Identifying drivers of water costs

A high-level framework for cost modelling

Report developed through collaboration between
Severn Trent, Tim Keyworth and Reckon

February 2018

WONDERFUL ON TAP

SEVERN TRENT

1. Executive summary

This paper aims to assist with the development and review of potential cost models for use at PR19 by providing a high-level framework for testing the extent to which different econometric models can be expected to capture underlying cost drivers. Attention is focused on water network plus botex, but, in principle, a similar approach could be adopted to other areas and categories of cost.

The framework includes the identification of eleven factors as underlying drivers of water network plus botex:

	Factors driving Water Botex Net+ costs		
1.	The distance water has to be transported;	7.	Opportunities for the achievement of economies of scale in water treatment;
2.	The number of customers to whom water is distributed;	8.	The extent and forms of treatment that are required;
3.	The quantity of water that has to/may have to be transported and treated;	9.	Regional differences in relevant input costs;
4.	The geography and topography over which water has to be transported;	10.	Service quality variations;
5.	Opportunities for economies of scale in the transportation of water;	11.	The significance of other customer characteristics.
6.	The extent to which transportation activities are affected by 'congestion';		

Application of the framework involves potential models being assessed in relation to the following question:

> **Does the set of explanatory variables included in the model provide an adequate means of addressing each and all of the eleven identified factors?**

This does not imply that models should include an explanatory variable to reflect every factor. Rather, the framework provides five potential responses to this question which can be used to explicitly consider and identify the means by and extent to which each factor is captured (or not) in a given model. This recognises that factors may be captured indirectly, or outside of the model (e.g. through pre-modelling adjustments), while also providing a basis for identifying where factors are significant, but not captured in one of these ways.

By making this process explicit, and pre-defining a set of underlying factors, the framework aims help guard against the risk of material factors being unwittingly omitted without a clear basis to do so. The identified list of factors can also provide a list against which coherence checks can made in terms of resulting model coefficients.

The paper highlights that company activity aimed at maintaining the long-term capability of the network also merits careful consideration under the framework, in terms of whether and how it is captured. This follows because econometric benchmarking of the kind that Ofwat is undertaking involves deriving relationships on the basis of past expenditure within a defined period. If relative expenditure levels within that period are driven to a significant degree by differences in levels of activity aimed at maintaining the long-term capability of assets that are not otherwise captured, then this could be a material source of 'noise' that undermines the accuracy of results.

When attention turns to the practical question of how underlying cost drivers might be measured, data availability and endogeneity issues are likely to soon emerge. Finding variables for which there is reliable data available, and that are not – at least in part – under the control of a company's management can be challenging. It is important that the materiality and sources of endogeneity concerns, and the extent to which these risks might be mitigated, are carefully considered. The remaining risks associated with endogeneity should then be balanced alongside the (potentially greater) risks associated with the adoption of alternative approaches such as omitting or inadequately reflecting an underlying cost driver in the modelling.

2. Why develop a high-level framework?

This paper aims to assist with the development and review of potential cost models for use at PR19. For a model to be regarded as acceptable for use, it will have to perform sufficiently well in statistical terms. There are material risks, though, that models which do perform well on certain statistical measures/tests may nevertheless identify spurious relationships. The relatively small size of the sample available to Ofwat for econometric benchmarking can make these risks particularly pronounced.

An important means of guarding against this issue is for the model development process to be informed by, and tested against, some underlying engineering and economic understandings of what would be expected to drive costs. In particular:

> **Does the model imply relationships that unduly ignore, or run counter to, underlying cost driver relationships that would be expected?**

This kind of sense checking exercise formed an important part of the CMA's review of the PR14 econometric models in its Bristol Water Inquiry, and can help to ensure that the modelling process remains sufficiently grounded.

The aim of this paper is to provide a high-level framework against which this type of cross-checking exercise can be undertaken. By establishing a defined set of expected fundamental cost driver relationships in advance, the framework can provide a pre-determined reference point against which models can be assessed. While there may be good reasons for apparent divergences between the relationships implied by a model and the high-level relationships set out in the framework, the aim is to help make the identification and assessment of such divergences more explicit and structured as models are developed and reviewed.

3. Which costs is the framework intended to apply to?

We focus our attention here on Botex for the overall Water Network+ business. That is, on:

- **Botex:** Operating costs + capital maintenance.
- **Water Network+:** Raw water distribution, Water Treatment & Treated Water Distribution.

3.1 Botex rather than Totex modelling – challenges with enhancement

Focusing on Botex is consistent with the position adopted by the CMA in its Bristol Water assessment for PR14, and is broadly in line with the position set out in Ofwat's PR19 final methodology document. That is, Ofwat said that its econometric models would cover base costs (referred to here as Botex), but that it may be appropriate include certain enhancement activities as well (for example, where the driver of enhancement costs was regarded as common and persistent for all companies). This approach has come to be referred to as Botex+ modelling and would involve the separate assessment of other enhancement activities that are more company-specific, irregular and difficult to predict. This approach is welcome as the idiosyncrasies that are often associated with enhancement expenditure could pose material challenges to development of a sufficiently reasonable set of totex models.

Drivers of enhancement expenditure have not been considered here, and the inclusion of some aspects of enhancement in the cost modelling, as well as Botex, may mean that additional questions should be added to the framework set out below. The importance of this will depend on the materiality of those additions, and the extent to which the relevant costs are correlated with other aspects of Botex that are already captured.

3.2 Network plus rather than wholesale – challenges with water resources

We have focused on Water Network+ rather than Wholesale Water. The separation of Water Resources removes a material potential source of complexity and idiosyncrasy. In part, this relates to abstraction charges, and - as Ofwat has highlighted – they can be isolated from cost modelling for separate treatment. More fundamentally, though, it is simply not clear that the drivers associated with differences in water resource costs across companies align closely with the drivers of Water Network+ costs.

Water resource costs can be expected to be linked to a range of physical and environmental characteristics that different companies face (which affect resource availability). There are also a range of historic factors (including the terms of long-standing bulk supply agreements) that have a direct and ongoing bearing on water resource costs, and that raise additional complex issues for cost benchmarking. As Water Resources accounts on average for around 12% of Wholesale Water Botex, if these characteristics are not addressed sufficiently then the inclusion of Water Resources in the costs to be modelled may introduce a significant degree of 'noise' into the process. In principle, this could be overcome through the inclusion of additional variables that capture the drivers of Water Resources costs to a sufficient degree. In practice though, given the relatively small sample size that is available, increasing the number of explanatory variables can put a different type of strain on the likely adequacy of the modelling. This point was emphasised by the CMA when highlighting concerns over the number of explanatory variables included (most notably in the PR14 full totex model).

3.3 Potential for alternative model scopes

Given the above considerations, we have focused on Water Network+ Botex modelling here. In principle, a similar approach could be adopted to the consideration of wastewater costs, and to the more disaggregated consideration of Water Network+ costs. We welcome Ofwat's intention to consider the role that more disaggregated cost assessments might play in the determination of cost benchmarks. Such assessments may provide an opportunity to more accurately identify the cost relationships. However, disaggregated modelling does raise some problems and complications for modelling that are avoided under a more aggregated approach. These include potential inconsistencies in cost allocation, and the inter-relationship between costs in different categories.

4. What are the underlying drivers of Water Network+ Botex?

The Water Network+ service involves the transporting of raw water (that has been secured by the Resources business unit) to a treatment works, addressing any treatment requirements to ensure the appropriate quality of water can be provided, and then distributing that treated water to consumers.[1] As Table 1 below shows, Treated Water Distribution is much the largest source of Network+ Botex. Overall, 'transportation' (the distribution of raw and treated water) accounts for around two thirds of Water Network+ Botex, with treatment accounting for the other third.

Table 1: 2016/17 Industry Water Network+ Botex (Source: Ofwat 2017 Cost assessment submission)

Business Unit	Botex (£m)	% of Network+
Raw Water Distribution	109	3%
Water Treatment	1,039	31%
Treated Water Distribution	2,168	65%
Water Network+ Total	**3,315**	**100%**

[1] For some groundwater sources these steps may be truncated with raw water being subject to only minimal treatment at the abstraction site ahead of distribution (given its already high quality).

We have sought to capture the underlying drivers of Network+ Botex by identifying a relatively short list of factors that would be expected to affect (efficient) costs given the activities of transportation and treatment that need to be undertaken. The list of 11 factors presented in table 2 aims to collect together the main high-level reasons why efficient levels of cost might differ between companies. We have sought to explain why each factor may have a material bearing on costs - primarily by reference to aspects of the scale and complexity of distribution and treatment activities. The list is intended to provide a checklist for assessing the extent and manner that different underlying factors are taken into account in the cost assessment process. As will be highlighted below, this does not imply that models should include an explanatory variable to reflect each factor.

Table 2: Checklist for key underlying factors affecting Water Network+ Botex

Key underlying factor	Why relevant to costs?
1. The **distance** water has to be transported (which will be affected among other things by the relative locations of customers and sources of water)	• Transporting water over longer distances will require more assets that will be need to be maintained. • Operating costs can also increase with distance (e.g. when more pumping is required).
2. The **number of customers** to whom water is distributed	• More customer connection points will require a greater scale of distribution infrastructure, including water mains, pumping equipment and service reservoirs.
3. The **quantity** of water that has to/may have to be transported and treated	• Peak transportation and treatment requirements will affect capital costs as they will affect the capacity that assets need to provide for. • The quantity of water will also affect operating costs (e.g. energy for pumping and chemicals for treatment).
4. The **geography** (including the topography) over which water has to be transported	• Pumping costs will be affected by topography. • Geography can affect costs in a range of other ways such as through effects of pipe maintenance requirements (e.g. as a result of corrosion).
5. Opportunities for the achievement of **economies of scale in** the **transportation** of water	• When customers are more spread out, more assets will typically be required per customer, and opportunities for economies of scale in network assets can be more limited (e.g. more, smaller pumping stations are likely to be required alongside a greater length of lower-diameter water mains).
6. The extent to which transportation activities are affected by **congestion** issues	• Operating in densely populated urban areas can put upward pressure on costs as a result of above and below ground congestion, and the implications this can have on the need for, and complexities and resource requirements associated with, carrying out maintenance work.
7. Opportunities for the achievement of **economies of scale in water treatment**	• There can be significant economies of scale in water treatment, and so the extent to which those economies are achievable in practice can have a material bearing on costs.

8.	The **extent and forms of treatment** that are required given the quality of raw water	• Treatment costs will be affected by what forms of treatment have to be provided, and that will depend on the quality of the raw water.
9.	Regional differences in relevant **input costs**	• Labour and energy costs can vary between regions in ways that have a material bearing on overall transportation and treatment costs.
10.	**Service quality** variations	• Costs can be affected by differences in the levels of service quality that different companies provide. • This could include the provision of environmental and related outcomes (such as leakage levels).
11.	The significance of **other customer characteristics** that may affect transportation and treatment requirements	• For example, meter penetration levels and the relative size of non-potable demand may affect cost materially.

A whole range of much more specific factors could also be identified that will affect Network+ Botex. However, the aim here was to develop a concise list of high-level factors that can be understood as fundamental drivers of cost. This need is driven in part from the relatively small sample size that is available. Given the relatively small number of observations, the inclusion of additional explanatory variables can undermine the reliability of the modelling process. This issue was highlighted by the CMA in its assessments of the PR14 econometric modelling in its Bristol Water Inquiry.[2]

4.1 Consideration of asset history and maintaining the long-term capability of assets

The opening condition and configuration of distribution and treatment assets raises a number of issues. Given the very long-lived nature of many distribution and treatment assets, it is clear that costs now may be affected by decisions made many years ago. Those past decisions will affect the relative condition of assets. It might be argued that this makes it important to address questions of asset health when undertaking econometric benchmarking. Alternatively, it might be argued that differences in the opening condition of distribution and treatment assets should not be allowed for in econometric models so as to protect the incentive properties of the price control framework. In this case, materially diverging from that position would need to be considered in a more bespoke company-specific manner.

The long-lived nature of distribution and treatment assets can also have other less direct implications for the assessment of efficiency, as it can mean there is a significant degree of 'path dependency'. That is, it affects the ease (and cost) with which the distribution and treatment arrangements of any given company can be re-configured, such that decisions made many years ago (including decisions that with the benefit of hindsight may look sub-optimal) can be effectively 'baked in' to a significant degree. For example, the fact that it was considered appropriate to build a treatment works on a particular site 50+ years ago can greatly affect the relative cost of alternative potential treatment options now, and the location of that site can have long enduring implications for distribution costs.[3] Given its structural significance, the long-lived nature of distribution and treatment assets has implications for the consideration of how the factors identified in Table 2 might be captured. As will be considered below, it is relevant to the assessment of endogeneity concerns.

[2] For example, in Paragraph 4.50(c).

[3] As, for example, the location of the plant may have ongoing implications for pumping costs.

Irrespective of the view taken on the relevance of asset condition/configuration to the determination of future allowances, it is important to consider how past asset management can affect the cost assessment process. This follows because econometric benchmarking of the kind that Ofwat is undertaking involves deriving relationships on the basis of past expenditure (with those relationships then used to make projections of what future allowance levels may be appropriate). If past expenditure differences across companies reflect differences in levels of activity aimed at maintaining the long-term capability of assets, and those differences are not captured by other explanatory variables in the model, then there is the potential for model results to be materially inaccurate because relevant variables have been omitted.

This implies that careful consideration should be given to the need to reflect different levels of activity aimed at maintaining the long-term capability of the network (during the data period used for model estimation) in order to avoid such differences unduly affecting the basis the upon which future expenditure requirements are forecast. To not do so may effectively leave a significant amount of 'noise', and undermine the accuracy of the results that are generated. While we have not considered that issue as an additional underlying driver of Network+ Botex (as per Table 2), it is an additional factor that merits assessment against the questions set out later (in Table 3).

5. Guarding against the omission of material factors when developing models

The modelling process will inevitably involve the adoption of a relatively simplistic view of what determines (an efficient level of) costs. However, regulatory benchmarking processes may then assume that all (or an identified portion of) differences between a company's actual costs, and the level of costs implied by modelling, are a reflection of the company's overall level of efficiency being higher or lower than average (and higher standards – such as upper quartile – then typically involve this same underlying assumption being applied).

In practice, this is clearly not the case. Econometric models will inevitably provide a simplified depiction of what drives costs, and some relevant explanatory factors will have been captured imperfectly. Given this, some of the gap between modelled and actual costs will be due to modelling error rather than inefficiency.

The use of multiple models can help mitigate this kind of risk, and make it more likely that the overall efficiency judgements made on the basis of the modelling are broadly reasonable. Much will still depend, however, on the adequacy of the models within the selected set, and on what that overall set looks like. For example, the close alignment of the predictions of a number of different models can be seen as demonstrating model accuracy, but this can be misplaced if those models are similarly specified, and systematically reflect cost conditions poorly.

The checklist of factors set out in Table 2 is intended to help guard against the risks described above. The checklist does not imply that models should include a single variable to reflect each of the factors in the table. Rather, the key test for any given model is intended to be as follows:

> **Does the set of explanatory variables included in the model capture each and all of the factors in Table 2?**

In practice, there are a range different ways in which the question of whether a given factor is being captured could be answered. Table 3 below sets out a range of potential responses. We suggest that these potential responses can be used as a means to assess proposed econometric models, in terms of their coverage of the underlying cost drivers. This assessment would not delve into the details of econometric and statistical results and would focus more on the design and rationale for proposed model specifications.

Table 3: Is each factor captured in the model specification? – Potential responses

Potential response	Details/Comments
a) Yes, Directly	The factor is explicitly captured through the inclusion of a given explanatory variable.
b) Yes, Indirectly	The factor is picked up as other explanatory variables provide a sufficiently close proxy, either individually or in combination.
c) No, but factor not sufficiently material	The factor has been identified as not having a sufficiently material effect on costs for inclusion to be necessary.
d) No, but ex-model adjustment to be made	A separate, targeted adjustment has been, or will be, made outside of the model (either a pre-modelling or post-modelling adjustment) so that the factor does not need to be captured through the model specification.
e) No – factor not captured	The factor looks likely to have a material bearing on costs, but is not picked up in either the model specification or in an adjustment outside the model.

The table includes four potential responses as to why a factor might not be reflected directly in a model. These are considered in turn below.

5.1 The factor is captured, indirectly

A factor may be captured indirectly because other variables are considered to provide a sufficiently close proxy, either individually or in combination. In such a situation, adding a further explanatory variable to reflect the factor may add little, while at the same time increasing risks of inaccuracy given the small sample size.

For example, there are strong correlations between factors 1 (the distance water has to be transported over), 2 (the number of customers to whom water is distributed) and 3 (the quantity of water distributed). All three are concerned with the scale of the company's operations. It is possible that within a particular model, any one factor -might be taken as a proxy for the other two. This may be particularly so where one or more other variable within the model is likely to capture material differences between the three scale variables (e.g. measures of how spread out customers are may capture the relationship between distance and number of customers). In practice, it will be an empirical question whether the degree of correlation is sufficient to justify such an approach.

5.2 The factor is not considered sufficiently material for inclusion in the model

The second potential response highlighted in Table 3 for why a factor might not be reflected directly in a model, is that it is not considered sufficiently material. Where that is the case, the model may be regarded as reflecting the factor adequately without any direct or indirect account being taken of it. Our starting point is that the factors in Table 2 are material drivers of costs, but it may be the case that more detailed analysis reveals that this is not the case and this could provide a rationale for excluding them from the modelling.

5.3 The factor is not captured, but adjustments will be made outside of the model

The third potential response highlighted in Table 3 for why a factor might not be captured directly in a model is that it has been addressed separately through an adjustment outside of the model.

Regional differences in relevant input costs (factor 9) can provide an illustration of when this kind of ex-model adjustment approach may be desirable. Regional labour cost differences can have material bearing on the relative costs that different companies face. However, given limitations with the available data (and the small sample size issues referred to above) this factor looks likely to be accounted for most effectively by making a separate adjustment outside of the model where labour costs are identified as materially different to average. It may be that a similar ex model approach would be considered appropriate to address service quality differences (factor 10) that are regarded as sufficiently material.

5.4 The factor is not captured in either the model specification or in an adjustment outside the model, notwithstanding its likely materiality

The final potential response highlighted in Table 3 as to why a factor might not be reflected directly in a model is that it has been omitted notwithstanding its likely materiality. Identifying when this is likely to be the case is important as it should affect the confidence that is placed on the model results, and on how the contribution that a given model makes to a broader overall selected set should be interpreted. The approach taken to data quality concerns and to endogeneity issues can have a significant bearing on the likely extent to which underlying factors are left not adequately reflected in models, and this is considered further below.

6. Data limitations and endogeneity

Table 2 was concerned with the high-level identification of underlying (what might be understood as fundamental) factors that drive cost. When considering how these underlying factors might be measured, data availability and endogeneity issues can often emerge. Using variables that are to some extent under a company's control raises some risks given that the objective is provide a forward-looking view of efficient expenditure requirements. These risks can, however, easily be overstated, and need to be considered alongside other modelling risks. It is likely to be difficult to find variables to address each of the high-level factors for which there is reliable data available, and that are not – at least in part – under the control of a company's management.

Factor 7 (from Table 2) can be used to illustrate these points. The factor rests on the expectation that overall Network+ Botex will be materially affected by the opportunities available to companies to benefit from plant-level economies of scale in water treatment. When presented in this high-level manner, that proposition ought to be relatively straightforward and uncontroversial. However, the question of how the scale of that 'opportunity' can be reliably and effectively calibrated for modelling purposes is much more complex and open to dispute.

Measures of customer density can have some attractions in this context. They offer the prospect of being able to use an exogenous measure, and clearly have some relevance to opportunities for economies of scale. However, customer density measures take no account of the structure of the supply-side (e.g. of where raw water is sourced from, and in what condition). Such factors clearly also have a major bearing on opportunities for economies of scale. This is perhaps most apparent when groundwater sources are considered.

In the absence of a compelling, wholly exogenous alternative, direct consideration of the scale of company treatment works provides an obvious alternative way of considering this issue. Reliable data on the number of works (of varying type) is readily available, and could be used, for example, to infer average works size. Using a measure of what companies have actually implemented in terms of treatment works size has some attractions, including that it captures 'revealed' views on the perceived opportunities for economies of scale: company executives (at some point) considered the opportunities sufficiently concrete to base long-term investment decisions on them.

As was discussed earlier, the long-lived nature of treatment assets tends to lessen the significance of endogeneity concerns given the resulting constraints it can imply for plant and network reconfiguration options. Nevertheless, as companies retain some flexibility over the evolution of plant size, concerns can arise that the use of such a measure may bias future operational decisions. While this merits careful consideration, it is not obvious that it should raise major difficulties. While it is clearly possible that expectations over likely future benchmarking approaches could influence company treatment works sizing decisions, there are considerable uncertainties surrounding how costs will be assessed in the future. This would be expected to affect the reliance that would be put on such expectations in the context of long term investment decisions.

Also, it would be open to Ofwat to directly seek to undermine confidence in the likely future usage of a measure that it considered expedient to use for PR19, but that gave rise to some concerns with respect to possible distortions to dynamic incentives. For example, those concerns could be explicitly recognised, and a programme aimed at developing an appropriate more exogenous variable for use in PR24 could be initiated and/or committed to.

When the implications of endogeneity concerns are being considered, therefore, it is important to:

- Carefully assess the materiality and sources of endogeneity concerns.
- Consider the extent to which the resulting risks can be mitigated (e.g. through regulatory actions that can be expected to diminish the significance of the relationships that are causing concern).
- Balance the remaining risks that may be associated with endogeneity alongside the (potentially greater) risks that may be associated with the adoption of alternative approaches, including those that involve omitting or inadequately reflecting an underlying cost driver in the modelling.

7. Testing for overall coherence

Most of the above has focused on the risks associated with omitting material factors from the modelling process. However, the checklist of underlying factors in Table 2 can also be used to consider a subsequent question of whether the coefficients that a model identifies should be understood as coherent. It is notable, in this context, that the CMA's assessment of PR14 totex models questioned whether some of the implied relationships 'made sense'.

Expectations over the values that coefficients would have will depend on a range of matters, including those associated with model specification which have not been considered here. However, in broad terms, it is straightforward to identify a number of cost relationships that would be expected. For example, other things equal, costs would be expected to increase:

- with transportation distance;
- with the extent of geographic 'challenges' (notably the extent of pumping required);
- where there are more limited opportunities for economies of scale in transportation;
- where transportation activities are more affected by congestion issues;
- with the quantity of water that has to/may have to be transported and treated;
- where there are more limited opportunities for economies of scale in water treatment;
- with the extent to which raw water needs to be subjected to more extensive/complex treatment processes.