Dŵr Cymru
Welsh Water

Linea
Fortran Road
St Mellons
Cardiff
CF3 0LT

Linea
Heol Fortran
Llaneirwg
Caerdydd
CF3 0LT

02920 740450
dwrcymru.com

12 January 2023

Ofwat 21

Bloomsbury St Bloomsbury

London

WC1B 3HF

By email only: CostAssessment@ofwat.gov.uk

Template and guidance for the submission of base econometric cost models ahead of the spring 2023 consultation

Dear Ofwat

On behalf of Dŵr Cymru I am writing in response to Ofwat's invitation to submit base cost econometric models in connection with PR24.

Overview

We welcome the consultation on the econometric cost models being developed at PR24 which builds on the robust PR19 process. Ofwat has reiterated its confidence in the final suite of models developed for PR19 and stated that there is a "high bar" to make changes. We note that some of the PR19 models are significantly weaker than they were, as a result of which we think some of them should be discarded from the modelling suite. Within our submission we propose three new models in the area of wastewater, all of which we think perform markedly better than the PR19 models and should be included in the modelling suite.

When using the models for determining cost allowances, the robustness of the models should be considered within the framework for cost assessment. Where the models are weak and have a large range of residuals upper quartile benchmarks may not be appropriate and consideration should be given to using less stringent benchmarks such as the 60th or 66th percentile to reflect the degree of uncertainty in the modelling.

The final methodology proposes different arrangement for England and Wales with regards to the inclusion water site-specific developer services in the price control. This has implications for the modelling of botex and the resulting cost allowances. We would welcome engagement with Ofwat on how the site-specific developer services costs for companies in Wales will be assessed at PR24.

Submission

The PR24 econometric cost modelling consultation comprises of the following documents:

1. This document outlining our modelling submission;
2. “Econometric Cost Model Submission Jan 2023 Stata do WSH.do”- The Stata do file used to run the econometric models. The do file is based on Ofwat’s published do file and has been edited for our submission.
3. “Econometric Consultation WSH Results.xlsx”- This file provides an output of the Stata do file to calculate the distribution of efficiency scores.

Dataset

The models presented use Ofwat’s v3 of the Cost Assessment Master Datasets published in November 2022 on Ofwat’s website.

Ofwat’s dataset and supporting Stata files adjusts for principal use, bioresources backcasting and exceptional items. The data quality of these adjustments, in particular the bioresources backcasting should be taken into account in the overall cost assessment framework. We are still reviewing and considering the application of these adjustments.

Wholesale Water

In line with Ofwat’s guidance, we began our analysis by reviewing the updated performance of the final PR19 suite of models, and considered where there may be a strong case either for adding a new model or abandoning an existing model.

Water Resources and Treatment

At PR19 Ofwat used two models, WRP1 and WRP2, the difference between the two being alternative variables for capturing the complexity of water treatment.

We observe that, in both cases, the models have weakened in several respects, notably in relation to the significance of coefficients and the range of model residuals. However, our investigations have not produced obviously superior alternatives. Consequently, we are not proposing any new models for Ofwat’s consultation at this time.

In the absence of clearly superior models, it may be appropriate for Ofwat to persist with WRP1 and WRP2. However, the poorer model fit and the wide range of residuals suggests that the results of the models should be treated with caution. In particular, we do not think that there is sufficient confidence in the two models to warrant the use of an upper quartile benchmark for setting cost allowances going forwards and a less stringent benchmark should be used such as the 60^{th} or 66^{th} percentile.

Treated Water Distribution

At PR19 Ofwat used one model, TWD1.

We have re-estimated this model using the up-to-date dataset. It continues to perform well.

We also examined alternative specifications using Average Pumping Head in place of Booster Pumping Stations per Length of Main. Whilst this produced reasonable models, they were not obviously superior to the PR19 TWD1. In addition, there remain concerns that the industry data on Average Pumping Head (APH) does not meet Ofwat's stated principle that "Data Used is Good Quality". Ofwat's APH review undertaken by Turner and Townsend and WRC[1] reports that there are varying degrees of measurement and estimation in the calculation of APH, with a number of companies estimating all lift values. The report finds that estimated data used in the calculation of APH may be overestimating pumping head for some companies. At this time, we do not believe that the data is of sufficient quality across the industry to be utilised in the modelling.

Wholesale Water

At PR19 Ofwat used two models, WW1 and WW2. We have re-estimated both models using the up-to-date dataset. We observe that, in both cases, the models have weakened slightly. However, our investigations have not produced obviously superior alternatives. Consequently, we are not proposing any new models for Ofwat's consultation.

We therefore recommend that Ofwat persists with WW1 and WW2 for now. However, the somewhat poorer model fit and the wider range of residuals suggests that the results of the models should be treated with caution. In particular, we do not think that there is sufficient confidence in the two models to warrant the use of an upper quartile benchmark for setting cost allowances going forwards: a 60th or 66th percentile might be more appropriate.

Wholesale Wastewater

In line with Ofwat's guidance, we began our analysis by reviewing the updated performance of the final PR19 suite of models, and considered where there may be a strong case either for adding a new model or abandoning an existing model. We are putting forward a one new model for sewage collection, one for sewage treatment, and one for wastewater network-plus.

[1] https://www.ofwat.gov.uk/wp-content/uploads/2022/05/Average-Pumping-Head-Data-Quality-Improvement-Final-Report-.pdf

Sewage Collection

At PR19 Ofwat used two models, SWC1 and SWC2. The CMA refined SWC2 by adding a squared term for weighted average density, a change which we support. We have not put this forward as a "new model" on the basis that it should already be amongst the pool of models under consideration by Ofwat.

We have re-estimated both models using the up-to-date dataset. Although SWC2 is marginally weaker (we found that the sensitivity of coefficients to removal of companies was not entirely robust) we think it should remain in consideration by Ofwat for PR24.

We are, however, proposing an alternative model to SWC1. The SWC1 model uses properties per length but does not control for the "U-shaped" density impact like the updated SWC2 model.

We are putting forward a new model, WSH_SWC1, which replaces sewer length with connected properties as a scale variable and accounts for density by using the same weighted average density variables as the post-CMA SWC2. The alternative scale driver produces a better model fit and improves the performance of the density variable. Details are presented below using the prescribed template.

Econometric model formula

WSH_SWC1: $\ln(\text{SWC BOTEX+}) = \alpha + \beta_1 \ln(\text{Connected properties}_{it}) + \beta_2 \ln(\text{pumping capacity per sewer length}_{it}) + \beta_3 \ln(\text{weighted average density}_{it}) + \beta_4 \ln(\text{weighted average density}_{it})^2 + \varepsilon_{it}$

Description of the dependent variable

The dependent variable is defined as per Ofwat's consultation analysis files i.e. the sum of:

- Power
- Income treated as negative expenditure
- Service charges / Discharge
- Bulk Discharge
- Renewals expensed in year (infrastructure)
- Renewals expensed in year (non-infrastructure)
- Other operating expenditure excluding renewals
- Maintaining the long-term capability of assets (infrastructure)
- Maintaining the long-term capability of assets (non-infrastructure)
- Transfer of private sewers and pumping stations
- Atypical expenditure

- Reducing flood risk for properties (OPEX and CAPEX)
- Network reinforcement (OPEX and CAPEX)

It excludes the following cost categories:

- Costs associated with the Traffic Management Act
- Industrial Emissions Directorate
- NRSWA diversions (non-S185)
- Other non-S185 diversions
- Developer services base cost adjustment
- Backcasting adjustment (between bioresources and sewage treatment).

This is consistent with Ofwat's PR24 methodology.

Description of the explanatory variables

- Connected properties (BN1178).
- Pumping capacity per length of sewer (S4029 divided by sewer length)
- Weighted average density LAD (code: WAD_LAD), as reported in the published wholesale dataset

Brief comment on the model

The model uses the full dataset for sewerage expenditures, i.e 11 years. It estimates cost as a function of the number of properties served, the degree of pumping station capacity on the network, and population density/sparsity modelled as a U-shaped function.

The model performs better than SWC1 from PR19: model fit is better, the coefficients are all significant at the 1% level, the statistical and sensitivity tests are passed, and the range of residuals is plausible.

	WSH_SWC1
Dependent variable	BOTEX+ (SWC)
Connected properties (log)	0.792*** (0)
Weighted average density (log)	-1.848*** (0.00453)
Weighted average density (log), squared	0.129*** (0.00286)

Pumping capacity per sewer length (log)	0.400*** (0.000131)
Constant	-0.334 (0.896)
Estimation method (OLS or RE)	RE
N (sample size)	110
R2 adjusted	0.924
RESET test	0.418
VIF (max)	403.9
Pooling / Chow test	0.807
Normality of model residuals	0.0559
Heteroskedasticity of model residuals	0.256
Test of pooled OLS versus Random Effects (LM test)	2.18e-05
Efficiency score distribution (min and max)	92% to 111%
Sensitivity of estimated coefficients to removal of most and least efficient company	G
Sensitivity of estimated coefficients to removal of first and last year of the sample	G

Sewage Treatment

At PR19 Ofwat used two models, SWT1 and SWT2. We have re-estimated both models using the up-to-date dataset. We observe that, in both cases, the models have weakened. This is especially the case for SWT1 where the variable “load treated in size bands 1-3 (%)” is no longer significant. This variable captures the economies of scale benefits from being able to operate large sewage treatment works in urban areas instead of having to operate numerous small remote treatment works.

We are therefore putting forward what we consider to be a better alternative model, namely WSH_SWT1. As the percentage of load treated in bands 1-3 is no longer significant, economies of scale from the ability to operate large sewage treatment works are controlled for through the weighted average density variable. We present our proposed new model in the template below.

Econometric model formula

WSH_SWT1: $\ln(\text{SWT BOTEX+}) = \alpha + \beta 1 \ln(\text{Load}_{it})) + \beta 2 \ln(\text{weighted average density}_{it}) + \beta 3$ (% load with ammonia below 3 mg/l$_{it}$) + ε_{it}

Description of the dependent variable

The dependent variable is defined as per Ofwat's consultation analysis files i.e. the sum of:

- Power
- Income treated as negative expenditure
- Service charges / Discharge
- Bulk Discharge
- Renewals expensed in year (infrastructure)
- Renewals expensed in year (non-infrastructure)
- Other operating expenditure excluding renewals
- Maintaining the long-term capability of assets (infrastructure)
- Maintaining the long-term capability of assets (non-infrastructure)
- Transfer of private sewers and pumping stations
- Atypical expenditure
- Reducing flood risk for properties (OPEX and CAPEX)

It excludes the following cost categories:

- Costs associated with the Traffic Management Act
- Industrial Emissions Directorate
- NRSWA diversions (non-S185)
- Other non-S185 diversions
- Developer services base cost adjustment
- Backcasting adjustment (between bioresources and sewage treatment).

This is consistent with Ofwat's PR24 methodology.

Description of the explanatory variables

- Load (STWDP125_21)
- Weighted average density LAD (code: WAD_LAD), as reported in the published wholesale dataset
- % load with ammonia below 3mg/l (sum of STWDA121 + STWDA122_21 divided by load)

Brief comment on the model

The model uses the full dataset for sewage treatment expenditures, i.e 11 years. It estimates cost as a function of load, population density, and the extent to which treatment works have to meet an ammonia standard higher than 3mg/l.

The model is a development of SWT1 from PR19. The coefficient on the load in bands 1-3 is insignificant in our re-estimation of the PR19 models, even at the 10% level. To ensure economies of scale are appropriate captured, weighted average density has been included in the model. The variable has the expected sign and is significant. The model performs better than SWT1 at PR19. The range of residuals is quite wide, though slightly narrower than the range for the re-estimated PR19 SWT1. We address this in the context of our proposed network plus model below.

	WSH_SWT1
Dependent variable	BOTEX+ (SWT)
Load (log)	0.646*** (0)
Weighted average density (log)	-0.202*** (0.00177)
% load with ammonia below 3mg/l	0.00692*** (0)
Constant	-2.102*** (6.77e-05)
Estimation method (OLS or RE)	RE
N (sample size)	110
R2 adjusted	0.879
RESET test	0.293
VIF (max)	7.172
Pooling / Chow test	0.999
Normality of model residuals	0.00453
Heteroskedasticity of model residuals	0.322
Test of pooled OLS versus Random Effects (LM test)	0
Efficiency score distribution (min and max)	91% to 147%
Sensitivity of estimated coefficients to removal of most and least efficient company	G

Sensitivity of estimated coefficients to removal of first and last year of the sample	G

Wastewater Network Plus

At PR19 Ofwat did not use a wastewater plus model. Modelling at different levels of aggregation allows for cost relationships between activities to be considered and it also mitigates any impact of cost allocations. Therefore, we think there is a case for an overall network plus model at least as a cross-check on the bottom-up models to ensure that infeasible benchmarks are not inadvertently being created. We have put forward what we consider to be a strong potential wastewater network plus model, WSH_NPWW1. We present the proposed model in the template below.

Econometric model formula

WSH_NPWW1: $\ln(\text{NPWW BOTEX+}) = \alpha + \beta 1 \ln(\text{Connected properties}_{it})) + \beta 2 \ln(\text{weighted average density}_{it}) + \beta 3 \ln(\text{pumping capacity per sewer length}_{it}) + \beta 4 (\%\text{ load with ammonia below 3 mg/l}_{it}) + \beta 5 \ln(\text{load per property}_{it}) + \varepsilon_{it}$

Description of the dependent variable

The dependent variable is defined as per Ofwat's consultation analysis files i.e. the sum of:

- Power
- Income treated as negative expenditure
- Service charges / Discharge
- Bulk Discharge
- Renewals expensed in year (infrastructure)
- Renewals expensed in year (non-infrastructure)
- Other operating expenditure excluding renewals
- Maintaining the long-term capability of assets (infrastructure)
- Maintaining the long-term capability of assets (non-infrastructure)
- Transfer of private sewers and pumping stations
- Atypical expenditure
- Reducing flood risk for properties (OPEX and CAPEX)
- Network reinforcement (OPEX and CAPEX)

It excludes the following cost categories:

- Costs associated with the Traffic Management Act

- Industrial Emissions Directorate
- NRSWA diversions (non-S185)
- Other non-S185 diversions
- Developer services base cost adjustment
- Backcasting adjustment (between bioresources and sewage treatment).

This is consistent with Ofwat's PR24 methodology.

Description of the explanatory variables

- Connected properties (BN1178).
- Pumping capacity per lengths of sewer (S4029 divided by sewer length)
- Weighted average density LAD (code: WAD_LAD), as reported in the published wholesale dataset
- Load per property (load (STWDP125_21)divided by connected properties BN1178))
- % load with ammonia below 3mg/l (sum of STWDA121 + STWDA122_21 divided by load)

Brief comment on the model

The model uses the following explanatory variables:

- connected properties as the scale variable;
- "load per property" to reflect the fact that companies differ in the extent to which they have large non-household dischargers, whether of sewage or trade effluent;
- the same weighted average density measure which featured in a number of Ofwat's PR19 models;
- pumping capacity per length of sewer, as in Ofwat's PR19 SWC models; and
- load treated at works with ammonia consents tighter than 3mg/l, again as used by Ofwat at PR19 and in our proposed WSH_SWT1 described above.

Together these variables capture the principal cost drivers for the wastewater network plus business.

The model uses the full dataset for wastewater network-plus expenditures, i.e 11 years. It generally performs well: the model fit is very good and the range of residuals is

acceptable. All of the coefficients are of the expected sign and magnitude and are significant at the 1% level except for “load per property” which is significant at the 5% level. The statistical tests are passed, though the “load per property” variable becomes less significant when the sensitivity tests are carried out. Note that we did investigate using the square of the weighted average density variable, as in the Ofwat/CMA SWC2 and our proposed WSH_SWC1, but the coefficient was insignificant.

	WSH_NPWW1
Dependent variable	BOTEX+ (NPWW)
Connected properties (log)	0.726*** (0)
Weighted average density (log)	-0.142*** (1.98e-08)
Pumping capacity per sewer length (log)	0.354*** (1.86e-05)
Load per property (log)	0.423** (0.0395)
% load with ammonia below 3mg/l	0.00575*** (0)
Constant	-3.550*** (2.53e-06)
Estimation method (OLS or RE)	RE
N (sample size)	110
R2 adjusted	0.957
RESET test	0.784
VIF (max)	7.470
Pooling / Chow test	0.982
Normality of model residuals	0.0689
Heteroskedasticity of model residuals	0.734
Test of pooled OLS versus Random Effects (LM test)	6.20e-09
Efficiency score distribution (min and max)	96% to 107%
Sensitivity of estimated coefficients to removal of most and least efficient company	A [load per property becomes less significant]
Sensitivity of estimated coefficients to removal of first and last year of the sample	A [load per property becomes less significant]

Bioresources and Bioresources Plus

We note Ofwat's aspiration to set separate efficiency factors for bioresources price controls. We have re-estimated Ofwat's two PR19 models using the updated dataset, but note that the models continue to perform poorly relative to other forms of aggregation. Further, we have been unable to identify any better bioresources-only models. It may be that this reflects genuine substitutability between costs incurred in sewage treatment and bioresources in the engineering choices that companies make. However, we intend to continue to examine this issue in the coming weeks, in advance of Ofwat's spring consultation.

By contrast, we observe that Ofwat's PR19 Bioresources Plus models continue to perform rather better than the Bioresources-only ones. We are not proposing any new Bioresources Plus models at this time.

Household Retail

At PR19 Ofwat used four bottom-up models and three top-down models for household retail.

We have re-estimated the models using the updated dataset. We observe that, in general, all of the models are weaker than at PR19. In particular the range of residuals produced is wide, suggesting that the models are not capturing all the main drivers of household retail cost. At this time we have not been able to put forward any models that perform materially better than the PR19 suite. At a minimum, this suggests that the results of the models should be treated with caution. In particular, we do not think that there is sufficient confidence in them to warrant the use of an upper quartile benchmark for setting cost allowances going forwards: a 60^{th} or 66^{th} percentile might be more appropriate.

We have a particular concern with Ofwat's RTC1 and, by extension, ROC1. Unlike RTC2, RTC3, and ROC2, there is no term allowing for economies of scale (i.e. the number of household connections variable), so the unit cost formulation imposes a constant returns to scale requirement. However, as part of our modelling for this exercise we have looked at models which take absolute cost as the dependent variable and have consistently observed that economies of scale are implied by the (highly significant) coefficients on the scale variables in each case. We therefore recommend that Ofwat drops ROC1 and RTC1 from its spring consultation.

Finally, we have given careful thought to the treatment of the abnormally high costs incurred in the period 2019 – 2022 associated with the Covid pandemic. We have identified that the use of year dummy variables for the "Covid years" significantly improves the performance of the models. However, this raises the question of what values are used for the year dummies when extrapolating forwards. It would be incorrect to apply a value of zero for the dummy variables because this would be tantamount to assigning a probability of zero to the possibility of a new pandemic (or a resurgence of the last one). It is also inappropriate to set a value of zero as the use of a dummy variable for Covid years will

capture other cost differences in those years that are not as a result of Covid. In our view the first best option is to model retail costs without year dummies (but using the smoothed data because this is more realistic, having the advantage of hindsight). This is why we have not proposed any models using year dummies. If, however, it becomes necessary to use such variables because it is not possible to estimate reasonable cost functions without them, non-zero values should be assigned to them for the purposes of extrapolating forwards. Otherwise companies would be bearing all the asymmetric risk associated with the possibility of a new pandemic in AMP8, which would warrant an adjustment to the allowed cost of capital at PR24.

If you would like to discuss the submission further, please do not hesitate to contact me.

Kind Regards

Charlotte Beale

Head of Economics