Affinity Water

PR24 Proposed Econometric Models

11th January 2023

Contents

Summary ... **3**

Template for submission of econometric models - Treated Water Distribution ... **3**

Template for submission of econometric models - Wholesale Water (1) ... **8**

Template for submission of econometric models - Wholesale Water (2) ... **12**

Appendix 1- Regional Wages ... **18**

Explanatory variables for regional wages ... 18

Discussion ... 19

Appendix 2 - Soil Type ... **22**

Appendix 3 - Alignment with Ofwat's Principles ... **25**

References ... **26**

Summary

This paper intends to inform development of econometric models for PR24 cost assessment. It extends the models and approach used at the last price review to include new explanatory variables - regional wages and soil type.

We present our work using the template for submission of econometric models and include separately our data files, Stata output files and our Stata do. instruction files alongside Excel files showing the construction of our explanatory variables.

Our results show that we have identified statistically significant and economically meaningful effects for regional wages and soil type both in the treated water distribution activity and at aggregate wholesale water botex level. We consider that our models enhance the predictive power of the PR19 specifications. Our models perform satisfactorily against sensitivity and stability tests and therefore, should go forward to the next stage of stakeholder consultation.

The appendices to the paper provide discussion and full details of our approach to construction of regional wages and soil type data. We accompany this paper with the spreadsheets that set out our approach and calculations for the variables.

Template for submission of econometric models - Treated Water Distribution

Econometric model formula:

AFWTWD1

ln(TWD botex plus NR_{it}) = a + β_1 ln(length of $main_{it}$) + β_2 (ln(booster pumping stations per length of $mains_{it}$) + β_3 ln(weighted average density LAD_{it}) + β_3 (ln(weighted average density LAD_{it}))^2 + β_4 ln(regional wages weekly LAD $basis_{it}$) + ε_{it}

AFWTWD2

ln(TWD botex plus NR_{it}) = a + β_1 ln(length of $main_{it}$) + β_2 (ln(booster pumping stations per length of $mains_{it}$) + β_3 ln(weighted average density LAD_{it}) + β_3 (ln(weighted average density LAD_{it}))^2 + β_4 ln(regional wages hourly LAD $basis_{it}$) + ε_{it}

AFWTWD3

ln(TWD botex plus NR_{it}) = a + β_1 ln(length of $main_{it}$) + β_2 (ln(booster pumping stations per length of $mains_{it}$) + β_3 ln(weighted average density LAD_{it}) + β_3 (ln(weighted average density LAD_{it}))^2 + β_4 ln(soil $type_{it}$) + ε_{it}

Description of the dependent variables

Treated water distribution botex including network reinforcement (code: Botex+NR_TWD) as reported in the published PR24 wholesale dataset.

Description of the explanatory variables

Length of Main km (code: lengthsofmain) as reported in the published wholesale dataset.

Number of boosters per km of main (code: boosterperlength) as reported in the published wholesale dataset.

Weighted average density (code: density) as reported in the published wholesale dataset.

Weighted average density squared, calculated value, the square of the weighted average density variable

Regional weekly wage (LAD basis), AFW generated variable

Regional hourly wage (LAD basis), AFW generated variable

Soil type (index of susceptibility to shrink/swell) AFW generated variable

Brief comment on the models

The number of years in the sample was consistent with the standard OFWAT dataset. This was from 2011/12 to 2021/22. There was no difference between the sample panel structure and the published do-file. The variable was assigned a unique code on the do-file and was added to the "Independent variables (cost drivers)" section, as well as section 5 "Prepare the macros for different regression specifications".

We found all variables to show a statistically significant result for Treated Water Distribution when tested through the OLS and Random Effects models. Ln(regional wages weekly LAD basis) was the strongest variable, which was proven to a 1% significance level. This was the same for the other measure of regional wages, Ln(regional wages hourly LAD basis). Ln(soil type) was a statistically significant result, to 10% significance. All models display an R-squared of 0.963 or more and this is higher than the adjusted R-squared 0.957 in Ofwat's PR19 specification. Form this we conclude that addition of our proposed variables has increased the predictive power of the models.

We studied the correlation between density, regional wages and soil type to determine whether the density variable already captures the effects of the variables we propose to add. For this we calculated the VIF statistic for each explanatory variable, based on OLS modelling. Where VIF values are close to 1 (and not greater than 5), we can conclude there is no significant multi-collinearity and any correlation between variables is unlikely to be problematic for the models. Our results are tabulated below. We find VIF values in the range 1 – 2 for both weekly and hourly regional wages variables and for the soil type variable. We conclude that to the extent that the variables we propose are correlated with other variables (including density), that this is not problematic for our proposed models.

OLS VIF Score for each explanatory variable	**AFWTWD1**	**AFWTWD2**	**AFWTWD3**
ln(length of main$_{it}$)	1.1	1.1	1.1
ln(booster pumping stations per length of mains$_{it}$)	2.2	2.2	2.2
ln(weighted average density LAD$_{it}$)	257.8	249.3	214.5
ln(weighted average density LAD$_{it}$)^2	261.3	252.8	210.8
ln(regional wages weekly LAD basis$_{it}$)	2.1		
ln(regional wages hourly LAD basis$_{it}$)		2.0	
ln(soil type$_{it}$)			1.2

	AFWTWD1	**AFWTWD2**	**AFWTWD3**
Dependent variable	Botex including network reinforcement	Botex including network reinforcement	Botex including network reinforcement
Constant	-12.181** {0.023}	-2.105 {0.322}	3.987*** {0.005}
Ln(length of mains)	1.092*** {0.000}	1.096*** {0.000}	1.085*** {0.000}
Ln(Wtd. Avg. Density)	-2.054*** {0.000}	-2.053*** {0.000}	-2.680*** {0.000}
Ln(Wtd. Avg. Density) squared	0.164*** {0.000}	0.162*** {0.000}	0.213*** {0.000}
Ln(booster pumping stations per length of main)	0.496*** {0.000}	0.497*** {0.000}	0.439*** {0.000}
Ln(regional wages weekly LAD basis)	7.772*** {0.003}	-	
Ln(regional wages hourly LAD basis)	-	4.156*** {0.000}	
Ln(regional wage annual SOC basis)	-	-	
Ln(soil type)	-	-	0.142* {0.051}
Estimation method (OLS or RE)	RE	RE	RE

N (sample size)	187	187	187
R2 adjusted	0.968	0.967	0.963
RESET test	0.030	0.034	0.073
VIF (max)	258.621	249.539	224.979
Pooling / Chow test	0.750	0.925	0.829
Normality of model residuals	0.002	0.023	0.859
Heteroskedasticity of model residuals	0.008	0.008	0.057
Test of pooled OLS versus Random Effects (LM test)	0.000	0.000	0.000
Efficiency score distribution (min and max)	Min: 0.850 Max 1.334	Min: 0.850 Max 1.299	Min: 0.807 Max 1.301
Sensitivity of estimated coefficients to removal of most and least efficient company	GREEN	AMBER	AMBER
Sensitivity of estimated coefficients to removal of first and last year of the sample	GREEN	GREEN	GREEN

Significance level: *** (1%), ** (5%) and * (10%)

Efficiency scores distribution

AFWTWD1

SWB	0.850
PRT	0.878
NWT	0.928
HDD	0.941
SRN	0.951
AFW	0.966
SEW	0.969
WSX	0.995
SES	1.009
SVE	1.043
NES	1.051
WSH	1.121
TMS	1.134
SSC	1.233
YKY	1.240
ANH	1.250
BRL	1.334

Median	1.009
Upper Quartile	0.951
Range	0.484

AFWTWD2

SWB	0.830
PRT	0.859
WSX	0.922
SEW	0.926
SRN	0.936
NWT	0.937
AFW	0.947
HDD	0.967
SES	0.976
NES	1.046
SVE	1.056
WSH	1.117
TMS	1.128
YKY	1.241
ANH	1.258
SSC	1.276
BRL	1.299

Median	0.976
Upper Quartile	0.936
Range	0.470

AFWTWD3

SWB	0.807
PRT	0.826
SRN	0.905
WSX	0.965
NWT	0.976
HDD	1.006
AFW	1.069
SVE	1.080
TMS	1.101
SES	1.108
NES	1.134
SEW	1.141

WSH	1.183
YKY	1.184
ANH	1.261
SSC	1.281
BRL	1.301

Median	1.101
Upper Quartile	0.976
Range	0.494

Template for submission of econometric models - Wholesale Water (1)

Econometric model formula:

AFWWW4

$\ln(\text{WW botex plus NR}_{it}) = a + \beta_1 \ln(\text{number of connections}_{it}) + \beta_2$ (percent of water treated in bands 3-6$_{it}$)+ β_3 (ln(booster pumping stations per length of mains$_{it}$) + β_4 ln(weighted average density LAD$_{it}$) + β_5 (ln(weighted average density LAD$_{it}$))^2 + β_6 ln(regional wages weekly LAD basis$_{it}$) + ε_{it}

AFWWW5

$\ln(\text{WW botex plus NR}_{it}) = a + \beta_1 \ln(\text{number of connections}_{it}) + \beta_2$ (percent of water treated in bands 3-6$_{it}$)+ β_3 (ln(booster pumping stations per length of mains$_{it}$) + β_4 ln(weighted average density LAD$_{it}$) + β_5 (ln(weighted average density LAD$_{it}$))^2 + β_6 ln(regional wages hourly LAD basis$_{it}$) + ε_{it}

AFWWW6

$\ln(\text{WW botex plus NR}_{it}) = a + \beta_1 \ln(\text{number of connections}_{it}) + \beta_2$ (percent of water treated in bands 3-6$_{it}$)+ β_3 (ln(booster pumping stations per length of mains$_{it}$) + β_4 ln(weighted average density LAD$_{it}$) + β_5 (ln(weighted average density LAD$_{it}$))^2 + β_6 ln(soil type$_{it}$) + ε_{it}

Description of the dependent variables

Wholesale water botex including network reinforcement (code: Botex+NR_WW) as reported in the published PR24 wholesale dataset.

Description of the explanatory variables

Total properties (code: properties) as reported in the published wholesale dataset.

Percent of water treated in bands 3-6 (code: pctwatertreated36)

Number of boosters per km of main (code: boosterperlength) as reported in the published wholesale dataset.

Weighted average density (code: density) as reported in the published wholesale dataset.

Weighted average density squared, calculated value, the square of the weighted average density variable

Regional weekly wage (LAD basis), AFW generated variable

Regional hourly wage (LAD basis), AFW generated variable

Soil type (index of susceptibility to shrink/swell) AFW generated variable

Brief comment on the models

The number of years in the sample was consistent with the standard OFWAT dataset. This was from 2011/12 to 2021/22. There was no difference between the sample panel structure and the published do-file. The variable was assigned a unique code on the do-file and was added to the "Independent variables (cost drivers)" section, as well as section 5 "Prepare the macros for different regression specifications".

We found all variables to show a statistically significant result for Wholesale Water when tested through the OLS and Random Effects models. Ln(regional wages hourly LAD basis) was the strongest variable, which was proven to a 5% significance level. The other variables, Ln(regional wages weekly LAD basis) and Ln(soil type) were statistically significant to 10% significance. All models display an R-squared of 0.972 and above.

	AFWWW4	**AFWWW5**	**AFWWW6**
Dependent variable	Botex including network reinforcement	Botex including network reinforcement	Botex including network reinforcement
Constant	-11.714** {0.021}	-6.158*** {0.001}	-2.040** {0.014}

Ln(number of connections)	1.081*** {0.000}	1.084*** {0.000}	1.075*** {0.000}
Percent of water treated in bands 3-6	0.004*** {0.000}	0.004*** {0.000}	0.005*** {0.000}
Ln(Wtd. Avg. Density)	-1.566*** {0.000}	-1.493*** {0.000}	-1.944*** {0.000}
Ln(Wtd. Avg. Density) squared	0.105*** {0.000}	0.098*** {0.000}	0.134*** {0.000}
Ln(booster pumping stations per length of main)	0.383*** {0.006}	0.384*** {0.003}	0.343*** {0.008}
Ln(regional wages weekly LAD basis)	4.700* {0.058}	-	-
Ln(regional wages hourly LAD basis	-	2.802** {0.018}	-
Ln(regional wage annual SOC basis	-	-	-
Ln(soil type)	-	-	0.101** {0.029}
Estimation method (OLS or RE)	RE	RE	RE
N (sample size)	187	187	187
R2 adjusted	0.972	0.973	0.973
RESET test	0.123	0.096	0.233
VIF (max)	278.361	262.808	230.152
Pooling / Chow test	0.969	0.968	0.887
Normality of model residuals	0.102	0.114	0.169
Heteroskedasticity of model residuals	0.000	0.000	0.000
Test of pooled OLS versus Random Effects (LM test)	0.000	0.000	0.000
Efficiency score distribution (min and max)	Min: 0.797 Max 1.295	Min: 0.791 Max 1.279	Min: 0.794 Max 1.330
Sensitivity of estimated coefficients to removal of most and least efficient company	AMBER	AMBER	AMBER
Sensitivity of estimated coefficients to removal of first and last year of the sample	GREEN	GREEN	GREEN

Significance level: *** (1%), ** (5%) and * (10%)

Efficiency scores distribution

AFWWW4

PRT	0.797
AFW	0.856
SEW	0.907
SSC	0.913
HDD	1.000
NWT	1.020
NES	1.021
SVE	1.033
YKY	1.043
SWB	1.072
ANH	1.088
WSX	1.095
TMS	1.115
WSH	1.141
BRL	1.189
SRN	1.292
SES	1.295

Median	1.043
Upper Quartile	1.000
Range	0.498

AFWWW5

PRT	0.791
AFW	0.836
SEW	0.862
SSC	0.936
NES	1.017
HDD	1.020
NWT	1.029
YKY	1.042
WSX	1.043
SVE	1.044
SWB	1.061
ANH	1.084
TMS	1.115

WSH	1.136
BRL	1.167
SES	1.252
SRN	1.279

Median	1.043
Upper Quartile	1.017
Range	0.488

AFWWW6

PRT	0.795
AFW	0.899
SSC	0.951
SEW	0.986
YKY	1.001
HDD	1.022
SWB	1.029
NWT	1.058
SVE	1.061
NES	1.062
ANH	1.085
TMS	1.090
WSX	1.124
BRL	1.150
WSH	1.165
SRN	1.254
SES	1.330

Median	1.061
Upper Quartile	1.001
Range	0.535

Template for submission of econometric models - Wholesale Water (2)

Econometric model formula:

AFWWW7

ln(WW botex plus NR$_{it}$) = a + β_1 ln(number of connections$_{it}$) + β_2 (ln(treatment complexity index$_{it}$)+ β_3 (ln(booster pumping stations per length of mains$_{it}$) + β_4 ln(weighted average density LAD$_{it}$) + β_5 (ln(weighted average density LAD$_{it}$))^2 + β_6 ln(regional wages weekly LAD basis$_{it}$) + ε_{it}

AFWWW8

ln(WW botex plus NR$_{it}$) = a + β_1 ln(number of connections$_{it}$) + β_2 (ln(treatment complexity index$_{it}$)+ β_3 (ln(booster pumping stations per length of mains$_{it}$) + β_4 ln(weighted average density LAD$_{it}$) + β_5 (ln(weighted average density LAD$_{it}$))^2 + β_6 ln(regional wages hourly LAD basis$_{it}$) + ε_{it}

AFWWW9

ln(WW botex plus NR$_{it}$) = a + β_1 ln(number of connections$_{it}$) + β_2 (ln(treatment complexity index$_{it}$)+ β_3 (ln(booster pumping stations per length of mains$_{it}$) + β_4 ln(weighted average density LAD$_{it}$) + β_5 (ln(weighted average density LAD$_{it}$))^2 + β_6 ln(soil type$_{it}$) + ε_{it}

Description of the dependent variables

Wholesale water botex including network reinforcement (code: Botex+NR_WW) as reported in the published PR24 wholesale dataset.

Description of the explanatory variables

Total properties (code: properties) as reported in the published wholesale dataset.

Index of treatment complexity (code: wac) as reported in the published wholesale dataset

Number of boosters per km of main (code: boosterperlength) as reported in the published wholesale dataset.

Weighted average density (code: density) as reported in the published wholesale dataset.

Weighted average density squared, calculated value, the square of the weighted average density variable

Regional weekly wage (LAD basis), AFW generated variable

Regional hourly wage (LAD basis), AFW generated variable Soil type (index of susceptibility to shrink/swell) AFW generated variable

Brief comment on the models

The number of years in the sample was consistent with the standard OFWAT dataset. This was from 2011/12 to 2021/22. There was no difference between the sample panel structure and the published do-file. The variable was assigned a unique code on the do-file and was added to the "Independent variables (cost drivers)" section, as well as section 5 "Prepare the macros for different regression specifications".

We found all variables to show a statistically significant result for Wholesale Water when tested through the OLS and Random Effects models. Ln(regional wages hourly LAD basis) was the strongest variable, which was proven to a 5% significance level. The other variables, Ln(regional wages weekly LAD basis) and Ln(soil type) were statistically significant to 10% significance. All models display an R-squared of 0.972.

	AFWWW7	AFWWW8	AFWWW9
Dependent variable	Botex including network reinforcement	Botex including network reinforcement	Botex including network reinforcement
Constant	-11.261** {0.016}	-6.453*** {0.000}	-3.001*** {0.000}
Ln(number of connections)	1.070*** {0.000}	1.074*** {0.000}	1.062*** {0.000}
Ln(wac)	0.353*** {0.005}	0.348*** {0.002}	0.347*** {0.007}
Ln(Wtd. Avg. Density)	-1.412*** {0.000}	-1.343*** {0.000}	-1.697*** {0.000}
Ln(Wtd. Avg. Density) squared	0.095*** {0.000}	0.088*** {0.001}	0.117*** {0.000}
Ln(booster pumping stations per length of main)	0.379*** {0.003}	0.381*** {0.002}	0.460*** {0.000}
Ln(regional wages weekly LAD basis)	4.098* {0.077}	-	-
Ln(regional wages hourly LAD basis)	-	2.472** {0.025}	-
Ln(regional wage annual SOC basis)	-	-	-

Ln(soil type)	-	-	0.078* {0.059}
Estimation method (OLS or RE)	RE	RE	RE
N (sample size)	187	187	187
R2 adjusted	0.972	0.972	0.972
RESET test	0.055	0.033	0.088
VIF (max)	261.797	249.752	222.794
Pooling / Chow test	0.908	0.893	0.829
Normality of model residuals	0.727	0.732	0.859
Heteroskedasticity of model residuals	0.000	0.000	0.000
Test of pooled OLS versus Random Effects (LM test)	0.000	0.000	0.000
Efficiency score distribution (min and max)	Min: 0.782 Max 1.335	Min: 0.775 Max 1.297	Min: 0.773 Max 1.382
Sensitivity of estimated coefficients to removal of most and least efficient company	AMBER	AMBER	AMBER
Sensitivity of estimated coefficients to removal of first and last year of the sample	GREEN	GREEN	GREEN

Significance level: *** (1%), ** (5%) and * (10%)

Efficiency scores distribution

AFWWW7

PRT	0.782
AFW	0.859
SSC	0.871
SEW	0.919
HDD	1.021
NWT	1.023
NES	1.026
ANH	1.054
SVE	1.056
YKY	1.058
SWB	1.060
WSX	1.085
WSH	1.126

TMS	1.129
BRL	1.147
SRN	1.268
SES	1.335

Median	1.056
Upper Quartile	1.021
Range	0.553

AFWWW8

PRT	0.775
AFW	0.841
SEW	0.878
SSC	0.890
NES	1.023
NWT	1.032
WSX	1.036
HDD	1.041
ANH	1.050
SWB	1.053
YKY	1.057
SVE	1.066
WSH	1.124
TMS	1.129
BRL	1.131
SRN	1.256
SES	1.297

Median	1.050
Upper Quartile	1.023
Range	0.522

AFWWW9

PRT	0.773
SSC	0.885
AFW	0.898
SEW	0.995
SWB	1.021
YKY	1.030

HDD	1.044
ANH	1.046
NWT	1.053
NES	1.060
SVE	1.085
WSX	1.105
BRL	1.106
TMS	1.111
WSH	1.141
SRN	1.231
SES	1.382

Median	1.053
Upper Quartile	1.021
Range	0.609

Appendix 1- Regional Wages

Differences in the costs of labour across regions have the potential to influence water company costs most obviously through wage rates paid directly to employees. We would also expect regional wages to influence the costs of services purchased by water companies, as suppliers (principally works contractors, but all other service providers) will naturally wish to recover their labour costs within the rates they charge.

In previous AMPs 3–4 Ofwat adjusted model outcomes for regional wages through cost adjustment claims and it included regional wages as an explanatory variable in its 2014 models. Ofwat changed the position in 2019, excluding regional wages from its determination, arguing that in the econometrics work, the regional wages variable had low predictive power and sometimes showed inconsistent (negative) sign. It further noted that the regional wage drivers in models proposed by Thames Water were not statistically significant and that its own models, that include density variables, capture the effect of regional wages as the two are correlated.

That said, in 2019 Ofwat's independent academic advisor Prof. Andrew Smith noted that areas for reflection going forward include 'further consideration of how to model regional wage effects.' Further, since 2019 Ofgem has published its ED2 determinations where it made regional adjustments to cost data prior to econometric estimation, so there is recent regulatory precedent for including regional wage effects in cost estimation work. Closer to home in water, there is now a longer time series of data available for study, so it may be possible to establish relationships not previously observable within shorter data series. Therefore we consider it right to re-examine the case for regional wages in cost assessment.

Explanatory variables for regional wages

Our approach has been to develop regional wages series from the ONS Annual Survey of Hours and Earnings (ASHE), for econometric testing. We consider the ASHE dataset an authoritative source for UK wages data. It is available for all years in the study period 2011/12 to 2021/22 and with suitable regional granularity to measure the phenomenon we wish to study. The datasets present information in multiple ways which means there are options for how best to construct regional wage series for econometric analysis.

- Weekly or hourly wages
- Median or mean wages
- Including or excluding overtime
- All occupations, or occupations most relevant to water
- By large, sub-national regions or local authority districts
- Part time, full time or all employees
- Male and female employees

Our preference is for weekly pay because water company staff are for the most part salaried rather than paid per hour. However, for completeness we estimated both approaches. We chose to follow CEPA's 2018 approach in measuring mean wages rather than median.

We have measured wages including overtime. For some salaried water workers, working longer hours than usual does not typically result in overtime payments. For others, overtime is an important part of remuneration, for instance, workers engaged in repair jobs that may overrun and for emergency work that occurs 'out-of-hours'. Water companies need to use overtime to compete for and secure certain types of labour and to manage temporary fluctuations in workload, such as that caused by adverse weather. We concluded that series that include overtime provide the best measure.

By selecting 'all occupations' we are likely to capture the influence of regional wages on costs as they are influenced by prices for bought-in services as well as that from directly employed staff.

We chose to measure the wages of 'all employees' because water companies employ both full and part time, and male and female workers. Data to weight ASHE statistics according to contracted hours or gender mix of workers across water companies is not publicly available.

Therefore we have collected data in two ways:

- Gross Weekly Wages, all employees by local authority area
- Gross Hourly Pay, all employees by local authority area

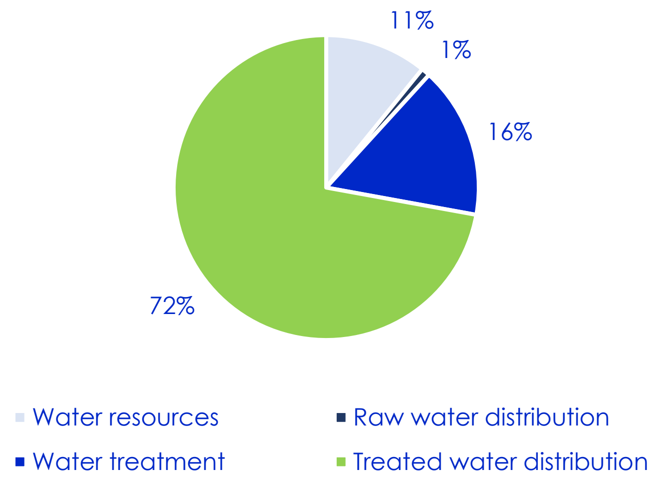
We obtained regional data for wages from the ASHE series, tables 8.1a for weekly wages, and table 8.5a for hourly earnings. These data are collected at local authority level, so it is possible to derive wage measurements in each company area, using the local authority allocation factors released by Ofwat in early 2021.

We considered the suitability of regional wage data based on 2-digit occupational codes, noting that this was the approach used by CEPA in 2018. Using occupational codes has the advantage that it can exclude occupation types that are unlikely to be used in water supply activities. However, to construct data necessarily requires assumptions about the mix of occupations represented in the water industry. Those mixes might be different across the retail, treated water distribution and water resources plus activities and are likely to differ across companies. We see a risk that regional data series constructed in this way would reflect the assumptions made about labour mix as much as they do differences in pay.

Discussion

We confined our regional wages modelling to the treated water distribution activity and to the aggregate wholesale water activity. This is because evidence from annual reports shows that water FTEs are concentrated in the treated water distribution segment, both for AFW and the industry overall. It follows that we are most likely to observe the relationship between costs and regional wages in this segment.

Affinity Water : Distribution of FTEs by Segment, 2020

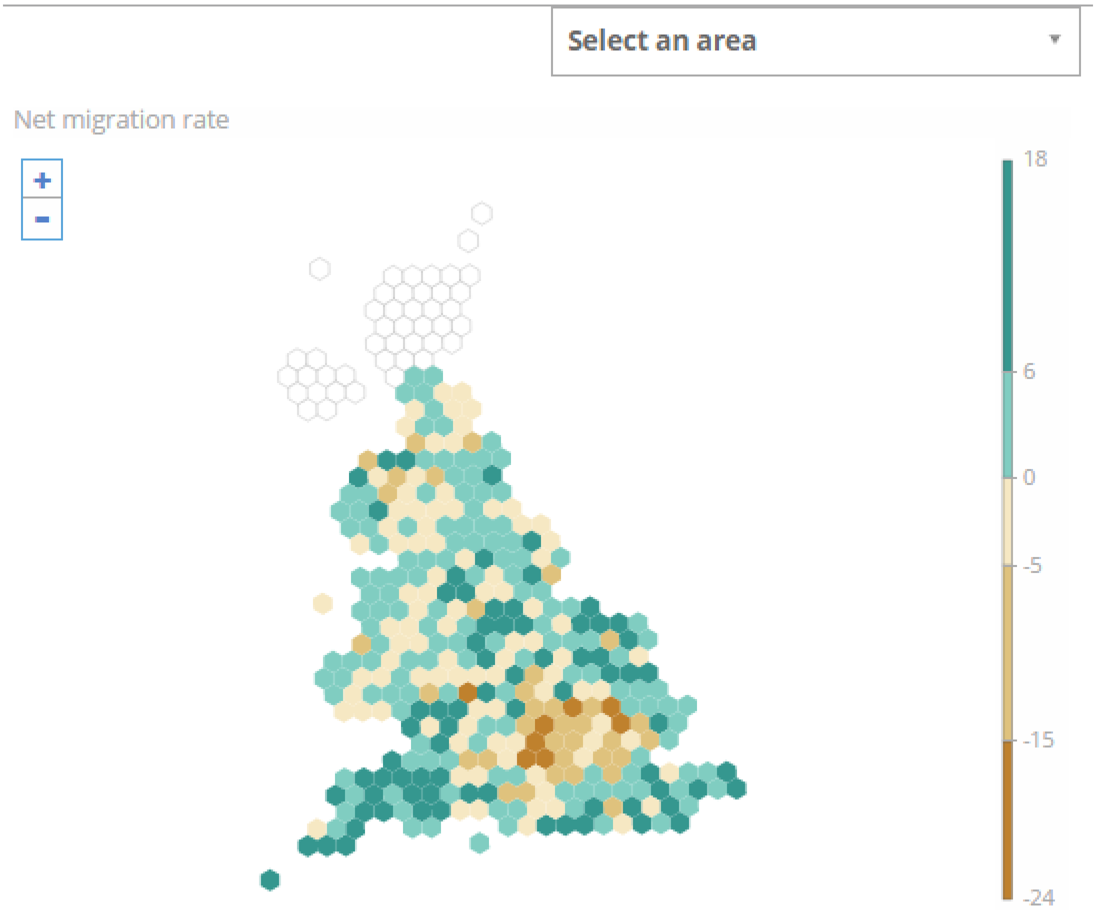


Following CEPA 2018, the extent to which costs are influenced by regional wages can be thought to depend on:

- The extent to which labour markets allow regional differences in wage rates to persist
- The extent to which companies can reduce their exposure to labour costs, for instance through capital substitution or sourcing labour from outside their areas of operation

Economic theory would predict that workers would be attracted by higher regional wages and that over time, labour mobility would compete away regional differences. However, we do not observe this effect in evidence. Instead, ASHE data is consistent with frictions in labour markets that allow regional wage differentials to persist over time.

ONS evidence (see chart) shows patterns of net migration. The largest net migration outflows are from the London area, particularly NW London. The largest inflows are into the Southwest. Three local authority areas with the largest net outflows include Brent, Ealing and Hounslow and there are significant outflows from Luton, Watford, Barnet and Harrow. All these local authorities are entirely, or largely within Affinity Water's supply area. Whilst economic theory would expect net migration from low wage areas toward higher, the evidence on net migration suggests flow in the opposite direction. This further suggests a diminishing pool of available labour in Affinity Water's (and some other companies) areas which may be exacerbating regional wage differentials rather than reducing them.



Source: Internal migration, ONS

Link : [Internal migration, England and Wales - Office for National Statistics (ons.gov.uk)]

We also noted research that shows that labour migration between regions is declining. So there is recent evidence that the ability of labour markets to compete away regional wage differentials has diminished. The P. McCann study (2013) notes "The majority of households do not relocate interregionally or internationally, such that when they do relocate, it tends to be within the same broad locality or region. Indeed, there is some evidence that over recent years interregional migration rates have actually been falling."

Our assessment of the evidence is that labour markets and inter-regional migration are not sufficiently strong forces to compete away regional differences, therefore the evidence suggests that water companies located in high wage areas will be exposed to higher costs. Accordingly we think regional wages are a valid candidate variable for econometric testing and further consultation.

Appendix 2 - Soil Type

Differences in the types and properties of soils across regions have the potential to influence water company costs as different soils are differently aggressive to underground water infrastructure. In common with others, (see box) we recognise the extent to which soils are susceptible to shrink/swell effects as being a key driver of burst pipes, leaks and subsequent repair activity, and a driver of mains replacement need. We consider that the relationship between shrink/swell effects and underground infrastructure is well established and accepted across the industry.

As part of the Cost Assessment Working Group activities, we proposed soil type as an area for summer 2022 additional information data collection, however Ofwat did not prioritise data collection for this item. Therefore, we have constructed our own soil type data from publicly available sources for econometric testing.

Susceptibility to Shrink/Swell

Clay soils, for instance, have a far greater volume when wet than dry. Houses, pipelines or roads built on clay are thus subject to greater ground movements and stresses in the summer and autumn when these soils shrink and swell with changing moisture conditions. *(Tim Farewell, Senior Research Fellow at Cranfield University)*

The main reasons why bursts in London are more frequent are…soil condition. In London, about 40% of soil is highly or very highly corrosive to iron mains and about 40% of soil is highly or very highly shrinkable, making it more susceptible to movement through changes in conditions; *(Thames Water Annual Performance Report 2021, p63.)*

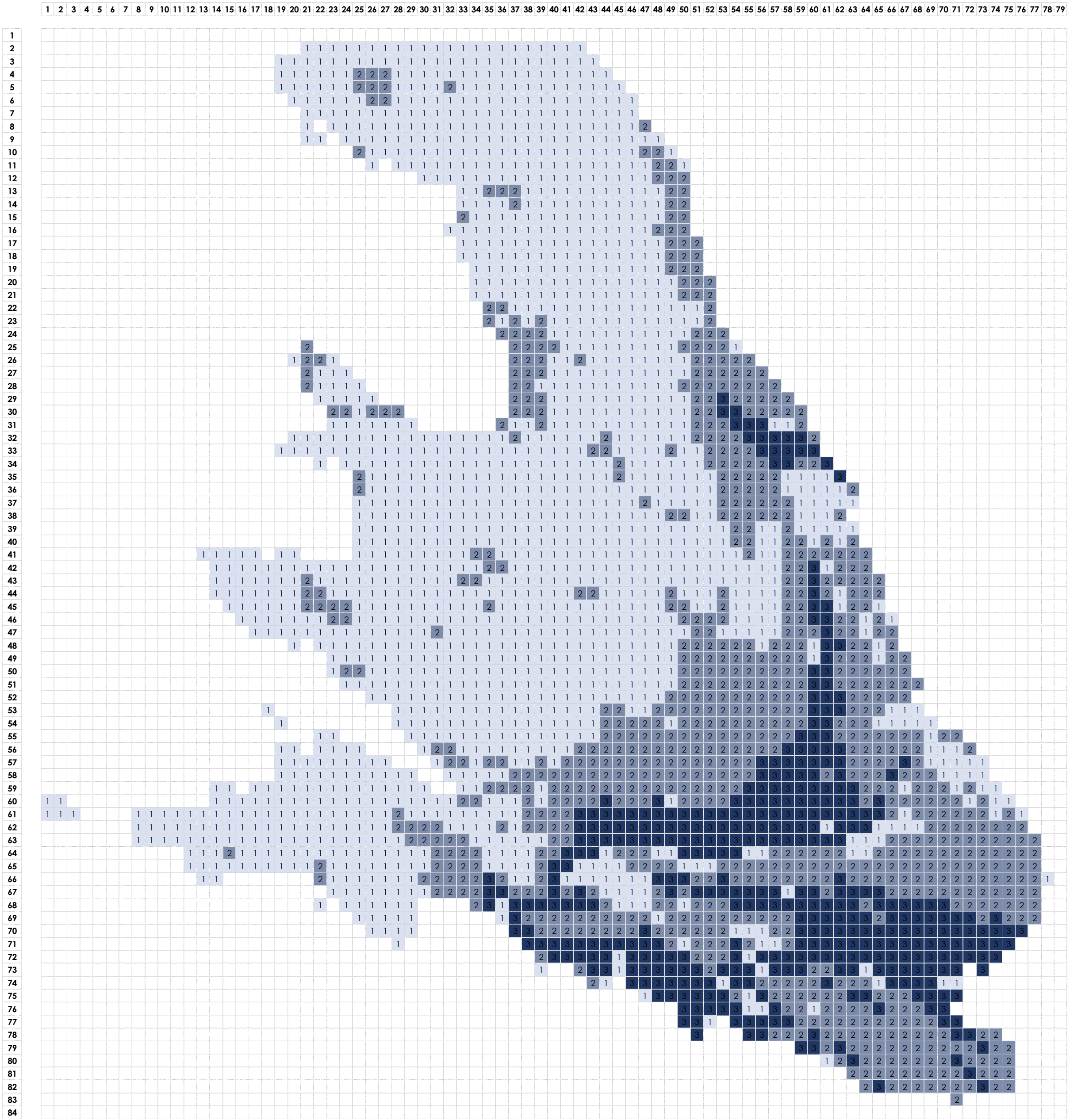
An overall conclusion is that there is a relationship between the number of bursts and hotter, drier summers shrinking certain soils, and the subsequent re-wetting and swelling of the soil in the autumn *(Soil and climatic causes of water mains infrastructure bursts – A report for Anglian Water 2012 Anglian Water Bursts (cranfield.ac.uk)*

Our approach has been to use data published by the British Geological Survey (BGS) to construct an index that characterises the soil conditions in each water company supply area in terms of soil susceptibility to shrink/swell. BGS rate soils as 1-3, corresponding with low, moderate or significant susceptibility and we have used this to calculate an average score for soil type in each company area.

We have further weighted this score to adjust it for the prevalence of ferrous mains in each water company area, since it is this pipe material that is most at risk from exposure to shrink/swell effects. To do this we have used APR data to calculate a weighting factor, being the percentage of total mains length accounted for by mains laid or substantially refurbished prior to 1961. We used 1961 as a threshold, since after that date, use of cast and ductile iron for water mains was in decline.

The maps on the next page show the distribution of soil types across the country and our mapping to water company areas. We consider that soil type is a valid candidate variable for econometric estimation and further industry consultation.

Map 1 - Shrink Swell - Level of Susceptibility to Ground Movement

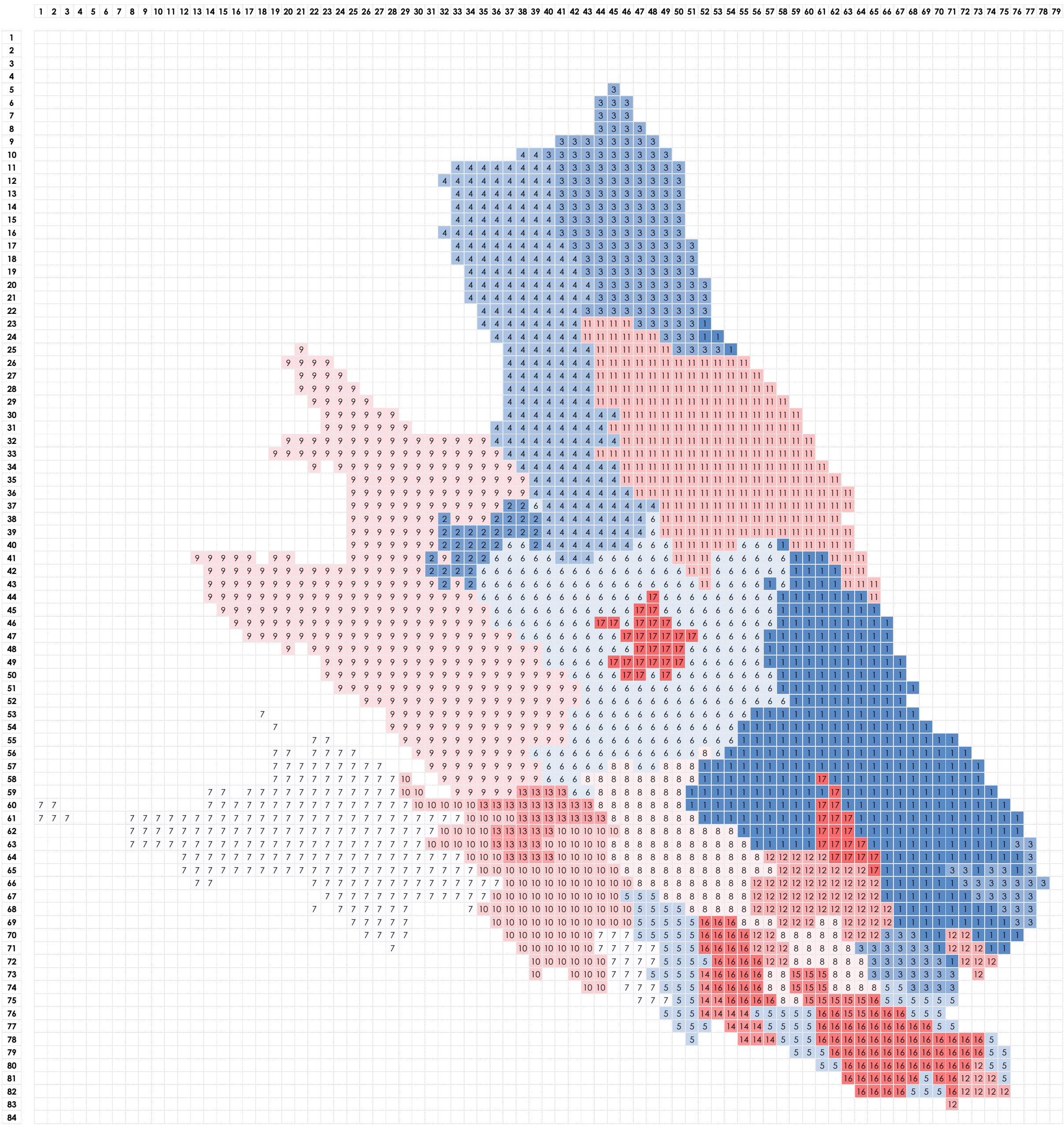


Key - Level of Susceptibility to Ground Movement

1 **- Low**
2 **- Moderate**
3 **- Significant**

Source : British Geological Survey (GeoSure 5km Hex v.8 dataset)

Map 2 - Water Company Water Supply Areas



Key - Water Supply Company

- 1 ANH - Anglian Water
- 2 HDD - Hafren
- 3 NES - Northumbrian (incl. Essex & Suffolk)
- 4 NWT - United Utilities
- 5 SRN - Southern
- 6 SVE - Severn Trent
- 7 SWB - South West
- 8 TMS - Thames
- 9 WSH - Welsh
- 10 WSX - Wessex
- 11 YKY - Yorkshire
- 12 AFW - Affinity
- 13 BRL - Bristol
- 14 PRT - Portsmouth
- 15 SES - Sutton & E Surrey
- 16 SEW - South East Water
- 17 SSC - South Staffs / Cambridge

Appendix 3 - Alignment with Ofwat's Principles

The table below summarises our assessment of our proposed additional variables against Ofwat's assessment criteria.

Regional Wages		
Ofwat Principle		**Notes**
Data is of good quality	G	Our data is derived from the ONS ASHE series, the authoritative source for UK wages data
Consistent with engineering, operational and economic rationale	G	There are sound economic reasons to think that regional wage differences will lead to different water company costs across regions. The existence and persistence of regional wage differences is overwhelmingly supported by evidence.
Sensibly simple and transparent	G	We have provided spreadsheets showing the construction of the variables we used and being from public data are 100% transparent
Focus on exogenous cost drivers	G	Labour market conditions are exogenous, and companies have limited capability to insulate themselves from these costs. Evidence shows that labour migration is not effective at competing away regional differences.
Robust econometric models	G	Our econometric estimations are summarised in the templates above. We found statistically significant estimators that have the signs expected and perform well against tests of sensitivity and stability.
Correlation with other variables (e.g. density)	G	We have tested the OLS VIF statistic for both regional wages variables and as our result is around 2.0, we conclude that multicollinearity between regional wages and other variables (e.g. density) is not problematic.
Soil Type (Index of susceptibility to shrink/swell)		
Data is of good quality	G	We consider the BGS soil maps an authoritative source for characterising soil types. Most water companies purchase BGS data for their areas to inform and support their operations.
Consistent with engineering, operational and economic rationale	G	We think the susceptibility of ferrous mains to different soil conditions is widely accepted by engineers across water companies and is supported by published research
Sensibly simple and transparent	G	We constructed our data from public sources, so is transparent
Focus on exogenous cost drivers	G	Soil type is not within company control, and the stock of ferrous mains is not controllable except over the very long time periods needed to renew ferrous mains with pipes made from modern materials.
Robust econometric models	G	Our econometric estimations are summarised in the templates above. We found statistically significant estimators that have the signs expected and perform well against tests of sensitivity and stability.
Correlation with other variables (e.g. density)	G	We have tested the OLS VIF statistic for soil type and as it is close to 1, we conclude that multicollinearity between soil type and other variables (e.g. density) is not problematic for our models.

References

Farewell, T.S., Hallett, S.H., and Truckell, I.G. (2012) Soil and Climatic causes of water mains infrastructure bursts. 111pp. NSRI Research report. NSRI, Cranfield University, UK.

McCann, P. (2013), Modern Urban and Regional Economics, second edition, Oxford: Oxford University Press

Thames Water, Annual Performance Report 2021, p63.