PR24 Draft methodology:

Severn Trent cost assessment response

7 September 2022

WONDERFUL ON TAP

SEVERN TRENT

Ofwat's PR24 Draft Methodology consultation

Question 6.1. Do you agree with our proposed approach to setting efficient expenditure allowances at PR24?

We consider that the fundamental premise of setting efficient expenditure allowances should be to ensure that companies are able to meet their statutory requirements and deliver good quality resilient services to customers and the environment at a cost that is fair. Making sure that cost allowances are fair and requirements / service commitments are deliverable requires the chosen regulatory cost assessment approach to be correctly specified. This means that legitimate costs are accounted for, and that efficiency challenges are reasonable and justifiable, rather than stretching without basis.

Using econometric cost models to challenge companies to be efficient, whilst allowing for required drivers of cost outside of companies control, has been tested and refined across two price reviews. We remain of the view that this is an appropriate platform to identify efficient expenditure requirements (provided they are complemented by other regulatory assessments where econometric analysis is less well suited).

However, it is critical that the way in which econometric models are developed and used remain aligned with their underpinning economic and engineering bases. Given the sensitivity of econometric model outputs, there is a risk that a spurious or inappropriate view of efficiency is revealed if the underpinning economic and engineering assumptions are poor or not valid. Based on the information provided in chapter 6 and appendix 9 of the PR24 draft methodology, we make four fundamental observations on the approach to econometric modelling that has been set out:

- **Ensuring model residuals are randomly distributed and appropriately account for observed changes in costs across the data panel** – It is sensible to use the PR19 models as a starting position. However, when reviewing the models (updated to 2020/21), patterns in the model residuals can be seen. This is particularly noticeable for water distribution and water wholesale models. This is a major concern for the validity of random effects models where errors should be independently and identically distributed (IID). There are a range of potential remedies that should be considered.
- **Developing richer models where engineering or economic logic supports rather than over emphasis on model parsimony** – There is a risk of 'overfitting' models particularly when using relatively small data series. This was a key driver for developing 'sensibly simple' models at PR19. However, this approach is likely driving a counteracting risk of omitted variables. With an increasing data panel length, there is a case to develop richer models where engineering or economic logic supports the inclusion of additional or more complex explanatory drivers. We consider that this will provide more protection against inaccurate efficiency benchmarks compared to triangulating multiple simplistic models that do not adequately account for fundamental cost drivers.
- **Setting out a clear basis for the efficiency challenges used, particularly if using forecast costs and drivers when constructing models** – The calculation and use of efficiency challenges should be driven by the quality of the models being used and reflect true efficiency as opposed to underspending and/or poor service delivery. This linkage should be explicitly set out to minimise the risk of inappropriate efficiency challenges being applied. This is particularly relevant if forecast costs and cost drivers are to be included. Inclusion of forecasts will also

increase the level of uncertainty in the model coefficients. Consequently, we think that forecasts should not be used if possible. However, if they are included, they should be acknowledged in efficiency challenge calculations. Consideration should also be given to making sure that companies are not disadvantaged where material variances between forecasts and cost outturns materially skew model outputs.

- **Following a transparent and consultative approach to developing the final PR24 cost models and allowing sufficient opportunity to develop adjustment claims after models are confirmed** – We welcome the opportunity share our thinking on how to improve models ahead of the spring cost modelling consultation. We have identified a series of areas where the predictive power of econometric models can be improved. These are identified below and we plan to provide more substantive evidence to support them when contributing to the cost modelling consultation. We also hope that the outputs of the consultation will allow for a more informed cost adjustment claim consultation in summer 2023.

These observations are set out in more detail below.

Ensuring model residuals are randomly distributed and appropriately account for observed changes in costs across the data panel

It is sensible to use the PR19 models as a starting position for constructing models for use at PR24. However, detailed analysis of the models (updated to 2020/21) suggests that they may suffer from some fundamental weaknesses which are reducing their robustness and ability to make accurate efficiency forecasts. These should be investigated and remedies sought.

The Random Effects estimator can be expressed as: $y_{it} = \alpha + X\beta + u_i + \varepsilon_{it}$. For the model to be robust, the unobserved heterogeneity (time-invariant) term, u_i, must be uncorrelated with the independent variables as well as independently and identically distributed (IID). The idiosyncratic (time-variant) error term, ε_{it}, must also be IID. In short, the components of the random errors produced by the random effects model therefore should not be related to one another, and shouldn’t exhibit a pattern, whether that be:

- over time within the data panel;
- against any of the model variables; or
- in relation to another potentially omitted variables.

Where this is not the case, the model is highly likely to have omitted variable bias. This means that the model is placing the effect of a legitimate driver of explainable costs into the error term and showing it as (in)efficiency. Failing to account for omitted variables can also overstate or understate both the significance and the magnitude of other cost drivers in the models. It will also have a significant impact on model quality. If it isn’t accounted for in models the efficiency challenge should then be adjusted accordingly.

Temporal autocorrelation

When reviewing the PR19 models, patterns in the model residuals can be seen. This is particularly noticeable for water distribution and water wholesale models. In the TWD model (see figure below) there is a clear pattern over time, with a downward trend in the errors until 2016 followed by an upward trend for all companies. The distribution also suggests that a structural break may exist in the data. Structural breaks are thresholds within the data series whereby the relationship between costs and cost drivers change. In summary, it is highly likely that the full distribution of costs is not being appropriately explained by the existing explanatory variables in the model.

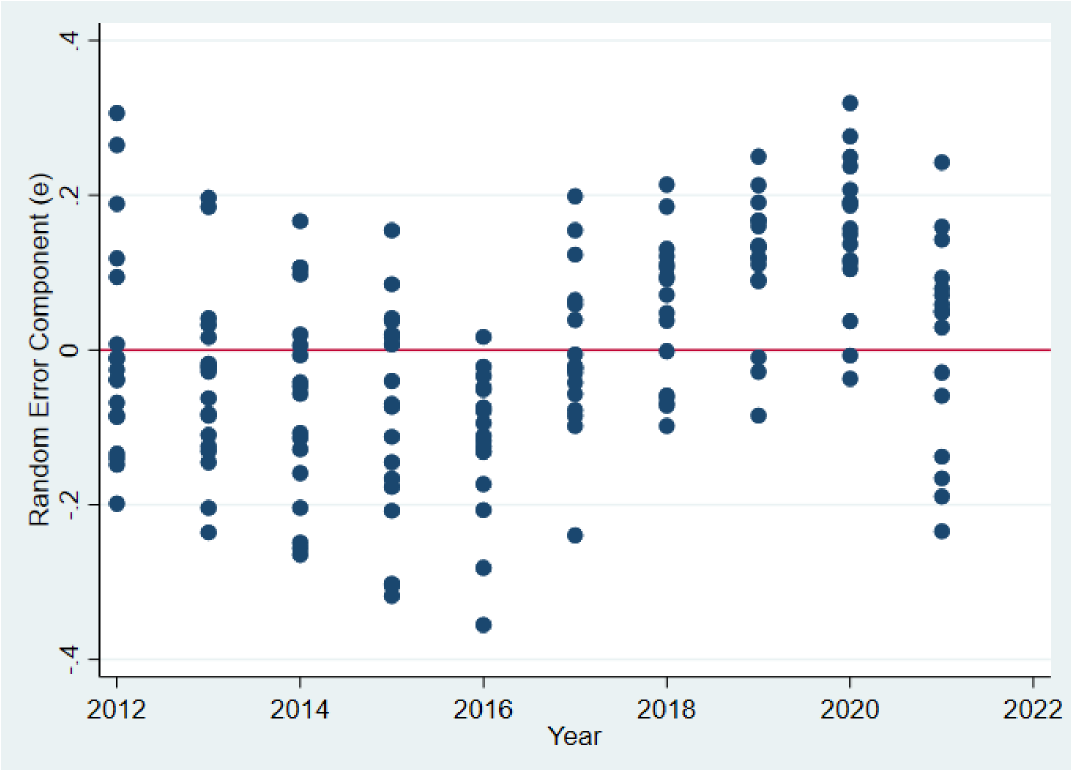

Figure: PR19 TWD model (with extended data panel to 2020/21)

Analysis of expenditure over the full regulatory time series back to 1989, suggests that a fundamental driver for the pattern may relate to the way in which expenditure is incurred across AMP periods (see figure below). Given that the modelling data panel only accounts for two AMPs (and only one AMP – AMP6 – from start to finish), this is hard to see in isolation. However, viewing the longer time series, two patterns become clear:

- The first year of each AMP has the lowest expenditure in the given AMP. This is almost exclusively driven by Capex spend and is likely to relate to the uncertainty associated with confirming the final determination and the associated challenges of procuring and contracting large capital programmes.
- There appears to be varying levels of expenditure incurred between AMPs, this is likely to relate to the specific policies and incentives in place in the relevant time period.

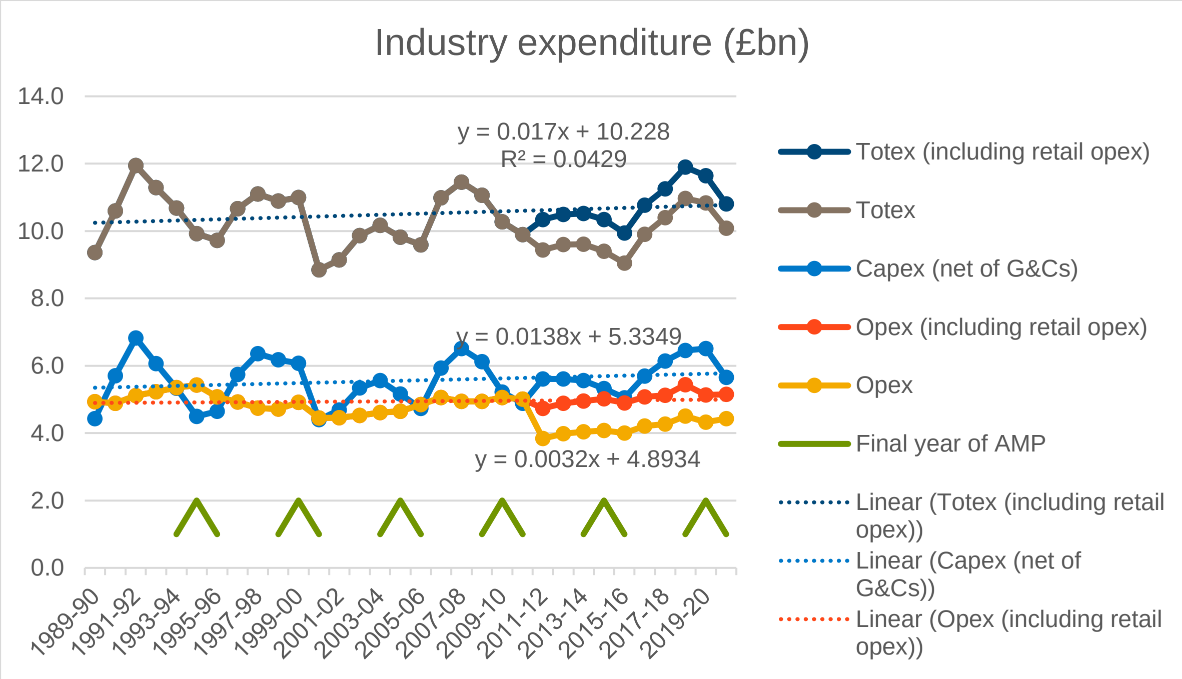

Figure: Timeseries showing the AMP cycle of expenditure (normalised to 2020-21 prices)

Spatial autocorrelation

We have also identified that model errors may show correlation with spatial attributes and other potential omitted variables. Both highlight that the important assumption of errors being independently and identically distributed (IID) in random effects may not hold. For example, regressing model errors with average pumping head and maximum temperature both reveal statistically significant relationships.

We have also been able to improve the performance of models by including regional attributes in models through the use of simple spatial econometrics. These analyses suggest spatial autocorrelation exist in models, most likely driven by regional circumstances not accounted for in the existing explanatory variables. Such spatially related characteristics could be explained by topographical, geological, or socio-economic variation across the country.

Potential remedies for generating models with independent and identically distributed errors

Where errors are shown to not be independent and identically distributed (IID), a remedy should be sought before the model can be considered robust. These could include:

- inclusion of additional explanatory variables (as described elsewhere) or interaction terms where they can be shown to reduce observed error patterns;
- use of ‘dummy’ or ‘trend’ variables to acknowledge the pattern without explicitly identifying the causal driver; or
- pre modelling transformations of the data (such as changing the length of the panel or applying a rolling average) to allow focus on explanatory variables where the causal relationship is clear.

In different ways, each of these remedy types will allow models to account for the exogenous increases in costs which are not accounted for by the independent variables currently in the models.

Each are recognised approaches in empirical literature. For example, there is regulatory precedent with Ofgem using a time trend in RIIO-ED2.

Developing richer models where engineering or economic logic supports rather than over emphasis on model parsimony

There is a risk of 'overfitting' models particularly when using relatively small data series. This was a key driver for developing 'sensibly simple' models at PR19. However, this approach is likely driving a counteracting risk of omitted variables. With an increasing data panel length, there is a strong case to develop richer models where engineering or economic logic supports the inclusion of additional or more complex explanatory drivers. This should help to improve model robustness and address identified vulnerabilities in the current specifications.

Improving the predictive power of models

The R^2values of the PR19 models become uncomfortably low when they are transformed to a unit cost specification. This indicates that the predictive ability of the non-scale explanatory drivers is limited. Given the extended data panel now available, we would like to see additional and potentially more complex variables added to the models to improve explanatory power.

We consider that the inclusion of weather variables should be explored. Temperature and precipitation are both undeniably exogenous, and historical data and forecasts can be produced independently without company influence. There is clear engineering logic and strong empirical evidence of weather-related cost pressures in both water and waste services, e.g. Peak demand pressure driven by extreme temperatures and intense rainfall creating pressure on wastewater networks. Given anticipated changes in weather over time (namely Climate change), this is also a good opportunity to help future proof cost models against growing cost pressures.

There are opportunities to improve the performance of models by including multiple explanatory variables for related cost drivers where they can each be justified from an econometric and statistical perspective (provided multicollinearity risks are controlled). A clear example would be the inclusion of several network complexity explanatory drivers in treated water distribution models to account for different aspects of complexity that are missed if only one variable is included. Reviewing the way in which network costs are incurred in practice supports the inclusion of average pumping head (APH) alongside the existing booster pumping stations per length metric. The former has a strong relationship with opex costs incurred, whereas the later provides a better description for capital maintenance cost pressures. Model runs also show that the statistical performance reinforces the engineering rationale to include both variables in the same model.

It is also important to test and challenge existing model drivers to give confidence that they are operating in line with the underpinning engineering and economic logic. This should include the way in which the current population density drivers interact and relate to economies of scale drivers and measures of network and treatment complexity. The specification of the existing water and waste treatment complexity drivers should also be reviewed – the existing thresholds may not accurately reflect the true distribution of treatment costs incurred across differing treatment processes. For example, differences in groundwater and surface water treatment complexity cost pressures and in wastewater where complying with tight P consents are not currently appropriately allowed for.

Improving the quality of models rather than relying on the triangulation of multiple more simplistic model specifications

We encourage a greater focus on improving the quality of individual models. All models should be tested to demonstrate that they are appropriately specified and therefore have high predictive capability. This is preferable to relying on the triangulation of multiple more simplistic models that will likely have a lower predictive power.

For example, triangulating two models with separate measures of network complexity (e.g. one treated water distribution model with average pumping head and a second with boosters per length) will not have the same effect of having a model that accounts for both drivers together (tests also show that multicollinearity is not a problem when both variables are included in the same model). The separate models will fail to capture the interactive costs associated with the other complexity variable. This will leave an element of network complexity in the error term by not allowing for the costs driven from having *both* high APH and boosters per length cost drivers present.

Adapting the granularity of models to better reflect cost drivers

Developing wastewater network plus models should be a priority to give a 'higher-level' view of costs following the proposed approach to bioresources and likely removal of the bioresources plus models. We note, however, that developing a sensible and robust model that captures wastewater network plus may prove to be difficult because existing variables act in opposing directions in sewage collection and sewage treatment, e.g. higher density is likely to be beneficial for treatment but more costly for collection. We therefore welcome Ofwat's invitation for companies to submit potential cost models in their modelling consultation.

We agree that the low reliability and poor performance of the 'retail bad debt' and 'other' cost models may justify focusing on just total cost models. However, we encourage the consideration of more advanced models to improve predictive performance and an attempt to produce a richer and broader total cost model suite should follow.

Setting out a clear basis for the efficiency challenges used, particularly if using forecast costs and drivers when constructing models

Setting an appropriate catch-up efficiency challenge

As described above, when using a random effects estimator it is standard practice to assume that the idiosyncratic component of the random error (ε_{it}) equates to (in)efficiency. For this to be accurate this assumes that the time invariant error component (u_i) is inconsequential as the material drivers of variance between observations have been described by all the explanatory variables included in the model, or the time-invariant error component is very small. If efficiency could be fully isolated by models, catch-up challenges could be set at a frontier level. However, models will never be perfectly specified and consequently the variance between model and observation is not a perfect proxy for efficiency. This means that efficiency challenges need to be moderated away from the frontier in proportion to the predictive performance of the models used. We consider that this is a key component in setting efficient cost allowances that are fair.

Given that the PR24 models will use the same estimator and their specification is likely to evolve from those used PR19, we assume that model quality is likely to approximate to that at PR19. Consequently, we consider that the starting point for the catch-up efficiency challenge should be upper quartile. This conforms with what was assumed at: PR14, PR19 draft determination, and the CMA's PR19 judgements (where a tightening of the efficiency challenge to the 4th company was rejected as an improvement in predictive capability could not be demonstrated). We believe that a commitment to this effect should be outlined in the final methodology before rather than after models are finalised.

We think that companies that are not delivering appropriate levels of core service should not contribute to the setting of the catch-up efficiency challenge. Ofwat's PR19 efficiency challenge methodology did not account for this potential ambiguity, and the PR24 draft methodology also does not set a clear expectation. To protect customers and companies that are delivering appropriate levels of service to customers, a clear distinction between efficiency and underspend should be set out.

Setting an appropriate frontier shift efficiency challenge and productivity assumptions

The setting of a frontier shift challenge should be based on macroeconomic analysis rather than unsupported judgement. Setting an appropriate frontier shift challenge using the current method (assessment using total factor productivity, TFP) is complicated by the fact that it cannot be accurately isolated from catch-up productivity improvements. This leads to an element of double counting between Ofwat's two efficiency challenges if TFP is assumed to equate to frontier shift.

Ofwat assert that that productivity in the water sector has stagnated since 2011. However, its analysis doesn't account for quality. Analysis from the WaterUK report by Frontier Economics (2017) suggests that the water industry had been more productive than the identified EU KLEMS comparator sectors once quality is accounted for.

Setting stretching efficiency challenges is further confused by the expectation to deliver performance improvements within base cost allowances. Improving performance from ongoing 'base' costs is itself a productivity improvement that can lead to reduced future costs or improved performance. It might be fair to expect one of these productivity improvements whilst delivering base expenditure. However, an expectation to deliver on both while also setting stretching catch-up and frontier shift challenges does not feel like an appropriate balance. This could have potential negative long-term impacts for both companies and its customers as unsustainable trade-offs could drive greater levels of risk, fuel short-termism and therefore erode future productivity improvements.

Cross sector productivity comparisons are challenging. Aside from the issues of definition that we have identified above (which are likely to increase the perceived strength of the frontier shift applied), a current study by Economic Insight (2022) indicates only two primarily technology driven sectors achieved the level of productivity that Ofwat set at PR19 since the financial crisis (1.1% pa).

It should also be noted that post-financial crisis productivity has stagnated for almost all industries in the UK. The Economic Insight report updates the TFP values to 2018 revealing declining productivity across the comparator sectors. The effects of subsequent major macroeconomic events such as Brexit, the Russia-Ukraine war, Covid and the imminent stagflation will exacerbate this further. Therefore, the PR19 frontier shift assumptions do not feel like they are at an appropriate level to set and challenging but achievable productivity target.

Using forecasts in the base models

In its draft methodology, Ofwat reiterates its intention to 'include more of a forward look in [its] cost models'. To do this it is considering using forecast costs in its base cost models or using these forecast costs to set a forward-looking efficiency challenge.

Using forecast costs and independent variables when constructing models will necessarily reduce predictive capability of the resultant models. This is because these values have not been realised and the inferred relationships between costs and drivers may not materialise. For this reason, we do not support the use of forecasts when developing models.

However, if forecasts are to be used in the models, both the forward- and backward-looking efficiency challenges should be set at a less stringent level. Where a forward-looking efficiency challenge is considered but the forecasts are not included in the models, the forward-looking challenge should be set at a less stretching level than a backwards looking challenge.

There also needs to be consideration of how to manage changes in cost efficiency benchmarks if they become materially impacted by overly stretching business plan forecasts of other companies which subsequently cannot be realised. As a principle, companies that submit realistic and well-evidenced business plans shouldn't be adversely impacted if the forecasts of others are found to be unsustainable. To manage this risk, it might be appropriate to encourage companies to set out their efficiency assumptions in plans so that pre- and post-efficiency views of expenditure can be considered. This would provide insight into the level of stretch assumed and allow efficiency to be assessed on a consistent basis.

Following a transparent and consultative approach to developing the final PR24 cost models and allowing sufficient opportunity to develop adjustment claims after models are confirmed

We welcome the opportunity to consult on and improve existing models and have identified several key risks inherent with the current (PR19) PR19 model specifications. These identify weaknesses with the underlying logic and statistical performance of the models. We will be presenting supporting evidence in our model submission ahead of the spring 2023 cost modelling consultation and remain willing to discuss and share our emerging findings in advance if required.

Identified Risk	Implication	Potential remedies being considered
Model errors are not Independently and Identically Distributed (IID)	**General model misspecification** Significant reduction in validity of model outputs	Manage IID risks though: • Inclusion of additional explanatory variables (including interaction variables and/or spatially weighted variables) • Use of 'dummy' or 'trend' variables; or • Pre modelling transformations (such as smoothing of cost data to account for AMP cycle)
Model outputs don't account for major differences in performance and cost allocation policies	**Poorly specified efficiency challenge** Efficiency challenges are poorly specified and incentivise cost reductions rather than efficiency (i.e. productivity)	Ensure that model costs conform with minimum service performance expectations and cost allocation requirements.

Weather variables not included despite being a major exogenous driver of cost	**Major cost driver omitted** No account for cost pressures incurred during peak demand or high rainfall intensity	Consider the use of weather metrics as exogenous explanatory variables. For example, maximum temperature as a driver of water costs and rainfall intensity as a driver of wastewater costs.
Population density drivers don't perform in line with engineering expectation	**Major cost driver partially accounted for** Models don't adequately account for sparsity cost drivers	Remove 'weighted average population density squared' driver where more focused variables are available that better describe direct cost pressures. • In Treated Water Distribution models, allow density to account for increasing costs in urban areas. Cost drivers in rural areas are better accounted for by network complexity drivers • In Water Resources Plus models, population density creates opportunities for economies of scale at treatment works. But this is complicated by fundamental differences in the size of surface and ground water assets. It may be more appropriate to account for directly rather than using the density proxy.
Water treatment complexity drivers don't differentiate observed costs at ground water and surface water assets	**Major cost driver partially accounted for** Assumes that treatment complexity does not materially drive costs at surface water treatment sites	Reconsider existing water treatment complexity drivers with more explicit consideration of groundwater and surface water cost drivers. Drivers need to account for clear differences in cost at ground and surface water works and separately account for economies of scale of assets of very different sizes.
Water network complexity drivers are too broad	**Major cost driver partially accounted for** Costs incurred in running and maintaining pumping assets in the network are poorly accounted for through existing driver	Include both APH and boosters per length in TWD model specifications to account for different drivers of water network complexity
Wastewater treatment complexity drivers are too broad	**Major cost driver partially accounted for** Increased costs for conforming to very tight P consents are poorly accounted for through existing Ammonia driver	Explore ways of accounting for wider wastewater complexity divers for example including drivers to account for costs incurred in operating at very tight P consents.