Private & Confidential
Sonia Brown
Ofwat
21 Bloomsbury Street
London
WC1B 3XD

29 April 2014

Dear Sonia,

PwC review of South West Water's doubtful debt cost models

Under the terms of Service Order SER/0379 you asked PricewaterhouseCoopers LLP ("us") to carry out a review of the econometric analysis undertaken on behalf of South West Water Limited ("South West Water") by Oxford Economic Research Associates Ltd ("Oxera"). Oxera's analysis was used to support a doubtful debt cost adjustment submitted in South West Water's Retail Business Plan for Asset Management Period Six ("AMP6") which covers the period from 2015/16 to 2019/20.

The modelling submitted by South West Water was revised several times following an iterative process throughout which Ofwat provided comments and feedback on the approach. In total, four reports were issued by South West Water containing different approaches. This letter describes the comments provided by PwC during this process and the amendments to the modelling approach that were undertaken.

Initial doubtful debt cost submission

The initial econometric analysis submitted by South West Water was presented in an Oxera report "Average cost to serve: effect of income deprivation and bill size on doubtful debt" in November 2013 ("Oxera 2013"). This paper developed a series of company level panel data models using a random effects approach. The preferred model (model 2) linked doubtful debt costs per unique customer to the index for multiple deprivation, the average combined bill level and a time trend. An alternative version of this model was also used as a sensitivity test. This was identical except that it included time fixed effects, rather than a linear time trend; it also provided very similar results.

These two models each provided an estimate of the "efficient" doubtful debt cost given the bill and deprivation levels in South West Water's area. The average result of the two models was used as the basis of the adjustment claim contained in South West Water's Retail Business Plan, with a glide path applied over AMP6 between the actual costs in 2012/13 and adjusted doubtful debt cost.

We reviewed the methodology used in this modelling approach and pointed out several concerns to Ofwat. Our methodological review did not include an audit of the underlying data or workings but we did replicate the findings by running Oxera's Stata modelling files and comparing the results with those reported.

A principle concern we raised was that the requested adjustment was based on a modelled 2012/13 value rather than a forecast value for AMP6. We felt that there might be differences in

the two due to changes in the underlying drivers in the future. A second concern was the use of a time trend or time fixed effects in the models. These time variables were used as a proxy for other factors which may change over time but we considered that they lacked theoretical justification and therefore we could not be sure they were an appropriate proxy for such other omitted variables. Using these time variables also made the models less suitable for forecasting as additional assumptions would need to be employed in order to forecast the future. For example, if the time trend was applied over the forecast period then this would implicitly assume that whatever drove the relationship with time historically would continue in the future. As it was not clear what drove the relationship between doubtful debt costs and time in the historic period used to estimate the models, it was not clear that this implicit assumption is applicable in the future.

We set out more detail on these points below:

The theoretical merits of the methodological approach used

1. The models were not used to produce forecasts; rather they were used to estimate a current "efficient" or modelled level of doubtful debt costs in 2012/13 given South West Water's characteristics. The 2012/13 values were used as the basis for the claim in their business plan. This approach implicitly assumes that 2012/13 doubtful debt levels are representative of the whole of AMP6. It is not clear that this is a reasonable assumption. Doubtful debt increased sharply during the recession as more consumers struggled to pay their bills, but it may be expected to decline as the economy recovers. If this is the case then the use of 2012/13 as a base year may overstate the value of the funding requirement over the course of AMP6. We saw no evidence from South West Water supporting the implicit assumption that was used (for example, they did not examine the persistence of doubtful debt costs after the early 1990s recession).

2. One model included a linear time trend variable as a driver of doubtful debt per unique customer and the other model included time fixed effects (these have similar properties to the time trend but allow the estimated relationship with time to vary each year). We raised several concerns over using these time variables in the models, rather than real data:

   • We tested the linearity assumption in the first model by including different powers of the time trend in the model. These tests implied that the time trend is non-linear since the square of the time trend was statistically significant. This is indicative of omitted variable bias potentially being present in the model and the model using the wrong functional form.

   • When time trends or time fixed effects are used in panel data models there is usually a theoretical hypothesis supporting their inclusion. For example they may be used as a proxy for some driver that evolves uniformly over time but for which data is not readily available (e.g. in models of real costs, time trends are commonly used as a proxy for productivity growth). It was not clear in the documentation what the time variables were intended to act as a proxy for.

   • If these models were used to produce forecasts for AMP6, then an additional assumption would be needed for how to treat the time variables going forward. For

example, continuing the time trend in the future would implicitly assume that whatever drove the relationship with time historically will continue. Since there was no evidence of what has caused the observed historic relationship, there was no support for this assumption.

- We suggested an alternative modelling approach would be to replace the time based variables with the data for the other economic variables for which the time trends act as a proxy. This might better explain what drove the observed relationship over time. In addition, using actual data might allow the AMP6 period to be forecast more robustly, if forecasts for the explanatory variables are available.

3. The use of the glide path from the current cost to the estimated "efficient" cost implies that customers will continue to fund the modelled cost inefficiency during this adjustment period in AMP6.

The results of statistical tests on the models[1]

1. There is a standard test to identify whether a random effects model is more appropriate than an Ordinary Least Squares (OLS) model (the Breusch-Pagan test). This test confirmed that the random effects approach was preferable which is consistent with the method used.

2. Several tests look at whether the statistical properties of model residuals (i.e. the difference between the actual and modelled values) are consistent with that expected under a given modelling approach. The Oxera models passed both the Shapiro Wilk test for residual normality and a test for skewness and kurtosis in residuals. These indicate that outlier issues were not present.

3. A further test on the statistical properties of residuals (Heteroskedasdicity) was failed. But the modelling approaches used an appropriate control for this issue (robust standard errors) so this had no impact on the reliability of the results.

4. A model specification test (Ramsey RESET) was failed. However, the results of this test should not be relied upon they are not valid when heteroskedasdicity tests are failed. If taken at face value, the results would suggest the models may have the wrong functional form (i.e. whether they are linear or not). This can also be indicative of omitted variables. Whilst this test was inconclusive as a result of heteroskedasdicity, we considered that omitted variables were an issue, as indicated by the significance of the non-linear time trend (as discussed in the preceding section).

5. A different model specification test (Linktest) was passed, which contradicted the result of the Ramsey RESET test.

6. It is important that there should be no correlation between the explanatory variables and the error terms in a model (an issue known as endogeneity). The Davidson-McKinnon test for endogeneity was passed by the models.

---

[1] The time trend and time dummy models both gave identical test results so no distinction is made between them in this section.

3

7. We also tested whether time fixed effects added explanatory power to the model. The test showed that they did. Note that this test does not provide any evidence of whether time fixed effects are preferable to a linear time trend.

8. We did not perform a test for autocorrelation since the data set had only five years of data. These tests would normally be conducted where there are ten or more years of data.

Overall, the models performed relatively well under most of the standard statistical tests and we considered that they provided evidence of the link between deprivation, average bills and doubtful debt charges.

Updated analysis 16 March 2014 (Oxera 2014a)

On 10 March 2014 Ofwat announced that South West Water had pre-qualified for enhanced status subject to further actions set out by Ofwat[2]. One of the actions related to the bad debt adjustment:

"Action 7: Although you provided sufficient evidence that we should make an adjustment to the ACTS to reflect your bad debt costs, we considered that you provided insufficient evidence to support the calculation of the size of adjustment. We therefore ask you to take the following actions.

*   Replace the proposed bad debt adjustment with one that is supported by sufficient and convincing evidence calculated using a suitable approach, in particular an approach that reflects the impact of changes in deprivation and bill levels over 2015-20. We ask that you do this after agreeing to a process for providing us with an updated adjustment by 7 April 2014. We will then review the revised adjustment.

*   Provide sufficient and convincing evidence to support customers paying for the proposed glide path from current costs to the efficient bad debt charge for South West. Set out the reasoning to support your view that bad debt costs will not change over the period in your area."

In response to these comments South West Water submitted an updated version of their analysis in a new paper by Oxera ("Oxera 2014a")[3]. This included five new approaches that sought to address Ofwat's comments by forecasting future bad debt costs in AMP6. Each of these approaches used the same underlying random effects panel model that was developed in Oxera 2013. The principle difference was that they forecast a value for doubtful debt in AMP6, rather than modelling a doubtful debt cost in 2012/13 (as used in the original adjustment request).

These approaches continued to use a time trend and so did not address our concern raised with regard to Oxera 2013. The new approaches used GDP growth to forecast the Index of Multiple

---

[2] http://www.ofwat.gov.uk/pricereview/pr14/prs_web140310pr14prequal
[3] "Impact on doubtful debt as a result of changes in deprivation and bill levels over 2015–20: an update", 16 March 2014.

Deprivation ("IMD") and we questioned whether this was an appropriate approach. Finally, several of the new approaches generated coefficients based on modelling both historic and forecast data (the latter from water companies business plans). We were concerned that generating coefficients using forecast data is not a standard econometric approach (as opposed to generating coefficients based on modelling historic data alone), and this approach introduced additional modelling error through the error inherent in the forecasts and circularity in the estimates generated.

Updated analysis 23-26 March 2014 (Oxera 2014b, Oxera 2014c)

In response to these comments, South West Water provided an updated version of their analysis in "Impact on doubtful debt as a result of changes in deprivation and bill levels over 2015–20: an update", on 23 March ("Oxera 2014b"). This paper was subsequently withdrawn and replaced with a revised version on 26 March 2014 (Oxera 2014c"). We only describe the revised 26 March report here. This contained four new approaches.

Approach I: used income deprivation (a sub-component of the overall index for multiple deprivation) and historic data between 2007/08 and 2012/13 to estimate the relationship between doubtful debt and deprivation and customer bills. Income deprivation was projected forward using forecast benefits expenditure and the historical relationship between benefits expenditure and income deprivation, which suggested that a 1% increase in benefits expenditure could result in a 0.4% increase in deprivation levels (and vice-versa). A separate time trend was estimated for the historic and forecast period.

Approach II: as approach I, except that all model coefficients were estimated using historic and forecast data combined.

Approach III: As approach I, except that the index for multiple deprivation was used rather than the income sub-domain.

Approach IV: As approach II, except that the index for multiple deprivation was used rather than the income sub-domain.

Overall we considered that all of the approaches taken in Oxera 2014c were an improvement when compared to Oxera 2013 and Oxera 2014a, in particular:

- The Oxera 2014c approaches changed the basis for forecasting deprivation by using a modelled link between total UK benefit expenditure and income deprivation levels at a local authority level for all models. We considered this approach to be a more appropriate proxy than the approach used in Oxera 2014a (using GDP growth). However, we noted that the model used to link benefit expenditure to deprivation was likely to suffer from omitted variable bias and it assumes that the historic modelled relationship from a single year will be applicable to the future. We also suggested alternative proxies that could be used to forecast deprivation, such as caseload figures for income related benefits.

- The Oxera 2014c approaches did not simply apply the historic time trend to the forecast period but instead estimated a different time trend for the future based on forecast data

from company's business plans. This was an improvement on the former approaches but did not address our concern over the introduction of circularity and the reliance on the accuracy of the business plan forecasts[4].

We also still considered that the use of a time trend as a proxy for omitted variables was not appropriately justified and the approach could have been improved by testing actual data rather than this proxy.

The four approaches provided similar results and shared the same underlying methodical approaches (and therefore concerns). Therefore we did not rank any of these four approaches as being inferior or superior to the others.

Additional sensitivity tests undertaken

In parallel to the process described above, Ofwat asked PwC to consider different options for forecasting how bad debt costs may evolve in AMP 6. We developed two approaches which built upon existing analysis which are described below:

Approach A: We used the original Oxera dataset (Oxera 2013) and tested an alternative model which replaced the time trend and deprivation index with data on unemployment[5]. The model retained average bills as the other explanatory variable.

- By removing the time trend this model removes the issue of how this should be forecast[6].

- Unemployment can be forecast using publically available projections[7].

- Average bill forecasts were sourced from data published by Ofwat on 14 March 2014[8].

- The model used the same random effects panel approach as Oxera 2013.

The underlying modelling file for this approach is contained in Appendix 1 of this letter. Ofwat will publish the dataset used separately.

Approach B: combined two modelling approaches, with one model used for the base year and a separate time series model used to generate a forecast profile.

- The 2012/13 doubtful debt cost base is derived from the fitted values from the Oxera 2013 model which uses a time trend (model 2).

---

[4] The reports did cite precedent for this approach from the UK regulated energy sector.

[5] We added two variables to the dataset: historic and forecast unemployment data, and forecast bill data.

[6] In both approaches we have assumed that the random effects component of the model is zero in the forecast period. This is because this value cannot be accurately computed outside of the period covered by the dataset and we believe this value is likely to capture other omitted variables and inefficiency.

[7] http://budgetresponsibility.org.uk/economic-fiscal-outlook-march-2014/

[8] http://www.ofwat.gov.uk/content?id=270d1e9b-ab63-11e3-8760-31cfc651602f

- The forecast profile is developed using a model PwC produced when reviewing how Thames Water's bad debt costs had evolved over time[9]. This model uses a data set assembled by FTI consulting[10].

- The forecast model uses a quantile regression approach to explain debt write-offs based on changes in employment deprivation.

- We developed a historic proxy for employment deprivation at a company level using data on the employment related benefits used to derive the measure. We also developed a forecast for this proxy using benefit caseload projections published by the Department for Work and Pensions (DWP)[11].

The key inputs and results of this approach are contained in Appendix 2 of this letter.

It is important to highlight several limitations which apply to both of these approaches:

- Both of these approaches used datasets provided by others (Oxera and FTI Consulting). We have not audited the underlying data from these data sets;

- We believe that the models used in our approaches suffer from methodological issues which bias the coefficients and so the results, in particular omitted variable bias; and

- Due to limited data availability on forecasts for unemployment and benefit claimants we applied national trends when producing company level forecasts.

Approach B has some additional limitations:

- This approach combined the outputs of two different models to obtain a forecast. This is not consistent with standard econometric methods. Moreover, the econometric models had different dependent variables so the approach did not compare like with like[12];

- The proxy indicator for employment deprivation used was not a perfect replica of the actual data. For example, data were not available for one of the components of the measure (participants in the New Deal programme); and

- The approach did not factor in forecast changes in the numbers of unique customers by company[13].

---

[9] The details of this model are contained in a letter from PwC to Ofwat entitled "PwC review of econometric analysis of bad debt" on 01 November 2013.
[10] "The impact of the deterioration in economic circumstances on Thames Water's bad debt" FTI Consulting, October 2013.
[11] This proxy includes Jobseekers Allowance, Incapacity Benefits, Severe Disablement Allowance and Employment Support Allowance but excludes New Deal participants.
[12] The dependent variables are debt-write offs and doubtful debt accounting charges.
[13] This is unlikely to have a material impact as customer numbers tend to change slowly over time.

Given the range of limitations above the purpose of this work was to provide an illustration of bad debt related costs as an additional check for the existing approaches. We do not argue that these approaches are superior to those submitted by South West Water.

The table below summarises the doubtful debt cost predictions from these two approaches, alongside the results from Oxera 2014c for comparison. The values for South West Water only are shown.

Estimated doubtful debt costs for South West Water under different modelling approaches (£ million, 2010/11 constant prices)

Approach	Reference	2015/16	2016/17	2017/18	2018/19	2019/20
I	Approach I from Oxera 2014c	11.6	11.2	10.8	10.4	10.0
II	Approach II from Oxera 2014c	11.4	11.3	11.1	10.9	10.8
III	Approach III from Oxera 2014c	11.6	11.2	10.8	10.4	10.0
IV	Approach IV from Oxera 2014c	11.4	11.3	11.1	10.9	10.7
A	PwC approach using unemployment rate and average bill size as explanatory variables	9.9	9.8	9.7	9.8	9.8
B	PwC approach using 2012/13 fitted value from Oxera 2013 model 2, and forecast profile from employment deprivation model.	10.6	9.8	9.5	9.3	9.3
	Simple average of six approaches	**11.1**	**10.8**	**10.5**	**10.3**	**10.1**

Yours Sincerely

PricewaterhouseCoopers LLP

Appendix 1: Stata do-file for unemployment model (approach A)

```
* Average cost to serve: Effect of unemployment and bill size on doubtful debt
* (c) PricewaterhouseCoopers LLP
* April 2014

/*

This code executes the simple PwC regression model based on average bill size and
unemployment rate.

Description of variables used in this code:
comp_ids: Unique company id
year: Year
ave_bill_comb: Average water and sewerage bills
debt_uc: debt per unique customer
debt_tot: total debt
n_unique_customers: number of unique customers
unemploymentrate: Rate of unemployment in areas served by the company, weighted by
population
debt_uc_new: Forecast rate of debt per unique customer
*/

clear
cd "<Insert directory name here>" /* e.g.: "C:\Panel Regression Model\" */
set more off
insheet using "Unemployment data file 29 Apr 2014.csv" /* use if a comma-separated
spreadsheet file, i.e. '.dta' file  is used */

*Declaring the dataset as a panel dataset*
egen comp_ids=group(comp_id), label /* This step generates a unique id for each
company. This is used to declare the dataset as a panel. */
xtset comp_ids year

*Generating variables:
gen lave_bill_comb=ln(ave_bill_comb) /* Log of average water and sewerage bills */
gen debt_uc=debt_tot/n_unique_customers /* Total debt per unique customer */
gen ldebt_uc=ln(debt_uc) /*Log of total debt per unique customer */
gen lagldebt_uc=l.ldebt_uc /*First lag of the log of total debt per unique customer*/
by comp_ids: replace unemploymentrate =l.unemploymentrate if year>2018 /*First lag
of the unemployment rate*/
```

9

* PwC model: panel regression
xtreg ldebt_uc lave_bill_comb unemploymentrate if comp!="Industry average" &
year<=2013, robust

* Forecasting

* Storing the coefficients from the regression above:
scalar a= _b[lave_bill_comb]
scalar b= _b[unemploymentrate]
scalar c=_b[_cons]

* Generating forecasts based on coefficients above and projected unemployment and
average bills:
gen debt_uc_new=0
replace debt_uc_new=scalar(a)*lave_bill_comb+scalar(b)*unemploymentrate+scalar(c)

*** smearing factor applied to the fitted value

predict e, e
replace e=exp(e)
replace debt_uc_new=exp(debt_uc_new)
summ e if e(sample)
replace debt_uc_new=debt_uc_new*r(mean) if(r(mean)>=1)
replace debt_uc_new=ln(debt_uc_new)
gen fitted_debt_uc_new=exp(debt_uc_new)*n_unique_customers

outsheet year comp_ids debt_tot fitted_debt_uc_new using "Macro_output.csv", comma
replace

clear

Appendix 2: key inputs and results of deprivation model (approach B)

Modelled value of doubtful debt costs (£m), 2012/13, (2010/11 prices)

Company	Doubtful debt cost, 2012/13, (£m)
Affinity	5.1
Anglian	28.8
B&W Hants	0.5
Bristol	1.7
Dee Valley	0.3
Dwr Cymru	15.2
Northumbrian	18.2
Portsmouth	0.5
Severn Trent	39.9
South East - Mid Kent	3.3
South West	11.8
Southern	18.7
Southstaffs-Cambridge	2.5
Sutton & E Surrey	0.8
Thames	50.1
United Utilities	46.4
Wessex	11.8
Yorkshire	27.6

Source: "Model 2 as presented in "Average cost to serve: effect of income deprivation and bill size on doubtful debt", November 2013 ("Oxera 2013").

The figures are taken from the stata log file provided by Oxera "Oxera acts analysis_November 2013.log'" for regression results for the random effects model with multiple deprivation (prediction made using "li comp year debt_tot pred_val l if year==2013, noobs separator(20)" command).

11

Proxy indicator for the Employment sub-domain of the Index for Multiple deprivation (percentage of working population in receipt of certain benefits)

Company	2008/09	2009/10	2010/11	2011/12	2012/13
Affinity	7.1%	8.2%	7.9%	8.0%	7.8%
Anglian	8.3%	9.5%	9.0%	9.2%	9.1%
B&W Hants	7.6%	8.6%	8.2%	8.1%	8.0%
Bristol	8.2%	9.1%	8.8%	9.0%	8.8%
Dee Valley	11.5%	12.5%	12.0%	12.0%	11.5%
Dwr Cymru	13.4%	14.2%	13.6%	13.7%	13.2%
Northumbrian	11.4%	12.5%	12.0%	12.2%	12.1%
Portsmouth	7.2%	8.1%	7.7%	7.9%	7.8%
Severn Trent	10.1%	11.3%	10.7%	10.9%	10.6%
South East - Mid Kent	6.1%	7.0%	6.7%	6.9%	6.7%
South West	9.5%	10.2%	9.9%	10.2%	10.0%
Southern	8.8%	10.0%	9.6%	9.7%	9.5%
Southstaffs-Cambridge	10.4%	11.9%	11.1%	11.3%	11.0%
Sutton & E Surrey	6.1%	7.1%	6.9%	7.1%	7.0%
Thames	8.4%	9.4%	9.2%	9.4%	9.1%
United Utilities	12.5%	13.5%	13.0%	13.3%	12.9%
Wessex	6.8%	7.6%	7.2%	7.3%	7.2%
Yorkshire	10.6%	11.9%	11.5%	11.8%	11.8%

Source: PwC analysis.

The proxy for the employment sub-domain of the index for multiple deprivation is estimated by combining data on benefit caseloads from 2008/09 to 2012/13. These data are mapped to company totals using local authority level data.

Econometric coefficients - PwC quantile version of FTI model

Variable	Coefficient
Employment deprivation (two year lag)	1.55
Constant term	-3.51

Source: Page 67, PwC letter to Ofwat "PwC review of econometric analysis of bad debt", 1 November 2013.

The model links bad debt write-offs to employment deprivation over time.

Projected proxy deprivation profile, 2012/13 to 2019/20

Variable	2012/13	2013/14	2014/15	2015/16	2016/17	2017/18	2018/19	2019/20
Employment deprivation proxy (Index 2012/13 = 100)	100.0	91.3	86.1	83.9	82.8	82.6	83.2	83.2
Cumulative profile, % change relative to 2012/13	0.0%	-8.7%	-13.9%	-16.1%	-17.2%	-17.4%	-16.8%	-16.8%

Source: PwC analysis based on DWP forecasts for national benefit caseload (December 2013).

Model outputs – estimated doubtful debt costs (£m), (2010/11 prices)

Company	2015/16	2016/17	2017/18	2018/19	2019/20
Affinity	4.4	4.0	3.9	3.8	3.8
Anglian	25.8	23.8	22.9	22.5	22.4
B&W Hants	0.5	0.4	0.4	0.4	0.4
Bristol	1.5	1.4	1.3	1.3	1.3
Dee Valley	0.3	0.3	0.3	0.3	0.3
Dwr Cymru	13.1	12.2	11.8	11.6	11.5
Northumbrian	16.4	15.3	14.8	14.5	14.4
Portsmouth	0.5	0.4	0.4	0.4	0.4
Severn Trent	34.9	32.4	31.3	30.7	30.6
South East - Mid Kent	2.9	2.6	2.5	2.4	2.4
South West	10.6	9.8	9.5	9.3	9.3
Southern	16.5	15.2	14.7	14.4	14.3
Southstaffs-Cambridge	2.2	2.0	1.9	1.9	1.9
Sutton & E Surrey	0.7	0.7	0.6	0.6	0.6
Thames	43.9	40.5	39.0	38.2	38.1
United Utilities	41.1	38.2	37.0	36.4	36.2
Wessex	10.2	9.3	9.0	8.8	8.7
Yorkshire	25.2	23.4	22.7	22.3	22.2

Source: PwC analysis

These results combine the estimated debt write-offs for 2012/13 with the projections generated from the employment deprivation proxy and econometric coefficients.