Homes &
Communities
Agency

DELIVERING BETTER VALUE FOR MONEY: UNDERSTANDING DIFFERENCES IN UNIT COSTS – TECHNICAL REGRESSION REPORT

June 2016

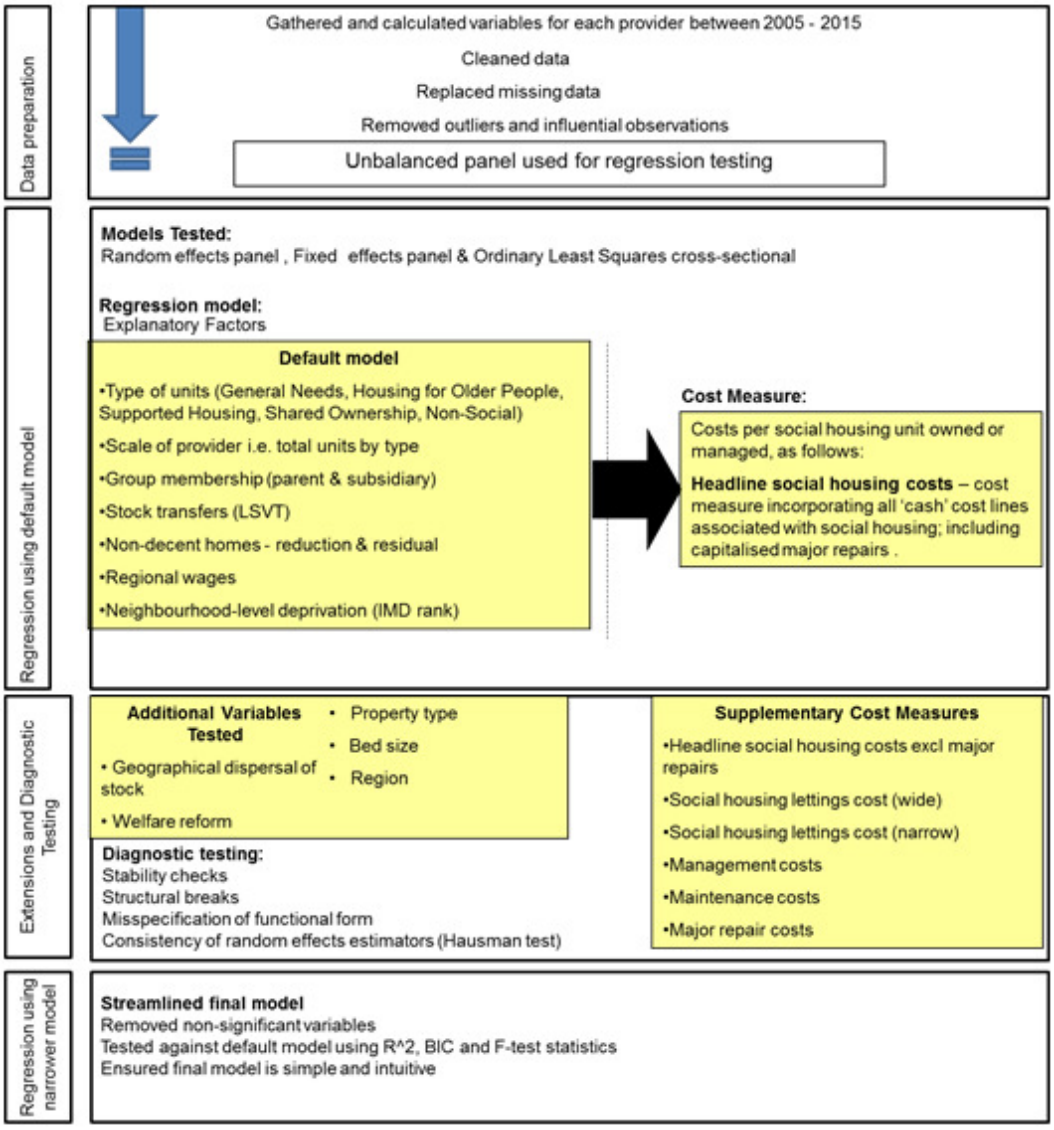
Contents

1. Introduction

1. This technical report sets out the detail of regression analysis undertaken by HCA Regulation on unit costs of private registered providers of social housing ('providers') in England. This report is intended for a technical audience and presents detail of data definitions and the statistical process. Key messages drawn from this analysis are set out in the accompanying unit cost summary report. The analysis set out in this report updates similar work published in 2012[1].
2. This analysis aims to understand how unit costs of providers are related to a range of measured explanatory factors. Regression analysis is the standard statistical method to estimate such relationships. This is a powerful analysis, drawing on 11 years' data for all providers in England with more than 1,000 social housing units. However, there are limits to the power of the work especially as there is a lack of data for many important factors that affect costs – for example, scope and quality of services or condition of stock. The aim of the analysis is to provide robust evidence to support debate within the sector, and to better understand unit costs.
3. The approach to analysis, summarised in the figure overleaf, is explained in this report in three main sections:
    - Unit cost data and explanatory variables – sources, definitions and descriptive statistics for key variables.
    - Regression analysis: headline results – results for the streamlined final model, which explains the relationship between unit costs and seven explanatory variables.
    - Regression analysis: additional testing – results from the default model, which explains cost variation using 19 explanatory variables and was used as the initial basis for modelling, plus testing of other explanatory factors and cost measures.
4. The Annex sets out detail of diagnostic testing, definitions of explanatory and additional variables.

[1] HCA 2012 unit cost analysis can be accessed via this link: http://udc.homesandcommunities.co.uk/news/hca-publishes-analysis-unit-costs-drivers-providers

Figure 1: Schematic outline of statistical process



2. Unit cost data and explanatory variables

5. This section describes the data used to run the regression analysis. This includes definition of unit costs, explanatory variables and the process of cleaning data.

Data overview

6. Analysis is based on a panel dataset of cost and contextual information for provider entities with more than 1,000 units for 2005 to 2015 inclusive. Data is complete for 3,838 observations between 2005 and 2015, which represents approximately 349 registered providers per annum on average. The number of providers varies year on year due to a number of factors including registrations of new providers, transfers from local authorities and mergers of existing providers. This dataset includes the vast majority of provider entities with at least 1,000 units (owned or managed) each year[2], which account for at least 95% of all the social housing units within the sector per annum.

7. Analysis is primarily based on data submitted by providers and already published by the regulator: for cost data this is the Global Accounts data, and for most explanatory factors it is the Statistical Data Return (SDR, previously the Regulatory and Statistical Return survey (RSR)). For a small number of variables this is supplemented with other available data including: the Continuous Recording of Lettings and Sales in Social Housing in England (CORE), the Annual Survey of Hours and Earnings (ASHE) for regional wages and the Index of Multiple Deprivation (IMD) for neighbourhood deprivation.

Data on unit costs

8. The **data used for unit costs is derived from the electronic accounts data returns** database from the year ending March 2005 to year ending March 2015. This is the same database the regulator uses to develop the annual Global Accounts report. All the cost measures were adjusted to 2015 prices, using the Consumer Price Index (CPI) for March each year.

Headline social housing costs

9. Analysis is based on 'headline social housing costs per unit' which is composed of several cost lines reported in the Global Accounts. As defined overleaf, it is made up of the main components of management and service charges, maintenance, major repairs and other social housing costs. It is sufficiently broad so as not to be affected by different cost apportionment approaches that can cause significant variance in narrower cost lines reported by provider. It aims to estimate 'proxy cash' costs of running social housing per annum. It excludes costs of sales (e.g. capital costs of building new social housing) and notional expenditure items – depreciation, impairment and bad debts – that can be sensitive to different accounting approaches and presentation of data within financial statements, which can lead to wide divergence in reported costs. It includes capitalised as well as expensed major repairs.

[2] A very small number of providers with more than 1,000 units are excluded from the dataset because cost data is known to be anomalous – for example, new providers may have zero costs or only have costs for a period for less than a year.

Table 1 Headline social housing cost

Section	FVA cost lines	**Headline social housing costs**	Components				
			Management	Service charge costs	Maintenance	Major repairs	Other social housing costs
Expenditure on social housing lettings	Management	Y	Y				
	Service Charge Cost	Y		Y			
	Care & Support Costs	2005-12		2005-12			
	Routine Maintenance	Y			Y		
	Planned Maintenance	Y			Y		
	Major Repairs Expenditure	Y				Y	
	Bad Debts						
	Lease Charges	2005-12					2005-12
	Depreciation of housing properties						
	Impairment of housing properties						
	Other Costs	Y					Y
Capitalised major repairs	Capitalised major repairs and re-improvements	Y				Y	
Exp. on other social housing activities	Other social housing activities - Other (expenditure)	Y					Y
	Other social housing activities - Charges for Support Services (expenditure)	Y					Y
	First tranche shared ownership sales	2008-					
Expenditure on non-social housing activities							

Cost definition considerations

10. There is not a single perfect measure of costs that can be derived from Global Accounts data. There are a range of options, and the pros and cons of each must be weighed against objectives of any particular exercises and the accounting approaches by providers. Headline social housing costs have been defined following careful consideration of alternative measures in light of the objectives of this particular analysis. This definition will be kept under review by the regulator.

11. One consideration is the treatment of service charge costs. Headine unit costs includes gross service charge costs (i.e. not discounting service charge income that may cover these costs). This partly reflects a practical consideration in that providers treat service costs differently in preparing their accounts – for example what some providers report as service charge costs, others may record as 'other' social housing costs. The definition is desined to allow a comparison of overall costs across providers that is not unduly affected by these reporting differences. More fundamentally, the aim of the headline measure is to focus on variation of a relatively simple unit costs, leaving aside considerations the level of net income or the scope or quality of services associated with costs. More in-depth analysis of VfM at a provider level may legitimately include these considerations.

12. Another consideration was bad debts and void costs, which are excluded from the headine unit cost measure. On the one hand, it is recognised that increased expenditure on management activities, for example on rent collection or lettings functions, could have an positive offsetting effect on bad debts. However the definition aims to focus on simple cost variation across providers, leaving aside questions of improved outcomes or performance that may be generated by additional expenditure. Bad debts and void costs essentially reflect the difference between gross and net rental income. Further, for bad debt costs different accounting approaches can lead to variation in or 'lumpy' bad debt charge figures that reduce comparability between providers. For these reasons the regulator has excluded these cost lines from the headline measure.

Time series considerations

13. Global Accounts data has changed between 2005 and 2015, due to amendments to data returns and accounting practices. Drawing from examination of individual accounts and internal peer review and discussion, costs measures were defined to allow greatest consistency over time. Definitions take account of the following changes:

    - Care & support costs, measured separately in accounts returns up to 2012 are assumed to be largely reflected in management and service charges cost lines after this date. Similarly lease charges are assumed to have been largely picked up in the other costs line.
    - The introduction of component accounting in 2011/12 means that some costs formally attributed to maintenance (planned and routine) and major repairs may have moved into capitalised major repairs and re-improvements.
    - Pre 2008 first tranche shared ownership sales were not split out from other social housing activities – other costs meaning that they are likely to be included within the headline measure from 2005 to 2007.

14. The accompanying summary report sets out two time series - management & service charge costs and maintenance & major repairs costs per unit at a sector level – which are both broadly robust to the changes outlined above. However, due to the effect of accounting for first tranche shared ownership sales, caution is required in interpreting the time series for the headline social housing costs per unit and for this reason it is not presented. Notwithstanding these issues, headline social housing costs per unit remains a valid measure for examining in-year variation in unit costs across providers and hence the regression analysis in this technical report.

Total social housing units

15. Unit costs for the purposes of regression analysis were generated using total social housing stock owned or managed, as reported in the SDR/RSR. This includes general needs, supported housing, housing for older people, care homes classified as social housing, general needs affordable rent units as well as low cost home ownership units where the tenant has acquired less than 100% of the equity. An average in-year figure was derived using the average from the end of current and previous years, to mitigate the impact of any large changes in stock. Intermediate rent units were not included as they were not defined as social housing pre 2012 and were therefore not included in previous iterations of the analysis and supported housing affordable rent units were not included as it is not possible to split out the housing for older people units. Due to the limited number of units in each of the tenures their exclusion was assessed to have an immaterial effect on results.

Supplementary cost measures

16. Along with the headline social housing costs, a number of secondary narrower cost measures were included in diagnostic testing. They were used to test the stability of the model to the removal of certain cost lines and to generate additional information on how cost drivers influenced various components of headline social housing costs. These supplementary cost measures were as follows, with definitions provided in Annex C:

    - Headline social housing costs excluding major repairs
    - Social housing lettings cost (wide)
    - Social housing lettings cost (narrow)
    - Management costs
    - Maintenance costs
    - Major repair costs

17. The electronic accounts data splits the majority of the cost lines by tenure, allowing the formation of general needs specific cost measures for social housing lettings cost (wide) and social housing lettings cost (narrow). These were tested to assess the stability of (non-supported housing based) explanatory variables to the removal of the high costs associated with supported housing.

Data on explanatory variables

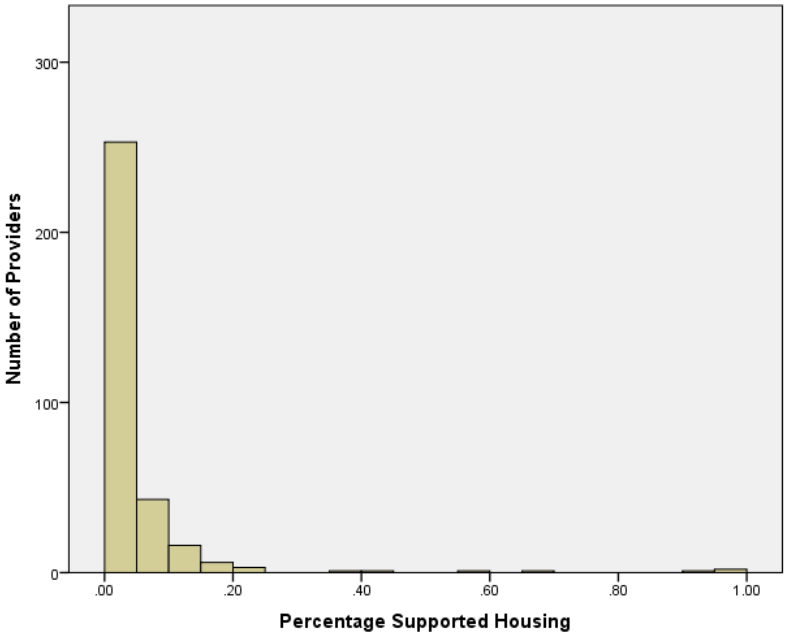
18. This section describes the key explanatory variables used in the final model. Unless otherwise stated, data is drawn from SDR/RSR submissions by providers to the regulator. A full variable definition list is included in Annex B. Descriptive data refers to providers with more than 1,000 units that complete electronic accounts returns and are hence included in this analysis. For a full profile of the provider sector, including the large number of providers with less than 1,000 units, please refer to the HCA Statistical Data Return (SDR) publication.

Supported housing (% total)

19. This measures supported housing (SH) units (excluding housing for older people) owned and managed, averaged over the current and previous year, as a proportion of average total social housing stock owned and managed in the current and previous year. The figure excludes a small number of Affordable Rent SH stock as it is not possible to separate it from Affordable Rent HOP stock within the Statistical Data Return.

20. There are a limited number of specialist supported housing provider with more than 1,000 units. In 2015, out of the 328 providers with complete data (and over 1,000 units) only seven had greater than 30% supported housing stock and for 77% of providers the figure was below 5%. However, there is no clear dividing line between 'supported housing' providers and others – 74% of providers own or manage at least one supported housing unit.

Figure 2: Number of providers by percentage of supported housing (2014/15)



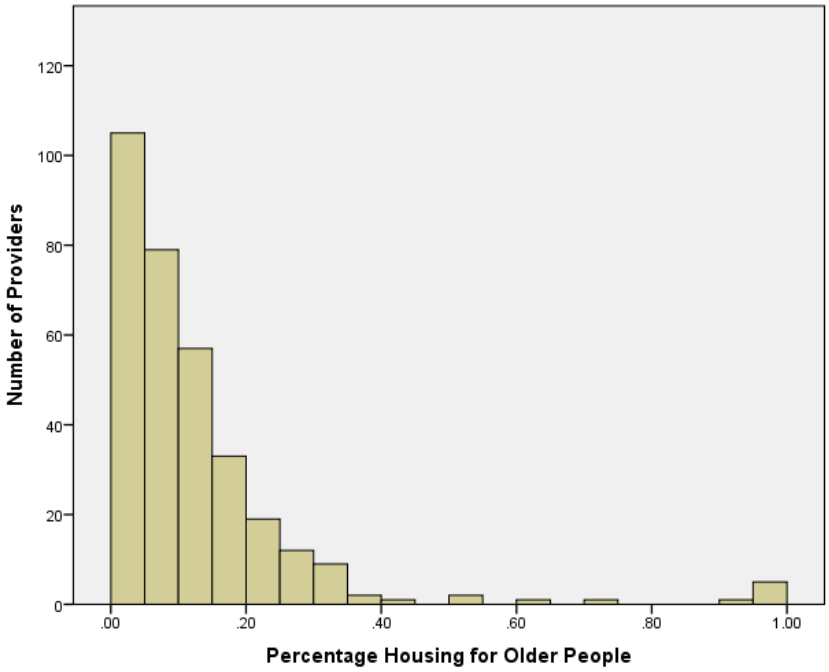
Housing for older people (% total)

21. This measures housing units for older people (HOP) owned and managed, averaged over the current and previous year, as a proportion of average total social housing stock owned and managed in the current and previous year. The figure includes care homes units recorded as social housing. It excludes

a small number of affordable rent HOP stock as it is not possible to separate it from affordable rent SH stock within the Statistical Data Return. .

22. There are more specialist HOP providers than specialist support housing providers; they are however still a minority. Out of the 328 providers in 2015 with complete data (and over a 1,000 units) only 23 had over 30% housing for older people stock. The median in the sector is 8.4% housing for older people stock and the mean is 12.7%, with 91% of providers holding some HOP stock.

Figure 3: Number of providers by percentage of housing for older people (2014/15)



% Reduction in non-decent stock

23. This measures the reduction in stock not meeting the Decent Homes Standard compared to the figure for the previous year, as a proportion of total social housing stock. This is to measure a key driver of major repairs expenditure, especially in earlier years of the sample. All recorded increases in non-decent stock owned by a provider during a year, due to transfers of stock from local authorities for example, are excluded.

24. The significance of the variable has reduced over the period of the analysis due to the decline in the number of units not meeting the Decent Homes Standard. The total number of units not meeting the Decent Homes Standard was 340,000 in 2005 reducing to 13,000 in 2015 for providers with more than 1,000 units.

Stock transfers (LSVTs)

25. A provider was defined as a stock transfer (or Large Scale Voluntary Transfer, LSVT) if over 50% of their stock was obtained through transfers, with the age of the LSVT being determined by the date of the largest transfer. Three time-dependent dummy variables categorise the maturity of LSVT. The categories are based on whether the provider transferred six or fewer years ago, seven to twelve years ago, or more than twelve years ago.

26. The LSVT sub-sector has matured significantly over the period of the analysis, driven by low transfer volumes in the last five years. In 2005, 23% (85) of providers were LSVT providers in their first six years after transfer; this had reduced to 3% (10) by 2015.

Table 2: LSVT breakdown by year (no. providers, with only 1,000+ units only))

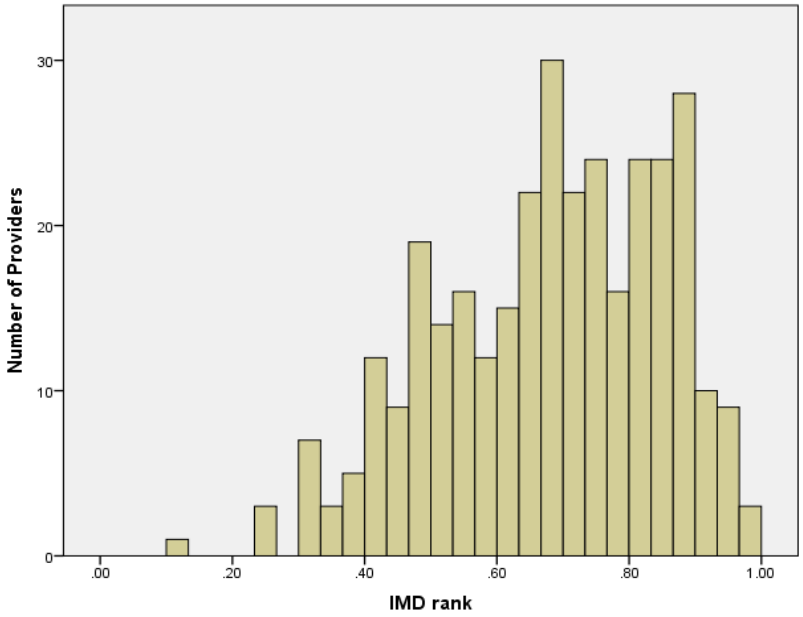
	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
LSVT <6 years	85	75	60	67	60	60	54	46	37	17	10
LSVT 7-12 years	57	65	73	67	82	78	73	63	60	72	63
LSVT >12 years	21	29	36	45	47	52	58	67	79	85	98
Total Providers	366	365	359	372	372	365	354	334	335	334	328

Neighbourhood deprivation (IMD)

27. The deprivation variable indicates the degree to which a provider operates in deprived neighbourhoods (Lower Super Output Areas, LSOAs). The neighbourhood stock profile of the provider is estimated using CORE data on general needs lettings by LSOA for each provider. This is combined with the Government's published Index of Multiple Deprivation (IMD) for each LSOA to generate a weighted IMD for each provider's general needs stock. The LSOA's percentile rank in the 2010 IMD was used; the variable was then multiplied by the provider's share of general needs. IMD measures levels of local deprivation across seven domains: income, employment, education, health, crime, housing and the living environment. This measure was included in the regression to pick up the potential effect of additional lettings, repairs, advisory services, anti-social behaviour or regeneration activities sometimes associated with operating in relatively deprived areas.

28. The sector, as a whole, is generally concentrated in relatively deprived areas in England. The distribution of IMD ranks exhibits a negative skew, with the median provider operating in neighbourhoods that are in the 31% most deprived in England (median of 0.69).

Figure 4: Number of providers by deprivation rank (2014/15)



Regional wage index

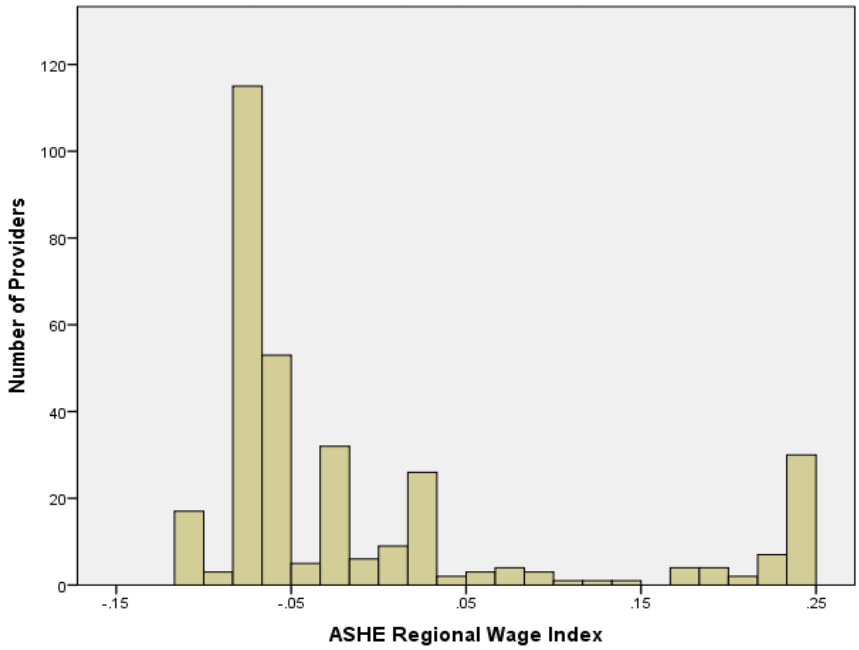
29. A regional index was calculated for each provider. This uses published Annual Survey of Hours and Earnings data (2006-2013), using the mean salary for full-time employees with a split of two thirds administration and one third construction workers in each region. The provider level wage index was then calculated by multiplying the provider's share of stock in each region by the index and subtracting 1 so that the England average was 0.00.

30. The vast majority of providers operate mainly in one region; with 93.3% of providers having over 50% of the stock that they own or manage in a single region. The index and the accompanying graph both have a positive skew caused by the fact that wages are significantly higher in London than in any other region. It shows that 68% of providers are predominantly operating in regions where the wage index is lower than the average for England.

Table 3: Percentage of providers by region and associated wage index (2014/2015)

	ASHE Regional Wage Index	% of providers
North East	**-0.11**	5.5%
North West	**-0.07**	19.4%
Yorkshire and The Humber	**-0.06**	6.1%
East Midlands	**-0.08**	5.5%
West Midlands	**-0.07**	11.8%
East of England	**-0.02**	10.3%
London	**0.24**	14.8%
South East	**0.03**	10.3%
South West	**-0.06**	9.7%
Mixed	**N/A**	6.7%

Figure 5: Number of providers by regional wage index (2014/15)



Missing data

31. Where possible all data was included in the analysis and there were few instances of missing data. The largest source of missing data was from the CORE dataset, which is used in the formation of the IMD variable. In cases where provider's CORE logs were available for at least one of the years, the average IMD score of the available years was used to replace the missing values. Analysis demonstrated this is a reasonable approximation in that most providers' lettings patterns and deprivation remained relatively constant over time. In the few cases where no CORE data was available, for any of the years a provider was in the sample, the missing data was replaced by the global average. This only affected a relatively small number of data points[3] and was deemed preferable to removing observations from the regression analysis.

[3] In the 2015 dataset 96% of providers had available CORE data.

3. Regression analysis – headline results

32. This section sets out the results of regression analysis of unit costs using the streamlined final model (seven explanatory variables).

Overview of regression

33. Results derive from standard Ordinary Least Squares (OLS) regression analysis for each year 2005 to 2015 inclusive. The summary results are drawn from the 'streamlined model' which shows the relationship between unit costs and the seven most important explanatory variables. The subsequent section shows results for the broader default model, which was used for initial testing, and provides a discussion of those variables where there was no clear evidence of any relationship with unit costs.

34. Random effects and fixed effects models were tested alongside OLS models. These models were used as the basis for results present in previous regression analysis published in 2012 – however, they were found not to be suitable for use in this round of analysis.

35. Analysis is based on an unbalanced panel, which means that all providers are potentially included each year. The alternative is a balanced panel in which only providers with data for all eleven years are included. Outliers and influential observations which can disproportionately skew regression analysis were removed. These were identified using standard thresholds for studentised and standardised residuals, Cook's distances and Leverage metrics. This resulted in the removal of 415 observations (10.8%) over the 11 years, meaning that analysis was ultimately based on a panel with 3,423 observations – on average 311 providers per annum.

36. In order to get assurance on the robustness of results, default and streamlined regressions were also run for models including outliers and influential variables and for the unbalanced as well as balanced panel. Results stated in this section were robust to these alternative forms.

Headline results

37. Headline results are based on the final streamlined OLS model. All quoted figures are based on the mean coefficients, over the latest five years' regressions (2011-2015) that are significant at 90% confidence levels or above.

Baseline unit costs (general needs)

38. The baseline unit cost for a provider is £3,300. This is based on a traditional provider with 100% general needs stock, with no non-decent homes, operating in an area with average deprivation and wages. This is based on the regression intercept (£2,700) plus the effect of average neighbourhood deprivation.

Support housing

39. Each unit of supported housing is associated with costs of £10,800 above general needs properties. This is likely to be associated with the high support costs and a broader scope of activities undertaken by organisations with a specialised focus. The precise estimate of associated costs is sensitive to the inclusion or removal of more specialised supported housing organisations – with model estimates varying from £8,400 to £14,000 per unit. There is likely to be considerable diversity with the cost associated with each supported housing unit in the sector, for example by the client group and the level of service supplied.

Housing for older people

40. Each unit is associated with cost of £1,800 above general needs units. This average is likely to include a range of service levels and costs, from very intensive facilities offering services akin to care homes to properties where costs may be broadly similar to other general needs properties. As with supported housing the coefficient is sensitive to the inclusion and/or exclusion of a small number of specialised providers (although not to the same extent).

Table 4: Average coefficients from 2011-2015 (using only coefficients are significant at >=90% confidence)

Year	Coefficient
Intercept	2.70
% Housing for older people	1.76
% Supported housing*	10.81
% Reduction non-decent	8.79
LSVT <7 years (DV)**	1.46
LSVT 7-12 years (DV)	0.33
Regional wage index (Combined)*	5.40
Index of Multiple Deprivation (% rank)	1.07
R-squared	0.52
Adjusted R-squared	0.51

* most influential factors using regression with standardised variables
** standardised regression highlighted as influential factor – however there are potential issues using dummy variables (DVs) in standardised regression models so results must be interpreted with caution

Regional wage effects

41. Cost differences in providers operating in different regions broadly follow the differences in underlying regional wages, once all factors are taken into

account. The average coefficient from the streamlined regression means that providers in London have average costs of £1,900 per unit above those in the North East. This is equivalent to 139% of the estimated ASHE wage differential between London and the North East. However, this coefficient should be interpreted very carefully as it is heavily influenced by the effect of supported housing.

42. The broader default model estimates separate wage effects for general needs and supported housing, with results in Table 6. The wage effect for supported housing is significantly above those for general needs - the most plausible explanation for this is that supported housing units in higher cost areas, particularly in London, are associated with more intensive support on average and the wage index is picking up this unmeasured factor. The regional wage index for general needs reduces the ratio to the estimated wage differential from 139% to 119% of the ASHE regional wage differential. This is in line with the equivalent figures for general needs wage effects from the 2012 analysis (116%). However, a single composite wage index is included in the streamlined model for simplicity and in particular so higher supported housing costs are picked up in the supported housing variable rather than a secondary wage measure.

43. There are at least two possible explanations for more than 100% of the wage differential feeding through into costs. First, the wage index constructed may not adequately reflect the differences in registered provider salaries between regions. For example, differences in executive pay between regions may be more marked than for general administrative or construction salaries. Second, there may be other costs, for example office rental, where cost differences are more marked between regions. Alternatively, higher social housing rents – correlated with regional wages, and not included in the analysis – may permit higher costs.

Stock transfers

44. Stock transfer providers have average headline costs of £1,500 per unit higher than traditional providers in years 1-6 post-transfer. This gap narrows to £300 per unit for providers in years 7-11, and disappears after 12 years. There are also significant differences between the cost profiles of LSVT and traditional providers at a more granular level. LSVTs have on average significantly higher maintenance and major repairs costs and lower management and service charge costs than the traditional sub-sector.

45. There are two main reasons for cost differences. Firstly, the higher maintenance and major repairs cost for many LSVT providers is in all probability attributable to high capital investment, regeneration and community engagement programmes typically undertaken in the first few years after transfer. The higher management and service charges of traditional providers can be explained, at least to some extent, by the greater proportion of supported housing and housing for older people units held.

Neighbourhood deprivation

46. Providers operating in neighbourhoods ranked in the most 1% most deprived according to the Index of Multiple Deprivation have costs on average £500 per unit higher than providers operating in an area with median levels of deprivation for England. This could be associated with a range of factors, including more extensive regeneration and community initiatives, higher voids and turnover, and potentially greater crime and anti-social behaviour.

Decent homes

47. Reduction in units not meeting the Decent Homes Standard by one unit is on average associated with costs of £8,800 per unit. The limited residual non-decent stock in the sector means this is only an important factor for a small minority of providers.

Standardised regression results

48. A standardised regression model was run with using normalised explanatory and unit cost variables. The results of which showed that the most powerful explanatory variables in terms of explaining unit cost variation were the supported housing, the regional wage index and LSVTs (< 7 years) variables. Variation in supported housing stock owned or managed accounted for example, accounted for four to five times more unit cost variation than the weakest explanatory variable (reduction in non-decent homes).

Explanatory power of the streamlined model (R^2)

49. The seven variables in the streamlined model can account for around half (R^2 of 52%) of the variation in unit costs on average over the last five years. This is only slightly lower than for the 19 variables in the default model (55%).

Table 5: Final streamlined model results – Coefficients show change in headline social housing cost per unit (£000s) associated with a one unit change in the each explanatory variable (holding other explanatory factors constant)

Year	2015	2014	2013	2012	2011	2010	2009	2008	2007	2006	2005
Intercept	3.01***	2.84***	2.45***	2.49***	2.73***	2.96***	2.54***	2.82***	2.17***	2.98***	2.49***
% Housing for older people	2.10***	1.69***	1.36***	1.99***	1.66***	1.02*	1.75***	1.21*	2.42***	1.21**	1.54***
% Supported housing	9.79***	10.32***	10.98***	10.43***	12.56***	11.16***	14.03***	11.37***	10.36***	10.64***	8.84***
% Reduction non-decent	-4.89*	13.47**	7.62***	9.38***	4.71***	5.22***	6.39***	2.94***	3.52***	3.47***	2.61***
LSVT <7 years (DV)	1.81***	1.00***	1.25***	1.42***	1.83***	1.92***	1.92***	1.62***	1.23***	1.21***	1.17***
LSVT 7-12 years (DV)	0.39***	0.47***	0.25**	0.17*	0.21**	0.33***	0.41***	0.18*	-0.04	0.02	0.06
Regional wage index (Combined)	5.88***	5.05***	5.17***	5.96***	4.95***	6.37***	6.23***	6.34***	5.46***	4.61***	5.62***
Index of Multiple Deprivation (% rank)	0.47	0.59**	1.39***	1.23***	0.57*	0.60*	0.98***	1.00***	1.87***	0.55	1.01***
N (total observations)	291	292	300	301	307	318	326	330	315	322	321
Mean of cost measure	3.85	3.79	3.94	3.99	4.00	4.26	4.32	4.33	4.19	4.25	4.04
Standard deviation of costs	1.15	1.11	1.28	1.38	1.41	1.40	1.55	1.49	1.38	1.39	1.31
Standard error of the regression	0.82	0.82	0.88	0.94	0.94	0.98	1.03	1.08	1.05	1.06	1.01
R-squared	0.49	0.47	0.54	0.55	0.56	0.53	0.57	0.49	0.44	0.43	0.41
Adjusted R-squared	0.48	0.45	0.53	0.54	0.55	0.51	0.56	0.48	0.42	0.42	0.40
	Unless indicated otherwise, figures presented in the main body of the table are the regression coefficients. DV indicates dummy variable. *** Significant at 95% confidence level ** Significant at 90% *Significant at 80% (standard t-tests)										

4. Regression analysis – additional testing

50. This section sets out the results of wider regression testing, above and beyond the streamlined final model presented in the previous section. It includes results from the broader default model (19 explanatory variables).

Explanatory factors with weak relation with costs

Economies of scale

51. There was no significant evidence of a clear relationship between scale of a provider and lower costs (once factors are controlled for). The presence of linear relationships between costs and the main stock types (general needs, supported housing, housing for older people and shared ownership) was tested with no consistent significant results. Different functional forms including squared, cubed and logged terms were also tested with results in line with those seen for the linear forms. The variables were also tested across the narrower cost lines: management, maintenance and major repairs with no significant evidence of a relationship between scale and cost found. This conclusion is drawn from extensive testing and is consistent with the finding from equivalent work in 2012.

52. One of the limits of the model is that important cost drivers - such as service levels or stock condition - cannot be measured. Therefore, the lack of evidence for economies of scale does not necessarily mean that they cannot be achieved or they are not being achieved by providers. It is possible that economies of scale are achievable, but may not be being delivered in practice. However, it is also theoretically possible that efficiency gains are reflected in higher service levels or better stock condition, which weakens the statistical link between unit costs and size. Cost data alone cannot show whether or not larger providers do deliver improved service in this way.

Non-social units

53. The non-social explanatory variable was shown to be significant from 2012 onwards. Given that non-social activity is stripped out of the dependent variable (both numerator, in terms of costs, and denominator, in terms of stock), the variable is not expected to exert an influence. Further investigation has highlighted two probable causes. Firstly, changes in data definitions when the data return moved from the RSR to the SDR in 2012 led to a significant change in non-social stock figures, in particular including large amounts of certain types of shared ownership for a small number of providers. The non-social variable was significantly positively correlated with the regional wage variable in each of the last five years and it is therefore probable that the variable is picking up some of the cost of operating in higher cost areas and therefore the results do not provide a reliable estimate.

Shared ownership units

54. There was limited evidence that the proportion of shared ownership stock held by a provider affects costs with significant results in only one of the eleven years. As shared ownership units were included in the total stock figure, this signifies there is no statistical evidence that the cost of shared ownership units differ from those of general needs units.

Group structures

55. There was no clear evidence of relationships between group membership and unit cost.

Table 6: **Headline social housing cost (£000s) – Default OLS model**

Year	2015	2014	2013	2012	2011	2010	2009	2008	2007	2006	2005
Intercept	2.89***	2.94***	2.16***	2.67***	2.97***	3.18***	2.53***	1.99***	2.13***	2.64***	2.09***
% Housing for older people	2.23**	1.57**	1.49**	0.89	0.70	1.96**	2.36***	2.84***	2.91***	2.14**	3.19**
% Supported housing	10.42***	11.22***	13.04***	12.52***	12.12***	10.44***	14.37***	11.92***	10.06***	10.02***	8.41***
% Shared ownership	0.13	-0.06	-0.07	0.40	2.21	0.69	1.48	5.27***	2.95	3.02	2.19
% Non-social	2.08***	2.07***	3.56***	2.99***	2.23**	0.98	0.40	2.24***	1.66	-0.62	-0.03
% Reduction non-decent	-3.95	9.78*	7.27***	5.30***	4.53***	4.43***	5.55***	3.38***	3.93***	3.47***	2.54***
% Residual non-decent	6.03	-5.75*	-1.90	6.09***	3.21***	2.18**	2.43***	-0.08	0.53	0.64	1.11***
LSVT <7 years (DV)	1.48***	0.91***	1.39***	1.55***	1.64***	1.61***	1.49***	1.55***	0.98***	1.07***	0.9***
LSVT 7-12 years (DV)	0.37***	0.37***	0.29**	0.31**	0.28**	0.22	0.24*	0.23*	-0.20	-0.03	-0.01
LSVT >12 years (DV)	0.00	-0.05	0.18	0.24**	0.05	-0.11	-0.08	-0.07	-0.41***	-0.14	-0.25*
Group parent (DV)	0.23	-0.05	-0.10	-0.07	-0.23*	0.05	-0.13	-0.13	-0.10	0.04	0.06
Group subsidiary (DV)	0.03	-0.15	-0.21*	-0.05	-0.32***	-0.25**	-0.17	-0.23*	-0.06	-0.03	-0.08
Regional wage index (GN)	4.52***	4.79***	4.31***	5.39***	5.65***	6.07***	6.38***	5.63***	5.11***	4.82***	5.00***
Regional wage index (SH)	44.30***	23.71***	29.96***	29.01***	31.84***	27.35***	20.97***	30.47***	15.32	3.83	16.35*
Index of Multiple Deprivation (% rank)	0.09	0.33	1.51***	0.98***	0.77**	0.44	0.79**	1.65***	1.69***	0.53	0.81**
General needs stock (000s)	0.06**	0.05***	0.04	-0.05*	-0.07*	0.01	0.08**	0.10**	0.06	0.10*	0.15***
General needs stock (000s) squared	0.00*	0.00**	0.00*	0.00***	0.00*	0.00	0.00	0.00*	0.00	0.00*	-0.01***
Housing for older people stock (000s)	-0.15	-0.02	-0.02	0.06	0.15	-0.26*	-0.14	-0.25*	-0.15	-0.21	-0.46*
Supported housing (000s)	-0.17	-0.24	-0.47**	-0.36*	-0.07	-0.10	-0.22	-0.02	-0.14	0.30	0.21
Shared ownership (000s)	-0.23**	-0.22*	-0.05	-0.16	-0.26	-0.17	-0.35	-0.45*	-0.43	-0.47	-0.27
N (total observations)	291	292	300	301	307	318	326	330	315	322	321
Mean of cost measure	3.85	3.79	3.94	3.99	4.00	4.26	4.32	4.33	4.19	4.25	4.04
Standard deviation of costs	1.15	1.11	1.28	1.38	1.41	1.40	1.55	1.49	1.38	1.39	1.31
Standard error of the regression	0.79	0.80	0.84	0.90	0.91	0.97	1.03	1.06	1.07	1.07	1.02
R-squared	0.56	0.51	0.60	0.60	0.60	0.55	0.58	0.52	0.44	0.44	0.43
Adjusted R-squared	0.53	0.48	0.57	0.58	0.58	0.52	0.56	0.49	0.41	0.40	0.39
	Unless indicated otherwise, figures presented in the main body of the table are the regression coefficients. DV indicates dummy variable. *** Significant at 95% confidence level ** Significant at 90% *Significant at 80% (standard t-tests).										

Additional variables tested

56. This section outlines additional variables that were tested as part of or alongside the default model but were not selected for addition to the final streamlined model.

Regional wage for supported housing

57. The regional wage (SH) variable looked to be producing non intuitive results for specialised supporting housing providers. This is due to the fact that the vast majority of cases have limited supported housing and therefore significant extrapolation is needed to calculate values for specialised providers. These higher than expected results are most likely to be because the regional wage (SH) variable is picking up other variation in the cost of managing supported housing not caused by regional changes. Potentially the level of support offered in high value areas is in general greater than that in lower cost areas as suggested in the final model discussion.

58. The combined regional wage index was therefore used in the final model providing more powerful and more intuitive results.

Region

59. These were variables indicating the proportion of stock in each of the former Government Office regions or larger super-regions. Significant correlations between regional wage index and region variables mean that they cannot be run concurrently. Models where regional variables replaced the single regional wage index were less powerful and efficient (due to the replacing of one wage index with a number of region variables) and therefore the wage index was chosen for the final model. The results of the region based models were broadly in line with models including the regional wage variable.

Geographic dispersal

60. There was an inconclusive outcome from the testing of the dispersal variables. When tested on the headline unit cost measure this gave rise to excessive multi-collinearity with a number of the existing variables, so it was decided to test on general needs specific cost lines only. Some significant results were discovered in the last couple of years suggesting that more dispersed stock led to greater costs but these could not be replicated in any of the earlier years and such there was not sufficient evidence from which to draw robust conclusions.

Property size (bed-size)

61. There was some evidence that one bed properties were more costly to maintain, however further examination suggests a strong correlation between one bed units and high cost areas, particularly in London. Due to lack of data, it was not possible to control for factors such as whether properties are flats or houses, or whether they form part of certain types of housing estates. It is likely that one bed units are to some extent picking up factors missing from the model – for example costs of additional facilities and infrastructure costs associated with high rise properties in certain urban areas, especially parts of London. Given the likely spurious correlation, property size was not included in the streamlined final model.

Property type (non-self-contained)

62. The number of self-contained supported housing (including housing for older people) units and non-self-contained bedspace supported housing (including housing for older people) units were tested to determine whether the property

type affected cost. The variables however were very closely correlated with the existing supported housing and housing for older people variables. The resulting replacement models were weaker than the existing model and so these variables were not included within the final model.

Residual non-decent units

63. The default model included two variables for decent homes – the residual number of units not meeting the Decent Homes Standard, as well the reduction in non-decent units. Due to non-decent stock reducing sharply up to 2015 the power of the two decent homes variables has also reduced, with only one statistically significant result in the last three years. Permutations of the two variables were tested and the results showed that the model containing only the % reduction non-decent variable performed strongest. Therefore the residual variable was dropped from the final streamlined model.

Welfare reform

64. The effect of estimates for stock affected by the Removal of the Spare Room Subsidy (RSRS, HCA Regulation internal modelling) on unit costs was tested in years 2013 to 2015.The variables were not used in the final model, however, as there was an absence of a clear and consistent relationship. Further investigation found that the welfare reform variables were significantly correlated with both the wage index and deprivation variables suggesting that the variable acts as a proxy for location, which is not to be unexpected as under occupation is more prevalent in weaker housing markets. This meant that the models were unstable and firm conclusions could not be inferred.

Supplementary cost measures

65. This section outlines the results from models run over a number of supplementary cost measures.

Social housing lettings costs

66. As the cost measures did not include major repairs expenditure, the LSVT and decent homes variables were no longer significant. Otherwise the significance and coefficients of the variables were in-line with those seen in the headline unit cost models highlighting the stability of the main model. Overall the headline social housing cost models were preferred due to the increased number of significant variables and higher model power.

Major repairs

67. Major repairs saw stronger results than the other narrow cost lines. R-squared values were above 45% from 2008-2013, which is intuitive as that is when decent homes activity was at its peak. Both LSVT and decent homes variables were significant throughout. Cost variation and model power have reduced significantly in the last couple of years, as vast majority of units are up to Decent Homes Standard and the number of new LSVT entrants has slowed.

Management & maintenance

68. Both management and maintenance models had limited power, with typical R-squared values being between 10-20%, and there were few continually significant variables.

General needs specific cost lines

69. The models provided lower R-squared values, this is however to be expected given that a number of the largest cost drivers, in particular supported

housing activities, were removed. Moreover, examination of case study accounts case some doubt over consistency of apportioning costs to property types. Overall, regressions using these cost lines did not provide a strong basis for conclusions.

Annex A – Details of diagnostic testing

70. This annex summarises the diagnostic testing on the streamlined and default models presented above. This is to ensure that assumptions necessary to draw statistical inferences are met and that results are stable to different formulations to the model.

Heteroskedasticity, autocorrelation & normality of residuals

71. Heteroskedasticity and Autocorrelation Corrected (HAC) standard errors were used throughout the analysis. Normality of residuals was tested and they failed the Jarque-Bera test at conventional confidence levels, with positive kurtosis and skew. This is largely expected with costs, given the zero floor and lack of an upper limit. However, standard statistical tests and inferences are still valid given the large sample size and the use HAC standard errors as outlined above.

72. A number of transformations were tested, including ln(x), ln (x+1) and sqrt(x), but that none of the transformations caused the residuals to pass the Jarque-Bera test. Linear models with HAC standard errors were used throughout.

Hausman test for validity of random effects model

73. The random effects model was the initial preferred model for analysis, given it allows combination of the all-years data in single powerful regression model. However, the random effects model relies on a certain set of assumptions that are tested in the Hausman test. The model failed the Hausman test for the vast majority of years at conventional confidence level, meaning the residuals were correlated with a number of the explanatory variables and therefore the random effects estimators would have been potentially biased. For this reason, results are instead drawn from OLS models run for each year.

Mulitcollinearity

74. Multicollinearity is where that two or more explanatory variables are closely linearly related. The primary concern is that as the degree of multicollinearity increases, the regression model estimates of the coefficients become unstable and the standard errors for the coefficients can get inflated. Multicollinearity was tested using the variance inflation factor (VIF) values for each of the explanatory variables in the default model, with no further action required. Each time new variables were added, the model was tested for multicollinearity. In instances where the new variables created problems of multicollinearity with existing variables the new model was carefully compared to the existing model to determine which of the variables should be included.

Structural Breaks

75. Structural breaks were tested on LSVTs, supported housing providers, and over time based on the Chow test and dummy variable equivalent. Breaks were found between LSVTs and traditionals - however further examination suggests this was tied specifically to the supported housing stock variable with LSVTs doing either less, or different types of, supported housing activity. Given the distinct effect is likely to be supported housing, separate models for LSVTs and other associations were not thought to be on balance support the objectives of analysis.

76. The extent to which providers with over 30% supported housing had different coefficients than other providers was tested. This group was statistically different, but this is to be expected given the nature of supported housing.

Due to the importance of obtaining a model which is intuitive and relatively easy to communicate to a wide audience and the relative stability of the model to the removal of the small number of providers with over 30% supported housing no changes were made on the basis of these findings.

77. A number of break points were found over different time series splits, suggesting that unit cost drivers varied year to year rather than there being particular time series breaks. This contributing to the decision to run the model using singular year OLS models.

Specification testing

78. A degree of misspecification within the default model is suggested by the Ramsey RESET test, which failed in most years under the default and streamlined models. Further testing and examination of significance of particular interaction variables in the Ramsey RESET test model suggests this is predominantly due to the effect of supported housing, with non-linear functions picking up the variability in costs associated with supported housing due to factors outside of the model (e.g. type of client supported housing client group).

79. Rather than a systematic non-linear relationship between supported housing stock owned, these relationships are likely to be simply describing these missing variables as they show up in the costs of particular associations that own significant supported housing stock. This conclusion is supported by tests on general needs only costs which passed the Ramsey RESET test and some significance of non-linear supported housing variables.

80. Given the Ramsey RESET test passed for log transformation, which generated equivalent results, results are judged to be robust to an apparent mis-specification of functional form (see below).

Functional Form

81. Log-linear and square root transformation models, with linear explanatory variables, were tested. The log-linear model in particular performed well in terms of explanatory power and generated findings that were consistent with the default model in terms of magnitude and significance. Moreover, the potential specification issues in the default model did not occur in the log-linear – the Ramsey RESET test was passed for the vast majority of years. However, the weakness of the log-linear transformation is added complexity and less intuitive results. For these reasons the linear functional form was retained for the default model – with comfort taken from the fact that the main inferences are consistent with feasible transformations.

Model selection

82. The aim of the process was to obtain a parsimonious model with non-significant variables omitted, with good explanatory power and that can be readily communicated and understood. Using the default model as the basis of testing, decisions to omit or add variables described in previous sections were made with reference to R^2, Bayesian Information Criterion (BIC) and F-test statistics.

Annex B – Full explanatory variable list

Table 7: List of explanatory variables to be used in regression analysis	
Variable name	**Description**
General needs (% total)	General needs (GN) units owned and managed, averaged over the current and previous year, as a proportion of average total social housing stock owned and managed in the current and previous year. Social and affordable rent general needs stock are included.
Housing for older people (% total)	Housing for older people (HOP) units owned and managed, averaged over the current and previous year, as a proportion of average total social housing stock owned and managed in the current and previous year. Only social stock is included in this variable, as affordable rent HOP stock is unidentifiable in the Statistical Data Return.
Supported housing (% total)	Supported housing (SH) units (excluding housing for older people) owned and managed, averaged over the current and previous year, as a proportion of average total social housing stock owned and managed in the current and previous year. Only social stock is included in this variable, as affordable rent SH stock is unidentifiable in the Statistical Data Return.
Shared ownership (% total)	Total shared ownership stock and other stock which is <100% leasehold (excluding housing for older people), as a proportion of total social housing stock which is owned and managed.
Non-social housing (% total)	Total non-social housing which is owned and managed as a proportion of total social housing stock which is owned and managed.
% of GN – bedroom properties	Owned general needs bedsit/one-bedroom units (used as a proxy for flats) as a proportion of total social housing stock which is owned and managed. All figures are averages over the current and previous year.
% of GN 2-bedroom properties	Owned general needs two-bedroom units as a proportion of total social housing stock which is owned and managed. All figures are averages over the current and previous year.
% of GN 3-bedroom properties	Owned general needs three-bedroom units as a proportion of total social housing stock which is owned and managed. All figures are averages over the current and previous year.
% of GN 4-bedroom plus properties	Owned general needs four-bedroom units as a proportion of total social housing stock which is owned and managed. All figures are averages over the current and previous year.
% reduction in non-decent stock	Reduction in non-decent stock owned since the previous year, as a proportion of total social housing stock. This is a proxy for major repairs. Therefore all recorded *increases* in non-decent stock owned by a provider during a year – due to transfers of stock from local authorities for example – are excluded.
% of non-decent stock	Units of stock which are non-decent at the end of the year, as a proportion of total social housing stock owned and managed.

LSVT < 7 years (DV)	Dummy variable to indicate where a provider has been a stock transfer organisation for under 6 years (i.e. =1 if a the provider is a stock transfer organisation & has been so for less than 6 year, =0 if not).
LSVT 7 - 12 years (DV)	Dummy variable to indicate where a provider has been a stock transfer organisation for between 7 to 12 years (i.e. =1 if a the provider is a stock transfer organisation & has been so for between 7 and 12 year, =0 if not).
LSVT > 12 years (DV)	Dummy variable to indicate where a provider has been a stock transfer organisation for over 12 years (i.e. =1 if a the provider is a stock transfer organisation & has been so for more than 12 year, =0 if not).
Group parent (DV)	Dummy variable to indicate whether the provider is the parent of a group (i.e. =1 if a parent, =0 if not).
Group subsidiary (DV)	Dummy variable to indicate whether the provider is a subsidiary in a group structure (i.e. =1 if a parent, =0 if not).
DV for HOP specialist	A dummy variable to indicate whether the provider can be termed a housing for older people specialist provider (=1 if supported housing for older people) is more than 30% of stock owned or managed, =0 if less).
DV for SH specialist	A dummy variable to indicate whether the provider can be termed a supported housing specialist provider (excluding older people's units) (=1 if supported housing (excl. older) is more than 30% of stock owned or managed, =0 if less).
Total social housing stock	Total stock which is owned and managed, including social, and social <100% leasehold housing. GN affordable rent stock is included, however SH and HOP affordable rent stock is excluded from is variable (because it is excluded from all numerator calculations).
GN stock (000s)	General needs stock in units of thousands which is owned and managed, averaged over the current and previous year. Social and affordable rent GN stock are included.
Shared ownership stock (000s)	Shared ownership stock and other stock which is <100% leasehold (excluding housing for older people) in units of thousands which is owned and managed, averaged over the current and previous year.
Non-social stock (000s)	Non-social stock in units of thousands which is owned and managed, averaged over the current and previous year.
HOP stock (000s)	Housing for older people stock in units of thousands which is owned and managed, averaged over the current and previous year. Only social stock is included in this variable, as affordable rent HOP stock is unidentifiable in the Statistical Data Return.
SH stock (000s)	Supported housing units (excluding housing for older people) in units of thousands which is owned and managed, averaged over the current and previous year. Only social stock is included in this variable, as affordable rent SH stock is unidentifiable in the Statistical Data Return.

Weighted wage index GN	A composite regional wage index has been calculated for every provider. This is based on a regional wage index (based on eight years of national Annual Survey of Hours and Earnings data for relevant occupations, 2006-2013) and the share of GN stock owned by each English region. In the wage index the England average is indexed at 1. In the final regression we subtract 1 off each variable, so the England average is 0. The figure is then multiplied by the proportion of GN stock.
Weighted wage index SH	A composite regional wage index has been calculated for every provider. This is based on a regional wage index (based on eight years of national Annual Survey of Hours and Earnings data for relevant occupations, 2006-2013) and the share of SH stock owned by each English region. In the wage index the England average is indexed at 1.0. In the final regression we take 1.0 off each variable, so the England average is 0.0. The figure is then multiplied by the proportion of SH stock.
Weighted wage index HOP	A composite regional wage index has been calculated for every provider. This is based on a regional wage index (based on eight years of national Annual Survey of Hours and Earnings data for relevant occupations, 2006-2013) and the share of HOP stock owned by each English region. In the wage index the England average is indexed at 1.0. In the final regression we take 1.0 off each variable, so the England average is 0.0. The figure is then multiplied by the proportion of HOP stock.
Weighted wage index combined	A composite regional wage index has been calculated for every provider. This is based on a regional wage index (based on eight years of national Annual Survey of Hours and Earnings data for relevant occupations, 2006-2013) and the share of GN, SH & HOP stock owned by each English region. In the wage index the England average is indexed at 1.0. In the final regression we take 1.0 off each variable, so the England average is 0.0.
Proportion of GN stock in North East	Proportion of GN stock in North East as proportion of total social stock.
Proportion of GN stock in North West	Proportion of GN stock in North West as proportion of total social stock.
Proportion of GN stock in Yorkshire	Proportion of GN stock in Yorkshire as proportion of total social stock.
Proportion of GN stock in West Midlands	Proportion of GN stock in West Midlands as proportion of total social stock.

Proportion of GN stock in East Midlands	Proportion of GN stock in East Midlands as proportion of total social stock.
Proportion of GN stock in East	Proportion of GN stock in East as proportion of total social stock.
Proportion of GN stock in South East	Proportion of GN stock in South East as proportion of total social stock.
Proportion of GN stock in South West	Proportion of GN stock in South West as proportion of total social stock.
Proportion of GN stock in Inner London	Proportion of GN stock in Inner London as proportion of total social stock.
Proportion of GN stock in Outer London	Proportion of GN stock in Outer London as proportion of total social stock.
Proportion of GN, SH & HOP stock in North East	Proportion of GN, SH & HOP stock in North East as proportion of total social stock.
Proportion of GN, SH & HOP stock in North West	Proportion of GN, SH & HOP stock in North West as proportion of total social stock.
Proportion of GN, SH & HOP stock in Yorkshire	Proportion of GN, SH & HOP stock in Yorkshire as proportion of total social stock.
Proportion of GN, SH & HOP stock in West Midlands	Proportion of GN, SH & HOP stock in West Midlands as proportion of total social stock.
Proportion of GN, SH & HOP stock in East Midlands	Proportion of GN, SH & HOP stock in East Midlands as proportion of total social stock.
Proportion of GN, SH & HOP stock in East	Proportion of GN, SH & HOP stock in East as proportion of total social stock.
Proportion of GN, SH & HOP stock in South East	Proportion of GN, SH & HOP stock in South East as proportion of total social stock.
Proportion of GN, SH & HOP stock in South West	Proportion of GN, SH & HOP stock in South West as proportion of total social stock.

Proportion of GN, SH & HOP stock in Inner London	Proportion of GN, SH & HOP stock in Inner London as proportion of total social stock.
Proportion of GN, SH & HOP stock in Outer London	Proportion of GN, SH & HOP stock in Outer London as proportion of total social stock.
Proportion of unavailable voids	Self-contained vacant but unavailable units as a proportion of total social housing stock owned and managed in the current year.
% of available voids	Self-contained GN vacant but available units as a proportion of total social housing stock owned and managed in the current year.
% of voids	All self-contained GN vacant units as a proportion of total social housing stock owned and managed in the current year.
% of GN relets lettings	General needs re-lettings per annum, as a proportion of general needs stock, multiplied by share of general needs of all stock.
Weighted Index of Deprivation	Weighted Index of Multiple Deprivation per annum for each landlord. Constructed by HCA on the basis of lettings per Lower Super Output Area (LSOA) (from CORE data) and the percentile rank from the Index of Multiple Deprivation (IMD) for each LSOA, multiplied by the average general needs stock as a proportion of average total social housing stock in the current and previous year. The 2010 IMD is used for all years.
Weighted Index of Deprivation (minus mean)	Weighted Index of Multiple Deprivation per annum for each landlord. Constructed by HCA on the basis of lettings per Lower Super Output Area (LSOA) (from CORE data) and the percentile rank from the Index of Multiple Deprivation (IMD) for each LSOA. The mean is subtracted from each of the cases and then multiplied by the average general needs stock as a proportion of average total social housing stock in the current and previous year. The 2010 IMD is used for all years.
% of GN stock affected by RSRS (1 bed)	The number of households affected by the Removal of the Spare Room Subsidy welfare reform (1 bed), as a proportion of general needs stock, multiplied by share of general needs of all stock.
% of GN stock affected by RSRS (2 bed)	The number of households affected by the Removal of the Spare Room Subsidy welfare reform (2 bed), as a proportion of general needs stock, multiplied by share of general needs of all stock.
% of GN stock affected by RSRS	The number of households affected by the Removal of the Spare Room Subsidy welfare reform (total), as a proportion of general needs stock, multiplied by share of general needs of all stock.
2015 (DV)	A dummy variable to indicate whether the data is for 2015 (i.e. =1 if 2015, =0 if another year).
2014 (DV)	A dummy variable to indicate whether the data is for 2014 (i.e. =1 if 2014, =0 if another year).
2013 (DV)	A dummy variable to indicate whether the data is for 2013 (i.e. =1 if 2013, =0 if another year).

2012 (DV)	A dummy variable to indicate whether the data is for 2012 (i.e. =1 if 2012, =0 if another year).
2011 (DV)	A dummy variable to indicate whether the data is for 2011 (i.e. =1 if 2011, =0 if another year).
2010 (DV)	A dummy variable to indicate whether the data is for 2010 (i.e. =1 if 2010, =0 if another year).
2009 (DV)	A dummy variable to indicate whether the data is for 2009 (i.e. =1 if 2009, =0 if another year).
2008 (DV)	A dummy variable to indicate whether the data is for 2008 (i.e. =1 if 2008, =0 if another year).
2007 (DV)	A dummy variable to indicate whether the data is for 2007 (i.e. =1 if 2007, =0 if another year).
2006 (DV)	A dummy variable to indicate whether the data is for 2006 (i.e. =1 if 2006, =0 if another year).
2005 (DV)	A dummy variable to indicate whether the data is for 2005 (i.e. =1 if 2005, =0 if another year).
Geographical dispersal – General needs	
Proportion of GN in pockets of 50 LA	Proportion of general needs stock owned in pockets of less than 50 per local authority, multiplied by the share of general needs of all social housing stock.
Proportion of GN in pockets of 100 LA	Proportion of general needs stock owned in pockets of less than 100 per local authority, multiplied by the share of general needs of all social housing stock.
Proportion of GN in pockets of 250 LA	Proportion of general needs stock owned in pockets of less than 250 per local authority, multiplied by the share of general needs of all social housing stock.
Proportion of GN in pockets of 500 LA	Proportion of general needs stock owned in pockets of less than 500 per local authority, multiplied by the share of general needs of all social housing stock.
Proportion of GN in pockets of 50 sub-region	Proportion of general needs stock owned in pockets of less than 50 per sub-region (approximately corresponding to counties), multiplied by the share of general needs of all social housing stock.
Proportion of GN in pockets of 100 sub-region	Proportion of general needs stock owned in pockets of less than 100 per sub-region (approximately corresponding to counties), multiplied by the share of general needs of all social housing stock.
Proportion of GN in pockets of 250 sub-region	Proportion of general needs stock owned in pockets of less than 250 per sub-region (approximately corresponding to counties), multiplied by the share of general needs of all social housing stock.
Proportion of GN in pockets of 500 sub-region	Proportion of general needs stock owned in pockets of less than 500 per sub-region (approximately corresponding to counties), multiplied by the share of general needs of all social housing stock.
Geographical dispersal – Housing for older people	

Proportion of HOP in pockets of 50 LA	Proportion of housing for older people owned in pockets of less than 50 per local authority, multiplied by the share of housing for older people of all social housing stock.
Proportion of HOP in pockets of 100 LA	Proportion of housing for older people owned in pockets of less than 100 per local authority, multiplied by the share of housing for older people of all social housing stock.
Proportion of HOP in pockets of 250 LA	Proportion of Housing for older people owned in pockets of less than 250 per local authority, multiplied by the share of Housing for older peopleof all social housing stock.
Proportion of HOP in pockets of 500 LA	Proportion of housing for older people owned in pockets of less than 500 per local authority, multiplied by the share of housing for older people of all social housing stock.
Proportion of HOP in pockets of 50 sub-region	Proportion of housing for older people owned in pockets of less than 50 per sub-region (approximately corresponding to counties), multiplied by the share of housing for older people of all social housing stock.
Proportion of HOP in pockets of 100 sub-region	Proportion of housing for older people owned in pockets of less than 100 per sub-region (approximately corresponding to counties), multiplied by the share of housing for older people of all social housing stock.
Proportion of HOP in pockets of 250 sub-region	Proportion of housing for older people owned in pockets of less than 250 per sub-region (approximately corresponding to counties), multiplied by the share of housing for older people of all social housing stock.
Proportion of HOP in pockets of 500 sub-region	Proportion of housing for older people owned in pockets of less than 500 per sub-region (approximately corresponding to counties), multiplied by the share of housing for older people of all social housing stock.
Geographical dispersal – Supported Housing (Excl. Housing for Older People)	
Proportion of SH in pockets of 50 LA	Proportion of supported housing (excluding older person's units) owned in pockets of less than 50 per local authority, multiplied by the share of supported housing (excluding older person's units) of all social housing stock.
Proportion of SH in pockets of 100 LA	Proportion of supported housing (excluding older person's units) owned in pockets of less than 100 per local authority, multiplied by the share of supported housing (excluding older person's units) of all social housing stock.
Proportion of SH in pockets of 250 LA	Proportion of supported housing (excluding older person's units) owned in pockets of less than 250 per local authority, multiplied by the share of supported housing (excluding older person's units) of all social housing stock.
Proportion of SH in pockets of 500 LA	Proportion of supported housing (excluding older person's units) owned in pockets of less than 500 per local authority, multiplied by the share of supported housing (excluding older person's units) of all social housing stock.

Proportion of SH in pockets of 50 sub-region	Proportion of supported housing (excluding older person's units) owned in pockets of less than 50 per sub-region (approximately corresponding to counties), multiplied by the share of supported housing (excluding older person's units) of all social housing stock.
Proportion of SH in pockets of 100 sub-region	Proportion of supported housing (excluding older person's units) owned in pockets of less than 100 per sub-region (approximately corresponding to counties), multiplied by the share of supported housing (excluding older person's units) of all social housing stock.
Proportion of SH in pockets of 250 sub-region	Proportion of supported housing (excluding older person's units) owned in pockets of less than 250 per sub-region (approximately corresponding to counties), multiplied by the share of supported housing (excluding older person's units) of all social housing stock.
Proportion of SH in pockets of 500 sub-region	Proportion of supported housing (excluding older person's units) owned in pockets of less than 500 per sub-region (approximately corresponding to counties), multiplied by the share of supported housing (excluding older person's units) of all social housing stock.

Annex C - Supplementary cost definitions

Table 9: Supplementary cost lines tested as part of the regression analysis	
Cost Measure	**Individual cost lines included**
Headline social housing costs excluding major repairs	Management, service charge costs, care & support costs (2005-2012), routine maintenance, planned maintenance, lease charges (2005-2012), other costs, other social housing activities - other (expenditure) & other social housing activities - charges for support services (expenditure)
Social housing lettings cost (wide)	Management, service charge costs, care & support costs (2005-2012), routine maintenance, planned maintenance, lease charges (2005-2012), other costs & other social housing activities - charges for support services (expenditure)
Social housing lettings cost (narrow)	Management, service charge costs, routine maintenance, planned maintenance & other social housing activities - charges for support services (expenditure).
Management	Management
Maintenance	Routine maintenance & planned maintenance
Major repairs	Major repairs & capitalised major repairs and re-improvements

homesandcommunities.co.uk
mail@homesandcommunities.co.uk
0300 1234 500

Homes and Communities Agency
Fry Building
2 Marsham Street
London SW1P 4DF

The Homes and Communities Agency is committed to providing accessible information where possible and we will consider providing information in alternative formats such as large print, audio and Braille upon request.

Publication date: June 2016