PR19 econometric model evaluation – suggestion for PR24

Table of Contents

Table of Contents 2
1. Introduction 3
2. Autocorrelation and the water base cost model 3
  2.1. Theoretical background and Wooldridge test 4
  2.2. Application in the case of PR19 base cost models for water 4
3. Direction of causality and theoretical background of PR19 retail cost models 7
  3.1. General about endogeneity 7
  3.2. Analysis of PR19 retail cost models 7
  3.3. Suggestion for alternative modelling approach in retail cost assessment 11
4. Conclusion 14
Appendix 15
References 22

List of tables

Table 1. Distribution of efficiency scores across models 7
Table 2. Evaluation of explanatory variables in the replicated PR19 retail cost model 10
Table 3. Summary of retail cost drivers to be tested 12
Table 4. PR19 water base cost models 15
Table 5. Wooldridge test for autocorrelation in panel data, ... 15
Table 6. Replicated PR19 water models with lagged dependent variable (AR1) 16
Table 7. AR models with average pumping head as alternative to booster pumping station per km of mains 17
Table 8. PR19 retail models replicated with data 2014-2022 18
Table 9. First manual test for reverse causality with 1-3 lags 18
Table 10. Second manual test for reverse causality, 2 lags 19
Table 11. Total retail cost models 20

1. Introduction

Price review is a continuously improved process that delivers fair outcome for water and wastewater customers while determining sufficient funding for the industry to deliver the service. Part of this process is the econometric modelling approach that estimate the efficient expenditure of water and wastewater services. Econometric modelling is highly data dependent. As more data are generated and available, the modelling approach may need to adapt to better reflect the reality. The PR19 econometric models were built during 2018-19 and on the basis of the best data available at the time. As time goes by and more years of data are collected, what was robust and unbiased in 2019 is no longer the case after three years of data are added. This paper aims at recommending the necessary adjustments in modelling technique that may be helpful for the next price review, PR24.

The paper replicates water base cost models and retail cost models using the PR19 model specifications, but with more recent data. Basic diagnostic tests are then applied to analyse and evaluate the virtues of the replicated models. Then the paper suggests techniques and adjustments that are standard practice found in the literature to improve the model quality.

For water base cost models, the tests have found that autocorrelation exists, hence needs addressing. The recommended technique is to add a lagged dependent variable to the model, one of the common measures in correcting for autocorrelation. However, correcting for one problem may cause another one. In this case, adding a lagged dependent variable turns some of the key explanatory variables insignificant. The paper then suggests an alternative variable that would still deliver the same virtue of explaining cost, but perform better econometrically. Similar procedure can be expanded to wastewater models.

For retail cost model, the replicated models show that the PR19 model specifications can no longer hold. Therefore, the suggestion is to rethink the modelling approach for retail.

The rest of the paper is organised as follow: Section 2 discusses autocorrelation problem and its application in water base cost models, Section 3 discusses the endogeneity problem and its relevance in the case of retail cost models, and Section 4 concludes the paper.

2. Autocorrelation and the water base cost model

Autocorrelation, whenever it exists, causes bias in the standard error and renders the model less efficient.

Let us examine a linear model with panel data

$$y_{it} = \alpha + X_{it}\beta_1 + Z_i\beta_2 + \mu_i + \epsilon_{it} \quad (1)$$

Where y_{it} is the dependant variable, X_{it} is a (1xK_1) vector of time-varying covariates, Z_{it} is a (1xK_2) is a vector of time invariant covariates, α, β_1, β_2 are parameters to be estimated, μ_i is the individual level effect, and ϵ_{it} is the idiosyncratic error. We say autocorrelation exists when μ_i is correlated with the X_{it} or the Z_{it}. In that case the standard errors of β_1 or β_2 will be biased and it will affect the forecast efficiency.

Vast body of literature can be found on the problems of and solutions for autocorrelation. This paper only focuses on the impact of autocorrelation in the context of water base cost modelling as part of the price review of water industry in England and Wales. The next section will present a test and a solution that may help address autocorrelation in the PR19 style econometric models.

2.1. Theoretical background and Wooldridge test

Wooldridge method as explained by Drukker (Drukker, 2003) starts with the following difference model:

$$y_{it} - y_{it-1} = (X_{it} - X_{it-1})\beta_1 + \epsilon_{it} - \epsilon_{it-1} \qquad (2)$$

$$\Delta y_{it} = \Delta X_{it} + \Delta\epsilon_{it} \qquad (3)$$

We estimate parameter *B₁* by regressing Δy_y on Δx_t and obtain the residual $\widehat{e_{it}}$. Based on Wooldridge procedure, if the ϵ_{it} are not serially correlated, then $Corr(\Delta\epsilon_{it}, \Delta\epsilon_{it-1}) = -0.5$. The procedure regresses the residual $\widehat{e_{it}}$ from the regression with first-difference variables on their lags and test that the coefficients on the lagged variables are equal to -0.5. The method of estimation can use variance clustering for robustness.

2.2. Application in the case of PR19 base cost models for water

The Wooldridge method for testing autocorrelation presented in Section 2.1 can be applied to the PR19 model specifications for water base cost.

The starting point of the testing exercise is to replicate all the five water base cost models in PR19, but with additional three years of data. The model specifications and estimation method stay exactly the same as in PR19 final determination. The modelling results are reported in *Appendix Table 4*.

The next step is to run the Wooldridge test for autocorrelation. The test procedure is run for all the five water base cost models and the results are reported in *Appendix Table 5*.

Based on the test results, first order autocorrelation exists in all five models, as the null hypothesis of no autocorrelation is strongly rejected.

There are various measures to correct for the bias and inefficiency caused by autocorrelation. One solution that addresses the autocorrelation is to transform the variable or the model. Adding a lagged dependent variable into the model is the approach tested in this paper. The suggested model would be:

$$y_{it} = \alpha + \gamma y_{it-1} + \beta_1 X_{it} + \varepsilon_{it} \quad (4)$$

The presence of the lagged dependent variable on the right-hand side would correct for the inefficiency in the standard errors of β_1, as the biases are now captured by γ.

The autoregressive models of order one (AR1) shown in equation 4 are run for all the water base cost models, and the results are reported in *Appendix Table 6*. The only difference between *Table 4* and *Table 6* models are that previous year relevant cost is added to the right-hand side as additional explanatory variables in *Table 6* models.

With the previous year cost added to the right-hand side as another explanatory variable, water resource plus models become weaker, as most of existing variables turn insignificant. The treated water distribution and the wholesale water model withstand the test. The only variable that becomes insignificant is the *number of booster pumping station per km of mains*.

One remark on the difference between the replicate PR19 models in Table 4 and the AR1 models in Table 6 is that the size of the scale variable coefficients has dropped significantly from the former to the latter. The reason is that much of the scale effects in the AR1 models have been captured by the lagged dependent variable. This difference shows another advantage of the AR1 models compared to the replicated PR19 models. In the replicated PR19 models, all coefficients of scale variables – number of properties and mains length – are greater than 1. This result is les reliable and may overestimate the scale effect. As it stands, the interpretation of this coefficient size would be that a 1% increase in scale (number of properties or mains length) would result in more than 1% increase in cost, which is not very intuitive in most cases.

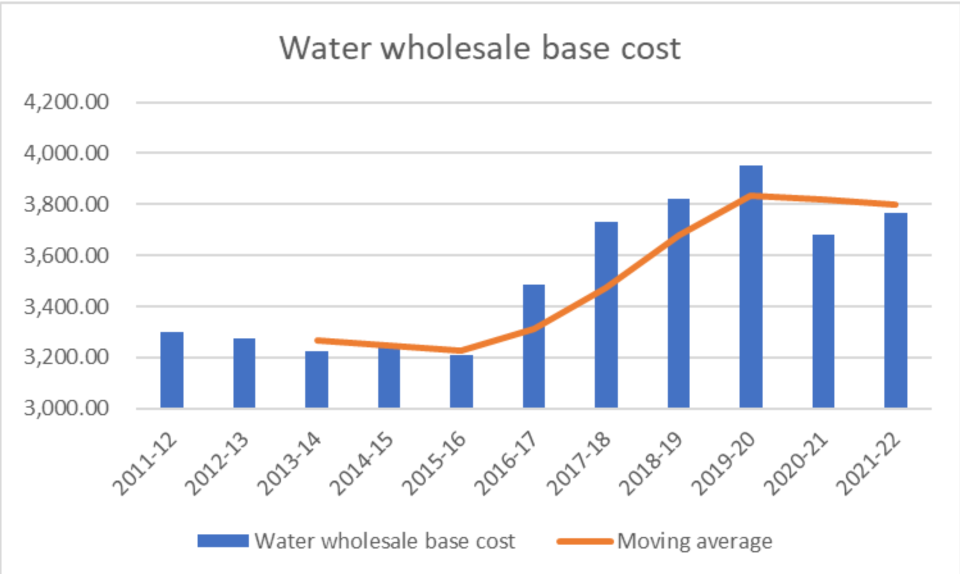
The next step in model improvement will be to rerun all the autoregressive (AR) models in *Table 6* that contain booster per mains length, but replace this variable (which is now insignificant) with average pumping head. The theory tells us that average pumping head (APH) is a valid energy cost driver, therefore using it as an alternative variable to booster per length is theoretically justifiable. The results of these AR models with APH, instead of booster per mains length are reported in *Appendix Table 7*. The relevant APH is used in the corresponding value chain. For example, APR in treated water distribution is used in the treated water distribution cost model, and APH for total water is used in water wholesale model. Regardless of the value chain or cost model, all the coefficients of APH are statistically significant, and perform better than booster per mains length.

A counter-argument in using lagged cost as another explanatory variable from the regulatory point of view can be that such a variable would cause perverse incentives to companies, telling them to spend more this year in order to be allowed to spend more next year. To verify this potential undesirable impact that the model may cause, we look at historical data of water industry base cost in real price over the 2012-2022 period, using

2017-18 as base year. *Figure 1* presents actual base expenditure and its three-year moving average that smooths out annual random variations. The trend shows that indeed expenditure increases by 14% during this period, even though the annual rates of increase seem to decelerate in more recent year. It can be due to investment cycle and some management decision in response to the price review cycle, but that is subject to further analysis.

The question to address here is whether adding lagged cost variable would cause perverse spending behaviour, or would it reflect the spending reality. Modelling work is a ex post action that uses the past to predict the future. In this case, the reality of increasing expenditure over time is the reality to be properly modelled. The industry water wholesale base cost has increased by 14% in real term over the last 10 years. This historical reality should be reflected in cost assessment going forward. From customer protection point of view, it is certainly not a desirable trend. However, both the industry and the regulator should look into this reality in designing assessment approach in order to balance everyone's interests. What needs to be furthered analysed next is the reason of increasing expenditure, whether it is enhanced quality, higher standard requirement, or inefficiency. But that should be a separate analytical work for policy decision.

Figure 1. Industry's water wholesale base cost trend, 2012-2022



When evaluating models for their predictive value, one of the indicators to look at is the distribution of the efficiency scores, the difference between the minimum (most efficient) to the maximum (least efficient) score, and the standard deviation. Table 1 compares these statistics across the three modelling approaches:

1) PR19 models replicated with new data added;

2) Booster per km of mains is replaced with average pumping head;

3) Lagged dependent variable (previous year's cost) added to correct for autocorrelation.

Every step from 1 to 3 is seen as an improvement in the efficiency score convergence, with smaller difference between the most and least efficiency company. Efficiency score dispersion is another outcome to be considered in evaluating models to be selected for cost assessment.

Table 1. Distribution of efficiency scores across models

Model	Max (least efficient)	Min (most efficient)	Standard deviation
PR19 replicated	1.40	0.77	0.15
Replace booster/length with APR	1.41	0.81	0.14
APH and previous year cost added	1.42	0.88	0.12

3. Direction of causality and theoretical background of PR19 retail cost models

3.1. General about endogeneity

Whenever we build an econometric model with one dependent variable and a number of explanatory variables, we need to justify the direction of causality and the impact as predicted by theory, if it exists, of every single variable. Without a clear justification, or a hypothesis, for having a variable in the model, any outcome would be subject to debate about what causes what and why. The PR19 retail cost models are typical example of specifications that are not supported by any theory of a causality check. Such a lack of theoretical background and justification for the direction of causality sharply weakens the predictive value of the models. The models can well suffer from endogeneity or spurious correlation that fails to explain the impact of the independent variables.

Endogeneity happens when the error terms of the regression model are correlated with one or more explanatory variables. When this happens, the coefficients of explanatory variables are no longer reliable in forecasting. This is among the most common problems in econometrics when we try to explain the impact of different variables on the dependent variables, and they all happen to impact each other.

3.2. Analysis of PR19 retail cost models

Appendix Table 8 shows the result of the PR19 models replicated with additional three years of data.

There are three dependent variables to be examined here: 1) non-deb cost per household, 2) deb-cost per household, and 3) the total of the aforesaid two, which is total cost per household.

For non-debt cost models, explanatory variables, or cost drivers, include the number of household properties, the percentage of dual service customers, and the percentage of metered customers. These models came out quite weak, as the adjusted R-squared are only in the range of 0.13-0.14, which means the models can explain only 13-14% of the data variation. This basic significance test alone would be sufficient to reject the models for their predictive value. However, this is not the main point for this paper.

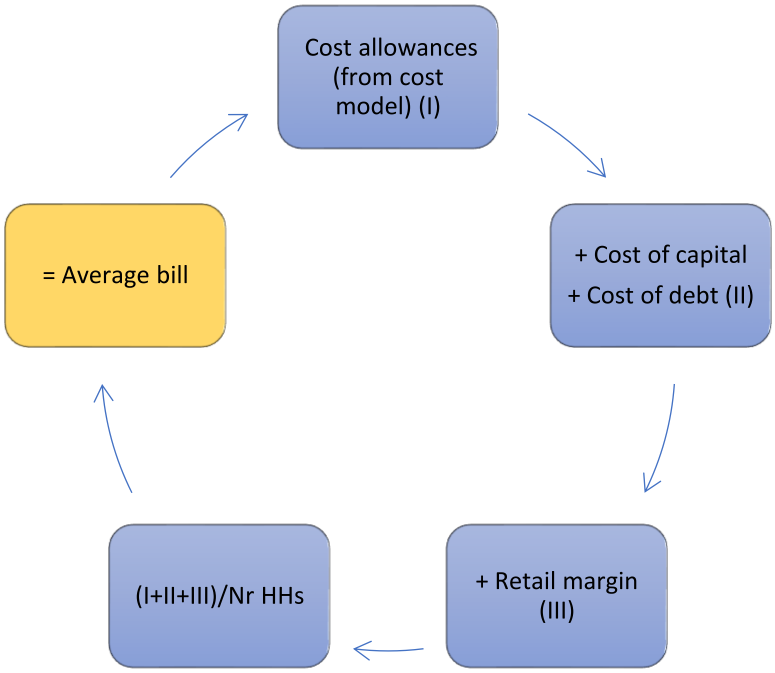
The deb cost models use revenue per household, some form of deprivation index, and total migration as explanatory variables. Here comes the potential endogeneity problem to be tested. The models use revenue as a predictor of cost without stating any theoretical background or some kind of justification. One can argue that cost comes first and causes revenue instead of the other way around.

In fact, the process of calculating allowed revenue starts with allowed cost, among other variables, as the input, and then adds cost of capital, other allowances, margin, etc. to arrive at the final allowed revenue. This practice in itself is causing endogeneity of the cost-revenue relationship to be modelled. Therefore, when the so-called average bill enters the cost model, it is already the result of cost allowances to a certain extent. One can argue that the revenue allowances is for both wholesale and retail, not just retail. However, retail is still part of the total revenue. *Figure 2* presents the full cycle of cost assessment and revenue allowances for retail.

Naming this variable as average bill is also misleading. It is a simple average revenue that companies are allowed to collect per household. This formula ignores consumption volume, which is the key driver of total cost, debt, and so many other operational and financial metrics. The actual bill that customers pay depends on their consumption and water charges, which is determined by the allowed revenue. Therefore, consumption should be in the picture as a cost driver.

The total cost models have similar specifications as the debt-cost models, with revenue per household as one of the explanatory variables.

Figure 2. Retail cost assessment and revenue allowance cycle



The results shown in *Appendix Table 8* indicate that all explanatory variables are problematic in one way or another. Some do not past the basic significance test, others have no supporting theory to justify their presence in the model, and the direction of causality is questionable. *Table 2* summarises the evaluation of all variables.

Table 2. Evaluation of explanatory variables in the replicated PR19 retail cost model

Variable	Outcome from model	Evaluation
Percentage of dual customer	Significant in non-debt cost model	Without supporting theory that states an expected sign of the coefficient, this variable is considered a spurious correlation.
Percentage of metered customer	Significant in non-debt cost models, but not in total cost models	Relevant for retail, as metered customers tend to make more contact about their metering (confirmed by companies' data on contact volume). Fragile significance across models, may need some transformation.
Number of household properties	Not significant	The supporting argument for this variable could be that it reflects scale if it is a total cost model. However, since it is unit cost in this case, scale variable fails to explain cost.
Deprivation index	Not significant	The variable was significant in the PR19 model data sample. It does not stand the test in the extended dataset with longer time series.
Total migration	Not significant	Same as above.
Revenue per household customer (or average bill)	Significant and contributes most part to the R-squared	Direction of causality is unclear. It is subject to the test to tell whether cost causes revenue or vice versa. A sign of endogeneity.

In order to test whether cost causes revenue, we have conducted two tests for reverse causality. The first test is to run a regression of revenue as a function of 1-3 lagged cost variable to check the significance of individual explanatory variable. The second test is to run a regression of revenue as a function of cost and its two lags, then do the F-test to see if all coefficients are jointly significant.

In the first test, we run a model with revenue as dependent variable and cost, both present and past, as explanatory variable. If the coefficients of cost in the past are significant in predicting revenue, then we can conclude that cost causes revenue to a certain extent. We consider the following model

$$R_{it} = \alpha + \sum_{k=1}^{3} \beta_i \, C_{i,t-k} + \varepsilon_{it} \tag{5}$$

Where R is revenue per household property, α is the model constant, C is total cost per household property, *i* stands for individual companies, *t* for time period, *k* for the number of lags to be tested, and ε is the error terms. The estimation approach is OLS with variance clustering for robustness. The test result is shown in *Appendix Table 9*. The test shows that cost of all the previous 3 years have significant coefficients. In the first test model, only one lag of cost is used, and the coefficient is significant. In the second test model, two lags of cost are used, and they are both significant. In the third test model, three lags of cost are used, and only the two-year lagged cost is not significant. This means that previous year cost indeed causes or has impact on present revenue. This is a complete reverse causality to what the PR19 retail cost model assumed or was meant to predict. This result confirms that endogeneity is present in this pair relationship.

In the second test, the model specification is slightly modified, with cost and its two lags, instead of lags only as in the first test. The regression looks like this

$$R_{it} = \alpha + \sum_{k=0}^{2} \beta_i \, C_{i,t-k} + \varepsilon_{it} \tag{6}$$

The result is shown in *Appendix Table 10*. This test also rejects the null hypothesis of no reverse causality.

The evaluation of all explanatory variables in the replicated PR19 retail cost models and endogeneity test have shed light on a need for serious rethinking of modelling approach for retail cost in PR24. The PR19 model specification with all variables are problematic cannot be used to assess cost efficiency. Cost assessment approach for retail needs to be completely rebuilt.

3.3. Suggestion for alternative modelling approach in retail cost assessment

The PR19 retail cost models seem to be heavily influenced by the same modelling approach applied for wholesale. In this approach, scale (or size), complexity, network characteristics are the key building blocks. While this structure works for wholesale, which is modelled on the total cost basis, it shows weaknesses in assessing retail cost on the unit cost basis.

Ofwat data request for retail service efficiency benchmarking in August 2022 shows a promising shift in modelling approach. The request was for data related to customer service activities, labour input, service quality and efficiency, etc. for the non-debt cost and affordability and payment related data for debt cost. While much of these activities or business characteristics can be within the management control, their normalised form still can be used for cost assessment in a relatively objective way. For example, number of inbound contact volume per customer service agent, or per household property, handling time per query, or dummy variable for outsourced call centre interacted with company's size. These are only a few suggestions. Without actual data

While the alternative assessment approach may better reflect the reality of retail activities, there is no guarantee that any econometric model built on this basis would produce meaningful results. No strong recommendation can be made before empirical work begins. In case this approach does not deliver a satisfactory result, or the models do not past the basis diagnostic tests, and the assessment still has to rely on the PR19 top-down style of modelling, it would make more sense to substantially reduce the weight of the PR19 style models in the overall assessment.

Due to the endogeneity and reverse causality shown earlier, if any endogenous variable would still be used, it is highly recommended to conduct the Granger causality test for panel data (Lopez and Weber, 2017) or some form of transformation and distributed lag model in order to address the bias.

While the data for bottom-up modelling approach are not available at the time of writing, an alternative modelling approach still can be tried. This paper recommends the total cost, instead of unit cost modelling, to be in line with wholesale modelling approach. The tested results are quite robust (*Appendix Table 11*).

Table 3. Summary of retail cost drivers to be tested

Cost to be modelled (Dependent variable)	Cost drivers (Independent variable)	Justification/Expected sign of coefficient
Total cost, non-debt cost, debt cost	Total consumption	Scale variable, not fully controlled by management, though long term trend may be downward as a result of water efficiency measures, not clear impact in short run. Positive.
Total cost, non-debt cost, debt cost	Number of connected households	Scale variable, not controlled by management, used in wholesale. Positive.
Total cost, non-debt cost	Number of metered households	Activity volume driver, endogenous in wholesale, but not in retail. Positive.
Total cost, non-debt cost	Migration	Activity volume, not controlled by management. Positive.
Total cost, debt cost	Percentage of households with default	Affordability driver, not controlled by management. Positive.
Total cost, debt cost	Consumption per household	Economies of scale for total cost, affordability for debt cost (more deprived households tend to consume less). Negative

Table 3 summarises the variables used in the suggested total retail cost models, the rationale and expected sign of the coefficients. In this total cost modelling approach, the same dependent variables, total retail cost, non-debt cost, and debt cost, are tested. The independent variables can be grouped into three categories: 1) scale, 2) economies of scales, 3) activity volume, and 4) affordability. The first two categories are similar to wholesale cost. The third is more specific and applicable to retail, as it affects debt related cost.

Scale variables tested in these models include total consumption and total number of connected households. Although consumption is not entirely out of control by the management in the long run, and it is expected to fall in line with water efficiency and environment performance commitments, it can still reflect scale and can be difficult to substantially influenced in the short run. The number of connected households is a similar scale variable used in wholesale cost model, and cannot be controlled by the management.

Variables that represent economies of scale in these models is average consumption (consumption divided by number of households). Average consumption is expected to have negative sign in the model, if the assumption of economies of scale should hold. This variable is introduced for testing and comparison purpose, rather than as a strong and highly recommended variable, as the reasoning behind its performance is subject to further analysis. It's presence in the model should be considered as empirical test only.

Number of metered household and total migration rate are meant to measure activity volume. This variable can be considered as endogenous in wholesale, as it is the management choice to decide on metering. However, it can be considered as more exogenous in retail, as metering has been decided outside retail activities, and retail as business unit still has to deal with the cost it causes in terms of contact volume and meter reading. The PR19 retail cost models use percentage instead of number of metered households, which does not perform well econometrically as discussed earlier. The logarithm form of absolute number of metered households performs better in the recommended total cost models. However, as metered households are part of total number of households, both cannot be used in the same model. Therefore, number of metered households are used only in models where total consumption is the scale variable for avoidance of multicollinearity.

Migration is the second variable that represents activities and is outside management control. It has been used in PR19 models with sound justification.

Affordability is measured by percentage of household with default. Affordability is outside management control and expected to affect debt related costs.

A few other explanatory variables have also been tested for comparison and robustness check. Those variables include density and its squared terms, council tax collection rate, and per capita consumption.

Density is significant in only one model where total cost is a function of number of connected households, density and its squared terms, and council tax collection rate. This result for density performance is considered unsatisfactory for robustness test, therefore not recommended.

Council tax has been tested in all three levels of cost. However, its coefficient is quite fragile, does not have an expected sign, therefore considered as failed to explain cost.

Per capita consumption, calculated as total consumption divided by population served, has also been tested. This variable is correlated with consumption per household, as expected. However, as population serve is not as stable as connected properties, and may not reflect the true impact on cost, hence not selected either.

Based on the rationale for variable selection explained above, a set of eight models are recommended for selection (Appendix Table 11). These are random effect models with variance clustering for robustness, the same approached applied throughout PR19. All variables are statistically significant. All of the results from diagnostic tests show that this modelling approach performs far better than the unit cost approach. A visual plot of residual versus fitted value is shown in *Figure 3*. These graphs do not indicate any pattern of relationship as a cause of concern for heteroskedasticity or correlation between the error terms and the fitted values.

The only aspect for further consideration is the dispersion of the efficiency score as predicted by the models. Some models predict while large difference between the most and the least efficiency companies. A triangulation of a few alternative specifications that capture different aspect of efficiency can help iron out such differences.

4. Conclusion

This paper has evaluated the water base cost and retail cost models from the PR19 price review. As more data are generated to be used in the next price review, the evaluation and review of previous modelling approach is necessary to improve the next modelling efforts. PR19 cost models achieved their objectives of making fair cost assessment to a certain extent at the time they were produced, based on the data available at the time. However, and as usual, more data added would change the model robustness and predictive value. Therefore, adjustments are needed.

For water wholesale model, autocorrelation was not an issue in PR19 models at final determination, based on rigorous diagnostic tests. However, with three more years of data, the test shows that autocorrelation does exist and needs proper measure to correct for it. The solution suggested in this paper is an autoregressive model of order one, meaning adding a previous year cost to the model as an additional explanatory variable. The correction causes some variable to turn insignificant, as usual. The solution suggested in this case is to replace that affected variable with an equivalent one that would still satisfy all other modelling criteria. If all the recommended measures are applied to improve the models, we can also see the narrowing in efficiency score as well, with smaller difference between the most and least efficiency company as predicted by the models.

For retail models, replicated models with additional data reveal the true weaknesses. The unit costs modelling approach seems to fail all the basis tests. The recommended total cost modelling approach provides a better technical alternative to assess retail cost efficiency.

Cost assessment is a complex process and one cannot expect to build a near perfect model. It is about trial and error in the learning process. Lessons learned from any price review are always useful in improving the next one.

Appendix

PR19 water base cost model replicated with dataset of 2012-2022

Table 4. PR19 water base cost models

	PR19 (1) Water Resource Plus 1	PR19 (2) Water Resource Plus 2	PR19 (3) Treated Water Distribution	PR19 (4) Wholesale Water 1	PR19 (5) Wholesale Water 2
Nr properties	1.074***	1.069***		1.065***	1.053***
	{0.000}	{0.000}		{0.000}	{0.000}
Water treated at 3-6	0.006***			0.004***	
	{0.000}			{0.000}	
Density	-1.614***	-1.412***	-2.921***	-2.119***	-1.865***
	{0.000}	{0.005}	{0.000}	{0.000}	{0.000}
Squared density	0.101***	0.087***	0.234***	0.150***	0.131***
	{0.000}	{0.009}	{0.000}	{0.000}	{0.000}
Weighted average complexity		0.377			0.425***
		{0.123}			{0.001}
Length of mains			1.066***		
			{0.000}		
Booster/length			0.500***	0.369**	0.370***
			{0.000}	{0.014}	{0.009}
Constant	-5.093***	-5.805***	4.895***	-1.341*	-2.326***
	{0.000}	{0.000}	{0.000}	{0.079}	{0.001}
Econometric_model	Random Effects	Random Effects	Random Effects	Random Effects	Random Effects
N	187	187	187	187	187
vce	cluster	cluster	cluster	cluster	cluster
R_squared	0.917	0.907	0.962	0.97	0.971
RESET_P_value	0.439	0.323	0.131	0.276	0.149
Prob > F(*)	0.005	0.003	0	0	0

*Significance level: *** (1%), ** (5%) and * (10%), p-values are in parentheses.*

(*) Result from Wooldridge test for autocorrelation. If the probability of the coefficient of the lagged residual from the regression with first- differenced variables is not equal to -0.5 is near zero, which is the case for all PR19 models, the autocorrelation is present.

Table 5. Wooldridge test for autocorrelation in panel data,

	PR19 (1)	PR19 (2)	PR19 (3)	PR19 (4)	PR19 (5)
Dependent variable in original model to be tested	Water Resource Plus 1	Water Resource Plus 2	Treated Water Distribution	Wholesale Water 1	Wholesale Water 2
Coefficient of lagged residual	-0.234**	-0.231***	-0.097	-0.072	-0.067
	{0.012}	{0.010}	{0.205}	{0.368}	{0.391}
F_stat	10.051	11.362	30.273	30.078	32.724
P-value	0.005	0.003	0	0	0

$F(1, 20) = 4.35$ at 5% level of significance

Null hypothesis: No first order autocorrelation

Decision rule: Reject null hypothesis if the correlation coefficient between the residual and its first order lag is not -0.5, as shown in the F-statistics that is greater than the F value.

Based on the test result in Table 3 above, the null hypothesis (no first order autocorrelation) is strongly rejected.

Table 6. Replicated PR19 water models with lagged dependent variable (AR1)

	AR1 WRP1	AR1 WRP2	AR1 TWD	AR1 WW1	AR1 WW2
	Water Resource Plus 1	**Water Resource Plus 2**	**Treated Water Distribution**	**Wholesale Water 1**	**Wholesale Water 2**
Nr properties	0.263***	0.235***		0.359***	0.359***
	{0.002}	{0.003}		{0.000}	{0.000}
Water treated at 3-6	0.002**			0.002***	
	{0.046}			{0.000}	
Density	-0.385	-0.219	-0.880***	-0.748***	-0.628***
	{0.135}	{0.249}	{0.000}	{0.000}	{0.000}
Squared density	0.024	0.013	0.070***	0.052***	0.044***
	{0.154}	{0.303}	{0.000}	{0.000}	{0.000}
Weighted average complexity		0.168*			0.243***
		{0.086}			{0.000}
Length of mains			0.325***		
			{0.000}		
Booster/length			0.089	0.075	0.079
			{0.119}	{0.148}	{0.193}
Previous year WRP cost	**0.760*** **	**0.785*** **			
	{0.000}	**{0.000}**			
Previous year TWD cost			**0.691*** **		
			{0.000}		
Previous year WW cost				**0.662*** **	**0.656*** **
				{0.000}	**{0.000}**
_cons	-1.386	-1.737*	1.256***	-0.568	-1.116***
	{0.148}	{0.052}	{0.000}	{0.172}	{0.002}
Econometric_model	Random Effects	Random Effects	Random Effects	Random Effects	Random Effects
N	168	168	168	168	168
vce	cluster	cluster	cluster	cluster	cluster
R_squared	0.966	0.965	0.98	0.983	0.983
RESET_P_value	0.257	0.312	0.808	0.844	0.965

*Significance level: *** (1%), ** (5%) and * (10%), p-values are in parentheses.*

Table 7. AR models with average pumping head as alternative to booster pumping station per km of mains

	AR1 TWD	AR1 WW1	AR1 WW2
	Treated Water Distribution	**Wholesale Water 1**	**Wholesale Water 2**
Length of mains	0.327***		
	{0.000}		
Average pumping head, TWD	0.059**		
	{0.031}		
Density	-0.920***	-0.777***	-0.669***
	{0.000}	{0.000}	{0.000}
Squared density	0.071***	0.054***	0.046***
	{0.000}	{0.000}	{0.000}
Nr properties		0.357***	0.348***
		{0.000}	{0.000}
Water treated at 3-6		0.002***	
		{0.001}	
Average pumping head, WW		0.073*	0.065
		{0.069}	{0.151}
Weighted average complexity			0.175***
			{0.002}
Previous year cost, TWD	0.690***		
	{0.000}		
Previous year cost, WW		0.671***	0.675***
		{0.000}	{0.000}
_cons	0.826**	-0.966**	-1.333***
	{0.043}	{0.024}	{0.001}
Econometric_model	Random Effects	Random Effects	Random Effects
depvar	lnrealbotextwd	lnrealbotexww	lnrealbotexww
N	168	168	168
vce	cluster	cluster	cluster
R_squared	0.98	0.983	0.983
RESET_P_value	0.794	0.724	0.915

*Significance level: *** (1%), ** (5%) and * (10%), p-values are in parentheses.*

Table 8. PR19 retail models replicated with data 2014-2022

	reROC2	reROC4	reRDC1	reRDC20	reRTC3	reRTC4	reRTC8
	PR19 non-debt cost 1	**PR19 non-debt cost 2**	**PR19 debt cost 1**	**PR19 debt cost 2**	**PR19 total cost 1**	**PR19 total cost 2**	**PR19 total cost 3**
Dual customer	0.002**	0.003***					
	{0.024}	{0.000}					
Metering	0	0			0.001	0.003	0
	{0.841}	{0.872}			{0.667}	{0.342}	{0.915}
Nr HH properties		-0.049				-0.081***	-0.068**
		{0.118}				{0.005}	{0.045}
Average bill			1.187***	1.162***	0.519***	0.623***	0.660***
			{0.000}	{0.000}	{0.000}	{0.000}	{0.000}
Deprivation, Equifax			0.024		0.011	0.021	
			{0.211}		{0.403}	{0.130}	
Deprivation, IMD				0.021			-0.002
				{0.388}			{0.856}
Migration				-0.015			0.004
				{0.515}			{0.723}
Constant	2.722***	3.354***	-4.893***	-4.281***	0.093	0.333	0.594
	{0.000}	{0.000}	{0.000}	{0.000}	{0.848}	{0.399}	{0.101}
Estimation_method	RE	RE	RE	RE	RE	RE	RE
depvar	lnsOC_hh	lnsOC_hh	lnDC_hh	lnDC_hh	lnsTC_hh	lnsTC_hh	lnsTC_hh
Adj_R_squared	0.132	0.14	0.614	0.604	0.617	0.64	0.605
RESET_P_value	0.683	0.121	0.094	0.136	0	0.006	0.009
N	152	152	152	152	152	152	152

*Significance level: *** (1%), ** (5%) and * (10%), p-values are in parentheses.*

Table 9. First manual test for reverse causality with 1-3 lags

	Test 1	Test 2	Test 3
Dependent variable		Average bill	
Total cost, 1 lag	1.156***	0.657***	0.622**
	{0.000}	{0.000}	{0.023}
Total cost, 2 lags		0.542***	-0.041
		{0.003}	{0.800}
Total cost, 3 lags			0.635***
			{0.006}
_cons	1.691***	1.536**	1.492**
	{0.007}	{0.022}	{0.029}
Estimation_method	OLS	OLS	OLS
N	133	114	95
vce	cluster	cluster	cluster
Adj_R_squared	0.603	0.623	0.653
VIF_statistic	1	5.585	9.195
F_stat	.	.	.

Table 10. Second manual test for reverse causality, 2 lags

Decision rule:

H0: Cost does not cause revenue (no reverse causality)

Reject H0 if F-statistic > F*

F* = F(3, 18) = 2.4

```
. reg lnrev_hh lnsTC_hh L.lnsTC_hh L2.lnsTC_hh, vce(cluster id)

Linear regression                               Number of obs     =        114
                                                F(3, 18)          =      16.94
                                                Prob > F          =     0.0000
                                                R-squared         =     0.6650
                                                Root MSE          =     .25044

                                   (Std. Err. adjusted for 19 clusters in id)
------------------------------------------------------------------------------
             |               Robust
    lnrev_hh |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    lnsTC_hh |
         --. |   .6124111   .1918466     3.19   0.005     .2093563    1.015466
         L1. |   .2064995   .1743537     1.18   0.252    -.1598041    .5728031
         L2. |   .4385414   .1632866     2.69   0.015     .0954889    .7815938
             |
       _cons |   1.363161   .6139692     2.22   0.039     .0732592    2.653062
------------------------------------------------------------------------------
```

```
. test lnsTC_hh L.lnsTC_hh

 ( 1)  lnsTC_hh = 0
 ( 2)  L.lnsTC_hh = 0

       F(  2,    18) =    8.95
            Prob > F =    0.0020

. return list

scalars:
                r(drop) =  0
                r(df_r) =  18
                   r(F) =  8.950825923544695
                  r(df) =  2
                   r(p) =  .0020018091327127
```

Stata command of the test

reg lnrev_hh lnsTC_hh L.lnsTC_hh L2.lnsTC_hh, vce (cluster id)

test lnsTC_hh L.lnsTC_hh

return list

Table 11. Total retail cost models

	tc1	tc2	tc3	tc4	tc5	tc6	tc7	tc8
	Total retail cost			**Non-debt cost**			**Debt cost**	
lnconsumption	0.834***		0.628***	0.480***				1.004***
(Total consumption)	{0.000}		{0.001}	{0.000}				{0.000}
eq_lpcf62	0.035**					0.079***		0.047**
(Household with default, %)	{0.021}					{0.000}		{0.028}
lnhh_t		1.053***			1.007***		1.172***	
(Nr connected households)		{0.000}			{0.000}		{0.000}	
lnavgconsumption		-0.448*				-0.238*	-0.763***	
(Consumption per household)		{0.054}				{0.065}	{0.006}	
lnmeter_nr			0.276*	0.418***		0.879***		
(Nr metered properties)			{0.073}	{0.000}		{0.000}		
totalmigration			0.041*	0.074***	0.021*			
(Total migration, %)			{0.066}	{0.000}	{0.058}			
_cons	11.547***	2.026***	11.335***	10.412***	2.471***	8.938***	-1.048	9.240***
	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.000}	{0.315}	{0.000}
depvar	lnTCsdebt_	lnTCsdebt_	lnTCsdebt_	lnsOCsdeb	lnsOCsdeb	lnsOCsdeb	lnDCsdebt_	lnDCsdebt_
Estimation_method	RE	RE	RE	RE	RE	RE	RE	RE
N	153	153	153	153	153	153	153	153
vce	cluster	cluster	cluster	cluster	cluster	cluster	cluster	cluster
Adj_R_squared	0.835	0.954	0.885	0.909	0.959	0.955	0.897	0.735
RESET_P_value	0.329	0.492	0.375	0.276	0.843	0.182	0.354	0.882

Notes:

*Significance level: *** (1%), ** (5%) and * (10%), p-values are in parentheses.*

In total nine models are presented, of which four are total cost, three are non-debt cost, and two are debt cost. Following Ofwat's suggestion of giving more weight (75%) to the top-down (total cost) model, and less weight (25%) to the bottom-up models, the overall efficiency score can be calculated as follows:

Predicted total cost = Sum of predicted cost in models tc1, tc2, tc3/3

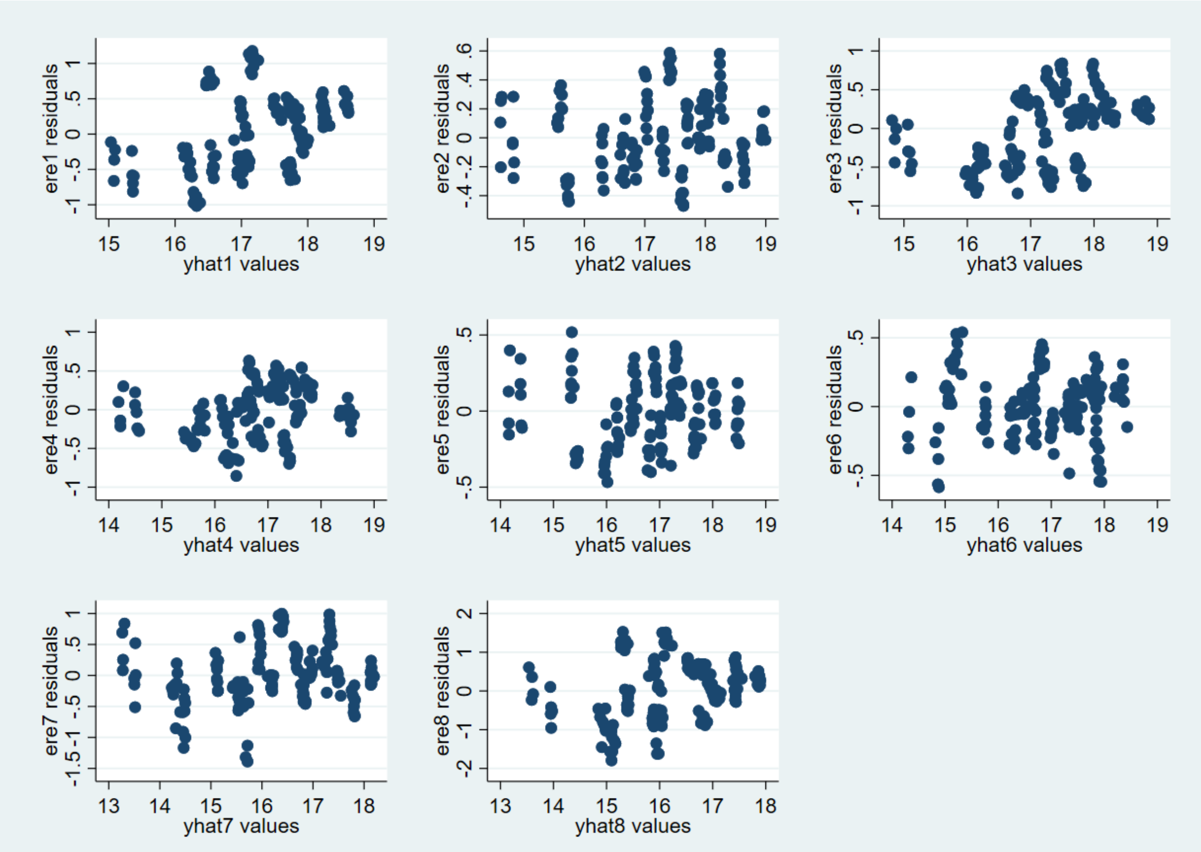
Predicted non-debt cost = Sum of predicted cost in models tc4, tc5, tc6/3

Predicted debt cost = Sum of predicted cost in models tc7, tc8/2

Triangulated predicted cost =

*Predicted cost * 0.75 + (Predicted non-deb cost + Predicted debt cost) * 0.25*

Figure 3. Residual versus fitted value plot for retail cost models



These graphs plot the residuals versus the fitted values for each of the eight selected models for retail cost. There is no meaningful pattern detected for the distribution of the residuals. Therefore, heteroskedasticity is not a concern from visual check, despite the LM test shows that it may be the case.

References

Drukker, D. M., Testing for Serial Correlation in Panel Data, The Stata Journal, 2003, 3, Number 2, pp 168-177

Lopez, L. and Weber, S., Testing for Granger causality in panel data, The Stata Journal, 2017, pp. 972-984