Table of Contents

# Dynamic Panel Data Modelling Approach for Enhancements – An Alternative Technique for PR24 and beyond

September 2021

**Section 1** Introduction ..... 3
Enhancement expenditure modelling ..... 3
**Section 2** Enhancement Modelling Approach ..... 5
A Enhancement Approach ..... 5
B Literature Review ..... 6
C Example 1: Metering ..... 7
Descriptive Statistics on Metering ..... 7
Ofwat PR19 FD Metering Enhancement Modelling Static Panel Approach: A summary ..... 9
PR19 Ofwat's Enhancement Metering Model at FD ..... 12
Dynamic Panel Models on Enhancement Metering ..... 14
D Example 2: Growth ..... 22
Descriptive Statistics on Growth (New Development & Connections) ..... 22
PR19 Ofwat's Growth Models at IAP ..... 24
Dynamic Panel Models on Enhancement Growth ..... 26
E Conclusion ..... 30
F References ..... 32

**ABSTRACT**

This report considers potential ways to design econometric models to estimate the efficient cost of enhancement expenditure. It builds upon our submission, as third party, to the CMA in its redetermination of the water price controls in 2020 and our involvement in the PR19 cost assessment process.

It has been produced by Thames Water for the purpose of helping Ofwat (and other interested parties) develop models which are as good as they can be, underpinned by robust statistical tests, sensible coefficients, and the most representative econometric technique to model capital expenditures such as is the case of enhancements. In this report we focus on Dynamic Panel Models as an alternative technique to model the lumpiness of enhancements. We show that empirical evidence supports the implementation of dynamic models versus the static approaches used in previous price reviews. Using different examples in metering and growth the results suggest a significant improvement when compared to previous or more traditional approaches such as cross-sectional or static panel models. This allows us to present models that are consistent with regulatory, economic, and econometric principles.

# Section 1 Introduction

## Enhancement expenditure modelling

1.1 A significant area of difference between Ofwat and companies at PR19 was over the appropriate level of enhancement expenditure. Industry requested £10.71bn while Ofwat provided allowances of £8.27bn, which is equivalent to a 23% gap between the allowances and business plans, whereas on base costs the gap at the industry level was only 0.4%.[1] Ofwat relied on relatively simple models to assess enhancements, for example the only cost driver for metering expenditure was the number of meters installed, which needed to be supported with deep dive assessments. In our view this is an area where econometric modelling could be used to improve and reduce the subjectivity of the deep dive approach across significant areas of expenditure. By using a technique that captures the dynamic and behaviour of enhancement expenditures would produce an efficient allowance that aligns with customers preferences and outcomes.

1.2 During PR19 we undertook dynamic panel modelling of growth expenditure, when this was being assessed outside of the botex models and found that the enhancement modelling was improved by this approach.

1.3 We have more recently considered this approach for the assessment of metering expenditure and found similar conclusions. The predictive power of the econometric models were significantly improved by the use of dynamic panel models.

1.4 The CMA has given credit to explore dynamic panel models as an alternative to the traditional approaches suggested and used in price reviews. This econometric technique controls for the lumpiness of this type of expenditure for areas where data is available such as metering, new developments & connections (growth) or even capital maintenance among others.

1.5 The comments made by the CMA can be found in their Final Report at https://assets.publishing.service.gov.uk/media/60702370e90e076f5589bb8f/Final_Report_---_web_version_-_CMA.pdf (p. 415, footnote 1429).

1.6 Hence, enhancement expenditure (where appropriate) should be considered or at least explored using dynamic panel data as an alternative to what has been used in the previous PR14 and PR19 reviews. This approach has the potential to improve the assessment of these expenditures by capturing the lumpiness of the expenditure, which will result in more reliable and objective results that will be beneficial for customers and the environment.

[1] See https://www.ofwat.gov.uk/wp-content/uploads/2019/12/PR19-final-determinations-Securing-cost-efficiency-technical-appendix.pdf
Ofwat PR19 final determinations: Securing cost efficiency technical appendix (updated April 2020) p. 167-68.

1.7 The *aim* of this paper is to illustrate how dynamic panel models could be an alternative in assessing efficient enhancement expenditure. The examples we provide in the paper are just an illustration of the potential that this econometric technique has for modelling purposes. The econometric models proposed in the examples of the document (e.g., metering and growth) can be extended by including other relevant cost drivers, as is the case of the metering example in section 2C.

1.8 The purpose of the document is to provide an objective analysis and to highlight the potential of this technique to understand how the lumpiness of expenditure can be captured or controlled by a lagged variable and how this should be treated from the econometric point of view. This tool can help us to understand capex from a different angle and to model closely the patterns observed in these types of expenditures versus traditional techniques such as static panel data models that perform better in assessing base costs.

1.9 In practical terms, this technique can be easily implemented in different statistical packages. Particularly, Stata is quite convenient for the easy implementation of the different estimators that are used to estimate these models.

1.10 The next sections will provide a brief summary of the dynamic panel data modelling approach and some applications in metering and growth:

- Section 2A highlights the key issue of lumpiness in modelling enhancement expenditure.
- Section 2B provides a literature review of the use of dynamic models.
- Section 2C provides an example using metering
- Section 2D provides an example using growth.
- Section 2E provides our conclusions.

Section 2

# Enhancement Modelling Approach

## A Enhancement Approach

2.1 Enhancement or investment models play a fundamental role in assessing cost allowances. In the 2018 Econometric consultation[2] Ofwat proposed Static Panel Data models for water and waste enhancement activities. For the draft determination (DD) and final determination (FD) Ofwat used a simpler set of econometric or unit cost models (e.g., in enhancement metering, a cross-section econometric model is used).

2.2 Ofwat recognises that the key issue for enhancement modelling is: *"the efficient level of enhancement costs is more difficult to estimate than for base costs. Due to their irregular nature, there is less opportunity to compare the cost of required enhancement solutions between companies, and in some areas the exact requirements may be subject to uncertainty"*[3]. The challenging part of enhancement modelling is therefore the irregular nature of investments/enhancements or the lumpy patterns observed across the industry on different types of enhancement activities in water and waste, within and between companies.

2.3 The aim of this paper is to provide an alternative approach to the cross-section, static panel data, or unit cost approaches used by Ofwat during the PR19 cost assessment enhancement process and also to what has been used in the past such as in PR14.

2.4 The alternative modelling approach we propose to be explored for a more objective assessment, relies on dynamic panel data models. The nature of enhancement investments expressed as lumpy or irregular levels across the industry can be modelled using these types of models.

2.5 The dynamic approach allows us to capture in a consistent way the different dynamic patterns (e.g., irregularities) of investments that each company faces at any particular period of time (e.g., yearly) by introducing the lagged dependent variable (e.g., the amount invested in metering in previous years, *t-1*, *t-2*, *t-3*, for example) that could captures the history, cyclical or lumpy patterns of investments within a company. It also allows us to understand the persistence of enhancement decisions of the past in the present, or to capture the magnitude of adjustment investments expenditures.

2.6 This paper provides two applications for the use of dynamic panel models using data relating to metering and growth enhancement expenditure. Metering represents a significant proportion of enhancement expenditure and it also reflects one of the largest

[2] See Section 4, p.102 Econometric Consultation 2018 at: https://www.ofwat.gov.uk/wp-content/uploads/2018/03/Appendix-1-Modelling-results_Final.pdf

[3] Ofwat, 'PR19 draft determinations: Securing cost efficiency technical appendix', July 2019, pp. 35-37, Section 4.1.

within company variations across the different enhancement activities in the water industry[4] (see section 2C). As an additional area of analysis, we also explore the use case of growth as a separate model as it was initially treated by Ofwat in the IAP. We provide our analysis of the use of dynamic models to assess growth expenditure in Section 2D.

2.7 The next section provides a brief review of the literature on how dynamic models are used to treat this type of expenditures. The literature in this topic is significantly large so we only introduce some practical examples.

## B Literature Review

2.8 Companies' expenditures show different cycles and patterns (e.g., lumpiness) over price control periods. For example, when a new development project or metering strategy is designed for three, four or five years there may be, during this time, some reallocation of resources that could over or underestimate the initial investment plan, yielding a lumpy investment series.

2.9 For example, Peck (1974) is one of the earliest articles that highlights the lumpy nature of investments in utilities. His model introduces dynamic components in the empirical specification such as the fixed lag model used to explain the lumpy investments made by the firm. His investigation is applied to a sample of 15 firms in the U.S. electric utilities industry between 1948-1969 using Bayesian econometrics.

2.10 Furthermore, in a regulatory framework, investments can also be driven by different regulatory incentives or macroeconomic shocks (e.g., changes in demand such as COVID-19). For instance, Cambini and Rondi (2011) use a dynamic panel model of investment in 15 EU Public Telecommunications Operators to account for investment adjustments. Similarly, Cambini et al. (2016) use a dynamic accelerator model of investment, to test the impact of output-based incentives on the investment rate and to see if this incentive on investment survives after controlling for other determinants using a dynamic panel data model. The information used in this paper is based on the largest electricity distribution operator in Italy with 115 distribution zones between 2004-2009.

2.11 We find that there is a large academic literature on how dynamic effects play a significant role on investment models in utilities and other areas by including the effect of the lagged dependent variable. For example, Nardi (2012) proposes an empirical analysis to assess if unbundling measures are related to the increase of grid investments focusing on 14 countries between 2001 and 2010. The author proposes a dynamic panel data model of interconnection investments.

2.12 As another example on the use of dynamic models, Poudineh and Jamasb (2016) develop a model that considers the main determinants of investment under incentive regulation in the Norwegian electricity distribution network. Their dynamic model includes the lag of investment to control for the lumpiness behaviour of investment.

[4] During PR19 we undertook dynamic panel modelling of growth expenditure, when this was being assessed outside of the botex models and found that the enhancement modelling was improved by this approach.

Large investment projects may take multiple years, so spells of high investment rates are followed by spells of zero investments. The main result is that due to the dynamic nature of investment decisions, a large part of the variation in the levels of investment of the firms is explained by investment rates in previous periods. The analysis is applied to 129 electricity companies in Norway between 2004 and 2010 using a Bayesian Model Averaging approach.

2.13 These examples illustrate that investment or enhancement expenditure in period *t* could be influenced by the dynamic effect of previous events/decisions at *t-1*, *t-2*…*t-N*, which basically reflects previous managerial expenditure decisions and cyclical or lumpiness investment patterns.

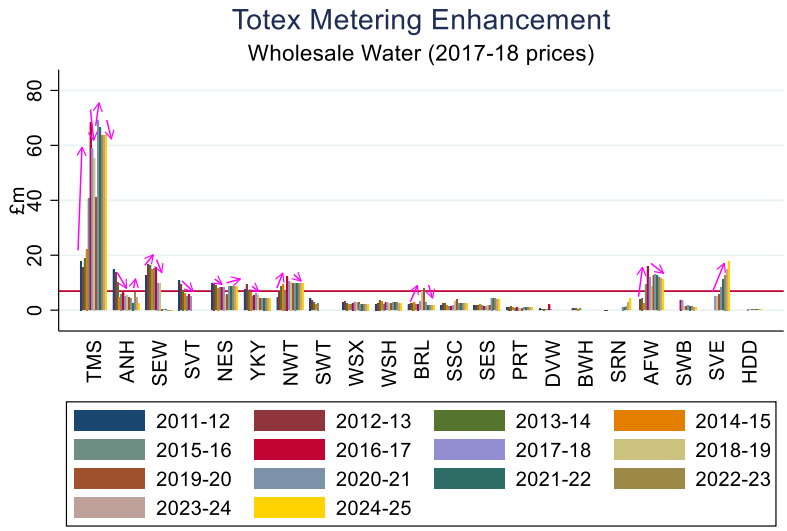
2.14 Therefore, a more realistic and appropriate empirical specification of enhancement expenditure should control for these dynamic patterns (e.g., lumpiness, irregularities). In the following sections, we illustrate with examples in metering and growth enhancement cases how a potential dynamic model could be implemented. In these sections we investigate the patterns of metering expenditure and growth but also, we compare the use of different econometric approaches such as static versus dynamic panel data models.

## C Example 1: Metering

### Descriptive Statistics on Metering

2.15 This section focuses on metering enhancement expenditure only. Figure 1 shows the investment/expenditures levels on metering for each water company between 2011-12 to 2017-18 and forecasts through to 2024-25. It shows the lumpy or irregularity of metering investment within and between companies across the industry. For example, ANH investment levels are quite volatile with a min and max of £2.3m and £15m, respectively. Similarly, TMS has a min of £15m and a max of £68m, whereas the industry has an average investment across the period of £7m per year. Many small companies have very low levels of metering investment; nearly zero (e.g., PRT has a min and max of £0.2m and £1.5m, respectively) which reflects the different strategies, conditions and heterogeneities that each company faces. For instance, between 2010 and 2015, SRN became the first water company to implement a 100% metering policy across their region due to their critical position as a water stressed zone.

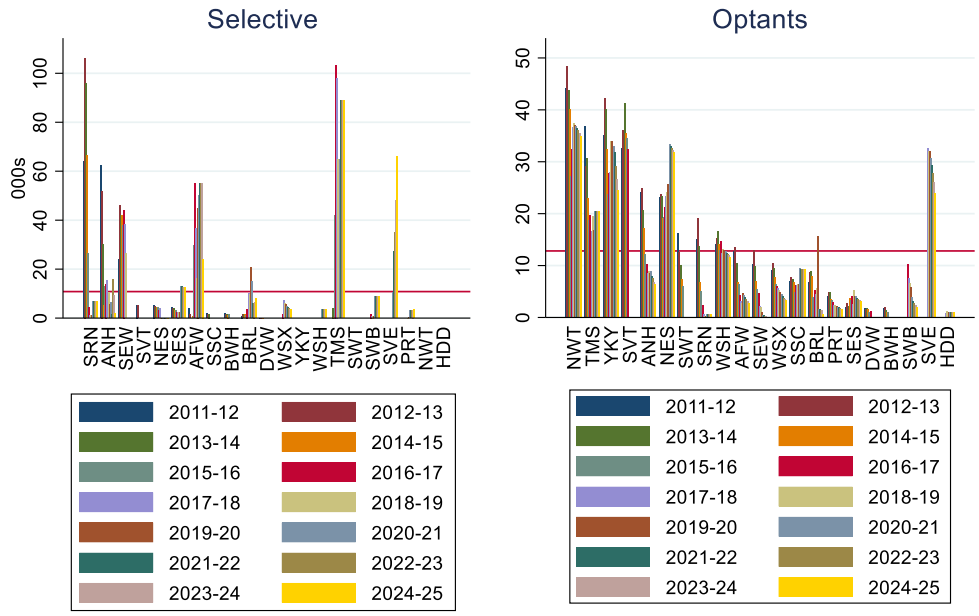
## Figure 1 – Totex Metering Enhancement




Source: Economic Regulation, Thames Water

2.16 Another example of the different strategies implemented by companies is shown in Figure 2. This figure depicts the different levels of meters installed during AMP6 and AMP7 (e.g., selective or optant) reflecting the different priorities that companies are facing such as leakage reduction (e.g., a better control and understanding of the network using smart meters could help to reduce leakage). In addition, some companies are trying to understand, and provide incentives for the reduction of water consumption (per capita consumption-PCC) as is the case of SRN.

## Figure 2 – Meters Installed over AMP6 and AMP7




Source: Economic Regulation, Thames Water.

## Ofwat PR19 FD Metering Enhancement Modelling Static Panel Approach: A summary

2.17 In this subsection, we introduce a summary of the PR19 FD approach used by Ofwat on modelling metering enhancement expenditure and how this approach shows some inconsistencies and weaknesses. We will also extend these models into a static panel data framework to see the potential for improvements from exploiting the characteristics of the panel data in the next section. In the last part of Section 2 we introduce the advantages of dynamic panel data models.

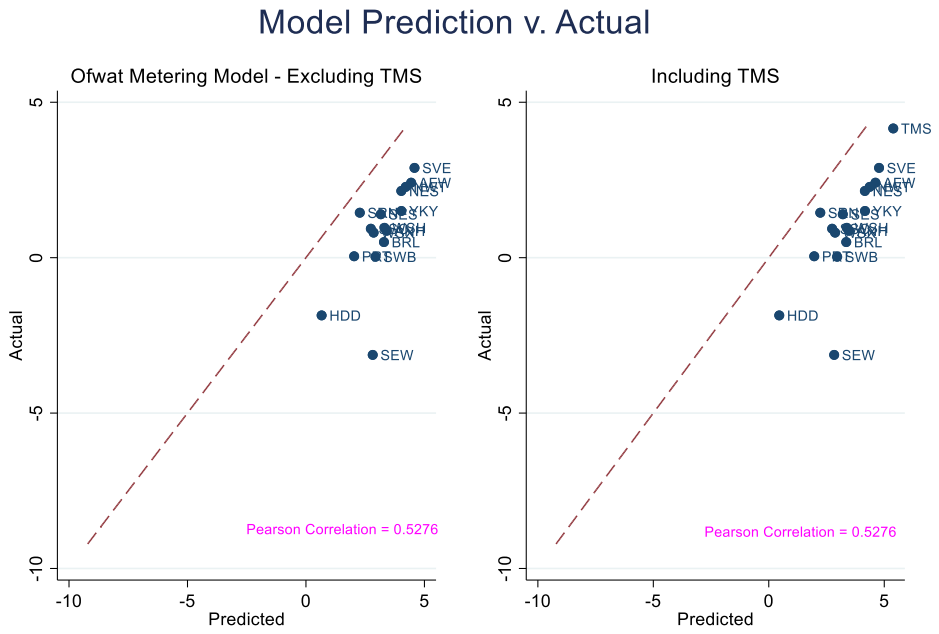
2.18 A brief summary of Ofwat's approach to modelling metering expenditure and key assumptions is presented as follows:

- Ofwat has used a unit cost model based on the forecast totex in metering expenditure submitted by companies for the period 2018 to 2025 (i.e., 7 years period).
- The main and only cost driver of the unit cost model is the combined number of optant and selective meters.
- The model is approached as a cross-sectional linear regression model at the levels and log specification.
- The model aggregates totex by adding all years of the seven-year period to obtain a single observation for each company and to run a statistical analysis on 15 observations. This is a small number of observations to establish a reliable model. Similarly, an aggregate of all optant and

selective meters is created to obtain the total number of new meters for the seven-year period for each company.

- The final modelling outcomes face an efficiency challenge.
- Deep dive is applied for those companies where there is a material gap between the model's forecast and companies' business plans.
- Thames Water (TMS) is excluded from the initial linear regression model as an outlier.
- The metering models explicitly exclude a control for meter technology.
- There is no consistency on the allowance prediction calculations. A time framework of seven years, aggregated as a single cross-sectional model and the estimated coefficients derived from this seven-year aggregation are used to estimate a period that only uses five years (e.g., AMP7). Basically, it imposes a slope that does not belong to the initial set of observations used to estimate the original model.
- Finally, Figure 3 below analyses the fit of the logarithmic model used in PR19. The chart depicts the 45-degree line for perfect correlation that indicates a good fitness of the model between what the cost model predicts and the actual cost. The correlation between actuals and predicted values is significantly low (0.57) with or without Thames Water (TMS). Ofwat claims that TMS is an outlier but the right chart on Figure 5 shows no indication of this claim. In fact, there are other observations such as SEW and HDD that are significantly underestimated and causing a poor performance of the model. Overall, all the predictions are underestimated across the industry showing a potential significant weakness of the model.

## Figure 3 – Model Prediction v. Actual (Ofwat PR19 Ln's Metering Model)




Source: Economic Regulation, Thames Water. Note: In this Ofwat's model metering version we exclude realloacations.

## PR19 Ofwat's Enhancement Metering Model at FD

2.19 In this sub-section we focus only on the logarithmic (Ln) version of the metering enhancement model used at PR19. The reason is to keep the analysis simple, but the implications of this analysis hold for any other treatment of the models such as when it is modelled in levels. Table 1 presents the econometric results of the aggregate or cross-section approach used at PR19 excluding (see column label as *OM_A*) and including TMS in the analysis (see column label as *OM_A_TMS*). All the models in Table 1 are based on the period 2018 to 2025.

## Table 1- Ofwat Metering Models and Extensions

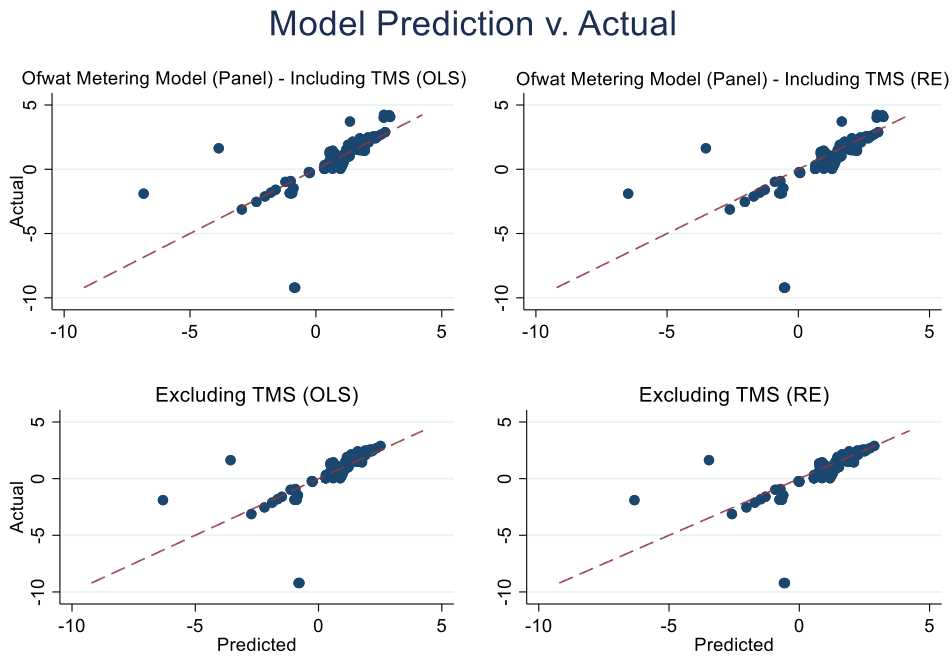
Ofwat Metering Enhancement Models and Extensions (Aggregated (A) and Panel (P) Versions)

| | OM_A<br>b/se | OM_A_TMS<br>b/se | OP1T_OLS<br>b/se | OP2T_RE<br>b/se | OP1_OLS<br>b/se | OP2_RE<br>b/se |
|---|---|---|---|---|---|---|
| ln_agg_meter | 0.992*** | 1.094*** | | | | |
| | (0.05) | (0.09) | | | | |
| ln_meters | | | 0.834*** | 0.828*** | 0.767*** | 0.800*** |
| | | | (0.20) | (0.27) | (0.19) | (0.28) |
| Constant | -1.414*** | -1.835*** | -1.002* | -0.685 | -0.939 | -0.724 |
| | (0.23) | (0.38) | (0.58) | (0.82) | (0.57) | (0.80) |
| R2_Adjusted | 0.9524 | 0.9333 | 0.5302 | | 0.4839 | |
| R2_overall | | | | 0.5336 | | 0.4879 |
| Observations | 16 | 17 | 137 | 137 | 129 | 129 |

Source: Economic Regulation, Thames Water.
* p<0.10, ** p<0.05, *** p<0.01

2.20 In these two cross-section models, we did not make any reallocations in order to be consistent with the panel data results. These models show a significant $R^2$ above 0.93 with or without TMS. However, as mentioned before in paragraph 2.18, the predictability of the model is quite poor as was shown in Figure 3 previously, with serious concerns about the effects of potential outliers. This result could suggest a significant harm on the final allowances. The last four columns presented in Table 1 are models that extend this analysis by using a static panel approach (no aggregation of the data). Columns OP1T_OLS and OP2T_RE are models that include TMS and are modelled by pooled OLS and Random Effects (RE). The last two columns present the same results but excluding TMS from the regression. The models with TMS included and exploiting the characteristics of the panel (RE) yield higher $R^2$ and a more accurate prediction than the ones that exclude TMS (see models OP1_OLS and OP2_RE in Figure 4).

Figure 4 – Model Prediction v. Actual (Static Panel Models OLS and RE)




Source: Economic Regulation, Thames Water. Note: In this Ofwat's model metering version we exclude realloacations.

2.21 However, Figure 4 shows again that a few observations are causing a potential problem as they are likely to be outliers. The rest of the observations tend to be closer to the 45-degree line and to reflect a higher correlation between actuals and predicted values in a more accurate way that the ones presented in Figure 3.

2.22 Table 2 summarises the predictions made by each of the models compared to the current PR19 model outputs on the log model presented in the first column of Table 1. If Ofwat had used the consistent approach of the prediction of the model, the allowances for this model would have been higher. By consistency we mean the prediction of the model over all the set of observations that the model estimates and it does not impose a change in the driver to the aggregation of five years to make the prediction yearly rather than the seven-year aggregated period. The second and third column in Table 2 that uses a panel structure of the data is consistent with the estimated coefficients[5].

[5] The totals produced by the consistent models can be divided by seven years and the yearly average then multiplied by five to get a consistent outcome with the estimations of the models. By doing this, the outcomes are still higher than the ones produced by Ofwat. This piece of analysis/example on the calculation of the allowances is to illustrate the potential weakness and underestimation of the final allowances and the significant harm that companies are facing in this enhancement case.

## Table 2 – Expenditure Predictions using Ln models

| | Expenditure Predictions using Ln models (Cross-Section v Static Panel Models), £m | | | | | | |
|---|---|---|---|---|---|---|---|
| | Section Aggregate and Inconsistent Predictions | Cross-Section Aggregate and Consistent Predictions | | Static Panel Models Predictions | | | |
| Company | Ofwat Metering PR19 model Ln (No TMS and Inconsistent) | Ofwat Metering PR19 model Ln (No TMS and Consistent) (OM_A) | Ofwat Metering PR19 model Ln (With TMS and Consistent) (OM_A_TMS) | OP1T_OLS with TMS | OP2T_RE with TMS | OP1_OLS No TMS | OP2_RE No TMS |
| AFW | 60 | 85 | 101 | 49 | 65 | 40 | 56 |
| ANH | 17 | 30 | 33 | 17 | 23 | 15 | 20 |
| BRL | 11 | 27 | 29 | 12 | 16 | 11 | 15 |
| HDD | 1 | 2 | 2 | 2 | 3 | 2 | 2 |
| NES | 38 | 56 | 64 | 34 | 45 | 28 | 39 |
| NWT | 42 | 68 | 79 | 36 | 49 | 31 | 43 |
| PRT | 6 | 8 | 7 | 7 | 10 | 7 | 9 |
| SES | 19 | 24 | 25 | 19 | 25 | 17 | 23 |
| SEW | 0 | 17 | 17 | 1 | 1 | 1 | 1 |
| SRN | 9 | 10 | 9 | 10 | 13 | 9 | 12 |
| SSC | 11 | 15 | 16 | 12 | 16 | 11 | 15 |
| SVE | 75 | 98 | 119 | 59 | 80 | 48 | 68 |
| SWB | 14 | 19 | 19 | 14 | 19 | 13 | 17 |
| TMS | 112 | | 219 | 89 | 119 | | |
| WSH | 19 | 28 | 29 | 18 | 25 | 16 | 22 |
| WSX | 9 | 17 | 18 | 10 | 14 | 10 | 13 |
| YKY | 34 | 56 | 64 | 30 | 41 | 26 | 36 |

Source: Economic Regulation, Thames Water.

2.23 The last set of columns produce the predictions under a static panel approach where the RE models suggest that the OLS cross-section approach is yielding unpredicted outcomes. In the next section we consider a potential way to tackle the irregularity or lumpiness of these types of investments or enhancement expenditures by extending the analysis into a dynamic panel framework.

## Dynamic Panel Models on Enhancement Metering

2.24 In this section we explore an alternative approach that allows us to capture the irregularities or lumpiness of the expenditure. We present the analysis in two parts. Firstly, we investigate the dynamic effect of enhancement expenditure based on historical data (2011-12 to 2017-18) showing the persistent and significant effect that this driver has in the models and secondly, we extend the analysis using only the future values of the panel (2018-19 to 2025-25) as has been proposed by the current PR19 assessment of the metering enhancement case.

2.25 To highlight the importance of dynamic effects in these types of expenditure and its implications on the general performance of the models we can follow Bond (2002) that says that '*even when coefficients on lagged dependent variables are not of direct interest, allowing for dynamics in the underlying process may be crucial for recovering consistent estimates of other parameters*'[6].

[6] Bond, S. (2002). Dynamic Panel Data Models: A Guide to Micro Data Methods and Practice. The Institute for Fiscal Studies, CEMMAP working paper.

## Historical Dynamic Panel

2.26 In this section we present the main results of adding the dynamic effect of enhancement across different model specifications ranging from pooled OLS to RE and Fixed Effects (FE), but first we describe briefly the mathematical expression of the dynamic panel data model. The model is written as follow[7]:

$$y_{it} = \delta y_{i,t-1} + x'_{it}\beta + u_{it}$$

2.27 Where $\delta$ captures the effect of the dynamic component of adjustment of previous levels of enhancement expenditure calculated in the lagged variable $y_{i,t-1}$ whereas $x'_{it}$ contains all other exogenous drivers such as scale controls, for example, meters installed (i.e., selective + optant) or the different types of technologies on meters installation (e.g., AMI (Advanced Metering Infrastructure) or AMR (Automatic Meter Reading)) and $u_{it}$ is the error component defined as $u_{it} = \mu_i + v_{it}$, where $\mu_i$ is the unobserved time-invariant heterogeneity of the firm and $v_{it}$ is the random noise both to be assumed i.i.d.

2.28 Dynamic panel data models estimated by OLS are biased and inconsistent by construction. Since the dependent variable $y_{i,t}$ is a function of the unobserved time-invariant firm heterogeneity effect $\mu_i$, it immediately follows that the lagged dependent variable $y_{i,t-1}$ is also a function of $\mu_i$, hence it will be correlated with the error term $u_{it}$[8]. Similarly, the within and GLS estimators for the Fixed and Random Effects models in a dynamic panel model approach are also biased and inconsistent[9].

2.29 Table 3 presents the results on a historical dynamic panel data set framework for enhancement metering (2011-12 to 2017-18). The first two columns present the static version of the panel models (EM_COLSh and EM_CREh where *h* means historical) estimated by OLS and RE. The rest of the table results show the dynamic version of the models by including the lagged effect of metering enhancement (L.ln_totex_metering or $Ln(y_{i,t-1})$).

[7] This is independent of the way the models are estimated by OLS, RE or FE.
[8] See Baltagi, B. (2005). Econometric Analysis of Panel Data. John Wiley & Sons Ltd. Third Edition, or Arellano, M. (2003). Panel Data Econometrics. Advanced Texts in Econometrics. Oxford University Press.
[9] See Cameron, C. and Trivedi, P (2005). Microeconometrics, Methods and Applications. Cambridge University Press

## Table 3 – Historical Static and Dynamic Panel Resutls

Historical Static and Dynamic Panels

| | EM_COLSh<br>b/se | EM_CREh<br>b/se | EM_OLSh<br>b/se | EM_REh<br>b/se | EM_FEh<br>b/se | EM_ABh<br>b/se | EM_BB1h<br>b/se |
|---|---|---|---|---|---|---|---|
| ln_meters | 0.457** | 0.457** | 0.213 | 0.209* | 0.206** | 0.223** | 0.309*** |
| | (0.20) | (0.20) | (0.14) | (0.12) | (0.09) | (0.09) | (0.10) |
| L.ln_totex_metering | | | 0.725** | 0.640** | 0.180 | 0.049** | 0.614** |
| | | | (0.26) | (0.29) | (0.13) | (0.02) | (0.23) |
| Constant | -0.902 | -0.902 | -0.083 | -0.137 | 0.412 | 0.565 | -0.552** |
| | (0.67) | (0.67) | (0.16) | (0.33) | (0.28) | (0.41) | (0.24) |
| R2_Adjusted | 0.0556 | | 0.7419 | | 0.3328 | | |
| R2_overall | | 0.1094 | | 0.7594 | 0.6364 | | |
| Observations | 124 | 124 | 102 | 102 | 102 | 82 | 102 |

Source: Economic Regulation, Thames Water. Note: AB=Arellano-Bond Estimator, BB=Blundel-Bond Estimator. All models include Time Dummies Year Effects.
* p<0.10, ** p<0.05, *** p<0.01

2.30 As mentioned before, dynamic model estimations by OLS, within and GLS estimators, for the pooled (EM_OLSh), Random Effects (EM_REh) and Fixed Effects (EM_FEh) models are by construction biased and inconsistent but provide information about the magnitude of the bias on the lagged dependent variable and its effects on other controls. The potential *upward* bias of the dynamic effect is provided by the OLS estimation (0.725) whereas the within estimation of the FE model provides the *downward* bias limit (0.180). Therefore, when correcting for the endogeneity issues the true value of the dynamic effect should be somewhere in between or not significantly higher than 0.725 or lower than 0.180. With respect to the other coefficients such as the number of meters installed (e.g., selective and optant, ln_meters) the results suggest that in the static models the estimation is biased with respect to the dynamic Random Effect model by a significant proportion (e.g., $\widehat{\beta_{Meters,RE\ Static}} = 0.457\ v.\ \widehat{\beta_{Meters,RE\ Dynamic}} = 0.209$). This result is an example of what Bond (2002) suggests on the effects that the omission of dynamic components in the models could have on other parameters of the model.

2.31 In order to tackle the endogeneity issue presented in the previous models of Table 3, the ***Arellano-Bond Estimator*** (columns EM_ABh and EM_BB1h) uses instruments to correct the issue by using information on the previous time periods, *t-2*, *t-3*, etc. By using these instruments and following the appropriate test such as the Hansen test of instruments validation, we conclude that the instrument validation hypothesis is not rejected for model EM_ABh (Arellano-Bond). However, the coefficient of the lagged variable seems to be unrealistic or too low compared to what we have observed in previous results. This could be explained potentially by weak instruments that could be caused by finite-sample biases[10]. The ***Blundell-Bond estimator*** (column EM_BB1) or system GMM provides an alternative to this issue by incorporating more informative moment conditions that reduces the bias dramatically on the last two columns of the table with more sensible results on what is expected (e.g., see form example the estimated results on $\widehat{\beta_{Meters,BB\ Dynamic}} = 0.309$ and $\widehat{\beta_{lagged,BB\ Dynamic}} = 0.614$).

[10] For a discussion about finite-sample bias see Blundell, R., and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models, Journal of Econometrics, 87, pp. 115-143.

2.32 These results based on historical data, show how important it could be to use a more appropriate approach on estimating a consistent enhancement type of expenditure that allows us to control for the irregularities or lumpiness that are presented within and between companies. In this metering example, the results indicate a statistically significant effect of the lagged dependent variable in almost all the different specifications presented in Table 3 (in particular, the unbiased model EM_BB1h last column on Table 3). This empirical evidence suggests how important is to control for this dynamic effect in this type of expenditure because omitting this dynamic control could also have significant effects on other coefficients and therefore on efficiency scores and allowances. This omission can end up in underestimated predictions as the one currently used in PR19, jeopardizing the performance of companies and service to customers. The next section will present a similar analysis but using the future or forward-looking values of the panel to mimic the approach used in PR19 and to highlight the potential areas to be improved and the difference in the econometric results.

Forward-looking Dynamic Panel

2.33 Similarly, as explained in the previous section, we carried out the same estimation procedures to understand the dynamic effect of the lagged dependent variable under a dynamic panel approach using the values that companies have put in the Business Plans or forward-looking values on metering enhancement for the period 2017-18 to 2024-25 (same period used by the PR19 models). This will allow us to have more consistent and comparable results between the two approaches.

2.34 Table 4 below shows the results of the static panel models OLS and RE (EM_COLS and EM_CRE) and all the biased estimations of the dynamic approaches OLS, RE and FE (see columns EM_olsD, EM_RED and EM_FED).

Table 4 – Static and Dynamic Panels Resutls (Biased)

Static and Dynamic Panels (stage 1: Biased Estimations)

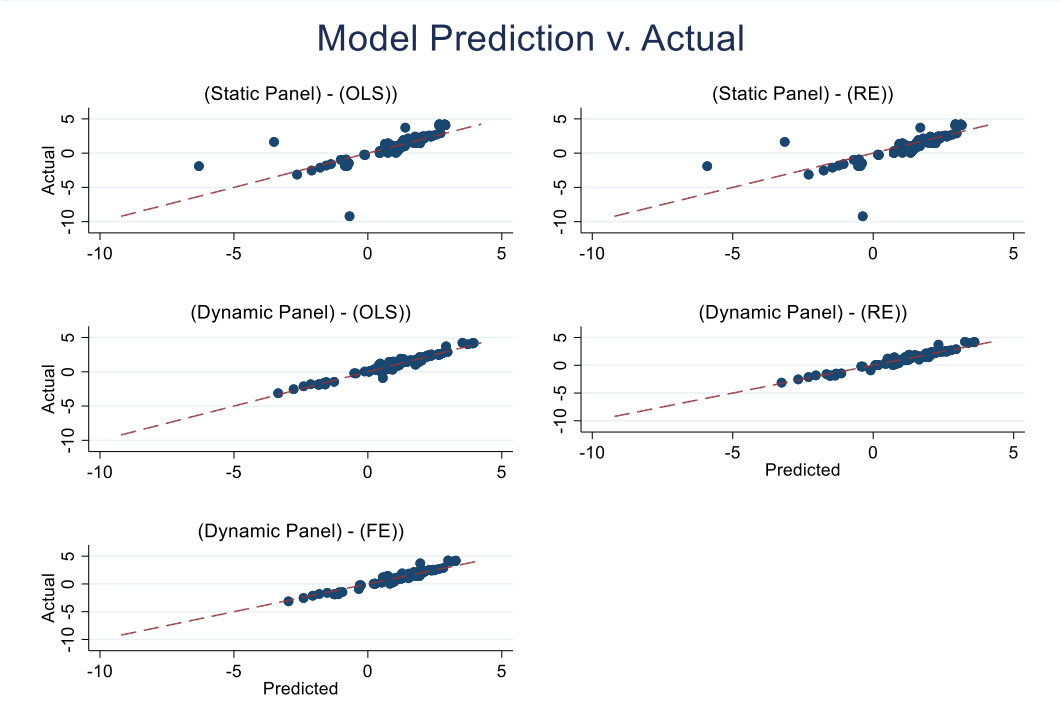
| | EM_COLS<br>b/se | EM_CRE<br>b/se | EM_olsD<br>b/se | EM_RED<br>b/se | EM_FED<br>b/se |
|---|---|---|---|---|---|
| ln_meters | 0.784*** | 0.772*** | 0.484*** | 0.645*** | 0.681*** |
| | (0.17) | (0.21) | (0.12) | (0.13) | (0.15) |
| L.ln_totex_metering | | | 0.576*** | 0.346** | 0.215* |
| | | | (0.15) | (0.16) | (0.10) |
| Constant | -0.817 | -0.503 | -0.760*** | -0.884*** | -0.831*** |
| | (0.47) | (0.66) | (0.15) | (0.19) | (0.26) |
| R2_Adjusted | 0.5903 | | 0.9581 | | 0.8518 |
| R2_overall | | 0.5934 | | 0.9511 | 0.9400 |
| Observations | 135 | 135 | 115 | 115 | 115 |

Source: Economic Regulation, Thames Water.
* p<0.10, ** p<0.05, *** p<0.01

2.35 The biased estimations of OLS and FE under the dynamic setting, suggest that the estimated effect of the lagged dependent variable of enhancement metering program over AMP7 ranges between 0.215 and 0.576. In addition, these results also suggest a potential bias on the scale parameter of numbers of meters installed.

2.36 Figure 5 depicts the 45-degree line between fitted and actual values. The top two charts warn us about potential outliers on the static panel results even after removing some potential outliers identified in Figure 4. These observations might cause potential issues on the estimations.

Figure 5 – Model Prediction v. Actual (Biased Models)




Source: Economic Regulation, Thames Water.

2.37 However, the dynamic biased models OLS, RE and FE (see columns EM_olsD, EM_RED and EM_FED) successfully control for these remaining potential observations and the 45-degree line produces more accurate results (see Figure 5 dynamic panel models). The Pearson correlation of these models illustrated in the charts presented in Figure 5 are calculated in Table 5:

## Table 5 – Correlations across models

| | ln_tot~g |
|---|---|
| ln_totex_m~g | 1.0000 |
| xb_cols | 0.6264 |
| xb_cre | 0.6264 |
| xb_olsD | 0.8877 |
| xb_reD | 0.8868 |
| xb_feD | 0.8435 |

Source: Economic Regulation, Thames Water

2.38 The improvement in the correlation coefficient between the static and dynamic models is substantial and we can easily see it on the predicted and actuals correlations as these are increased from 0.62 to around 0.88. Nevertheless, all these estimated results are significantly biased as explained in the previous section, regarding OLS, FE or RE effects models. To overcome the bias estimations presented in Table 4, we use the ***Arellano-Bond*** estimator (EM_AB) and the ***Blundell-Bond*** estimator (EM_BB). The results of these dynamic models are presented in Table 6.

## Table 6 – Dynamic Panels Resutls (Unbiased)

Dynamic Panels (stage 2: Unbiased Estimations)

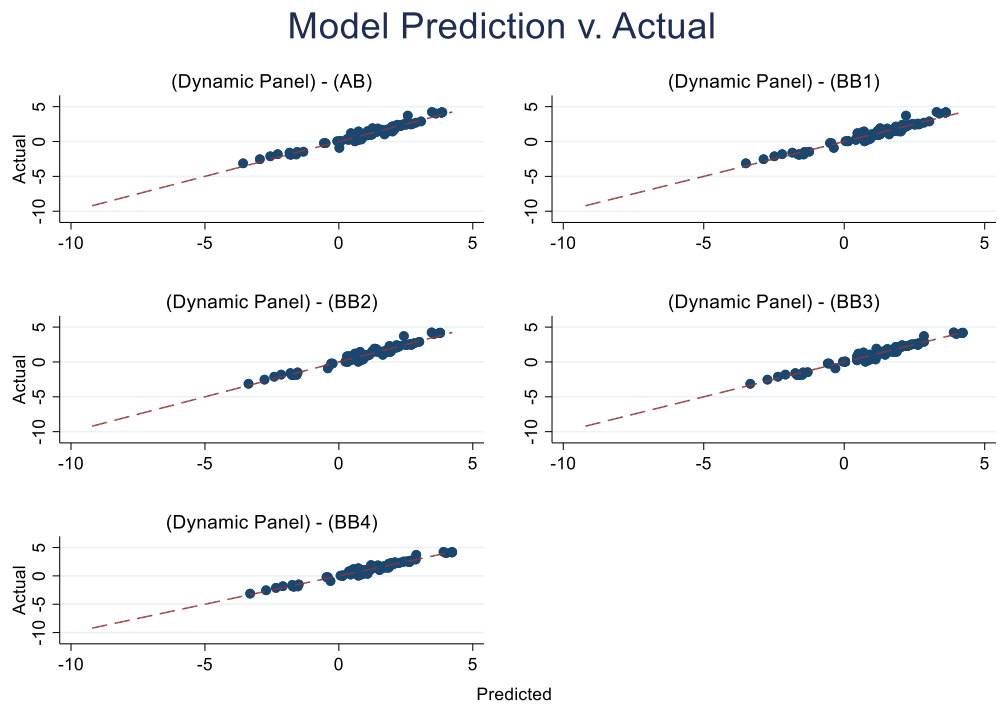
| | EM_AB<br>b/se | EM_BB1<br>b/se | EM_BB2<br>b/se | EM_BB3<br>b/se | EM_BB4<br>b/se |
|---|---|---|---|---|---|
| L.ln_totex_metering | 0.437* | 0.298* | 0.281* | 0.286* | 0.292** |
| | (0.24) | (0.15) | (0.15) | (0.14) | (0.14) |
| ln_meters | 0.638*** | 0.728*** | 0.690*** | 0.707*** | 0.683*** |
| | (0.16) | (0.12) | (0.12) | (0.12) | (0.11) |
| ln_density | | | 0.183* | | 0.080 |
| | | | (0.10) | | (0.08) |
| ami | | | | 0.806*** | 0.689*** |
| | | | | (0.19) | (0.16) |
| amr | | | | 0.144 | 0.096 |
| | | | | (0.11) | (0.10) |
| Constant | -0.983*** | -1.051*** | -2.258*** | -1.115*** | -1.603** |
| | (0.18) | (0.15) | (0.74) | (0.19) | (0.61) |
| AB_Autocorr_order2 | -1.605 | | | | |
| AR1_p_value | | 0.2712 | 0.3286 | 0.2634 | 0.3000 |
| AR2_p_value | | 0.141 | 0.142 | 0.130 | 0.124 |
| Hensen_Test_Overid~f | | 0.995 | 0.936 | 0.984 | 0.952 |
| Number_Instruments | | 6.000 | 7.000 | 8.000 | 9.000 |
| N | 98.000 | 115.000 | 113.000 | 115.000 | 113.000 |

Source: Economic Regulation, Thames Water. Note: AB=Arellano-Bond Estimator, BB=Blundel-Bond Estimator
* p<0.10, ** p<0.05, *** p<0.01

2.39 The first two columns in Table 6 present the results for the EM_AB and EM_BB1 models. Both models show a significant effect of the lagged dependent variable. The scale effect (e.g., meters) is slightly lower for the dynamic panels presented in Table 4 (e.g., EM_RED and EM_FED) when compared to the results in Table 6, indicating a slightly marginal downward bias. The lagged dependent variable suggests a significant

improvement on the estimations presented on Table 6 ranging between 0.215 and 0.576 as suggested by the OLS and FE models presented in Table 4.

Figure 6 – Model Prediction v. Actual (Unbiased Models)

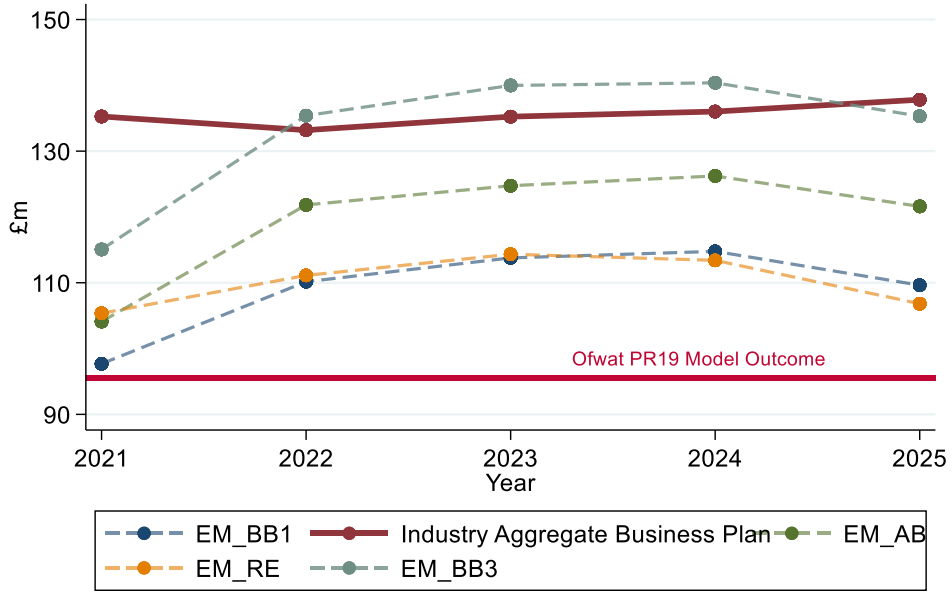



Source: Economic Regulation, Thames Water. Note: AB=Arellano-Bond Estimator, BB=Blundel-Bond Estimator

2.40 Figure 6 presents the correlation between actual and predicted outcomes from all the unbiased dynamic panel estimations. For all models, the correlations are between 0.80 and 0.87, presenting a consistent and better fit of the models. We have also extended the analysis by adding density (see model EM_BB2) and technology (see model EM_BB3) as other potential drivers that could help to explain the metering expenditures. We construct a technology meter variable to mitigate the effect of omitting this driver and to move away from the PR19 assumption of not including any technology parameter in the models. For example, model EM_BB3 introduces the effect of technology represented by AMI and AMR meters[11]. All these unbiased estimations are examples on how the model predictability is improved by controlling for the right dynamic patterns. Dynamic panels help us to avoid econometric techniques used in previous price reviews when modelling enhancement, such as moving average calculations that smooth the original data but adversely affect the consistency of the totex or botex identities (see PR14 or PR19 assessment as examples).

[11] Where AMI=Advance Metering Infrastructure and AMR=Automatic Meter Reading.

Figure 7 – Industry Prediction Expenditure Aggregates on Enhancement Metering




Source: Economic Regulation, Thames Water. Note: Ofwat outcomes are divided by the total number of years in AMP7.

2.41 Finally, in terms of prediction, the outcomes of model EM_BB3 at the industry level are in line of what is being requested by the industry in the business plans (see Figure 7), reflecting the relevance of the meter technology on the final outcomes of the model. Moreover, the effect of technology suggests a statistically significant effect on the enhancement metering expenditures (see AMI coefficient, for instance) which supports and provides empirical evidence on controlling for this cost driver and its potential as a cost driver. Figure 7 also puts in context the prediction of different models presented across this section. It shows the potential underestimation predicted by the cross-sectional model (see horizontal red line in Figure 9)[12]. At PR19 the weakness of the modelling approach was recognised and additional allowances provided through the deep dive approach. The benefit of an improved modelling approach is that the subjective and time-consuming deep dive can be avoided.

2.42 The rest of the models such as EM_BB1 or EM_AB which are the product of consistent estimations using dynamic panel models also show significant higher outcomes in terms of the allowances predicted and the variance of the efficiency scores. These results suggest that the modelling approach can be substantially improved if the right econometric methodology is chosen. The results also suggest that significant econometric biases are having a substantial material impact on the industry regarding metering enhancement expenditure. We believe that most of the enhancement expenditure (when appropriate) should be modelled or at least explored using dynamic panel data models as an alternative to what has been proposed in the past PR14 and PR19 reviews. This approach has the potential to improve the assessment of these

[12] We have divided the total allowance produced by this model by five, to represent the yearly average produced by this model, as the cross-sectional model can't produce a yearly prediction being a panel model.

expenditures and to bring more reliable and objective outcomes, that will be beneficial for customers, the environment and the long-term view of efficient costs. The dynamic approach is a promising technique in controlling for the irregularities of enhancement expenditures.

# D Example 2: Growth

## Descriptive Statistics on Growth (New Development & Connections)

2.43 This section is based on enhancement growth or new development & connections expenditure therefore we only focus on the water industry. The results of this section are based on the initial PR19 approach of modelling growth as a separate type of expenditure case, before it was added into the botex models. This analysis was developed before the IAP, so the results only cover the period 2011-12 to 2016-17.

2.44 As an introduction we explore all the types of expenditure enhancements. During this review we find that there are more than ten water enhancements activities in the water industry. Across them there is a significant within company variation and between companies variation in the industry. In particular, the activities that register most of these variations are new developments & connections (see table 7 below) and supply-demand activities.

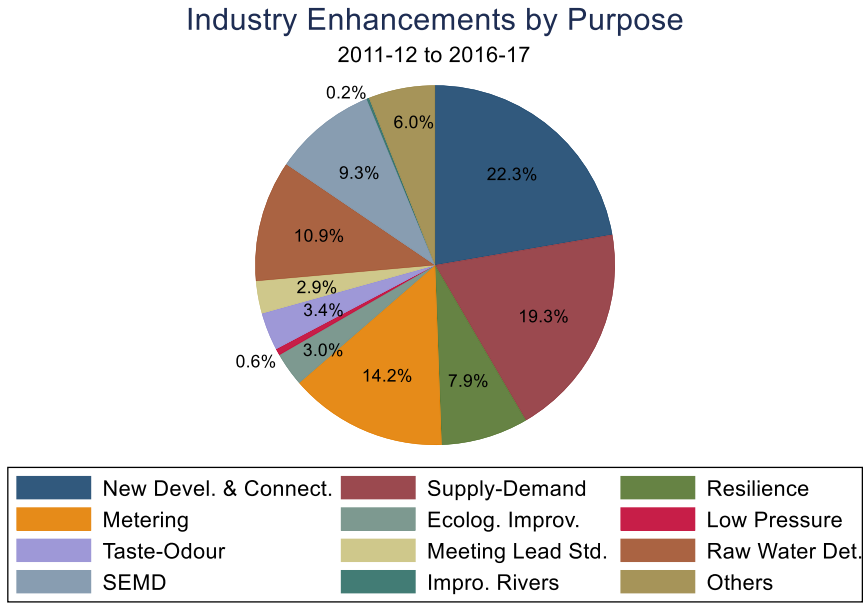
Table 7 – Growth Expenditures Descriptive Statistics

| Variable | | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| Industry Agg. New Dev. & Conn. | overall | 167.19 | 34.76 | 117.60 | 207.99 |
| | between | | 9.45 | 166.81 | 207.99 |
| | within | | 34.53 | 117.98 | 208.37 |

2.45 In Figure 8, new developments & connections represents the highest proportion (22.3%) of capex enhancement across the industry over the period 2011-12 to 2016-17, followed by supply-demand (19.3%) and metering (14.2%). With nearly 3% of enhancement expenditures, meeting lead standards has one of the lowest proportions across the industry followed by addressing low pressure (0.6%) and improvements to river flows (0.2%) activities. There are other relevant expenditures in the middle of the enhancement activities that represent a significant level of investment and within variation across the industry such as raw water deterioration (10.9%).

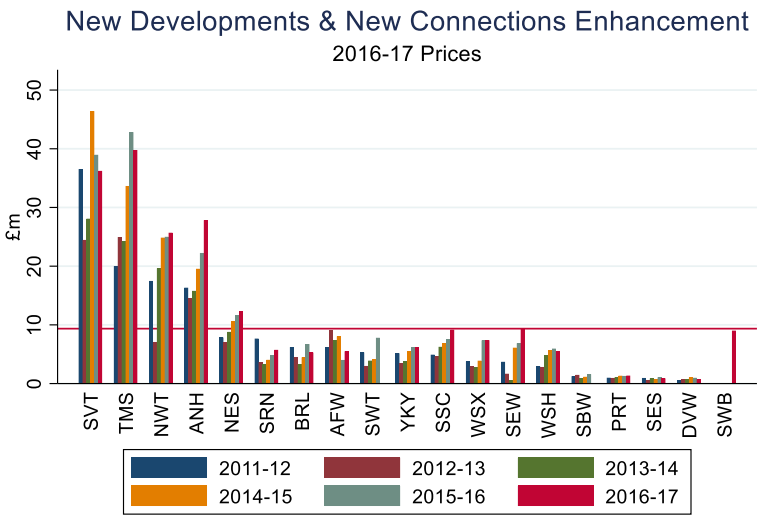
2.46 Figure 9 shows the investment levels on new developments & connections for each company. It shows how lumpy the investment is between 2011-12 and 2016-17 across the industry and within companies. For example, SVT investment levels are quite volatile with a min and max of £24m and £46m, respectively. Similarly, TMS has a min of £20m and a max of £43m, whereas the industry has an average investment across the period of £9m per year. On the other hand, as a small company example with nearly zero levels of investment, PRT has a min and max of £0.95m and £1.35m, respectively.

Figure 8 – Enhancements Cases by purposes 2011-12 to 2016-17




Source: Economic Regulation, Thames Water

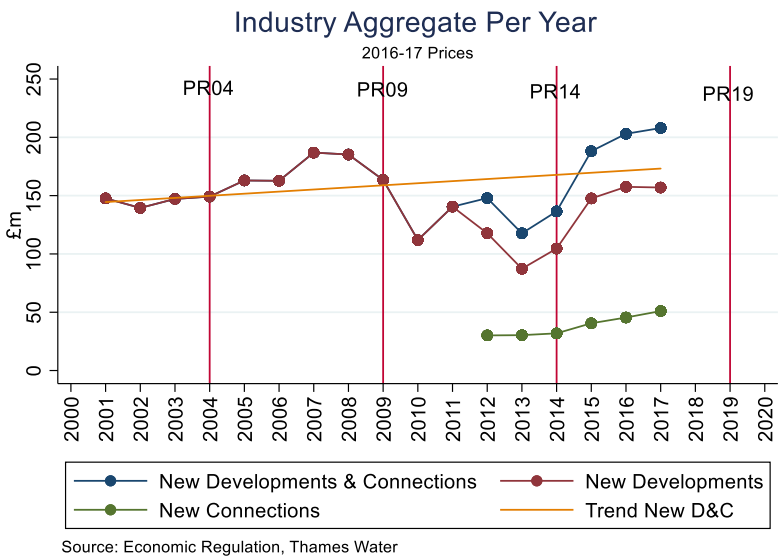
Figure 9 – Enhancements Cases by purposes 2011-12 to 2016-17




Source: Economic Regulation, Thames Water

2.47 Figure 10 shows the industry aggregated levels of investment in new developments & connections from 2001 to 2017. At PR09, the financial crisis could explain the cyclical pattern and significant drop in the trend that persists between PR09 and PR14. Onwards, the level of investment moved back towards the long-term trend. Exogenous changes at macroeconomic level could affect the behavioural level of investment suggesting a dynamic pattern of investment in the industry as another potential effect on growth.

Figure 10 – Enhancements Cases by purposes 2011-12 to 2016-17




## PR19 Ofwat's Growth Models at IAP

2.48 Table 8 reproduces the enhancement models OE4 and OE5 on new developments & connections presented at the IAP. In these Static Panel Data Random Effects models the variables are transformed to a moving average of three years distorting the actual data observed on: (i) the dependent variable (e.g., £m capex on new developments & connections) and the explanatory variables (ii) *population* served and total number of *households and non-households*. The main reason why this approach was used was to try to smooth the lumpiness of the expenditures but without really controlling for it.

2.49 This transformation is adopted to be consistent with the empirical specification. However, there are some questions to be considered: Are the coefficients biased? Is there any empirical misspecification? Do we have time inconsistency effects on the totex outcomes by using a moving average approach on capex? How robust is the predictive power of these models?

Table 8 – Ofwat's Growth Models at PR19 (IAP)

| | OE4<br>b/se | OE5<br>b/se |
|---|---|---|
| Ln_Pop_Served_Mov_~g | 1.061*** | |
| | (0.10) | |
| Ln_HH_NoHH_Mov_Avg | | 1.040*** |
| | | (0.12) |
| Constant | -6.498*** | -0.242 |
| | (0.79) | (0.24) |
| R2_overall | 0.8231 | 0.8149 |
| Joint_Statistic | 104.6084 | 71.4339 |
| RESET_P_value | 0.228 | 0.736 |
| Corr_company_effect | 0.866 | 0.862 |
| PooledvsRE_P_value | 0.000 | 0.000 |
| Autocorre_P_value | 0.000 | 0.001 |
| Hausman_P_value | 0.000 | 0.948 |
| Observations | 70.000 | 70.000 |

* p<0.10, ** p<0.05, *** p<0.01

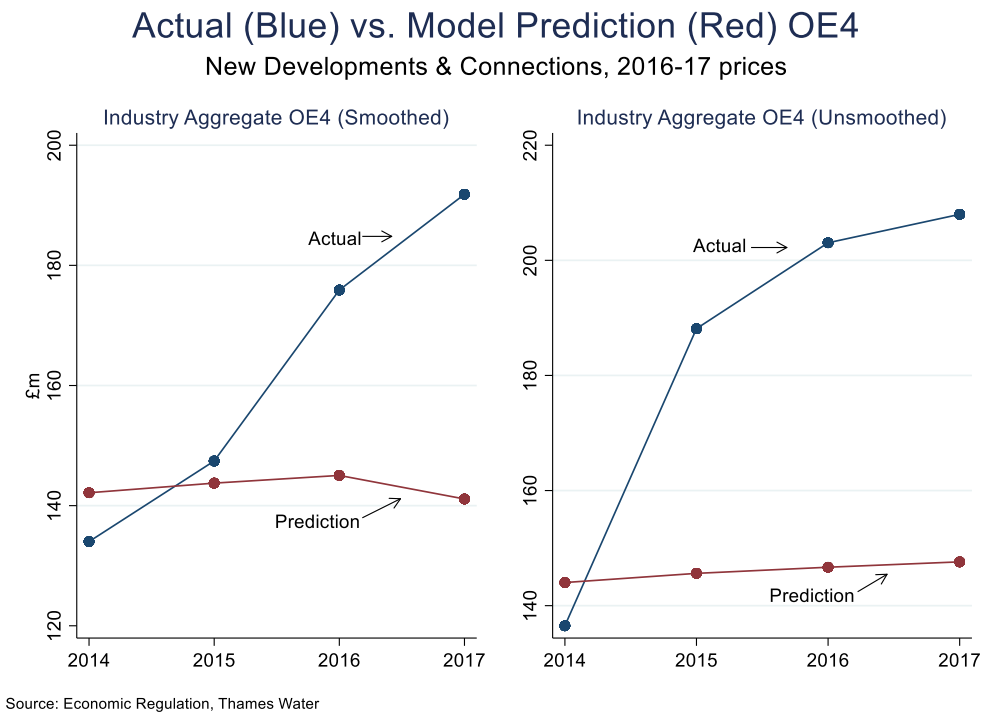
2.50 Figure 11 explores the predictive power of these two models (i.e., OE4 and OE5) at the industry level v the actuals. We follow an analysis that proposes a different econometric approach that could help to provide insights about the bias of the coefficients and predictive power of the models in the next paragraphs on dynamic models.

IAP models predictions

2.51 Model OE4 uses a moving average version of the scale driver *total population served* in this enhancement investment model. The charts in Figure 11 depict the industry aggregates of the actual levels and model OE4 predictions in smoothed and unsmoothed versions to visualise the outcomes of the different approaches of the models.

2.52 The prediction seems to be poor versus what is observed in the industry aggregates. For the smooth version the industry gap is on average £19m per year, whereas for the unsmoothed version the gap is around £38m per year. This gap between the model and the actual spend is significant and increasing suggesting there are potential efficiency issues with the models and potentially the models misrepresent the investment levels in the industry. This potential is explored further below.

## Figure 11 - Growth Models at PR19 (IAP)




## Dynamic Panel Models on Enhancement Growth

2.53 The first two columns on Table 9 show the original random effect static panel data model OE4 smoothed version (OE4_Smt (RE)) and its unsmoothed version results (OE4_Un_ (RE)). The estimated coefficient seems to be very similar for all of the approaches taken. However, there is a loss of observations, a potential misspecification and biased estimations as well as some time inconsistency with the totex definition under the static models.

## Table 9 – Ofwat's Growth Models at PR19 (IAP)

| | Static Panel Data Models | | Dynamic Panel Data Models (Unsmooth) | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *Dep. Variable:* ***Ln_New_Developments_Connections_(t)*** | **OE4_Smt (RE) (Ofwat)** | **OE4_Un (RE)** | **OE4 (OLS)** | **OE4 (FE)** | **OE4 (RE)** | **OE4 (Arellano-Bond)** | **OE4 (Arellano-Bond)** | **OE4 (Blundell-Bond)** | **OE4 (Blundell-Bond)** |
| Ln_Population_Served_Smoothed | 1.061*** | | | | | | | | |
| | (0.10) | | | | | | | | |
| Ln_Population_Served_Unsmoothed | | 1.045*** | 0.326*** | 11.783*** | 0.475*** | 11.287*** | -3.297 | 0.343 | 0.517** |
| | | (0.10) | (0.11) | (3.61) | (0.11) | (3.20) | (5.56) | (0.20) | (0.19) |
| Ln_New_Developments_Connections_(t-1) | | | 0.690*** | 0.190*** | 0.545*** | 0.366** | 0.371* | 0.662*** | 0.504*** |
| | | | (0.10) | (0.05) | (0.09) | (0.15) | (0.20) | (0.17) | (0.16) |
| Constant | -6.498*** | -6.358*** | -1.959** | -89.108*** | -2.877*** | | | -2.037 | -3.283** |
| | (0.79) | (0.75) | (0.69) | (27.66) | (0.72) | | | (1.26) | (1.29) |
| R2_overall | 0.8231 | 0.7666 | 0.8762 | 0.7618 | 0.8741 | | | | |
| Time dummies | No | No | No | No | No | No | Yes | No | Yes |
| AR1 (p-value) | | | 0.8515 | 0.7869 | | 0.183 | 0.109 | 0.141 | 0.071 |
| AR2 (p-value) | | | 0.6553 | 0.1004 | | 0.296 | 0.282 | 0.269 | 0.200 |
| Hansen Test of Overidentification | | | | | | 0.389 | 0.359 | 0.099 | 0.374 |
| Number of Instruments | | | | | | 11 | 15 | 7 | 11 |
| Observations | 70 | 107 | 89 | 89 | 89 | 70 | 70 | 88 | 88 |

* p<0.10, ** p<0.05, *** p<0.01

**Source:** Economic Regulation, Thames Water

2.54 The dynamic models in Table 9 are all unsmoothed. Columns 3 to 5 labelled as OE4 (OLS), OE4(FE) and OE4(RE), respectively, estimate the dynamic version of the static model OE4 (e.g., OE4_Smt (RE) (Ofwat) and OE4_Un (RE)). The first three *dynamic models* are estimated by OLS, within and GLS estimators, for the pooled, Fixed Effects (FE) and Random Effects (RE) models. These models by construction are biased and inconsistent as it was explained in the metering example. However, they are useful as a reference as they provide information about the magnitude of the bias on the lagged dependent variable and its effects on other controls. The potential upward bias of the dynamic effect is provided by the OLS estimation (0.690) whereas the within estimation provides the downward bias limit (0.190). Therefore, when correcting for the endogeneity issue the true value of the dynamic effect should be somewhere in between, i.e. not significantly higher than 0.690 or lower than 0.190. With respect to the other coefficient of *population served* the results suggest that in the static models the estimation is biased with respect to the dynamic random effect model approach by a significant proportion (e.g., $\widehat{\beta_{Pop}} = 1.045\ v.\widehat{\beta_{Pop}} = 0.475$). This result illustrates what Bond (2002) suggests. The ***Arellano-Bond estimator*** (columns 6 and 7) uses instruments to correct the endogeneity issue by using information on the previous time periods, *t-2*, *t-3*, etc.

2.55 The Hansen test of instruments validation is not rejected for model OE4 (Arellano-Bond). However, the coefficient of the population served seems to be far from what is expected. This could be explained by the weak instruments used in the ***Arellano-Bond estimator*** that could cause large finite-sample biases (see Blundell and Bond (1998b)). The Blundell-Bond estimator or system GMM provides an alternative to this issue by incorporating more informative moment conditions that reduces the bias dramatically on the last two columns of the table with more sensible results (e.g., $\widehat{\beta_{Pop}} = 11.287\ or - 3.27\ v.\widehat{\beta_{Pop}} = 0.517$, for example).

<u>Dynamic Panel Models Predictions for Model O4</u>

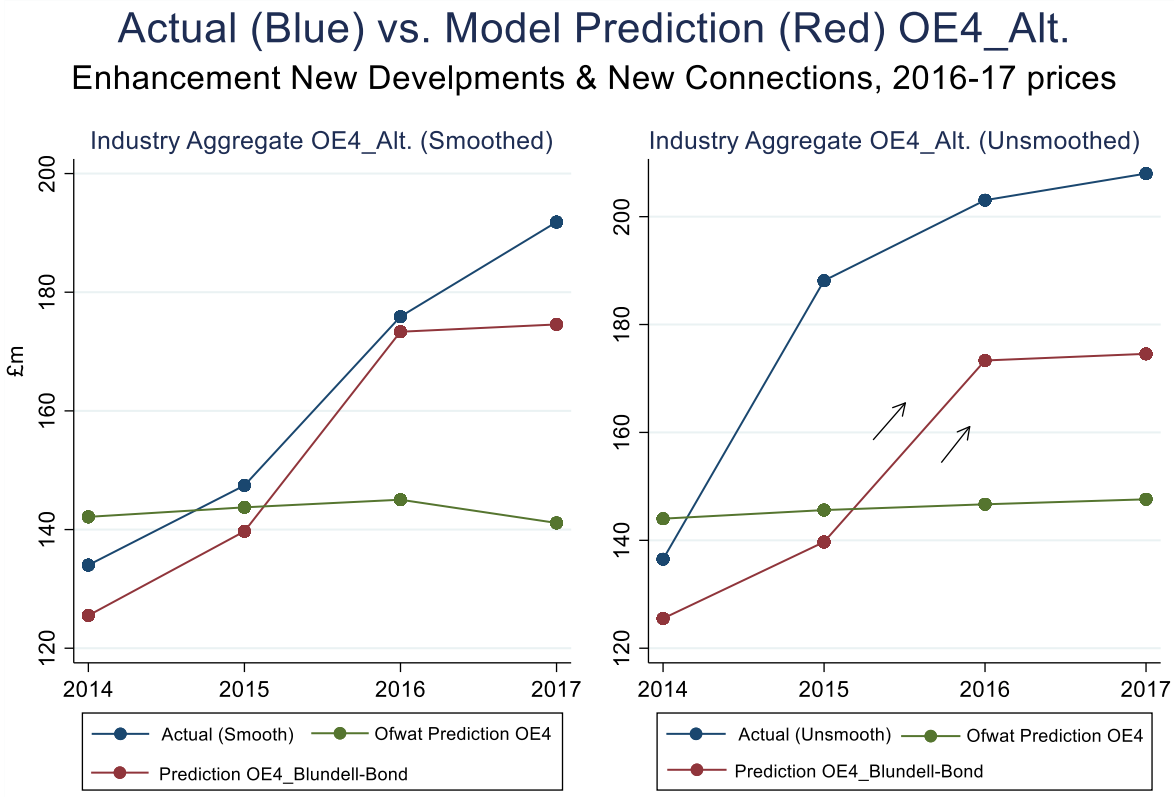
2.56 Figure 12 presents the results of model OE4 unsmoothed Static Panel Data model at the industry level against the Dynamic Panel approach (e.g., Blundell-Bond estimator) with time dummies and the actual unsmoothed aggregate levels of enhancements. Several advantages are found in this dynamic approach.

2.57 The number of observations lost in the estimation procedure is reduced substantially against the smoothed approach.

2.58 The new developments & connections enhancement expenditure is time consistent with the botex and totex definitions mitigating any misallocation of time on the totex cost allowances for AMP7.

2.59 The predictive power of the dynamic model improves the prediction and trend outcomes compared to what the static model produces. This is due to the correction of the specification and bias in the model (e.g., $\widehat{\beta_{Pop}} = 1.045\ v.\widehat{\beta_{Pop}} = 0.517$). For example, in 2017 the gap between the actual and the static model is £60m whereas between the actual and the dynamic the difference is £33m.

Figure 12 – Ofwat's Growth Models at PR19 (IAP)




Source: Economic Regulation, Thames Water

## Model OE5

2.60 Similarly, the results for model OE5 are also improved under a dynamic approach. The static panel data approach yields an estimated coefficient for total number of household and non-household new connections of 1.040 for the smoothed version of the data (e.g., model OE5) versus a 0.950 in the unsmoothed version. Under the dynamic panel model this coefficient is quite different (ranging from -0.245 to 0.749) from the static approach, suggesting a potential biased estimation in this scale driver.

2.61 The dynamic parameter of the model, New_Developments_Connections (*t-1*) ranges between 0.230 (FE) and 0.698 (OLS) as expected and suggested by Bond (2002). These results are showing that the estimated coefficient of Number_HH_NoHH is potentially affected/biased. In order to reduce this issue, the Arellano-Bond estimator provides a first approximation to resolve this empirical problem. However, given the lumpiness of the data the use of instruments at the levels in these models is affected and it could produce unrealistic estimations (e.g., -0.245.

## Table 10 – Ofwat's Growth Models at PR19 (IAP)

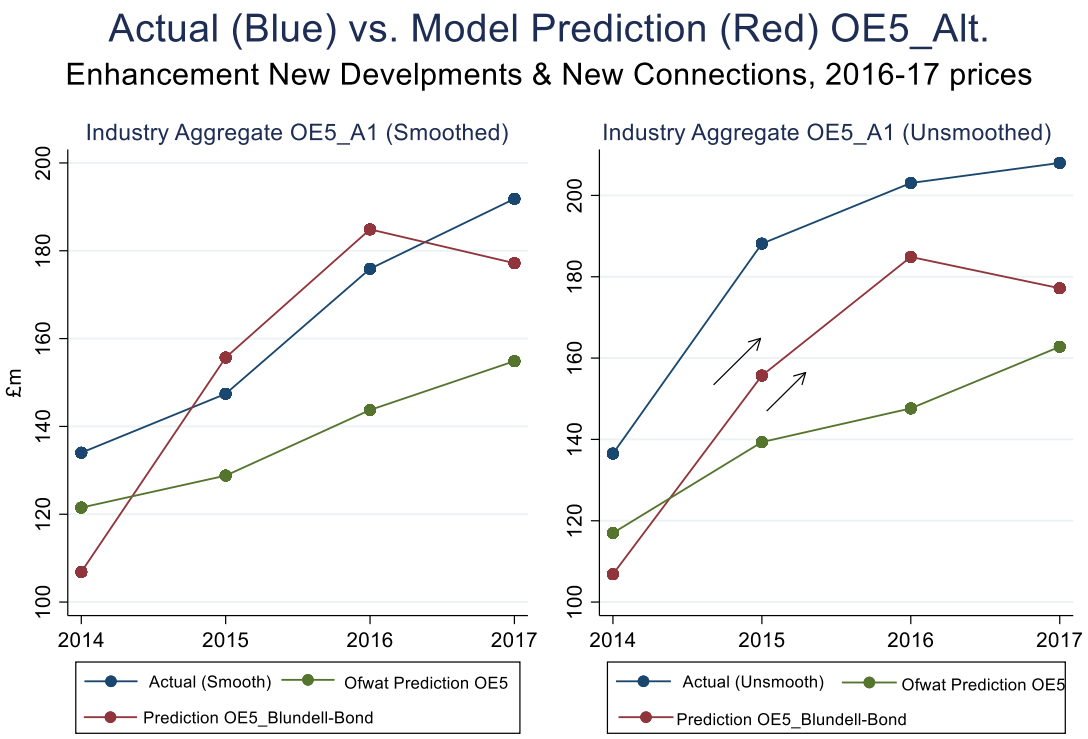
| | Static Panel Data Models | | Dynamic Panel Data Models | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Dep. Variable: ***Ln_New_Developments_ Connections_(t)*** | **OE5_Smt (RE) (Ofwat)** | **OE5_Un (RE)** | **OE5 (OLS)** | **OE5 (FE)** | **OE5 (RE)** | **OE5 (Arellano-Bond)** | **OE5 (Arellano-Bond)** | **OE5 (Blundell-Bond)** | **OE5 (Blundell-Bond)** |
| Ln_Number_HH_NonHH_Smoothed | 1.040*** | | | | | | | | |
| | (0.12) | | | | | | | | |
| Ln_Number_HH_NoHH_Unsmoothed | | 0.950*** | 0.316*** | 0.716** | 0.486*** | 0.749*** | -0.245 | 0.660 | 0.582*** |
| | | (0.12) | (0.09) | (0.08) | (0.11) | (0.27) | (0.29) | (0.14) | (0.14) |
| Ln_New_Developments_ Connections_(t-1) | | | 0.698*** | 0.230*** | 0.525*** | 0.395*** | 0.355*** | 0.338*** | 0.405*** |
| | | | (0.08) | (0.33) | (0.09) | (0.12) | (0.20) | (0.10) | (0.12) |
| Constant | -0.242 | -0.079 | -0.046 | -0.029 | -0.085 | -0.347 | 1.733** | -0.095 | -0.148 |
| | (0.24) | (0.23) | (0.08) | (0.50) | (0.13) | (0.42) | (0.73) | (0.16) | (0.15) |
| R2_overall | 0.8149 | 0.7571 | 0.8769 | 0.8244 | 0.8723 | | | | |
| Time dummies | No | No | No | No | No | No | Yes | No | Yes |
| AR1 (p-value) | | | 0.7525 | 0.5093 | | 0.125 | 0.100 | 0.133 | 0.090 |
| AR2 (p-value) | | | 0.6564 | 0.0825 | | 0.337 | 0.216 | 0.341 | 0.303 |
| Hansen Test of Overidentification | | | | | | 0.241 | 0.689 | 0.245 | 0.747 |
| Number of Instruments | | | | | | 11 | 15 | 16 | 20 |
| Observations | 70 | 107 | 89 | 89 | 89 | 70 | 70 | 88 | 88 |

* p<0.10, ** p<0.05, *** p<0.01

**Source:** Economic Regulation, Thames Water

2.59 The ***Blundell-Bond estimator*** or ***system GMM*** exploit better the initial condition information in generating efficient estimators for dynamic panel data models where the number of time series observations is small (see Baltagi (2005), p. 148). Hence, a result between 0.338 and 0.405 is sensible, whereas for the scale driver a more likely unbiased estimation ranges between 0.582 and 0.660, versus the static approach that ranges between 0.950 and 1.040.

## Figure 13 – Growth Models at PR19 (IAP)




Source: Economic Regulation, Thames Water

2.60 Figure 13 presents the results of Ofwat's OE5 unsmoothed Static Panel Data model at the industry level against the dynamic panel model results (e.g., Blundell-Bond estimator) with time dummies and the actual unsmoothed aggregate levels of enhancements.

2.61 The predictive power of the dynamic model improves the outcomes and trends compared to what the static model produces. For example, the gap between the aggregate yearly average actual and the equivalent in the static model is £42m whereas the difference with respect to the dynamic model outcomes is £28m.

2.62 The dynamic panel data model estimated by the Blundell-Bond estimator provides better predicted outcomes at the aggregate industry level. These results are similar to the extended version of model OE4.

2.63 We are not presenting efficiency scores outcomes at this stage, but this is something we have calculated and compared between models. Given that the aim of the paper is to show the relevance of the dynamic technique we leave the efficiency analysis for another scenario.

## E Conclusion

- Specification and Estimation: Lumpy enhancement investments are common capital expenditure data series in utilities (see Peck (1974)). Static Panel Data estimations are likely to be biased on enhancement activities if the right methodology is not taken into account. The static panel or the cross-section approach v the dynamic panel models presented in this paper suggest that the current allowances in different enhancment cases might be underestimted putting at risk companies ability to properly serve their customers. Hence, a dynamic panel data methodology provides a better approach for these types of investment by reducing issues of misspecification, inconsistency and bias. For instance, the dynamic extension model estimation results for models OE4 and OE5 in the growth example, suggest the potential bias of the estimated scale coefficients in the static panel results, suggesting further exploration of the dynamic models. Furthermore, the predictive power of models OE4 and OE5 are substantially improved under the dynamic panel approach at the industry and company level. The estimated coefficients between the static and dynamic approaches suggest significant differences that are crucial for the forecast calculation of enhancements in future AMPs.

- Time inconsistencies: To tackle the lumpiness of enhancement, Ofwat has used CEPA PR14 methodology of smoothing the series by a moving average approach of the financial data in the growth case when assessed at the IAP. In addition, Ofwat has followed CMA suggestions on the smoothed process of not only the financial data but also the non financial data in order to be consistent with the econometric enhancement model (e.g., consistency in the capex enhancement models). However, this approach could produce two issues or unknowns:

- An unknown time effect inconsistency issue on the totex definition and final cost allowances in AMP7 (e.g., $Totex_t = Opex_t + Capex_t$, v. an average of $\overline{Capex}_3 = \frac{\sum_{i=1}^{3} Capex_i}{3}$ ).
- There is no evidence that by changing the series or actual data approach yields consistent results from an econometric perspective. In fact, the actual data is manipulated and adjusted to reduce observed decisions of companies on investments.

- A dynamic panel data model mitigates these risks and brings consistency on the totex definition and data observed.
- Academic literature supports the use of dynamic approaches in the treatment of capital investments (e.g., enhancements) and the use of dynamic panel data models to control for lumpiness patterns (e.g., Arellano-Bond, Blundell-Bond estimators among others).
- In practial terms, these models are easily run in Stata for estimation purposes.
- Implementing this econometric technique improves the information on the calculation of capital expenditure efficiency allowances that are needed to meet customers outcomes and enviromental targets.

We welcome feedback on the findings set out in this report. Please contact Carlos Pineda at carlos.pinedabermudez@thameswater.co.uk

# F References


- Arellano, M. (2003). Panel Data Econometrics. Advanced Texts in Econometrics. *Oxford University Press*.
- Baltagi, B. (2005). Econometric Analysis of Panel Data. *John Wiley & Sons Ltd*. Third Edition.
- Blundell, R., and Bond, S. (1998a). Initial conditions and moment restrictions in dynamic panel data models, *Journal of Econometrics*, 87, pp. 115-143.
- Blundell, R., and Bond, S. (1998b). GMM estimation with persistent Panel Data: An application to Production Functions. The Institute for Fiscal Studies, working paper, No. W99/4.
- Blundell, R., and Bond, S. (2000). GMM estimation with persistent panel data: An application to production functions, *Econometric Reviews*, 19, pp. 321-340
- Blundel, R., Bond, S., Devereux, M., and Schiantarelli, F. (1992). Investment and Tobin's Q, Evidence from company panel data. *Journal of Econometrics*. Vol. 51, pp. 233-257.
- Bond, S. (2002). Dynamic Panel Data Models: A Guide to Micro Data Methods and Practice. The Institute for Fiscal Studies, CEMMAP working paper.
- Bond, S. and Meghir, C. (1994). Dynamic Investments Models and the Firm's Financial Policy. *The Review of Economic Studies*. Vol. 61, pp. 197-222.
- Cameron, C. and Trivedi, P (2005). Microeconometrics, Methods and Applications. *Cambridge University Press*.
- Cambini, C. and Rondi, L. (2011). Capital structure and investment in regulated network utilities: evidence from EU telecoms. *Industrial and Corporate Change*, Vol. 2, Number 1, pp. 31-71.
- Cambinini, C., Fumagalli, E. and Rondi, L. (2016). Incentives to quality and investment: evidence from electricity distribution in Italy. *Journal of Regulatory Economics*, Vol. 49, pp. 1-32.
- Lyon, T., and Mayo, J. (2005). Regulatory opportunism and investment behaviour: evidence from the U.S. electric utility industry, *RAND Journal of Economics*, Vol. 36, No 3., pp 628-644.
- Nardi, P. (2012). Transmission network unbundling and grid investments: Evidence from UCTE countries. *Utilities Policy*. Vol. 23, pp. 50-58.
- Peck, S. (1974). Alternative Investment Models for Firms in the Electric Utilities Industry. *The Bell Journal of Economics and Management Science*, Vol. 5, No. 2, pp. 420-458.
- Poudineh, R. and Jamasb, T. (2016). Determinants of investment under incentive regulation: The case of the Norwegian electricity distribution networks. *Energy Economics*, pp. 193-202.

Thames Water