

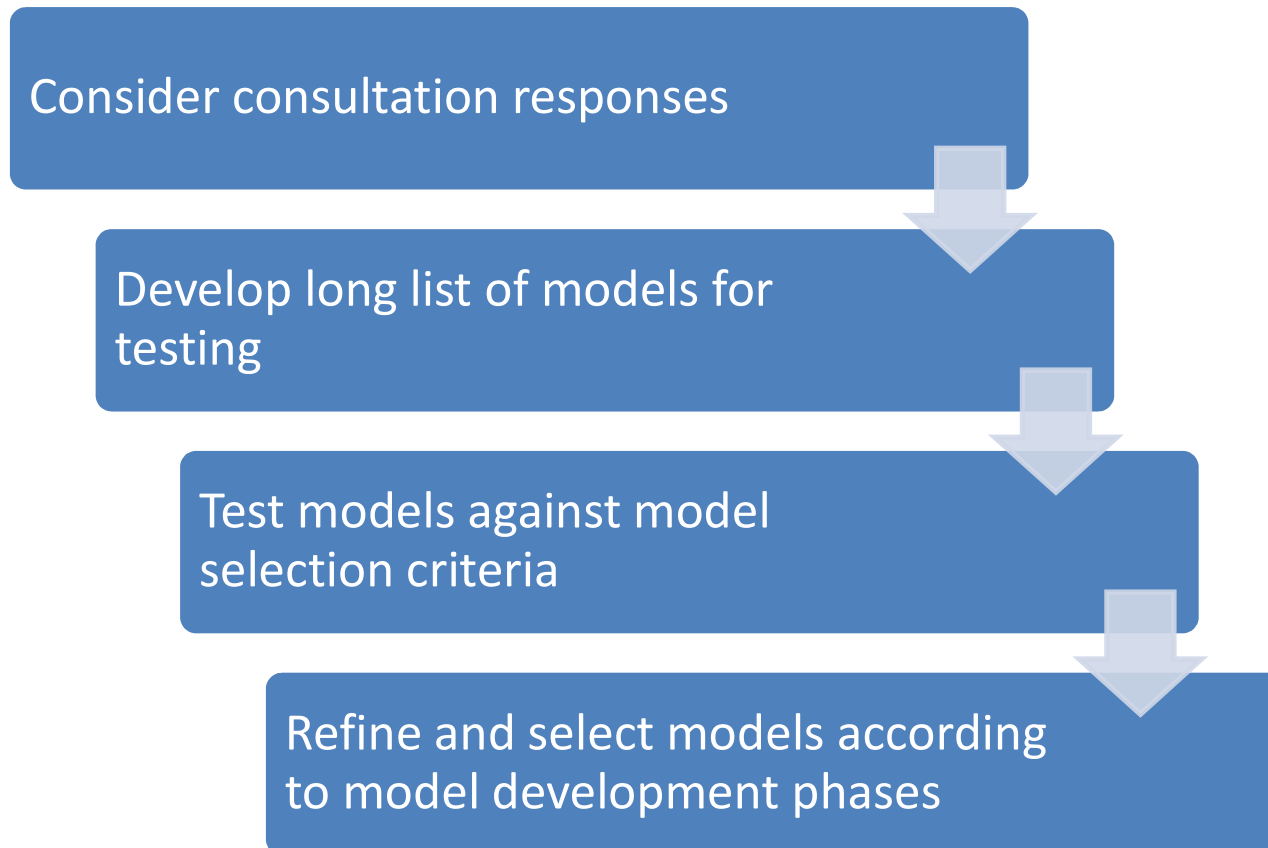
RIO2 Cost Assessment Working Group

Meeting 11



1. Summary of *RIO-2 tools for cost assessment* consultation responses
2. Update on model testing
 - SFA testing
 - MEAV update
 - Model testing
3. Future meetings
4. AOB

- Ofgem – to share the output of the regional factor modelling presented at the CAWG (Cadent suggestion to share models using fully normalised data, and the data with some normalisations removed)
- Ofgem – circulate the updated CAWG terms of reference for GDN comment
- Ofgem – pick up CISBOT discussion in asset management working group on 7th August



Summary of RII0-2 tools for cost assessment consultation responses



Most respondents support the continued use of COLS, however WWU suggested we should also be considering SFA and DEA.

*"The preference for cost assessment based upon COLS is now viewed as a key element of the regulatory framework and is well understood and accepted. As such it provides a highly valuable level of stability to the framework and in a low risk and return environment can be seen to be adding significantly to the strong incentive properties of the framework which will drive both short and long term value for customers" – **NGN***

*"...simplicity and transparency should not be given disproportionate weight compared to the accuracy of estimating relative efficiency in deciding the appropriateness of a method...we consider it extremely unlikely that DEA or SFA would not yield new insights compared to COLS" – **WWU***

Most respondents supported the continued use of workload cost drivers.

*"It is better to use a driver that makes economic and engineering sense, rather than one which makes less sense but is entirely outside of the companies' influence" – **Cadent***

*"...given limitations in the available data, it may be better, in some cases, to include an explanatory variable that carries risks of endogeneity, than to fail to take any account of potentially important differences between companies" – **CMA, Bristol Water (2015) (via Cadent)***

*"If the regulator is concerned about inflated workload forecasts then this can be dealt with through the use of PCDs and volume drivers" – **NGN***

*"...there are some categories where the workload drivers that were used at GD1 do not have a good fit with the cost data...Despite these issues, we think there is a valuable role for analysis based on workload for cost categories where there is a large volume of broadly repetitive work which would be expected to have similar unit cost" – **SGN***

Respondents suggested options for cost pooling, while maintaining a preference for more aggregated models.

*"...to assess the opex plus costs using one approach, and the bulk of repex and mains reinforcement using another, would lead to a significant level of inconsistency and cherry-picking, and deliver results that were not robust and low confidence" – **Cadent***

*"Emergency, Repair and Operations Management have a strong correlation and trade-offs within opex, as well as clear issues with cost boundary complexities. Operations Management could be a substitute for Emergency and Repair – stronger management functions could lead to a reduced need for operational employees" – **NGN***

*"'Opex plus', if it is used, should be used as an additional cross-check to the totex modelling, not as a replacement" – **SGN***

Respondents supported assessing gross business support costs at the group level, and provided benchmarking suggestions.

*"Regression analysis would have the benefit that it would take account of any economies of scale that might apply to larger groups that can spread their BSCs across a number of licensees. The current ratio analysis does not account for possible economies of scale" – **SGN***

*"The composite cost driver is reasonable and has the advantage of being broadly comparable across many organisations, not just utilities. We also support the use of MEAV as a cost driver, but only within each regulated sector, as it will not be available for the external world, and will not be comparable across regulated sectors" - **Cadent***

We received mixed responses to how we should estimate and model the impact of regional factors.

*"We do not believe there are sufficient data points, with eight GDNs, to attempt to apply within-model adjustments" – **Cadent***

*"We agree with Ofgem's prior expectation that density is more likely than labour to be suitable for within-model adjustments but the options will need to be fully tested by Ofgem" – **SGN***

*"...imposing symmetrical adjustments is essentially arbitrary and has no economic rationale" – **SGN***

*"...even if pre-modelling adjustments are used, regression analysis may be helpful to test for significance of a regional factor (in addition to other cost drivers)" – **WWU***

*"We disagree with Ofgem that each RPE needs to be shown to be material relative to both totex and CPI. ...If an RPE makes up a small proportion of totex but deviates significantly from CPI for a sustained period of time, this could result in a material impact on costs" - **SGN***

Mixed views on whether RPE indices need to be forecastable – on one hand indexation avoids reliance on forecasts, yet the ability to forecast may imply robustness as well as providing foresight to consumers/companies.

*"If labour is split into direct and contract labour, it would be difficult to attain appropriate indices for some specialized contract labour categories. Moreover, as it is a company's choice to employ direct or contract labour, providing an RPE allowance for contract labour may incentivise firms' employment choices" – **WWU***

*"Approaches such as SFA and DEA can be considered to decompose productivity growth achieved by the industry into its constituent parts and therefore isolate the impact of frontier shift" - **WWU***

*"...which suggests that RPEs and productivity improvements will offset each other, leading to our recommendation of a zero RPEs allowance and a zero ongoing efficiency target for RIIO-T2." – **SPEN***

Respondents supported placing the same or greater weight on top-down models.

*"The totex approach is the only approach to balance trade-offs between the different activity, structural and reporting choices open to GDNs, and is therefore the approach in which Ofgem should have the highest level of confidence" – **Cadent***

*"If Ofgem continues to see value in using both aggregated and disaggregated models, we recommend ensuring first that the disaggregated models are suitably robust and accurate. If this can be achieved, an equal weighting should be applied to each" – **SGN***

*"No one knows what the true underlying relationship is and reliance on only one model risks that the outcome may be due to the idiosyncrasies of the particular model used. As such, multiple models and multiple levels of aggregation should be considered" – **WWU***

Model estimation techniques

- Test SFA, compare against COLS models

Cost drivers

- Test new cost drivers (including both scale and workload drivers) and compare with existing models

Cost pooling

- Test cost pooling options and compare with existing models

Business support costs

- Test econometric approach

RPEs and ongoing efficiency

- Test direct and indirect approaches to measurement of ongoing efficiency

Combining the elements of our cost assessment

- Test a range of models before deciding on model combinations

- Option to use a future meeting (in 2020) to discuss and share evidence
- Regional wages
 - Update regional wage data
 - Is it still appropriate to use the three digit SOC codes?
 - GDNs to complete occupation table in BPDT (2.13 for December submission) – this will enable us to update the indices
 - Is the area inside the M25 still the best proxy for the London region?
- Urbanity/sparsity
 - Test density measure within models
- Assess company-specific factors separately

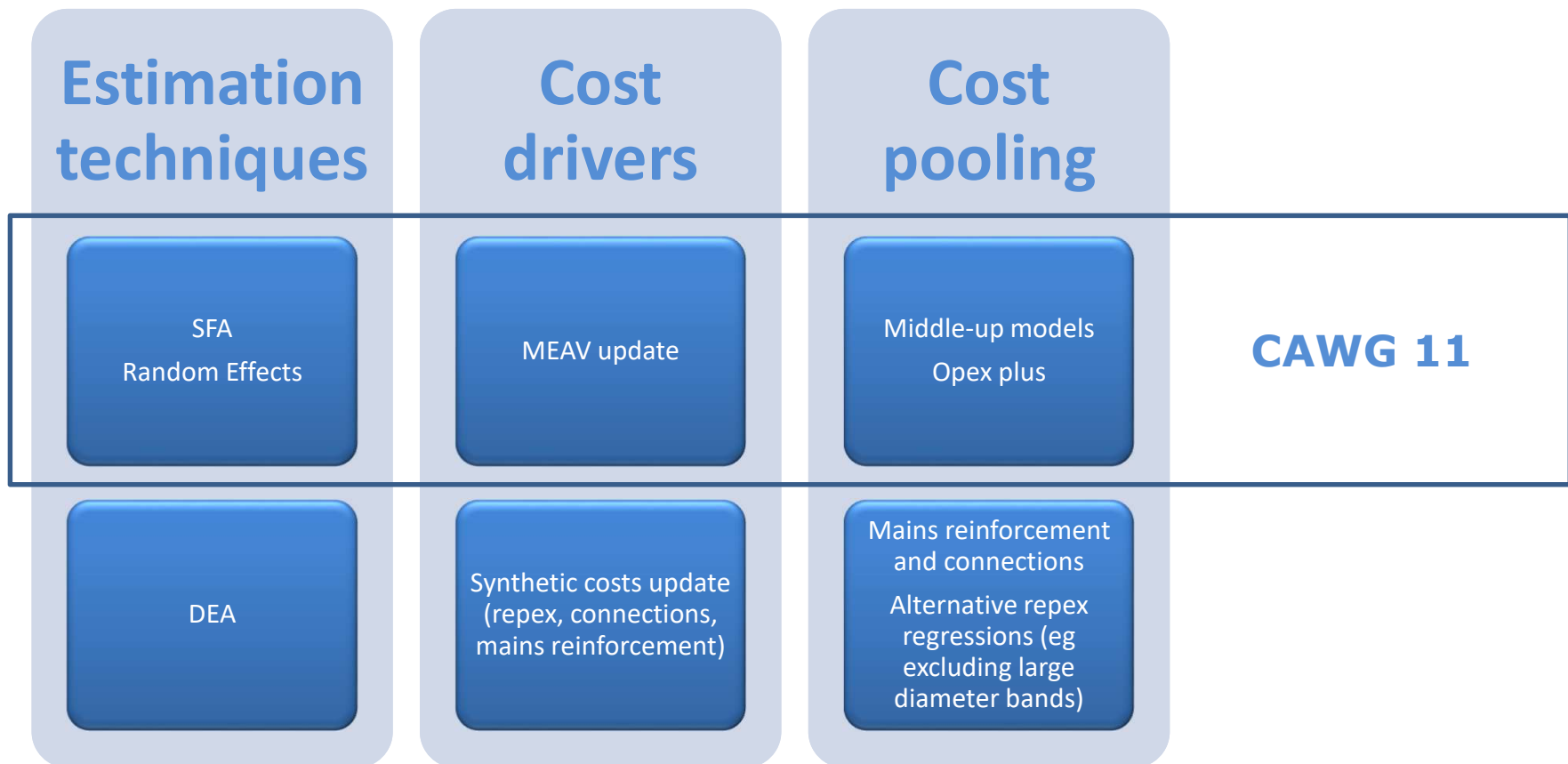
Cadent suggested it would be helpful to build a further formal step into the process.

"Given where we are, we believe it would be helpful to build a further formal step into the process, to carry out an Initial Thoughts consultation in the spring of 2020, to share Ofgem's developing thinking and modelling results prior to the Draft Determination, which would allow for two iterations of the approach before the Final Determination, rather than only one"

- Intention is to continue engagement via CAWGs in 2020
 - Targeted discussions, quicker feedback

Update on model testing





- **Parametric technique** to estimate production or cost frontier (ie maximum output or minimum cost)
 - Differently from the Corrected OLS (COLS) approach, Stochastic Frontier Analysis (SFA) allows for separation between statistical noise and inefficiency
- Many models available, depending on the **assumptions** on
 - Distribution of the error term
 - Efficiency over time (time-varying vs. time-invariant)
- Used by other regulators (eg Germany)
 - Data intensive

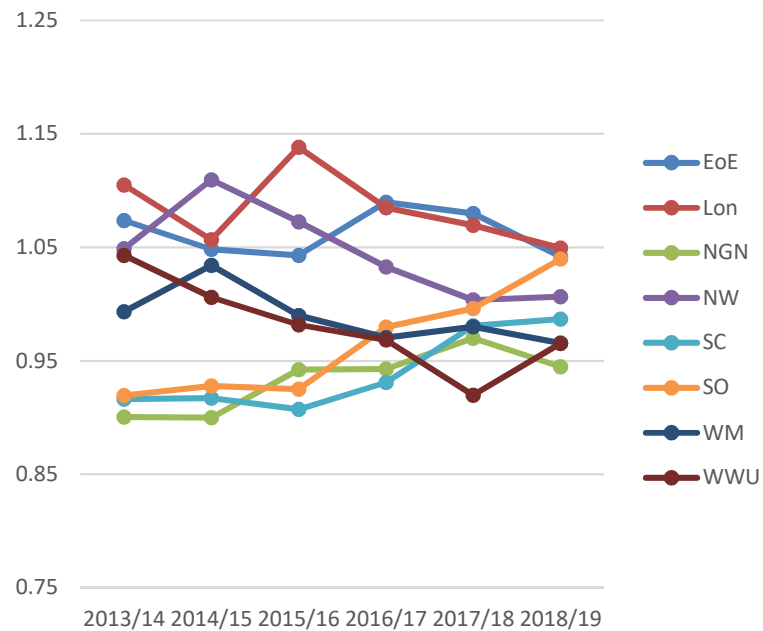
- We compared different estimation techniques using **RIIO-GD1 historical data** (2013/14 to 2018/19)
 - Cross-section analysis: **OLS**
 - Panel data analysis: Random effects (**RE**) and time-varying efficiency SFA model (**SFA_P** - Battese and Coelli, 1992)

Log_totex	OLS	RE	SFA_P
Log_totexcsv	.787*** (.033)	.736*** (.065)	.815*** (.035)
Trend	-.020*** (.006)	-.020*** (.003)	-.007 (.005)
Constant	-.045 (.228)	.305 (.447)	-.421 (.260)
Log likelihood	64.585	80.376	85.295
R2	.927	(.930)	

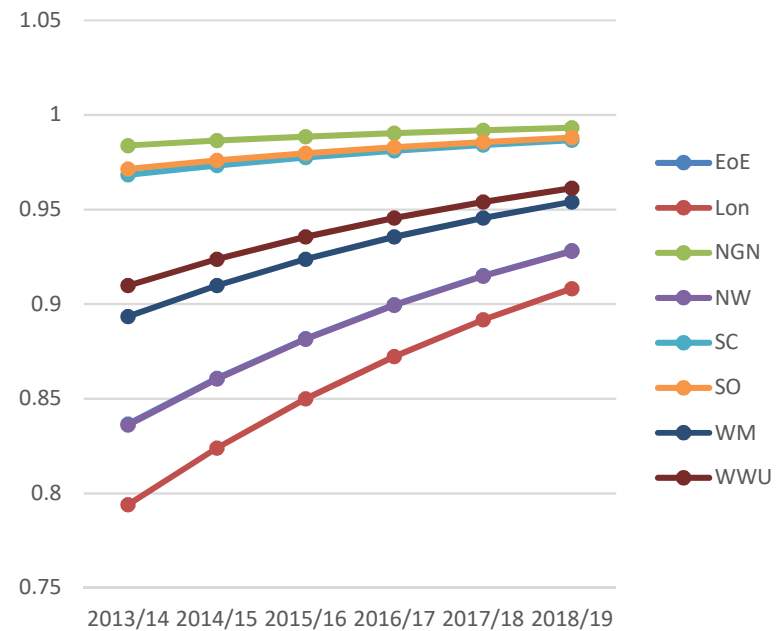
Similar results
independently of the
approach

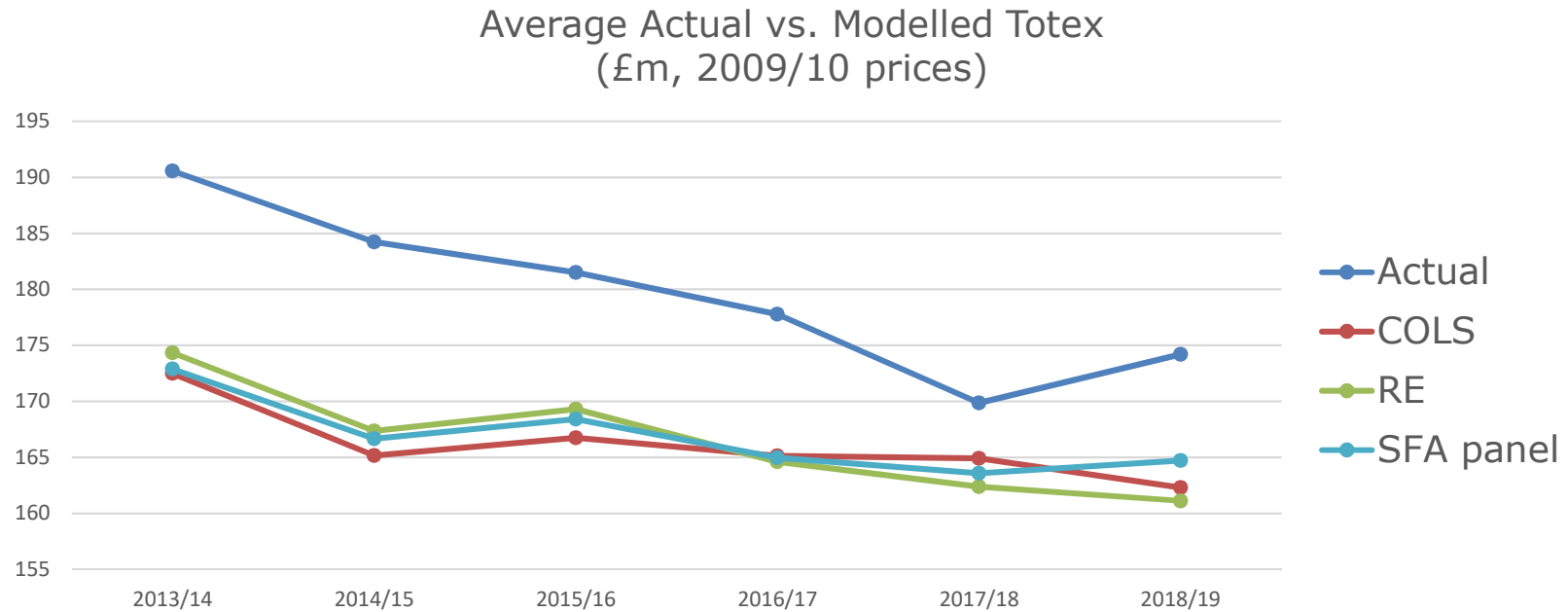
Statistically significant,
decreasing trend of costs

OLS



SFA panel



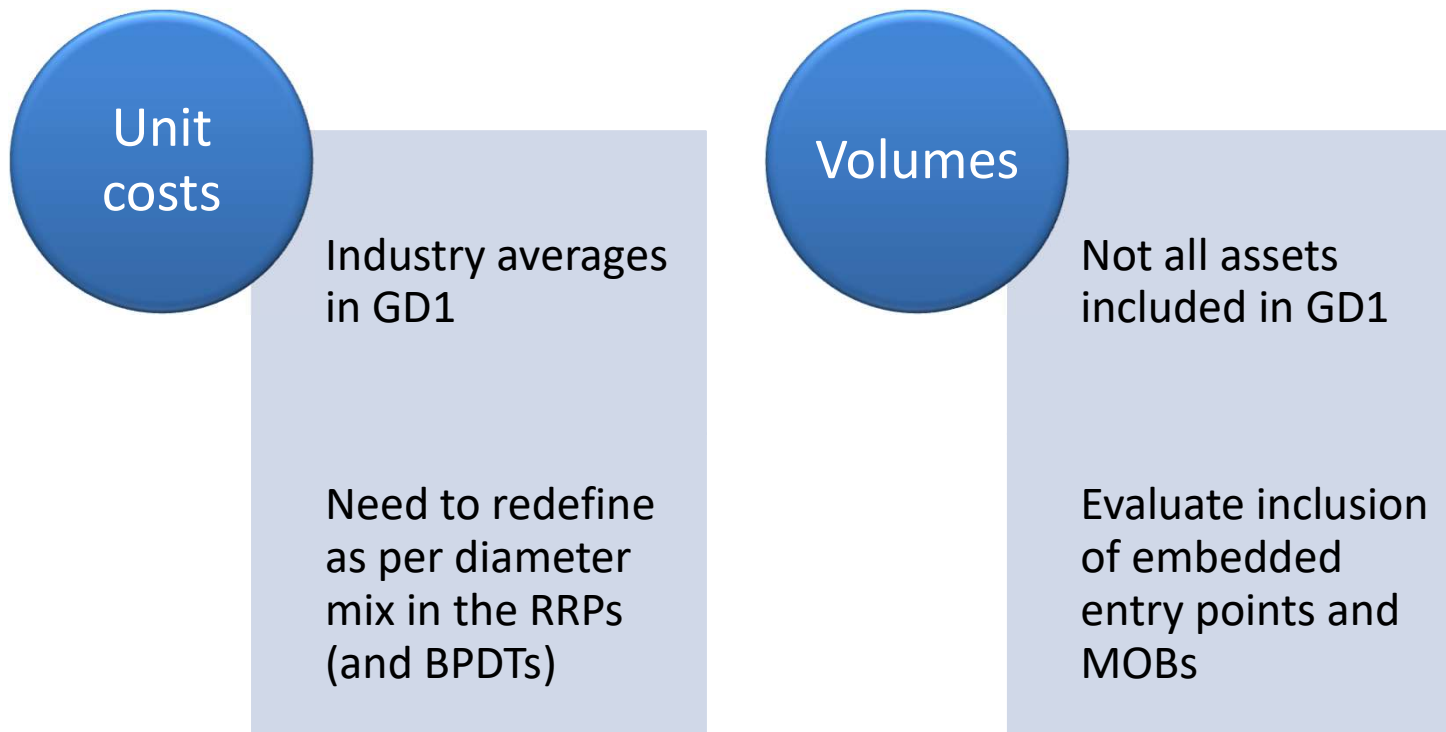


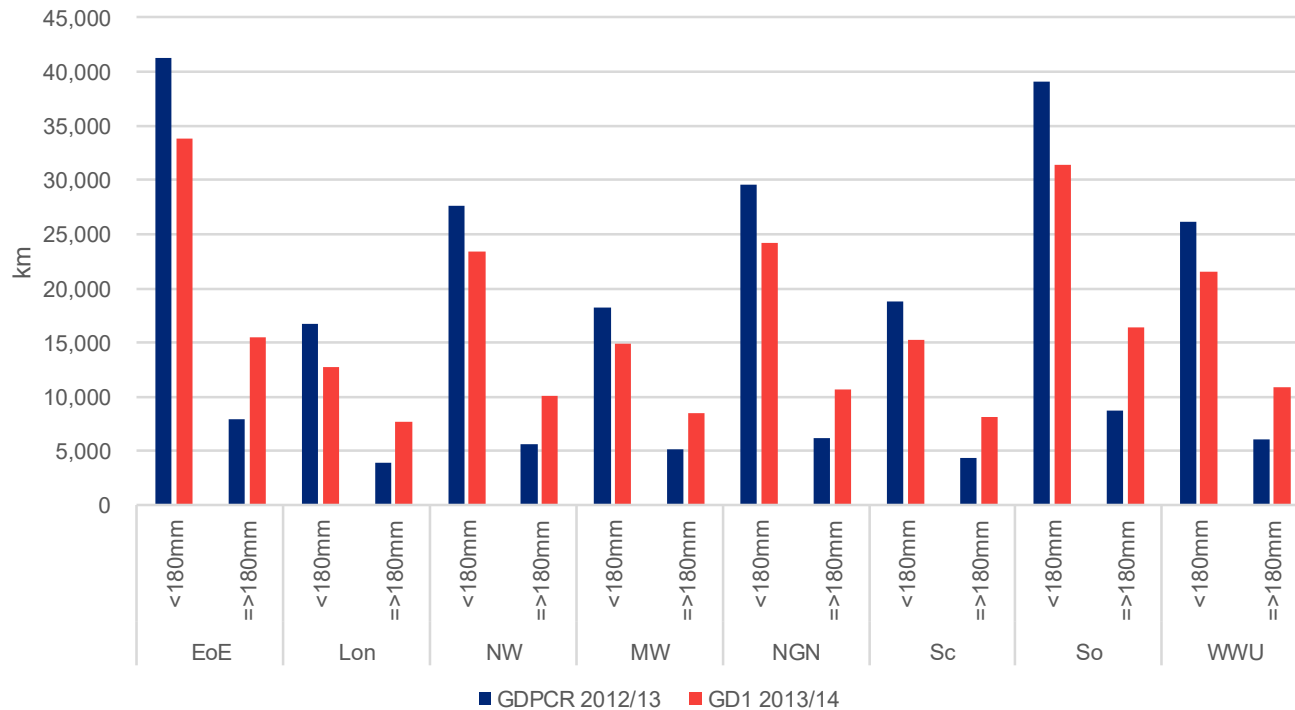
SFA **results very similar** to COLS and RE
But need to compare with other SFA models

COLS (with UQ view) seems to provide a slightly **tougher efficiency benchmark**

- **Similar rankings and efficient totex** with different estimation techniques
 - SFA worth exploring as a robustness check for OLS, but can it really be an alternative given the strong distributional assumptions and data requirements?
 - Moreover, the same analysis would be difficult to implement for bottom-up models
- **Next steps**
 - Additional SFA models, also on GD2 data
 - DEA model
 - With one output, testing a variable returns to scale model is the only reasonable option
 - Exploring whether a DEA model with one input (adjusted totex) and several outputs (eg individual components of totex CSV) could be an additional robustness check

- **MEAV** (Modern Equivalent Asset Value) is the current replacement value of an asset



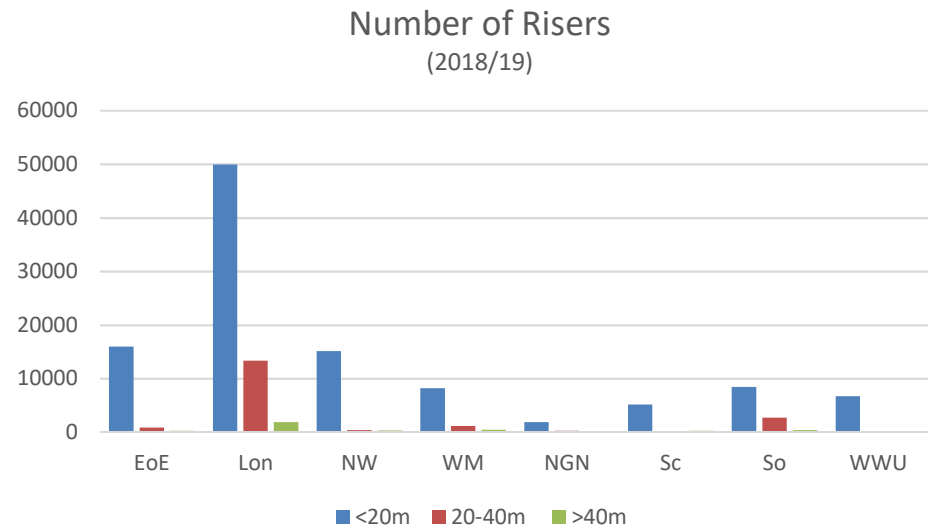


- During the GD1 period, the length of mains has been collected on different diameter bands compared to those used for setting allowance at GD1
- On average mains km below 180mm are 19% lower than in GDPCR, while mains km above 180mm are 80% higher -> higher MEAV

- To adapt to new diameter bands, proposed **linear interpolation** based on:
 - assigning the 2009/10 unit cost to the mid-point of historical bands
 - interpolating the unit costs based on the difference between the midpoint of the old bands and the new bands

Band	Dimensions	Mid-point of diameter range	Linear interpolation	Cadent by fifths approach
GDPCR band	< 76mm	37.5	73,300	73,300
Band A	To 109mm	54.5	74,783	73,562
GDPCR band	101.6 - 127mm	114.3	80,000	80,000
Band B	110-179mm	144.5	98,667	87,333
GDPCR band	152.4 - 177.8mm	165.1	111,400	111,400
Band C	180-212mm	196	180,134	131,811
GDPCR band	203.2mm- 228.6mm	215.9	224,400	224,400
Band D	213-249mm	231	239,809	229,532
Band E	250-267mm	258.5	267,872	271,878
GDPCR band	254mm- 304.8mm	279.4	289,200	289,200
Band F	268-399mm	333.5	375,778	381,015
GDPCR band	304.8mm- 457.2mm	399	480,600	480,600
Band G	400-499mm	449.5	562,024	600,455
GDPCR band	457.2mm- 609.6mm	533.4	697,300	697,300
Band H	500-709mm	604.5	739,206	758,260
GDPCR band	> 609.6mm	904.8	916,200	916,200
Band I	>710mm	955	967,032	996,296

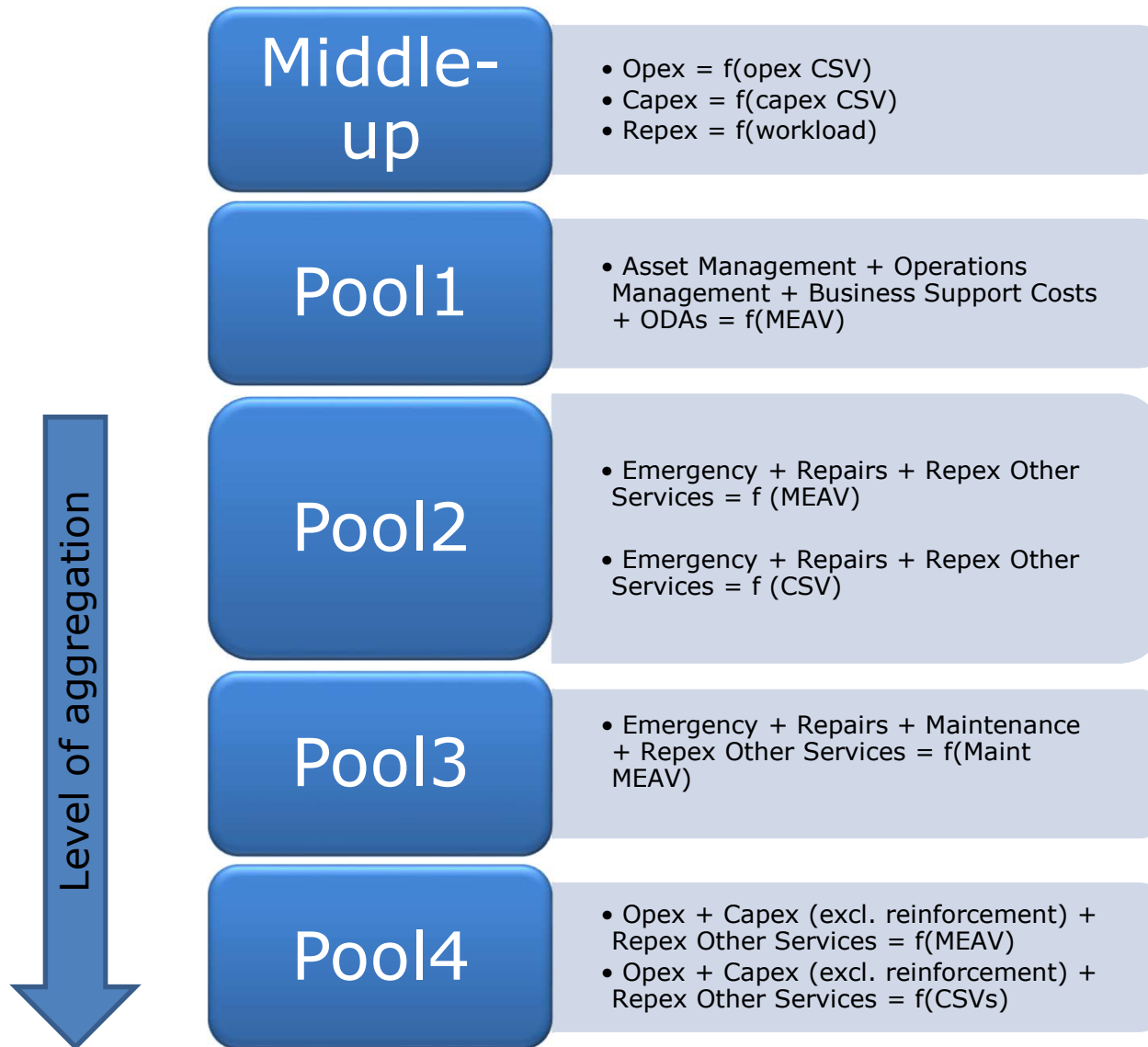
- The inclusion of **Embedded Entry Points** looks reasonable
 - Number increased over GD1. Why?
 - But...Can we get replacement data? Cadent's proposal (£1m > 7bar, £.5m < 7bar) based on adoption rates
- The inclusion of **MOBs** in MEAV also seems reasonable provided they are relevant cost drivers
 - Major effect on London
 - Do risers numbers reflect scale, complexity and costs?
 - Regional factors?
 - Need to solve reporting issues
 - Is data reliable?
 - Separate assessment?



	Log_work_mgt	Log_work_mgt
Log_MEAVold	.712*** (.121)	
Log_MEAVnew		.719*** (.032)
Trend	-.062*** (.020)	-.037** (.018)
Constant	-3.138*** (1.093)	-3.499 (.202)
Observations	64	64
Adj. R2	.474	.496

Similar magnitude of
coefficients, but
slightly higher R2

The model passes all
diagnostic tests



GD1 data

All variables
in natural
logarithm

All
specifications
include a
time trend

Results are
preliminary

	<i>Totex</i>	<i>Opex</i>	<i>CostPool1</i>	<i>Emergency</i>	<i>Maintenance</i>	<i>Repairs</i>	<i>Work Management</i>	<i>Opex + Repex Oth. Serv.</i>	<i>CostPool2</i>	<i>CostPool3</i>	<i>Repex</i>	<i>Connections</i>	<i>Reinforcement</i>	<i>Capex</i>	<i>CostPool4</i>
Totex	1														
Opex	0.83	1													
CostPool1	0.69	0.93	1												
Emergency	0.82	0.9	0.85	1											
Maintenance	0.74	0.81	0.61	0.6	1										
Repairs	0.76	0.78	0.62	0.83	0.55	1									
Work Management	0.56	0.84	0.91	0.72	0.57	0.48	1								
Opex + Repex Oth. Serv.	0.84	0.99	0.92	0.92	0.79	0.82	0.82	1							
CostPool2	0.79	0.79	0.67	0.89	0.5	0.96	0.52	0.84	1						
CostPool3	0.88	0.92	0.74	0.89	0.82	0.9	0.62	0.94	0.91	1					
Repex	0.94	0.68	0.54	0.75	0.58	0.74	0.39	0.71	0.78	0.8	1				
Connections	0.6	0.33	0.32	0.4	0.28	0.21	0.32	0.37	0.36	0.38	0.57	1			
Reinforcement	0.4	0.13	0.01	0.15	0.28	0.15	-0.01	0.14	0.17	0.25	0.34	0.57	1		
Capex	0.7	0.28	0.13	0.29	0.44	0.28	0.04	0.29	0.29	0.41	0.62	0.62	0.64	1	
CostPool4	0.97	0.89	0.77	0.84	0.8	0.75	0.65	0.9	0.78	0.91	0.83	0.56	0.36	0.67	1

Green shading – correlation greater than 0.5

Aside from capex mains and connection, the expenditure categories are largely correlated

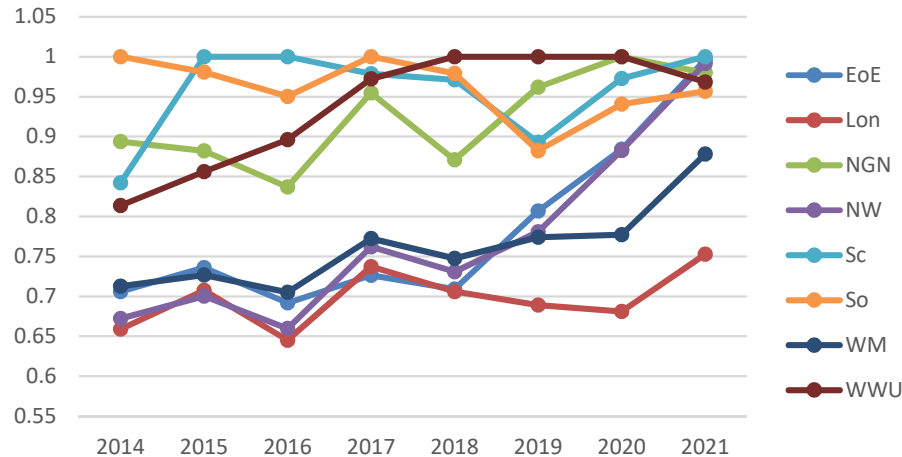
Correlations indicate that mains and connections are likely to move in a different way to other expenditure
This may support the use of a opex + base capex pooled model

	Opex	Capex	Repex
Opex CSV	.940*** (.001)		
Capex CSV		.801*** (.000)	
Repex_Wkl			.810*** (.000)
Trend	-.008 (.293)	.075** (.024)	.006 (.341)
Constant	-4.666** (.023)	-2.713** (.025)	.951** (.043)
Observations	64	64	64
R2	.774	.558	.878

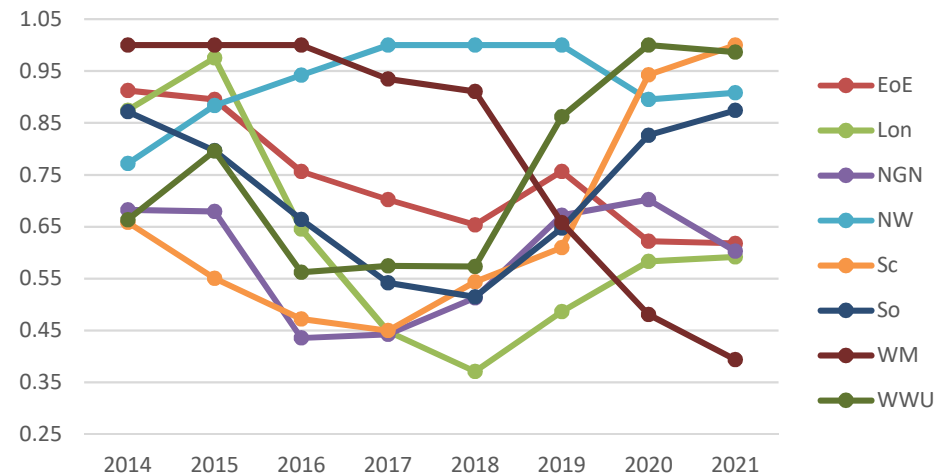
Coefficients as
expected

No clear trend for
opex and repex,
increasing trend
for capex

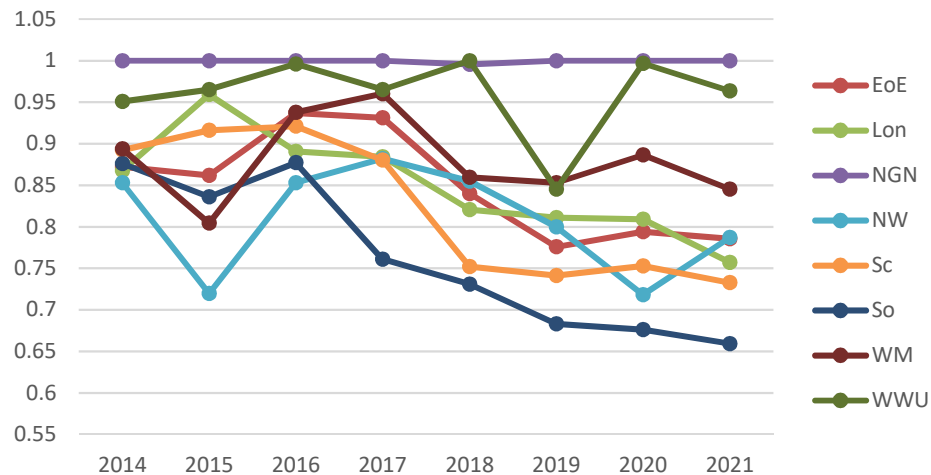
Opex Efficiency



Capex Efficiency



Repex Efficiency



Measure of technical efficiency
reflecting distance from the frontier

As expected, high capex efficiency
variation (no smoothing)

Decreasing efficiency variation for
opex, increasing for repex

Log_CostPool	Pool1 (Asset Mgt, Oper Mgt, BSC, ODAs)	Pool2 (Emerg., Repairs, Repex Other Services)	Pool3 (Emerg., Repairs, Mainten., Repex Other Services)	Pool4 (Opex, Capex excl. Reinf., Repex Other Services)
Log_MEAV	.725** (.039)	.860** (.047)		.771*** (.000)
Log_Maint Meav			1.063*** (.000)	
Trend	-.020 (.207)	-.013* (.056)	.310* (.055)	.010 (.306)
Constant	-2.849 (.319)	-4.406 (.233)	-6.176*** (.003)	-2.210* (.077)
Observations	64	64	64	64
Adj. R2	.447	.378	.761	.749

Models with MEAV
as a cost driver
seem to perform
better for higher
levels of cost
aggregation

Capturing well
overall scale of
operation

Log_CostPool	Pool2 (Emerg., Repairs, Repex Other Services)	Pool4 (Opex, Capex excl. Reinf., Repex Other Services)	Pool4 (Opex, Capex excl. Reinf., Repex Other Services)
CSV (cust., PREs, ext. cond. rep.)	1.156*** (.000)		
CSV (cust., throughput, net. length)		.707*** (.002)	
Totex CSV			.742*** (.000)
Trend	-.006 (.465)	.007 (.448)	.002 (7.83)
Constant	-10.276*** (.000)	-3.576** (.077)	-.324 (.633)
Observations	64	64	64
Adj. R2	.829	.701	.783

Not much difference
between a CSV of
scale drivers only and
a CSV combining
scale and workload
drivers

Test	Description/ issue [Refer to Smith (2019) 'Note for Ofgem on Diagnostic Tests in Efficiency Benchmarking Studies', June, for more details]
Statistical significance of the coefficients	This test is asking whether we can be confident that there is a relationship between the explanatory variable and cost – or more formally can we (statistically) reject the proposition that there is no relationship (i.e. that the coefficient is zero). Statistical significance is generally measured at the 1%, 5%, or 10% levels, though there is no set rule on which is the most appropriate.
RESET test	A test for whether there is some non-linear relationship in the model that is not been captured. In the cost modelling literature this is normally dealt with directly by considering a translog specification which captures these non-linearities directly. Reported as p-value; values above 0.05 indicates that we cannot reject the null hypothesis that the model is correctly specified.
Heteroscedasticity	With respect to heteroscedasticity / autocorrelation, violations of the assumptions in OLS impact only on the standard errors (that is the standard errors are biased and this impacts on our ability to gauge whether a variable is statistically significant or not) and do not cause the estimates themselves to be biased. We use robust standard errors when assessing statistical significance.
Normality	The impact of nonnormality only has implications for the ability to use finite sample inference – that is, making judgements about the statistical significance of the parameters in finite samples. The breakdown of normality would not generally be seen to be a problem in most applications, though it could cause issues for very small sample sizes. However, the models used by Ofgem incorporate several firms over several years and thus we would not expect a breakdown of normality to be a serious problem. Looking for values above 0.1.
Pooling	For each model, we test whether the true coefficients of the pooled OLS model are significantly different from the true coefficients of the same model run on each individual cross-section of the data (i.e. 2011/12, 2012/13, etc.). If this test fails (p-value of less than 0.01) this provides evidence that panel data analysis may not be appropriate.

Diagnostic Tests on Estimated Models

Test/Model	Opex	Capex	Repex	Pool1	Pool2	Pool3	Pool4	
Stat. Signif.								
RESET								Passed
Heterosced.								Some Passed
Normality								Not Passed
Pooling								
Adj R2	.774	.558	.878	.447	.378 .829	.761	.749 .701 .783	

Which value for the R squared would be **sufficiently high**?

Generally the models are **performing well**, although not all of the new ones

The majority of models **do not pass the RESET tests**
We will test whether squared terms affect the results

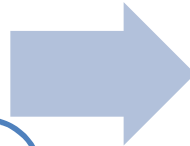
Check how any new expenditure pooled models fit within the **broader modelling**

Any additional models to be tested?
Ongoing work/thinking that could be shared?

Phase 1

- Identify preferred models based on:
- Predictive power (adjusted R-squares)
- Statistical robustness
- Economic/Engineering rationale
- Are the results consistent with the rest of the price control?

- **Pool 2-4, Opex, Repex**



Phase 2

- Put preferred models through robustness testing:
- Removal of years/companies from panel
- Random effects
- Within-sample forecasting (when appropriate)

- **Pool 1, Pool4, Capex**

Future meetings



- Model testing and refinement
- Regional factors
- RPEs and ongoing efficiency

Any other business



Our core purpose is to ensure that all consumers can get good value and service from the energy market. In support of this we favour market solutions where practical, incentive regulation for monopolies and an approach that seeks to enable innovation and beneficial change whilst protecting consumers.

We will ensure that Ofgem will operate as an efficient organisation, driven by skilled and empowered staff, that will act quickly, predictably and effectively in the consumer interest, based on independent and transparent insight into consumers' experiences and the operation of energy systems and markets.