

4th May 2018

Dear Ofwat

Cost assessment for PR19: a consultation on econometric cost modelling

This letter sets out our response to Ofwat's consultation on PR19 econometric cost modelling. We welcome the approach Ofwat has taken to developing the cost models for PR19 and the effort that has been taken to develop a quality data set. The engagement throughout the process has allowed companies to consider what will be an appropriate set of cost models, and this is reflected in the large pool of models that were included in the consultation. The consultation approach has then allowed us to review this range of modelling, including those that Bristol Water contributed to the process.

We also support the overall approach to the consultation and the criteria used to select models. No model will be perfect, and a set of models that to a greater or lesser degree have engineering, operational or economic rationale, as well as statistical test validity, allow a topic that has a degree of judgement to be explored robustly. This has led our own approach to cost model development at PR19, in particular the development by NERA of their innovative use of Monte-Carlo simulation to screen the explanatory variables considered in their modelling. As we explain further below, we found the CEPA report and the model statistical information that was published particularly helpful in preparing our response to this consultation.

Our approach to responding to the consultation

We do not comment on wastewater models, as a water only company these are not areas of our expertise.

Our response has been informed by the work that we undertook on water wholesale costs with NERA residential retail modelling jointly with Wessex Water through Economic Insight. We published reports of our modelling on our website¹ and as part of the consultation. We do not repeat the technical details of this modelling in this response, although we do highlight how we think this approach to modelling compares to the other models proposed.

As part of our response we reviewed Ofwat's and CEPA's modelling against our work with NERA and Economic Insight. We carried out further testing of our modelling with these independent experts to

¹ <http://www.bristolwater.co.uk/about-us/regulatory-policy-and-consultations/>

consider how they performed against the model tests suggested by CEPA, including the relative importance of the individual tests. Having reviewed Ofwat and CEPA's models, we also then considered the other proposed models against the models that Ofwat, CEPA and ourselves had developed.

We established a set of criteria for reviewing all of the models relevant to us in the consultation. We have added our assessment of the models against these criteria in the consultation response template, as well as providing comments and an overall "RAG" rating. Our view is that given the extensive amount of work that had been put into the model generation across the industry, and that the CEPA analysis of statistical validity in the consultation appeared broadly sound, we specifically focussed in our criteria on how to choose between the set of models in the consultation. We also wanted to consider the specific question on level of aggregation that Ofwat raised in the consultation in light of the modelling results. Our assessment criteria are:

- Statistical validity – we use CEPA's criteria for assessing concerns on the statistical validity of individual models (Red), or where there were questions arising (Amber).
- Better model – is there clearly a better, more statistically significant model that improves explanatory power or model logic, for instance by including additional variables. This is slightly different to the criteria and statements on model approach on page 10 of Ofwat's consultation – we think there needs to be a clear justification for ignoring significant variables, especially if this is just to improve what is perceived as the prediction power for other estimation variables. Although we do not disagree with this principle, it is a significant risk when applying it to efficiency estimation in practice if there are valid models with significant variables that are screened out on this basis. For individual companies, this could make a significant difference to how the models are used for cost estimation, even if Ofwat felt the logic was better for the industry as a whole. We prefer NERA's approach to model selection which avoids these judgements for this reason as there less risk of this type of judgement leading to questions on the validity of the efficiency interpretation.
- Appropriateness – do the results make the model useful for the purposes of cost assessment at PR19 (in particular, is the spread of efficiency results plausible?). In some cases we identified that presumptions on the correct sign of estimation variables result in an implausibly wide range of efficiency outcomes – variables may be missing when economic or engineering logic has been applied in an over-simplistic fashion.
- Rationale – this criterion considers whether the explanation of the economic and engineering approach to developing the model is sound. Is the logic justifying the model variables sound? Does its use at this level of disaggregation work? Is the explanatory variable disaggregated, justifying the use of the level of aggregation for the cost modelling rather than more aggregated versions
- Consequences – Is there any obvious bias in either the explanatory variables included or excluded in a model, or adverse implications for the future efficiency and effectiveness of the

industry should this driver of cost be used. An example could be a model that assumed more metering resulted in lower on-going cost, which could imply a disincentive to meter more properties. Another example would be where the form of the model appears to be trying to reflect a company specific situation without explaining why this logic applies across the industry, for instance in geographic area or population variables.

Overall this means we assess many models in the consultation as “Red”. This does not mean that the models themselves have no value. Rather, that after considering them there is a good reason to prefer other models or different levels of cost aggregation, having reviewed the model results in the round. We think the models we have rated as Amber and Green are worth considering further. Green models effectively are those that we think unambiguously should be considered as part of a suite of models used at PR19 – given the wide range of models these are inevitably relatively few in number compared to “Amber” assessments where some questions are apparent. In some cases where models are rated as “Red”, there are models within CEPA’s report which appear to better meet our criteria, including statistical validity. We make indicate through our comments in the response spreadsheet where CEPA models should be considered, as Ofwat did not ask for assessment of the CEPA models directly in the consultation. We think this was a potential gap in the consultation approach that is worthy of further consideration before finalising the models for PR19.

We prefer our appropriateness criterion to the tests used by CEPA for stability of efficiency rankings and stability of inefficiency ranges. All the models could be equally “wrong” on the CEPA tests, so on its own these tests do not demonstrate the appropriateness of a particular model. We prefer therefore an assessment that recognises that a post modelling judgement as well as pre-modelling presumption on economic, operational or engineering logic is required. Effectively we put a higher weighting on policy in our review of the models, but a lower weighting on presuming this in the model selection, consistent with the NERA model selection approach.

CEPA in their modelling also presume that aggregated rather than unit cost models should be deployed, indicated by the adjusted R-squared criteria for model selection of 80%. Other models proposed in the consultation based on unit cost forms should not be screened out based on this criterion. Therefore we disagree that the relative importance of R-squared should be very high, except for aggregated cost models.

Overall observations

We set out below our overall observations and conclusions arising from the consultation:

- It was not clear to us from the consultation how Ofwat had used the CEPA report to develop the proposed models, or the process through which the CEPA models had been compared to Ofwat’s models. In general the main advantage of the Ofwat models appeared to be that they were simpler in form or had fewer explanatory variables. However, we would have expected

more exploration of the differences in the consultation document. One example is the use of water treatment complexity where the CEPA models use 4-6 level complexity and the Ofwat models use level 3-6 complexity.

- We conclude that the modelling has identified a range of plausible models. What was not clear from the consultation is whether Ofwat plan to select a single preferred model or use a range of models that are equally as plausible. We think that in many areas of expenditure there are a range of models that should inform the final assessment. Our own view is that those we have assessed as Green, together with many of those we have rated as Amber, are likely to be useful in combination for assessing costs at PR19. As there are many plausible models in the consultation, there did not appear to us to be many advantages of preferring one single model.
- For the retail bad debt models, we think the models we suggested that had been developed by Economic Insight perform better than the Ofwat models, because of the choice of deprivation variable. Equifax credit ratings appeared to us to be at least partially inside of management control, and therefore we prefer the statistically significant and high-performing models which use standard ONS deprivation data.
- In many cases the models presented in the consultation appear to perform better using our criteria on more aggregated data rather than separated into separate expenditure sub-categories. This is particularly apparent for water resources and the sub-modelling within retail. The evidence suggests that modelling at wholesale total and retail total cost level produces better models. We think this is because industry planning decisions are taken across wholesale water and residential retail and therefore this is the right level to measure efficiency, particularly between water resources and water treatment and between retail bad debt and other retail costs. The logic of selecting simple disaggregated models is not apparent from the data, because the economic, operational and engineering logic is likely to be, for “botex” models in particular, at a wholesale total and retail total level.
- Ofwat conclude that the Retail “other than bad debt” modelling does not lend itself to a “frontier” or “upper quartile” approach to efficiency targeting, rather than an industry average. With our own modelling, and in reviewing the other models in the consultation, we reach a different conclusion, particularly if Retail total cost models are used. There are clear risks from setting a Bad Debt efficiency challenge based on a different approach to other residential retail costs from an operational perspective (for instance whether customer engagement and satisfaction helps to avoid future bad debt risk by understanding and meeting individual customer’s needs). Ofwat should consider this point in the context of the form of control before concluding on the aggregation of residential retail costs.
- We agree with the conclusion that excluding regional wage factors entirely from the cost assessment appears to be supported by the modelling evidence, leaving this to company cost adjustment claims where there is an exceptional impact.
- For Water Resources modelling, we think there is a case for further consideration of what costs are included and excluded from the cost modelling. There is logic in third party and abstraction charges to be excluded from the modelling, but with increasingly diverse cross-regional buying and selling of water there are other similar costs which are embedded in charges not excluded

from the modelling that may be apparent. This may be a logical reason not to use disaggregated modelling as these factors are far less material at a wholesale total level.

- Ofwat will need to consider carefully the time period used in the efficiency modelling in two ways. Firstly Ofwat will need to consider how to assess the models companies have proposed that include time dummy variables. These dummy variables appear in some of the models that better meet the assessment criteria we have set out, including where random effects (RE) modelling forms have used. Secondly the stability of the model needs to be tested on shorter and more recent time periods. This will be particularly important if Ofwat follow the approach to assuming future efficiency changes that were implied in the KPMG / Aqua Consulting analysis for Ofwat. We responded previously to some of the challenges with this analysis, but if Ofwat do assume that totex and outcomes have enabled innovation that can be reflected in a substantially higher frontier shift than has been apparent in the past, this may limit the time period of data being used in “catch-up” efficiency models such as included in this consultation and their application.
- We do not think the random effects (RE) and OLS methods should be categorically preferred. Our assessment of the CEPA report and other models in the consultation suggests that in many cases RE models are improvements in terms of the use for assessing relative efficiency, in particular when using panel time series data. Some company models used dummy variables as a partial alternative, but in comparison to the simpler Ofwat models we do think that RE models perform better in many cases.

Summary of our assessment of the wholesale water models

- None of the Water Resource models meet the criteria we set out in this consultation response. CEPA model 4a appears to have the most potential. The YKY models with reservoir capacity appeared to perform well, but we have concerns that this factor is not outside of management control, and in general not all capacity at reservoirs is in practice usable, which may lead to data consistency concerns.
- For Water Treatment many of the models in the consultation had potential, although the most simplistic models proposed by Ofwat (OWT1 and 2) were dominated by more complex CEPA models. In general the Ofwat models may improve through use of better differentiation of treatment works complexity (e.g. levels 4 and 5) rather than using levels 3-6.
- For “Water Resources plus” a number of Ofwat models (OWRP4, 5 and 6) best meet our criteria. In addition there are CEPA RE models that perform well and should be considered (e.g. valid with use of treatment levels 4-6 rather than 3-6).
- For Treated Water Distribution our main concern was whether using “length of main relined or refurbished” was in practice sufficiently outside of management control. Our concerns on this are reduced in more aggregated modelling, as in practice the better models for this level of disaggregation all included this variable. This informed our view that more aggregated modelling may be preferred.

- For “Water Network plus” we identified a range of models that met our criteria. We prefer the CEPA models to those proposed by Ofwat as the modelling form appears to have more appropriate explanatory power, with ONPW5 and 6 there are insignificant variables that CEPA’s RE approach found to be significant. There are a wide range of models with different drivers and therefore it appears that considering how to triangulate between the models we rated as “Green” and “Amber” may be appropriate.
- “Wholesale Water” model conclusions are similar to network plus, and therefore on balance we think this supports an aggregated approach to modelling. If, as we believe, there are no viable water resource models, then aggregation at wholesale water may be an appropriate way forward.

Summary of our assessment on enhancement models

- We did not think the enhancement models proposed were logical, in particular the use of a smoothed “water delivered” explanatory variable for lead. There is likely to be significant innovation required in the industry to meet future lead standards, and pilots of novel approaches required. Given the small element of plan costs, perhaps the overall “Water Wholesale” position could reflect the efficiency challenge for lead programmes to avoid unwanted consequences and to encourage innovation (as lead programmes will reduce on-going water treatment costs as plumbosolvency dosing is reduced).
- We think the need for new development costs and new connection costs to be modelled as enhancement is not evident, except as a simple unit cost model. Considering this, there appear to be valid Water Botex plus models that could be used.

Summary of our assessment on residential retail models

- We think there are a range of plausible models for modelling residential retail costs. The best models appear to use standard ONS data on multiple sources of deprivation (IMD). Use of unemployment rates appear crude measures of deprivation in comparison. We prefer published national data to Equifax default rate or credit status data, at least in part because these factors are partly within management control and imply an approach to debt collection that may not be in the interests of vulnerable customers. They reflect actual defaults, not the underlying customer characteristic and should not be considered to have predictive value. The Ofwat Bad Debt models focus on default propensity, which wider customer service and sensitive treatment of the individual characteristics, including social tariff offers, in practice can and should be managed. This is the most critical point in the modelling consultation that Ofwat should re-consider very carefully as it concerns matters of public policy and interest.

- Models should include either the distinct number of water and wastewater customers, or a suitable dummy variable. Indicators of whether customers are dual or single billed (rather than who serves them) are substantially within management control.
- We prefer Retail total expenditure modelling. Although there are valid models at a disaggregated level, retail businesses should look at their total cost to serve, as it is within management control whether to combine or separate different retail activities. Given that customer service is measured at an aggregated level, and similar explanatory variables apply in the aggregate and disaggregated modelling, this provides an objective reason to model at total retail cost level.
- We think it is the model specification and the limited number of explanatory factors explored that leads to Ofwat's conclusion that "costs other than bad debt" should be on an average cost to serve basis. We think there are models in the consultation that suggest otherwise and that aggregated modelling should be considered. We note that Saal and Nieswand came to similar conclusions in their commentary on Anglian Water's phase 2 modelling.
- The Ofwat Retail "totex less bad debt" models use time dummy variables, rather than exploring other explanatory variables. They are problematic to use in practice, in particular because of Ofwat's PR19 approach to making separate allowances for inflation and productivity.
- In the Retail models scale is uniformly important. This lends itself to using total connected properties alongside a further dummy variable or control for single service customers. The models suggest that there is a structural difference between retail costs for WoCs and WaSCs and therefore approaches that control for this should be a feature in the retail models. There appear to be a range of approaches in the modelling which are plausible, but we generally conclude that a single service dummy variable is the simplest and most effective approach.

We hope our contribution supports Ofwat's final conclusions on the approach to econometric modelling at PR19. If you require any further clarification on our response, please do not hesitate to contact us.

Yours sincerely

A handwritten signature in blue ink, appearing to read "Iain McGuffog".

Iain McGuffog
Director of Strategy & Regulation