MODELLING THE PROPENSITY TO DEFAULT ON PAYMENT OF WATER BILLS

Final report prepared for Thames Water

5 October 2018

Jayanthi Ezekiel
020 7031 7061
jay.ezekiel@frontier-economics.com

Rob Francis
020 7031 7096
rob.francis@frontier-economics.com

CONTENTS

EXECUTIVE SUMMARY

As part of PR19, Ofwat will use econometric modelling to set cost baselines, including bad debt costs, for residential retail services. Thames Water has commissioned Frontier Economics to investigate the drivers of the cost of retail bad debt, particularly the propensity to default. This report sets out our findings and recommendations for PR19 and beyond.

Robust models are necessary for promoting consumer outcomes

Developing econometric models is central to Ofwat's approach to setting companies' baseline efficient cost allowances. It is important that the models are as robust as possible so that water companies are compensated for efficient costs while incentivising them to maximise efficiency to the benefit of their customers. The figure below summarises the characteristics that we used to guide our analysis

Figure 1 Characteristics of robust models for price controls

Characteristics	Ofwat principles
1 Capture all relevant drivers ▪ A weak model means that some of the differences between actual and forecast costs are because the model does not sufficiently capture the underlying drivers of efficient costs rather than because a company is inefficient. This could lead to Ofwat setting the cost allowance too low. **2 Use drivers that are outside the control of the company** ▪ Using exogenous drivers is more important than ensuring as good a fit as possible (i.e. it is more important than simply maximising the statistical properties of the model). ▪ Ensures correct efficiency incentives. **3 Ability to predict the value of the drivers with reasonable accuracy** ▪ Allows better forecasting of reasonable costs.	• Engineering, operational and economic understanding. • Robustness of the estimated coefficients and model specification. • Estimated coefficients are of plausible sign and magnitude.

Source: Frontier Economics and Ofwat

Since there is unlikely to be a single ideal model of bad debt costs, any results should be interpreted in the wider context of the regulatory control, including the other tools available to incentivise efficiency.

The focus of our work was to review the propensity to default and its impact on bad debt costs. We have used Ofwat's analysis of bad debt costs from its March 2018 consultation document, which considered a range of candidate explanatory factors, as the starting point for our analysis.

Ofwat's analysis is a starting point

Ofwat presented six specifications of the model of bad debt costs and found that each were appropriately specified, offered statistically significant results and had coefficients with economically intuitive signs and magnitudes. Ofwat found that average bill size is an important driver of bad debt costs and there is some evidence of economies of scale. Ofwat modelled the propensity to default on payment using three distinct measures: default on bills or credit risk score (as proxies of this propensity); and data on income deprivation from the index of multiple deprivation (IMD, as a driver of this propensity).

Models using default or credit risk data may provide valid results for PR19

Ofwat's models that include default rates and credit risk scores (ORDC1 and ORDC2) are statistically robust. The results are consistent with the economic intuition that if a household has a poor credit rating it will be more likely to default (since that rating is based on historic levels of default and other characteristics such as level of income). Historic levels of default may also be a good predictor of future levels of default to the extent that the characteristics that led a household to default persist in the future. . They are both based on detailed data available on a time series basis.[1] For price controls beyond PR19, while there is a potential drawback of using default rates, we consider that this could be mitigated. That is, to some degree default rates may be within the control of water companies and this could reduce the incentive properties of using this model to set cost allowances. However, this is much less of an issue for PR19 as the inclusion of default as a driver of bad debt costs would not have been known to companies. Further, Ofwat has other tools available to it to provide incentives for efficiency (e.g. form of control, or adjustments made outside of the model).

For models using credit risk scores, improved models may be possible if the scores were adjusted to reflect the specific risks faced by water companies. In particular, water companies are less able to manage their credit risks as they are required to provide water to households even when they do not pay their bills.

Available data on income deprivation is limited

The Ofwat model that uses data on the income component of the IMD produces statistically significant results that accord with economic intuition. However, it is limited by the data currently available which is based on the 2011 census and does not consider year to year variations. Further, income deprivation alone is unlikely to be the only driver of the propensity to default. For example, the income component of the IMD does not consider different degrees of deprivation (only whether a household is deprived or not). It also does not account for housing costs or for variation within area that would not be reflected by a simple average. Therefore, for PR19, the Ofwat models using default rates and credit risk scores are likely to be more appropriate than the Ofwat models using income deprivation as they can potentially capture a wider range of relevant factors.

Deprivation and vulnerability are the main underlying drivers of default

Economic intuition suggests that both deprivation and vulnerability are core drivers of the propensity to default since such circumstances can make it harder for households to pay their bills. Deprivation can be defined in various dimensions including income, education, health and employment. Ofwat has described vulnerability as customers without reasonable opportunity to access and receive an inclusive service.[2] Ofwat identifies those with water bills accounting for more than 3% of their disposable income as being either at risk or actually failing to pay their water bills. These are more likely to be low-income households, working-age adults living alone, lone parents and single pensioners.

We have modelled bad debt costs using data that better accounts for both vulnerability and deprivation than the data currently available on income deprivation. Using the proportion of households that are lone parent families as a proxy for vulnerability produces statistically significant and robust results. This

[1] We restrict attention to models that do not account for scale but note that our findings/observations extend to models that do (e.g. ORDC3 and 4).

[2] The Ofwat Vulnerability Focus report from February 2016 defines a vulnerable customer as “A customer who due to personal characteristics, their overall life situation or due to broader market and economic factors, is not having reasonable opportunity to access and receive an inclusive service which may have a detrimental impact on their health, wellbeing or finances.”

factor could be a proxy for wider groups at risk of affordability issues given its correlation with the incidence of them (working age adults living alone and people with long-health problems and disabilities, although not with single pensioners).

We sought to model other characteristics of deprivation and vulnerability including absolute income levels and housing costs. There are two main reasons why taking account of how these multiple characteristics vary both across and within regions (and thus across companies) is challenging. First, data on these characteristics is not always available, particularly on an up-to-date basis or as a time series. There are also issues with mapping the data onto company areas. Second, the sample size is limited to the number of companies and the number of years modelled. This limits the number of variables than can be included as well as the ability to independently estimate a robust composite measure. As we describe below, going forward, there is significant scope to develop modelling of the propensity to default.

We have four main recommendations for PR19 and beyond

The complexity of the factors that drive the propensity of a household to default on water bills means that no single model will meet all the relevant criteria and principles. Therefore, several econometric models and a range of information will need to be considered together to ensure that water companies can recover efficiently incurred costs in PR19. Given this, we have four recommendations.

First, the Ofwat models that use defaults and credit risk scores could be used as an interim measure for the purposes of PR19. These models are based on sound economic and operational understanding; produce statistically significant results and have robust model specifications; and have coefficients that are of plausible sign and magnitude. Even so, the results should be used with care to mitigate the risk of penalising companies for the weaknesses of the model rather than inefficiency. Also, specific characteristics that cannot be captured in a model but that affect one or more companies to a materially greater extent (e.g. transience, and housing costs relative to income), should also be taken into account (e.g. through adjustments to the cost allowances).

Second, future models should consider vulnerability and deprivation as the underlying drivers of the propensity to default. The income deprivation measure considered by Ofwat in ORDC5 is based on data that is out of date and unlikely to reflect household characteristics over the duration of the price control. It is also unlikely to capture sufficiently the variance of income deprivation within a water company area, as well as other drivers such as other types of deprivation, housing costs and income levels. Other things being equal, this means it is likely to be a weaker model than those using defaults or credit risk scores. As it stands, wider types of vulnerability and different types of households at risk of affordability issues can be proxied by using the proportion of households that are lone parent families. This offers statistically significant results and with a clear underlying economic intuition.

Third, better data is needed on income levels and volatility, living costs including housing costs, triggers of the risk of vulnerability, and non-income components of deprivation. While Ofwat's models using income deprivation are intuitively appealing and statistically robust, the models do not allow for the consideration of these other potentially important drivers. Improvements include ensuring data is collected and reported regularly; sufficiently geographically disaggregated;

comparable across all water company areas; sufficiently detailed for consideration of the variation of the data (not just simple averages); and specific to water bills.

Fourth, the issues associated with a limited sample size should be addressed as this materially restricts the scope for developing robust models. One way could be to increase the sample size by encouraging water companies to publish their bad debt costs on more geographically disaggregated basis. Separately from increasing the sample size, developing a composite measures of deprivation and vulnerability would mean fewer explanatory variables but would still allow for the consideration of relevant factors in a statistically robust way.

1 INTRODUCTION

As part of PR19, Ofwat will use econometric modelling to set cost baselines for residential retail services, including bad debt costs. Thames Water has commissioned Frontier Economics to investigate options for modelling the propensity to default on payment of water bills as a driver of bad debt costs.

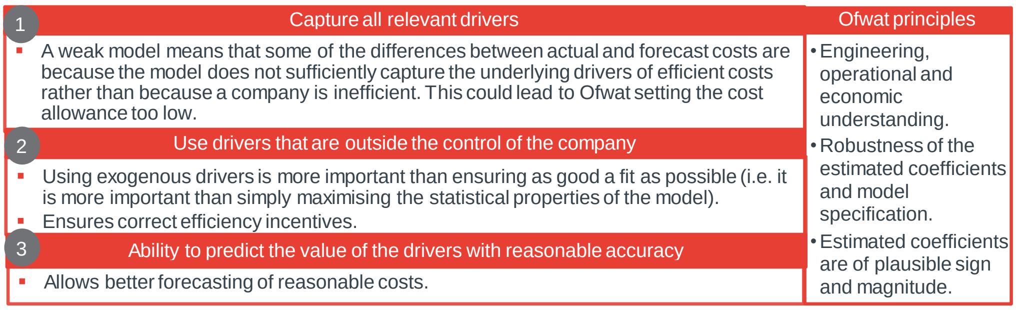
This report sets out our findings on this aspect of Ofwat's econometric modelling as presented to date and whether improvements could be made. We also consider how econometric modelling needs to be evaluated not just in statistical terms but also within the wider context of the price control and the impact it has on the efficiency incentives for water companies.

In the rest of this section, we describe objectives of econometric modelling for the purposes of a price control.

1.1 Objectives of econometric modelling for PR19

The objective of the econometric analysis in the current context is to develop a model that allows the estimation of the efficient level of bad debt costs faced by each of the water companies throughout the PR19 period. This has three main implications for the model; these are described in the figure below.

Figure 2 Characteristics of robust models for price controls



Source: Frontier Economics

In the context of modelling bad debt costs, the external factors facing companies are less clear-cut than other areas of operation. For example, waste management costs, sludge treatment costs or pumping costs can be modelling using operational and asset data from the companies. In contrast, the likelihood of water bill default largely comes from wider societal factors (and subsequent behaviour by households). There are many drivers of payment default, and the techniques for measuring these drivers are more varied still.

1.2 Interpreting the modelling for a price control

For a given benchmarking model, the estimated coefficients on drivers provide a predicted value of bad debt costs for each company, which when compared to the actual level implies an historical level of efficiency. Ofwat decides how to translate historical efficiency into a forward-looking assessment, through its choice of the appropriate efficiency frontier (historical average efficiency, or upper-quartile efficiency, etc). Combining forecasted values of independent variables and the chosen level of efficiency provides an estimated efficient cost allowance for each company throughout the PR19 period.

When setting the cost allowances, the models should not be used mechanistically. In particular, there should be careful consideration of the strengths and weaknesses of models relative to alternative specifications and wider information available. This is important so that the cost targets reflect scope for true efficiency gains rather than weaknesses of chosen models, and hence allow companies to recover efficiently incurred costs.

1.3 Report structure

In the rest of this report we set out:

- Ofwat's approach to modelling the cost of retail bad debt (Section 2);
- Our review of the evidence from the water and other sectors relating to the drivers of default on bills (Section 3);
- A description of our modelling approach in terms of principles, criteria for assessing different specifications and the data we have relied on (Section 4);
- The results of our econometric analysis of modelling bad debt costs, with a focus on propensity to default (Section 5); and
- Our recommendations for PR19 and beyond (Section 6).

Detailed modelling results are provided in annexes to this report.

2 OFWAT APPROACH TO MODELLING THE COST OF BAD DEBT

Ofwat has stated that it will use econometric models to set a base line efficient level of costs for each water company to deliver its business plan outcomes. In this section we describe the approach that Ofwat presented in its March consultation document in terms of the scope of bad debt models (Section 2.1); and the economic drivers of bad debt costs and how Ofwat has considered these (Section 2.2).

2.1 Scope of the bad debt models

Ofwat developed three econometric models for residential retail services over a four year period from 2013/14 to 2016/17: bad debt costs; other retail costs; and total retail costs (sum of bad debt and other retail costs).

Ofwat defines bad debt costs as all costs associated with managing bad debt and collecting outstanding customer revenues. It considers bad debt and debt management costs together as they are closely interlinked due to the "inherent operational choices and interactions between them".[3] Ofwat considers bad debt costs and other retail costs in separate models in order to "better capture the specific relationship between debt related costs and their unique drivers, such as bill size and deprivation." Given potential cost allocation issues, Ofwat (in addition to separate models for bad debt and other retail costs) also developed total retail cost models (i.e. models including bad debt, debt management and other retail costs). We consider that drivers that explain bad debt should be included to estimate efficient total retail costs as bad debt costs represent roughly half of total retail costs.[4]

2.2 The economic drivers of bad debt costs

Ofwat specifies the dependent variable in all its retail models as retail cost per connected household. This is because it considers retail costs are driven primarily by the number of customers. The figure below summarises the four main drivers of bad debt costs that Ofwat identified and the data that it used in its models.

[3] "Cost assessment for PR19: a consultation on econometric cost modelling", Ofwat consultation document, March 2018

[4] We have not reviewed the Ofwat models in detail to assess whether these are the correct disaggregation of retail costs. In particular, we have not considered whether splitting bad costs and other retail costs is the most appropriate way to model these costs.

Figure 3 Drivers of the cost of bad debt and data used by Ofwat

Driver	Description	Data/proxy used
Economies of scale	Average cost falling with scale, any decline likely to be modest.	Total number of households served.
Average bill size	Revenue at risk if there is a default. Higher bills increase the cost but may also increase the probability of default.	Average revenue per household.
Changes in household occupancy (transience)	May reduce ability to recover unpaid bills and, in turn may further increase the propensity to default.	Ofwat does not include a measure of transience in its consultation document.
Propensity to default on payment	Positive correlation with bad debt costs.	Modelled using three types of data ▪ Different types of deprivation as captured in the index of multiple deprivation (IMD) sourced from official government statistics. ▪ Credit arrears risk data from Equifax (provided by UU). These are constructed from customer data to predict the likelihood of non-payment. ▪ Percentage of households that defaulted on payment of water bills in previous years from Equifax.

Source: Cost assessment for PR19: a consultation of econometric cost modelling

Ofwat's models include data for the first and second drivers (number of households and average bill size), did not include a proxy for the third driver (transience), and proxied the fourth (propensity to default on payment). The latter is the most complex variable because it captures multiple characteristics and these are challenging to capture in a model. This is the focus of this report.

We broadly agree with these drivers. They are mainly exogenous to the behaviour of the company apart from some actions that management can take to influence the propensity to default on water bill payment (such as debt collection practices, social tariffs and water efficiency initiatives). We focus on whether the Ofwat models consider the "efficient" propensity to default (that is, default that is driven by factors that are exogenous to the behaviour of water companies).

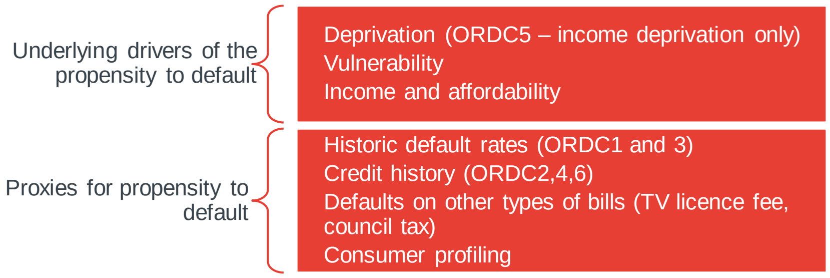
3 DRIVERS OF DEFAULT IN WATER AND OTHER SECTORS

The drivers of water bill default are varied and complex, and the reasons for defaulting will vary across households. In this section, we review the Ofwat models and existing evidence from a range of sources regarding the drivers of bad debt in water and other sectors and how those drivers are defined. We focus on the propensity of customers to default. Using Ofwat's models as a starting point, we considered:

- The models submitted by other water companies and the comments they made on Ofwat's approach;[5]
- Ofwat's wider research into affordability and vulnerability; [6] and
- Potential lessons from television licence fee evasion and council tax default.

We start this section by describing the distinction between using proxies versus the underlying drivers for the propensity to default (Section 3.1). The remainder of this section (as summarised in Figure 4 below) then describes potential explanatory variables, as well as their relative strengths and limitations. In Sections 3.2 to 3.4.1 we evaluate the measures considered by Ofwat. We then consider potential alternatives (Sections 3.4.2 to 3.9). We use the findings from this review to focus our econometric analysis (Sections 4 and 5).

Figure 4 Summary of potential explanatory variables



Source: Frontier Economics

3.1 Using data on underlying drivers versus proxies

We make a distinction between the two approaches that can be used to model the propensity to default on the payment of water bills. First, the model can consider proxies for the propensity to default on water bills by using historic data on default rates for (water) bills and on credit risk scores. This is what Ofwat modelled in all but one of the models it presented (ORDC5 being the exception). While such proxies do not necessarily ***cause*** households to default on their water bills, they may be strongly ***correlated*** with default. Second, the model can consider the

[5] https://www.ofwat.gov.uk/wp-content/uploads/2018/03/Appendix-1-Modelling-results_Final.pdf
This includes models from Severn Trent Water, South West Water, Wessex Water and Bristol Water, Yorkshire Water and South East Water. Our scope of work was to review the Ofwat proposed models and therefore we did not consider the models suggested by the other parties in depth (including the relative merits of including transience).

[6] Affordability and debt 2014-15, Ofwat, December 2015.

underlying drivers of the propensity to default on water bills (i.e. the factors that cause or contribute) such as income, deprivation and vulnerability. This is what Ofwat modelled in ORDC5 using data on the income component of the index of multiple deprivation (IMD).

Compared to using proxies, the main advantage of considering the underlying drivers of the propensity to default is that it is based on variables that are beyond the control of the water companies. The existing Ofwat model does this using data on income deprivation. Using this to set costs for the price control can provide water companies with strong incentives to reduce the level of defaults. However, care needs to be taken to ensure that the data used sufficiently accounts for the various factors that affect default propensity and does not focus overly on income alone. It is also important that the model is not "over-fitted" – i.e. that the specification with the strongest statistical properties is chosen at the expense of a specification that has better economic intuition and is more suitable for forecasting.

3.2 Historic defaults on bills

Ofwat models, ORDC1 and ORDC3, proxy the propensity to default by using the percentage of households that have defaulted over the time period modelled. It is not made explicit from the data provided whether the default rates relate to defaults only on water bills or whether they relate to defaults on all types of bills (such as mobile phone bills, credit card payments, other utility bills and so on). Therefore, in our analysis, we consider both possibilities.

3.2.1 Defaults on water bills

The intuition behind using data on historic defaults on water bills is that it directly measures the propensity to default on water bills and this could potentially reflect the future propensity to default on water bills. Unsurprisingly, the coefficient is of the correct sign and magnitude and statistically significant. The model also performs well against other statistical tests (R2, VIF(max) and Reset test). The data on historic defaults is available on a very granular level and for each of the time periods modelled. It may therefore be a suitable interim measure for the purposes for PR19 if interpreted with care.

For price controls beyond PR19, while there is a potential drawback of using default rates, we consider that this could be mitigated. In particular, default rates may, to some degree, be within the control of water companies and this could reduce the incentive properties of using this model to set cost allowances (although they would not be able to recover debts that were larger than average or inefficiently high debt collection costs). However, this is unlikely to be relevant in PR19 as companies would have not known in the past that the model would be used. Also, Ofwat has other tools available to it to provide incentives for efficiency (e.g. the form of control, or adjustments made outside of the model).

3.2.2 Defaults on all bills

Similarly to using historic data on water bills only, the intuition behind using historic data on defaults on all types of bills is that it may reflect future propensity to default on water bills. Using defaults on all types of bills rather than just water bills could to some degree mitigate the incentive issue described above. However, if people are more likely to default on water bills than they are on other types of bills then the mitigation effect may be more limited since defaults on water bills will account for a greater proportion of defaults.

At the same time, using data on the defaults for all types of bills could, to some degree, reduce the ability to explain variation in bad debt costs across companies. This is because the data will reflect characteristics of the provision of those goods and services that are different to water (mainly relating to the ability of the providers of those goods and services to better manage their credit risk). This will depend on the extent to which other types of goods and services share the specific characteristics of water bills.

- Water companies do not select their customers and are required to offer services to all customers. Therefore, water companies are exposed to a higher risk of default than most retailers (such as mobile phone companies or mortgage providers) that are able offer their products and services selectively.
- Water companies are not allowed to cut off household water supply in case of non-payment of bills. This means that vulnerable individuals may be more likely to default as the consequences will be limited.
- Water companies are prevented from using prepayment meters to stop water supply as a way of managing the risk of non-payment in the way that other companies might be able to (e.g. mobile phones or electricity companies).

These characteristics mean that households are more likely to deprioritise paying water bills in favour of other bills where the consequences of non-payment are more immediate. While this does not necessarily invalidate models that rely on default on all bills as a proxy of propensity to default, it suggests that a better specification may be possible.

3.3 Credit risk scores

Both Ofwat and Severn Trent included credit risk scores in their econometric modelling and found statistically robust models.[7] In principle, appropriate credit risk scores could mitigate the incentive problem of using default rates (whether defaults on all bills or water bills only) because they are based on data on a wider range of factors. Typically, such factors include total existing debt, credit history, age of accounts and types of accounts. While credit risk scores would still include water bill defaults, the impact of water bill defaults in themselves on credit risk scores is likely to be limited.

However, our understanding is that large retail financial services companies do not typically use the credit risk scores in the way that they have been compiled by credit agencies such as Equifax. This is because the scores do not reflect the credit risks that are specific to their customers. Instead, these companies use the underlying data to construct their own ratings. As described above, there are specific characteristics of retail water customers which would mean a bespoke

[7] Ofwat calculates an average credit score rating for each water company area and uses this in ORDC2.

rating would better capture the propensity to default on water bills. Also, many of the customers that are more likely to default on payment would also be unable to access the consumer credit market. While these characteristics do not necessarily invalidate using credit score as a proxy for propensity to default,[8] they do suggest that better model specifications may be possible if sufficient data were available.

3.4 Deprivation and vulnerability

Given the issues described in Sections 3.2 and 3.3, we consider the underlying drivers of the propensity to default. In particular, we look at how the socio-economic circumstances of customers underpins their ability to pay (water) bills. This can be in two main ways: deprivation and vulnerability. While both these terms can be defined in various ways, we describe below their likely link with the propensity to default.

The figure below summarises a number of dimensions of deprivation and their links with the propensity to default.

Figure 5 Dimensions of deprivation and likely link with the propensity to default

Dimension	Link with propensity to default
Income	Low and unpredictable income can lead to affordability issues
Education	Low levels of educational attainment may make it difficult to manage bills and personal finances more generally
Health	Poor health may reduce the ability and time available to deal with bills
Employment	Unemployment or unstable employment affects income. It may also affect access to credit or support.

Source: Frontier Economics using IMD components of deprivation

Vulnerability, while linked to deprivation, relates to how likely individuals are to suffer detriment due to their personal circumstances. This could either be longstanding (such as long term illness) or temporary in nature (such as recent bereavement or short term illness). Measures of deprivation may not pick up such factors.

3.4.1 Index of multiple deprivation

Ofwat models the impact of the income deprivation in its ORDC5 model as one aspect of different types of deprivation. However, as we describe below, we consider that data on income deprivation alone is unlikely to be sufficiently complete to allow for the consideration of all the relevant drivers of the propensity to default.

The English and Welsh IMD take data across a wide range of categories (called "domains" by the ONS), weighs them according to their perceived importance and create a ranking of each Lower-layer Super Output Area (LSOA). [9]

[8] As in ORDC2,4 and 6.

[9] An LSOA is a small administrative geographic area containing 1,500 people on average. IMD rankings are not statistically interpretable as they only show rankings between areas, and not the scale of the difference. Therefore, Ofwat's models use data from the underlying IMD domains instead of the rankings.

The IMD includes seven domains of deprivation (see figure below). For some of these domains, the link with the affordability of water bills and propensity to default is relatively clear. For example, low levels of skills may make it harder for households to manage bills. However, for others the link may be less direct. For example, although high levels of crime could be indicative of vulnerability, it may be better to look at measures of vulnerability that affect affordability more directly.

Figure 6 IMD domains

	Description
Income Deprivation	Proportion of the population experiencing deprivation relating to low income. The definition of low income used includes both those people that are out-of-work, and those that are in work but who have low earnings (and who satisfy the respective means tests).
Employment Deprivation	Proportion of the working-age population in an area involuntarily excluded from the labour market. This includes people who would like to work but are unable to do so due to unemployment, sickness or disability, or caring responsibilities.
Education, Skills and Training Deprivation	Lack of attainment and skills in the local population. The indicators fall into two sub-domains: one relating to children and young people and one relating to adult skills.
Health Deprivation and Disability	Risk of premature death and the impairment of quality of life through poor physical or mental health. The component measures morbidity, disability and premature mortality but not aspects of behaviour or environment that may be predictive of future health deprivation.
Crime	Risk of personal and material victimisation at local level.
Barriers to Housing and Services	Physical and financial accessibility of housing and local services. The indicators fall into two sub-components: 'geographical barriers', which relate to the physical proximity of local services, and 'wider barriers' which includes issues relating to access to housing.
Living Environment Deprivation	Quality of the local environment. The indicators fall into two sub-components. The 'indoors' living environment measures the quality of housing; while the 'outdoors' living environment contains measures of air quality and road traffic accidents.

Source: National Statistics and Department for Communities and Local Government, available online; https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015

Ofwat uses data on income deprivation from the IMD in its ORDC5 model. The bullet points below set out some of the limitations of using this data on income deprivation to explain water bill defaults. These limitations are in addition to those that relate to the IMD more generally as described further below.

- **It does not reflect the full distribution of incomes within each area**
  By measuring the proportion of people in an area who claim income-related benefits or tax credits, the IMD income score is only able to measure those households beneath a specific point in the income distribution. A variable that measures the proportion of people in each water company area beneath a certain income should be able to be tested at different income-level thresholds to ensure it is capturing the appropriate relationship. Such characteristics may also contribute to the propensity to default.
- **It is a partial measure of deprivation**
  When calculating the full IMD rankings, the ONS puts only a 22.5% weight on income deprivation scores, recognising the importance of many other

characteristics – such as employment, crime or living environment – which contribute to local deprivation.

- **It does not include non-income measures that affect propensity to default**
  When estimating the propensity to default for water bills, non-income measures such as housing affordability should also be considered, otherwise those with high incomes but proportionally higher rents will not be accounted for. This may not be reflected in deprivation data.

- **It does not consider the volatility of income**
  A significant driver of default on water bills may not just be the overall level of income available, but could also be the volatility of income across time periods. Using deprivation as measured at a single point in time makes it impossible to capture this relationship.

While the IMD provides granular data on other domains of deprivation, it has a number of limitations that prevent it from fulfilling this purpose. These are set out in the bullet points below.

- **It does not vary by year over the sample period**
  Using data from a fixed point in time makes variation over time impossible to observe or estimate. Further, it is based on relatively old data from the last census (2011). This is particularly problematic as the economic characteristics of areas can change quickly resulting in dramatic shifts in relative deprivation across areas. For example, parts of Sheffield and Nottingham have moved from the top 25% most deprived to the 25% least deprived between 1999 and 2009.[10]

- **Changes in deprivation cannot be forecast using current data**
  Without data varying by year, any attempt to project the expected path of deprivation throughout the PR19 period would be arbitrary.

- **Comparable data on most non-income elements is not available for Wales**
  Data for Wales is compiled on a slightly different basis to data for England. Ofwat has made adjustments to take account of this for the income component. While Welsh Water has commissioned work in this area we have not reviewed this for this project given the other limitations of the dataset described above.

3.4.2 Measuring vulnerability and affordability in the water sector

Research commissioned by Ofwat found that certain categories of households had higher levels of water bill default. In particular, Ofwat states that "Low-income households, working-age adults living alone, lone parents and single pensioners are more likely to have problems paying their bills, and are more likely to be in debt.".[11] As described below, household composition appears to be correlated with bad debt costs. This may be a useful proxy for vulnerability.

The availability and take-up of social tariffs could also provide a proxy if these are in place in areas where bad debt is particularly problematic and if reliable data was

---

[10] These figures are taken from a review of local economic deprivation by the ONS. This is available online. https://www.gov.uk/government/statistics/tracking-economic-and-child-income-deprivation-at-neighbourhood-level-in-england-1999-to-2009

[11] Affordability and debt 2014-15, December 2015, Ofwat

available. However, such schemes remain in their infancy and are therefore unlikely to have yet had a significant impact on default rates.[12]

In its 2016 practitioners' pack to accompany its vulnerability focus report, Ofwat identifies triggers to help water companies to identify customers whose circumstances make them vulnerable and who therefore do not have a "reasonable opportunity to access and receive an inclusive service which may have a detrimental impact on their wealth, wellbeing or finances".[13] These are summarised in Table 7 below.

Table 7 Triggers of risk of vulnerability

Category	Examples of triggers
Personal characteristics	Income assistance/financial vulnerability Old age (may be related to specific issues such as health issues or difficulty accessing information) Health conditions – especially those requiring higher use of water Disability – may affect access to information, domestic situation or job
Changes in life events	Hospitalisation, job loss, divorce/separation, moving from another country Increase caring responsibilities Economic changes (e.g. interest rate changes, redundancies) Changes to benefits entitlement
Difficulty understanding or accessing information	Mental health problems, learning difficulties, limited literacy (including financial literacy), dementia

Source: Ofwat[14]

As we describe Section 3.4.1 above, the IMD, accounts for some of these triggers. Nevertheless, data on specific triggers of vulnerability issues may also be informative, particularly if they are a useful proxy for other factors (e.g. the proportion of lone parent households can capture both low income and changes in life events). We also find that when considering affordability across regions, housing costs as a proportion of income are likely to have a significant impact (see Section 5.2.2).

Ofwat found that "affordability risks emerge when a household spends more than … 5%, of their disposable income on water and sewerage bills." It found that 11% and 24% of households spend more than these thresholds respectively.[15] For most categories of households, 13-15% of those households faced affordability issues.[16] However, a higher proportion of the following categories of households face

[12] The design and eligibility for social tariff schemes will vary from company to company. While this is likely to reflect underlying differences in the customers of different companies, it can also make comparison more challenging. Their existence, design and take up is also likely to be within the control of the water company, particularly in the medium to long term.

[13] Practitioners' pack for water companies to accompany Ofwat's vulnerability focus report, Ofwat, February 2016.

[14] Practitioners' pack for water companies to accompany Ofwat's vulnerability focus report, Ofwat, February 2016.

[15] Affordability and debt 2014-15, Ofwat, December 2015. This was based on data from the Department for Work and Pension's (DWP) annual Family Resources Survey (2013-2014).

[16] These other categories consist of pensioner couples, couples with children, couples without children, multi-unit and other (such as two working age adults sharing a property).

affordability issues: lone parents (40%); working age adults living alone (45%); and single pensioners (38%). [17]

Intuitively, the proportion of households that are lone parent could be a reasonable proxy for the proportion of households that are vulnerable to financial difficulty and default (see figure below).

Figure 8 Lone parent households as a proxy for vulnerability

Potential characteristic	Description
Lower income	This could be for a number of reasons: ▪ There is only one person earning or receiving benefits ▪ That person is typically a woman and therefore will typically earn less than a man ▪ That person will also likely have childcare responsibilities and may there need to work flexibly or part-time. This reduces earning potential. ▪ The fact they are not living with grandparents/other relatives suggests there is little family support available
Income may also be more volatile and unpredictable	This could be because unpredictable childcare responsibilities limit formal employment opportunities. Maintenance payments from the absent parent may be irregular.
Higher costs of living compared to households without children	They will have dependents. They will face childcare costs.
Higher water usage and therefore higher bills	Compared to single occupancy households Compared to households with only adults More time at home if the children do not attend nursery

Source: Frontier Economics

While being a lone parent household does not directly imply default, it may strongly correlated with other factors.[18] This could make household composition a suitable proxy for these other factors.

Based on our data analysis, we find that the concentration of lone parent families is not only an intuitive driver of bad debt costs, but also an effective proxy for other vulnerable populations. In particular, there is a strong relationship between lone parent households and households in which at least one person has a disability. The proportion of lone parent households and the proportion of working-age adults living alone are also correlated (see Figure 26 in Annex B). These correlations become stronger as the rates are aggregated to a local authority or company level.

An exception to this, however, is that single pensioner households appear to be negatively correlated with other vulnerable populations at lower levels of geographic granularity, and less strongly correlated at the water company level. This means that a single variable of vulnerability cannot fully capture all demographics of interest. That said, lone parent families (because of positive

17 The LCF survey data includes disposable income https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/bulletins/householddisposableincomeandinequality/financialyearending2017

18 Other factors may include income volatility, transience, low income, or low income once other costs like childcare have been taken into account.

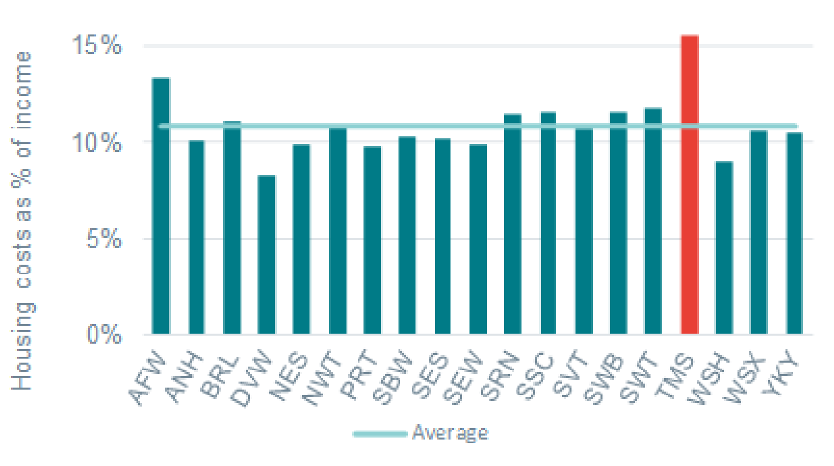
correlation) would pick up on the effect of most vulnerability categories on bad debt costs.

We set out the results of including lone parent households in in the econometric modelling in Section 5. This shows a strong and robust relationship between lone parent families and bad debt costs. However, similarly to the IMD data, the data on these local demographic measures is from the 2011 Census. In that regard it shares the same drawback as IMD in terms of potentially out of date information and also the challenge in forecasting values for the PR19 period.

3.4.3 Taking account of housing costs

All else being equal, we would expect a higher level of water bill default if housing costs account for a greater proportion of income. Households risk eviction for non-payment of rent or mortgages whereas water supply cannot be turned off. This means that housing costs would take a higher priority than water bills if a household faces affordability issues. Therefore, in some areas, if regulatory cost allowances do not take account of differences in housing costs relative to income across water company areas, water companies may not be able to recover efficiently incurred costs. The figure below shows housing costs as a percentage of average weekly income. On average, customers in the Thames Water and Affinity areas face much higher housing costs relative to income.

Figure 9 Housing costs as a percentage of mean income



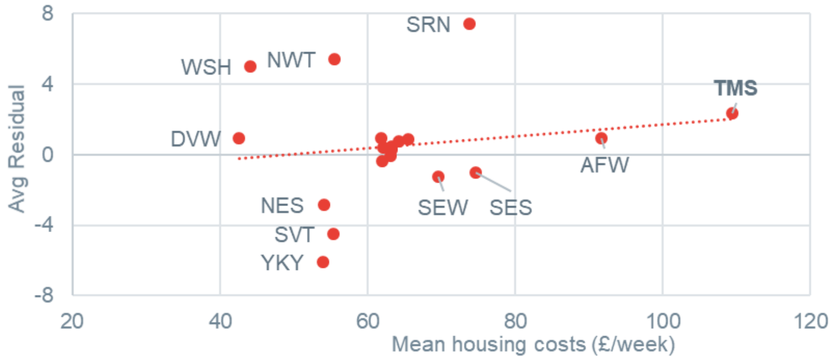
Source: Frontier Economics aggregation of ONS Small Area Model-Based income estimates for the financial year ending 2014[19]

Figure 10 below shows the relationship between mean housing costs and the residuals from the ORDC5 benchmarking model (i.e. controlling for both average bill size and income deprivation). Although the slight positive trend is weak, it does suggest that taking account of housing costs could strengthen the model if sufficient data were available. Therefore, we recommend that that more granular information would need to be considered before ruling out the impact of housing costs. In our analysis, we have relied on data the ONS has published experimental

[19] https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/articles/smallareamodelbasedincomeestimatesenglandandwales/financialyearending2014

statistics on incomes and housing costs for small areas[20] using model-based estimates of results from the Family Resources Survey (FRS).[21] However, regular, reliable and granular data on local housing-related costs is not made publicly available in England and Wales. While house prices, rental rates and mortgages are closely tracked by the ONS, VOA and HM Land Registry, these are only occasionally transformed into official statistics for small areas that can be mapped onto the water company areas. Better data would allow for the better consideration of the drivers of the propensity to default both between areas and over time.

Figure 10 Housing costs and Ofwat model residuals



Source: Frontier Economics recreation of Ofwat's ORDC5 model.

Note: The y-axis displays the average difference between modelled and actual log of debt cost per household for each water company. A higher residual implies less efficient bad debt costs, and a relationship between mean housing costs and residuals would suggest its value as an included variable for benchmarking. The dotted line is a simple linear fit of the two variables.
The modelled values are from a Frontier recreation of the ORDC5 model in the Ofwat March 2018 Consultation on Retail Cost Benchmarking.

3.5 Measures proposed by other companies

The figure below summarises some of the other options for modelling the propensity to default that other water companies proposed in their submissions to Ofwat. We consider that these measures typically only partially capture the relevant drivers of the propensity to default. Most importantly, they do not take account of regional differences in living costs or income volatility and how either of these can affect the affordability of water and wastewater services. As such, we consider that these measures are not good alternatives to other measures discussed in this section (such as default, credit score and lone parent families).

[20] Middle Layer Super Output Area – MSOAs are geographic areas with a minimum population of 5,000 and a mean size of 7,200. They consist of groups of contiguous Lower Layer Super Output Areas. There are around 35,000 LSOAs and around 7,000 MSOAs in England and Wales.

[21] For each MSOA, the ONS has published for 2011 and 2014 the estimated mean weekly income before housing costs and after housing costs and, as a residual, housing costs themselves. We have compiled this data first into local authorities and then into water company areas, at each stage taking the mean of the smaller geographies. We therefore have an estimate for the mean of income before housing costs, income after housing costs and housing costs themselves for each company area.

Figure 11 Measures of propensity to default suggested by other water companies

Measure	Comments
Bill to income ratio (Severn Trent)	This may help to provide a better indication of affordability rather than considering income alone
Proportion of private rental properties/ private renters (Severn Trent/ Yorkshire)	This captures neither the proportion of people living in social housing compared to those living in their own homes nor the underlying costs of housing which represents a large proportion of household expenditure.
Property repossessions (Wessex and Bristol)	This will capture affordability issues among home owners, but ownership is not typically associated with very low income.
Unemployment (South East)	Although high unemployment rates may be associated with low income, it does not take account of how income and affordability may vary for those on the lowest incomes (including those on benefits) who are most likely to face affordability issues. There may also be people that are not registered as unemployed but have very low income or unpredictable income (e.g. those on zero hour contracts).

Source: Ofwat consultation document

3.6 Council tax default

South West Water considered council tax default rates as part of its modelling. While South West Water attributes its data source to DCLG data on a local authority basis, we were unable to find such information publicly available. Nevertheless, if data availability issues could be addressed, we consider that this could be a useful proxy for the following reasons.[22]

- Like water companies, councils do not select their "customers" in that they do not control who can live where.
- While councils have various legal powers to require households to pay their council tax, they are limited in their ability to withhold services from households that do not pay. This is particularly true where services provided are public goods (e.g. road repairs, arts and leisure, street cleaning) or where it would be administratively costly to stop service provision to a single household (e.g. bin collection). Also, council tax may pay for services that the household itself does not receive such as care for vulnerable adults and children.
- Although council tax is typically paid in advance the usefulness of advance payment credit risk management tool is restricted by the limited legal powers that councils have to withhold services from households that do not pay.

3.7 Evasion of TV licence fees

TV licence fees also share a number of characteristics with water bills:

[22] One of the drawbacks of using defaults on council tax is that it is a proxy which limits its ability to forecast how defaults on water bills over the price control period.

- The BBC does not control which households have access to its broadcasting and digital services and it is unable to cut off supply to those households; and
- Prepayment is not a feasible way of managing credit risk.

This means that this could be a useful proxy for the propensity to default on water bills if sufficiently granular data on non-payment of TV licence fees were available across water company areas. In principle, such data could be available on a postcode level basis. Given issues of data protection such information is not made publicly available. However, it might be possible for the BBC to aggregate such data in way that it could be used as a measure of propensity to default in Ofwat's models without raising data protection issues. Another potential source could be the National Audit Office as it has carried out a review of licence fee evasion.[23]

3.8 Transience

None of the Ofwat models directly consider the impact of transience on the cost of bad debt even though Ofwat identified this as a driver in principle. While Ofwat identified transience as having an impact on the cost of chasing unpaid bills, it is also likely that customers are more likely to default if they know that it will be harder to water companies to chase them when moving property (in particular when moving abroad or to another water company area). Given these two channels for driving bad debt costs (with transience potentially raising both debt management and debt write offs), we recommend transience as an area for further modelling work in future price controls if it is not already accounted for through other measures (see Section 6). Also, it may be appropriate to consider this as a company-specific circumstance (see Section 6.5). We summarise the models that include transience that companies (Thames Water and others) submitted to Ofwat in Annex B.1.

3.9 Customer profiling

Agencies such as YouGov and ACORN develop customer profiles using data on a wide range of factors. These are typically compiled for purposes other than debt management or credit rating. In particular, we note their use for targeted marketing campaigns and to understand political voting behaviours. The classifications use data on the characteristics of factors that could be potentially relevant for these purposes. This means that they contain data on a number of characteristics that may have no clear link with the propensity to default on water bills. For example, it is relatively clear to see how an individual who engages actively with social media would be more responsive to marketing over the platform rather than traditional media.

Another limitation is that the data set available for modelling the cost of bad debt is small relative to consumer profiling datasets. This means that even if we were able to access the underlying data and only include relevant characteristics, it would not be possible top capture of all these in a statistically robust way given the small size of other samples used.

[23] Our request for this information under the Freedom of Information Act was denied.

3.10 Summary

Ofwat's analysis is a useful foundation in identifying the potential factors that either directly explain the propensity to default on water bills (income deprivation) or proxy for the propensity to default (defaults on all bills and a credit risk score). While each of the models proposed by Ofwat have their merits, there are a number of ways in which these models consider the propensity to default could be improved. In particular, modelling the underlying drivers of the propensity to default is preferable to using proxies as it considers factors that are beyond the control of the company. This would also allow for more robust forecasting.

Intuitively, deprivation, vulnerability and transience are likely the main drivers of the propensity to default. However, finding sufficiently detailed and up to date data to support this is challenging. This means that a practical approach for the purpose of PR19 will be to use proxies (such as default rates and credit scores). In contrast, the IMD data that Ofwat uses is relatively old and only relates to one year. It also does not allow for the consideration of the variance of income within and between regions as it is based on headcount rather than income levels.

Attempting to capture triggers of risk of vulnerability may allow to better model propensity to default (in future review periods). In particular, using the proportion of households that are lone parent families can capture both low income and changes in life events. Nevertheless, more work would be required to consider other populations at risk of vulnerability (e.g. single pensioners) that are not correlated with lone parent families. We investigate these areas further in Sections 5 and 6.

4 MODELLING APPROACH

Based on our review of how the propensity to default can be modelled (as described in Section 3 above), we find that a variety of model specifications and estimation procedures could be used to set the level of efficient bad debt costs. In the rest of this section, we set out the principles we have used in our modelling approach (Section 4.1) before discussing a number of aspects of modelling specification, including:

- Use of time series data (Section 4.2.1); and
- Rationale for not including economies of scale and scope (Section 4.2.2).

4.1 Modelling principles

To ensure that any models we produce align with the purpose of fair and accurate benchmarking and forecasting, we have followed the principles set out by Ofwat, as summarised in the figure below.

Figure 12 Ofwat principles

Principle	Description
Use of engineering, operational and economic understanding	This will be used to specify an econometric model and form expectations about the relationship between costs and cost drivers in the model
Sign and magnitude of estimated coefficients	Assess whether the estimated coefficients are of the right sign and of plausible magnitude
Robustness of estimated coefficients	Assess whether they are stable and consistent across different specifications and whether statistically significant
Consequences of cost drivers under management control	Particular focus will be given to the risk of any perverse incentives
Statistical validity of the model more widely	Consider whether the model performs well in terms of statistical tests and diagnostics
Appropriate estimation method	Consider whether to use more complex panel data estimation (such as a random effects model) or ordinary least squares (OLS).

Source: Ofwat consultation document

In addition to the Ofwat principles, we have also used the principles set out in the figure below.

Figure 13 Additional modelling principles

Principle	Description
Planning and extending our analysis on the basis of economic principles and operational experience	The intuitive and theoretical basis of a model is more important that the testing and statistical validity of it. This is particularly important given the wide variety of potential variables that could be used to explain or proxy for default risk, deprivation or transience.
Selecting a small set of potential models that have different strengths and weaknesses	This is done in favour of designing and extending a single preferred model, we have. This is similar to the approach used by Ofwat in the March 2018 Consultation, and to that used by many of the water companies that made submissions.
Preference for simple models over more complex ones wherever possible	Although complex models may provide some insight, they are more sensitive to assumptions and harder to interpret. In many cases, the simpler model can act as the 'default' model where other approaches are less reliable. This is consistent with the CMA's preference as expressed in the Bristol determination.

Source: Frontier Economics

4.2 Discussion of model specification

Our modelling process has built directly on the models put forward by Ofwat in the March consultation, since simple changes and extensions are easier to reconcile with Ofwat's existing thinking. In doing so, we focused on models of bad debt costs as these are most relevant to modelling propensity to default.

Since bad debt costs represent roughly half of total retail costs, we consider that drivers that explain bad debt (such as average bill size or propensity to default on payment of water bills) should be included in models used to estimate efficient total retail costs. For example, if the average bill size is an important driver of bad debt, it is also likely to be an important driver of total retail costs.

The starting point of our econometric work was to reproduce the Ofwat model results presented in the consultation document and to check that our results aligned with Ofwat's estimates. These checks are set out in Annex A of this report.

As described in Section 4.2.1 below, although we do not consider the use of time trends we have tested panel models, both fixed effect and random effects, where appropriate. Also, as described in Section 4.2.2, we opted not to model economies of scale and scope. As noted previously, we focus on how best to model propensity to default in bad debt cost models.

4.2.1 Use of time series data

The dataset we have used to estimate the benchmarking models consists of four years of data for each water company[24]. When running models on all of these observations (as opposed to company-level averages) we have used clustered standard errors, consistent with Ofwat's practice to manage correlation of residuals across years.

We have not considered time trends because using them would make the models less suitable for forecasting as additional assumptions would be required in order to do so. Our approach is in line with the PwC review of the models of doubtful debt submitted by South West Water as part of PR14.

We have used data for each year since using data averaged over the modelled time period would remove the available data with which to carry out the analysis. Having said this, we accept that Ofwat's use of the four year average values for ORDC6 provides some useful insights into the stability of the results over time.

4.2.2 Economies of scale and scope are not modelled

Ofwat has not typically allowed for economies of scale when determining the efficient level of costs when there has been scope for achieving such economies (e.g. through outsourcing or consolidating certain activities with another company). This is because it could be argued that achieving such costs savings are within the control of the water company and therefore not truly exogenous. Therefore, we do not consider this in our analysis.

If there is a fixed cost that relates to the provision of both water and waste water then there would be economies of scope. We have followed Ofwat (and most other companies) in not accounting for the scope of services in bad debt models. This is because we have been asked by Thames Water to focus on the costs of bad debt and most companies (such as UU, South Staffs Water, Welsh Water, and Yorkshire) only considered the scope of services when modelling total retail costs rather than for bad debt specifically. Only one submission considered the scope of services specifically in the context of bad debt costs: Wessex Water and Bristol Water. The submission, however, noted that modelling the effects of this led to coefficients that were too difficult to interpret as some companies do not have any dual service customers.

[24] As of 2017, the dataset contains 17 unique water companies, however the merger between South West and Bournemouth means that the full dataset includes 19 unique water company names

5 MODELLING RESULTS

In this section, we set out the results of our econometric modelling results. For each of the model specifications investigated, we describe the interpretation of the model, and discuss whether that interpretation accords with economic intuition. We have mainly focused on extending the models proposed by Ofwat in its March consultation, and we compare results of extended models with those of Ofwat's models. Therefore, although we have tested other functional forms,[25] we have preferred models where the relationship between average bill size and debt cost per household hold at just over one, and where the interpretation of other explanatory variables is clear and plausible.

In this section, we set out:

- A discussion of the Ofwat modelling results (Section 5.1);
- Our findings relating to deprivation and income (Section 5.2);
- Our findings relating to the use of the proportion of lone parent families/households as a proxy for vulnerability (Section 5.3); and
- A summary of our modelling results (Section 5.4).

5.1 Ofwat modelling results

Ofwat presented six different specifications of bad debt models (see Figure 14 below for a summary). Each of these included average bill size as an explanatory variable. The additional explanatory variables are listed in the table below. ORDC6 averages the data for each of the variables over the four year period (hence the smaller sample size compared to the other models) in order to smooth for year-on-year volatility in the reporting of costs.

Ofwat found that each of the specifications delivered statistically significant results with coefficients with signs and magnitudes that are economically intuitive. The results of the R squared, VIF (max) and reset tests for each of the models showed that the specifications were appropriate.

The models show the following.

- There is a strong correlation between average bad debt costs and average bill size and that as the bill size increases, not only does the cost of default increase but customers also appear to become marginally more likely to default. This is in line with what we would expect.
- There is some evidence of economies of scale with the coefficient on number of households being as expected but weakly significant at best.
- The proxies for the propensity to default are found to be significant.
  - There is a negative correlation with both historic default rates and credit risk scores.

[25] These include the log of total debt costs, and the debt cost per household without logs.

- There is a positive correlation with the income deprivation component of the IMD.

- The results remain broadly consistent when considering a four year average of costs (as per ORDC5) implying that year-on-year volatility in cost reporting is not substantial.

Figure 14 Ofwat model results

Consultation model ID	ORDC1	ORDC2	ORDC3	ORDC4	ORDC5	ORDC6
Dependent variable	--- ln(bad debt per household) ---				sample avg	
Ln(number of households)			-0.128* (0.083)	-0.032 (0.629)		-0.053 (0.601)
Ln(bill size)	1.160*** (0.000)	1.138*** (0.000)	1.341*** (0.000)	1.183*** (0.000)	1.095*** (0.000)	1.168*** (0.000)
HHs with default (%) (Eq lpcf62)	0.050*** (0.006)		0.068*** (0.004)			
Income deprivation domain (%)					0.058** (0.032)	
Credit risk score (Eq rgc102)		-0.032** (0.034)		-0.034** (0.034)		-0.036* (0.067)
Constant	-5.479***	0.393	-5.204***	0.888	-4.580***	1.467
R2 adjusted	0.79	0.773	0.803	0.771	0.774	0.789
VIF (max)	1.03	1.078	2.843	2.152	1.178	2.221
Reset test	0.146	0.257	0.153	0.352	0.018	0.477
Estimation method	OLS	OLS	OLS	OLS	OLS	OLS
N (sample size)	71	71	71	71	71	17

Source: March 2018 consultation, Appendix 1: modelling results

There appears to be a formatting error in the table presented by Ofwat and that only ORDC6 is based on four year averages of all (independent and dependent) variables. In particular, we understand that ORDC4 considers economies of scale without any averaging over the four year period. We also consider that ORDC5 is based on the full four years as we have been able to replicate the results on that basis.

5.2 Deprivation and income

As we describe in Section 3.4.1, the income domain of the IMD is a relatively crude measure to use in modelling the propensity to default on water bills. Therefore, in this sub-section we consider whether this measure (and therefore models including it as a driver of propensity to default) could be improved in the following ways:

- By considering the proportion of an area's population that are very deprived rather than the simple average of income deprivation using the Ofwat model (ORDC5)– see Section 5.2.1;

- Using data on absolute (mean) income rather than on income deprivation (Section 5.2.2); and

- Using data on absolute (mean) income while accounting for housing costs (Section 5.2.2).

In each model the dependent variable is log of bad debt costs per household. All models are estimated with OLS, with clustered standard errors. Each of the figures below show standard errors in parentheses.

5.2.1 Variations in income deprivation within areas

ORDC5 uses data on the income component of the IMD. This is the average score for the income component (measuring the proportion of the population experiencing deprivation relating to low income) of the areas within each water company on. Using the average score means that a water company with an equal number of very high and very low deprivation areas will have the same average income score as a company with average levels of deprivation throughout. However, we would expect the former to have higher levels of default on water bills.

Therefore, in order to test whether the distribution of income deprivation is important to bad debt costs, we replicated the Ofwat model but instead used the proportion of a water company's customers that live in an area of high deprivation as measured by the proportion of households in each company's area that are in the top decile for income deprivation in England or Wales.

Figure 15 compares the results of Ofwat's default model with our adjusted model. It can be seen that the adjustment has limited impact on the relationship between average bill size and bad debt costs. While the coefficient is positive (as we would expect), and other measures of the adequacy of the model are strong (such as the R2 or RESET test), this model is inferior to Ofwat model as the coefficient lacks any statistical significance.

Nevertheless, the presence of pockets of very income deprived areas may still be an important driver. In densely populated and highly diverse areas like London or other urban centres, such pockets could be as small or smaller than an LSOA area. This would mean that the income IMD is insufficiently granular to capture the drivers of the propensity to default at the water company level.

Figure 15 Income deprivation model estimates

	Average income deprivation score ORDC5	Population in top 10% income deprivation ORDC5 adjusted
Log Avg. Bill Size	1.095***	1.172***
	(0.094)	(0.106)
IMD Income Score – area average	5.810**	
	(2.502)	
IMD Income Score – Proportion of population in 1st decile		1.395 (1.433)
Constant	-4.323***	-3.585***
	(0.590)	(0.990)

Source: Frontier Economics analysis and replication of Ofwat modelling

5.2.2 Mean income before and after housing costs

The income deprivation component of the IMD is based on a headcount and therefore does not reflect how the levels of incomes for those people might vary. It also does not allow us to consider whether the propensity to default changes as levels of income change. Therefore, in this sub-section we use an estimate of mean absolute incomes at a local level to analyse the relationship between income and bad debt costs in more detail.[26]

We also considered whether the cost of housing relative to average incomes could lead to affordability issues. For example, although average incomes might be higher in the South of England, disproportionately higher housing costs may mean that consumers have relatively lower disposable income available to pay water bills.

However, as described below, while both models have coefficients in the correct direction and with reasonable intuition, neither provide a better fit than those using income deprivation scores (see the second column of Figure 16).

Figure 16 Mean income as a driver bad debt costs

	Before housing costs	After housing costs
Log Avg. Bill Size	1.159***	1.138***
	(0.119)	(0.119)
Log Mean income before housing costs	-0.549 (0.625)	
Log Mean income after housing costs		-0.705 (0.662)
Constant	-0.644	0.39
	(4.199)	(4.399)

Source: Frontier Economics

We make the following observations.

- The coefficient on average bill size is as expected in both models.
- For mean income before housing costs (first column), the coefficient predicts that a 1% increase in average income would reduce bad debt costs by 0.55%. This impact is in the expected direction, but the standard error is too large for us to conclude that it is significantly different from zero.
- For mean income after housing costs (second column), the coefficient is slightly more significant and impactful. In this model, a 1% increase in income after housing costs translates into a 0.7% reduction in bad debt costs per household. Since the estimated coefficient is further from zero and the standard errors remain similar, the model using income after housing costs seems slightly stronger than income before housing costs. However, the differences are still too small to draw strong conclusions.

While there are conceptual merits in considering variants of income deprivation (as discussed above), the models we considered do not produce statistically significant results. This is likely to be because of the limitations of data available, and the complexity of the relation between aspects of income deprivation (including housing costs) and bad debt costs. In Section 6, we provide recommendations on how data could be improved to potentially support better modelling of the propensity to default in bad debt models.

[26] The ONS produces an model-based statistic of mean incomes at the Middle-layer Super Output Area (MSOA) level. MSOAs include 5 LSOAs on average, i.e. around 6,000 households.

5.3 Proxying for vulnerability

As discussed in more detail in Section 3.4, certain demographic groups face a much higher risk of vulnerability and are therefore more likely to default on water bills. Including proxies for these drivers in the modelling of bad debt costs produces robust and strong impacts, and allows to account for factors that are well recognised (including by Ofwat) to affect households facing affordability problems.

5.3.1 Lone parent families

As described in Section 3.4.2. Ofwat found that certain categories of households were more likely to default on their water bills and it identified lone parent families as being more likely to face affordability issues. Therefore, we use the baseline model to consider the importance of lone parent families in two separate ways:

- First, we consider lone parent families as a percentage of all families; and
- Second, we consider lone parent families as a percentage of all households – that is, lone parent families that do not live with other family members such as grandparents.

We do not consider that being a lone parent household in itself leads to a higher propensity to default. While for a given region, the proportion of households that are lone parent is correlated with other characteristics such as vulnerability and higher costs associated with bad debt, this does not imply any causation between these variables.

Figure 17 shows the results of our analysis. We make the following observations.

- The coefficient on average bill size remains significant and close to one.
- Both models find the vulnerability measure (i.e. lone parent families / households) to be statistically significant. There is a substantial difference, however, in the coefficients for these measures. The coefficient on lone parent families (households) implies that a percentage point increase in the proportion of lone parent families would increase bad debt costs per household by 0.76% (1.99%). One reason for the substantial difference could be that using lone parent households more effectively identifies vulnerable families without parental or extended family support. It may also be the case that lone parents as a proportion of households are a better proxy for the vulnerable populations of interest, even though the propensity to default of the two populations is identical.

Figure 17 Estimated models with vulnerability variables

	Lone parent families	Lone parent households
Ln(Average bill size)	1.164***	1.169***
	(0.102)	(0.097)
Lone parent families (%)	7.593***	
	(2.190)	
Lone parent households (%)		19.875***
		(6.521)
Constant	-5.482***	-5.548***
	(0.677)	(0.737)

Source: *Frontier's calculations of Ofwat data and census data*

Note: *Lone parent families are the proportion of families in the area that are lone parent. Lone parent HHs is the proportion of households in the area that are lone parents with dependent children, and no other inhabitants.*

5.3.2 Assessing vulnerability versus income deprivation

However, given that vulnerable populations tend to have lower incomes on average, there is a risk that the lone parent families/households variable is significant because it is a proxy for income deprivation rather than for vulnerability itself. As we describe in Section 3.4, there is an important distinction between deprivation and vulnerability.

We have assessed whether household composition is proxying for income deprivation rather the vulnerability by including household composition in Ofwat's income deprivation model – the results are shown in Figure 18. These specifications show coefficients on the vulnerability variable significantly reduced, but still larger than those on income deprivation and significant in one of the models. The VIF statistic, which tests for multicollinearity (i.e. correlation between the explanatory variables), is less promising in these models than in those with just vulnerability or income deprivation. However, while these models may not be suitable as final models for benchmarking, they indicate that vulnerability – for instance, proxied by lone parent families – is a significant proxy for the propensity to default.

Figure 18 Modelling vulnerability while controlling for income deprivation

	ORDC5 plus lone parent families	ORDC5 plus lone parent households
Ln(Average bill size)	1.129***	1.131***
	(0.099)	(0.096)
Lone parent families (%)	5.473*	
	(2.854)	
Lone parent households (%)		13.224
		(7.799)
IMD Income	2.352	2.549
	(3.682)	(3.751)
Constant	-5.242***	-5.234***
	(0.551)	(0.583)

Source: Frontier's calculations of Ofwat data and census data

The size and sign of the coefficients on the percentage of lone parent families/households offers intuitive results. An increase in the number of lone parent families/households leads to a plausible increase in total bad debt costs. The presence of lone parent families/households is strongly correlated with other types of vulnerable households (as described in 3.4.2). Therefore, any increase in bad debt costs is also likely to be attributable to the increase in those types of households (not just to lone parent families/households).

Nevertheless, it is likely that other vulnerable populations that are not correlated with lone parent families/households should also be accounted for (such as single pensioners) However, the sample size of the econometric model is limited by the number of water companies, making it difficult to include and test more than one measure at a time. Consequently, as discussed in Section 3.4, we consider that it would be appropriate to work on designing a composite measure of vulnerable demographics. In the interim, vulnerability as proxied by lone parent families or households is demonstrably an important factor in determining bad debt costs which merits further consideration.

5.4 Summary of modelling results

Overall, we find that the three most robust models are Ofwat's model using defaults, Ofwat's model using the credit risk score and the Frontier model using lone parent families. While each of the models could be improved with better data (see Section 6), they each meet the principles set out by Ofwat. In particular, they:

- Are based on sound economic and operational understanding;
- Produce statistically significant results and have robust model specifications; and
- Have coefficients that are of plausible sign and magnitude.

The Ofwat model using data on income deprivation (ORDC5) is unlikely to provide a reasonable model of bad debt costs mainly because there are likely to be additional relevant factors that should be accounted for (such as income volatility, income levels, housing costs, transience and so on). Further, as it based on relatively old data for only one period in time, forecasting for the purposes of PR19 will be problematic. While attempting to control for housing costs relative to income does not produce statistically significant results, given the limitations of the data available, we consider that there is insufficient evidence available to conclude that these are not relevant drivers.

Compared to the Ofwat models, the Frontier model has the following strengths in relation to the additional principles we set out in Figure 13 in Section 4.1.

- **Captures many relevant drivers**
  While it does not cause bad debt in itself, there are clear economic reasons why lone parent families may be more likely to face affordability issues compared to other household types (see Figure 8).

  It is strongly correlated with other factors associated with vulnerability and affordability issues (including working age people living alone and households with at least one person with a disability, although not with single pensioners).

  It is not specific to any one product or set of products meaning that it will not be biased by the purchasing characteristics of products that are different to water.
- **Exogenous to the water company's actions**
  It is beyond the control of water companies and therefore using it to inform cost allowances for PR19 and beyond will have stronger incentive properties than using historic default rates.
- **Potentially able to predict with reasonable accuracy**
  Data on lone parent households may be available on a more frequent basis from private agencies like Equifax (we have used data from the census). This may make forecasting less problematic compared to forecasting with the IMD.

6 RECOMMENDATIONS

In this section we set out our recommendations for interpreting and improving the Ofwat econometric modelling for the purposes of determining regulatory cost allowances for the PR19 period. We focus on how to make the models more robust as well as the need to capture relevant drivers (for which values can be forecast with reasonable accuracy) and provide incentives for efficiency gains.

6.1 Summary of recommendations

Although the Ofwat models capture a robust relationship between debt costs, bill size and the measures for propensity to default included by Ofwat, there are a number of areas in which the models can be improved. As such, they should be used with care when setting cost allowances for PR19. The bullet points below summarise our recommendations which we then describe in more detail in the rest of this section.

- The use of default rates offers statistically robust and economically intuitive results and could provide a suitable interim measure for PR19. However, in future review periods and in the absence of other measures, it may lower efficiency incentives as default rates are partly within the control of companies (see Section 6.2).
- Future models should consider vulnerability and deprivation as the underlying drivers of the propensity to default. Further work should be carried out to understand how to best model these factors (see Section 6.3).
- Given the limitations of the data (both cost data and drivers) currently available in terms of both the level of granularity and the frequency of reporting, we recommend that subsequent price controls seek to address these issues (see Section 6.4).
- Cost allowances should be set with caution to ensure that companies are not penalised for the weaknesses of the model rather than true inefficiency (see Section 6.5).

6.2 Default rates as an explanatory variable

As described in Section 5, it is difficult to develop a model of the external drivers of bad debt costs that is both comprehensive and statistically robust. Historical default rates can be used as a proxy for the propensity to default as an interim solution for the purposes of PR19. Given that the Ofwat models that include defaults and credit risk scores produce similarly plausible models, and that the theoretical case for historical defaults appears stronger (see Section 3.3), we consider that default rates is the more appropriate variable to include. Although using the income domain of the IMD leads to a statistically strong model, income deprivation is a relatively crude measure that offers limited scope for forecasting efficient costs over the price control period. Therefore, for PR19, the Ofwat models using default rates and credit risk scores are likely to be more appropriate than the

Ofwat models using income deprivation as they can potentially capture a wider range of relevant factors.

Using the default rate may provide a useful interim solution if the model results are interpreted with caution given our findings on the wider drivers of the propensity to default on water bills. One way to do this could be through the design of the price control, such as using a glide path to mitigate the risk of penalising water companies for the weaknesses of the model rather than the inefficient behaviour over which they have control.

Another option could be to combine results from different models. Ofwat's consultation is very useful in this regard. Recognising that a single best model is unlikely to exist, it may be appropriate to use a similar approach to the triangulation approach that Ofwat recommends for other parts of the business plan. Further, we recommend that decisions about how to interpret the models and translate results into efficient cost allowances should reflect the assessed quality of the models.

Nevertheless, we consider that using default rates in future price controls has an important drawback. In particular, the default rate – even across all bills – is something over which the water company has some control. This means that it cannot be considered an external factor (i.e. it suffers from the problem of endogeneity). In particular, if a higher forecast default rate resulted in a higher cost allowance, companies would have limited incentive to reduce the default rate as it would be possible to recover costs associated with defaults. Therefore, if it is used as an interim measure, making clear that it will not be used for future price control periods will help to incentivise efficiency over the PR19 period.

6.3 Vulnerability

In order to model bad debt costs, some measure of or proxy for vulnerability should be included. In our modelling we have explored the impact of lone parent families, either as a proportion of all families or as a proportion of households. Our results show that, while correlated, the size of vulnerable populations is distinct from income deprivation as a driver of bad debt costs. We have also shown that the proportion of vulnerable demographics in a water company area (e.g. lone parent families with dependent children, which is also correlated with other vulnerable household types such as working age adults living alone) has a large and statistically significant impact on estimated bad debt costs. While variables measuring the proportion of lone parent families do not fully capture the demographics of vulnerability, even this indirect measure has a large impact on modelling results.

Given its importance as a driver of default and bad debt costs, not including some vulnerability measure in a benchmarking model risks creating an inaccurate set of cost allowances. This could, in particular, disadvantage companies serving a more vulnerable customer base.

6.4 Further development of data

There are substantial limitations to the data currently available in relation to different components of deprivation and vulnerability. We recommend that, while not feasible for PR19 given the time constraints, work in subsequent price controls should seek to address these. Also, the number of water companies limits the sample size and hence the scope for considering multiple variables. Therefore, we recommend focussing on developing appropriate composite measures of deprivation and vulnerability. Requiring water companies to collect and provide more granular detail on the levels of bad debt costs would also help to increase the sample size. This would also make it easier to develop a robust composite measure of deprivation and vulnerability. Data on the following areas could support the development of a composite index. This data may also be relevant to other utility regulators, making co-ordination between them beneficial.

Figure 19 Areas where additional data is required

Income levels and income volatility	Living expenses including housing costs	Transience	Triggers of risk of vulnerability (as set out in Table 7)	Components of deprivation from the IMD (as set out in Figure 6)

Source: Frontier Economics

Specific characteristics of suitable data include those set out in the table below.

Figure 20 Ideal features of data on the drivers of the propensity to default

Regular collection and reporting Data that is at least annual allows the models to consider variation in characteristics in each year. It also makes forecasting easier	**Consistent** A composite measures of vulnerability and deprivation should be developed in parallel so they are complementary and not overlapping or contradictory.	**Comparability across all water company areas** Data should be recorded and processed consistently across all geographic areas including England and Wales
Possibility to consider the variability as well as averages for a given area For example, while two areas may have the same average level of income, it is possible that one has very high and very low incomes and that the other has more uniform incomes. Households in the former area are more likely to face affordability issues but this is not reflected in the use of averages.	**Specific to water bills** The use of a composite measure of deprivation or vulnerability should use weights for the individual components that reflect the characteristics of the propensity to default on water bills (unlike the IMD which uses weights designed to reflect general deprivation).	**Sufficient geographic disaggregation** This is to allow for the mapping of data to water company areas at a minimum. If data on bad debt costs is available at a more granular level then the data on explanatory variables should also be at that level.

Source: Frontier Economics

6.5 Company-specific circumstances

In determining the cost allowances for water companies, there should be due consideration of the specific circumstances that affect individual water companies (for example, through cost adjustment claims). In particular, the small sample size means that it is not feasible to develop a statistically robust model that accounts for factors affecting only one company or a small number of companies to a material extent. In Section 5, we gave the example of how housing costs as a proportion of income varies across geographic areas and, through its impact on affordability, this could potentially affect water companies to differing extents. Another example that we did not consider in our report is transience. Furthermore, it would be reasonable for Ofwat to take into account the data limitations identified in this report in assessing the claims for individual company-specific circumstances.

ANNEX A REPRODUCING OFWAT'S ANALYSIS

Our modelling has primarily built on, and tested extensions to, the models put forward by Ofwat in the March consultation (the ORDC models). To do so, we have recreated the ORDC specifications and checked that our results align with Ofwat's estimates. These are set out in an annex to this report. Figure 21 shows side by side comparisons of the Ofwat results for the ORDC models and our results, between which there are no differences. That said, Figure 21 also shows small but important differences in how we have reported our results that should be kept in mind.

First, any coefficient on a percentage variable (such as % defaults or income component score) is 100 times higher in our estimates, as Ofwat has multiplied the proportion value before use to put in percentage terms. This makes our effects appear larger, but reflects an identical statistical relationship.

Second, we have reported standard errors where Ofwat reports the p-value on a test of whether the coefficient is different from zero. We consider that the significance of a coefficient's difference from zero is already captured by the asterisk notation, and that standard errors give a better picture of the uncertainty involved in our estimates.

Figure 21 Comparison of ORDC results

	ORDC1 (Ofwat)	ORDC1 (Frontier)	ORDC5 (Ofwat)	ORDC5 (Frontier)
Log Avg. Bill Size	1.160***	1.160***	1.095***	1.095***
	(0.000)	(0.090)	(0.000)	(0.094)
% Defaults	0.050***	4.998***		
	(0.006)	(1.598)		
IMD Income Score			0.058**	5.810**
			(0.032)	(2.502)
Constant	-5.479***	-5.479***	-4.580***	-4.580***
	(NR)	(0.697)	(NR)	(0.573)
Adj. R2	0.79	0.79	0.774	0.774
VIF	1.03	1.03	1.178	1.178
RESET	0.146	0.146	0.018	0.018
N	71	71	71	71

Source: *Frontier calculations of Ofwat data and the Ofwat March Consultation on retail cost benchmarking*

Note: *In the columns labelled "Ofwat", the numbers reported in parentheses are the p-values on tests that the coefficient is significantly different from zero. In the columns labelled "Frontier", the values in parentheses are the standard error of the coefficient. The asterisks denote equivalent levels of significant difference from zero in both sets of estimates.*

ANNEX B MODELLING RESULTS

In this Annex we set out further detail on:

- How other water companies have considered transience in their responses to the Ofwat consultation;
- Comparison of ORDC3 with the "baseline" model we used as the starting point for our analysis;
- Detailed comparison of the Ofwat model and the results of our models using the number of households in the 10% most deprived areas;
- Detailed comparison of the Ofwat model and the results of our models using mean absolute income before and after housing costs; and
- Correlation between the different proxies for vulnerability and a detailed comparison of the Ofwat model and the Frontier lone parent family model.

B.1 Consideration of transience

Figure 22 Summary of models submitted that consider transience explicitly

Submission	Data used and comments
Thames Water	ONS data on: • Total internal migration - propensity of people to migrate from/to UK local authorities, sum of inflows and outflows; and • Total international migration - propensity of people to migrate from/to UK local authorities and abroad, sum of inflows and outflows (ONS). Thames Water's econometric analysis shows that transience has the expected effect on the cost of bad debt. The significance of these two types of transience (internal, international) depends on the model specification.
Wessex and Bristol	The Wessex and Bristol analysis considers internal population total flow (%). It also finds that transience has the expected effect on the cost of bad debt. The magnitude of the coefficient on this variable (0.091) is broadly in line with the coefficient estimated by Thames (0.03 - 0.291, depending on the specification).
Yorkshire Water	Census data extrapolated forward for 2016 using regional data from the ONS on: • Proportion of the population privately renting for the bad debt model; and • Proportion of population in social housing for the other retail costs model. Yorkshire's analysis finds these variables to have the expected sign yet without being significant.

Source: Ofwat Appendix 1 – modelling results, March 2018

B.2 Comparison of ORDC3 with "Baseline" model

Figure 23 Simple estimate of log debt costs per household

	Baseline	ORDC3
Log Avg. Bill Size	1.220***	1.341***
(standard error)	(0.129)	(0.122)
% Defaults		6.850***
(standard error)		(2.051)
Log of number of HHs		-0.128*
(standard error)		(0.070)
Constant	-4.480***	-5.204***
(standard error)	(0.714)	{0.636}
Adj. R2	0.733	0.803
VIF	1	2.843
RESET	0.002	0.153
Observations	71	71

Source: Frontier calculations using Ofwat data

Note: The baseline model is identical to the ORDC models, but without any explanatory variables controlling for deprivation or propensity to default. The VIF test in the baseline model is exactly equal to 1 as it tests for the presence of multicollinearity and there are no other regressors.
The ORDC3 model is Frontier's recreation of the ORDC3 model reported by Ofwat in the March Consultation – there may be slight differences in estimated coefficients due to rounding errors. The coefficient on the % Defaults variable is 100 times that reported by Ofwat, as we have maintained it as a proportion variable, whereas Ofwat converted the proportions to percentages.

B.3 Income deprivation

Figure 24 Income deprivation model estimates

	Average income deprivation score ORDC5	Population in top 10% income deprivation ORDC5 adjusted
Log Avg. Bill Size	1.095***	1.172***
	(0.094)	(0.106)
IMD Income Score – area average	5.810**	
	(2.502)	
IMD Income Score – Proportion of population in 1st decile		1.395
		(1.433)
Constant	-4.323***	-3.585***
	(0.590)	(0.990)
Adj. R2	0.741	0.74
VIF	1.096	1.175
RESET	0.217	0.309
Observations	71	71

Source: Frontier Economics analysis and replication of Ofwat modelling

B.4 Mean income

Figure 25 Mean income as a driver bad debt costs

	Before housing costs	After housing costs
Log Avg. Bill Size	1.159***	1.138***
	(0.119)	(0.119)
Log Mean income before housing costs	-0.549	
	(0.625)	
Log Mean income after housing costs		-0.705
		(0.662)
Constant	-0.644	0.39
	(4.199)	(4.399)
Adj. R2	0.739	0.744
VIF	1.181	1.229
RESET	0.279	0.439
Observations	71	71

Source: Frontier Economics

B.5 Vulnerability

Figure 26 Correlation of vulnerability variables

% households where…	A working age person lives alone	Lone parent lives with only dependent children	At least one person has a disability	A pensioner lives alone
A working age person lives alone	1			
Lone parent lives with only dependent children	0.89	1		
At least one person has a disability	0.72	0.71	1	
A pensioner lives alone	0.57	0.41	0.85	1

Source: Frontier calculations of NOMIS Data from 2011 Census

Note: These correlations are at the water company level. At the LSOA level, roughly 1000-1500 people per area, these correlations become significantly less strong, for example as lone parents and single working age households may disproportionately live nearby but not in the exact same neighbourhoods.

Figure 27 Estimated models with vulnerability variables

	Lone parent families	Lone parent households
Ln(Revenue per household)	1.164***	1.169***
	(0.102)	(0.097)
Lone parent families (%)	7.593***	
	(2.190)	
Lone parent HHs (%)		19.875***
		(6.521)
Constant	-5.482***	-5.548***
	(0.677)	(0.737)
Adj R2	0.782	0.778
VIF	1.029	1.027
Reset	0.105	0.12
Observations	71	71

Source: *Frontier's calculations of Ofwat data* and census data

Note: *Standard errors are in parentheses*
In each model the dependent variable is log of bad debt costs per household. Lone parent families are the proportion of families in the area that are lone parent. Lone parent HHs is the proportion of households in the area that are lone parents with dependent children, and no other inhabitants.
All models are estimated with OLS, with clustered standard errors.

Figure 28 Vulnerability measures controlling for income deprivation

	ORDC5 plus lone parent families	ORDC5 plus lone parent households
Ln(Revenue per household)	1.129***	1.131***
	(0.099)	(0.096)
Lone parent families (%)	5.473*	
	(2.854)	
Lone parent HHs (%)		13.224
		(7.799)
IMD Income Score	2.352	2.549
	(3.682)	(3.751)
Constant	-5.242***	-5.234***
	(0.551)	(0.583)
Adj R2	0.782	0.778
VIF	2.736	3.303
Reset	0.048	0.023
Observations	71	71

Source: *Frontier's calculations of Ofwat data and census data*

Note: *Standard errors are in parentheses*
In each model the dependent variable is log of bad debt costs per household. Lone parent families are the proportion of families in the area that are lone parent. Lone parent HHs is the proportion of households in the area that are lone parents with dependent children, and no other inhabitants.
All models are estimated with OLS, with clustered standard errors.

www.frontier-economics.com