Ofwat draft Cost Model commentary by Anglian Water

Due to time constraints, we have written detailed commentary on only the Ofwat Wholesale models for the four wholesale cost controls plus the two wholesale integrated cost models.

We have made brief comments on the other Ofwat wholesale models along with the other wholesale models submitted by other companies on the accompanying spreadsheet.

Rather than give detailed commentary on our own models, we point to our recent Cost Modelling Report, published in March.

We intend to provide commentary on the retail and enhancement models shortly.

1. Water Resource Models

Ofwat Models OWR1 and OWR2

Traffic Light Rating -Red

At PR19, Ofwat has introduced a Water Resources price control with the intention of facilitating competition in water abstraction and providing further information with regard to water abstraction costs. Given this, the use of connected properties as the only output is surprising, given that water abstraction or some other volumetric measure of water usage is the logical output measure, as defined in the Regulatory Accounting Guidelines.

Furthermore, these models are overly parsimonious and in no way adequately capture the drivers of water resource costs, which are driven by factors such as the type of water abstracted, the scarcity of supply, the density of population etc. In fact, the model presented by Ofwat does little more than show a correlation between the size of a firm as measured by its connections and the size of its water resource expenditures.

In our two cost modelling reports, published in September 2017 and March 2018, we and our academic advisors have highlighted issues with regard to data quality, definition and separability for this modelled activity. These initial models suggest that Ofwat has also found it difficult to model cost at this level of disaggregation. However, it is notable that almost all the other models presented by companies are superior to those presented by Ofwat, in that they appear to use more appropriate drivers of water abstraction botex.

2. Network Plus Water Models

CEPA Models ONPW1-ONPW8

Traffic Light Rating -Yellow

Conceptually, Network Plus Water includes all the wholesale operations required to supply water with the exception of water abstraction: In other words, it includes Raw Water Transport, Water Treatment, and Treated Water Distribution costs.

Given this, we believe that these models have an inherent weakness of attempting to model a complex multiple output production process, after taking the restrictive modelling decision to preclude the use of more than six variables, and adopting an excessively conservative approach to multicollinearity, which will severely limit a model's ability to capture the impact of cost interactions between multiple outputs. With 107 observations available, it should be feasible to better define a tractable yet interpretable model of the complex inter-relationship between the volume and type of water treated, the number of connections served, distribution losses, metering, transportation distances, density and size of population served, etc. We therefore believe that to improve its modelling of Network Plus activities, Ofwat must first remove the artificially strong constraints imposed by the modelling approach adopted by CEPA.

Focusing in more detail, we note that ONPW1 to ONPW8 appear to share a common assumption that a single output measure, augmented with only explanatory factors can adequately explain the determinants of costs in what is a complex multiple output production process. Thus, in models ONPW1 to ONPW4 Water Network Plus costs are primarily driven by the single output measure of connected properties, while models ONPW5 to ONPW8 assume that the single primary output of mains length is sufficient. Moreover, ONPW1 to ONPW4 are so highly restricted that none of these models include any direct control for what is generally accepted to be a nonlinear relationship between the geographic distribution of connected properties in a water company's operating territory and costs. While ONPW5 to ONPW8 do include a logged density variable, this is in fact a highly restrictive specification (which is readily demonstrated to be a log log specification), when for example compared to the inclusion of both the log of density and its square in the Southern Water's Network Plus models, or the alternative use of two density variables capturing high and low density areas employed in South East Water's models.

Similarly, despite the significant botex contribution of water treatment volumes within Water Network Plus activities, as well as the presence of known and substantial differences in the costs of treatment and distribution related to water abstraction source, no volumetric or share based measures by water abstraction type variables are included in any of the eight models produced by CEPA for

Ofwat. This is consistent with what appears to be a primary focus on modelling distribution expenditures as the remaining explanatory variables can primarily be seen as factors influencing distribution costs and largely excluding treatment related variables.

Overall, while the models' explanatory factors are of merit, a more comprehensive approach that might better allow for how different water treatment sources and demographic and geographic costs influence overall modelled water systems costs would be desirable.

3. Wholesale Water Models

CEPA Models OWW1-OWW12

Traffic Light Rating -Yellow

Conceptually, Wholesale Water includes *all* of the vertical supply chain from Water Abstraction to Treated Water Distribution. As such, these models must include appropriate output variables to control for water abstraction, treatment and distribution costs and should allow for appropriate interactions to deal with how differences in companies' system configurations influence costs. Ofwat's Wholesale water models are quite similar in specification to its Network Plus water models. In assessing them, one could argue that within CEPA's modelling framework, which assumes that a single scale variable coupled with the inclusion of other explanatory variables is adequate, all necessary aspects of the water supply chain are modelled. Thus for example, operating characteristic variables related to distribution, water treatment and water resource and treatment pumping are included in the model.

However, given other models reported in this cost assessment consultation, as well as a general understanding that system configurations and costs are influenced by a complex interaction between available water resource type, water scarcity, settlement patterns and density, it is less than clear that a specification which does not allow for such factors to interact with outputs in determining cost interactions is appropriate. To put it another way, as CEPA's modelling approach effectively assumes that costs can be measured primarily with a single "scale" relationship, it is not feasible to capture nonlinear relationships between population density and serving connections, or how differences in water abstraction type influence the actual elasticity of multiple output production to costs. We therefore suggest a relaxation of some of the restrictive modelling assumptions imposed by CEPA to test if a more flexible approach to modelling cost interactions between water abstraction, treatment and abstraction better accounts for differences in water company costs.

4. Bioresources Models

CEPA Models OBR1-OBR3

Traffic Light Rating -Red

The Bioresources models developed by CEPA for Ofwat are inadequate.

Given that Bioresources involves the expenditures covering the transport, treatment and disposal of sludge resulting from sewage treatment, OBR1 is unacceptable given its use of properties as a proxy for an appropriate output. This conclusion is supported by Ofwat's own Regulatory Accounting Guidelines.

At this basic level, OBR2 and OBR3 at least include an appropriate output for the modelled activity as they use sludge produced as a basic output measure. However OBR3 should be excluded on basic modelling principles as its only difference compared to OBR2 is the inclusion of a single statistically insignificant variable. Moreover, as the difference in sludge treatment and transportation costs results from complex relationships driven by geography and population settlement patterns, it is very disappointing that modelling to capture these issues appears to have been limited to considering this simple variable.

Finally, turning to OBR2, we have a model with a single output and single control variable indicating the share of treated sludge disposed to farmland. As this model, can in no way capture how the size and location of sewage treatment plants, which is in turn driven by population settlement patterns influence sludge treatment costs, this model is also inadequate Beyond this, the models are entirely inadequate and fail to grapple with the complex factors driving sludge costs, both within and between companies. It is particularly disappointing that CEPA has not attempted to employ Ofwat's recent approach to measuring sparsity and density in an interactive way so as to capture these effects.

5. Network Plus Wastewater Models

CEPA Models ONPW1- ONPW10

Traffic Light Rating -Red

Network Plus Wastewater includes all activities and expenditures related to the collection of sewage and its treatment at sewage treatment plants. It excludes Bioresources. As such, this modelled activity represents a complex multiple output production process involving the collection of sewage, its transportation in the sewerage network, and its treatment. Given that this is the case, the multiple outputs to be modelled must capture how interactions between number of connections, transportation distance, and treatment loads influence overall expenditures. Furthermore, a noteworthy feature of the overall system is how relatively high sewage transportation costs cause companies to carefully balance the benefits of plant size economies in sewage treatment against the network transportation costs required to achieve sufficient volumes to exploit such economies. Thus, firms with large urban populations rely on larger sewage plants that justify more network and interconnecting. In contrast, firms with more scattered populations operate a much larger number of smaller sewage plants, with no network connection between small distinct population centres. Given this discussion, a model which purports to model Network Plus Wastewater must capture how geography and settlement patterns influence sewage transportation costs, and how this influences cost interaction influences the siting and scale of sewage treatment facilities.

Given this discussion, the models developed by CEPA for Ofwat are too parsimonious, and suffer from the general weakness of CEPA's modelling approach of relying on a single output variable and deliberately excluding modelling approaches that may be required to capture what are complex cost interactions driving differences in firm costs. To be parsimonious in our own review of these models, we first exclude consideration of ONPW3, ONPW4, ONPW7, and ONPW8 which all can be excluded relative to comparable models based on appropriate application of the Adjusted R Squared statistic. This leaves 3 model variants, demonstrated with 2 different "output" specifications, respectively employing properties served and sewage treatment load as the only output proxy. Unsurprisingly, given the high correlation between connected properties and treatment load, these different "output" specifications yield virtually no difference in estimated coefficients, or other statistics. This therefore leaves consideration of the model variants, and for this discussion we focus on the models employing treatment loads, for ease of exposition, as opposed to any preference for this approach. These variants all include a variable indicated high capex post 2001 as the % length of sewers laid post 2001, and then either includes a log(density) variable (thereby imposing what is really a peculiar and highly restricted log log relationship between botex costs, sewage loads, length of mains, and connections) or a an ammonia consent control designed to capture higher sewage treatment standards. As there is no model which includes all these controls, it is therefore impossible to tell if either of these variants are statistically preferred. At a minimum a general model in which the models presented are nested should have been provided, otherwise we are left to our own devices to determine whether the inclusion of density or treatment controls is statistically preferred.

However, beyond these statistical issues it is unclear how these models capture how differences in WaSC operating environments cause them to structure their sewage collection and treatment systems appropriately so as to minimize overall system costs. Given this, it would be appropriate for CEPA to firstly abandon its highly restrictive *a priori* modelling rules, and then carefully consider how other models presented in this consultation strive to capture the impact of geography and settlement patterns on sewage system design and costs, both within and between companies.

6. Wholesale Wastewater Models

CEPA Models OWWW1- OWWW8

Traffic Light Rating -Red

Wholesale Waste Water includes Network Plus Wastewater and Bioresources. As such, modelling at this level of activity must control even more thoroughly than is the case for Network Plus for the complex nature of the multiple output system discussed above. In general, Ofwat's models at this level resemble that done for Network Plus Wastewater, and provides restricted variants without providing the general models required to test between them. Thus, for example models including controls for sludge disposed to farmland and including controls for trade effluent customers are included, but not models including both. In summary, and relying on arguments made above, while all of the control variables included in these models are plausible and at face value are appropriate, it is not clear that the models provide a conceptually and statistically adequate model that can explain how differences in system configurations driven by geography and demographics result in differences in expenditure required to operate complex multiple output systems. Most fundamentally, while it is clear that firms do operate distinct sewage treatment systems, each of which balances the trade-off between scale in sewage and sludge treatment against the cost of network transportation costs required to achieve a given treatment plant scale, it is not clear that these models have adequately captured how such system level differences influence difference in required firm level expenditures.