Anglian Water

8C. VIVID ECONOMICS: LOG MODEL PREDICTION ERROR

Draft Determination Representation, August 2019

RESPONSIBLE BUSINESS OF THE YEAR 2017
BUSINESS IN THE COMMUNITY

THE QUEEN'S AWARDS FOR ENTERPRISE: SUSTAINABLE DEVELOPMENT 2015

love every drop
anglianwater

Adjusting log-log predictions for IAP enhancement models

Key points

The absence of an adjustment to log-log model predictions at IAP is theoretically invalid and has material consequences for allowances, but could be addressed at draft determinations.

- statistical theory shows that log-log models systematically underestimate cost unless allowances are suitably adjusted;
- for the IAP enhancement models, the issue is highly material, potentially affecting industry allowances by more than £450m;
- there is regulatory precedent for making such adjustments to model predictions at PR14 and RIIO-1;
- the 'conditional mean' or 'smearing' adjustments are in general less risky than 'alpha factors', which has impossibly small or large values when models are misspecified;
- the range of alpha factors observed across the IAP enhancement models shows that it is unsuitable in this context and provides further evidence of misspecification, particularly for wastewater growth.

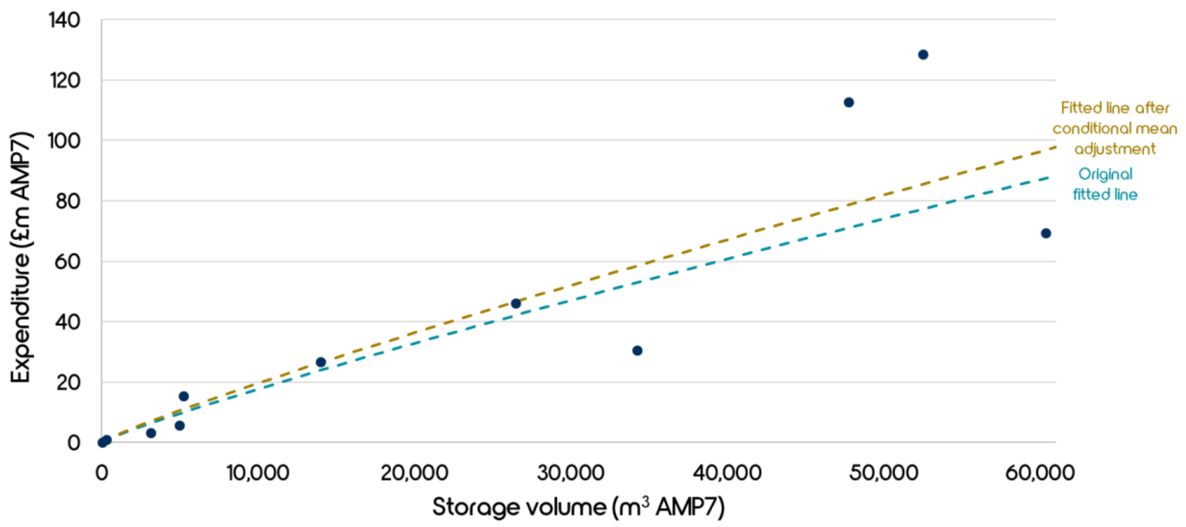
Overview of findings

Statistical theory states that an adjustment factor is required to obtain predictions of costs from log-log models and this factor should be greater than one.. Modelled estimates of *logged* costs cannot simply be transformed by the exponential function to arrive at an estimate of costs, as in logged models the error term, expressed in unlogged terms, has a skew distribution with a positive expected value.[1]

None of the IAP make an appropriate adjustment to account for this fact, meaning that costs are systematically underestimated. Of the IAP models, only the flow to full treatment model accounts for this fact in any way, and does so in a theoretically incorrect manner. *All other logged models underestimate industry costs*. As Figure 1 illustrates, the extent of such an underestimate can be significant and will be larger for companies with larger programmes.

[1] See page 212 of Introductory Econometrics: A Modern Approach, Wooldridge 2012 – link

Figure 1 A failure to adjust estimated costs in the logged storm tanks model 1 materially underestimates of cost



Note: Results shown are for storm tanks model 1, which regresses log capex on log storage volume
Adjustment shown is based on the conditional mean estimator, and accounts for the lognormal distribution of model predictions, in the case when errors are normally distributed

Source: Enhancement cost assessment modelling for the PR19 Initial Assessment of Plans, Vivid Economics

There are 3 common adjustments to log model predictions presented in the literature[2][3]**:**

1. **'Conditional mean':** is biased but consistent (converges to the true value as sample size grows), when the error term is normally distributed. The latter assumption is required in diagnostic tests used by Ofwat as well as the maximum likelihood definition of the OLS estimator, though it may be false in practice. The adjustment factor always takes values greater than one, consistent with theory.[4]

2. **'Smearing':** is biased but consistent when the error term is independent of the explanatory factors. The latter is a weaker assumption than required for the consistency of the conditional mean factor and is required to motivate the use of OLS estimators across the IAP models. The adjustment factor is always greater than one.[5]

3. **'Alpha factor':** is an alternative to the 'smearing' factor, which is also consistent but biased when the error term and explanatory factors are independent. Unlike the other adjustment factors – and inconsistent with statistical theory – it can produce values less than or much larger than one when the required assumptions fail.[6] The fact that it relies on a common assumption to the IAP models means the alpha factor can be used as a diagnostic test for model misspecification: for instance, the IAP wastewater growth 'forecast' data model has an alpha factor of 0.72.

Both the 'alpha factor' and the 'smearing' factor have regulatory precedent in England & Wales. At PR14, Ofwat selected the 'alpha factor' approach from the four approaches presented by CEPA (three approaches from above, and 'no adjustment'). In base cost assessment, values ranged from 0.99 – 1.02 in water and 1.00

[2] Naming convention follows Annex 7, page 108 of Cost Assessment – Advanced Econometric Models, 2014 – link

[3] See pages 212 – 214 of Introductory Econometrics: A Modern Approach, Wooldridge 2012 – link

[4] It is given by $exp\left(\hat{\sigma}^2/_2\right)$.

[5] It is given by $\frac{1}{n}\sum_{i=1}^{n} exp(\hat{u}_i)$.

[6] The adjustment factor is given by $\left(\sum_{i=1}^{n}\hat{m}_i{}^2\right)^{-1}\left(\sum_{i=1}^{n}\hat{m}_i y_i\right)$, where $\hat{m}_i = exp(\widehat{log y_i})$. The 'alpha factor' is interpreted as the slope coefficient in a regression with no constant of y_i, actual cost, against $exp\left(\widehat{log y_i}\right)$, the naïve prediction

– 1.02 in wastewater. In enhancement, the factor had a wide range from 0.84 – 2.34. Ofgem used the same approach at ED1, GD1 and DPCR5, although the 'smearing' 'factor was used at the draft determinations for RIIO ED1. While the CMA did not use adjustments in its determination for Bristol Water, this was on the grounds of materiality rather than theoretical validity. Furthermore, in this case the CMA used bottom-up engineering estimates of enhancement cost, so enhancement log model prediction error was not an issue in its assessment of Bristol Water's efficient costs.

The three adjustment approaches all have a highly material impact on industry allowances in the 7 IAP enhancement areas containing log-log specifications. As shown in Table 1, under the 'conditional mean' or 'smearing' approaches, the industry's enhancement allowance is £430 – £470m higher than under 'no adjustment'. This represents around a quarter of the £1,730m modelling shortfall identified in Vivid Economics' review of the IAP enhancement models[7].

Table 2 sets out the materiality impacts of the 3 approaches for Anglian Water.

Alpha factors for some models provide further evidence of misspecification and do not improve unadjusted estimates of costs. Results particularly highlight misspecification of the wastewater growth models – model residuals are large, leading to a sizeable adjustment in the 'conditional mean' and 'smearing' approaches, and an 'alpha factor' less than one.

Table 1 Industry allowance impacts of the 3 log model prediction adjustment approaches

	Conditional mean	Smearing	Alpha factor
Chemicals removal	10.0	5.8	6.2
Flow to full schemes	16.5	2.5	2.5
Metering	4.1	3.6	5.2
Sanitary parameters	54.7	39.2	38.9
Spill frequency	12.5	9.9	-2.4
Storm tanks	38.3	26.0	85.4
Wastewater growth	334.4	347.8	-209.4
Overall	**470.6**	**434.7**	**-73.7**

Table 2 Anglian Water allowance impacts of the 3 log model prediction adjustment approaches

	Conditional mean	Smearing	Alpha factor
Chemicals removal	1.2	0.4	0.7
Flow to full schemes	3.3	0.6	0.3
Metering	0.1	0.1	0.2
Sanitary parameters	4.8	3.4	3.0
Spill frequency	0.3	0.3	-0.1
Storm tanks	8.7	5.8	19.6
Wastewater growth	57.7	57.6	-47.1
Overall	**76.1**	**68.2**	**-23.3**

Note: All values in 2017/18 CPIH £m
Source: Vivid Economics analysis of PR19 Initial Assessment of Plans models

[7] See pages 14 – 15 of Enhancement cost assessment modelling for the PR19 Initial Assessment of Plans, Vivid Economics – link

The 'conditional mean' factor or 'smearing' factor approaches offer more robust ways of obtaining consistent cost estimates from log-log models. Statistical theory and modelling evidence demonstrate the case for making an adjustment to log-log model allowances. The 'alpha factor' approach serves as a useful model diagnostic, for instance highlighting potential misspecification issues in wastewater growth models. However, it is much more volatile than the two other approaches. Moreover, the use of 'alpha factors' below or significantly greater than one is not justifiable – in this case, the 'alpha factor' does not correct for log-log model misspecification, since it just adjusts all cost estimates downwards in parallel, and exacerbates model prediction error.