



# Index of comparative operating costs of Registered Social Landlords in Scotland

December 2007



Indepen is a management and economic consultancy. We understand and have experience of government, regulation and investors, as well as business and other forms of enterprise. We work to make business sense of regulation to produce better results for stakeholders.

We use our knowledge to challenge constructively and our thinking is independent, distinctive and rigorous. We work in this way to promote public and private value with clients in the UK, EU and elsewhere.

Further information can be found at [www.indepen.co.uk](http://www.indepen.co.uk).



## Table of Contents

Executive Summary.....	1
1 Introduction.....	3
2 New data available.....	4
2.1 Capitalised major repair costs.....	4
2.2 Scottish Housing Quality Standard .....	6
2.3 Proportion of transferred stock.....	7
2.4 Data used in analysis .....	7
3 Estimation of the OCI .....	9
3.1 Validity of pooling .....	9
3.2 Cost drivers .....	11
3.3 Conclusions.....	12
4 Results and interpretation .....	13
4.1 Preferred cost relationship .....	13
4.2 Predicted costs and residuals .....	15
4.3 Illustrative analysis .....	18
5 Conclusions.....	24
Appendix A: Table of results .....	25
Appendix B: Allowing for inflation .....	31
Appendix C: Balanced and unbalanced panels.....	32
Appendix D: Pooling tests .....	33
Appendix E: Glossary.....	37



## Executive Summary

### Purpose

This paper reports the results of estimating the Operating Cost Index (OCI) for Scottish RSLs using data for 2005/06.

### Approach

The statistical analysis uses the data on general needs RSLs and LSVTs to generate an average or predicted level of operating costs for each RSL depending on its circumstances. This is then compared against its actual costs. The difference between the predicted cost and the actual cost, known as the residual, depends on a number of factors including the ability of the RSL to convert inputs (costs) into outputs (service provisions). Given the possible presence of other factors in the residual, it is important to note the fact that if an RSL has higher than predicted costs, this does not necessarily imply it is inefficient or vice versa. We have distinguished between

- the (housing) services that RSLs provide to tenants and
- the (property) services that RSLs provide to the assets.

For the former, we have used information on the different types of stock as the types of services provided to tenants largely depend on the type of stock they live in.

RSLs' compliance with the Scottish Housing Quality Standard potentially provides a measure of the latter. In conjunction with forthcoming information on capitalised major repairs, it should be possible to gain a better understanding of the property services aspects.

### Conclusions on the approach

We can conclude that

- it is statistically valid to pool data for different types of RSL and also over time in an unbalanced panel data model, and
- the structure of the cost function appears robust over time; with the inclusion of data for 2005/06, the coefficients of the previously significant variables remain significant and similar in magnitude.

### Results

The main results are as follows.

- Allowing for inflation, the costs of Scottish RSLs are rising at a rate of just over 2% per annum.
- Other things being equal, the operating costs of LSVTs are higher than those of general needs RSLs.
- The mix of housing types (self-contained, bed spaces and non-self contained) has a significant effect on cost, in line with expectations. This is also true when considering shared ownership and sheltered stock.
- Variables describing the physical nature of the stock (type of building, age and size) are collectively significant with coefficients in line with expectations.
- As in the previous study, compared with houses, RSLs with stocks in high rise blocks have higher operating costs and RSLs with units in blocks of four have lower operating costs.



We have illustrated a way of analysing the effect on operating costs of mergers, while normalising for other relevant circumstances. There are not enough instances of merger to draw conclusions at this stage.



# 1 Introduction

Communities Scotland commissioned Indepen to develop the Operating Cost Index (OCI) as a means of making meaningful comparisons of operating costs across the RSL sector in Scotland.

Communities Scotland has stated that its overall aim is “*to understand how an RSL can improve its efficiency, effectiveness and the quality of its services and products*”<sup>1</sup>. This requires an assessment of the level of RSLs’ operating costs given their operating conditions and the services they provide. As such, Communities Scotland uses the OCI as one part of a rounded picture of RSL performance. It has not previously integrated this into its regulatory process but now intends to introduce it to all regulatory assessment as a basis for further engagement with RSLs to understand better the reasons why their costs are higher or lower than predicted.

This report contains the results of the analysis including 2005/06 data. The structure of the report is as follows.

- Section 2 examines changes in the data that are available, the composition of RSLs in the sector and the classification of RSLs.
- Section 3 considers certain aspects of the OCI methodology. For the most part the approach is as described in our previous report<sup>2</sup>.
- Section 4 presents the results of the analysis and their interpretation.
- Section 5 contains our conclusions and recommends areas for further development.

---

<sup>1</sup> Communities Scotland briefing paper on the RSL operating cost index, July 2005, [http://www.communitiesscotland.gov.uk/stellent/groups/public/documents/webpages/cs\\_010588.hcsp](http://www.communitiesscotland.gov.uk/stellent/groups/public/documents/webpages/cs_010588.hcsp)

<sup>2</sup> Indepen (2006), “Index of the comparative operating costs of Registered Social Landlords in Scotland. Analysis incorporating 2004/05 data”. Available on Communities Scotland website.



## 2 New data available

The 2005/06 data include for the first time information on

- capitalised major repair costs, which might permit us to adopt a more consistent measure of operating cost
- compliance with the Scottish Housing Quality Standard (SHQS), a measure of the quality of RSLs' social housing stock
- the proportion of each RSL's stock that has been transferred from public bodies (including local authorities), to better reflect the challenges provided by such transfers.

We discuss these new data items in turn.

### 2.1 Capitalised major repair costs

RSLs are free to decide whether major repair costs are capitalised and depreciated over a period of time or expensed in the year in which they are incurred. Other things being equal, an RSL that capitalises its costs will appear to have lower operating costs than one that expenses the same amount.

We have considered two specifications of the net operating costs variable: one that includes these costs and one that excludes them. They are as follows.

Net operating cost		Net operating costs including capitalised major repairs	
=	Total operating costs	=	Total operating costs
		plus	Capitalised major repairs
less	Impairment and revaluation adjustments	less	Impairment and revaluation adjustments
	Depreciation		Depreciation
	Income from factoring		Income from factoring
	Income from other management services		Income from other management services
	Miscellaneous income from other sources		Miscellaneous income from other sources

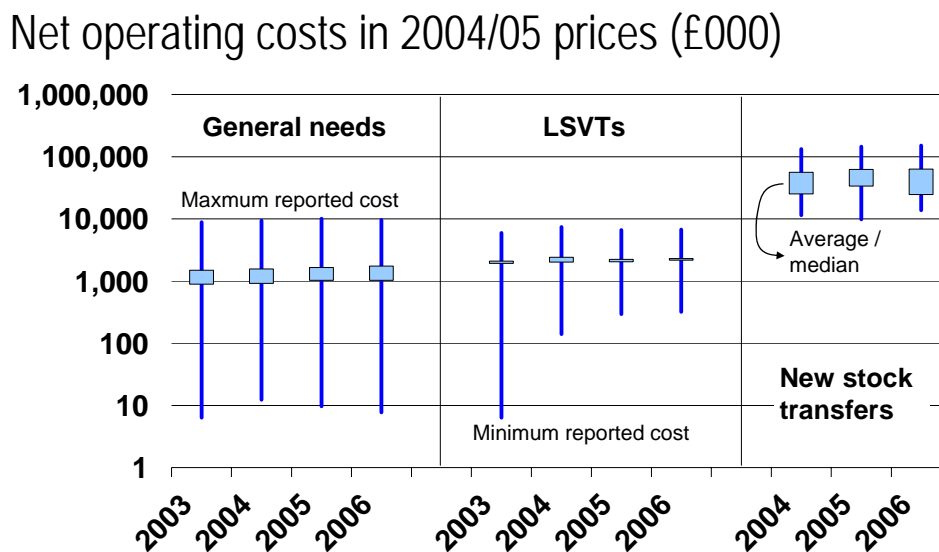
The cost definition that includes capitalised major repairs costs is better but some RSLs did not provide these figures in 2005/06. Furthermore, the capitalised costs provided are based on financial projections and are un-audited. We understand that there will be a statutory requirement for all RSLs to disclose total expenditure in each year on works to existing properties, split between the amounts capitalised and the amounts charged to the income and expenditure account. We have therefore retained the net operating cost specification for the purposes of this study but recommend incorporating major repair costs in future OCI work.

Summary statistics for net operating costs are in Figure 2.1. The length of the vertical bars shows the range of net operating costs. In 2002/03, for example, this was from around £6,000 to £10 million. The range remained relatively unchanging in real terms throughout the four years. For LSVTs, on the other hand, Moray Housing Partnership showed an increase in costs from around £6,000 in 2002/03



to around £140,000 in 2003/04<sup>3</sup>, resulting in the large decrease in the range of costs as shown in Figure 2.1. For the new stock transfers, the range of net operating costs has remained relatively constant in real terms over the three year period. The blue boxes on the vertical lines denote the median and the mean. For both general needs RSLs and LSVTs, average net operating costs have changed little in real terms. For the general needs RSLs, average operating costs were £1.7m in 2005/06 (£2.3m for LSVTs).

Figure 2.1. Annual net operating costs in 2004/05 prices



Note: X-axis in logarithmic scale

Taking the total costs of providing RSL services to be the sum of net operating costs and capitalised major repairs, capitalised major repair costs ranged between 0.07% and 24% of total costs for the 13 LSVTs that reported capitalised major repairs. For general needs RSLs, 49 RSLs capitalised their major repairs and 73 did not. The lowest reported figure was £5,000 by West Granton, which was around 1% of its net operating costs. At the other end of the scale, Oak Tree reported £2.1m in capitalised major repair costs, compared to a total cost of £4.6m.

While there are differences between the RSLs' capitalisation policies, this new data represents a useful step. In Figure 2.2 we show those RSLs that have major repair costs that amount to 10% or more of their total costs.

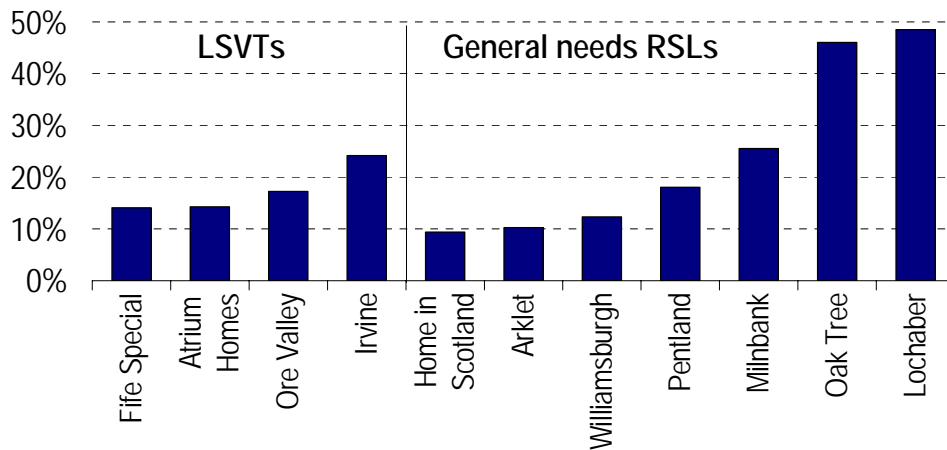
<sup>3</sup> Moray HP more than doubled their housing stock from 125 social housing units in 2002/03 to 266 in 2003/04.



Figure 2.2. RSLs with significant capitalised major repair costs in 2005/06

### Capitalised major repair costs in 2005/06

(% Total costs)

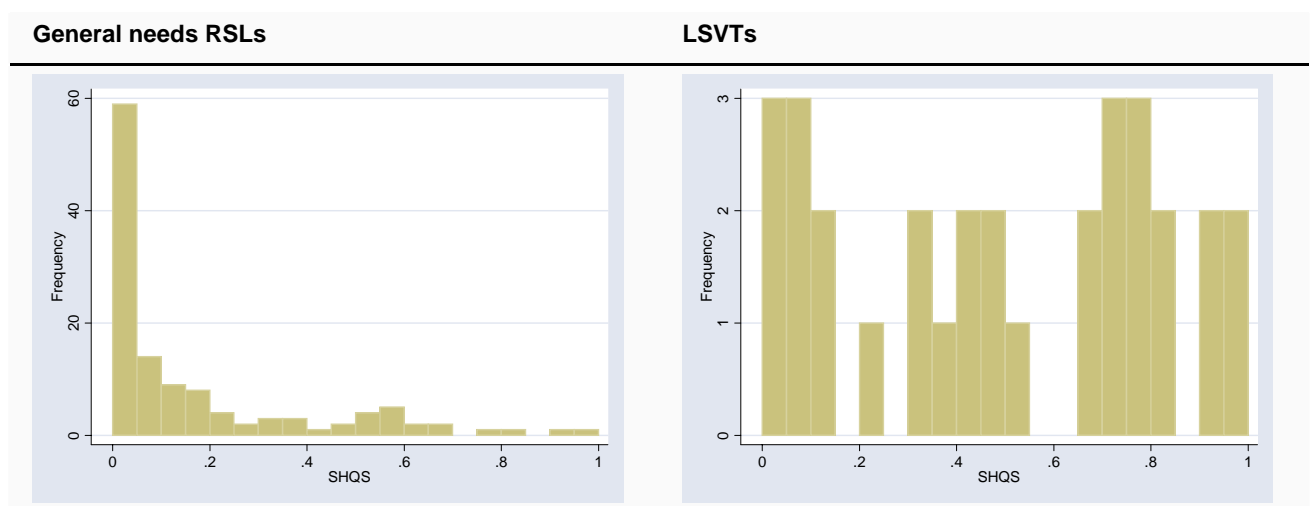


## 2.2 Scottish Housing Quality Standard

Social landlords (local authorities and RSLs) are required to meet the Scottish Housing Quality Standard (SHQS) by 2015.<sup>4</sup> This represents a minimum standard for the social housing stock that RSLs manage. Using data on the capitalised major repair costs, we can compare the expenditure by RSLs against the work that RSLs are required to carry out on their assets.

Data on RSLs' compliance became available in September 2007 and is summarised below. Figure 2.3 shows that around 60 general needs RSLs have failure rates of between 0% and 5% and a small number have a failure rate between 95% and 100%. Across the 33 LSVTs, the distribution of failure rates is evenly spread between 0% and 100%.

Figure 2.3. Reported failure rates of the SHQS by all RSLs



<sup>4</sup> <http://www.scotland.gov.uk/pages/news/2004/02/SECD392.aspx>



## 2.3 Proportion of transferred stock

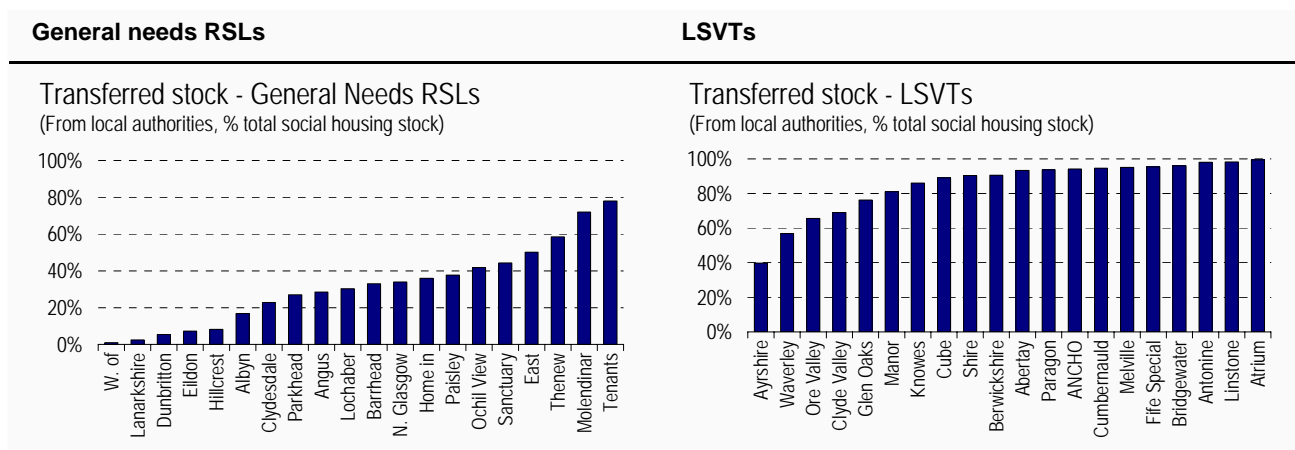
In the Annual Performance and Statistical Return (APSR) submissions, RSLs are classified into nine groups: 3 general needs according to the number of their social housing units<sup>5</sup>, 2 LSVT types depending on size<sup>6</sup>, 3 specialist types<sup>7</sup> and an additional class for the new stock transfers of 2003/04<sup>8</sup>.

The standard way of testing whether these different types of RSLs have different cost characteristics is by creating nine variables, one for each type. For example, the “small LSVT” variable would take a value 1 if the RSL was an LSVT with less than 1,000 units, and 0 for all other RSLs.

This approach may suffer from the fact that the classification reflects the status of RSLs at the time they were registered and there may have been subsequent changes in portfolios. For example, an RSL classified as general needs may have stock that has been transferred from public bodies after registration and an RSL registered as an LSVT may have acquired new stock.

We have used information from the APSR about the percentage of stock, as at 31 March 2007, held by each RSL that had been acquired through stock transfers. This is the first time this information has been available. Our initial analysis shows that of the 122 RSLs classified as general needs 21 had stock that had been acquired from public sector bodies. This ranged from 1% (West of Scotland) to 27% (Parkhead), 44.3% (Sanctuary Scotland), and 78% (Tenant First Housing). Similarly of the 31 landlords categorised as LSVTs, only 10 had 100% transferred stock and one had as little as 39.72% (Ayrshire Housing) of transferred stock. Our review identified also two RSLs that should not be categorised as LSVTs because their stocks are all new build. These are Aberdeenshire and Moray Housing Partnership.

Figure 2.4. Stock acquired through local authority transfers in 2005/06



## 2.4 Data used in analysis

As the new potential cost drivers are only available for 2005/06 we have used a single year equation to explore their significance. Our conclusions are as follows.

<sup>5</sup> General needs with less than 250 units, between 250 and 1,000 units, and greater than 1,000 units.

<sup>6</sup> LSVTs with less than 1,000 units, and LSVTs with greater than 1,000 units

<sup>7</sup> RSLs specialised in providing services to Older Clients, RSLs with Supported Accommodation and Other specialist RSLs

<sup>8</sup> This category has 3 RSLs: Glasgow Housing Association, Scottish Borders Housing Association and Dumfries and Galloway Housing Association



- Capitalised major repair costs can be included in our operating cost definition although in the single year analysis the explanatory power is slightly reduced compared with the equation that excludes them. This requires further investigation when more data become available. We have used the definition of net operating cost that excludes major repair costs in the estimation of the 2005/06 OCI.
- A higher rate of failure to meet the SHQS is positively related to higher levels of operating costs.<sup>9</sup> Since this data is only available for 2005/06 and housing quality compliance changes over time, we have not included it in this year's analysis but recommend that it should be taken into account in the future.
- Using the percentage of stock transferred gives an equation with similar explanatory power (adjusted  $R^2$ ) to that using the LSVT dummy variable. However, the coefficients imply that the impact of transferred stock is greater than if the LSVT dummy variable is used. This suggests that, unsurprisingly, the continuous variable better represents the effect of transferred stock because it takes into account those general needs RSLs who have transferred stock. In the light of this, we use the proportion of stock transferred instead of the LSVT dummy variable. Discussion with Communities Scotland indicated that for the majority of the RSLs this proportion stays relatively constant over time and so we can use the same value for each year. Incorporating it into our analysis now means that if the proportions change in future, this will be reflected in future estimates.

---

<sup>9</sup> It should be noted that the regression coefficient is only significant at the 80% confidence level.



### 3 Estimation of the OCI

The methodology is to a large degree the same as described in previous reports. We estimate a cost relationship for RSLs, which is a description of how different cost drivers affect the level of RSLs' net operating costs across the sector as a whole. We then use this cost relationship to calculate the predicted level of costs for individual RSLs, and compare these with their actual costs. We use the cost definition as given in Section 2.1.

In this section we comment on two aspects of the approach

- the validity of pooling the data across RSL types and over time, and
- the cost drivers.

We provide a glossary for the terms used in this report in Appendix E.

#### 3.1 Validity of pooling

In constructing the operating cost index one of the fundamental issues we need to address are whether the assumption that the relationship between the costs of RSLs and a set of cost drivers is the same for all types of RSLs and whether it remains unchanged over time.

It is important to note that if we erroneously pool the data, the cost relationship will be mis-specified and the estimated coefficients will be biased. For example, if general needs RSLs and LSVTs with different cost relationships are pooled together in a single equation, the estimated effects of the cost drivers will be the weighted average for the two types. In the extreme case where a cost driver had a positive impact on general needs RSLs' costs but a negative impact of LSVTs' costs, the estimated coefficient would be statistically equivalent to zero if there were the same number of RSLs of each type. This might lead to the cost driver being dropped from the final preferred equation. Differences due to this misspecification will carry over to the OCI estimates and the residuals. In this example, the estimated cost relationship would represent a "hybrid" cross between general needs RSLs and LSVTs and comparisons made against actual general needs or LSVT types would not be meaningful.

We apply two pooling tests before we opt for a panel data model. Details of these tests are described in Appendix D.

##### 3.1.1 Pooling different types of RSL

Before we can draw inferences we test the hypothesis that data from the two types of RSL can be pooled. If we do not reject this hypothesis then we infer that our proposed cost model is valid for both types. If we reject the pooling hypothesis then we infer that there are differences in how costs are determined for at least one of the cost drivers. For example, in the previous study, discussions with Communities Scotland and an RSL focus group concluded that care and support providers might have different cost and funding structures. Therefore, these types of RSL were excluded from the analysis. In this study we have not returned to this issue and continue to focus our attention on general needs RSLs and LSVTs. We note, however, that further work in this area might enable the analysis to be adapted to include such providers.

There are two equivalent ways of testing the pooling hypothesis.



- Estimate the cost function separately for general needs RSLs and for LSVTs and evaluate the degree to which these separate regressions are statistically better than a single regression including data from both types.<sup>10</sup>
- Estimate a single regression but augment the cost function with a set of interaction and intercept dummy variables. The hypothesis that the data can be pooled then translates into a test of whether the set of additional parameters are jointly equal to zero.

With only 33 LSVTs in 2004/05 and 2005/06, we cannot estimate a valid cost specification using LSVT data alone and we have therefore used the second method. It shows that the interaction variables are jointly insignificant and we infer that we can validly pool the data.

### 3.1.2 Pooling over time

In general, the larger the number of data points, the better will be our ability to draw inferences from the analysis. In our earlier studies we analysed two years' data together in an approach known as panel data analysis. As for pooling over RSL types, to see whether it is statistically valid to pool or combine the data from more than one year, we ask whether or not the cost relationship is the same for each of the years we are considering. If it is we can use a panel and if not, we would use single year equations.

To test this we have estimated the August 2006 cost specification using the pooled data from different years and compared them against single year equations. For example, for Model 1 in Table 3.1 we compare the residuals from the 3-year equation using data from 2002/03 to 2004/05 against the combined residuals from three single-year equations. We then calculate an F test statistic for each equation and compare it with a "critical" value for this statistic. If the test statistic is larger than the critical value, we reject the hypothesis that the coefficients from the two single year equations are not the same, and vice versa. The results are summarised below.

**Table 3.1. Results from pooling tests of RSLs over time**

	<b>Model 1</b>	<b>Model 2</b>	<b>Model 3</b>	<b>Model 4</b>
Years covered	2002/03 to 4/05	2002/03 to 5/06	2003/04 to 5/06	2004/05 to 5/06
Years covered	3	4	3	2
Observations	462	615	461	308
F statistic	2.4	3.5	1.5	0.5
Critical value at 99% confidence level	1.9	1.9	1.9	1.9
<b>Conclusion</b>	Cannot combine	Cannot combine	<b>Can combine</b>	<b>Can combine</b>

- Models 1 and 2 include data from 2002/03, i.e. the year before the three local authority transfers. The statistics suggest that only the models that exclude the 2002/03 data can be combined. That is, at the 99% confidence level we reject the hypothesis that the coefficients for the panels are statistically the same as the coefficients in the combined equation for the equations that include the 2002/03 data. This means that it is not valid to pool the data from 2002/03 to either 2004/05 or 2005/06.

<sup>10</sup> Formally, this is known as an F-test on the residuals, and is same approach taken for testing the validity of pooling across time.



- The results suggest that we can combine data from 2003/04 to 2005/06 or from 2004/05 to 2005/06.

### Balanced versus unbalanced panels

In addition to considering pooling over time, we need to consider whether to adopt a balanced panel where the composition of data remains the same over time, or an unbalanced panel. The social housing sector is characterised by mergers and transfers of local authority stock and as a result the number and mix of RSLs changes, sometimes substantially. The recent history of stock transfers and mergers is as follows.

- In 2003/04 there were three large scale whole stock transfers<sup>11</sup> involving almost 40% of the total RSL housing units in Scotland. There were also six mergers.
- There were no mergers involving general needs RSLs or LSVTs in 2004/05.
- In 2005/06 there were four mergers, three involving general needs RSLs.

We can back-date the information relating to the merging entities to obtain a balanced panel of data, i.e. one with the same entities represented in each year. However, the back-dating process risks becoming somewhat cumbersome if the OCI is to continue in future years and in this study we have adopted an unbalanced panel approach. This means that we aggregate the cost data of the merging RSLs in the year in which the merger takes place. We have illustrated the effects of using balanced and unbalanced panel approaches in Appendix C.

## 3.2 Cost drivers

The data on many of the potential cost drivers are from the APSR. They include

- type of organisation – general needs or whole stock transfer
- the proportion of RSLs' stocks that are transferred from public bodies
- the types of accommodation that an RSL manages – self contained units, bed spaces, and non self contained units
- the extent to which the units have been adapted to accommodate different types of tenants – general needs, sheltered and other specialist units - we note that care and supported RSLs have been excluded from this analysis, but nonetheless there are some RSLs with specialist housing units
- the age of the property and other physical characteristics of the accommodation – houses, high rise flats, tenements, four-in-a-block, other flats or maisonettes
- the location of the properties – disaggregated according to local authority areas.

In the previous analysis, because data on proportions of transferred stock in RSLs' portfolios was not available, so we used the APSR classification of LSVTs which is equivalent to assuming that all RSLs classified as LSVTs have 100% transferred stock and those not so classified have no transferred stock. For this study we have a continuous variable that measures the proportions of transferred stock regardless of an RSL's classification.

As in the previous study, although we have data on the location of RSLs' stock, we have not used it in the cost specification. This is because the data show local authority areas, which vary in size. Including it in our analysis would give ambiguous results which it would be difficult to interpret.

<sup>11</sup> Scottish Borders Housing Association, Dumfries and Galloway Housing Partnership and Glasgow Housing Association.



Similarly, our measure of stock concentration uses local authority data and faces similar interpretation issues.

We did not include a cost driver for those RSLs that reported sheltered or specially adapted housing units as in the previous study because the majority of care and supported providers have been excluded from the analysis. However, on revisiting this issue we found a small number of general needs RSLs and LSVTs that reported sheltered and specially adapted units. Table 3.2 shows the top five RSLs for each category.

**Table 3.2. Proportion of sheltered and specially adapted units reported**

ID	Name	2003	2004	2005	2006
<b>% sheltered units</b>					
164	Orkney	34%	34%	34%	34%
107	Eildon	44%	40%	38%	39%
153	Loreburn	40%	41%	42%	43%
83	Castlehill	47%	47%	45%	47%
114	Arklet	67%	67%	59%	58%
<b>% specially adapted units</b>					
232	Rural Stirling	8%	7%	9%	19%
325	Moray HP	29%	21%	21%	21%
301	Bridgewater	17%	18%	20%	22%
238	Easthall Park	24%	25%	24%	24%
261	Muirhouse	11%	13%	13%	26%

Given this data, we test whether these two cost drivers have significant impacts on the general needs RSLs and LSVTs included in this analysis.

### 3.3 Conclusions

Given the preceding analysis we conclude that to estimate the 2005/06 cost relationship we should

- adopt an unbalanced panel
- use three years' data from 2003/04 to 2005/06
- use data on the percentage of stock transferred instead of the LSVT dummy variable
- use data for general needs RSLs, LSVTs and large scale whole stock transfers (LAWSTs) – we need to include binary variables for the LAWSTs and a variable describing the proportion of stock acquired through stock transfers, but there is no need for interaction dummy variables<sup>12</sup>
- use data on the proportion of sheltered and specially adapted social housing units
- use net operating costs as our dependent variable, expressed in 2004/05 prices.

<sup>12</sup> See Appendix 5D.2 for a description of intercept and interaction dummy variables



## 4 Results and interpretation

### 4.1 Preferred cost relationship

As with previous studies, we have estimated a number of possible cost relationships using the data available. The preferred cost function is a total cost specification for general needs RSLs, LSVTs and the three large scale whole stock transfers (LAWSTs). It is as follows.

Table 4.1. Estimated cost relationship

Variable	Coefficient	Standard error	p value
<b>Ln (Total social housing units)</b>	<b>0.870</b>	<b>0.025</b>	<b>0.000</b>
<b>% sheltered units</b>	<b>0.546</b>	<b>0.194</b>	<b>0.005</b>
% specially adapted units	0.047	0.301	0.877
<i>% bed spaces</i>	<i>-1.621</i>	<i>0.989</i>	<i>0.101</i>
% non self contained units	-1.811	2.984	0.544
<b>Ln (Shared units)</b>	<b>0.041</b>	<b>0.014</b>	<b>0.003</b>
<i>Log (Non social housing units)</i>	<i>0.019</i>	<i>0.014</i>	<i>0.182</i>
<b>% high rises</b>	<b>0.714</b>	<b>0.208</b>	<b>0.001</b>
<b>% tenement</b>	<b>0.174</b>	<b>0.098</b>	<b>0.074</b>
% block	-0.152	0.240	0.526
% other flats	-0.033	0.128	0.799
% built between 1919 and 1944	0.141	0.158	0.370
% built between 1945 and 1964	0.133	0.136	0.327
% built between 1965 and 1982	0.172	0.139	0.215
% built after 1982	-0.057	0.115	0.622
% 2 bedroom units	-0.227	0.479	0.636
<b>% 3 bedroom units</b>	<b>-0.954</b>	<b>0.423</b>	<b>0.024</b>
% 4 bedroom units	0.218	0.439	0.620
% 5 bedroom units	0.406	0.629	0.518
<b>% Stock acquired through stock transfers</b>	<b>0.284</b>	<b>0.089</b>	<b>0.001</b>
<b>New stock transfer binary variable</b>	<b>0.414</b>	<b>0.159</b>	<b>0.009</b>
Years since registration	-0.002	0.011	0.861
<i>Year squared</i>	<i>0.000</i>	<i>0.000</i>	<i>0.091</i>
<b>Year</b>	<b>0.021</b>	<b>0.010</b>	<b>0.042</b>
<b>Constant</b>	<b>8.219</b>	<b>0.439</b>	<b>0.000</b>

Note: Cost drivers statistically significant at the 10% level are highlighted in **bold**, and at the 20% level in *Italics*.



The logarithmic transformation is necessary for cost drivers that have a large variance, such as the number of social housing units. Variables that are expressed in percentage terms are bounded between 0 and 1 and we do not need the transformation.

The predicted costs of individual RSLs are provided in Appendix A. The interpretations of the cost drivers are discussed below.

#### **4.1.1 Time effects**

We have included a time trend, which measures the rate of change of net operating costs in real terms (since we have already allowed for inflation in the net operating costs definition) from one year to the next. This is estimated to be 2.1% per annum, and is statistically significant.

We also included a time trend that depends on the age of RSLs but this was not statistically significant.

#### **4.1.2 Transferred stock**

An LSVT with 100% transferred stock has costs that are 28.4% higher compared to an otherwise equivalent general needs RSL with no transferred stock. RSLs with some transferred stock have a proportionate increase in their costs, regardless of their classification. For example, Ochill View, a general needs RSL with around 40% transferred stock and Ayrshire, an LSVT with a similar amount of transferred stock, are predicted to have costs which are 11.36% ( $= 28.4\% * 40\%$ ) higher than an RSL that has no transferred stock.

In addition, there is a positive coefficient on the new stock transfer variable indicating that, other things being equal, the fixed costs of LAWSTs are 41.4% higher than a LSVT with 100% transferred stock that is similar in all other respects. Equivalently, since a local authority whole stock transfer RSL is by definition an LSVT with 100% transferred stock, it would have fixed costs around 70% higher than a general needs RSL with no transferred stock.

#### **4.1.3 Variables describing the type of social housing**

The number of social housing units, which is the sum of the number of self contained units, bed spaces and non self contained units has a significant positive effect on the net operating costs of RSLs. One of the types, self contained units, is taken as the point of reference and is not included in the equation; otherwise the proportions would add to 1. For every 10% increase in the number of social housing units owned and managed by an RSLs, the net operating cost increases by 8.7%. This indicates economies of scale where the increases in units do not lead to proportionate increases in operating costs.

A 10% increase in shared ownership units increases costs by an additional 0.41%. The impact of having non social housing units is not statistically significant and is smaller than the impact of shared units.

To estimate the effect on costs of having the different types of social housing unit (bed spaces, self contained and non self contained) we include the proportions of these types. The coefficients show that bed spaces are less costly than self contained units. On average non self contained units appear to be less costly than self contained units, although the majority of RSLs tend to have self contained social housing units.

Similarly, the impact of having sheltered units (the sum of sheltered bed spaces, self contained and non self contained sheltered units) is that it increases RSLs' costs. Each 10% increase in this type of



units in an RSL's portfolio increases its costs by 5.46%. This is relevant to only a small number of RSLs. The impact of specially adapted units is not statistically significant.

#### 4.1.4 Variables describing physical characteristics of the housing stock

We have identified three variables that affect costs.

- Type of building in which the unit is located, i.e. house, high rise block, tenement, four-in-a-block or other flats and maisonettes.
- Age of the unit, as described by the period in which the unit was built, i.e. before 1919, between 1919 and 1944, between 1945 and 1965, between 1965 and 1982 and after 1982.
- Size of the unit, as measured by the number of bedrooms, i.e. one, two, three, four or five or more bedrooms.

For each characteristic, one of the variables is taken as the point of reference and is not included in the equation otherwise the proportions would add to 1.<sup>13</sup> Although not every variable in describing a particular characteristic is statistically significant, taken together they are. For example, the coefficient on the percentage of units built after 1982 is not statistically significant but the probability that the coefficients on all four age variables are zero is low. This means the variables are jointly statistically significant.

As in the previous study, compared with houses, RSLs with stocks in high rise blocks have higher operating costs and RSLs with units in blocks of four have lower operating costs.

Compared with RSLs whose social housing units were built before 1919, RSLs which have social housing units built after this period tend to have higher costs. Smaller units are less costly than larger ones as shown by the signs on the coefficients of the two and three bedroom units compared with the four and five bedroom units.

## 4.2 Predicted costs and residuals

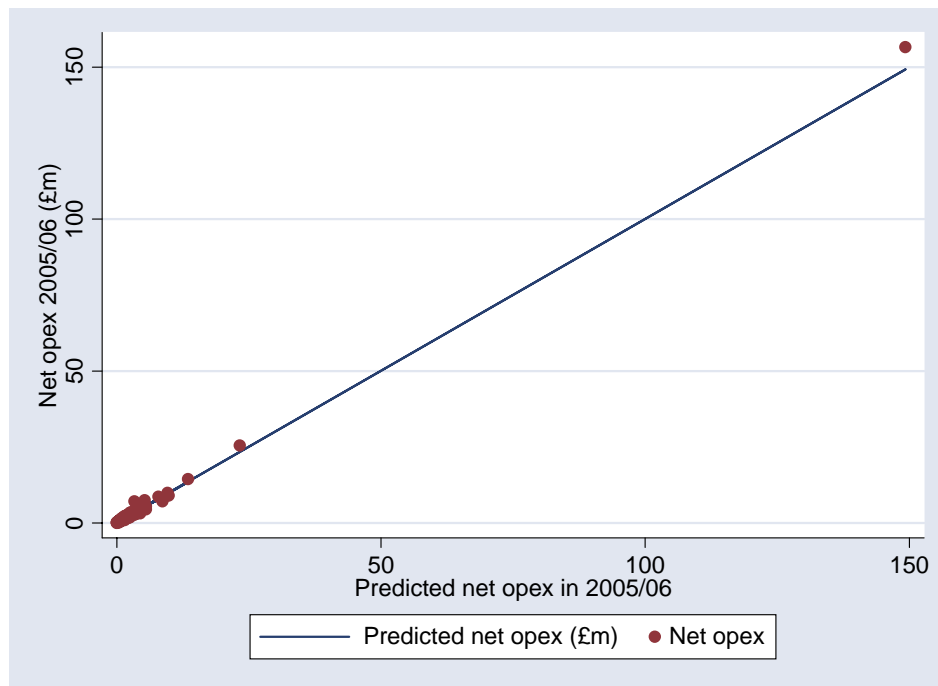
The prediction (solid line) and residuals for 2005/06 are shown in Figure 4.1. The dots represent actual costs. Those that are above the line represent RSLs that have actual costs higher than predicted, and vice versa. We have used data for the three years to estimate the cost relationship, and used this relationship to predict RSLs' 2005/06 costs.

---

<sup>13</sup> For example, the proportion of houses is taken as the point of reference and is not included in the equation.



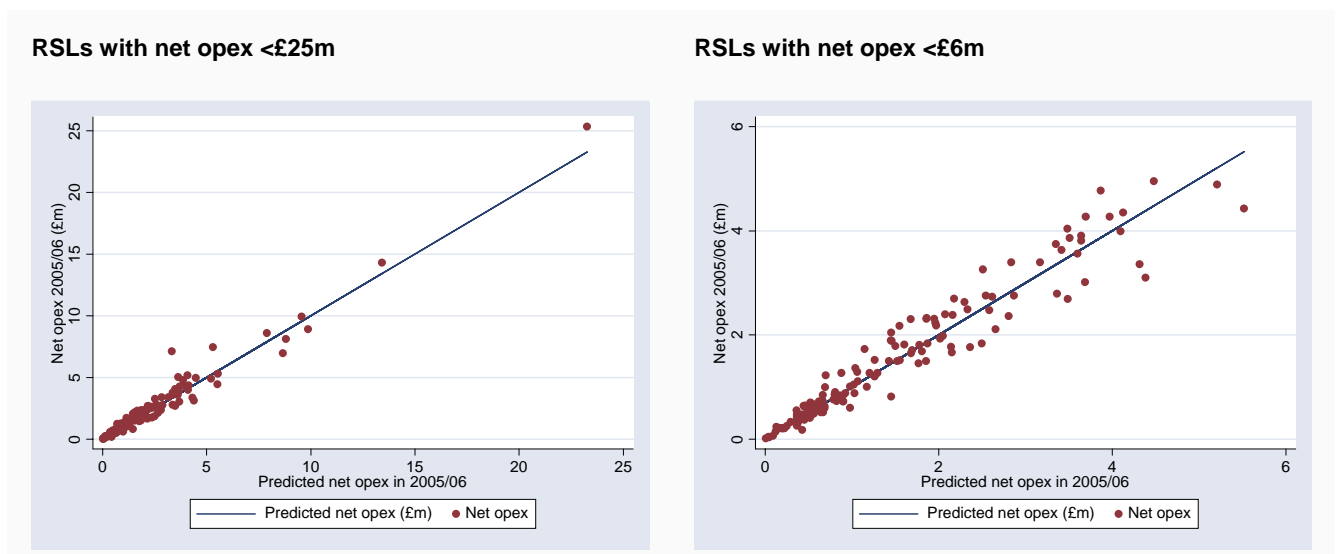
**Figure 4.1. Actual versus predicted costs for 2005/06: All RSLs**



In Figure 4.1 we see that the RSL with £150 million predicted net operating costs in 2005/06 has actual net operating costs just above what is predicted.

For rest of the sector, we need to “zoom in” and consider the RSLs with lower net operating costs. Figure 4.2 shows close ups of the distribution of RSLs’ costs around the predicted cost line. We note that the scattered dots are evenly spread around the line.

**Figure 4.2. Actual versus predicted net operating costs for 2005/06**





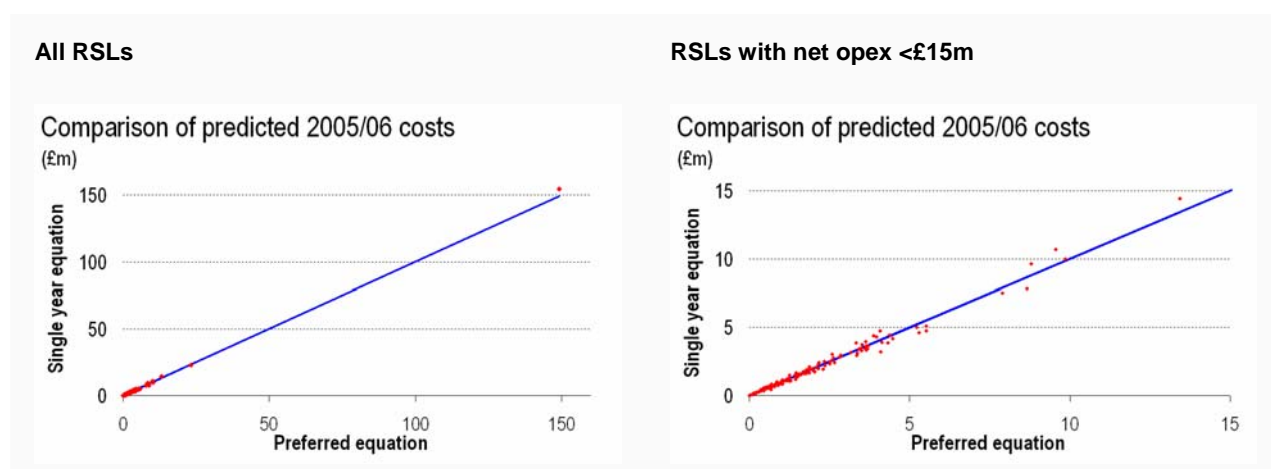
### 4.2.1 Tests on preferred specification

If we have included all the relevant cost drivers and the functional form is correct, the residuals from the relationship will show no systematic pattern and there should be no correlation between the residuals and any of the cost drivers. In this section we describe the tests of alternative specifications that lead to the preferred specification.

#### Single year equation

We test whether the data support the hypothesis that the cost relationship is unchanged from 2003/04 to 2005/06 and that we can pool the data. Figure 4.3 compares the 2005/06 predicted costs from the preferred equation with those from the single year equation using 2005/06 data.

Figure 4.3. Comparison of predicted costs: preferred versus 2005/06 equation



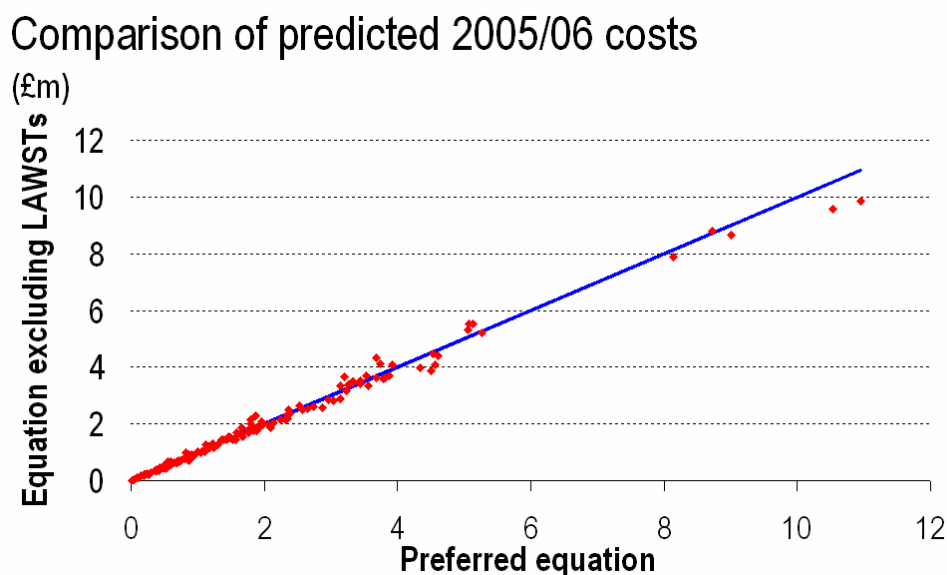
We observe that for most RSLs, the predicted cost from the single year equation is similar to that obtained from the preferred equation. In the right hand side of Figure 4.3, we focus on RSLs with operating costs of less than £15 million in 2005/06. These results indicate that while the single year approach is valid, where the pooling tests indicate that it is valid to combine the years together, we should do so. This is because it allows us to use all the information available to understand the way RSLs' costs move over time and begin to ask questions when observed costs vary significantly without corresponding changes in the cost drivers.

#### Local Authority Whole Stock Transfers

We have tested whether the estimated coefficients and therefore the predicted costs and residuals for RSLs in general are significantly affected by the inclusion of the three LAWSTs. We find that there is little change in the predicted costs, confirming that the LAWSTs do not have an undue influence on the estimated relationship. Figure 4.4 illustrates.



Figure 4.4. Comparison between predicted costs: Final equation versus equation that excludes LAWSTs



#### Other potential cost drivers

Not all of the items on the list of potential cost drivers have been included in the preferred equation. The following were not statistically significant and were excluded from the preferred equation

- index of stock concentration
- number of local authorities where units are located
- proportion of units that are sheltered or specially adapted
- tenancies managed
- houses factored, and
- voids.

#### Interaction variables

The assumption made in the estimated cost relationship is that the cost drivers affect RSLs' net operating costs in the same way regardless of the RSL's type. To test this, we added interaction variables and tested whether they were jointly significant in the pooled model. For both the LSVTs and the LAWSTs, they were not and the data support pooling over the RSL types.

### 4.3 Illustrative analysis

Using the model we have estimated, we can analyse the effect of changes in various parameters. The process is to compare actual costs with those predicted by the preferred equation.

#### 4.3.1 Comparator groups

We consider two types of RSLs: general needs and LSVT and group each type according to the relative size (percentile ranges) of the RSLs. Statistics for the type/size combinations are in Table 4.2.

**Table 4.2. Characteristics of general needs RSLs and LSVTs**

	General needs RSLs			LSVT		
	Small (25 <sup>th</sup> percentile)	Medium (50 <sup>th</sup> )	Large (75 <sup>th</sup> percentile)	Small (25 <sup>th</sup> percentile)	Medium (50 <sup>th</sup> )	Large (75 <sup>th</sup> percentile)
# Housing units	303	544	1,210	611	948	1,662
# Shared units	0	16	50	0	0	0
# Non social units	1	2	14	5	98	191
% Sheltered units	1%	3%	17%	2%	2%	17%
% Specially adapted units	0%	0%	1%	0%	0%	2%
% High rises	0%	0%	0%	0%	0%	0%
% Tenement	61%	72%	55%	56%	31%	30%
% In a block	8%	0%	3%	0%	16%	13%
% Other	3%	5%	8%	2%	6%	1%
% House	28%	22%	34%	42%	48%	55%
% 1919-44	0%	0%	0%	0%	0%	6%
% 1945-64	2%	0%	12%	57%	62%	55%
% 1965-82	0%	2%	1%	14%	8%	37%
% Post 1982	42%	34%	34%	0%	5%	4%
% Pre 1919	56%	64%	54%	29%	25%	0%
% 2 bedroom	23%	36%	37%	13%	10%	14%
% 3 bedroom	41%	42%	44%	54%	44%	45%
% 4 bedroom	23%	17%	14%	26%	34%	33%
% > 5 bedroom	5%	3%	1%	4%	4%	3%
% 1 bedroom	8%	1%	3%	2%	8%	4%
% Transfer stock	0%	0%	0%	98%	92%	94%
Years since registration	15	15	23	11	8	9

We make the following observations on the characteristics of these RSLs:

- General needs RSLs are typically smaller than LSVTs in terms of the number of units managed.
- For both types, housing units are mainly self contained units, with negligible numbers of bed spaces and non-self contained units.
- General needs RSLs have been registered for longer than LSVTs.
- General needs RSLs have some shared units and few non social housing units, while the LSVTs tend to have no shared units and a larger number of non social housing units.
- The majority of RSLs' units are in tenements and houses, with less than 10% in total in high rises, in a block or other flats. This is true for both general needs RSLs and LSVTs.



- Most of the general needs RSLs' social housing stock was built before 1919, whilst for the LSVTs, the majority was built between 1945 and 1964 and some between 1965 and 1982.
- The majority of general needs RSLs' self contained stock are 2 or 3 bedroom units, whilst those of LSVTs are 3 or 4 bedroom units.
- Most general needs RSLs do not have transferred stock. LSVTs stock is not all transferred.

The total and unit costs for these type/size combinations are in Table 4.3.

**Table 4.3. Actual total and unit net operating costs of RSLs (2004/05 prices)**

	General needs RSLs			LSVT		
	Small	Medium	Large	Small	Medium	Large
<b>Actual total net operating costs (£m)</b>						
2003/04	0.46	0.97	1.84	0.96	1.83	3.16
2004/05	0.54	1.03	2.02	0.94	1.95	3.30
2005/06	0.49	1.22	2.10	1.07	2.00	3.46
<b>Actual unit net operating costs (£/unit)</b>						
2003/04	1,521	1,780	1,519	1,579	1,931	1,899
2004/05	1,771	1,886	1,667	1,542	2,061	1,987
2005/06	1,632	2,246	1,738	1,749	2,112	2,085

We analyse the following.

- Comparison of general needs and LSVT costs.
- The effect of having additional transferred stock, i.e. scale effects. We assume that the additional stock does not change the proportions of units in different housing types, ages and sizes, so that the only changes we consider are in the number of housing units and the subsequent change in the proportion of transferred stock in an RSL's portfolio.
- The effect of mergers.

### 4.3.2 General needs and LSVTs

Using the preferred equation, we can estimate the total and unit costs for each of these type/size combinations, shown in Table 4.4.

**Table 4.4. Predicted total and unit net operating costs of RSLs (2004/05 prices)**

	General needs RSLs			LSVT		
	Small	Medium	Large	Small	Medium	Large
<b>Predicted total costs (£m)</b>						
2003/04	0.45	0.92	2.10	0.95	1.62	2.85
2004/05	0.47	0.96	2.19	0.97	1.66	2.94
2005/06	0.49	1.00	2.29	1.01	1.71	3.03
<b>Predicted unit net operating costs (£/unit)</b>						
2003/04	1,500	1,698	1,733	1,549	1,710	1,717
2004/05	1,556	1,769	1,812	1,597	1,757	1,768
2005/06	1,615	1,846	1,896	1,648	1,807	1,821

When we compare this with actual costs we note the following.

- For large general needs RSLs, actual costs are lower than predicted for all three years.
- For small general needs RSLs, actual costs are higher than predicted for all three years.
- For medium sized general needs RSLs, actual costs are lower in 2003/04 but higher in 2004/05 and 2005/06
- Actual costs are higher than predicted for medium and large LSVTs, even after taking into account the different proportions of transferred stock in their portfolio.
- The predicted costs of LSVTs are higher than those of general needs RSLs in all size type combinations.

A possible explanation for the difference between costs of general needs RSLs and LSVTs is the state of repair in the transferred stock managed by LSVTs. We understand from previous studies that transferred stock tends to be in need of repair. Without taking into account the quality indicators in the estimated relationship, we cannot determine whether or not LSVT costs are higher because of a higher requirement for repairs and maintenance. The results of the single year analysis indicates that there is some correlation between costs and quality compliance, and the statistics show that LSVTs, on average have a higher failure rate than general needs RSLs.

**Table 4.5. Scottish Housing Quality Standard failure rates (2005/06)**

	Number	Average	25 <sup>th</sup> percentile	50 <sup>th</sup> percentile	75 <sup>th</sup> percentile
LSVTs	30	43%	7%	44%	75%
General needs RSLs	122	16%	0%	5%	22%

In future years when more housing quality data become available for inclusion in panel data analysis, it may be possible extract the variation in costs that is caused by repairs related to compliance with the standard, and see whether the difference between general needs RSL and LSVT costs still remains.



### 4.3.3 Effect of transferred stock

The new variable we use in this study is the percentage of transferred stock in an RSL's portfolio, and this replaces the previous LSVT dummy variable.

We know from the estimated coefficients that a 1% increase in stock would lead to a 0.876% increase in costs, assuming it does not change the overall proportion of stock that are in different locations, of different ages and sizes. In this example, we assume that RSLs have a 10% increase in transferred stock for each of these size types and calculate what the costs would be for 2005/06. We compare these costs against the 2005/06 costs in Table 4.6.

**Table 4.6. Predicted net operating costs of RSLs for 2005/06 (2004/05 prices)**

	General needs RSLs			LSVT		
	Small	Medium	Large	Small	Medium	Large
Predicted costs (£m)	0.55	1.12	2.56	1.09	1.86	3.29
Predicted unit net opex (£/unit)	1,799	2,058	2,113	1,791	1,967	1,982
% increase in costs	11.4%	11.5%	11.5%	8.7%	8.8%	8.8%

We see that for general needs RSLs, taking on an additional 10% of stock transferred from public bodies, costs would increase proportionately more. This is the reverse for LSVTs that already have existing stock that are transferred from public bodies.

### 4.3.4 The effect of merger

There were three mergers involving general needs RSLs in 2005/06. Table 4.7 shows

- actual and predicted costs for the component RSLs of each merger for 2004 and 2005 and a subtotal of these costs
- actual and predicted costs for the merged entity for 2006.

In two of the three cases, the aggregate of pre-merger costs was greater than the comparable aggregate predicted cost. For these RSLs, the post-merger costs were lower than predicted, resulting in a negative residual.

These results illustrate an approach to exploring the effect of mergers. It is not possible to draw conclusions from a sample of only three mergers over a short period but a more robust analysis should be possible with a larger number of merger examples and more years of data.



Table 4.7. Comparison of pre- and post-merger predicted costs

ID	Name	Year	Net operating costs in 2005/06 prices (£m)	Predicted costs (£m)	Residual (£m)
81	Castle Rock	2004	5.26	5.52	-0.26
106	Edinvar	2004	2.30		= 7.56
81	Castle Rock	2005	7.00	5.71	1.29
106	Edinvar	2005	2.97		= 9.97
106	Merged Castle Rock Edinvar	2006	7.83	8.80	-0.97
77	Lorne Area	2004	0.83	0.86	-0.02
170	Port of Leith	2004	2.53		= 3.36
77	Lorne Area	2005	0.93	0.90	0.03
170	Port of Leith	2005	2.71		= 3.64
170	Merged Port of Leith	2006	2.99	4.38	-1.39
111	Dunedin	2004	3.68	2.53	1.16
116	Canmore	2004	3.98		= 7.66
111	Dunedin	2005	3.89	2.60	1.30
116	Canmore	2005	3.91		= 7.80
116	Merged Dunedin Canmore	2006	8.30	7.88	0.41



## 5 Conclusions

We have extended the OCI to include 2005/06 data and explored the implications of including additional variables for which data are now available.

The estimated relationship is similar to that estimated in the August 2006 study with three additions

- proportion of transferred stock instead of the simple LSVT dummy variable
- time trend to evaluate the change in RSLs' costs over and above inflation
- proportions of sheltered and specially adapted units

The relationship appears robust over time. The coefficients of the previously significant variables remain significant and similar in magnitude. It is not surprising that some of the coefficients that were insignificant (i.e. not statistically different from zero) have changed from being positive in the last report to being negative in this, and vice versa.

In this study we continued with the panel data model and also have used single year analysis to explore the significance of new variables. The advantages of adopting a single year cross-section analysis are that it allows flexibility to include additional cost drivers and it is not made more difficult by the occurrence of mergers and de-registrations. A panel data approach, on the other hand, makes better use of the available information and provides a simple way of analysing changes over time. We believe the panel data model continues to be appropriate.



## Appendix A: Table of results

The table summarises the results based on the preferred cost relationship. The predicted net operating costs and residuals are rounded to the nearest £10,000, and predicted unit costs and residuals are rounded to the nearest £10.

**Table A.1. Predicted costs of RSLs for 2005/06**

ID	RSL	Actual opex (£000)	Opex in 2004/05 prices (£000)	Predicted net opex (£000)	Predicted unit net opex (£/unit)	Residual (£000)	Unit residual (£/unit)
64	Albyn	4,274	4,120	3,700	2,020	580	320
65	Angus	2,790	2,690	3,360	2,190	-570	-370
70	Barrhead	1,496	1,442	1,520	2,000	-30	-30
73	Blythswood	175	168	430	2,780	-250	-1,640
76	Broomhouse (1986)	234	226	130	1,350	100	1,060
80	Calvay	883	851	1,030	2,360	-150	-340
83	Castlehill	3,809	3,672	3,640	2,550	170	120
84	Cassiltoun	729	702	820	1,860	-90	-210
85	Cathcart & District	596	575	690	2,450	-90	-320
86	Clydebank	1,773	1,709	2,140	2,040	-370	-350
87	Govan	2,759	2,659	2,540	2,100	220	180
88	Charing Cross	1,266	1,220	1,200	2,110	60	110
90	Home in Scotland	3,992	3,848	4,100	2,090	-100	-50
91	Cloch	2,181	2,102	1,970	1,810	210	190
93	Clydesdale	1,521	1,467	1,260	1,890	260	390
95	Craigdale	513	494	670	1,940	-150	-450
98	Dalmuir Park	1,890	1,821	1,460	2,150	430	630
99	Ferguslie Park	1,728	1,666	1,150	1,470	580	740
103	East Lothian	1,931	1,862	2,020	2,130	-90	-90
104	Wishaw and District	1,648	1,588	1,680	2,060	-30	-40
106	Castle Rock Edinvar	8,121	7,828	8,800	1,720	-680	-130
107	Eildon	4,349	4,192	4,130	2,320	220	120
108	Elderpark	1,761	1,697	2,360	2,160	-600	-550
110	Forth	742	716	660	1,380	80	170
112	Four Walls	41	39	40	3,600	0	90
114	Arklet	707	682	670	3,330	40	200



ID	RSL	Actual opex (£000)	Opex in 2004/05 prices (£000)	Predicted net opex (£000)	Predicted unit net opex (£/unit)	Residual (£000)	Unit residual (£/unit)
116	Dunedin Canmore	8,608	8,297	7,880	2,160	720	200
117	Govanhill	3,394	3,272	3,170	1,900	230	140
120	Grampian	5,164	4,978	4,080	1,840	1,080	490
125	Hillcrest	8,898	8,577	9,850	1,950	-960	-190
126	Glasgow West	1,784	1,719	1,500	2,030	280	380
127	Hjaltland	756	729	790	2,420	-30	-100
128	Horizon	1,266	1,221	1,300	1,830	-30	-40
129	Fairfield	513	494	640	1,990	-130	-400
130	Hunters Hall	473	455	420	2,080	60	270
137	Oak Tree	2,489	2,399	2,330	1,590	160	110
142	Kingdom	7,098	6,842	3,330	1,400	3,770	1,580
145	Langstane	4,771	4,599	3,870	1,680	900	390
148	Link Group	9,935	9,577	9,560	1,670	380	70
149	Linthouse	3,257	3,139	2,510	2,190	750	650
150	Lister	367	354	450	2,460	-90	-460
151	Lochaber	719	693	900	1,850	-180	-370
153	Loreburn	3,748	3,613	3,350	1,870	400	220
159	Maryhill	1,706	1,645	1,700	2,100	10	10
161	Milnbank	1,457	1,405	1,770	490	-310	-80
162	New Lanark	144	138	120	2,630	30	560
163	West Highland	602	580	980	1,990	-380	-770
164	Orkney	836	806	900	2,020	-60	-130
166	Paisley South	2,393	2,307	2,080	1,860	320	280
167	Parkhead	2,112	2,036	2,650	2,080	-540	-420
168	Partick	3,904	3,763	3,640	2,140	260	150
170	Port of Leith	3,101	2,989	4,380	2,160	-1,280	-630
171	Hawthorn	409	395	520	1,640	-110	-340
172	Queens Cross	7,437	7,168	5,300	2,520	2,130	1,010
173	Reidvale	2,309	2,226	1,860	1,980	450	480
174	Rosehill	817	787	1,450	1,810	-630	-790
176	Rutherglen and Cambuslang	1,509	1,455	1,550	2,000	-40	-60



ID	RSL	Actual opex (£000)	Opex in 2004/05 prices (£000)	Predicted net opex (£000)	Predicted unit net opex (£/unit)	Residual (£000)	Unit residual (£/unit)
183	Shettleston	2,759	2,659	2,870	1,920	-110	-70
185	Drumchapel	1,227	1,183	700	1,460	530	1,100
186	Southside	1,897	1,828	1,450	2,730	450	840
187	North Glasgow	4,433	4,273	5,520	2,600	-1,090	-510
190	Strathclyde (Camphill)	204	196	220	4,200	-20	-350
193	Thenew	5,288	5,097	5,520	2,050	-240	-90
195	Cunninghame	3,395	3,273	2,840	1,580	560	310
197	Tollcross	1,815	1,749	1,600	2,370	210	320
201	West of Scotland	4,886	4,709	5,210	2,080	-320	-130
202	Lanarkshire	1,207	1,163	1,260	1,630	-50	-70
203	West Whitlawburn	1,290	1,244	1,060	1,950	230	420
204	Prospect Community	2,043	1,969	1,450	1,650	590	670
205	Whiteinch and Scotstoun	2,228	2,147	1,960	2,050	270	280
207	Williamsburgh	1,837	1,771	2,500	1,800	-660	-470
208	Yoker	1,114	1,074	1,070	1,960	40	80
209	Yorkhill (1988)	628	606	570	1,760	60	190
211	New Gorbals	2,383	2,297	2,160	1,900	220	200
212	Trafalgar	486	469	560	1,860	-70	-240
213	Ochil View	1,665	1,605	2,150	1,830	-490	-410
214	Gardeen	421	406	460	1,780	-40	-140
216	Blairtummock	635	612	640	1,580	-10	-20
219	Ardenglen	1,270	1,224	880	1,470	390	660
220	Cube	6,969	6,717	8,650	2,590	-1,680	-500
223	Blochairn	551	531	360	1,650	190	860
224	Perthshire	1,805	1,740	1,780	1,480	30	20
225	West Granton	474	457	470	1,560	10	20
226	Garrion Peoples	420	405	370	1,680	50	240
230	Cernach	614	592	490	1,570	130	410
231	Pineview	904	871	810	1,620	100	200
232	Rural Stirling	600	578	550	1,330	50	120
233	Copperworks	641	618	460	1,610	180	630



ID	RSL	Actual opex (£000)	Opex in 2004/05 prices (£000)	Predicted net opex (£000)	Predicted unit net opex (£/unit)	Residual (£000)	Unit residual (£/unit)
234	Kingsridge Cleddans	312	301	380	1,590	-70	-280
235	Bellsmyre	1,004	968	1,170	2,010	-170	-290
236	Ore Valley	720	694	620	1,520	100	240
237	Faifley	699	674	520	1,500	180	510
238	Easthall Park	633	610	690	1,470	-50	-120
241	Glen Oaks	2,361	2,275	2,810	2,020	-440	-320
242	Provanhall	454	438	440	1,460	20	50
245	Tenants First	2,635	2,540	2,300	1,950	340	290
246	Kincardine	208	200	180	2,180	30	320
251	Kendoon	259	250	370	1,320	-110	-390
258	South Uist	64	62	90	1,670	-30	-510
259	Cordale	552	532	560	1,550	-10	-20
260	Dunbritton	879	847	920	1,510	-40	-70
261	Muirhouse	615	593	590	1,290	30	60
263	North Harris	30	29	50	1,660	-20	-590
264	Beechwood	204	196	200	1,740	0	20
267	Glen	615	593	460	1,320	160	460
268	Lochfield Park	567	546	540	1,360	30	80
269	North View	1,361	1,312	1,040	1,500	320	470
270	Cadder	1,211	1,167	1,260	1,840	-50	-70
271	Forgewood	466	449	370	1,680	100	450
272	Barra and Vatersay	43	41	40	1,890	0	-20
273	Pentland	722	696	900	2,190	-180	-440
274	Molendinar Park	1,006	970	980	2,070	30	60
275	Abronhill	635	612	440	1,970	190	860
276	Berneray	8	8	10	1,290	0	-290
277	Ruchazie	328	316	290	1,680	40	200
279	East Kilbride and District	662	638	630	1,550	40	90
280	Irvine	3,012	2,903	3,680	1,820	-670	-330
281	Wellhouse	537	518	520	1,310	10	40
282	Muirneag	205	198	200	1,280	0	10



ID	RSL	Actual opex (£000)	Opex in 2004/05 prices (£000)	Predicted net opex (£000)	Predicted unit net opex (£/unit)	Residual (£000)	Unit residual (£/unit)
284	Manor Estates	2,305	2,222	1,680	1,850	630	690
285	Almond	3,361	3,240	4,320	1,830	-960	-410
286	Melville	3,557	3,429	3,600	2,220	-40	-30
289	Berwickshire	3,631	3,500	3,420	1,970	210	120
291	Clyde Valley	4,273	4,118	3,970	1,620	300	120
293	Larkfield	1,046	1,008	1,020	2,410	20	60
295	Fife Special	4,952	4,773	4,480	1,900	470	200
296	Shire	1,977	1,905	2,050	2,280	-70	-80
297	Abertay	4,045	3,899	3,480	1,960	560	320
298	Paragon	2,735	2,636	2,620	1,850	120	80
299	Linstone	3,862	3,723	3,510	2,170	350	220
300	Knowes	2,322	2,238	1,860	1,820	460	450
301	Bridgewater	1,833	1,767	1,870	2,680	-40	-50
302	Sanctuary (Scotland)	2,691	2,594	3,490	1,610	-800	-370
304	Ayrshire	1,501	1,447	1,850	1,800	-350	-340
305	Atrium Homes	1,686	1,625	1,800	2,220	-120	-150
306	ANCHO	1,498	1,444	1,430	2,030	70	100
307	Antonine	858	827	800	2,320	60	160
309	Cumbernauld	5,024	4,843	3,630	1,980	1,390	760
310	Thistle	2,313	2,229	1,950	2,080	370	390
311	Homes for Life	332	320	360	1,220	-20	-80
312	Aberdeenshire	999	963	690	1,110	310	500
313	Scottish Borders	14,310	13,794	13,420	2,170	890	140
314	Port Glasgow	665	641	570	2,280	90	360
315	Dumfries and Galloway	25,322	24,409	23,270	2,140	2,050	190
317	Glasgow	156,523	150,878	149,300	2,010	7,220	100
318	West Lothian	246	238	250	1,090	-10	-20
319	Capital City Homes	211	203	160	1,720	50	520
321	Fyne Homes	2,695	2,598	2,180	1,700	520	410
323	Spire View	835	805	860	1,570	-30	-50
324	Lochalsh & Skye	843	813	660	1,690	180	460



ID	RSL	Actual opex (£000)	Opex in 2004/05 prices (£000)	Predicted net opex (£000)	Predicted unit net opex (£/unit)	Residual (£000)	Unit residual (£/unit)
325	Moray HP	433	418	440	1,640	0	-10
326	Hillhead 2000	2,175	2,096	1,550	1,820	620	730
327	Waverley	2,473	2,384	2,580	1,650	-110	-70



## Appendix B: Allowing for inflation

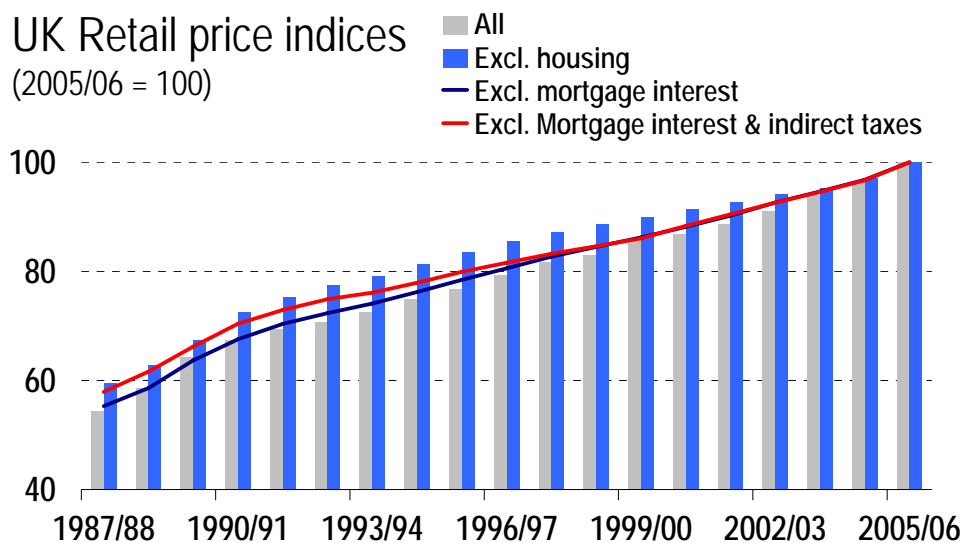
In order to compare costs over time, we need to extract the effects of inflation. To convert RSLs' reported annual operating costs into "real" terms, we need to divide through by an appropriate price index.

$$\text{Operating costs in constant prices} = \frac{\text{Operating costs in nominal prices} * 100}{\text{Price index}}$$

The most common price index used is the Retail Price Index (RPI). It is an indicator of inflation and measures the average change from month to month in the prices of goods and services purchased. The statistics published by the Office of National Statistics<sup>14</sup> is only available for the United Kingdom as a whole, and does not provide regional breakdowns of price indices. It is the most general purpose domestic measure of inflation and is used by the Government uses it for uprating of pensions, benefits and index-linked gilts and by the private sector for uprating of maintenance payments and housing rents.

There are different measures of retail prices, some excluding housing costs and interest. Figure B.1 show that the differences between the four measures have shown very little variation since 2002/03. As such, we use the all items RPI data to deflate RSLs' operating costs into constant prices and we have rebased operating costs to 2004/05 prices in line with the previous study.

Figure B.1. UK retail price indices, 1987/88 to 2005/06



Source: Office of National Statistics

<sup>14</sup> <http://www.statistics.gov.uk/rpi>



## Appendix C: Balanced and unbalanced panels

A data set containing observations on a single unit (i.e. an RSL) which is recorded over many time periods is referred to as a time series. A data set containing observations on a sample of RSLs which are observed for a single period is referred to as a cross-section. A dataset which records the values of one or more variables (i.e. costs) for a sample of RSLs which are observed over time is referred to a panel.

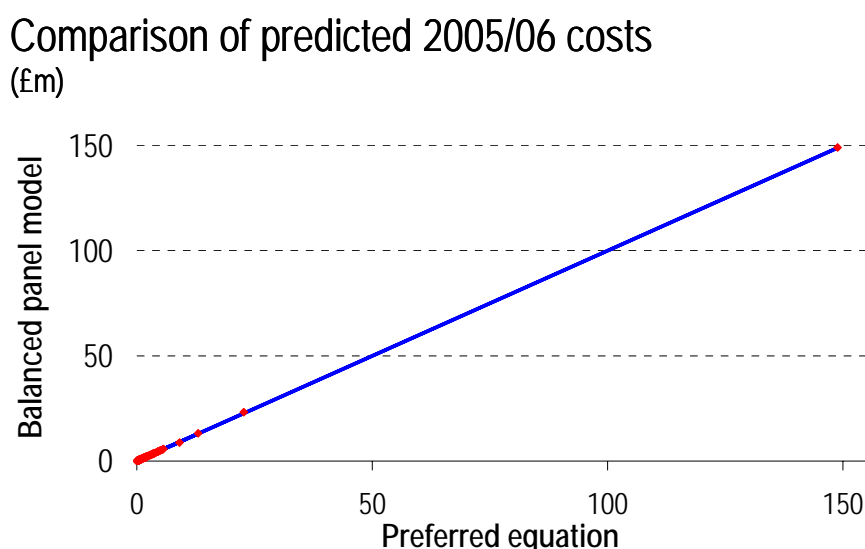
We can broadly differentiate between two types of panel data. A balanced panel contains information on the same number of units for all time periods. In the case of an unbalanced panel, the number of units observed in each period may differ. The causes of this vary according to the application. For example, is a panel where the unit of observation is the individual may be unbalanced if after a number of time periods certain individuals do not respond to the survey questions. In the context of the current study, merger activity, as outlined in Section **Error! Reference source not found.** will generate an unbalanced panel.

The consequences of working with an unbalanced panel are manifest in that a number of formulae for calculating standard errors of estimated parameters and test statistics need to be modified to account for the fact that not all units are observed in each time period. In the case of most software packages, including the one we use for this analysis, the adjustments to the requisite statistics are automatic once the user indicates that the panel is unbalanced.

We provide a comparison of the results when we adopt balanced versus unbalanced panels based on the mergers that occurred in 2005/06

- Equation 1: Using 2003/04 and 2004/05 data with pre-merger RSLs, i.e. with 153 RSLs
- Equation 2: Using 2003/04 and 2004/05 data but with the backtrack mergers, i.e. with 150 RSLs

Figure C.1. Comparison of 2005/06 predicted costs



The solid line represents cases where the predicted costs from the final equation are the same as those from the balanced panel model. The red dots represent estimates for individual RSLs, dots below the line represent those RSLs whose predicted costs from the final equation are higher than the estimates from the balanced panel model and vice versa. We note that most of the estimates line roughly on the solid line, with no systematic pattern of being above or below the line.

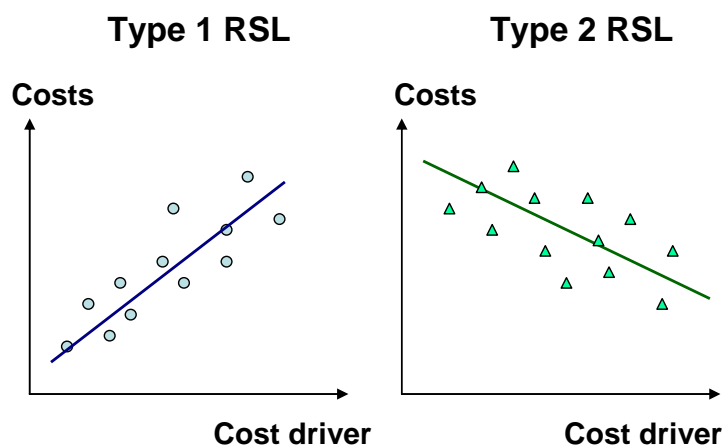


## Appendix D: Pooling tests

Due to the different types of RSLs in the social housing sector, we want to consider whether their characteristics per se have any impact on the costs of providing services to tenants. In estimating a single equation for all RSLs we are assuming that the same cost relationship holds for each RSL type. This implies that the cost drivers for general needs RSLs are the same as those of LSVTs and the impacts of these cost drivers on operating costs are the same for every cost driver. These are strong assumptions and their validity needs to be tested.

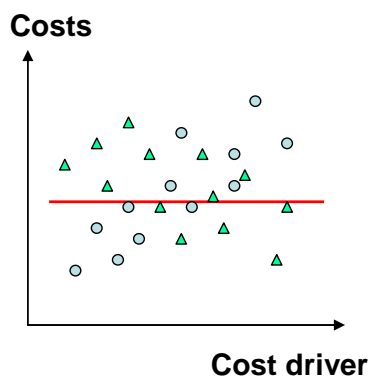
Note that the impact of erroneously pooling RSLs of different types in a single cost relationship is that the estimated equation will be mis-specified, and the errors carry over into the predicted costs and residuals. For example, if the underlying relationship between RSLs of two types are as follows:

Figure D.1. Illustrative cost relationship for two different RSL types



However, if we merge the data from the two RSL types together and estimate a single equation without taking their individual characteristics into account, we may end up with a result where the cost driver does not appear to have any impact on RSLs' costs, i.e. we estimate a horizontal line where the coefficient is statistically equivalent to zero.

Figure D.2. Illustrative cost relationship for a merged dataset



In order to avoid making such errors, we test the validity of pooling before adopting a panel data approach. There are two equivalent ways of testing this:



- An F-test on the residuals. We estimate the cost function separately for general needs RSLs and for LSVTs and evaluate the degree to which these separate regressions are statistically better than a single regression with data from the two types.
- A dummy variable test. We estimate a single regression but augment the cost function with a set of interaction and intercept dummy variables. The hypothesis that the data can be pooled then translates into a test of whether the set of additional parameters are jointly equal to zero.

These are described in turn.

## D.1 F-test on residuals

The first step to this test is to see whether the impacts of the cost drivers are the same across the two types of RSLs or not, i.e. whether the data from the two types of RSLs can be pooled or not. We utilise an F-test to determine whether the coefficients in a regression model are the same in sub samples  $S_1$  and  $S_2$  where sub sample  $S_1$  has  $n_1$  observations and  $S_2$  has  $n_2$  observations. To carry out this test we run three regressions that differ only in terms of the number of observations included:

- Equation 1: Regression with all the observations,  $n_1 + n_2$ . We obtain the residual sum of squares, RSS.
- Equation 2: Regression with only  $n_1$  observations. We obtain the residual sum of squares  $RSS_1$ .
- Equation 3: Regression with only  $n_2$  observations and obtain the residual sum of squares  $RSS_2$ .

The regression sum of squares is the difference between the total sum of squares and the residual sum of squares. In the case of our estimated cost relationship, the total sum of squares is the total amount of variability in the net operating costs of RSLs and the residual sum of squares is the variability that still cannot be accounted for after the regression model is fitted. The regression sum of squares is the amount of variability in the response that is accounted for by the regression model. By comparing the residual sum of squares of different models we can see which one has the ability to explain more of the variation in costs.

Note that equation 1 is a restricted version of equations 2 and 3 combined where the estimated coefficients are restricted to be the same across the two samples. If sub samples  $S_1$  and  $S_2$  are indeed different, the combination of equations 2 and 3 is a less restrictive specification and therefore one would expect them to be able to explain more of the variation of the observations. Equation 1 would essentially “average out” the effects of the cost drivers across the two samples, and the residuals RSS would be large compared to the sum of  $RSS_1$  and  $RSS_2$ . On the other hand, if the behaviour of observations in the two samples is similar, then equations 2 and 3 will still be able to explain more than equation 1 alone, although the difference between them might be statistically insignificant. To test the significance of the difference, we use the F statistic, expressed as:

$$F_{k, n_1+n_2-2k} \sim \frac{RSS - \{RSS_1 + RSS_2\}}{k} \bigg/ \frac{RSS_1 + RSS_2}{n_1 + n_2 - 2k}$$

If the calculated value is greater than the F statistic, then we reject the hypothesis that the coefficients of the cost drivers from equations 2 and 3 are the same.

## D.2 Dummy variable test

We create a set of intercept and interaction dummy variables. An intercept dummy variable for LSVTs takes a value 1 for RSLs 1, 6, 7 and 8, and 0 for the others; the “General needs” dummy variable gives the opposite values. If we sum these two dummy variables we get a column of 1s. For this



reason we either include a constant and one of the dummy variables (the other being the *reference category*), or we exclude the constant and include both the dummy variables.

**Table B.2. Illustration of a dummy variable**

RSL	Constant	Units	Classification	LSVT	General needs
RSL1	1	25,620	LSVT	1	0
RSL2	1	2,321	General needs	0	1
RSL3	1	569	General needs	0	1
RSL4	1	6,582	LSVT	1	0
RSL5	1	2,541	General needs	0	1
RSL6	1	1,593	LSVT	1	0
RSL7	1	985	LSVT	1	0
RSL8	1	745	LSVT	1	0
RSL9	1	689	General needs	0	1

*Interaction dummy variables* are created by multiplying the RSL type by each cost driver. Table 3.2 provides an example of an interaction dummy for the number of units.

**Table B.3. Illustration of an interaction dummy variable**

RSL	Constant	Units	Classification	LSVT	General needs
RSL1	1	25,620	LSVT	1	25,620
RSL2	1	2,321	General needs	0	0
RSL3	1	569	General needs	0	0
RSL4	1	6,582	LSVT	1	6,582
RSL5	1	2,541	General needs	0	0
RSL6	1	1,593	LSVT	1	1,593
RSL7	1	985	LSVT	1	985
RSL8	1	745	LSVT	1	745
RSL9	1	689	General needs	0	0

We create similar intercept and interaction dummy variables for the three local authority whole stock transfers (LAWST) RSLs in the same way.

We estimate two cost functions that include all general needs RSL, LSVTs and the three whole stock transfer RSLs with  $k$  cost drivers:

- A simple cost function that does not distinguish between the different types of RSLs

$$\text{Costs} = \alpha + \beta_1 * \text{Cost driver}_1 + \dots + \beta_k * \text{Cost driver}_k$$



- An “augmented” cost function that includes all the intercept and interaction dummy variables. In effect, we have an equation that has an intercept, 2 dummy variables and  $3 \cdot k$  cost drivers

$$\begin{aligned} \text{Costs} = & \alpha + \beta_1 \cdot \text{Cost driver}_1 + \dots + \beta_k \cdot \text{Cost driver}_k \\ & + \alpha_1 \cdot \text{LSVT} + \gamma_1 \cdot \text{LSVT interaction dummy variable}_1 + \dots + \gamma_k \cdot \text{LSVT interaction dummy variable}_k \\ & + \alpha_2 \cdot \text{LAWST} + \delta_1 \cdot \text{LAWST interaction dummy variable}_1 + \dots + \delta_k \cdot \text{LAWST interaction dummy variable}_k \end{aligned}$$

We then test whether  $\alpha_1 = \gamma_1 = \dots = \gamma_k = 0$  and  $\alpha_2 = \delta_1 = \dots = \delta_k = 0$ . If we reject this hypothesis, it means that we would not be able to adopt the simple cost function to describe the costs of general needs RSLs, LSVTs and LAWSTs. However, if the augmented cost function satisfies the underlying assumptions of the statistical analysis (e.g. constant variance in error term, no evidence of excluded explanatory variables etc) then we can still work with the augmented equation for these RSLs. We can then simplify the augmented equation by iteratively excluding the interaction dummy variables that not statistically different from zero. In our preferred equation, we find that we do not need to include the interaction dummy variables for both the LSVTs and LAWSTs, i.e. we find that  $\gamma_1 = \dots = \gamma_k = \delta_1 = \dots = \delta_k = 0$ .



## Appendix E: Glossary

The glossary explains the econometric terms used in the report. It does not repeat those definitions that are included in the text, for example of the various statistical tests.

Coefficient	A value expressing the extent to which the <i>dependent variable</i> changes with respect to changes in an independent variable. This value is estimated using regression analysis.
Average	The simplest measure of the central tendency of a set or list of data, and refers to a measure of the "middle" of the data set. The most common method, and the one generally referred to simply as the average, is the arithmetic <i>mean</i> , which is the sum of all the members of the list divided by the number of items in the list.
ARCS	Communities Scotland's Accounts Review Calculation System (ARCS), which holds data from the published accounts of RSLs.
Binary variable	Also known as a dummy variable. It marks or encodes a particular attribute and has the value zero or one for each observation, e.g. 1 for male and 0 for female. It is included in a regression where one expects the effect on the dependent variable is different for different types of observations.  If a constant term is included in the regression, it is important to exclude one of the dummy variables from the regression, making this the base category against which the others are assessed. If all the dummy variables are included, their sum is equal to the constant term, resulting in perfect <i>multicollinearity</i> .
Coefficient of determination	Usually referred to as $R^2$ (R squared), and is the proportion of a sample variance of a response variable that is "explained" by the <i>explanatory variables</i> when a <i>linear regression</i> is done. It is a measure of how well the statistical relationship "explains" the variation in the <i>dependent variable</i> .
Conditional mean	The expected value of a random variable, adjusted to take into account values of other variables that are thought to affect it.
Cost drivers	A factor related to an activity that changes the volume or characteristics of that activity, and in doing so changes its costs. Activities can have multiple cost drivers.
Cost function or equation	The relationship, expressed as an equation, between a cost and a one or more cost drivers. In choosing a cost function both economic plausibility and goodness fit are relevant.
Cross section data	A type of one-dimensional data set. Cross section data refers to data collected by observing many subjects (such as individuals, firms or countries/regions) at the same point of time, or without regard to differences in time. Analysis of cross sectional data usually consists of comparing the differences among the subjects.
Dependent variable	In regression, the variable whose values are to be explained by changes in the other variable (the independent or explanatory variables). In a cost function, the dependent variable is the cost itself.
Distribution	The pattern of variation of a variable is called its distribution, which can be described both mathematically and graphically. In essence, the distribution records all possible numerical values of a variable and how often each value occurs (its frequency). The most well known example of a distribution is the bell-shaped curve.
Elasticity	Percentage change in the dependent variable due to a percentage change in the explanatory variable.



Error	The difference between the actual costs of an RSL and the cost predicted by the econometric relationship. Synonymous with <i>residual</i> .
Expected value	The sum of all possible values for a random variable, each value multiplied by its probability of occurrence. For a symmetrically distributed random variable the <i>mean</i> is the same as the expected value and is the same as the <i>median</i> .
Explanatory variable	In regression, an explanatory variable (also known as a regressor or independent variable) is a variable which appears on the right hand side of the equation. Its function is to explain the evolution of the <i>dependent variable</i> . In a cost function, the dependent variables are the cost drivers.
Fitted values	The value of the <i>dependent variable</i> that is predicted by the regression analysis. It is the conditional <i>mean</i> or <i>expected value</i> of the <i>dependent variable</i> conditional on the values of the <i>explanatory variables</i> .
F test	The F test employs the statistic (F) to test various statistical hypotheses about the <i>mean</i> (or means) of the distributions from which a sample or a set of samples have been drawn. The t-test is a special form of the F test.
Fixed costs	Elements of cost that do not vary with the number of units the RSL has. In an econometric cost equation, the <i>intercept</i> is interpreted as the fixed cost.
Heteroskedastic and Homoskedastic	In statistics, a sequence or a vector of random variables is homoskedastic if all random variables in the sequence or vector have the same finite variance. This is also known as homogeneity of variance. The complement is called heteroskedasticity. The assumption of homoskedasticity simplifies mathematical and computational treatment and may lead to good estimation results (e.g. in data mining) even if the assumption is not true.
Intercept	The point at which the econometric relationship crosses the vertical axis when plotted in two dimensions. In a <i>cost function</i> , the intercept is interpreted as the fixed cost.
Kurtosis	Kurtosis is an attribute of a distribution, describing its 'peakedness'. Data sets with high kurtosis tend to have a distinct peak near the mean, decline rather rapidly, and have heavy tails. Data sets with low kurtosis tend to have a flat top near the mean rather than a sharp peak. A uniform distribution is the extreme case.
Linear regression	A regression analysis where the relation between the <i>dependent variable</i> and the <i>explanatory variables</i> is a linear function of the <i>coefficients</i> .
Logarithmic transformation	The transformation is the logarithm of an original number. The logarithm is the power to which a base, such as 10, must be raised to produce a given number. If $n^x = a$ , the logarithm of a, with n as the base, is x; symbolically, $\log_n(a) = x$ . For example, $10^3 = 1,000$ ; therefore, $\log_{10}(1,000) = 3$ . The natural logarithm is a special case, which is the logarithm to the base e, where e is equal to 2.718281828459... (continuing infinitely).
Mean	Technical term for the <i>arithmetic average</i> . For a set of observations it is calculated by adding up the observations and dividing by their number.
Median	In statistics, the middle value of a set of numbers or data points; half the figures will fall below the median and half above.



Multicollinearity	<p>Multicollinearity refers to linear inter-correlation among variables. If nominally "different" measures actually quantify the same phenomenon to a significant degree, i.e., different variables with numerical representations but correlate highly with each other, then one of them is redundant.</p> <p>A principal danger of such data redundancy is that of overfitting in regression analysis models. The best regression models are those in which the predictor variables each correlate highly with the dependent (outcome) variable but correlate at most only minimally with each other. Such a model is often called "low noise" and will be statistically robust (that is, it will predict reliably across numerous samples of variable sets drawn from the same statistical population).</p>
Ordinary least squares (OLS)	A method of estimating a relationship between a dependent variable and one or more explanatory variables which minimises the sum of the squares of the residuals
p value	The probability that the estimated value is statistically equivalent to zero. Typically one would accept a p value of up to 20 per cent as being evidence that the estimated value is significant in explaining the dependent variable.
Panel data	<p>Panel data is a two-dimensional data, for example, data on a population over time.</p> <p>A balanced panel exists where there are data for all members of the population for each year. If this is not the case, there is an unbalanced panel.</p>
Panel data analysis	Method of statistical analysis combining <i>cross section data</i> and <i>time series data</i>
Post-estimation tests	Statistical tests carried out on the residuals to make sure the OLS assumptions are valid.
Predicted costs	Same as <i>fitted costs</i> .
Quantile	A quantile is the value that corresponds to a specified proportion of an (ordered) sample of a population. For instance a very commonly used quantile is the <i>median</i> , which is equal to a proportion of 0.5 of the ordered data.
Quantile regression	Quantile Regression seeks to complement classical linear regression analysis. In <i>Ordinary Least Squares</i> (OLS) the primary goal is to determine the conditional mean of random variable Y, given some explanatory variables, reaching the expected value. Quantile Regression goes beyond this and enables one to pose such a question at any quantile of the conditional distribution function.
Regression analysis	A statistical technique used to find relationships between variables for the purpose of explaining the observed variation in these variables and predicting future values.
Residual	The residual is a positive or negative numeric difference between two numbers. The residual of a cost function is the difference between the actual costs and the cost predicted by the econometric relationship.
Skewness	Skewness is an attribute of a <i>distribution</i> describing its symmetry. A distribution that is symmetric around its mean has skewness of zero, and is 'not skewed'. The skewness of a normal distribution is zero and any symmetric data should have skewness near zero. Negative values for the skewness indicate data that are skewed left (i.e. the left tail is heavier than the right tail) and positive values for the skewness indicate data that are skewed right. Some measurements with a lower bound (e.g. of zero) are skewed to the right.
Standard deviation	Measure of the dispersion of observations around their <i>mean</i> , the higher the standard deviation the more disperse the observations.



Standard error	A common measure of the uncertainty associated with a numerical estimate. In a regression analysis, standard errors are often reported with the <i>coefficient</i> estimates. A large standard error suggests a wide range of possible values and therefore high level of uncertainty associated with the estimate.
Statistical inference	The process of making conclusions about a population using data collected from a sample.
Time series data	A time series is a sequence of data points, measured typically at successive times, spaced apart at uniform time intervals. Time series analysis comprises methods that attempt to understand such time series, often either to understand the underlying theory of the data points (where did they come from? what generated them?), or to make forecasts (predictions).
Unconditional mean	Same as <i>average</i> , and is sum of all the members of the list divided by the number of items in the list.
Univariate regression	A regression analysis where the dependent variable is determined by only one explanatory variable.
Variables	A series of comparable observations or characteristics of a phenomenon taken as a single set of data; a listing of specific characteristics of a population or a number of observations taken over a specific period of time which may reasonably be expected to vary from observation to observation. In a cost function, the behaviour of the costs is related to or explained by the explanatory variables (or cost drivers).