Comparing estimated heat loss with recorded energy usage

Dr John Paul Gosling[1]

14th August 2014

Overview

In this document, we investigate and formally analyse data on heating energy usage that has been collected by Logicor (GROUP) Ltd for the Logicor Clear Heating System. Throughout, we assume that we are interested in how the systems perform across all dwellings in the UK. We consider performance with respect to estimated heat energy loss. In the next section, we describe some of the assumptions in the heat loss estimates. In subsequent sections, we explore the dataset, perform a statistical analysis and check the robustness of our results.

Heat loss calculations

The heat loss calculations used in this study are performed using the method from an earlier pilot study [1]. The 'U' values and assumed number of air exchanges used in the calculations are given in the appendix. Two key assumptions used in the calculations are as follows: there is no heat loss between adjacent heated rooms and the rooms are unfurnished. The second assumption here can have an appreciable impact on heat loss given the insulation that furnishings provide.

Both internal and external temperatures are required to calculate the heat loss estimate. Temperatures were recorded throughout the property to arrive at an average internal temperature. The external temperature was also recorded and cross-checked with Met Office records.

The output from these calculations is an estimate of heat loss for each surveyed property on the day that the temperatures were recorded in kWh. These estimates provide us with a baseline from which to compare the actual energy usage of the Logicor Clear Heating Systems. Given their approximate nature and assumptions, we would expect that the actual usage is lower than the estimates heat energy lost. Using formal statistical techniques, we can say how much lower we expect the actual usage of the heating systems will be.

Recorded energy usage and the associated dataset

Fifty-three properties were surveyed and had their energy use recorded. Properties were selected at random by the Logicor Clear Heating System distributors contacting the property owners that had the system and booking appointments for the survey to be carried out. This random selection process seems reasonable if we want to make inferences about the entire UK housing stock provided that the properties that have this heating system are representative of UK housing. Figure 1 shows the breakdown of the properties by type. Comparing this breakdown with the 2008 housing stock profile [2], the number of flats and terraced houses seems to be consistent, but there is a bias in the data set towards bungalows and away from detached houses (9.4% bungalows in housing stock profile against 26.4% in the dataset, 17.4% detached houses in housing stock profile against 7.5%). However, the sample is so small that this is not enough to question the assumption of random selection.

[1] Lecturer in Statistics, School of Mathematics, University of Leeds, Leeds. johnpaulgosling@gmail.com

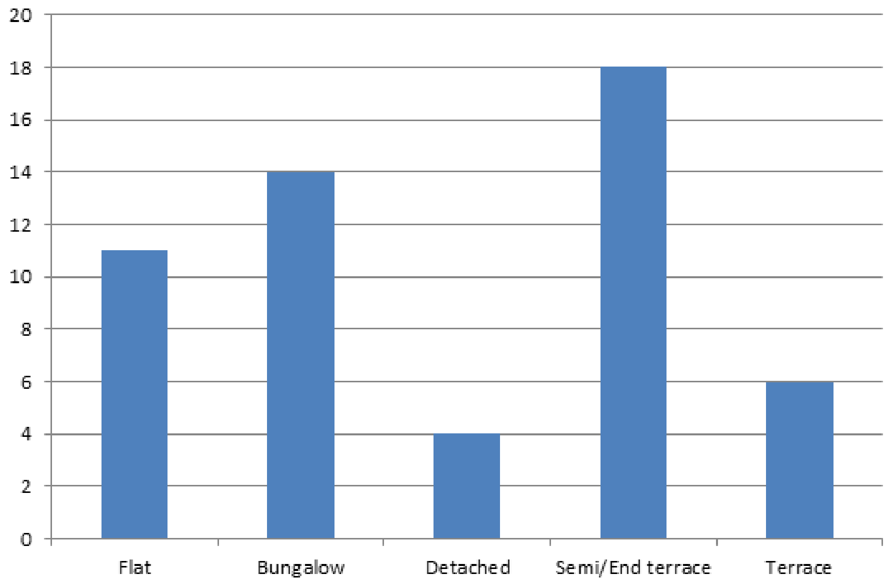

Figure 1: Number of properties by property type.

Figure 2 shows the breakdown of the properties by number of rooms. Again, the breakdown in the dataset seems to be consistent with the expected breakdown for the types of properties prevalent in the UK.

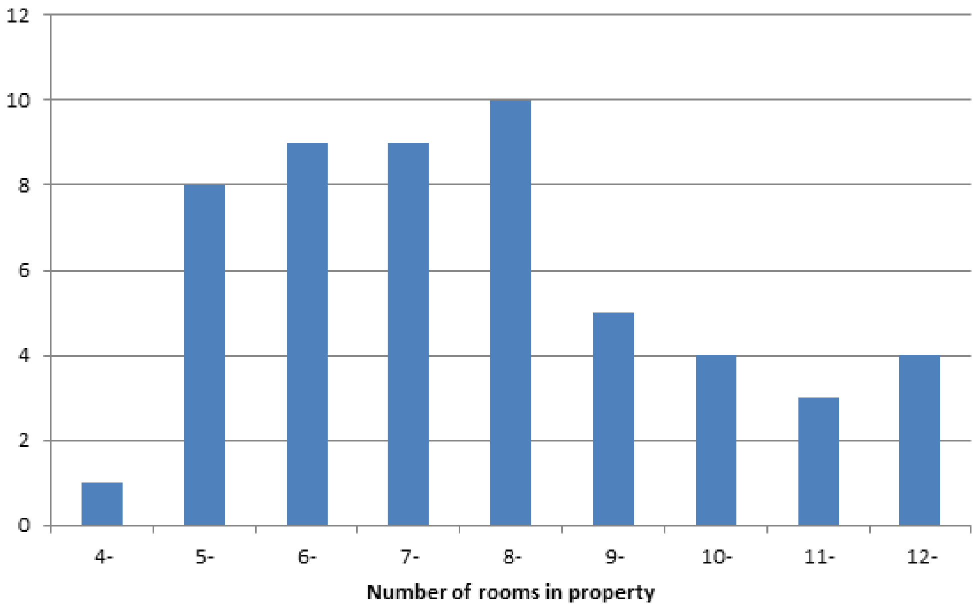

Figure 2: Number of properties by number of rooms.

The surveyed properties are customers of Logicor (Group) Ltd and, as such, are not spread across the UK. Figure 3 shows the location of the 53 surveyed properties. Given the climatic differences from the west to the east of the region covered, we can be reasonably confident that the results will carry for all but the properties facing the more extreme climatic conditions. Perhaps special cases might occur for properties on the south coast, in wet and hilly rural areas, or in London.

Figure 3: Map showing locations of surveyed properties (red arrows indicate postcodes with multiple surveyed properties) [3].

In Figure 3, two of the property locations are marked with red arrows rather than blue. This indicates multiple properties being surveyed at the same postcode. Later, we need to invoke the assumption of independence between properties in our sample. This assumption does not seem fair if properties on the same street are used (especially, if they were surveyed on the same day). To eliminate this problem, we take the worst performing property from the two postcodes and discard the information from the other properties. This way we have removed the potential dependence and we cannot be accused of deciding to keep only the best results.

Finally, Figure 4 is a plot of recorded energy use on the day of the survey against estimated heat loss. This gives an impression of the spread of the data and highlights some potential outliers. The two properties with large recorded energy usages were a farm house and a sizeable detached property. In our subsequent calculations, these outliers will be brought into line with the other data points because we will be considering relative differences.

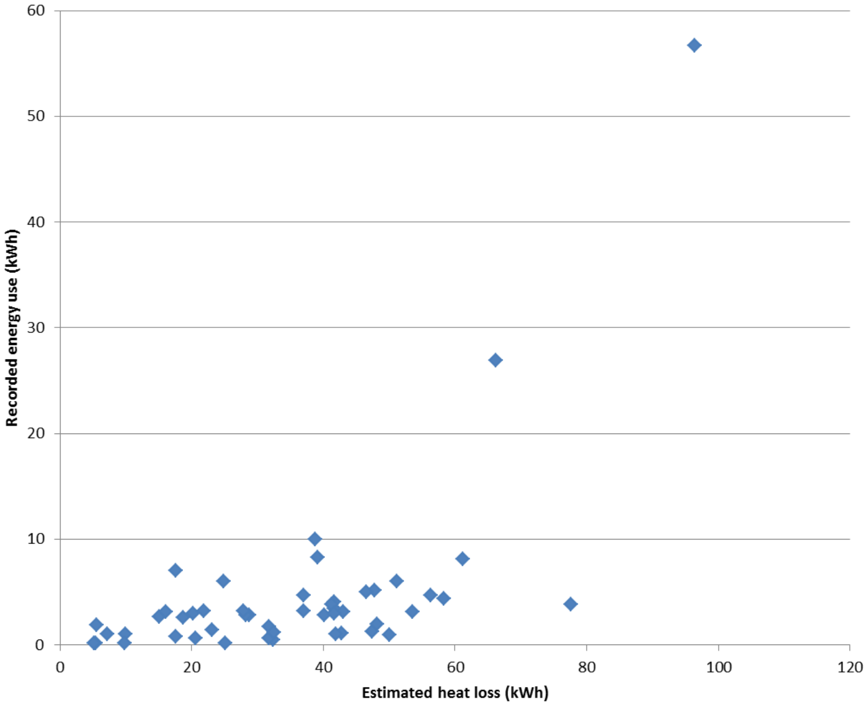

Figure 4: Scatterplot of recorded energy use against estimated heat loss.

Statistical analysis

We want to use statistical methods to estimate the mean of

$$\frac{100 \times (\textbf{Estimated heat loss} - \textbf{Recorded energy usage})}{\textbf{Estimated heat loss}},$$

which we will call the **% difference**. This quantity is a measure of performance against theoretical heat losses. It is not a measure of savings against competing heating systems. The larger the value of this quantity, the better the heating system is performing with respect to the estimated heat losses. Also, because we divide by the estimated heat loss, we are correcting our performance measure for property size and temperature differences.

By making some assumptions, we can produce a confidence interval (CI) for the mean percentage reduction. A CI is a measure of reliability of some estimate based upon a sample. When producing a CI, we also need to specify a confidence level (usually given as a percentage) that we will set to 99% for our analysis. The interpretation of a 99% confidence interval is as follows: if we were able to construct 99% CIs for many separate analyses of randomly selected properties, the proportion of such intervals that contain the true value of the parameter will be 0.99. In statistics, we often use 90%, 95% or 99% CIs where the bigger the percentages, the more confidence we have of our interval including the true mean.

When producing a CI for a mean, there are a number of assumptions that need to be considered. First, the data must be a random sample from the population that we are estimating a mean for. Given the method of property selection described earlier and the different locations and types of property, this seems to be a fair assumption. The properties in the sample must be independent. Again, because we have removed properties from the same postcode, this seems to be a fair assumption especially as there was a range of temperatures being considered. The final assumption is that the sample mean is normally distributed. This can occur in two ways: (1) the underlying data are normally distributed or (2) the sample size is large enough to invoke the central limit theorem (a rule of thumb is that the sample size should be over 30). Figure 5 is a histogram of the percentage differences. It is clear to see that the data are heavily skewed and the assumption of normality would be poor. However, with a sample size of 49, we can use the central limit theorem.

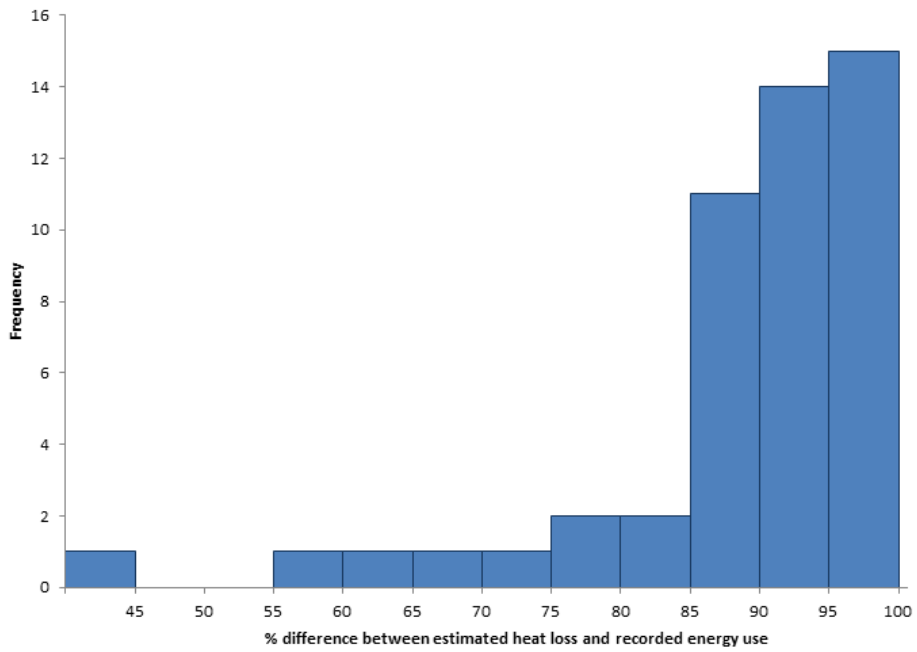

Figure 5: Histogram of % differences.

From the data, we calculate that the mean % difference is 88.3% and the standard deviation is 11.5%. The standard error for our estimate of the mean is therefore 1.6%. With this information, we can construct a 99% confidence interval for the mean: (83.9%, 92.7%). Given this information, we can be very confident that the mean % difference is far away from 0%. Roughly speaking, we would be surprised if the mean % difference was less than 80%, but we would not be that surprised if it was in the low nineties.

<u>Robustness analysis</u>

When doing a statistical analysis of this type, it is worth considering what alternative results we could get if the underlying assumptions were incorrect. One way to do this is to consider the effect of the heat loss calculations being overestimates and the recorded readings being underreported. In both of these cases, we would have a reduction in the mean % difference.

In our robustness analysis, we reproduced the analysis of the previous section after making adjustments to the heat loss estimates and the recorded energy usages. We decided to consider the reduction of all of the estimated heat losses simultaneously at four levels: -5%, -10%, -25%, -50%. The worst case there being -50%, which would correspond to the situation that out estimated heat losses are double what they actually are. Alongside this, we considered increasing all of the recorded energy usages simultaneously at five levels: +5%, +10%, +25%, +50% and +100%. Here the worst case for us is +100%, which corresponds to the actual energy usage being double what was recorded. This might be a little extreme, but some of the recorded usages were based on a simple average over a number of days. The robustness analysis results are given in the following table.

		Adjustment to recorded energy usage					
		Unchanged	+5%	+10%	+25%	+50%	+100%
Adjustment to heat loss estimate	Unchanged	88.28 *(83.88, 92.68)*	87.70 *(83.08, 92.32)*	87.11 *(82.27, 91.95)*	85.35 *(79.86, 90.85)*	82.43 *(75.83, 89.02)*	76.57 (67.77, 85.37)
	-5%	87.67 *(83.04, 92.30)*	87.05 *(82.19, 91.91)*	86.43 *(81.34, 91.53)*	84.58 *(78.80, 90.37)*	81.50 *(74.56, 88.45)*	75.33 *(66.07, 84.60)*
	-10%	86.98 *(82.09, 91.87)*	86.33 *(81.2, 91.46)*	85.68 *(80.30, 91.06)*	83.73 *(77.62, 89.84)*	80.47 *(73.14, 87.81)*	73.96 *(64.19, 83.74)*
	-25%	84.38 *(78.51, 90.24)*	83.6 *(77.44, 89.76)*	82.82 *(76.36, 89.27)*	80.47 *(73.14, 87.81)*	76.57 *(67.77, 85.37)*	68.76 *(57.03, 80.49)*
	-50%	76.57 *(67.77, 85.37)*	75.40 *(66.16, 84.63)*	74.22 *(64.55, 83.9)*	70.71 *(59.71, 81.71)*	64.85 *(51.65, 78.05)*	53.14 *(35.54, 70.73)*

In the table, the result of our analysis from the previous section is given: a mean of 88.28% with a 99% CI of (83.88%, 92.68%) when both the recorded energy usages and heat loss estimates are unchanged. An important result to consider from the table is from the extreme worst case scenario where we have doubled all of the recorded energy usages and halved all of the heat loss estimates. Even in this case, we have a mean estimate of 53.1% and a 99% CI of (35.5%, 70.73%). If this was truly the situation, we would be very surprised if the mean % difference was less than 30%. This robustness analysis gives us confidence that the Logicor Clear Heating System is performing very well against the heat loss estimates.

Looking at the data (as shown in Figure 4), there are a number of properties with recorded energy usages close to 0 kWh. Given that we are looking at % differences, these could be skewing our results by giving values close to 100%. If we made a bold statement that we believe all recorded energy usages of less than 1 kWh are incorrect and remove them from our analysis, we would have 40 properties left to do our statistical analysis on. Doing this, we arrive at a mean of 86.2% and a 99% CI of (81.2%, 91.3%), which are very close to the results of our full analysis of the previous section. Again, this gives us more confidence that our original analysis is correct and that the heating system is performing well.

Conclusions

The heat loss estimates used in this analysis are based on the assumptions given in the second section and the equations provided in the earlier report [1]. This type of estimation procedure can be very sensitive to the choices that are made for the number of air exchanges and the ‘U’ values. A more comprehensive survey of the properties and heat loss determination could help to estimate uncertainty on the heat losses, but the cost could be too great. However, we have some confidence in our results given that the robustness analysis showed favourable results and that the modelling assumptions are backed up by government publications.

An alternative to using the imperfect heat loss estimates as a baseline would be to get data on energy use from conventional systems for each property. These may provide a more informative baseline, but there would still need to substantial assumptions about the temperature profiles both internally and externally and there could be issues with separating this information from overall domestic energy usage. These problems could be lessened if annual data were used instead of daily data (provided the years had similar climatic conditions).

Overall, the results that we have produced show the Logicor Clear Heating System in a good light. Properties with this heating system can expect a % difference of 80 to 90% from the estimated heat loss and they should be very surprised if the % difference they observed was less than 30%.

Appendix: assumed ‘U’ values and air exchanges

The following table gives the ‘U’ values that were used in the heat loss calculations [4,5,6].

Material	'U' value set (W/m^2K)
Cavity brick wall with insulation	0.45
Cavity brick wall without insulation	1.37
Party wall	0.59
Solid/stone wall	2.23
Plain wooden floor	1.41
Carpeted wooden floor	1.41
Concrete floor with laminate	0.43
Concrete floor with carpet	0.43
Tiled roof with insulation (0-100mm)	0.34
Tiled roof with insulation (greater than 100mm)	0.12
Tiled roof without insulation	2.51
Slate roof with insulation (0-100mm)	0.34
Slate roof with insulation (greater than 100mm)	0.12
Slate roof without insulation	2.51
Concrete ceiling	1.36
Plasterboard ceiling	0.82
Flat felt roof (insulated)	0.32
Flat felt roof (w/o insulation)	1.69
Wood DG windows	2.80
Plastic DG windows	2.80
Aluminium DG windows	3.40
Secondary double glazing	2.80
Wood SG windows	4.80
Metal SG windows	5.70
External wooden door	3.00
External UPVC door	2.90

The number of air exchanges per hour was set depending on the room type. Entrance halls and porches were assumed to have 2 per hour, bedrooms and home offices were assumed to have 1 per hour and all other rooms were set at 1.5 per hour [5].

Acknowledgement of funding

This work was wholly funded by Logicor (GROUP) Ltd.

References

[1] Gosling, J.P. (2013). Heat loss calculations for three properties. *Technical report* for Logicor (GROUP) Ltd.
[2] Department for Communities and Local Government (2010). *English Housing Survey: Housing stock report 2008.*
[3] Google (2014). Google maps. Online: https://www.google.com/maps.
[4] DTLR (2010). *The Building Regulations 2010 Part L: Conservation of fuel and power*. The Stationery Office.
[5] CIBSE (2013). *Domestic Heating - Design Guide*. The Yale Press Ltd..
[6] Pilkington UK Ltd. (2013). Understanding the Government's data on U values. Online: http://www.pilkington.com/~/media/Pilkington/Site%20Content/UK/Reference/TableofDefaultUValues.ashx