant differences in water resource and treatment costs attributable to the type of water source employed, regional variation in water scarcity, and substantial variation in company efforts to reduce water demand and scarcity via leakage control and water metering are only a few examples that quickly come to mind. Thus, compared to both the CMA's models in the Bristol Water Price review and the academic literature, Ofwat's models do not appear to have controlled sufficiently for the complexity of water supply. ***Ofwat's models should therefore not be considered as robust until alternative models with reasonable control variables are demonstrated to be either inappropriate and/or consistent with its results.***

- **Ofwat's models are unlikely to accurately capture the impact of differences in population density on water supply costs**. In all of its models Ofwat relies exclusively on the specification of a single scale variable, and its density controls to capture these effects. While significant in all other models, the density variables are not individually significant in the WRP2 Water Resources model. Moreover, its discussion of its estimated elasticity of cost with respect to density and the figure it provides illustrating this (Ofwat, 2019x, p. 15) suggest that the elasticity of cost with respect to density is always increasing. This is contrary to the general understanding that in network industries, costs generally decrease as density increases before some critical level where congestion leads to increased costs. Moreover, we note that Ofwat's limited conceptual discussion of how density influences costs also contradict this general understanding. Thus, it focuses solely on density providing potential scale economy benefits in treatment scale being offset by higher property, rental, and access costs (Ofwat, 2019b, p.143), while completely ignoring that serving properties with lower density is likely to require higher network transportation costs to serve customers. Thus, we interpret Ofwat's interpretation of how density influences costs in its models as justifying its results, rather than insuring that its results are consistent with the J-shepaed impact that economists, managers and engineers would expect.

Moreover, there is an extremely well-established academic literature on modelling and controlling for economies of density in multiple output network industries that provides alternative and widely accepted approaches to that employed by Ofwat.[9]

We also note that, given the statistical insignificance of its density controls in the WRP2 Water Resource model, and our discussions above and below which suggest that weighted average density measures based on Local Area District measures provide too coarse a measure of density to accurately capture the population settlement patterns influencing water recycling system design, it is highly likely that different measures of density would yield different cost assessment results. Moreover, we are able to demonstrate this to be the case through the simple inclusion of the "lnboosterperlength" variable used by Ofwat in its other models, in its Water Resource Models. The inclusion of this variable, which is highly negatively correlated with Ofwat's weighted density variable (because low density requires higher pumping costs) causes the density variable in both of Ofwat's Water Resource Models to become jointly insignificant, thereby demonstrating that Ofwat's Water Resource Models are misspecified.

None of Ofwat's models should be considered until alternative specifications capturing density effects are employed and compared to those of its models

- **It is likely (and our below extensions of Ofwat's models strongly suggest) that Ofwat's models are under-specified and therefore do provide biased estimates because they do not control for the multiple output characteristics of water supply.** Water distribution and Integrated Wholesale Water are characterized by multiple output production These

[9] A full literature review is beyond the scope of this review of Ofwat's modelling. However, we briefly note a few examples from this literature. Abbot & Cohen (2009) is only one of several literature review on academic and regulatory cost modelling in water which has itself received 262 google scholar citations since its publication, illustrating the scope and continuing relevance of this literature. Within this literature, Torres & Morrison-Paul(2006) and Garcia and Thomas (2001) are just two examples of many studies that have modelled overall scale economies and the impact of density on firm costs in an alternative, and we believe conceptually more appropriate, manner to that employed by Ofwat, Moreover, these two examples have respectively received 129 and 268 google scholar citations since their publication, showing their wider relevance.

outputs include the provision of a network transportation output and outputs which are provided with that network. The outputs provided by the network can be variously modelled with inclusion of output proxies such as population served, volumetric water delivered proxies, and/or proxies for connections served. Moreover the same well-established academic literature that provides viable alternatives to the density approach employed by Ofwat suggests that cost interactions between network transportation outputs, volumetric outputs, and connection outputs substantially influence costs.

- **A striking characteristic of all of Ofwat's Water Models, is that they do not include the key output delivered by water companies. i.e. water!** While we are sympathetic to Ofwat's justification that using the number of households instead of water volumes allows it to use an output proxy that will allow better controls and incentives for leakage reduction and water efficiency schemes (Ofwat, 2019x, p. 12), we are not convinced by its modelling on several levels. Firstly, Ofwat does not actually use total households in its modelling but uses total properties including non-households. This variable is a good proxy for the number of connections served by a company, but may or may not proxy for water volumes depending on variation in industrial water demand. Moreover, many other viable alternative scale variables could be employed, while also allowing for incentives for efficient water usage. Thus, population served, water delivered (distribution input less distribution losses) and Effective Water (Distribution Input less Total Leakages) are all equally valid measures that would meet Ofwat's stated criterion of using an output proxy that incentivizes for leakage reduction and water efficiency schemes. Moreover, within the scope of the time allowed for this review, we have in fact demonstrated via models replacing Ofwat's property-based scale variable with Effective Water, that such alternatives provide viable alternatives models to Ofwat's properties-based approach. ***Ofwat's models should therefore not be considered robust, and/or should be triangulated with alternative specifications that employ alternative and potentially more appropriate proxies for the key water output delivered by companies***

- **Moreover, we further emphasize that Environment Agency concerns that within 25 years England will not have enough water to meet demand, are particularly inconsistent with Ofwat's failure to control for past differences in water scarcity, leakage levels, metering, and other factors that have already resulted in additional costs for water stressed companies in its models. E.g. this omission suggest Ofwat's models do not adequately address past nor future implications of steadily tightening water demand balances on water supply costs in England and Wales.**

  This concern is particularly well illustrated by Ofwat's models of water resource plus costs, which have no controls for the water scarcity faced by firms, nor their past efforts to deal with this issue via water supply and demand management activities.

  While beyond the scope of this review to carefully develop, Ofwat should more carefully consider the incentive implications of its models with regard to leakage controls and water demand reducing activities such as metering. These are both costly activities that will result in lower water volumes. Models which use simple scale proxies as their primary drivers may therefore provide perverse incentives that suggest companies that have higher outputs should have higher costs, when the relationship is much more complex when water scarcity, leakage and water demand management are considered properly. Moreover, as argued by Garcia and Thomas (2001) there are tradeoffs between the increased cost of leakage reduction and water demand management which increase network costs, and the savings to be gained in water resource plus. Furthermore, Brea-Solis et al (2017) modelled English and Welsh water data for the period 1996-2010 while controlling for water losses, providing evidence that not only demonstrates the feasibility of controlling for such factors, but s;dp evidence which is nformative with regard to how regulatory policy might be improved to reduce leakage in the face of increasing water scarcity.

- **Even when it does provide alternative models for Water Resources Plus and Integrated Wholesale Water, Ofwat does not carry out appropriate triangulation at the model level**. This comment is related to the previous comment, but focusses more on issues related to what the key explanatory factors in Ofwat's models are, and how this should influence its triangulation approach. Thus, even though 87.8 to 94.8 percent of the variance of costs is explained by the single scale variable included in each of its models, the triangulation carried out by Ofwat literally amounts to testing two alternative controls for water treatment complexity in its Water Resource Plus and Wholesale Water Models. Thus, Ofwat's "triangulation" amounts to modifying control variables which provide little of the overall explanatory power in its models. In contrast, we would argue that triangulation should focus first on those variables that provide most of the explanatory power in the model. Thus, triangulating models with alternative feasible and appropriate scale variables would be necessary. Similarly, as literally all of Ofwat's models rely on the same variable and specification of density, and density contributes far more to the explanatory power of the models than treatment complexity controls, Ofwat should also provide triangulation of models which capture density through different modeling approaches. ***Ofwat's model should therefore not be considered robust until appropriate triangulation of its models is carried out.***

- **Ofwat offers a single distribution model, and therefore does not triangulate its estimation of distribution costs.** This is even though the underlying model estimated by Ofwat treats the number of pumping stations as an output which is effectively estimated to be responsible for 46.5 percent of all water distribution costs. In contrast, the number of properties serviced, or the volume of water delivered on the network are effectively assumed to have zero cost elasticity. ***Ofwat, should, therefore, at a minimum triangulate its water distribution models with alternative specifications.***

Given the conclusions of this general review, we now turn to a brief review of Ofwat's models, which, in the interest of brevity, will refer back to the arguments we have already made here. We emphasize that our purpose is not to provide exhaustively and fully specified alternatives to Ofwat's models, as that is beyond the scope of this review. Instead, the purpose is to demonstrate the characteristics of Ofwat's models, and how simple changes to those models yield either alternative specifications of equal validity to its models, or demonstrates issues such as under or mis-specification and/or questionable parameter estimates within Ofwat's models.

We therefore emphasize that while we simply accept Ofwat's modelling approach, do not test alternative control or density variables, and simply replace and/or add additional scale variables to its models, with more time we would be likely to develop alternative models to properly address the concerns raised above.

Water Resource Plus

Given our above arguments, the below table provides five additional models to the two that Ofwat specified for Water Resources Plus. The first three columns provide scale variable only specifications respectively employing properties, effective water volumes (as defined above) and length of mainsas potential alternative scale proxies.

We focus first on the scale only specification with length of mains, which is provided for two reasons: Firstly it provides consistency with the tables and discussions that are provided below for Water Distribution and Integrated Wholesale Water, where we will test multiple output extensions of Ofwat's models. However, for our immediate purposes, this specification reveals a potential pitfall in Ofwat's modelling approach, and the related potential issue with regard to employing properties as a proxy for water volumes. Thus, while we do not believe that length of mains is an appropriate output proxy for water resources, using it in a scale only model yields a higher R-Squared than specifications with properties and effective water. Thus, spurious correlation rather than appropriate conceptually-based output definition is a danger with Ofwat's modeling approach, which may or may not influence Ofwat 's choice of the properties variable as a proxy for water volumes in its Water Resource Models. This brings into question the extent of engineering input and appropriate conceptual understanding in Ofwat's model development

In contrast, focusing on the scale only specification with Effective Water Volumes readily demonstrates the feasibility of employing a conceptually appropriate water volume-based alternative to properties. Moreover, this variable is specified in a manner which reduces output by removing leakage from distribution input and thereby output in this model will only increase if more water is effectively delivered to customers. In other words, it is consistent with Ofwat's justification for the use of properties as the only scale variable in its models. Moreover, as demonstrated in the specifications WRP1-Alt and WRP2-Alt, simply replacing the properties variable with Effective Water in Ofwat's models yields output specifications that are broadly equivalent to its models, and also consistent with its practice of including variables at lower than standard levels of statistical significance which we assume is attributable to its use of clustered standard errors. Thus, we can readily demonstrate two alternative specifications which provide appropriate models with which to triangulate Ofwat's models via replacement of the scale variable that provides most of the explanatory power in Ofwat's models. **Moreover, we believe that at the conceptual level these water volumes-based models are superior to Ofwat's WRP1 and WRP2 models.**

With regard to Ofwat's own specifications, we must note that in WRP2, both of the coefficients on Ofwat's chosen density measure fall below the 20 percent significance level, which is well below any conventional significance threshold , even if we acknowledging that it uses coefficients above the standard 10 percent level given the clustered standard error approach it relies on. This is also the case with our WRP2-Alt specification. We emphasize that this provides strong evidence that Ofwat's exclusive reliance on LAD weighted density is inappropriate, That is to say, if the simple replacement of one control for water treatment complexity with another drives both the parameter coefficients to statistical insignificance, we should not be confident that this density measure adequately controls the impact of density on costs. We therefore question if Ofwat's models accurately capture the determinants of water resources plus, particularly as they do not allow for many of the standard controls such as difference in the source of water or pumping costs.

Water Resource Plus Models Models

	Scale only Models			**Ofwat Models**		**Alt Output to Ofwat**	
	properties	Effective Water Volumes	length of mains	WRP1	WRP2	WRP1-Alt	WRP2-Alt
lnproperties	0.992***			1.014***	1.014***		
	{0.000}			{0.000}	{0.000}		
lneffwater		1.008***				1.048***	1.050***
		{0.000}				{0.000}	{0.000}
lnlengthsofmain			1.012***				
			{0.000}				
pctwatertreated36				0.008***		0.007***	
				{0.001}		{0.001}	
lnwedensitywater				-1.360**	-0.701	-0.942**	-0.345
				{0.013}	{0.200}	{0.044}	{0.483}
lnwedensitywater2				0.083**	0.036	0.05	0.007
				{0.034}	{0.372}	{0.144}	{0.840}
lnwac					0.443***		0.363**
					{0.002}		{0.020}
lnboosterperlength							
boosterperlength							
_cons	-9.790***	-2.232***	-5.744***	-5.316***	-7.605***	1.084	-0.943
	{0.000}	{0.000}	{0.000}	{0.008}	{0.000}	{0.503}	{0.588}
Observations	124	124	124	124	124	124	124
Parameters	2	2	2	5	5	5	5
Degrees of Freedom	122	122	122	119	119	119	119
OLS R-Squared	0.877	0.868	0.902	0.936	0.929	0.929	0.921
RE R_squared	0.878	0.869	0.903	0.934	0.921	0.928	0.916
VIF_statistic	1	1	1	228.784	211.876	227.732	210.643
OLS RESET_P_value	0.026	0.006	0.041	0.017	0.01	0.008	0.027
RE RESET_P_value	0.694	0.992	0.398	0.35	0.117	0.508	0.347

(reported coefs are random effects as per Ofwat's modelling choices)

Given this concern we pursued a simple test and have been able to demonstrate that neither of Ofwat's water resource models is properly specified, via the simple inclusion of the logged booster stations per length variable it employs in its other models. The logic of this inclusion was twofold as 1) we noted that this variable was strongly negatively correlated with Ofwat's density measures but excluded from its water resources models, and 2) its inclusion may be conceptually appropriate, because low density areas require higher pumping to provide water resources.[10]

Thus, as demonstrated in the table below, including the booster pumping variable into both of Ofwat's specifications causes its density variables to become jointly insignificant, while in the subsequent models including only the booster pumping variable and no density controls all variables are statistically significant.

We believe that implications of this last specification suggest that Ofwat's Water Resource Plus models require substantial revision.

[10]We note that we have simply employed the pumping variable Ofwat used elsewhere in its modelling, so as to readily demonstrate this point. Ofwat might argue that limiting pumping to raw water distribution, abstraction, and treatment activities is appropriate for Water Resources Plus modelling is appropriate. However, a modelling framework that takes potential cost interactions seriously might also test how pumping costs in their treated water distribution networks influence overall costs when choosing which abstraction sources to employ. E.g. we believe it is highly likely that cost interactions between water resources plus and distribution occur because companies choose to minimize overall system costs, including the transportation of treated water when choosing what abstraction sources to employ.

Water Resource Plus Models Denonstrating Density Variables are Insignficant

	Ofwat Models		Adding Pumping		Add Pumping Remove Density	
	WRP1	WRP2	WRP1-Pump	WRP2-Pump	WRP1-No Density	WRP2-No Density
lnproperties	1.014***	1.014***	0.958***	0.945***	0.945***	0.943***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
lneffwater						
lnlengthsofmain						
pctwatertreated36	0.008***		0.006**		0.005***	
	{0.001}		{0.015}		{0.007}	
lnwedensitywater	-1.360**	-0.701	-0.875	-0.298		
	{0.013}	{0.200}	{0.155}	{0.620}		
lnwedensitywater2	0.083**	0.036	0.06	0.021		
	{0.034}	{0.372}	{0.155}	{0.624}		
lnwac		0.443***		0.316**		0.338**
		{0.002}		{0.047}		{0.029}
lnboosterperlength			0.535*	0.659**	0.620***	0.658***
			{0.055}	{0.018}	{0.001}	{0.000}
boosterperlength						
_cons	-5.316***	-7.605***	-4.344**	-5.737***	-6.885***	-6.773***
	{0.008}	{0.000}	{0.019}	{0.007}	{0.000}	{0.000}
Observations	124	124	124	124	124	124
Parameters	5	5	6	6	4	4
Degrees of Freedom	119	119	118	118	120	120
OLS R-Squared	0.936	0.929	0.939	0.935	0.935	0.935
RE R_squared	0.934	0.921	0.939	0.931	0.935	0.931
VIF_statistic	228.784	211.876	250.614	220.412	1.104	1.136
OLS RESET_P_value	0.017	0.01	0.008	0.027	0.008	0.027
RE RESET_P_value	0.35	0.117	0.649	0.537	0.65	0.53
P-value Joint sig of Density Variables	0.000	0.005	0.3624	0.8845		

(reported coefs are random effects as per Ofwat's modelling choices)

Water Distribution

Ofwat offers a single model which includes only lengths of main, weighted density controls (which are added in a nonintractive manner implying that density is assumed to not influence the elasticity of cost and hence economies of scale but only shifts the cost relationship up or down), already discussed and ln(booster pumping stations/length of mains) as a "control" variable.

Ofwat's model for water distribution can therefore be presented as the following model

$$\text{Ln}(Botex) = \alpha + +\delta \ln(length) + \ \beta \ln\left(\frac{booster\ stations}{length}\right) + \gamma \ln(weighed\ pop\ density)$$
$$+\theta \ln(weighed\ pop\ density)$$ (Ofwat TWD1)

However, this can be demonstrated to be mathematically and empirically equivalent to the following model, which effectively treats booster pumping stations as an output in a log linear output specification/

$$\ln(Botex) = \alpha + (\delta - \ \beta) \ln(length) + \ \ \beta \ln(booster\ stations) +$$
$$\gamma \ln(weighed\ pop\ density) \ \ +\theta \ln(weighed\ pop\ density)$$ (Ofwat TWD1')

Thus, as demonstrated in our below table of alternative Water Distribution Models, the actual underlying model, that is nonetheless fully empirically equivalent to the model estimated by Ofwat for Water distribution is as follows:

$$\ln(Botex) = 5.777 + 0.549 \ln(length) + \ \ 0.465 \ \ln(booster\ stations) \quad -$$
$$3.068 \ln(weighed\ pop\ density) \ \ +0.245 \ln(weighed\ pop\ density)$$ (Ofwat TWD1')

Thus, as specified this model is not really consistent with including a pumping control, but actually specifies a model in which lengths of mains as an output cost elasticity of 0.548 and pumping booster stations have an output cost elasticity of 0.465. This is frankly untenable, as it implies that the cost response for pumping is nearly as great as that for network length. Stated differently this model is really an inappropriate two-output-model that attributes approximately 46.5 percent of a company's cost response to increases in its "scale" of distribution activities to a "booster pumping station output" and approximately 54.8 percent to a network length output.

Moreover, careful consideration reveals that the model actually implies that increasing the numbers of properties served and/or the water delivered through the network has no direct impact on costs, so it is difficult to understand how this model could be used to project costs forward into the PR19 control period. **We therefore believe that this model specification appears to rest more on correlation than well conceptualised engineering principles yielding appropriate cost elasticity estimates.**

We therefore test Ofwat's model in two ways. Firstly, as there is no *a priori* reason to include the booster station control in logs, we simply respecify the model with this variable expressed in levels, thereby breaking the mathematical link between the estimated length and booster station variables in TWD1. The resulting TWD-Booster model yields an underlying cost elasticity with respect to network length, that is 1.027, which is more appropriate as a scale elasticity, but of course yield different estimates that could be used to triangulate Ofwat's model, which is an improvement on

Ofwat's model. However, the model still effectively assumes a zero cost elasticity for increases in properties and/or water demand, again making it difficult to see how this model can be used to project costs forward into the PR19 control period.

We therefore proceed by attempting to estimate what we believe are conceptually appropriate two-output-models as simple extensions of TWD-Booster. Thus, as discussed above both the academic literature as well as water company managers and engineers understand that Water distribution and Integrated Wholesale Water are characterized by multiple output production, that is to say,the provision of a network transportation output, and outputs which are provided/served within that network. Thus TWD1-Properties adds properties to the model thereby providing a model in which network length serves as proxy for water transportation as an output, and properties serve as a proxy that effectively combines the connection and volumetric outputs of serving households and non-households. In contrast, TWD1- Effective Water similarly treats network length as a proxy for transportation output services, but uses an output proxy which assumes an additional output related to the amount of effective water supply provided by the company.

In considering these two models we must first emphasize that while these models are informative, with regard to the reliability of Ofwat's models their specifications, they result from a simple test of Ofwat's modelling and are not fully developed multiple output models. Such models should account for the complex interaction between volumetric, transportation, and connection outputs which are further influenced by complex settlement pattern differences that vary both between and within firms.

Nevertheless, both TWD1-Properties and TWD1-Effective Water both suggest that Ofwat's Distribution model is underspecified and suffers from omitted variable bias, as the properties and effective water output proxies are both statistically significant additions to these models. Moreover, the inclusion results in a dramatic fall in the statistical significance of the length of mains variables, suggesting that Ofwat's single output model does not capture fully the determinants of distribution costs.

Emphasising again that these extended two-output-models, are therefore not fully developed specifications, we note that the statistical insignificance of the mains length variable in TWD1-Effective Water, means that this model should not be further interpreted without further development. However, we emphasise that TWD1-Properties yields a statistically significant cost elasticity estimate of 0.575 for properties and a 0.46 cost elasticity estimate for network length, which has a statistical significance level greater than many of the coefficients retained in Ofwat's models. Moreover, we ote that these cost elasticity estimates are broadly consistent with engineering and managerial expectations and the academic literature. Thus, a fully developed model such as this is more likely to give appropriate projections forward into PR19 vial allowing a nonzero, and appropriate estimate of the cost response to property expansion. Moreover, as we are about to see, similar two output extensions of Ofwat's integrated wholesale wastewater models, provide pilot models in which all variables are statistically significant.

Treated Water Distribution Models

	Scale only Models			Ofwat Model		Ofwat-change booster spec	Two Output Models	
	properties	Effective Water Volumes	length of mains	TWD1	TWD1'	TWD-Booster	TWD1-Properties	TWD1-Effective Water
lnproperties	1.063***						0.575*	
	{0.000}						{0.097}	
lneffwater		1.099***						0.861***
		{0.000}						{0.000}
lnlengthsofmain			1.046***	1.013***	0.549***	1.027***	0.46	0.207
			{0.000}	{0.000}	{0.001}	{0.000}	{0.184}	{0.374}
pctwatertreated36								
lnwedensitywater				-3.068***	-3.068***	-3.022***	-2.463***	-1.923***
				{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
lnwedensitywater2				0.245***	0.245***	0.241***	0.192***	0.147***
				{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
lnwac								
lnboosterperlength				0.465***				
				{0.002}				
lnbooster					0.465***			
					{0.002}			
boosterperlength						29.605***	33.810***	37.460***
						{0.008}	{0.001}	{0.000}
_cons	-10.506***	-2.522***	-5.808***	5.777***	5.777***	3.127**	-0.919	2.613*
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.032}	{0.715}	{0.064}
Observations	124	124	124	124	124	124	124	124
Parameters	2	2	2	5	5	5	6	6
Degrees of Freedom	122	122	122	119	119	119	118	118
OLS R-Squared	0.945	0.948	0.897	0.967	0.967	0.966	0.97	0.972
RE R_squared	0.945	0.948	0.898	0.968	0.968	0.967	0.97	0.973
VIF_statistic	1	1	1	222.233	222.233	227.415	333.864	358.523
OLS RESET_P_value	0	0	0.066	0.064	0.064	0.039	0.012	0.003
RE RESET_P_value	0.068	0.324	0.994	0.526	0.526	0.555	0.565	0.42

(reported coefs are random effects as per Ofwat's modelling choices)

Integrated Wholesale Water

Given the already demonstrated weaknesses of Ofwat's water resource plus models we find it remarkable that Ofwat chooses to assess whole integrated system costs by simply adding a single pumping capacity variable to its water resources plus model. Thus, moving from assessing circa 25 percent of the value chain to 100 percent, moving from a "simple model" of acquiring water and treating it, to one where that water is distributed over networks, to properties, where leakage occurs, meters are installed, etc is accomplished by the addition of a single variable to its models and changing the dependent variable in its models. Frankly, we do not understand how such models can be viewed as conceptually appropriate, as if all those activities and all those costs have been added to the model, should it not be the case that model specification needs to be more significantly expanded to capture the resulting increase in complexity that must be modelled?

Given these arguments, we pursue a simple approach to demonstrate that Ofwat's models are under specified and therefore likely to be biased due to omitted variables bias.

To begin with, we develop these arguments using the below table with an approach that the reader should now be familiar with. Thus, three single-output-scale-only-models are presented using properties, effective water volumes, and length of mains as outputs. These models demonstrate that all are plausible proxies to capture scale effects. In other words, they capture the fact that larger companies have larger costs and larger scale.

We next present Ofwat's two specifications (WW1 and WW2) which only differ in their specification of water treatment complexity. As with its water resource models we therefore question if Ofwat provides appropriate triangulation of its WW models if it does not test alternative specifications of its output scale and density variables.

We note to the reader that as Ofwat's WW models do not include the log of length of mains, the mathematical link that implied the effective estimation of a two-output-model with length and boosters in the distribution models is not present here, so the WW model coefficients can be directly interpreted. However, Ofwat's WW models also predict a very high cost elasticity with respect to booster capacity per length of main, which may be deemed excessive by industry experts.

As there is no *a priori* reason to assume the pumping capacity control needs to be logged, the WW1-booster and WW2-booster specifications allow a cross check on Ofwat's models, with the booster station control entered in levels. However, as we will next demonstrate that Ofwat's Integrated Wholesale Water models are underspecified via the use of models with two scale variables that include network length, our real purpose in providing this model is to illustrate a base case, that will allow us to add network length to Ofwat's models without encountering the difficulties in interpretation illustrated above for Ofwat's TWD1 model.

Before proceeding, we emphasise that in contrast to Ofwat's approach, the two output scale variables we are about to discuss imply a significant difference in modelling at the integrated water level, relative to Ofwat's approach. That is to say, this approach means that as we move from explaining part of the value chain to all of the value chain, we make a much larger shift in modelling to reflect the increased complexity that needs to be modelled.

This is most easily seen by considering the WW1-properites and WW2-properties specifications which add network length in addition to the properties output already included in Ofwat's WW models. When compared to Ofwat's approach of simply adding a single control variable to its water resource models, this is not a minor change, because it involves adding an additional explanatory variable capable of explaining a substantial part of costs on its own.

Thus, conceptually we can see these models as an integrated approach that models water resources with the properties output (as Ofwat does in its water resources model) and distribution with the network length variable (as Ofwat does in its distribution model). This contrasts with Ofwat's approach, which is conceptually inconsistent: while Ofwat assumes that different outputs are needed at the disaggregated level, it makes the contradictory assumption that both outputs are not required when modelling these activities together. This is an unfortunate consequence of Ofwat's apparent blind adherence to a modelling framework that insists that all activities must be modelled with a single output variable regardless of the complexity or level of aggregation it is modelling. Moreover, this is in complete contradiction to the academic literature on modelling water industry costs (citation needed)

Direct consideration of the WW1-properites and WW2-properties specifications reveals that the length of mains specification is statistically significant in both specifications. **This alone strongly suggests that Ofwat's WW models suffer from omitted variables bias.** Moreover, these models provide cost elasticity estimates for network length of 0.558 to 0.576 and 0.422 to 0.429 for properties, which are reasonable, and reflect the distinction between providing network capacity and providing and maintaining connection services and water to households and non-households.

Consideration of the WW1 Effective Water and WW2 -Effective Water specifications yields alternative but similarly conceptually appropriate models with Effective Water replacing properties as the output proxy for providing water services to households. Thus, our simple extension of Ofwat's modelling framework, has not only demonstrated that Ofwat's models are underspecified and therefore subject to omitted variables bias, it has also demonstrated that it is feasible to provide meaningful triangulation at the aggregated WW level, via alternative specifications of the assumed scale variables that are the primary explanatory variables in Ofwat's models.

We finally note, that as these two output models include outputs for both network length and properties/or effective water, they are by far more suited for projecting costs into the PR19 cost control period than Ofwat's model.

We thus conclude our review of Ofwat's Integrated Wholesale Water models, and emphasise that our discussion has not only demonstrated that Ofwat's models are underspecified. We have also demonstrated that it is not only plausible but feasible to provide stronger triangulation of models at aggregate level, and also improve the conceptual link between disaggregated and integrated modelling, and thereby provide stronger models for regulatory cost assessment than Ofwat has provided.

Integrated Wholesale Water Models

	Scale only Models			Ofwat Models		Ofwat-change booster spec		Two Output Models			
	properties	Effective Water Volumes	length of mains	WW1	WW2	WW1-booster	WW2-booster	WW1-properties	WW2-properties	WW1 - Effective Water	WW2 - Effective Water
lnproperties	1.027***			0.993***	0.984***	1.003***	0.991***	0.422	0.429*		
	{0.000}			{0.000}	{0.000}	{0.000}	{0.000}	{0.155}	{0.054}		
lneffwater		1.055***								0.497***	0.407**
		{0.000}								{0.008}	{0.015}
lnlengthsofmain			1.024***					0.576**	0.558**	0.518***	0.593***
			{0.000}					{0.049}	{0.010}	{0.003}	{0.000}
pctwatertreated36				0.003***		0.004***		0.004***		0.003***	
				{0.002}		{0.000}		{0.000}		{0.000}	
lnwedensitywater				-1.711***	-1.473***	-1.684***	-1.384***	-2.194***	-1.870***	-1.888***	-1.719***
				{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
lnwedensitywater2				0.126***	0.109***	0.124***	0.102***	0.169***	0.146***	0.146***	0.135***
				{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
lnwac					0.371***		0.457***		0.474***		0.378***
					{0.000}		{0.000}		{0.000}		{0.000}
lnboosterperlength				0.515***	0.517***						
				{0.000}	{0.000}						
boosterperlength						35.605***	36.526***	31.803***	32.349***	34.480***	34.430***
						{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
_cons	-9.420***	-1.665***	-5.010***	-1.273	-2.267**	-4.255***	-5.526***	-0.403	-1.83	2.011**	0.869
	{0.000}	{0.000}	{0.000}	{0.303}	{0.022}	{0.004}	{0.000}	{0.841}	{0.217}	{0.026}	{0.200}
Observations	124	124	124	124	124	124	124	124	124	124	124
Parameters	2	2	2	6	6	6	6	7	7	7	7
Degrees of Freedom	122	122	122	118	118	118	118	117	117	117	117
OLS R-Squared	0.948	0.946	0.927	0.977	0.979	0.978	0.98	0.98	0.981	0.98	0.981
RE R_squared	0.948	0.946	0.928	0.978	0.979	0.979	0.98	0.98	0.982	0.981	0.982
VIF_statistic	1	1	1	250.614	220.412	251.748	224.756	355.276	334.675	417.034	379.634
OLS RESET_P_value	0.007	0	0.602	0.38	0.387	0.263	0.335	0.332	0.614	0.228	0.657
RE RESET_P_value	0.169	0.747	0.816	0.588	0.618	0.657	0.721	0.782	0.952	0.897	0.855

Assessment of Ofwat's Wholesale Wastewater Models

Overall Assessment

In our opinion, the striking characteristic of Ofwat's wastewater modelling for PR19 is its lack of a clear and well-developed conceptual relationship to the complex engineering, demographic, spatial, and other characteristics that determine the costs of providing wastewater services. Moreover, this has resulted in a failure by Ofwat to improve on the modelling which it presented as part of its 2018 Cost Assessment consultation.

This failure is most explicitly demonstrated by Ofwat's apparent inability to develop models of Integrated Wholesale Wastewater, and its resulting failure to provide an aggregate wastewater model with which to properly triangulate the cost assessment of disaggregated activities. Moreover, this is despite the fact that we were able to develop several robust and parsimonious models of integrated wholesale wastewater suitable for regulatory application, within approximately one day, while also deliberately adhering to Ofwat's modelling framework. We are therefore somewhat puzzled with regard to why Ofwat did not seek to adopt or further develop any of the many alternative models of wholesale wastewater that companies submitted to the 2018 Cost Assessment consultation, particularly given that all of the Wholesale Wastewater models presented by Ofwat to the consultation failed model specification tests.

This has left Ofwat to rely on the following disaggregated models in its Initial Assessment of Plans:

- Sewage Collection models that can be readily challenged based on a correct interpretation of their underlying parameters.

- Bioresources Models that rely only on a measure of density, in addition to a scale variable to model the transportation, treatment and disposal of sludge. This has resulted in one bioresources model that fails basic model specification tests, and another that relies on a single marginally statistically significant explanatory factor in addition to output. Moreover, we emphasize that Ofwat provides no compelling explanation about why other factors were not considered in its bioresources modelling, despite its modelling framework suggesting the need for further controls for topography and complexity.

- Sewage Treatment and Bioresources Plus models that are superficially robust but identically specified. As discussed below, in our opinion these models do not sufficiently control for the complex determinants of the costs of these activities, and we also demonstrate that the Bioresources model appears to be under specified. Moreover, as the Sewage Treatment and Bioresources Plus models are identical, by definition, the Bioresources models do not allow for appropriate controls to capture the implications of transporting, treating, and disposing of sludge and should therefore be suspect on these grounds alone.

Thus, in sum, we believe that Ofwat's failure to provide an Integrated Wholesale Waste water model, coupled with the overall relatively low quality of its disaggregated models, has resulted in

Ofwat employing a set of models that does not yet meet what we believe to be appropriate standards for use in regulatory cost assessment.

Wastewater System Optimization and Its Implications for Cost Modeling

Over the past two years, our engagement with Anglian Water's managers and engineers has deepened our understanding of the complex engineering, demographic, spatial, and other characteristics that determine the costs of providing wastewater services. The resulting models of integrated wholesale wastewater costs formed part of Anglian Water's 2018 cost assessment report (Anglian Water, 2018), were presented at the May 2018 Centre for Productivity and Performance Workshop attended by Ofwat and almost all companies (Saal, Nieswand, and Arocena, 2018) and are currently being developed for academic publication.

We have come to understand that cost interactions between sewage treatment, sludge treatment, and sewage collection are fundamental to wastewater system design and cost minimisation. Moreover, these cost interactions operate at a highly localized level, with cost minimizing managers and engineers choosing to operate many distinct sewage treatment plants and hence distinct wastewater systems in England and Wales.

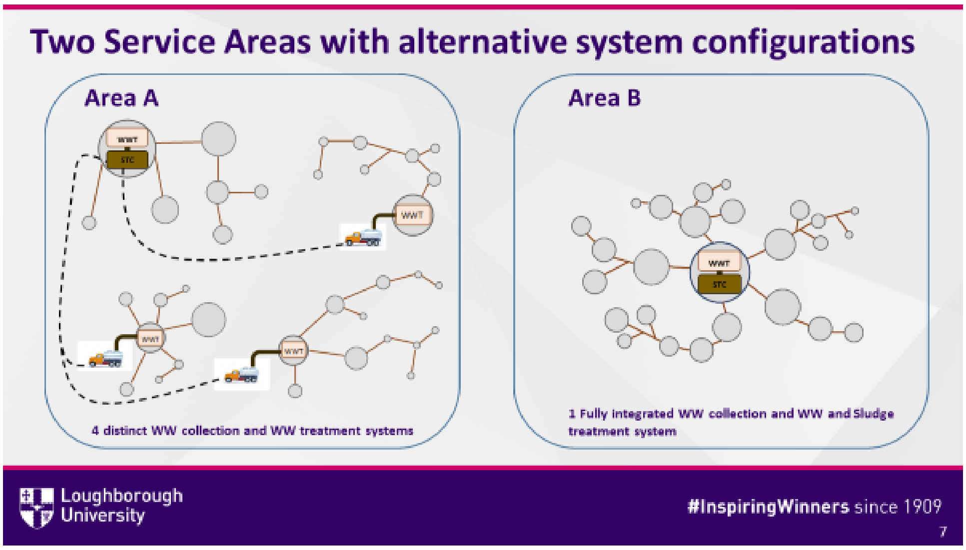
The fundamental factor influencing this localized decision making is the trade-off between the benefits of increased plant level scale economies in the treatment of sewage and the treatment of sludge, and the cost of transporting sewage and sludge to achieve these scale economies. Thus, areas with low population density and/or other operating characteristics that result in high sewage and sludge transportation costs optimize the overall cost of wastewater collection and treatment by operating small distinct waste water systems, thereby trading off reduced transportation costs against higher treatment costs due to small plant sizes. In contrast, areas with high density and/or other operating characteristics that result in relatively low sewage and sludge transportation costs optimize the overall cost of wastewater collection and treatment by operating large waste water systems[11], that also benefit from reduced treatment costs due to large treatment plant sizes.

We therefore emphasise that appropriate modelling of wastewater systems must build from an understanding that managers do not directly optimise wastewater costs at company level, but instead optimise at the local system level. Moreover, this optimisation **requires** them to design local wastewater systems that balance sewage and sludge transportation costs against the benefits of achieving treatment plant scale economies. This of course implies that it is not feasible to disentangle sewage collection, sewage treatment, and sludge treatment costs because local systems are designed in a process that balances the costs and benefits of transportation costs against treatment costs.

[11] We note that our discussion in this document refers to waste water systems, by which we mean the area served by the collection system connected to a sewage treatment work. For clarity, we note that it has been pointed out to us by Anglian Water that the in the industry, a more typical language describing this would focus on the catchment area and sewage collection system that drains to a given sewage treatment work.

Stated differently, the WaSCs of England and Wales do not choose to have over 6300 sewage treatment plants, and hence wastewater systems flippantly. Instead they optimize whole system costs at an appropriate local level, thereby resulting in this waste water system structure.

The implications of this highly localised structure can be demonstrated with the below illustration developed in Saal, Nieswand, and Arocena (2018). If we assume that at aggregate level two areas have the same amount of land area, population, properties, load, etc they can nonetheless operate entirely different wastewater systems. Thus, in Area B population is more concentrated and/or network transportation costs of sewage are sufficiently low that wastewater system optimisation results in a single integrated system in which only one plant treats sewage, and sludge is also indigenously treated at the same plant.



In contrast, in area A population is more dispersed and/or network transportation costs of sewage are sufficiently high that wastewater system optimisation results in the operation of 4 distinct wastewater systems, trading off increased treatment costs due to diseconomies in sewage treatment against the resulting reduced sewage transportation costs. Moreover, given the high cost of sewage transportation, but the relatively higher minimum efficient scale for sludge treatment, managers also choose to incur sludge transportation cost via the use of road tankers to move sludge from three of the small sewage treatment plants to the largest sewage treatment plant operated in the area.

These conceptual examples illustrate that sewage treatment costs, sludge treatment costs, sewage collection costs, and sludge transportation costs are integrally linked, and determined in a process that should seek to minimise whole system costs for wastewater services.

This has several important implications which influence our review of Ofwat's wholesale wastewater modelling:

- Because wastewater system optimisation requires trade-offs between network and sludge transportation costs and treatment plant scale economies, integrated cost assessment models should be best able to capture and control for such interactions.

- Wastewater Costs are not optimized at company level, so even if costs are modelled at company level, models should control for variation in local operating characteristics at a level consistent with the level of disaggregation evidenced by the operation of over 6300 sewage treatment plants in England and Wales
- Models of sewage collection at company level will be difficult to specify, because companies operate thousands of distinct sewage collection systems designed to optimise local whole system costs.

- Because of the trade-off between sewage and sludge transportation costs and achieving plant level economies of scale in sewage and sludge treatment is fundamental in local wastewater system design, factors such as low population density and increased pumping requirements due to local geographic conditions will influence the choice of sewage plant size, the resulting choices with regard to sludge transportation and sludge treatment, and hence sewage treatment costs and overall bioresources plus costs .

- Non-indigenous treatment of sludge is a clear indicator providing information with regard to the trade-offs faced by managers between transportation costs and achieving plant level scale economies. That is, it provides information suggesting that managers have adapted their approach to sewage and sludge treatment, because sludge is being transported from sewage plants of insufficient scale to justify co-treatment of sewage and sludge. It is likely to be an indicator associated with increased costs because non-indigenous treatment of sludge is more likely with small sewage treatment plants, and requires sludge transportation costs to be incurred

Sewage Collection

It is noteworthy that Ofwat's sewage collection models, as conceived within Ofwat's modelling framework (see below for discussion of how alternative mathematical specifications of these models differ from Ofwat's conception/presentation of them), do not in fact control for the actual sewage load transported, and/or the actual properties served by the sewage collection system , but instead rely on the natural log of the network length used to provide these outputs as a "scale" proxy. **Hence, at the most basic level, the conceptual model employed by Ofwat may reward companies who have inappropriately sized networks for the actual service they provide to customers.**[12]

As discussed in our above review of Ofwat's methodology, while both models also control for "topography", Ofwat's reliance on pumping capacity as the sole measure to "capture" how distinct operating environments influence costs, is subject to challenge. Within sewage collection there is no evidence provided by Ofwat that they have considered alternative specifications, which for example might consider how size and fragmentation of sewage systems, prevalence of combined sewers, prevalence of gravity fed sewers, etc influence system costs, and our own modelling with Anglian Water suggests these factors are statistically significant and appropriate from an economic, operational and engineering perspective. Moreover, within the sewage specifications, it is readily demonstrable that Ofwat's control for "topography" actually treats pumping capacity as a non-interactive "output" rather than the control factor that Ofwat claims it be. Furthermore, in at least one of these models (SWC2) the resulting cost elasticity estimate is suspiciously high relative to the estimated cost elasticity for network length. **This inappropriately high relative cost elasticity is suggestive of a modelling approach relying on correlation rather than appropriate engineering, economic, and/or managerial causal arguments as the basis of model selection.**

We finally consider the only variable that differs in Ofwat's two sewage collection models SWC 1 and SWC2, which is their "density control". We discuss this issue further in our below, detailed discussion of the two models, but we do not believe that either provides a robust application suitable for determining appropriate regulated revenues**. Suffice it to say that SWC1 actually provides a model in which a large and inappropriate negative cost elasticity for network length is disguised through Ofwat's specification.** While SWC2's coefficients are superficially superior to SWC1, its inappropriately large cost elasticity for pumping capacity, as well as its reliance on the excessively coarse Local Authority density measure that has already been discussed in our review of the water models, suggest that this model also does not properly capture the determinants of the over 6300 sewerage collection systems operating in England and Wales. This conclusion is strongly supported by the fact that the SWC2 Random Effects R-Squared reported by Ofwat is 0.819, which is a small increase relative to the 0.791 of variance explained for a model which only includes Ofwat's chosen length of mains scale variable. **Thus, sound consideration of both engineering and managerial**

[12] For clarity, we emphasise that this comment applies to the interpretation of the underlying coefficients in Ofwat's models, and not to the appropriateness of controlling for network length in a model of sewage collection costs. However, we would emphasise that care needs to be taken in considering how to model collection, if each company actually operates hundreds of distinct sewage collection systems. e.g., simply assuming that companies operate a single aggregated system is unlikely to be appropriate.

factors influencing costs and of the explanatory power of the SWC2 model strongly suggest that this model is also inadequate.

Ofwat would of course dispute the above criticism and point to its controls for population density and "topography" as effective and understandable controls for differences in the average operating environment faced by firms. However, detailed consideration of the models reveals that such arguments are not sustainable.

Detailed Comment Sewage Collection Model 1 (SWC1)

When the alternative underlying log-linear specification of SWC1 is considered and interpreted, it becomes clear that the model actually includes an inappropriate and strong negative estimated relationship between network length and costs. This model should therefore be disqualified from any use as an appropriate model for regulatory cost assessment.

Ofwat's first sewage collection model is presented as the following model

$$\ln(Botex) = \alpha + +\delta \ln(length) + \beta \ln\left(\frac{pumpingccap}{length}\right) + \gamma \ln\left(\frac{properties}{length}\right)$$ (Ofwat SWC1)

but the simple use of algebra demonstrates that this can be equivalently stated as:

$$\ln(Botex) = \alpha + (\delta - \beta - \gamma) \ln(length) + \beta \ln(pumpingccap) + \gamma \ln(properties)$$ (Ofwat SWC1')

This reveals that the true underlying log-linear specification does not really control for pumping capacity intensity and density as Ofwat suggests, but in fact specifies a model in which mains length, pumping capacity and properties are treated as outputs in a log-linear specification. Thus, this specification is inconsistent with the modelling framework that Ofwat purports to employ. Moreover, the true underlying specification implies that there is not really an appropriate control for the variable impact on density: Length of mains and properties effectively impact costs with fully separable impact.

To assuage the concerned reader, who might not trust that these two econometric specifications will result in identical empirical results, we demonstrate the exact empirical correspondence between (Ofwat SWC1) and (Ofwat SWC1') in a table at the end of this section. However, for our immediate purposes it is sufficient to simply note that the parameters of the underlying model Ofwat has estimated with random effects are as follows:

$$\ln(Botex) =$$

$$-8.907^{***} - 0.901^{***} \ln(length) + .170^{**} \ln(pumpingccap) + 1.471^{***} \ln(properties)$$

That is, Ofwat has based 50 percent of its cost assessment for sewage collection on a model which suggests that all other things being equal (pumping capacity and properties served) increasing the length of your sewerage network will REDUCE costs. In other words, the model does not pass basic tests of having appropriately signed coefficients. Moreover, if one considers the magnitude of the implied cost elasticity for network length, this is not a minor issue. To drive this point further home, we emphasize that the regulatory implication of this model is that all companies should simply INCREASE their network length to reduce costs, an implication with obvious conclusions with regard to the suitability of this model for projecting costs forward into the PR19 review period.

We would note that this unacceptable model is likely to result from poor model design and the flawed conceptual approach to modelling we have discussed above. **However, the sewage collection models' quality may also be compounded by underlying issues with the quality and reliability of the network length variable employed**. Moreover, this issue may also influence the low positive sign on network length, we will discuss below for Ofwat model SC2. Thus, Ofwat employs a measure including all legacy and adopted sewers. However, our discussions with Anglian Water suggest that most companies have little knowledge of the extent of the network they adopted, suggesting measurement errors. In addition, we believe there is a credible argument, supported by our own modelling of integrated wholesale wastewater costs, that the legacy network may better reflect the network design that actually influences the collection process and its relationship with density.

Furthermore, we might also note that if the model is one of sewage collection costs, it is unclear why storm sewers should be included in the total, as they do not deliver sewage to sewage treatment plants, tend to be gravity fedand otherwise differ from sewerage collection networks. Similarly, it is unclear why substantial differences in connected sewage system design, size and costs related to the extent of combined sewers, ,which impact the required level of pumping within the sewage collection network, and gravity sewers are not controlled for in either of Ofwat's models. As these characteristics are likely to significantly influence sewage collection costs, as well as sewage treatment costs (as the volume and intensity of sewage will vary as a result), this point serves as an example of how Ofwat's models are unlikely to have adequately controlled for the complex determinants of sewage collection costs.

We finally note that as the extent of pumping will be significantly influenced by such system design issues, there is further potential that Owat's reliance on pumping as the only factor influencing "topography" may be confusing correlation with causation. **Moreover, as pumping is effectively an input used by companies to help provide sewage collection services, rather than an output, it is interesting that Ofwat effectively specifies it as an output variable rather than as a control variable in both of its sewage collection models.**

Detailed Comment Sewage Collection Model 2 (SWC2)

At face value, Ofwat's second sewage collection is very similar to model 1, and only differs in its definition of the control used for density. However, Model 2 actually does differ significantly because it employs an exogenously determined density measure, and therefore really does impose the fully separable impact of density that Ofwat's modelling framework is designed to impose.

However, as with Model 1, consideration of the true underlying specification employed by Ofwat is still revealing. Thus, Ofwat's second sewage collection model is presented as the following model

$$\ln(Botex) = \alpha + +\delta \ln(length) + \beta \ln\left(\frac{pumpingccap}{length}\right) + \gamma \ln(weighed\ pop\ density)$$

(Ofwat SWC2)

As above, this model can be demonstrated to be mathematically and empirically equivalent to the following model.

$$\ln(Botex) = \alpha + (\delta - \beta)\ln(length) + \beta\ln(pumpingccap) + \gamma\ln(weighed\ pop\ density)$$

(Ofwat SWC2')

We would emphasize that in concert with our above discussion of Ofwat's methodological approach, this specification imposes very strong assumptions on how density influences costs, as it can only shift the fitted relationship between what are Ofwat's two (length and pumping capacity, not only length as they claim) effective "scale" variables and costs up and down but does not otherwise influence the estimated cost elasticities in the model. Stated differently, Ofwat assumes that output elasticities are the same once density is controlled for, which is a very strong assumption that is difficult to reconcile with careful consideration of engineering and managerial constraints and how they influence the cost of adding a customer in rural, suburban, and urban environments.

Given this discussion, the random effects parameters of this true underlying model estimated by Ofwat in SWC2 are as follows

$$5.037 + 0.368\ln(length) + 0.346\ln(pumpingccap) + 0.256\ln(weighed\ pop\ density)$$

(Ofwat SWC2')

Thus, in contrast to Model 1, Model 2 does provide an underlying model in which estimated coefficients have plausible signs. However, when considering those coefficients, we would firstly question Ofwat's assumption that increased density always raises sewage collection costs, as it is generally accepted that low density areas have higher costs, and that increases in density then reduces costs up to some point where further increases in density subsequently increase costs.

Furthermore, the model provides a suspiciously low cost elasticity for network length, while in contrast, it provides an extremely high cost elasticity for pumping capacity, virtually equal to that for network length, which is Ofwat's specified scale variable. Moreover, the 0.739 sum of the two effective scale variables employed by Ofwat have a relatively low impact on costs. Stated differently, we are unconfident that this low cost elasticity estimate is consistent with engineering and managerial knowledge with regard to the cost implications of network expansion.

Furthermore, the model does not include them, it actually assumes that adding properties, volumes collected and/or or population to an existing network has no direct impact on costs! **We therefore emphasize that while this model may or may not provide an appropriate model to judge companies' past cost performance, it is unlikely that such a limited specification will provide accurate estimates of how output expansion will influence costs in the PR19 control period.**

On top of all these concerns we finally re-emphasize our above conclusions with regard to Ofwat's reliance on weighted density measures calculated at with Local Area District level, which is by far too coarse a measure of density. As Wastewater systems operate at scales and within geographic areas that are much smaller than LAD's, a more accurate model capturing the impact of density on system design by engineers and manages requires a much finer definition of density. Alternative approaches such as the measure of sparsity and density developed by Ofwat, but abandoned in their final RP19 modelling without adequate justification, should therefore be explored, as well as more standard academic approaches for modelling network industries. Moreover, these conclusions are supported by the comparison of the R-Squared for SWC2 with that for a Model that only includes Ofwat's scale variable, which suggests that this model does little more than capture the obvious correlation between cost and scale, but does not adequately capture further operating characteristics that influence sewage collection costs.

Table of Reported Model Parameters with Direct Comments

Sewage Collection Models						
		Ofwat's Models			**Equiv. Model Specs**	
	Ofwat Scale Variable Only	**SWC1**	**SWC2**		**SWC1'**	**SWC2'**
lnsewerlength	0.768***	0.739***	0.714***		-0.901***	0.369***
	{0.000}	{0.000}	{0.000}		{0.003}	{0.000}
lnpumpingcapperlength		0.170**	0.346**			
		{0.019}	{0.022}			
lndensity		1.471***				
		{0.000}				
lnwedensitywastewater			0.256**			0.256**
			{0.019}			{0.019}
lnpumpingcap					0.170**	0.346**
					{0.019}	{0.022}
lnproperties					1.471***	
					{0.000}	
_cons	-3.655***	-8.907***	-5.037***		-8.907***	-5.037***
	{0.000}	{0.000}	{0.000}		{0.000}	{0.000}
Observations	70	70	70		70	70
Parameters	2	4	4		4	4
Degrees of Freedom	68	66	66		66	66
OLS R-Squared	0.788	0.906	0.832		0.906	0.832
RE R_squared	0.791	0.907	0.819		0.907	0.819
VIF_statistic	1	1.15	1.566		27.945	2.664
OLS RESET_P_value	0.003	0.249	0.126		0.249	0.126
RE RESET_P_value	0.451	0.35	0.128		0.35	0.128
(reported coefs are random effects as per Ofwat's modelling choices)						

SWC1 - The underlying negative lnsewerlength coef should disqualify this model

SWC2 – Comparison of the R-Squared statistics suggest that marginal additional explanatory power of this model relative to including only Ofwat's scale variable undermines its validity

While Ofwat does not discuss multicollinearity in its January 2019 modelling documents, the documents provided at the cost assessment consultation, suggest strongly that its largely unchanged modelling framework was driven to a large extent by an excessive focus on multicollinearity, which we believe led to its excessively parsimonious model specifications. . Comparison of SWC1 and SWC1', which yield identical econometric models, despite the difference in parametrisation , demonstrate that a narrow focus on the VIF statistics, is not a valid approach. E.g, as Ofwat's model with "low multicollinearity" can be estimated in an empirically equivalent format with "high multicollinearity" this example demonstrates that it is the quality of the econometric model and the interpretation of its underlying coefficients that matters, and not multicollinearity alone

The generally low VIF by standard academic standards as well as the large number of degrees of freedom remaining suggest that Ofwat's approach is excessively parsimonious and could be readily expanded to consider alternative modelling approaches.

Bioresources

Ofwat chooses to model Bioresources (sludge transportation, treatment and disposal) with two models which have exactly two variables in them: a total sludge produced variable and two alternative measures related to density. Moreover, in BR1 the weighted local area density parameter employed is statistically insignificant, while in BR2 the logged ratio of sewage treatment works per property is marginally statistically significant, but the model fails Ofwat's Ramsey Reset test rule**. Stated differently, both BR1 and BR2 can be judged as not meeting our understanding of the minimum requirements for use in regulatory cost assessment on statistical grounds alone.**

BR1 and BR2 are also the product of Ofwat's modelling approach, which emphasizes the need to control for the size of treatment works, "topography" and "complexity", in addition to output scale and density variables. However, in its Bioresources modelling, Ofwat appears to entirely abandon this approach and offers the following three sentence explanation of these modelling choices:

> *"For bioresources models, we have selected the weighted average density and the number of sewage treatment works per properties. Our models show that bioresources costs decrease with density. A bioresources provider can use larger sewage treatment works for larger population centres with lower associated unit costs, due to economies of scale". (Ofwat (2019a), p. 23)*

We now consider this brief justification sentence by sentence.

The first sentence states that Ofwat uses two different density measures in their models.

The second sentence states that their models show that bioresources costs are related to density, and effectively restates the first sentence, while not providing any justification with regard to why other factors may influence bioresources costs, nor why it has abandoned its dedication to modelling with density, topography, and complexity in its bioresources models

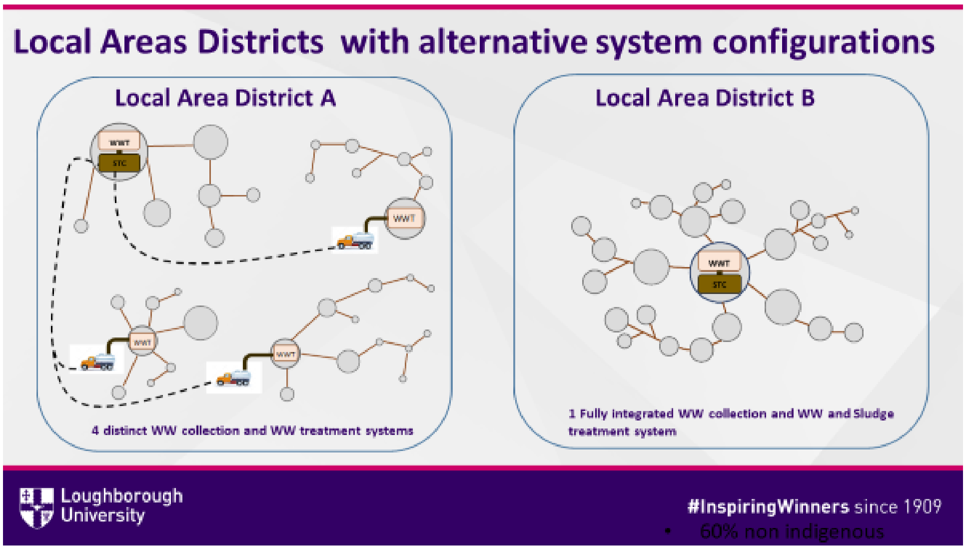
In our opinion, the third sentence reveals a lack of real understanding of how economic, engineering, geographic, and spatial factors influence how and why sludge is transported, treated and disposed and how this determines Bioresources costs, and how an **appropriate** measure of density would influence this.

Thus, the third sentence emphasizes only the scale economies of **sewage** treatment plants and how density of **population centres** allow the benefits of economies of scale. However, sludge is in fact only treated indigenously in sewage treatment plants with sufficient scale to also support sludge treatment. In contrast, in population centres with insufficient size and/or high enough sewage transportation costs to justify larger scale sewage treatment plants, sewage is treated in small scale facilities appropriate for their location, and the sludge is transported from smaller scale sewage treatment plants to larger scale plants for treatment. Moreover, it is even sometimes first transported to a dewatering plant to reduce overall sludge transportation costs, and hence overall bioresources costs. Thus, in contrast to Ofwat's modelling approach, an appropriate conceptual understanding of bioresources costs would highlight and attempt to model and control for what are extremely complex relationships between settlement size, and the trade-offs between sewage treatment plant scale economies, sludge treatment scale economies, and both the network transportation costs and sludge transportation costs, which influence overall cost minimizing wastewater system design (and hence Bioresources costs) in each population centre served by each population centre served by a company.

Stated differently, Ofwat's empirical models, as well as its three-sentence conceptual justification of those models, reveals insufficient consideration of the factors influencing Bioresources costs.

We finally consider Ofwat's density measure, and emphasize that a properly conceived and empirically specified model of bioresources costs (as well as sewage treatment and bioresources plus costs) could include **appropriate** controls for density, and should employ an appropriate specification that can captures how population settlement patterns influences sewage and sludge treatment costs. **However, while Ofwat is correct to emphasize a relationship between the size of population centres and wastewater costs, its chosen LAD weighted average density measure in the BR1 model (and elsewhere where it is used in its modelling) is not an appropriate means with which to control for this effect, as, it uses density data measured at by far too coarse a level to capture the population density of the population centres served by English and Welsh WaSCs.**

Thus, Ofwat relies on a density measure calculated at the level of the UK's 398 Local Areas. However, the LAD level is far too aggregated to capture the actual variation in density influencing wastewater system design in the actual geographically distinct population centres served by the 10 WaSCs. This is revealed by the 10 WaSCs' reliance on over 6300 distinct sewage treatment plants, and, hence, over 6300 distinct wastewater systems, which suggests that variation in density at a level much finer than LAD level influences system design and hence costs.



This argument can be conceptually demonstrated with the above figure adapted from the paper we presented at the May 2018 Centre for Productivity and Performance workshop. (Saal, Nieswand and Arocena, 2018). This figure illustrates that two LADs which are assumed identical in area, population and density, and all other population settlement characteristics when measured at aggregate level, can nonetheless have substantial variation in the number, size, and density of population settlements within their boundaries. Moreover, it illustrates that these differences can result in entirely different wastewater system configurations and substantial variation in the number and scale of sewage treatment plants, and hence sludge treatment and transportation costs, because of differences in density measured at below the LAD level. Furthermore, it illustrates that with such variation, the differences in the average size of sewage treatment plants (as in Ofwat's model BR2) is not sufficient to capture variation in bioresources costs: This is because such a measure not only does not capture the need for sludge transportation costs (inter siting costs) with small plants, but also cannot capture the lost benefits of continuous integrated sewage and sludge treatment, when small population settlement size precludes the indigenous treatment of sludge in the same treatment plant where it was produced.

Thus, our conclusion is that the both the modelling approach, and the density measures employed in both model BR1 and BR2 are likely to be incapable of properly controlling for the spatial characteristics that actually influence bioresources costs.

Brief Detailed Comment Bioresources Model 1 (BR1)

BR1 includes the **statistically insignificant** weighted LAD density measure that Ofwat favours. It is unsurprisingly statistically insignificant, because the required level of density measurement to capture the key differences in sludge treatment is almost certainly below the local authority district (LAD) that Ofwat has used a weighted measure. In other words, Ofwat's weighted density variable quickly fails both statistical significance and conceptual appropriateness tests.

Brief Detailed Comment Bioresources Model 2 (BR2)

Turning to BR2, we note that this model fails Ofwat's own model specification **tests (which interestingly are not reported in the cost assessment report "technical appendix**", but are only recoverable from running the Stata codes provided by Ofwat), as it does not pass Ofwat's own random effects Ramsey Reset test. This model test failure implies this model should not be used by Ofwat to assess performance.

Table of Reported Model Parameters with Direct Comments

Bioresource Models				
			Ofwat's Models	
	Ofwat Scale Variable Only		**BR1**	**BR2**
lnsludgeprod	0.889***		1.058***	1.183***
	{0.000}		{0.000}	{0.000}
lnwedensitywastewater			-0.28	
			{0.121}	
lnswtwperpro				0.320*
				{0.077}
lnproperties				
lnWorks				
_cons	-0.466		0.749	0.746
	{0.266}		{0.325}	{0.182}
Observations	70		70	70
Parameters	2		3	3
Degrees of Freedom	68		67	67
OLS R-Squared	0.762		0.790	0.795
RE R_squared	0.766		0.796	0.801
VIF_statistic	1		2.325	3.697
OLS RESET_P_value	0.005		0.002	0.002
RE RESET_P_value	0.608		0.406	0.039

(reported coefs are random effects as per Ofwat's modelling choices)

BR1 – The logged weighted density variable is statistically insignificant, implying the model is statistically equivalent to one with an output variable only. This is likely to be influenced by the fact that density variable employed by Ofwat is too coarse to be appropriate for determining sludge treatment costs

BR2 – Ofwat's Ramsey Reset test for random effects demonstrates a model misspecification issue, and the coefficient on the alternative measure of density is only marginally significant at the 10 percent level, suggesting the need for further explanatory factors.

Comparison of the R-Squared statistics for both models suggest that their marginal additional explanatory power relative to including only Ofwat's scale variable, should undermine their validity as appropriate models for regulatory cost assessment

The very low VIF by standard academic standards as well as the large number of degrees of freedom remaining, suggest that Ofwat's approach is excessively parsimonious and could be readily expanded to consider alternative modelling approaches.

Sewage Treatment and Bioresources Plus

Direct Assessment of Ofwat's Models

Ofwat's Sewage Treatment and Bioresources Plus modelling is based on an identical specification of cost drivers. Thus, all models include sewage treatment load as a scale variable and a control for ammonia consents to capture the cost implications of higher level treatment. The only difference in models is the inclusion of either a control for the share of load treated in smaller band 1 to 3 plants (SWT1 and BRP1) or the share of load treated in large band 6 plants (SWT2 and BRP2).

As indicated in the colour coded assessment of these models in the below table, these models have coefficients that are always statistically signficant and also pass Ofwat's Random Effects Ramsey Reset test. Moreover, the estimate parameters all have the expected signs. Thus, these models appear to be robust at first glance. However, several issues can be raised with these models, at the basic statistical level.

Firstly, comparison of the R-squared statistics for Ofwat's chosen models relative to auxiliary models including only a scale variable suggests an extremely modest increase in model predictive power. We suggest that this results from the inability of Ofwat's models to capture much of the complexity that actually determines sewage and sludge treatment costs, only some of which we have discussed in our above commentary.

Secondly, consideration of the further auxiliary models we have built which include both the pctband13 and the pctband6 variables reveals an interesting difference between the Sewage Treatment and Bioresources Plus models. In both cases, both the coefficients are individually highly significant, thereby explaining why Ofwat would legitimately not have included such a model in its assessment. However, when we consider the joint significance test for these two variables together, we find that while they are jointly insignificant in the sewage treatment models, they are jointly significant in the Bioresources Plus models.

This difference may seem minor, but it provides evidence for what we believe is a fundamental flaw in these models, which is that it is inappropriate to model bioresources costs with the same model as sewage treatment because the former includes the cost of sludge transportation, treatment, and disposal while the latter does not. Hence, the joint statistical significance of these parameters provides clear evidence of under specification of the Bioresources Plus models. While the individual insignificance precludes using these models, one might speculate that the increased information on the share of load at different plant sizes in the auxiliary model for Bioresources Plus captures some of the impact of sludge treatment that we discussed above in our section on wastewater system optimisation.

We therefore conclude that in general Ofwat's models appear under-specified, and that there is clear evidence provided by a minor extension of its models for Bioresources Plus that its models do not adequately model bioresources costs. We therefore proceed by briefly considering an extension of Ofwat's bioresources models, which we believe unequivocally demonstrates the under-specification of its models, and simply note that we do not pursue this extension for the sewage treatment models due to both time limitations, and the arguments we made above on system optimisation suggesting that sewage and sludge treatment costs are too intricately linked to allow separate estimation

Sewage Treatment and Bioresource Plus Models

	Sewage Treatment Models				Bioresource Plus Models			
		Ofwat's Models				Ofwat's Models		
	Ofwat Scale Variable Only	SWT1	SWT2	Include Both pctband variables	Ofwat Scale Variable Only	BRP1	BRP2	Include Both pctband variables
lnload	0.746***	0.795***	0.779***	0.801***	0.799***	0.788***	0.770***	0.790***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
pctbands13		0.045**		0.025		0.039**		0.017
		{0.038}		{0.577}		{0.024}		{0.614}
pctbands6			-0.012*	-0.006			-0.010**	-0.007
			{0.079}	{0.601}			{0.026}	{0.478}
pctnh3below3mg		0.004***	0.004***	0.004***		0.005***	0.005***	0.005***
		{0.000}	{0.000}	{0.000}		{0.000}	{0.000}	{0.000}
_cons	-4.608***	-5.498***	-4.203***	-5.014***	-4.969***	-5.107***	-3.944***	-4.538***
	{0.000}	{0.000}	{0.000}	{0.003}	{0.000}	{0.001}	{0.000}	{0.009}
Observations	70	70	70	70	70	70	70	70
Parameters	2	4	4	5	2	4	4	5
Degrees of Freedom	68	66	66	65	68	66	66	65
OLS R-Squared	0.840	0.863	0.847	0.866	0.892	0.919	0.915	0.918
RE R_squared	0.842	0.866	0.849	0.859	0.894	0.919	0.918	0.920
VIF_statistic	1	4.775	4.114	8.763	1	4.775	4.114	8.763
OLS RESET_P_value	0	0.002	0	0	0	0	0	0
RE RESET_P_value	0.032	0.391	0.537	0.444	0	0.111	0.223	0.141
P-value Joint sig of pctband13 and pctband6				0.131				0.046

(reported coefs are random effects as per Ofwat's modelling choices)

Demonstration of the Under specification of Ofwat's Bioresources Plus Models

We have chosen to further test Ofwat's bioresources Plus models, because our above empirical demonstration supports our assessment at the conceptual level. As argued in the above section on waste water system optimisation and its implications for cost modelling:

- Because the trade-off between sewage and sludge transportation costs and achieving plant level economies of scale in sewage and sludge treatment is fundamental in local wastewater system design, factors such as low population density and increased pumping requirements due to local geographic conditions will influence the choice of sewage plant size, the resulting choices with regard to sludge transportation and sludge treatment and hence sewage treatment costs and overall bioresources Plus costs .

We therefore proceed to assess Ofwat's bioresources models by testing the appropriateness of including the following variables to its specification:

- The measure of weighted density Ofwat prefers on the grounds that it may capture how population settlement patterns influence the dispersion of population settlements, thereby impacting sewage transportation costs and the scale method and cost of sewage treatment and sludge treatment

- The measure of pumping capacity per length of main Ofwat prefers as a measure of topography on the grounds that it provides evidence of the dispersion in network transportation costs related to topography which also influence network transportation costs and the scale method and cost of sewage treatment and sludge treatment

- The share of sludge treated non-indigenously to both control for sludge transportation costs and as a proxy that is unambiguously related to local system level optimisation by managers with regard to how best to trade off the costs network and sludge transportation costs against the benefits of increased scale in sewage and sludge treatment

Before proceeding, we note that as our approach mimics Ofwat's use of density variables to explain water resource plus costs, it is fully consistent with Ofwat's demonstrated modelling approach. Moreover, we emphasize that the approach employed is fully consistent with Ofwat's approach of modelling with a single scale variable and including potential controls for density, complexity and topography. However, this approach differs from Ofwat's in that it does properly include variables that can capture how cost interactions between network and treatment can and do significantly influence treatment costs, and also includes the indigenous treatment of sludge variable so as to account for sludge costs in the bioresources plus model.

While time limitations preclude a detailed discussion of these results, we present a colour coded table which indicates green for significant parameters and passing Ramsey reset tests, red for parameter estimates that are highly insignificant and failing Ramsey reset test, and yellow for parameter estimates with significance between the 10 and 20 percent level, which re flects Ofwat's practice of allowing moderately insignificant coefficients, as result of the tendency for its use of clustered standard errors to inappropriately exclude variables that may contribute significantly to the overall fit of the model (Glass and Glass, 2018)

In a nutshell, appropriate testing down from the general specification provided in the Reviewer BRP2 model, and rejecting all models with highly statistically insignificant parameters and those that exclude statistically significant parameters results in a conclusion that the model labelled Reviewer BRP5, is an appropriate specification statistically, as well as conceptually.

Moreover as this specification results from the testing down of a general specification in which Ofwat's models are nested, it also demonstrates that Ofwat's bioresources models are underspecified, and therefore likely to provide an inappropriate and biased assessment of bioresources Plus costs.

Extension of Ofwat's Bioresource Models by the Reviewers

Model and Rating	scale only	Include Both pctband variables	Ofwat BRP1	Ofwat BRP2	Reviewer BRP1	Reviewer BRP2	Reviewer BRP3	Reviewer BRP4	Reviewer BRP5	Reviewer BRP6	Reviewer BRP7
lnload	0.799***	0.790***	0.788***	0.770***	0.647***	0.758***	0.750***	0.728***	0.738***	0.699***	0.707***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
pctbands13		0.017	0.039**			0.042**	0.005				
		{0.614}	{0.024}			{0.031}	{0.614}				
pctbands6		-0.007		-0.010**		0.011		0.002			
		{0.478}		{0.026}		{0.150}		{0.713}			
pctnh3below3mg		0.005***	0.005***	0.005***	0.005***	0.006***	0.006***	0.006***	0.006***	0.005***	0.006***
		{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
nonindgloadshr					0.103	0.362**		0.196	0.184*	0.235**	0.052
					{0.441}	{0.020}		{0.155}	{0.098}	{0.031}	{0.686}
lnpumpingcapperlength						0.253***	0.223***	0.264***	0.247***	0.263***	
						{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	
lnwedensitywastewater						-0.150**	-0.133*	-0.129**	-0.121**	nested	-0.151*
						{0.018}	{0.079}	{0.014}	{0.017}	sig coef	{0.092}
_cons	-4.969***	-4.538***	-5.107***	-3.944***	-3.207***	-4.762***	-3.628***	-3.552***	-3.609***	-3.991***	-2.877***
	{0.000}	{0.009}	{0.001}	{0.000}	{0.001}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Observations	70	70	70	70	70	70	70	70	70	70	70
Parameters	2	5	4	4	4	8	6	7	6	5	5
Degrees of Freedom	68	65	66	66	66	62	64	63	64	65	65
OLS R-Squared	0.892	0.918	0.919	0.915	0.893	0.951	0.941	0.942	0.943	0.934	0.906
RE R_squared	0.894	0.92	0.919	0.918	0.894	0.954	0.944	0.946	0.946	0.938	0.91
VIF_statistic	1	8.763	4.775	4.114	5.081	19.814	5.315	7.612	6.79	5.268	6.713
OLS RESET_P_value	0	0	0	0	0	0.019	0.009	0.002	0.019	0.002	0
RE RESET_P_value	0	0.141	0.111	0.223	0.025	0.001	0.281	0.251	0.299	0.11	0.001

(reported coefs are random effects as per Ofwat's modelling choices)

Integrated Wholesale Wastewater

Comment on Ofwat's Failure to Provide a Model for Integrated Wholesale Wastewater

As part of the March 2018 Cost Assessment Consultation, Ofwat provided 8 models of integrated wholesale wastewater, all of which failed its Ramsey Reset specification test. (Ofwat, 2018x), p. 29) However, rather than providing improved models, Ofwat has chosen simply to not provide nor use any model of integrated wholesale wastewater in its IAP assessment. **We believe that this is an extremely significant omission which undermines Ofwat's overall assessment of wastewater costs fundamentally, because Ofwat has failed to provide an appropriate integrated model with which to triangulate its disaggregated cost assessments**.

if we consider that CEPA(2011) advised Ofwat to adopt triangulation between aggregated and disaggregated cost assessments so as to better control for cost interactions between different activities, Ofwat should be very aware of the need for such triangulation. Moreover, it is unclear why Ofwat decided it was necessary and appropriate to do this in its water modelling, but that this was not the case for wastewater. Similarly, this omission for wastewater contradicts there comment that:

> *"To mitigate risks of error and bias we do not rely on a single model. Rather, we use a diverse set of models, with different drivers and different levels of aggregation, in triangulation (Ofwat, 2019x, p. 5)*

Furthermore, as we have personally been involved in developing integrated wholesale water models for this price review, and even presented two such models at a May 2018 Centre for Productivity and Performance workshop that Ofwat attended (Saal, Nieswand, and Arocena 2018) , we actually know that it is feasible to develop integrated wholesale wastewater models with the available regulatory data. We also emphasize that similar models that we developed were submitted by Anglian Water to the Cost Assessment consultation (Ofwat 2018x), and were published in Anglian Water's 2018 cost assessment report, (citation needed). However, we can see no evidence in Ofwat's IAP documents that it even considered the Wholesale wastewater models submitted by Anglian nor those provided by other companies to the consultation. We must admit that we are disappointed by this, as we believe the models we developed for Anglian Water were sufficiently parsimonious and understandable for regulatory application, and also carefully blended an academic approach to cost modelling in network industries (where cost interactions matter), with a deep understanding gained from managers and engineers of the most important factors influencing wastewater costs.

We must acknowledge that the approach we employed differs from that of Ofwat in that we employed interactions between variables. Moreover, given that we are experts in the econometrics of cost modelling, we also transformed our data to both improve the interpretability of our models and also control for factors such as the number of works operated by a firm while not using up excessive degrees of freedom or causing severe multicollinearity.

Stated more simply, while we do not accept the validity of Ofwat's modelling approach because of the severe and unjustified restrictions it places on the modeller, we can only speculate that Ofwat appears to have rejected any model that does not mechanistically adhere to its modelling framework. That is, one that does not model with a single output and use the limited number of control variables from the categories that Ofwat has defined (Scale, Complexity, Topography, and Density).

We therefore proceed in our review of wholesale waste water modelling by accepting Ofwat's modelling framework and assessing its assertions that it was not possible to develop an appropriate integrated wholesale waste water model.

Demonstration of Feasibility to Provide a Wholesale Wastewater Model While Maintaining Ofwat's Approach to Modelling

In doings this, our approach to this is similar to that employed for our extension of Ofwat's Bioresources Plus models. We therefore considered only the variables it considered in its various wastewater models, except for our inclusion of the nonindigenous sludge treatment control discussed extensively above. We have also chosen to employ the weighted average density measure preferred by Ofwat, noting our concerns that this variable is unlikely to capture density at the level of granularity required to capture how density influences wastewater system design.

The results of this exercise are reported in the below table, with the colour coding based on the same criteria discussed above. A general specification including all the controls is given in model 2. While all the coefficients in this model are statistically significant supporting the need for multiple variables to capture the complexity of wastewater system costs, the Ramsey Reset test for this model suggests it may be mis specified.

We therefore proceed by systematically excluding variables and judging the appropriateness of the resulting specifications. Thus, the consistent insignificant parameters for the pctban6 and pctband13 variables, as well as the statistical insignificance of the weighted density variables are consistent with its discussion of these variables. (Ofwat (2019a), p. 23)

However, if one simply includes controls for indigenous treatment of sludge as an appropriate complexity control, and then tests down from the general specification, a model of integrated wholesale wastewater is readily estimated. Thus, model 13 in the below table demonstrates that wholesale waste water costs can be modelled with lnload(scale), lnpumpingcapperlength (topography), pctnh3below3mg (complexity) and nonindgloadshr (complexity)

However, in order to generate such a model, one needs a strong conceptual understanding of the factors that influence wastewater costs, but also needs to understand that many of the factors being modelled as controls will be closely correlated, as demonstrated by the following table of correlations.

	pctnh3below3mg	pctbands6	pctbands13	lnpumpingcapperlength	lnwedensitywastewater	nonindgloadshr
pctnh3below3mg	1					
pctbands6	0.6829	1				
pctbands13	-0.6635	-0.927	1			
lnpumpingcapperlength	-0.1399	-0.4552	0.4437	1		
lnwedensitywastewater	0.7733	0.7301	-0.6754	-0.1924	1	
nonindgloadshr	-0.5487	-0.4463	0.2345	-0.2117	-0.396	1

Consideration of this table is consistent with Ofwat's conclusions with regard to the insignificance of the scale control variables it employed, which are highly correlated with not only the density measure but also the ammonia consents control. Stated differently, if Ofwat modelled with the ammonia control, the plant size controls and the density measure, it is not surprising it got unusable results because of the high correlations between these variables. Moreover, careful consideration by Ofwat should have revealed that all of these variables are in fact strongly related to density, as plant size increases with density, and requirements to treat sewage at higher levels tend to be higher in larger plants.

However, while the density measure is strongly correlated it also does not perform well in these models when entered without the scale size variables, which may result from it being based on Local Area district densities which may be too coarse a level to model with, and/or its close correlation to the ammonia treatment control variable.

Nevertheless, we are confident that model 13 has demonstrated a model that is consistent with not only Ofwat's modelling framework, but also the conceptual model that we would develop based on our understanding of the determinants of waste water system costs.

Moreover, we have also been able to demonstrate that Ofwat could also triangulate such as model by employing alternative definitions of the scale variable it employs, which always provides most of the explanatory power in its specifications and should therefore be tested for proper triangulation. We therefore specified alternative models, using the property scale variable Ofwat used in some of its Wholesale wastewater models provided in the 2018 cost assessment consultation (Ofwat, 2018x), p. 29). We also tested a specification using population served as a control variable. A summary of these results is provided in the final table in this section, with the same specification of the control variables being the preferred specification.

Thus, we believe it is an unambiguous conclusion that Ofwat not only could have, but should have provided triangulated models of integrated wholesale wastewater, that could and should have been used to triangulate its disaggregated models of wholesale wastewater costs.

Moreover, the absence of such models calls Ofwat's entire assessment of Wholesale Waste Water costs into question.

Reviewers' Identification of an Integrated Wholesale Wastewater Model Within Ofwat's Modelling Framework

Model and Rating	1	2	3	4	5	6	7	8	9	10	11	12	13
lnload	0.765***	0.738***	0.698***	0.664***	0.699***	0.671***	0.699***	0.702***	0.683***	0.705***	0.694***	0.693***	0.677***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
pctnh3below3mg		0.006***	0.005***	0.005***	0.005***	0.005***	0.005***	0.005***	0.005***	0.005***	0.005***	0.005***	0.005***
		{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
pctbands6		0.026***	0.008	0.001		0.002		0.004	0.001				
		{0.000}	{0.369}	{0.745}		{0.498}		{0.583}	{0.768}				
pctbands13		0.086***	0.031		0.006		0.002	0.015		0.004			
		{0.000}	{0.317}		{0.684}		{0.893}	{0.570}		{0.764}			
lnpumpingcapperlength		0.278***	0.251***	0.275***	0.247***	0.289***	0.264***	0.239***	0.253***	0.233***	0.260***	0.236***	0.257***
		{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
lnwedensitywastewater		-0.147***				-0.048	-0.031	-0.077	-0.069	-0.059	-0.041	-0.068	
		{0.002}				{0.236}	{0.591}	{0.173}	{0.198}	{0.399}	{0.471}	{0.281}	
nonindgloadshr		0.605***	0.269*	0.215	0.199*	0.204	0.189				0.187		0.203*
		{0.000}	{0.079}	{0.102}	{0.072}	{0.142}	{0.129}				{0.127}		{0.058}
_cons	-4.063***	-5.481***	-4.213***	-3.166***	-3.515***	-3.000***	-3.288***	-3.291***	-2.818***	-3.096***	-3.139***	-2.867***	-3.227***
	{0.000}	{0.000}	{0.001}	{0.000}	{0.000}	{0.000}	{0.004}	{0.005}	{0.000}	{0.004}	{0.000}	{0.000}	{0.000}
Observations	70	70	70	70	70	70	70	70	70	70	70	70	70
Parameters	2	8	7	6	6	7	7	7	6	6	6	5	5
Degrees of Freedom	68	62	63	64	64	63	63	63	64	64	64	65	65
OLS R-Squared	0.912	0.968	0.958	0.943	0.946	0.947	0.946	0.952	0.945	0.946	0.946	0.946	0.944
RE R_squared	0.913	0.971	0.954	0.943	0.945	0.946	0.945	0.949	0.944	0.945	0.946	0.945	0.944
VIF_statistic	1	19.814	15.86	5.283	5.385	7.612	7.025	8.97	5.203	5.315	6.79	5.199	5.268
OLS RESET_P_value	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
RE RESET_P_value	0.000	0.004	0.000	0.347	0.273	0.150	0.000	0.000	0.444	0.633	0.358	0.532	0.178

(reported coefs are random effects as per Ofwat's modelling choices)

Alternative Integrated Wholesale Wastewater Models Developed by the Reviewers

	Properites as Scale Variable				Population Served as Scale Variable				Treatment Load as Scale Variable			
lnproperties	0.801***	0.762***	0.757***	0.724***								
	{0.000}	{0.000}	{0.000}	{0.000}								
lnpopserved					0.748***	0.745***	0.741***	0.713***				
					{0.000}	{0.000}	{0.000}	{0.000}				
lnload									0.738***	0.694***	0.693***	0.677***
									{0.000}	{0.000}	{0.000}	{0.000}
pctnh3below3mg	0.005***	0.005***	0.005***	0.005***	0.006***	0.005***	0.005***	0.005***	0.006***	0.005***	0.005***	0.005***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
pctbands6	0.027***				0.026***				0.026***			
	{0.000}				{0.000}				{0.000}			
pctbands13	0.084***				0.073***				0.086***			
	{0.000}				{0.000}				{0.000}			
lnpumpingcapperlength	0.262***	0.265***	0.236***	0.269***	0.222***	0.218***	0.188**	0.227***	0.278***	0.260***	0.236***	0.257***
	{0.000}	{0.000}	{0.001}	{0.000}	{0.000}	{0.004}	{0.011}	{0.001}	{0.000}	{0.000}	{0.000}	{0.000}
lnwedensitywastewater	-0.205***	-0.082	-0.110*		-0.216***	-0.074	-0.104*		-0.147***	-0.041	-0.068	
	{0.000}	{0.233}	{0.064}		{0.000}	{0.206}	{0.064}		{0.002}	{0.471}	{0.281}	
nonindgloadshr	0.432***	0.213*		0.239**	0.442***	0.207		0.232**	0.605***	0.187		0.203*
	{0.001}	{0.081}		{0.011}	{0.001}	{0.116}		{0.036}	{0.000}	{0.127}		{0.058}
_cons	-7.300***	-5.135***	-4.785***	-5.182***	-6.992***	-5.560***	-5.208***	-5.600***	-5.481***	-3.139***	-2.867***	-3.227***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Observations	70	70	70	70	70	70	70	70	70	70	70	70
Parameters	8	6	5	5	8	6	5	5	8	6	5	5
Degrees of Freedom	62	64	65	65	62	64	65	65	62	64	65	65
OLS R-Squared	0.968	0.942	0.943	0.934	0.964	0.94	0.941	0.931	0.968	0.946	0.946	0.944
RE R_squared	0.97	0.938	0.943	0.931	0.966	0.935	0.94	0.928	0.971	0.946	0.945	0.944
VIF_statistic	19.68	6.774	4.83	5.667	19.714	6.575	4.731	5.498	19.814	6.79	5.199	5.268
OLS RESET_P_value	0.0030	0.0010	0.0010	0.0000	0.0000	0.0000	0.0000	0.0000	0.0010	0.0000	0.0000	0.0000
RE RESET_P_value	0.014	0.443	0.429	0.374	0.001	0.368	0.411	0.204	0.004	0.358	0.532	0.178

(reported coefs are random effects as per Ofwat's modelling choices)

References

Abbott, M., & Cohen, B. (2009). Productivity and efficiency in the water industry. *Utilities Policy*, *17*(3-4), 233-244.

Anglian Water (2017) "Water Industry Cost Modelling: Anglian Water's Approach and Initial Assessment" Available here.

Anglian Water (2018) "Water Industry Cost Modelling: Update Report" Available here

Brea-Solis, H., Perelman, S., & Saal, D. S. (2017). Regulatory incentives to water losses reduction: the case of England and Wales. *Journal of Productivity Analysis*, *47*(3), 259-276.

Frontier Economics (2017) Productivity Improvement in the Water and Sewerage Industry in England Since Privatisation. A Report commissioned and published by Water UK. (https://www.water.org.uk/publication/productivity-improvement-in-the-water-and-sewerage-industry-in-england-since-privatisation/)

Fletcher R. (2018) "Why change is needed" Speech given by Rachel Fletcher, Chief Executive Moody's UK Water and Regulated Networks 10 October 2018. Available here

Garcia, S., & Thomas, A. (2001). The structure of municipal water supply costs: application to a panel of French local communities. *Journal of Productivity analysis*, *16*(1), 5-29.

Greene, W. H. (2012). Econometric analysis 7th edition Edinburgh*: Pearson Education Ltd.*

Ofwat (2019a): Supplementary technical appendix: Econometric approach.

Ofwat (2019b): Technical appendix 2: Securing cost efficiency.

Ofwat (2018): Cost assessment for PR19: a consultation on econometric cost modelling.

CEPA (2011) Cost Assessment- Use of Panel and Sub-Company Data". Report commissioned and published by the Office of Water Services, May 2011

CEPA (2018): PR19 econometric benchmarking models, March 2018.

Saal, D. S., Arocena, P., Maziotis, A., & Triebs, T. (2013). Scale and scope economies and the efficient vertical and horizontal configuration of the water industry: a survey of the literature. *Review of network economics*, *12*(1), 93-129.

Saal, Nieswand, & Arocena (2018) Wastewater Botex Modelling for PR19: Developing a Robust Model for Regulatory Cost Assessment from an Appropriate Conceptual Model of How Cost Interactions Drive Integrated Expenditures. Paper presented at the Loughborough University Centre for Productivity and Performance Workshop on Water Industry Performance and Cost Assessment, 23rd May 2018 (https://www.lboro.ac.uk/media/wwwlboroacuk/external/content/schoolsanddepartments/sbe/sbewebsite/downloads/events/DavidSaal_Maria%20Nieswand_Pablo.pdf

Saal (2018): Comments on CEPA's methodological approach in its PR19 econometric benchmarking models for Ofwat. *(https://www.ofwat.gov.uk/wp-content/uploads/2018/03/Professor-David-Saal-consultation-submission.pdf)*

Stone and Webster (2004) "An Investigation into OPEX Productivity Trends and Causes in the Water Industry in England & Wales – 1992-93 to 2002-03" Report commissioned and published by the Office of Water Services, June 10th 2004.

SUMICSID (2007): Development of the benchmarking models for German electricity and gas distribution.

Torres, M., & Paul, C. J. M. (2006). Driving forces for consolidation or fragmentation of the US water utility industry: a cost function approach with endogenous output. *Journal of Urban Economics*, *59*(1), 104-120.

Wooldridge, J.M. (2006): Introductory econometrics – A modern approach. Thomson South West, 3rd edition.