MORE AND BETTER DATA

How companies and Ofwat
need to adapt their approach

2020

MORE AND BETTER DATA

HOW COMPANIES AND OFWAT NEED TO ADAPT THEIR APPROACH

WHAT COULD THE WORLD OF WATER LOOK LIKE WITH MORE AND BETTER DATA?

In the water sector, we can envisage a world where water companies have the data capabilities shown in Figure 1 below. In addition to the data capabilities within water companies, open data can create opportunities for third parties to develop applications and insights that drive efficiency.

Although significant steps have been taken and some companies are more advanced than others in this area, this future world will require substantial investment. Nevertheless, Figure 2 illustrates how data insights have the potential to transform the way we plan, deliver and analyse water services across the whole supply chain from resource to retail.

One of the main issues at both company and industry level that we observed at PR19 is the scope to improve integration between the companies' regulatory and operations teams. This means that even today with the limited amount of data available, the sector is not fully utilising data to create valuable insights. For example:

- Data collected for operational purposes often does not find its way into regulatory submissions, which can lead to positions that cannot be supported by a strong evidence base. This is particularly the case for cost and service quality special factor claims and includes demonstrating that companies optimise their own processes.

- Analysis of customer behaviours (e.g. contacts) in response to different service quality issues is often not integrated with customer research results, so we do not know if customers' views are supported by their behaviour. This means that we do not use a vast amount of data and information on customer behaviour and rely instead on "stated preferences".

- Ofwat's approach to benchmarking is clearly limited by the volume of data it collects, and the benchmarking dataset represents only a fraction of the data that

EXEC SUMMARY

Technological progress has created new, exciting opportunities to gather and analyse data on a scale that was unthinkable a decade ago. Automated collection of a vast, wide range of data by robots and the ability to analyse it using modern computational techniques, such as machine learning and artificial intelligence (AI), are some examples of what is now possible.

Over the next decade we expect data collection and analysis to evolve further to provide faster and better insights to manage businesses. Digital twins are starting to be developed. Digital twins are a virtual replica of the physical world and can include machine learning and AI to continuously learn and update themselves with the ultimate aim of optimising operations. Trends such as augmented analytics (automatically finding the most important insights or changes to optimise decision-making) and natural language processing (asking questions verbally instead of via text or code) show how data analytics is not just becoming more sophisticated but also more user-friendly.

companies gather about their operations. This means that the benchmarking models are limited in their accuracy and precision. While operational data currently may not be subject to the same degree of assurance, it can still provide useful insights into the underlying cost functions.

FIGURE 1   FUTURE WATER COMPANY DATA CAPABILITIES
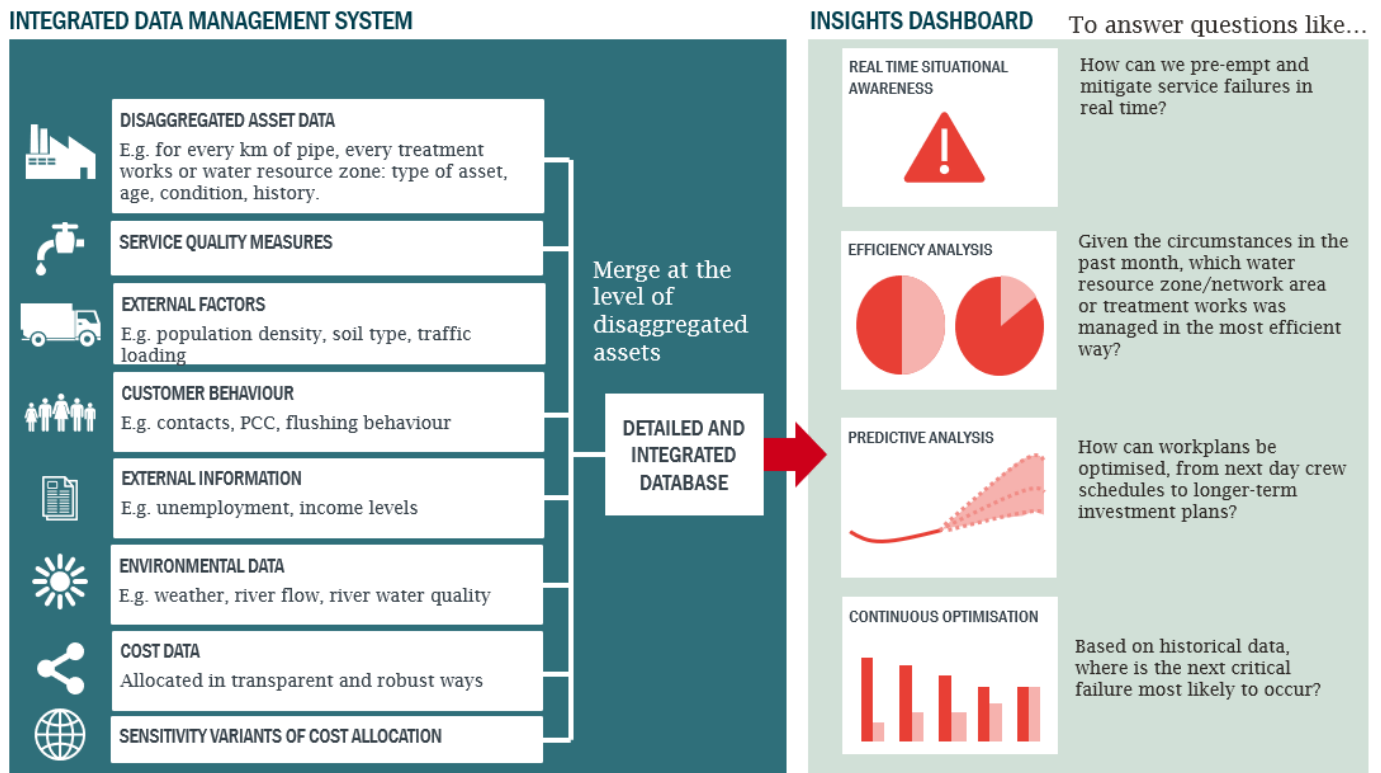
Source: Frontier Economics

Combining the opportunities created by the "data revolution" with our experience at PR19, this raises two important questions:

- How can companies and Ofwat use more and better data to improve economic regulation in the future?

- Is the current regulatory approach incentivising the right amount and type of investment in more and better data?

The objective of this paper is to address the first question. We provide our initial high-level views on the second question in the conclusion. With regard to the first question, we have identified three areas where new and better data can transform the current approach:

1   How can companies use more and better data to improve the quality their business plans;
2   How can companies apply economics and behavioural economics to more and better data to improve efficiency; and
3   The opportunities and risks from using more and better data for Ofwat's benchmarking.

1 HOW CAN COMPANIES USE MORE AND BETTER DATA CAN TO IMPROVE THE QUALITY OF THEIR BUSINESS PLANS

WHAT OPPORTUNITIES DOES MORE AND BETTER DATA PROVIDE?

Business plans that are based on better evidence increase the likelihood that Ofwat will accept the company's specific plans. While the approach to economic regulation over the next decade will evolve, it is clear that, regardless of the approach Ofwat takes, more and better data gives water companies the opportunity to drastically change the quality of their business plans and the engagement with Ofwat at price controls. While all companies might be expected to make significant progress in the next 10 years, some will get there sooner and will benefit from being ahead of the curve. The specific opportunities are:

- **Better evidence on customer behaviour and views** – integrate customer contact datasets (and other customer data) with operational data to develop a clearer understanding of customers' behaviour and attitudes. This can fundamentally shift the evidence away from surveys and "stated preferences" towards observing people's actual behaviour and responses in the context of service failures or improvements. For example, a peak in customer contacts on an unplanned interruption after 4 hours or at 8pm can reveal when the impact of not having water becomes more significant. Companies that use these techniques can be a lot more confident about their customers' opinions. This means that service quality targets can be informed by better evidence and are better able to withstand challenge.

- **Better evidence on efficient opex** – better data on explanatory factors and use of new techniques (such as machine learning) to benchmark different areas within each water company can vastly improve the evidence base on the efficient level of opex. This can help companies set their own efficiency targets and inform the level of costs to include in the business plan. While Ofwat will always use industry modelling, better evidence from companies could lead to Ofwat putting greater weight on the company plans.

- **Better evidence on cost and service special factors** – while the current regulatory methodology allows only for special cost adjustment claims, Ofwat's future approach could evolve to include special service factor claims. More and better data provides an opportunity to present more thoroughly grounded claims, including:

  - Better evidence on how companies have optimised their processes and performance given the circumstances

  - More robust modelling on why a particular company is different – more granular data that is made available by third parties can play an important role here

  - More accurate quantification of the claim

  - Better evidence to prove to what extent the differences are already included in Ofwat's modelling.

There is a feedback loop, as Ofwat's approach to benchmarking will also evolve in tandem with more and better data. There are ways in which the regulatory models could develop with better data. One option would be more comprehensive modelling that could make the process for special factors redundant.

Another option would be to have better, but simpler models combined with a special factor or cost adjustment process based on robust data analytics.

- **Better evidence on enhancement projects** – more comprehensive data on how customers are impacted if an enhancement project does not go ahead, better quantification of the wider benefits of enhancement projects and more systematic modelling to support the selection of particular projects can all be achieved with more and better data and analyses. As many companies already use investment optimisation models, the potential for gains might look more incremental but based on our experience with developing business plans there are significant gains that could be realised by improving the link between company level benchmarking models, sector-wide comparisons and investment optimisation models.

- **Better evidence on service quality targets** – better evidence on customer behaviour can provide stronger support for bespoke targets that are tailored to maximising local customers' satisfaction. Company-internal cost benchmarking can include service quality to identify areas that are particularly efficient with regard to costs and service. These insights can then be used to support more achievable company-wide targets.

Figure 2 below shows why a regulatory data strategy is important to achieve the overall objectives at each price control.

FIGURE 2   WHY COMPANIES NEED A REGULATORY DATA STRATEGY

Achieving companies' objectives at a price control...
- Increased cost allowance at price control
- More achievable service quality targets at price control
- Better quality business plan - higher chance to be fast-tracked

...requires better evidence...
- Efficient level of opex
- Why enhancement projects are needed
- Special factors
- Service quality targets
- Customer views

...delivered by a regulatory data strategy

Source: Frontier Economics

HOW DO WE DEVELOP AN EFFECTIVE REGULATORY DATA STRATEGY?

To create a better evidence base and a higher quality business plan, companies need to develop effective regulatory data strategies. Figure 3 below summarises the most important steps. The key points are:

- **Start with a list of the things you need to evidence** –in a high quality plan it is important that the data strategy is driven by the key questions that you need to answer, otherwise there is a risk that data analyses does not serve a clear purpose.

- **Take stock of existing data and analyses but also consider new data** – new is not limited to data collected by each water company but also needs to include the increasing amount of publicly available data. This needs to be future-focused to reflect the needs of companies in the medium-term.

- **Set up effective teams and governance** – as data collection and analyses will be owned by many different people within each water company, it is important to create be clear about who will create the insights and how these insights translate into decision making. A matrix structure is likely to work best in most large companies where people from different teams can come together to ensure that data is effectively used.

FIGURE 3   STEPS TO DEVELOP A REGULATORY DATA STRATEGY
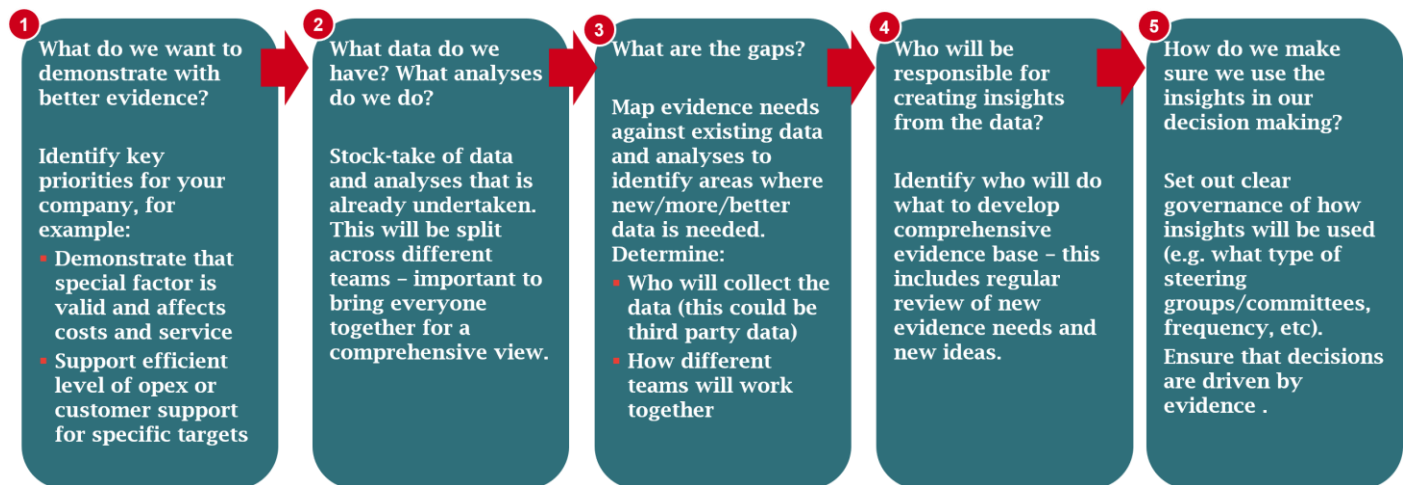
Source: Frontier Economics

2   HOW COMPANIES CAN APPLY ECONOMICS AND BEHAVIOURAL SCIENCE TO MORE AND BETTER DATA TO IMPROVE EFFICIENCY

Engineers, asset managers and scientists in water companies are always looking for ways to improve efficiency, and often closer to the practical considerations of data collection and analysis. The purpose of this section is therefore not to provide an overview of how data can drive efficiency in general but to look specifically at how that can be accomplished by combining regulatory and behavioural economics with operational data. There are important feedback loops, as the way in which companies assess and manage their own efficiency has implications for Ofwat's approach and affects the quality of business plans at the price control. We can identify three key opportunities:

- **Benchmarking efficiency both within water company boundaries and between companies** – as discussed in the following section, there is a wide range of new techniques that could be applied with more and better data. Water companies can use the same techniques to assess relative efficiency internally (e.g. between different zones or areas) and to derive specific insights on company comparisons. While the current methods are not sufficiently accurate, precise and granular to inform management decision-making, future techniques will generate genuine insights and therefore help drive efficiency.

- **Analyse how customer behaviour affects your costs to help shape initiatives to change people's habits** – one area that is often overlooked is how patterns of behaviour drive costs. More and better data can address this issue by making it easier to track customers' habits. For example, water efficiency initiatives are generally based on very little knowledge about customers actual behaviours and habits. Data is needed that tracks customers water usage on a device basis. This means we need to understand who uses water for what purpose over a sufficiently long time and large sample to test the effectiveness of different messages. Technological progress opens up a range of opportunities to track actual behaviour via apps and device-based technology such as smart appliances. This has the potential to transform the way companies engage with customers and to reduce costs, as they can roll out effective messages to all their customers.

- **Analyse customer views on real-time operations to optimise customer satisfaction** – the integration of customer behaviour data from all channels (including voice, web chat, social media, etc), operational data and survey-based data on satisfaction, value for money, etc. provides a valuable opportunity to analyse:

  - To what extent customer satisfaction reflects customers' experiences – for example, some of the research at PR19 showed that people who had suffered major disruptions, such as a sewer flooding incident were sometimes the most satisfied because they were impressed with the company response. But it is not clear that this view holds in the long-run.

  - What type of communication and interaction is more likely to leave people feeling satisfied with their water company – new data techniques can provide systematic analysis of all types of communication, whether verbal or online, and detect what type of responses work best.

  - How customer satisfaction can be maximised in the future – this can feed into resource planning in the short-run as well as longer-term business planning.

  - How companies should deal with anticipated failures so that customers' satisfaction is maximised.

Companies that harness better and more data and apply new analytical techniques can gain a significant advantage. There is also a feedback loop to the understanding of customer views for the business plan.

3 OPPORTUNITIES AND RISKS OF USING MORE AND BETTER DATA FOR OFWAT'S BENCHMARKING

The key question with regard to Ofwat's benchmarking approach is: how can we use more and better data to improve the approach to benchmarking without creating "bad" incentives?

A data-rich world provides enormous opportunities to improve our understanding of how external factors influence companies' costs and service quality. More and better data opens up new possibilities for integrating benchmarking of costs and service. This could address one of the most pressing questions for future price controls, underlining the value of the prize that more and better data offers. New approaches can also improve accuracy and precision and therefore reduce the risk of misallocating cost allowances, leading to greater confidence in the results.

However, at the same time we need to be careful that the availability of more and better data and new techniques does not dictate the approach to benchmarking. Because a new approach is possible, this does not mean that it is necessarily better. There are a number of risks that the industry needs to be aware of:

- Complex approaches with **little transparency can reduce trust** in the regulatory approach. We need to make sure that the new methodology is not a black box that prevents a dialogue between companies and Ofwat on the results.

- Collecting more-comparable data on an industry-wide basis requires **substantial efforts as the data has to be assured**. The costs need to be weighed up against the benefits.

- The water sector has traditionally used a set of explanatory variables that are exogenous to the companies, i.e. they cannot be influenced by the companies. Population density is a good example of an exogenous variable. Exogenous variables capture the cost drivers that are outside of companies' control. In a world with more and better data, there is a risk that the modelling approaches include explanatory variables that are within their control. This creates perverse incentives as companies can try to influence explanatory factors instead of focusing on cost efficiency. We therefore need to make sure that the **current approach of using exogenous variables is maintained**.

- We also need to remember that the purpose of the benchmarking is to set an overall efficient level of costs given the external circumstances of each company. Over the past two price controls, Ofwat has used modelling of (mostly) aggregated costs to set totex allowances. In our view, it is important to **maintain the incentive properties of the high-level approach to totex**. This ensures that companies have appropriate incentives to focus on the overall value for money they offer and to be innovative, finding new ways of achieving efficiency gains. Any new modelling approach that is driven by more and better data or new data analysis techniques needs to be aligned with these principles.

- We need to **avoid cherry-picking** when interpreting the results. With more disaggregated approaches, there is a risk that Ofwat will set the efficiency challenge for each company on the basis of the specific challenge for each data point (e.g. water resource zone). This could lead to an unachievable efficiency target when aggregating for each company. The results of new models therefore need to be interpreted with caution.

As Ofwat may revise its benchmarking approach with more and better data anyway, companies need to start engaging with the regulator now to maximise the opportunities and manage the risks. Keeping these risks in mind, we have identified a number of ways in which more and better data can improve the approach to benchmarking.

IMPROVING CURRENT APPROACHES

Even without considering new data analytics, more and better data will lead to an increase in the number of observations by using more granular inputs. This could be achieved in a number of ways such as using greater frequency (e.g. monthly instead of annual data) or greater geographic granularity (e.g. using water resource zones). The resulting increase in the sample size means that we can include a broader set of explanatory variables in a meaningful way. This reduces the "noise" in the estimates and has the potential to increase the accuracy and precision of the estimates and overall confidence in the results.

One of the practical challenges with this approach is the question of how costs are allocated. Consistent cost allocation is not easy to achieve but is essential if more- granular approaches are to work. This is particularly important if the proportion of fixed costs is high. The industry should consider the pros and cons of different types of granular approaches.
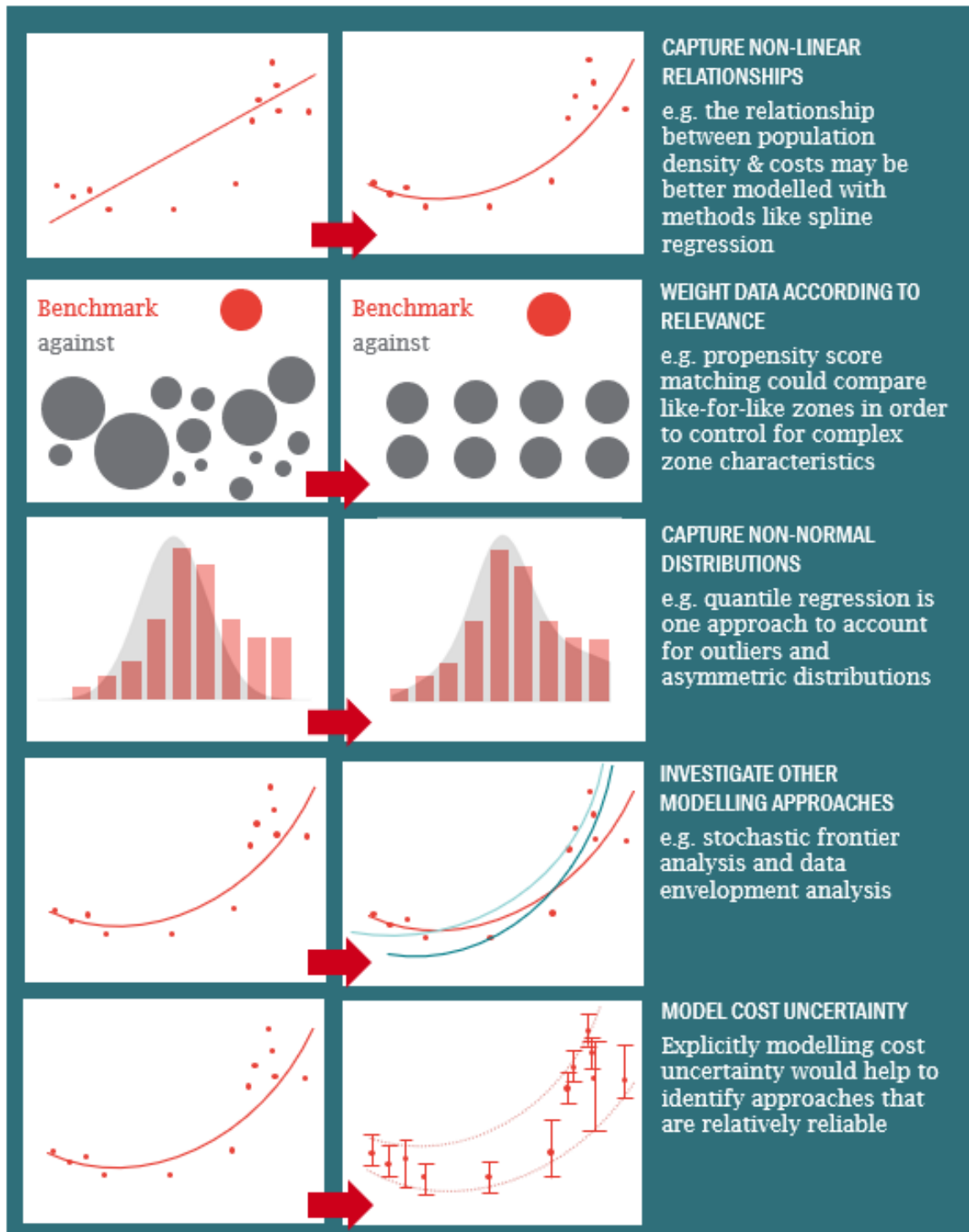
Another challenge with this approach is that all of the data points from one company are under the same management so they are not truly independent observations. This is not necessarily an issue, as company-

specific and data point-specific efficiency can be separated using statistical techniques. In any case, we note this also applies to the current approach, which includes observations from the same company over several years. Overall, an increase in the sample size is therefore worth exploring.

ENABLING NEW APPROACHES

The availability of more and better data could also open the door to new approaches, all of which could integrate cost and service quality modelling. One of the exciting things about the new data techniques is how straightforward it is to develop and test different types of models. Provided that the data quality is sufficient, in the new world we can explore many approaches very quickly, leaving more time to assess the pros and cons and incentive properties of various models.

Figure 4 (below) presents the parametric and non-parametric approaches that could be used with more and better data.

FIGURE 4   NEW APPROACHES ENABLED BY MORE AND BETTER DATA

CAPTURE NON-LINEAR RELATIONSHIPS

e.g. the relationship between population density & costs may be better modelled with methods like spline regression

WEIGHT DATA ACCORDING TO RELEVANCE

e.g. propensity score matching could compare like-for-like zones in order to control for complex zone characteristics

CAPTURE NON-NORMAL DISTRIBUTIONS

e.g. quantile regression is one approach to account for outliers and asymmetric distributions

INVESTIGATE OTHER MODELLING APPROACHES

e.g. stochastic frontier analysis and data envelopment analysis

MODEL COST UNCERTAINTY

Explicitly modelling cost uncertainty would help to identify approaches that are relatively reliable

Source: Frontier Economics

OPPORTUNITIES CREATED BY MORE AND BETTER DATA

The key opportunities are:

- **Relax linearity assumptions** – we would be able to explore a more flexible set of models that control for non-linear effects of the explanatory variables, for example the relationship between population density and costs. One simple approach is stratification, which involves estimating separate models for different groups of observations - for instance, all water resource zones that are fed by a large reservoir. More sophisticated methods would also be possible, such as spline regression (fitting a flexible non-linear function to data).

- **Relax the weighting assumptions** – under the current approach, each company is benchmarked against an estimate formed from all companies weighted equally. With more and better data, it would be possible to benchmark zones specifically against areas that have similar characteristics (differential weighting). This could help to set a more accurate efficiency challenge by allowing the model to control for more complex effects of zone characteristics. For example, differential weighting could capture geographic proximity and similarity in population density. One possible method is propensity score matching.

- **Relax distributional assumptions** – more and better data would allow us to test the current assumption that statistical errors are normally distributed. We could then modify the statistical approach as needed to address non-normality, so that the results are robust even if there are, for instance, many outliers or skewed outliers. Quantile regression is one possible method.

- **Estimate different types of models** – the current approach estimates efficiencies using random effect models. With more and better data, it would be possible to explore the feasibility of parametric and non-parametric models commonly used in benchmarking (e.g. in the energy sector in Germany). These models, such as stochastic frontier analysis (SFA), true fixed/random effect and data envelopment analysis (DEA), can have advantages over random effect models. For example, SFA is able to separate inefficiencies from idiosyncratic errors. DEA is able to identify for each inefficient resource zone/company a set of best-practice peers. True fixed/random effect models are able to distinguish persistent inefficiency from latent heterogeneity and idiosyncratic errors. Other more sophisticated options include models that allow inefficiency to vary over time. These models have been considered in various price control reviews but have not been taken forward due to the complexity and data limitations. More and better data and the ability to quickly try different techniques could make them a realistic option for the water sector.

- **Modelling of cost variability** – rather than simply estimating only the expected cost of delivering services in each area, cost variability could also be projected, enabling the full distribution of costs to be assessed. This would allow us to test which approaches are reliable when taking into account cost variability.

Relaxing these statistical assumptions could be implemented *manually* or *automated,* using, for example, unsupervised machine learning to group together (cluster) areas with similar characteristics and to train separate models for each group. In a manual approach, the modeller chooses the set of explanatory variables and the statistical form in which they affect cost. In a machine-learning approach, we specify a broader set of possible relationships between explanatory variables and costs and then use an algorithm to find the model within that set that has the best explanatory power. The advantages of this approach are:

- A large number of models can be tested relatively quickly, leaving more time to consider the outputs.

- Formalising the procedure for model selection may help to transparently communicate the range of models considered and the criteria for choosing the final one. A more objective approach would instil greater confidence in the results.

- The ultimate outputs are still transparent and just as easy to interpret as results derived "manually". This is an important point. The models are not a black box: they can still be interpreted and explained in the same way as the models today.

While many of these techniques would use more granular data, it is important to ensure that these are used in ways that maintains the incentives created by the totex approach. All of these approaches enable cost allowances to be set at the aggregate company level so that companies have discretion over how to improve average efficiency across different zones. Water companies and Ofwat need to start exploratory work on new approaches now to identify the direction of travel and ensure that we maximise the opportunities and manage the risk from new approaches to benchmarking.

DATASETS THAT COULD INFORM FUTURE BENCHMARKING

" The UK Water Industry Research (UKWIR) National Failure Database for Mains and Sewers brings together data on the water mains and sewer asset stock for the majority of the UK water companies, together with failure records spanning over 15 years. It has been developed to support collaborative industry research and to inform company asset management practices, but (with industry support) it could potentially have a future role in a more data-rich approach to regulation.

Use of detailed industry data could be a step towards answering UKWIR's 12th Big Question: "What is the true cost of maintaining assets and how do we get this better reflected in the regulatory decision making process?". If used appropriately, it could allow a more complete and consistent assessment of investment needs across the industry.

Also worth noting is the Geospatial Commission's National Underground Assets Register (NUAR) project, which is developing a web portal for making utility data accessible to engineers undertaking street works. Although the approved use-cases are very narrow, the development of the data sharing standards and IT protocols required for NUAR provides a step towards making a more data-rich approach to regulation feasible from a technical perspective."

GEORGE HEYWOOD
OVARRO

SO WHAT?

HOW CAN COMPANIES AND OFWAT USE MORE AND BETTER DATA TO IMPROVE ECONOMIC REGULATION IN THE FUTURE?

More and better data has the potential to transform the way we plan, deliver and analyse water services. This paper has explored the how companies and Ofwat can use more and better data to improve economic regulation in the future. We have identified three areas where new and better data can transform the current approach:

1   How can companies use more and better data to improve the quality their business plans
2   How can companies apply economics and behavioural economics to more and better data to improve efficiency
3   The opportunities and risks from using more and better data for Ofwat's benchmarking.

While the future world of more and better data will not be realised in one price control period, it is important for the sector to have a clear vision of how data will be used in the longer term. This will ensure that we make the right decisions today. We therefore think that:

- Companies need to develop regulatory data strategies so that future business plans and price review submissions are based on better evidence (see Figure 3).

- Companies need to apply the latest techniques and insights developed from economic regulation to new operational data to generate regulatory and operational insights (e.g. efficiency benchmarking).

- The sector needs to develop a clear, high-level vision of how costs and service should be benchmarked at the next price controls review. If it does not, there is a risk that data is not comparable, that the incentives to collect relevant data are not sufficient and that it will be too difficult at PR24 and subsequent reviews to adopt a new approach. A longer-term vision for benchmarking costs and service would create a clear way forward and PR24 can then be approached in this context. A joint vision can also ensure that high level incentives provided by the totex approach are maintained.

ARE REGULATORS INCENTIVISING THE RIGHT INVESTMENT IN MORE AND BETTER DATA?

Investment in more and better data collection and analysis requires substantial costs and effort over multiple AMPs. Similarly, the efficiency gains from better insights will be realised over the course of several AMPs. With a challenging PR19 Final Determination and the significant problems created by the COVID-19 pandemic, investment in data will be difficult to justify. The current regulatory approach creates short-term incentives that may not be compatible with investments and benefits that stretch over several AMPs. However, in our view more and better data is essential to drive long-term efficiency, which in turn is critical to the legitimacy of the water sector. Companies therefore should not be discouraged from commercially attractive options to invest in data. This means we need to examine more closely to what extent:

- Ofwat's £200m innovation fund could play a role in spurring investment in more and better data.

- Consistency of regulatory approach and commitment to future methods can provide sufficient certainty for companies to make investment decisions regardless of specific allowances.

- Existing regulatory incentives could be modified, or new incentives created outside the price control, to facilitate more investment in data.

- The methodology for PR24 could be more explicit in rewarding companies for efforts in this area.

AUTHORS

ANNABELLE ONG
Associate Director

CHRIS CUTTLE
Manager

FULVIO BONDIOLOTTI
Consultant

KATHARINE LAUDERDALE
Consultant