

# EVALUATING AN ‘ESTIMATED IMD’ APPROACH TO MODELLING THE IMPACT OF DEPRIVATION ON DEBT-RELATED RETAIL COSTS

A report for Welsh Water

Client report | March 2018



At PR19, Ofwat will use an econometric approach to set allowed costs for household retail. When setting allowed costs in relation to doubtful debt and debt management, Ofwat has further indicated that it will explore including deprivation as a driver of efficient costs. Given that there are a wide range of ways in which deprivation could be reflected in cost assessment, care must be taken to ensure that any approach does not (inadvertently) set allowed costs above, or below, the efficient level. Here, one important issue is that deprivation is a multifaceted concept, where its various dimensions, even if correlated, might individually impact costs in distinct ways. A specific related issue is that the overall Index of Multiple Deprivation (IMD), which captures the ‘broader’ dimensions of deprivation, is not available on a consistent basis for England and Wales. In this paper, we establish that robust estimated overall IMD scores can be developed for Wales, and that these are also robust when included in cost assessment models.

## 1. Introduction and executive summary

### 1.1 Key context

At PR19, Ofwat will set allowed costs for household (HH) retail using an econometric approach. In its final methodology, Ofwat set out that it would develop models across three categories of costs:

- total residential retail costs;
- bad debt plus debt management costs (which, here, we refer to as ‘*debt-related retail costs*’); and
- total residential retail costs, *less* bad debt and debt management costs.

In addition to describing the overall categories of costs being modelled, Ofwat provided guidance as to some of the potential drivers of (efficient) costs that it would control for within its econometric modelling. Here, Ofwat stated: “*our models will account for the impact of dual bill versus single bill customers and differences in the cost to serve a metered customer versus an unmetered customer. We will consider bill sizes and **deprivation levels** in areas served by water companies when assessing bad debt levels.*”<sup>1</sup> (emphasis added).

The inclusion of socioeconomic deprivation as a potential explanatory variable in setting cost allowances reflects the fact that, as is widely accepted, deprivation can

---

<sup>1</sup> [‘Delivering Water 2020: Our final methodology for the 2019 price review Appendix 11: Securing cost efficiency.’ Ofwat \(2017\); page 21.](#)

materially drive debt-related retail costs in a manner that is (primarily) outside of efficient management control.<sup>2</sup>

In practice, at PR19 there are a number of ways in which deprivation can be measured and included in cost benchmarking models. However, it is difficult to determine, a priori, which of these measures might *best capture* the way in which deprivation impacts company debt-related retail costs. An issue relating to this is that one potential measure – the overall IMD – which is the UK Government’s official measure of deprivation, *is not available on a consistent basis for England and Wales*.

Whilst the absence of a (comparable) overall IMD score for Wales does not prevent the development of econometric debt-related cost assessment models, it matters because the overall IMD is a ‘*broad*’ measure of deprivation (which, as we explain subsequently, is intended to reflect the fact that deprivation is multifaceted). In contrast, most other measures of deprivation are ‘*narrower*’. Therefore, there is a potential concern that, if the broader aspects of deprivation drive efficient debt-related costs (either for *specific companies*, or across *the industry more broadly*) econometric models that rely on narrower definitions of deprivation could, inadvertently, result in allowed costs being set above, or below, the efficient level.

## 1.2 Scope of our work

There are various reasons to suppose that the broader aspects of deprivation do affect debt-related retail costs. As such, Welsh Water (Welsh) commissioned Economic Insight to take forward analysis to explore this issue. Consequently, the main scope of our work was to:

- (i) develop *estimated* overall IMD scores for the company; and
- (ii) demonstrate that these estimated values can be robustly used within retail cost assessment models.

We have also undertaken some *initial* modelling work to explore whether, objectively, our estimated IMD values for Welsh perform better than other potential deprivation measures at explaining variation in debt-related retail costs (by local authority, within Wales). This is not, however, the main focus of our paper.

## 1.3 Summary of key findings

The evidence and analyses set out in this paper establish the following:

- **That it is possible to estimate overall IMD scores for Welsh (both for the company as a whole and by local authority area).**
- **Further, that our predicted IMD estimates are robust and plausible,** insomuch that: (i) the preferred model we use to generate the overall IMD has good statistical properties; and (ii) the predicted values from our preferred model accord with credible existing evidence on deprivation – including a prior independent academic study, which also predicts IMD values for Wales (on a like-for-like basis with the overall IMD for England).

---

<sup>2</sup> This is consistent with Ofwat’s Final Determinations at PR14, whereby the regulator approved special cost factor (SCF) adjustment claims relating to the impact of deprivation on debt-related retail costs.

- That **our estimated overall IMD value for Welsh is also robust when used in bad-debt related retail cost assessment models**. Specifically, models including the overall IMD (drawing on our estimated value for Welsh) have high R2.
- **Initial analysis further suggests that, objectively, our estimated overall IMD scores better explain variation in debt-related retail costs** within Wales than alternative deprivation measures. This is consistent with the intuition that the multidimensional aspects of deprivation do drive debt-related retail costs – and that, therefore, alternative deprivation measures might result in allowed costs being set above, or below, the efficient level. Additional work would be required in order to: (i) further verify this finding; and (ii) determine the extent to which this issue is ‘industry wide’, or is more ‘company specific’.<sup>3</sup>

Drawing the above conclusions together, our recommendations are:

- » Ofwat should give careful consideration to the ‘breadth’ of deprivation as a concept and the ways in which its different dimensions might impact ‘efficient’ debt-related retail costs at PR19.
- » Ofwat should further consider the extent to which the ways in which these ‘broader’ aspects of deprivation impact costs, might be *company specific*, or are *industry wide*.
- » Relatedly, within the context of developing a suite of models for retail cost assessment at PR19, Ofwat should include the overall IMD, using predicted values for Welsh (for example, using the data and approach outlined in our report) within the pool of deprivation variables it considers.
- » Reckon’s existing estimate of the IMD score for Welsh is implausibly low, being inconsistent with existing evidence. Therefore, it should not be used for setting Welsh’s allowed debt-related costs. This does not, however, imply any broader criticism of Reckon’s work for United Utilities (UU).

#### 1.4 Structure of our report

Our report is structured as follows.

- Firstly, we describe how the IMD is constructed, and what it measures.
- Secondly, we set out a ‘first principles’ assessment of how deprivation impacts debt-related costs, considering the various dimensions of deprivation.
- Thirdly, we identify the key implications for debt-related cost assessment arising from the fact that the overall IMD is *not available for England and Wales on a consistent basis*.
- We then set out the details of our analysis, which we have structured around the following questions:
  - » **Can robust estimates of the overall IMD for Welsh be identified?** This matters because, unless reasonable measures of the (overall) IMD for

<sup>3</sup> As this would determine the extent to which the appropriate solution was to: (a) ensure ‘broader’ measures of deprivation, such as the overall IMD, were used in ‘cost assessment’, and / or, (b) whether the impact of the breadth of deprivation would need to be assessed more case-by-case, say through cost adjustment claims (these approaches are not mutually exclusive).

Welsh can be derived, this possibility for cost modelling is irrelevant. Here, we therefore consider both: (i) the statistical validity of the models developed to estimate the overall IMD for Welsh; and (ii) the intuitive reasonableness of the approaches identified.

- » **How do our estimated IMD scores for Welsh compare to other available evidence?** This is to help establish the overall ‘plausibility’ of our estimated results.
- » **Is our estimated overall IMD score for Welsh robust when used in retail cost assessment models?** Clearly, one practical use of our estimated IMD score for Welsh is to use it within industry benchmarking models – and so we need to establish if the measure performs well within such models.<sup>4</sup>

- Finally, we set out our conclusions and recommendations.

## 2. Overview of the IMD

The IMD is the UK Government’s official measure of deprivation. It is published both by the Department for Communities and Local Government (GCLG) for England; and Statistics for Wales (for Wales). The IMD provides a set of measures (indices) and an overall measure of relative deprivation for small areas (Lower-layer Super Output Areas - LSOAs).

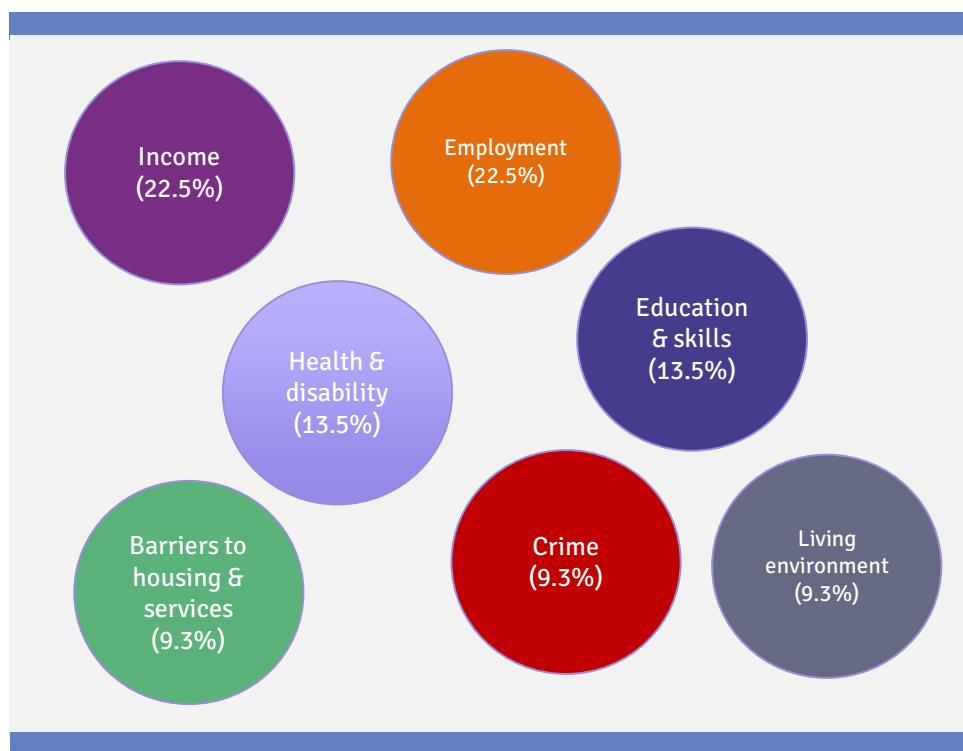
Focusing on the IMD for *England*, the ‘overall IMD’ score by LSOA is based on seven ‘domains’ (or types) of deprivation. The main methodological steps to deriving the overall IMD score are as follows:

- A ‘score’ is constructed for each of the seven domains of deprivation (each domain is made up of a wide basket of ‘indicators’ – or metrics).
- For each domain, the LSOAs are ‘ranked’, based on their score.
- The domain ranks are transformed to a specified exponential distribution.
- Finally, the exponentially transformed domain scores are combined using domain weights to form the overall IMD score, as shown in the following figure.

---

<sup>4</sup> *Noting, as previously set out, that the appropriateness of this turns primarily on whether one can determine if the ‘breadth’ of deprivation (being multidimensional) affects efficient retail costs for all companies, or is a more company specific issue.*

Figure 1: IMD domains



Source: *The English Indices of Deprivation 2015: Technical Report – Department for Communities and Local Government (September 2015)*.

The IMD was most recently published in England in 2015. For each indicator, the most ‘up-to-date’ data was used – which in most cases relates to 2012/13.

For Wales, the IMD was most recently published in 2014, with further revisions published in 2015. Whilst the general approach taken to construct the IMD for Wales is *broadly similar* to that for England, the overall IMD scores by LSOA in Wales cannot be compared to those in England. This is because:

- the IMD for Wales has eight domains, rather than the seven for England;
- related to the above, the ‘weights’ across overlapping domains therefore differ from those used to calculate the overall IMD scores in England (e.g. income has a weight of 22.5% in England, and 23.5% in Wales);
- due to the fact that the overall scores are derived from ‘ranks’, knowing the ranks of LSOAs ‘within Wales’ provides no information as to their rank within England; and finally
- the indicators (i.e. metrics) used to derive scores within certain domains can vary from those used in England.<sup>5</sup>

Collectively, the above means that (as is widely accepted) overall IMD scores cannot be compared across England and Wales. Two specific domains within the overall IMD do, however, appear to be comparable, albeit with (minor) caveats – as follows:

- The ‘employment’ score measures the proportion of the working-age population that is (involuntarily) excluded from the labour market. It is calculated as the proportion of (working-age) population that meet certain criteria where, from the

<sup>5</sup> All information taken from *Welsh Index of Multiple Deprivation 2014 (WIMD 2014): Technical Report.* Statistics for Wales (2014).

respective IMD Technical Reports, those criteria seem to be identical across England and Wales (save for some ambiguity regarding whether the carer's allowance is included in the latter).<sup>6</sup>

- The 'income' score captures the proportion of the population experiencing deprivation relating to low income. It is, again, calculated as the proportion of the population meeting certain criteria. On reviewing the Technical Reports, those criteria appear to be identical, or very similar, across England and Wales.<sup>7</sup> However, in Reckon LLP's report for UU, Reckon state that IMD income is not comparable, because the measure includes the proportion of families below 60% of median income (and median income will vary between England and Wales).<sup>8</sup> We disagree with this conclusion. Specifically, and set out in the previous footnote, only one of the indicators *within* the income domain includes the median income of England / Wales (the child tax credit criteria). Therefore, the impact of median incomes between England and Wales differing *only affects this narrow aspect*. Furthermore, as the respective Technical Reports also explain, this sub-criterion within the income domain only captures the proportion of the population *not already captured by other criteria* (which should further mitigate differences in median income between England and Wales).

*'The central idea of the Index of Multiple Deprivation is that deprivation is multidimensional and can be experienced in relation to a number of distinct domains.'* – IMD (England) Technical Report

Following from the above, we think both the IMD income and employment scores are sufficiently similar across England and Wales that it would be reasonable to explore using either or both when evaluating differences in deprivation (and its impact on debt-related costs) across companies. However, in relation to the income score, one might expect this to be somewhat understated for Wales.

The fact that two specific domains of the IMD may be considered sufficiently comparable is separate from the main issue of interest in our paper, however. Specifically, with reference to the overall IMD scores (which, as previously explained, cannot be compared across England and Wales) it is critical to keep in mind 'why' it has been constructed in the way that it has. Here, the Technical Report for England states: *"the central idea of the Index of Multiple Deprivation is that **deprivation is multidimensional** and can be experienced in relation to a number of distinct domains."*<sup>9</sup>

The above goes to the heart of the question we are exploring in this paper. Namely: **do the different ways in which individuals experience deprivation also translate into multiple ways in which the dimensions of deprivation could impact debt-related retail costs?**

<sup>6</sup> In England, the Technical report lists the criteria as: claimants of Jobseeker's Allowance, aged 18-59/64; Claimants of Employment and Support Allowance, aged 18-59/64; Claimants of Incapacity Benefit, aged 18-59/64; Claimants of Severe Disablement Allowance, aged 18-59/64; and Claimants of Carer's Allowance, aged 18-59/64. In Wales, the relevant Technical Report lists the same criteria, but does not explicitly list the Carer's Allowance.

<sup>7</sup> In England, the technical reports criteria as: Adults and children in Income Support families; Adults and children in income-based Jobseeker's Allowance families; Adults and children in income-based Employment and Support Allowance; Adults and children in Pension Credit (Guarantee); Adults and children in Working Tax Credit and Child Tax Credit families not already counted (that is those who are not in receipt of Income Support, income-based Jobseeker's Allowance, income-based Employment and Support Allowance or Pension Credit (Guarantee) and whose equivalised income (excluding housing benefit) is below 60 per cent of the median before housing costs; Asylum seekers in England in receipt of subsistence support, accommodation support, or both.

<sup>8</sup> *'Capturing deprivation and arrears risk in household retail cost assessment: Working paper for United Utilities on Wednesday 10 May 2017.'* Reckon LLP (2017); page 10.

<sup>9</sup> *'The English Indices of Deprivation 2015: Technical Report.'* – Department for Communities and Local Government (September 2015). Page 16.

Put simply, if one believes the answer to the above question to be ‘yes’, then there is a risk that ‘narrower’ measures of deprivation, which can be compared across England and Wales, will not properly reflect how deprivation affects efficient debt-related costs for companies. In this case, the absence of an overall IMD score for Wales might be of particular concern (especially when setting debt-related costs for Welsh).

### 3. How the various dimensions of deprivation can affect debt-related retail costs

Following from the previous section, given that the very basis for the IMD is that deprivation is multifaceted, it is essential to consider whether and how these various dimensions might be expected to impact debt-related costs (in a way that is outside of efficient management control). We subsequently explore this issue statistically – but also consider it important to consider this question from ‘first principles’.

In the following, we briefly set out our assessment of this. Here, it is important to be mindful of the fact that the various elements of deprivation are likely to be highly correlated (e.g. income and employment). So, what we are most interested in, is identifying *distinct ways* in which we might expect each aspect of deprivation to impact debt-related costs.

- **Income deprivation.** It is clear why the number of customers experiencing ‘very low’ income levels will directly impact debt-related costs. Specifically, customers facing very low incomes are clearly more likely to fall into arrears or default (on account of them facing severe budget constraints) which, all else equal, will increase doubtful debt costs for companies.
- **Employment deprivation.** Employment deprivation may be strongly associated with income deprivation (because, for example, unemployed customers may be more likely to have very low incomes than employed customers). However, it is also possible that this aspect of deprivation could impact debt-related costs in ways that are *separable* from income effects. For example, for an equivalent amount of income, the frequency, stability and predictability of income would clearly vary, depending on whether a customer was employed, or unemployed. As such, for an equivalent income level, one might nonetheless expect the probability of default or arrears (and therefore doubtful debt costs) to vary by employment status.
- **Education and skills deprivation.** Intuitively, customers with lower education and skills are more likely to drive higher debt-related costs for companies in a way that is distinct from pure ‘income’ effects. Specifically, such customers may have *poorer financial literacy* and, therefore, might be less capable of managing household budgets. Consequently, for two households of an equivalent ‘low income’, if one had lower education and skills, one would expect that household to have a higher probability of arrears or default. In addition, poor financial literacy would likely imply that such customers would find it harder to evaluate the benefits of ‘payment plans’ or ‘social tariffs’ when offered them. This too, would likely increase arrears or default probability, *independent of income*.
- **Health and disability deprivation.** One might expect there to be a relatively strong correlation between ‘health and disability’ within local areas and their respective ‘income’ and ‘employment’ related deprivation (ergo, implying that there is likely to be an *association* between ‘health and disability’ and debt-related costs). However, we might also expect ‘health and disability’ outcomes to impact

WHILST THERE MIGHT BE HIGH CORRELATIONS BETWEEN THE VARIOUS ASPECTS OF DEPRIVATION, A CAREFUL FIRST PRINCIPLES ASSESSMENT ALSO IDENTIFIES VARIOUS DISTINCT WAYS IN WHICH THE DIMENSIONS OF DEPRIVATION COULD AFFECT DEBT-RELATED COSTS.

debt-related costs *in ways that are distinct from income and employment*. For example, customers with poor health or a disability may be more difficult to contact (i.e. because their access to different forms of communication may be more limited). As such, companies might have to incur more effort, and therefore more cost, in relation to debt management activities for these customer groups (again, *independent of income*).

- **Barriers to housing and services deprivation.** Clearly, lack of access to key services and housing is likely to be strongly associated with other deprivation measures. However, there are intuitively sound reasons to suppose this aspect of deprivation also has distinct impacts on debt-related costs. Specifically, ‘access to housing’ might be relatively *fluid* for some customers (e.g. customers might be on a waiting list for social housing, meaning that lack of access is temporary; or equally, might have access withdrawn). As such, a lack of access to housing might make debt management activities ‘more difficult’ for companies (i.e. because it is harder to trace such customers, if it is associated with a high propensity to change address). In addition, this aspect of deprivation includes ‘housing overcrowding’. As such, areas with higher ‘scores’ on this metric will be associated with: (i) higher average household sizes; therefore (ii) higher average water usage; and thus (iii) higher average water bill sizes (again, controlling for income). It is widely accepted that average bill size drives doubtful debt costs, because – for an equivalent default probability - the absolute £s cost of arrears or default is higher where bill sizes are higher. Again, this effect is independent of income deprivation.
- **Crime deprivation.** This aspect of deprivation includes various metrics that might affect debt-related costs, in ways not captured by other dimensions. For example, households that experience ‘burglary’ might face an immediate budgetary shortfall that could reduce their ability to pay their water bill in the short-term (which would not be captured by the income domain).
- **Living environment deprivation.** Of the various domains of deprivation, there are perhaps less obvious ways in which living environment might impact debt-related costs in ways that are independent of the other aspects. However, even here we note that the domain includes ‘houses without central heating’. This, in turn, might be associated with customers’ water usage patterns, and thus bill sizes and expected doubtful debt costs.

Following from the above, the key point to note is that, whilst statistically there might be high correlations between certain dimensions of deprivation, there are identifiable reasons to suppose that these individual elements might also impact debt-related costs *in ways that are distinct*. Of course, it might be the that the magnitude of these distinct impacts varies (both in totality, but also by company). However, the key implication is that, in order to ensure that allowed company costs reflect the efficient level, one needs to: (i) carefully examine the impact of using ‘broad’ versus ‘narrow’ measures of deprivation; and relatedly; (ii) test how this might impact an assessment of ‘efficient’ debt-related costs, either for the industry, or specific companies.

In fact, as we subsequently set out in Section 5, when we follow a strict ‘general-to-specific’ econometric approach to explaining debt-related costs by local authority across England, we find that measures across *all of the above* dimensions remain statistically significant and intuitively signed.

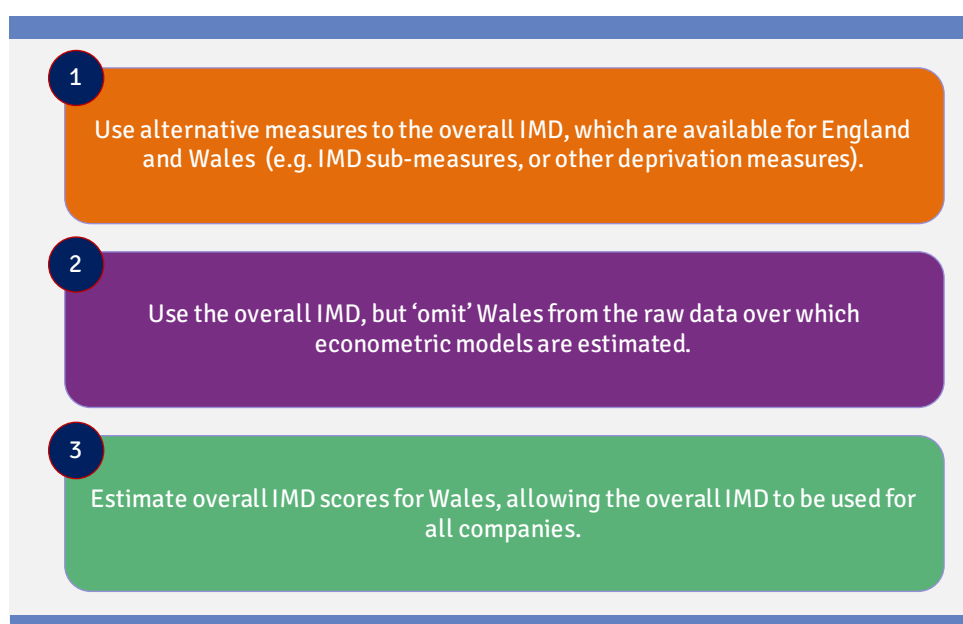


#### 4. Implications and considerations for approaches to cost assessment

Drawing the previous two sections together, the absence of an overall IMD score for Wales does not, in of itself, mean that credible econometric cost benchmarking models cannot be developed to capture the impact of deprivation on debt-related costs. For example, Ofwat could:

- use one of many *alternative measures of deprivation* to the IMD within econometric models, which are available for both England and Wales; and / or
- estimate econometric models using the IMD, but ‘*omitting*’ Wales in the data from which its models are estimated; and / or
- create estimated values for the overall IMD for Wales, allowing it to use this measure on a like-for-like basis with companies in England.

Figure 2: Broad options for capturing deprivation in debt-related cost models



Source: *Economic Insight*

From an ‘in principle’ perspective, none of the above approaches are necessarily ‘more right’ or ‘more wrong’ than the other. However, there are circumstances whereby they might inadvertently lead to costs being set below, or above, the efficient level (either for Welsh specifically, or the industry more broadly). These are summarised in the following table.

Table 1: How various approaches ‘could’ result in efficient costs being under or over-stated

Modelling options	How this <u>could</u> understate how deprivation affects (efficient) costs
<p>1. Use alternative deprivation measures that are available for both England and Wales (for example, sub-measures of the IMD, such as IMD income; benefit claimant rates etc).</p>	<p><b>If</b> the alternative measures understate how deprivation affects debt-related costs, relative to the overall IMD (either for Welsh specifically, or the industry), allowed debt-related costs might be set ‘too low’ <u>for Welsh</u> or for <u>all companies</u>. For example, and as described in the previous section of this report, one potential concern is that that the overall IMD is a ‘broad’ measure of deprivation, whereas alternatives could be ‘narrower’.</p>
<p>2. Use overall IMD scores (for England), but therefore omit Welsh from the raw data used to estimate the econometric benchmarking models.</p>	<p><b>If</b> Welsh has ‘higher’ deprivation than other companies and /or if the impact of this on costs is ‘higher’ than for other companies, omitting Welsh from the raw data could result in the coefficient on deprivation being ‘too low’. The result is that the models might understate efficient debt-related costs <u>for all companies</u>.</p>
<p>3. Estimate an overall IMD score for Welsh.</p>	<p><b>If</b> the precise way in which the overall IMD score for Welsh did not fully capture the ways in which the multiple aspects of deprivation affect debt-related costs (for Welsh) then this could result in allowed debt-related costs being set ‘too low’ <u>for Welsh</u>.</p>

Source: *Economic Insight*

As emphasised above, there is no inherent reason to suppose that any one of the above approaches is more likely than another to better predict companies’ ‘efficient’ debt-related costs. Rather, it depends very much on the specifics of how those approaches are implemented; and the issues identified in the table above (i.e. the “ifs”).

These, then, are the issues we seek to explore in our paper. Accordingly, in the following sections, we now set out our analyses, addressing the key questions in turn:

- » Can robust estimates of the overall IMD for Welsh be identified?
- » How do our estimated IMD scores for Welsh compare to other available evidence?
- » Is our estimated overall IMD score for Welsh robust when used in retail cost assessment models?

## 5. Can robust estimates of an overall IMD score for Welsh be derived?

In order to establish whether the overall IMD objectively better captures how deprivation drives debt-related costs, the first task, of course, is to establish *whether* robust measures of the overall IMD for Welsh can be estimated.

To address this question, our approach was to:

- identify potential ‘explanatory’ variables for the overall IMD score; and
- run regressions *across local authorities in England*, where ‘overall IMD’ is our dependent variable – and our various deprivation related measures are our explanatories; then
- use the results of those regressions to *generate various ‘estimated’ overall IMD scores for Welsh* (where we generate estimated overall IMD scores both by local authority area within Wales; and for Welsh as a whole).

In the following subsections, we provide further details of each of these steps in turn.

### 5.1 Identifying explanatory variables for the overall IMD score

As our goal here is to ‘predict’ overall IMD scores for Welsh, we wanted to ensure that any explanatory variables we used aligned, as closely as possible, to how the IMD itself is constructed (which we described previously in Section 2). As such, we started by identifying a wide range of deprivation measures that:

- align to the domains of the IMD (thus we included specific metrics *actually used within the IMD*, but also metrics not used in the IMD, but mapped to IMD domains); and
- are available for both England and Wales on a consistent basis (i.e. because otherwise the model results could not be used to estimate overall IMD scores for Welsh).

By adopting this approach, we were able to identify a wide range of potential measures, which are summarised in the table overleaf.

Table 2: Summary of IMD related measures available for both England and Wales

IMD domain	Measures available for England and Wales by domain	Of which are used in the overall IMD
Income	IMD income.	IMD income.
Employment	Jobseeker's Allowance (JSA); Employment and Support Allowance (ESA); Incapacity Benefit and Severe Disablement Allowance; Carer's Allowance (all % of local population).	Jobseeker's Allowance (JSA); Employment and Support Allowance (ESA); Incapacity Benefit and Severe Disablement Allowance; Carer's Allowance.
Education, skills & training	KS2 SATs % achieving above level 4 (average, English & Mathematics); GCSE average capped points; secondary school absences (% sessions); school leavers/16-year olds not in education or training (NEET); adults with poor English (% local population); adults with no qualifications (% local population).	KS2 attainment (average point score); GCSE average capped points; absences; school leavers not in education/training; adults with poor English; adults with no qualifications.
Health & disability	Deaths before age 75 (% local population); people with limited day-to-day activities (% local population); male life expectancy at birth (years); suicides (% local population).	Years of potential life lost; comparative illness and disability ratio; mood and anxiety disorders. <sup>10</sup>
Crime	Police Recorded Crime: Violence; theft; criminal damage; burglary (ratio to local population).	Violence; theft; criminal damage; burglary (relative to at-risk population).
Barriers to housing & services	Population density (people per hectare); occupancy rating (number of rooms); eligibility for housing assistance (% local population); house price to income ratio.	Occupancy rating; eligibility for housing assistance; housing affordability (more detailed measure including renting).
Living environment	Homes without central heating (% households); road traffic accidents (relative to local population).	Homes without central heating; road traffic accidents (involving injury to pedestrians and cyclists).

Source: Economic Insight

<sup>10</sup> Available data are proxies for the health and disability measures included in IMD.

## 5.2 Developing regressions of overall IMD scores across local authorities in England

Following from the above, our next step was to develop econometric analyses, whereby our general approach was to regress the overall IMD *across local authorities in England* against the various IMD related measures we had identified. We modelled at a local authority, rather than LSOA, level because this is the lowest level on which most explanatory variables were available on a consistent basis. However, this still gave a large underlying dataset size (N =324).

Here, we adopted a pragmatic approach, whereby we wanted to balance pure statistical criteria against including a wide set of measures that reflect the 'breadth' of the IMD itself. As such, we developed four models, as described below:

- **Model 1:** This model began with *all of the variables set out on the previous page*. We removed any variables that were incorrectly signed, but otherwise included the full set of explanatory variables, regardless of their level of statistical significance.
- **Model 2:** This began with the same set of variables as model 1, *but used a general to specific approach*, to generate a more parsimonious version – eliminating variables until all were statistically significant at 5%.
- **Model 3:** This model began with *only those variables that matched very precisely against a measure that was used in the calculation of the IMD*. We again removed any variables that were incorrectly signed.
- **Model 4:** This began with the *same set of variables as model 3, but used a general to specific approach*, again eliminating variables that were not statistically significant at 5%.

Consistent with the above, the table overleaf summarises the variables included in our four models.

Table 3: Variables included in our models

IMD domain	Model 1	Model 2	Model 3	Model 4
Income variables	IMD income	IMD income	IMD income	IMD income
Employment variables	JSA; ESA; Carer's Allowance; Incapacity benefit.	JSA; ESA.	JSA; ESA; Incapacity benefit.	JSA; ESA.
Education, skills & training variables	KS2 SATs; GCSE; Absences; Adults with poor English; Adults with no qualifications.	KS2 SATs; GCSE; Adults with poor English; Adults with no qualifications.	KS2 SATs; GCSE; Absences; NEET; Adults with poor English; Adults with no qualifications.	KS2 SATs; GCSE; Adults with poor English; Adults with no qualifications.
Health and disability variables	People with activities limited; suicides; male life expectancy.	People with activities limited; male life expectancy.	Premature deaths; people with activities limited; suicides.	Premature deaths.
Crime variables	Violence; theft; criminal damage; burglary.	Violence; burglary	Violence; theft; criminal damage; burglary.	Violence; theft; burglary.
Barriers to housing and services variables	Population density; occupancy rating; housing assistance; house price-income ratio.	Housing assistance; house price-income ratio.	Occupancy rating; housing assistance; house price-income ratio.	Housing assistance; house price-income ratio.
Living environment variables	Households without central heating; traffic accidents.	Households without central heating; traffic accidents.	Households without central heating; traffic accidents.	Households without central heating; traffic accidents.

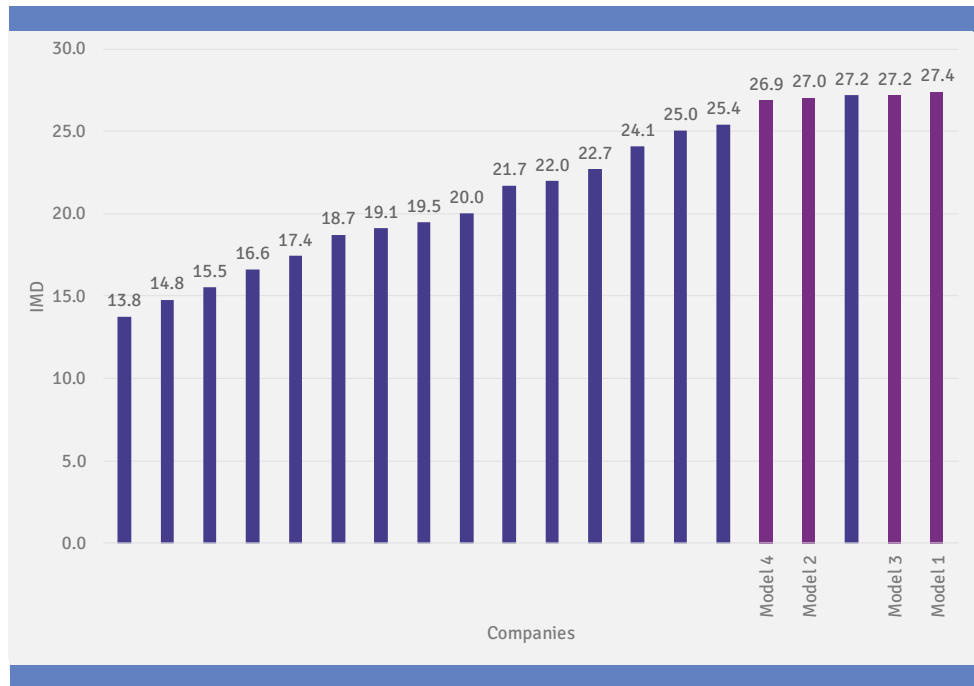
Source: *Economic Insight*

### 5.3 Predicted IMD scores for Welsh

Using the four models we developed, we were able to calculate ‘estimated’ overall IMD scores for Welsh (as a company overall, but also by local authority area within Wales). As can be seen from the following figure, our models:

- result in broadly similar predicted IMD scores for Welsh as a whole; and
- suggest that the company’s IMD score is one of the highest across the water industry.

Figure 3: Summary of predicted overall IMD scores for Welsh



Source: *Economic Insight*

In the following subsection, we set out the results of the four models used to generate the above scores, including diagnostics.

## 5.5 The statistical robustness and explanatory power of our models for estimating Welsh’s overall IMD score

The table below summarises the results of our four models used for predicting the overall IMD score across local authorities in England (as then used to estimate the overall IMD for Welsh). For consistency, we have report the same set of diagnostic tests as set out in Ofwat’s cost assessment consultation notification.

Table 4: Summary of model explanatory power and diagnostic results

Test	Brief explanation of test	Model 1	Model 2	Model 3	Model 4
R2	Proportion of variation in dependent variable explained by the model.	0.9868	0.9864	0.9861	0.9857
RESET Test	Test for specification errors using non-linear combinations of fitted values.	0.0344	0.0584	0.0544	0.1277
Heteroscedasticity	Test for non-constant variance in residuals.	0.003	0.001	0.0012	0.0003
VIF	Variance inflation factor - indicates presence of multicollinearity.	Mean: 10.92 Max: 77.38	Mean: 7.71 Max: 44.24	Mean: 8.93 Max: 56.62	Mean: 7.02 Max: 42.28

Source: *Economic Insight*

Following from the above, key points to note are that:

- All four models have high explanatory power, with R2 values of 0.99 (indicating that they explain almost all of the variation in the overall IMD across local authority areas in England).
- All of the models except for 1 ‘pass’ the specification test, although models 2 and 3 are somewhat ‘marginal’ in this respect.
- There is some evidence of heteroscedasticity across the models. In practice, however, using robust standard errors does not materially affect our results.
- Across the models, the income and employment variables had high VIFs. Of course, this reflects the fact that the IMD domains themselves are often highly correlated with each other. For the purpose of ‘predicting’ the overall IMD, therefore, we do not consider this to be an important consideration.

In summary, we conclude that, from a ‘statistical performance’ perspective, objectively Model 4 is the most robust. Note, as shown in Figure 3, this gives the lowest estimated IMD score for Welsh out of our models.



## 6. How do our estimated IMD scores for Welsh compare to other available evidence?

In order to help further inform the reasonableness and robustness of our estimated IMD scores for Welsh, we have sought to verify them against other available sources. Here, we have looked at:

- Other existing ‘estimated’ overall IMD scores for Welsh / Wales (where we have identified an independent academic study that used a statistical approach to derive values; and also, Reckon’s work for UU, which provides estimated overall IMD scores for the company).
- Broader official data on deprivation by region of the UK. This, whilst clearly not directly comparable to the IMD, is nonetheless informative at helping to understand whether our estimated IMD score for Welsh (in relative terms) is plausible.

We have also undertaken some *initial* modelling, examining the extent to which our estimated IMD scores for Welsh explain variation in debt-related retail costs, relative to other deprivation measures, by local authority area within Wales.

### 6.1 Existing estimates of IMD scores for Welsh

We are aware of two existing estimates of IMD scores for Wales / Welsh, one from an academic paper by Abel et al (2016); the other by Reckon (2017), in relation to its work for UU.

#### 6.1.1 Abel et al academic study

Estimates of the overall IMD for Wales on the same basis as the English IMD are available from a paper: *‘Adjusted indices of multiple deprivation to enable comparisons within and between constituent countries of the UK, including an illustration using mortality rates’* by Abel et al (2016).<sup>11</sup>

The objective of the paper is to provide measures of (IMD) deprivation at the LSOA level that are consistent across the whole of the UK. These estimates were derived in the context of exploring the impact of socioeconomic factors on health outcomes; and the paper therefore explores the relationship between IMD and mortality rates across LSOAs. The methodology followed by the paper is to develop LSOA-level regressions for the overall IMD score against the income and employment domains. We note that this is consistent with our view that these dimensions are comparable across England and Wales.

To transform, say, IMD scores for Wales into ‘English’ versions, the paper generates predicted values for Welsh LSOAs, using coefficients from the English regression. It then adds to this the standardised residual for the same LSOA from the ‘Welsh’ regression, scaled by the standard deviation of the residuals in the English regression.

---

<sup>11</sup> ‘*Adjusted indices of multiple deprivation to enable comparisons within and between constituent countries of the UK including an illustration using mortality rates.*’ Gary A Abel, Matthew E Barclay, Rupert A Payne. *BMJ Open* (2016).

### 6.1.3 Reckon’s work for UU

Reckon’s (2017) analysis for UU consisted of two main phases. In the first phase, Reckon explored (both qualitatively and through econometric modelling) relationships between various Equifax variables and: (i) measures of deprivation, including the Department for Communities and Local Government (DCLG)<sup>12</sup> deprivation measures (i.e. the IMD); and (ii) measures of UU’s bad debt costs at the LSOA level. In the second phase of work, Reckon made use of postcode level Equifax data for 29 shortlisted variables from 2006 to 2015. Using this data, Reckon firstly confirmed the results of the Phase 1 work. Reckon then undertook four further steps:

- » Calculated a weighted average Equifax value for each company’s supply area.
- » Developed econometric models to ‘predict’ the deprivation measures (the IMD) across England and Wales and for different years (this is the step of relevance to our work).
- » Developed econometric regressions to explore how variance in UU’s debt costs across LSOAs could be explained by the identified variables.
- » Finally, developed econometric models of company level debt-related costs, using the identified variables.

In relation to the second step (the one most relevant to this paper) variables used to construct the estimated IMD scores were: *Percentage of households with zero reported payment issues in the last six months; Equifax proprietary credit risk score; Average number of County Court Judgments per household; Percentage of population with no educational qualifications; Percentage of population that are inactive for employment purposes due to sickness; and Percentage of households in Council Tax Band A.*

### 6.1.4 Comparison of our results against existing studies

The table below compares the results from the above existing studies to our (preferred) estimated IMD score for Welsh (from our Model 4).

Table 5: Comparison of our estimated overall IMD score for Welsh relative to existing studies

	EI Model 4	Abel et al	Reckon
Welsh predicted IMD score (2015)	26.9	25.5	19.7

Source: *Economic Insight; Abel et al (2016); Reckon (2017)*<sup>13</sup>

As shown above, our estimated IMD score for Welsh is very similar (albeit slightly higher) to that reported by Abel et al, in their independent academic study. However, the Reckon estimate is materially lower than either the Abel et al, or our own, value.

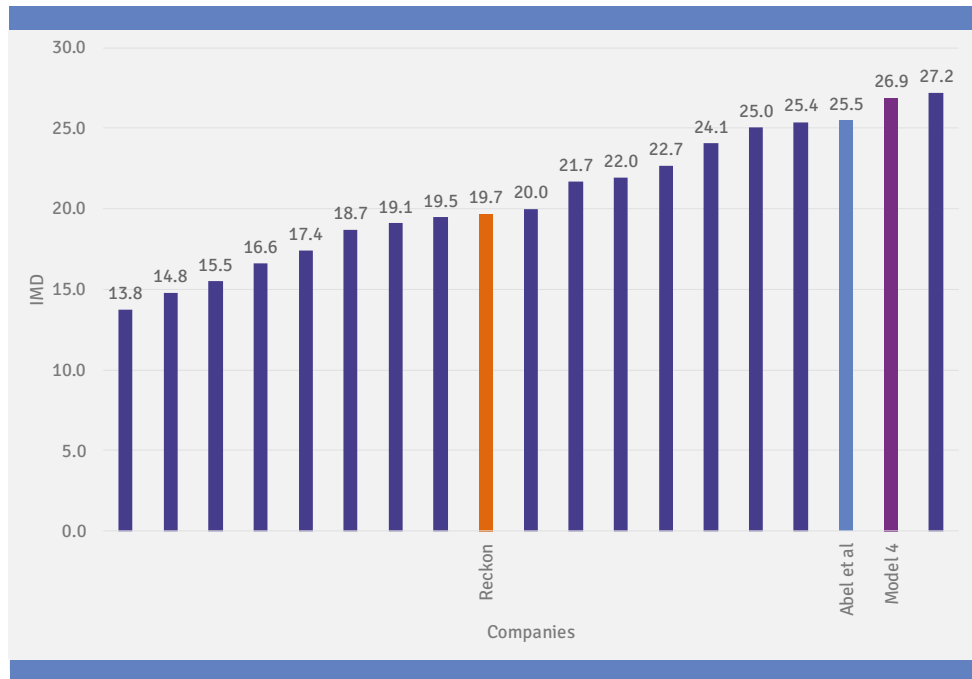
When one looks at the above in the context of overall industry rankings on IMD – as shown in the following figure – it is clear that the Reckon figure implies a materially lower relative deprivation position for Welsh than either our, or the Abel et al, estimate.

<sup>12</sup> Note, Reckon refers to this as the ONS.

<sup>13</sup> *‘Dataset for Deprivation and arrears risk in cost assessment.xls.’ Reckon (2017).*

OUR ESTIMATED IMD SCORE FOR WELSH IS VERY SIMILAR TO A PREVIOUS ACADEMIC STUDY. IT IS ALSO CONSISTENT WITH A WIDE RANGE OF INDEPENDENT EVIDENCE, SUGGESTING THAT WALES IS AMONGST THE MOST DEPRIVED REGIONS OF THE UK.

Figure 4: Overall IMD scores by company – estimated Welsh values highlighted: EI Model 4; Abel et al; Reckon



Source: Economic Insight

As can be seen, both our Model 4 result, and the Abel et Al study, imply that Welsh would be ranked 2<sup>nd</sup> on deprivation, compared to industry peers. In comparison, the Reckon result implies a much lower relative IMD position for Welsh, of 11<sup>th</sup> most deprived. We further note that, in its report on retail benchmarking for Ofwat, PwC also made use of the Abel et al study to infer an overall IMD score for Wales (albeit using a simplified methodology to the one we applied).<sup>14</sup> Consistent with our result, PwC find Welsh to be ranked 2<sup>nd</sup>, after UU, on the IMD.<sup>15</sup>

In light of the considerable existing evidence on relative deprivation in Wales (see next subsection, where official government data consistently finds Wales to be one of the, if not the, most deprived regions in the UK) the Reckon figure for Welsh is plainly implausible.

It should be emphasised that this is not a criticism of Reckon’s work more broadly, nor does it imply that their method is not sound, nor that IMD estimates for other companies are necessarily inappropriate. However, it does strongly suggest that, if the Reckon value were used within setting cost allowances for Welsh, the company’s allowed debt-related costs would likely be set substantially below the efficient level.

There are multiple possible explanations as to why the Reckon figure is materially lower than the Abel et al study, or our own estimate. However, we consider that the most likely explanation is that our approach is explicitly based on seeking to use explanatory variables that align as closely as possible to the original construction of

<sup>14</sup> Specifically, our approach has been to use the regression results reported by Abel et al to calculate the IMD scores by local area with Wales, and then to weight these by property numbers, in order to arrive at the overall score. In contrast, PwC actually start from an alternative source (IMD estimates produced by Carmarthenshire Council) and then ‘adjust these’ using elements of the Abel et al method, in order to derive an overall IMD score.

<sup>15</sup> See figure titled ‘Average Days Sales Outstanding vs Deprivation.’, on page 21 of ‘Retail Services Efficiency benchmarking.’ PWC (2017).

SET AGAINST THE NON-CONTENTIOUS OBSERVATION THAT WALES IS HIGHLY DEPRIVED, RECKON’S IMD ESTIMATE FOR WELSH IS IMPLAUSIBLY LOW. THIS IS NOT, HOWEVER, A CRITICISM OF RECKON’S WORK MORE WIDELY.

the IMD (which, as we explained, reflects the multifaceted nature of deprivation) - whereas the Reckon approach, for entirely legitimate reasons, did not.

## 6.2 Broader comparisons of Welsh’s relative deprivation position

To help further understand the validity of our estimate, we have also looked at broader evidence on Welsh’s relative position as regards deprivation, focusing on official Government published metrics. Before setting out the data, it is worth highlighting various independent findings of relevance – as follows:

*‘The level of poverty in Wales has been static, and Wales is currently behind only London.’*  
- Inquiry into Poverty in Wales.

- In 2015 the results of an extensive Government inquiry into poverty in Wales were published. It concluded: **“since the early 2000s, the level of poverty in Wales has been static, and Wales is currently behind only London... with 23% of the population living in poverty.”**<sup>16</sup>
- In May 2017, *The Economist* set out an analysis of Programme for International Student Assessment (PISA) data (which allows educational outcomes to be compared like-for-like, geographically). The article found: **“the [educational] results of Welsh 15-year-olds were similar to those of their peers in Latvia and the Czech Republic, and far below those in England, Scotland and Northern Ireland.”**<sup>17</sup>
- In relation to the ONS’ most recent release of regional economic performance data, *WalesOnline* commented: **“Figures released today by the Office for National Statistics reveal the yawning gap between Wales and other parts of the UK in economic terms. The only other UK nation or English region to see a similar figure was Northern Ireland - although Northern Ireland's figure is fractionally better than Wales.”**<sup>18</sup>

The table below sets out Wales’ relative rank on a number of measures relating to deprivation. To avoid creating a circularity with our Model 4, we have intentionally not reported any sources that overlap with our explanatory variables.

Table 6: Welsh’s relative position on broader deprivation related measures

Measure / source and Wales rank	Average earnings	Economic inactivity rate	Educational attainment at 15	Life expectancy
Sources	ONS: Annual Survey of Earnings and Hours (2017)	ONS: Regional labour market statistics in the UK (February 2018)	PISA (2015)	ONS: Life Expectancy at Birth & at age 65 by Local Areas in England & Wales (2017)
Metric	Mean weekly earnings	Inactivity rate 16-64	Attainment score	Life expectancy (years)
Wales Rank; (1 = worst)	1	=2	1	3
Comparators	Regions of England & Wales (out of 10)	Regions of England & Wales (out of 10)	Nations of the UK (out of 4)	Regions of England & Wales (out of 10)

Source: various, see table

<sup>16</sup> *‘Inquiry into Poverty in Wales: Poverty and Inequality.’ National Assembly for Wales Communities, Equality and Local Government Committee. (June 2015).*

<sup>17</sup> *‘Down in the valleys: The struggle to improve the worst education system in Britain.’ The Economist (May 2017).*

<sup>18</sup> *‘Shocking economic stats reveal just how the regions of Wales are doing.’ Wales Online February 8<sup>th</sup> 2018.*

Reviewed in the round, the key points to note are:

- A wide range of existing data and evidence consistently indicates that Wales is one of the most deprived regions of the UK.
- Moreover, this is a widely accepted, and non-contentious, point.
- As such, this also supports the overall validity of our estimated IMD score for Welsh.

### 6.3 Initial evidence regarding the performance of estimated IMD scores in relation to predicting variation in debt-related costs, relative to alternative deprivation measures

As a further means of testing the ‘validity’ of our estimated overall IMD scores for Welsh, we have undertaken an additional, *preliminary*, econometric modelling exercise. Here, we compared the explanatory power of our estimated IMD scores for Welsh (by local authority area within Wales) in relation to predicting the company’s debt costs at a local authority level, relative to alternative measures of deprivation.

This analysis is set out in Annex B. However, in summary, we find that our predicted IMD scores objectively perform better at explaining the variation in costs at a local authority level within Wales than the alternatives we tested (which included: IMD income; benefit claimant count; GVA per capita; economic activity count; and the employment rate). This finding is consistent with the overall IMD better capturing debt-related cost variation. However, further work is needed to: (i) verify this finding; and (ii) establish whether it holds ‘industry wide’, or is more ‘company specific’.

## 7. Is our estimated overall IMD score for Welsh robust when used in retail cost assessment models?

The final step in our work was to establish whether our estimated IMD score for Welsh was robust when included in retail cost assessment models. Here, we focus on its impact in relation to debt-related retail costs.<sup>19</sup> Our approach was as follows:

- We took two of our existing models of bad-debt-related retail operating costs using pooled data, one estimated using OLS, the other using GLS (which, in our original retail cost assessment model suite, are termed models A2 and A6 respectively). In their original specifications, the deprivation related variable used in these models is IMD income.
- We then re-ran both models, replacing IMD income with our estimated *overall IMD* (as per Model 4) for Welsh (and using the actual 2015 IMD for all other companies). For the purpose of this report, we term these alternate models: A2 (overall IMD); and A6 (overall IMD).

The results of this analysis are set out over the following two tables, consistent with Ofwat’s template for cost assessment model submissions. Full details of variable definitions and sources are contained in Annex A.

---

<sup>19</sup> This does not imply that we do not also consider it appropriate to include deprivation in models of total retail operating costs. Rather, the focus on debt-related costs in this paper is because it is the most logical aspect of cost to examine, when assessing the impact of ‘broad’ versus ‘narrow’ deprivation measures.

Table 7: HH Retail debt-related cost model results - OLS

Model name	A2		A2 (overall IMD)	
	(Original model, using IMD income)		(New model, using overall IMD, where value for Welsh is predicted from our Model 4)	
Description of model	Pooled OLS linear regression model, including separate dual service and single service customer number variables			
Description of dependant variable	Natural logarithm of bad debt related operating costs, including doubtful debt and debt management			
Variables	Coefficients	(P values)	Coefficients	(P values)
Ln(single service customers)	0.535***	0.000	0.646***	0.000
Ln(dual service customers)	0.121***	0.000	0.0831***	0.002
IMD income	0.189***	0.000		
IMD			0.140***	0.000
Ln(average wholesale bill)	1.744***	0.000	1.895***	0.002
Population transience (%)	0.0909***	0.001	0.117***	0.000
Constant	-14.37***	0.000	-16.27***	0.000
Model results and diagnostics	Result	Implication	Result	Implication
R2 adj.	0.9333	Model explains c. 93% of variation in dependent variable	0.9398	Model explains c. 94% of variation in dependent variable
RESET test	0.0004	None	0.0096	None
Variance Inflation Factor (VIF)	Mean: 3.55 Max: 6.78	Low concern with respect to multicollinearity	Mean: 3.98 Max: 7.67	Low concern with respect to multicollinearity
Method (e.g. OLS or RE)	Pooled OLS			
Sample size (N)	89	89	89	89
Companies	18	18	18	18
Years	5	5	5	5

Source: Economic Insight

Table 8: HH Retail deb-related cost model results - GLS

Model name	A6		A6 (overall IMD)	
	(Original model, using IMD income)		(New model, using overall IMD, where value for Welsh is predicted from our Model 4)	
Description of model	Random effects GLS linear regression model, including separate dual service and single service customer number variables			
Description of dependant variable	Natural logarithm of bad debt related operating costs, including doubtful debt and debt management			
Variables	Coefficients	(P values)	Coefficients	(P values)
Ln(single service customers)	0.532***	0.000	0.609***	0.000
Ln(dual service customers)	0.184***	0.003	0.168***	0.009
IMD income	0.136***	0.008		
IMD			0.0900***	0.008
Ln(average wholesale bill)	1.235***	0.002	1.282***	0.002
Constant	-10.25***	0.000	-10.90***	0.000
Model results and diagnostics	Result	Implication	Result	Implication
R2 overall.	0.9260	Model explains c. 93% of variation in dependent variable	0.9277	Model explains c. 93% of variation in dependent variable
RESET test	0.0014	None	0.0019	None
Variance Inflation Factor (VIF)	Mean: 3.81 Max: 6.78	Low concern with respect to multicollinearity	Mean: 4.27 Max: 7.62	Low concern with respect to multicollinearity
Method (e.g. OLS or RE)	Random effects GLS			
Sample size (N)	89	89	89	89
Companies	18	18	18	18
Years	5	5	5	5

Source: Economic Insight

## 7.1 Implications of model results

In relation to the above, key points to highlight are that:

- The amended debt-related models, incorporating our estimated overall IMD score for Welsh, have strong explanatory power, with high R2 values (the R2 values are also *fractionally* higher than those in the original models, using IMD income).
- The overall IMD variable is statistically significant at the 1% level<sup>20</sup> and is appropriately signed.
- With respect to diagnostic tests results, we attach little weight to RESET tests, in the context of models constructed for the purpose of efficiency benchmarking. This is because, in benchmarking models, one *intentionally* omits explanatory variables that might ‘explain’ more of the cost variation, *if they are deemed to be within management control* (because in efficiency benchmarking, we only want to control for factors outside of management control). Put another way, if one included additional explanatory variables, which explained further variation in cost, but were within management control: (i) models would be more likely to pass RESET tests; but (ii) by including the additional variables that are within control, the residuals provide a less reliable measure of relative efficiency. Turning the variance inflation factors (VIFs), all models have maximum values below 8, with mean values around 4 – indicating low concern with respect to multicollinearity.
- In totality, the above results are consistent with it being credible to use our estimated overall IMD score for Welsh within retail cost assessment at PR19.

## 8. Conclusions and recommendations

Our conclusions are as follows:

- It is widely accepted that socioeconomic deprivation impacts company debt-related costs in a manner that is (primarily) outside of efficient management control. Consistent with this, **Ofwat is proposing to include deprivation within the variables it explores when developing econometric cost assessment models for HH retail.**
- **In practice, there are a number of ways in which deprivation can be measured and included in cost benchmarking models.** However, it is difficult to determine, a priori, which of these measures might best capture the way in which deprivation impacts company debt-related retail costs.
- **An important issue is that one potential measure (the overall IMD – which is the ‘official’ measure of deprivation) is not available on a consistent basis for England and Wales.** Consequently, without more information, it is impossible to evaluate the appropriateness of using this measure, relative to alternatives.

---

<sup>20</sup> Note, our reported p-values are based on heteroscedasticity-robust standard errors for pooled OLS models and unadjusted standard errors for random effects GLS models – because this method already takes account of the correlation of errors within firms. This is consistent Ofwat’s approach at PR14.



- To help address the above, we establish that **it is possible to estimate overall IMD scores for Welsh (both for the company as a whole; and by local authority area).**
- **We further find that our predicted IMD estimates are robust and plausible,** insomuch that: (i) the preferred model we use to generate the overall IMD has good statistical properties; and (ii) the predicted values from our preferred model accord with credible existing evidence on deprivation – including a prior independent academic study, which also predicts IMD values for Wales.
- **Our estimated overall IMD value for Welsh is also robust when used in bad-debt related retail cost assessment models.** Specifically, models including the overall IMD (drawing on our estimated value for Welsh) have high R2.
- **Initial analysis further suggests that, objectively, our estimated overall IMD scores better explain variation in debt-related retail costs** within Wales than alternative deprivation measures. This is consistent with the intuition that the multidimensional aspects of deprivation do drive debt-related retail costs – and that, therefore, alternative deprivation measures might result in allowed costs being set above, or below, the efficient level. Additional work would be required in order to: (i) further verify this finding; and (ii) determine the extent to which this issue is ‘industry wide’, or is more ‘company specific’.

Drawing the above conclusions together, our recommendations are:

- » Ofwat should give careful consideration to the ‘breadth’ of deprivation as a concept and the ways in which its different dimensions might impact ‘efficient’ debt-related retail costs at PR19.
- » Ofwat should further consider the extent to which the ways in which these ‘broader’ aspects of deprivation impact costs might be *company specific*, or are *industry wide*.
- » Relatedly, within the context of developing a suite of models for retail cost assessment at PR19, Ofwat should include the overall IMD, using predicted values for Welsh (for example, using the data and approach outlined in our report) within the pool of deprivation variables it considers.
- » Reckon’s existing estimate of the IMD score for Welsh is implausibly low, being inconsistent with existing evidence. Therefore, it should not be used for setting Welsh’s allowed debt-related costs. This does not, however, imply any broader criticism of Reckon’s work for United Utilities (UU).

## 9. Annex A: description of variables and sources

The table below summarises the variables used in the econometric cost models described in the main body of our report. All variables that were not from the company data share (IMD, flats and population transience) have been mapped from local authority-level data to company supply areas. This was done on the basis of the population-weighted overlap between company water supply areas and local authority boundaries.

Table 9: Data description

Variable	Description	Source
Single service customers (000)	Number of water-only and wastewater-only customers	Company data share APR Table 2F Number of Customers Lines 1+2+4+5
Dual service customers (000)	Number of water and wastewater customers	Company data share APR Table 2F Number of Customers Lines 3+6
IMD income (%)	Income domain from English and Welsh Indices of Multiple Deprivation	Ministry of Housing, Communities and Local Government; Welsh Government Statistics
Average wholesale bill (£)	Wholesale charge revenue per customer. Figures for 2011/12 – 2014/15 estimated using average wholesale proportion from 2015/16 to 2016/17.	Company data share APR Table 2F (Wholesale Charges Revenue Line 7)/Number of customers Line 7)
Flats (%)	Percentage of dwellings that are flats, maisonettes or apartments	Office for National Statistics
Population transience (%)	Sum of population inflows and outflows between local authorities, divided by total population.	Office for National Statistics

Source: *Economic Insight*

## 10. Annex B: Initial evidence on the performance of our estimated IMD scores for Welsh in relation to predicting deb-related costs, relative to alternative deprivation measures

As outlined in the main body of this paper, there are various *'in principle'* reasons to suppose that the overall IMD might have advantages over other measures. This is primarily because the concept of deprivation is an inherently a multifaceted one; and, intuitively, one might expect each of its various dimensions to impact efficient debt-related costs in different ways. Resultantly, a concern might be that the use of *'narrower'* measures of deprivation would seem to have the potential to either over or understate efficient debt-related costs, because they *'omit'* some of the ways in which deprivation and costs interact.

To help further explore the above, we have undertaken a *preliminarily* additional econometric analysis. Specifically, we have regressed Welsh's debt costs (measured as outstanding debt per property) by local authority area within Wales against alternative deprivation measures – so that we could then compare their explanatory power and robustness, *relative to using our predicted IMD scores from Model 4 at the local authority level*. The alternative deprivation measures we tested included:

- IMD income;
- benefit claimant count;
- GVA per capita;
- economic activity count; and
- employment rate.

The general form our regression(s) was as follows:

$$\text{Debt costs LA}_i = \beta_0 + \beta_1 \text{Dep}_i + \beta_2 \text{single serve}_i + \beta_3 \text{flats}_i + \varepsilon_i$$

Where:

- debt costs LA = debt costs per property by local authority area;
- Dep = alternative deprivation measures (tested to compare against our IMD scores)
- Single serve = proportion of single service (water-only or waste-only customers)
- Flats = proportion of flats, maisonettes and apartments.

Accordingly, we had six regressions in total, one for each of the above alternative measures of deprivation (and one for our estimated overall IMD score by local authority area). The following table summarises our results.

Table 10: Results of regressions predicting variation of debt costs within Wales – comparison of alternative deprivation models to our predicted IMD score from Model 4

Result	IMD income	Benefit claimant count	GVA per capita	Economic activity rate	Employment rate	Overall IMD (as predicted from Model 4)
P-value	0.000	0.000	0.016	0.000	0.000	0.000
R <sup>2</sup>	0.4076	0.3819	0.3245	0.3465	0.4205	0.4258

Source: *Economic Insight*

Regarding the above, key points to note are as follows:

- There is some variation in the explanatory power of alternative measures of deprivation on debt costs (by local authority area within Wales), although all of the measures are statistically significant.
- We note that GVA per capita and the economic activity rate have materially lower R2 values than other approaches – indicating that they are likely to be *poorer* explainers of efficient debt costs (in isolation).
- The IMD income score and employment rates have near identical R2 values, at 0.41 and 0.42 (rounded). Importantly, these are similar to, but just below, the R2 obtained if the predicted overall IMD scores from our Model 4 are used to explain within area variation in debt costs.

In light of the above, in relation to explaining variation in debt-related costs, objectively we find that the overall IMD score derived from our modelling is to be ‘preferred’ to the alternative measures explored here. However, we note that the explanatory power of the overall IMD score is only ‘slightly better’ than using IMD income or the employment rate (and so this finding is ‘marginal’). We further note that:

- » As the above analysis relates to ‘within Wales’ variation in debt costs; strictly speaking, we can infer the overall IMD score better explains variation in debt costs *for Welsh*.
- » In turn, this suggests that, if alternative measures were used, they may not properly capture efficient costs *for Welsh*.
- » We cannot comment on whether this finding would hold across the industry, based the scope of this analysis.

**Economic Insight Limited**

125 Old Broad Street

London

EC2N 1AR

0207 100 3746

[www.economic-insight.com](http://www.economic-insight.com)

*Economic Insight Ltd is registered in England No. 7608279.*

*Whilst every effort has been made to ensure the accuracy of the material and analysis contained in this document, the Company accepts no liability for any action taken on the basis of its contents. Economic Insight is not licensed in the conduct of investment business as defined in the Financial Services and Markets Act 2000.*

*Any individual or firm considering a specific investment should consult their own broker or other investment adviser. The Company accepts no liability for any specific investment decision, which must be at the investor's own risk.*

*© Economic Insight, 2018. All rights reserved. Other than the quotation of short passages for the purposes of criticism or review, no part of this document may be used or reproduced without express permission.*

