Evidence, drivers and sources of distortions in the distribution of building energy ratings prior to and after energy efficiency retrofitting

Matthew Collins
Economic and Social Research Institute
Whitaker Square, Sir John Rogerson's Quay
Dublin 2
and Sustainable Energy Authority of Ireland
Wilton Park House
Wilton Place, Dublin 2
Ireland

John Curtis
Economic and Social Research Institute
Whitaker Square, Sir John Rogerson's Quay
Dublin 2
and Trinity College Dublin
College Green, Dublin 2
Ireland

Keywords

energy performance certificates retrofit, bunching

Abstract

Energy performance certificates (EPC) provide a measure of and raise the awareness of the energy efficiency of homes. The Sustainable Energy Authority of Ireland (SEAI) operates a grant aid scheme to incentivise residential energy efficient retrofits known as the Better Energy Homes (BEH) scheme, which was implemented in 2009. Since June 2010, participating homes have been required to undertake independent Building Energy Rating (BER) assessments of the home prior to and after the completion of energy efficient works. This study analyses the distribution of pre- and post-works BERs among participant households, using a regression discontinuity design to examine the significance of discontinuities at each BER grade threshold and to estimate the number of affected BERs in our sample. We find evidence of bunching at the more efficient side of thresholds of post-works BERs, while no evidence of bunching on the more efficient side was found among pre-works BERs. There exists slight evidence of bunching on the less efficient side of certain thresholds in the pre-works distribution. Counter-factual distributions around each threshold are estimated to examine the number of dwellings which may have been affected by potentially incorrect assessments. We analyse whether adjustment of BER assessments is systemic and whether market forces provide an incentive to adjust assessments. Results show significant evidence of the misrepresentation of Building Energy Ratings but this is not found to be systemic. We also examine potential sources of adjustment, finding discontinuities in certain parameters coinciding with the areas where bunching is found to occur.

Introduction

The European Union, through the introduction of the Energy Efficiency Directive, has set a target of achieving a 20 % reduction in greenhouse gas emissions and achieving energy savings of 20 % by 2020 (EU, 2012). With varying patterns of energy consumption across Europe, policies aimed at meeting 2020 targets are implemented at a national level with each state required to develop a National Energy Efficiency Action Plan (NEEAP). Ireland's third, and latest NEEAP concluded that Ireland had met 39 % of its 2020 target by the end of 2012 (DCENR, 2014). With almost 40 % of final energy consumption occurring in buildings, two thirds of which is used for space heating, improving the energy efficiency of the building stock provides a significant policy opportunity to help meet these targets (EC, 2011). This is particularly true of Ireland, where homes have the lowest level of double-glazing in northern Europe (Balaras et al., 2007) and where roughly 50 % of homes in Ireland possess a Building Energy Rating (BER) between D1 and G (CSO, 2015), which are the lowest six grades of a 15-point scale. One such policy to help improve energy efficiency in the residential sector was the introduction of a residential energy efficiency retrofit grant scheme, now known as the BEH scheme. Administered by SEAI, the BEH scheme provides grant aid for home owners to engage in retrofit activities to improve the energy efficiency of their homes.

The Energy Performance of Buildings Directive (EU, 2002), which was transposed into Irish law in 2006 (DEHL, 2006), established a methodological framework for calculating energy performance, which has been implemented in similar manners across the EU. This framework provides for a standardised ranking of the energy performance of homes using energy performance certificates (EPCs). The Irish BER uses a 15-point scale ranging from A1 to G, where A1 is the most energy efficient. This rating system is discussed in more detail below. As part of the BEH scheme, participant households are required to conduct independent BER assessments of their homes and submit a pre-works and post-works assessment to SEAI. As will be discussed in below, discontinuities appear to exist in the distribution of post-works assessments which are not present in the equivalent pre-works distribution. We implement a regression discontinuity design to examine the significance of these discontinuities.

EPCs provide various market benefits, particularly the reduction of information asymmetry. With such a system, agents looking to buy or rent are able to identify the energy performance of buildings which would otherwise be unknown. Provided consumers value energy efficiency, for comfort gains, monetary savings through reduced energy usage, environmental concerns or otherwise, this should lead to an increase in demand for more energy efficient homes. A body of research exists to show that this is the case in various countries, including Ireland (Hyland at al., 2013), England (Fuerst et al., 2015), Wales (Fuerst at al., 2016), Germany (Cajias and Piazolo, 2013), and the Netherlands (Brounen and Kok, 2011). BERs also provide knowledge of the energy efficiency status of a nation's building stock, which allows policy makers to identify where policy implementation may need adjustment. For example, certain categories of the building stock might require greater investment than others, or the extent to which energy efficiency improvements may be required could indicate whether grant aid or financing may be suitable policy options. For these benefits to be most effectively translated to the market, performance ratings must be accurate and dwellings appropriately labelled.

An EPC system with discrete performance thresholds may give rise to perverse incentives, i.e. an incentive may also lead to unintended and undesirable outcomes. We believe bunching to be evidence of such a perverse incentive, as the introduction of an incentive to improve a building's energy efficiency may have also caused an incentive to misrepresent energy efficiency ratings. Bevan and Hood (2006) define three types of perverse incentives, being ratchet effects, threshold effects and output distortions. *Ratchet effects* refer to the incentive not to exceed targets in a given year if targets are based on performance during the previous year. *Threshold effects* refer to the use of minimum performance standards which incentivise improved performance for those below the threshold but lead to stagnation of those above the threshold. *Output distortions* refer to distortions in economic output caused by threshold and ratchet effects. We hypothesise a threshold effect, whereby homes on the less preferable side of a grade threshold are encouraged to improve, although this gives rise to incentives to misrepresent energy performance. Given that a BER assessor can retrospectively adjust the final assessment, and given that the property market values energy efficiency, there is potentially a perverse incentive to marginally falsify BER assessments (i.e. adjusting the assessment in such a manner to move a rating from the less desirable side of a BER threshold to the more desirable side). If the market values energy efficiency, adjustment may allow property owners to sell or rent their home at a higher price, extracting undue rents. BER assessors are hired by home owners to conduct an assessment and, as such, assessors may be either persuaded or incentivised to give more favourable assessments. If these perverse incentive adjustments occur in a substantial number of cases we might expect to see disproportionately more bunching of BER assessments immediately on the more efficient side of the BER grade thresholds and disproportionately fewer assessments on the less efficient side of the BER grade thresholds.

Bunching analysis has been applied to various strands of literature, a summary of which is provided by Kleven (2016). Bunching analysis was originally developed in the area of tax policy and enforcement. For example, Chetty et al. (2011) found significant evidence of bunching in real earnings around kink points in tax thresholds in Denmark while Bastani and Selin (2014) found no bunching of earnings in Sweden, despite the presence of kink points in the tax schedule. Larger levels of bunching were found to exist in the distribution of earnings among the self-employed, who possess greater scope for tax avoidance (Saez, 2010; Kleven and Waseem, 2013). In welfare economics, Camacho and Conover (2011) examine bunching in poverty index scores of families in Colombia at the threshold for identifying recipients eligible for a variety of social welfare programs. In labour economics, Gourio and Roys (2014) look at bunching in the distribution of the number of people employed by firms. They found that, as firms with 50 or more employees faced more stringent regulation than those with less than 50 employees, significant bunching occurred in the distribution, with a disproportionally high number of firms employing 49 people.

With regard to EPCs, bunching analysis has been applied to the study of the effects of energy and efficiency labelling in the car market. Sallee and Slemrod (2012) examine bunching in fuel economy ratings, including the introduction of a rebate program for energy efficient cars in Canada. They matched cars with equal specifications in the years prior to and after introduction of the rebate, using a logistic regression to model the likelihood that a car moved from the less favourable side of the threshold to the more favourable side upon introduction of the scheme. As part of an analysis of price effects of energy efficiency in the Swiss car market, Alberini et al. (2014) use a regression discontinuity design to examine bunching of prices among cars of varying efficiency labels, finding strong evidence that A-grade cars extracted a price premium. Pierce and Snyder (2012) investigate systemic manipulation of vehicle emissions testing in New York state by using a regression discontinuity design. They examine the distribution of test scores before and after a change in the test score required to pass. They find statistically significant evidence of bunching in the distribution of test scores at the passing thresholds for five out of six emissions tests. They concluded that manipulation by testers was likely the cause, as cars could be re-tested until receiving a pass.

The research issue explored in this paper is to identify whether there is potential misrepresentation of building energy ratings as evidenced by statistically significant discontinuities of the distribution of these ratings. As both pre- and post-works ratings are calculated following the completion of energy efficient retrofit works, a greater level of discontinuity of post-works ratings may indicate gaming of the system, given the benefits of an improved BER and may indicate the need for further tailoring the means by which assessments are targeted for audit.

We find that prior to the completion of energy efficiency retrofit works, there is no evidence of bunching on the more efficient side of any grade thresholds and that there appears to be evidence of slight bunching on the less efficient side of the C2/C3 and D1/D2 thresholds. Significant evidence of bunching was found at most thresholds of the post-works distribution and was found to be stronger at thresholds where the letter grade changes. We estimate a counter-factual distribution of BERs surrounding each threshold and estimate that over 3.5 percent of assessments may have been adjusted to pass into more efficient grade labels. We examine drivers of the adjustment of assessments, although no evidence of systemic drivers of adjustment have been found. We investigate sources of adjustment, finding that the low energy lighting parameter used in BER assessments might be a commonly used tool to improve a property's rating.

Data

The Irish BER is an energy label pertaining to the energy efficiency of a home. Homes are assigned to an alpha-numeric grade on a 15-point scale ranging from A1 to G, with A1 being the most energy efficient. SEAI provide a guide to the BER for home owners, outlining who requires a BER and an outline of BER calculation[1]. A BER takes account of energy requirements for space heating, ventilation, water heating and lighting, less savings from energy generation technologies, measured in kilowatt hours (kWh/m^2/year). This is based on a standardised occupancy, with living areas heated to 21 °C and other rooms to 18°C (SEAI, 2013). This calculation requires assessment of a home's dimensions, orientation, insulation and space and water heating system efficiencies. BER assessment does not include the use of electrical appliances such as cookers, washing machines, etc, although pumps and fans for heating and ventilation are included. BER assessments are carried out by independent assessors who have completed an accredited training course, including national examination, and are registered with SEAI. BER assessments are based on calculated energy use, based on a standard expected level of heating and lighting in the home, and does not reflect measured consumption. For the purposes of the BEH scheme, a BER assessment is carried out following completion of energy efficiency retrofit works. The result of this assessment is registered as that home's post-works BER value. The BER assessor then retrospectively discounts the parameters pertaining to the retrofit measures undertaken under the grant aid scheme to estimate what the BER would have been before retrofitting so that energy savings can be estimated. We refer to this estimation as the 'pre-works' BER.

Our dataset comprises all applications to the SEAI's Better Energy Homes scheme. Since June 2010 it has been mandatory for all households in receipt of grant aid to have an independent BER assessment performed to assess the home's energy efficiency both before and after retrofit works were undertaken. Our dataset therefore possesses pre- and post-works BERs for successful applicants from June 2010 through to October 2015 and those who choose to undertake a pre- and post-works BER assessment prior to June 2010. This allows for comparison of the distributions of household BERs. Properties that have had more than one BER assessment over time are represented only by their latest assessment so we discarded BEH applications where a property was assessed subsequently for other purposes (sale, letting, etc.), which led to the loss of 10,862 homes, or 8.8 % of homes who participated in the grant aid scheme.

Figures 1 and 2 show the distribution of pre- and post-works BERs, respectively, detailing the number of homes in each one-kWh band. The pre-works distribution appears to be reasonably smooth and is right-skewed, with more homes possessing E, F and G grades than A or B grades. There does not appear to be any visual evidence of bunching of households within any bands outside of general noise across the distribution. The post-works distribution is much less smooth, with apparent bunching of households in many grade labels. A trend appears to exist whereby a disproportionally large number of homes appear to possess BERs that are marginally on the more energy efficient side of grade thresholds, while proportionally lower numbers of homes appear to possess ratings on the less efficient side. This is particularly noticeable at grades B3 to D2, which possess higher numbers of homes. Evidence of bunching at the most efficient grades is not apparent but there are very few homes with such grades. For example, only 0.26 % of homes possess A or B1 grades after having energy efficiency works undertaken.

Methodology

We take a similar approach to Pierce and Snyder (2012) by using a regression discontinuity design to estimate the significance of the discontinuity at the threshold for each discrete building energy rating grade. This is done for the distribution of both pre- and post-works building energy ratings. The distribution surrounding each grade threshold is taken as a function of the bin number and a pooled polynomial regression, as described by Lee and Lemieux (2009), is modelled as follows:

$$y_j = \alpha + \tau.T + \sum_{i=1}^{p} [\beta_{i1}.(X_j - c)^i + \beta_{i2}.T.(X_{j-c})^i] + \epsilon$$

X_j represents the bin number along the distribution and c represents the bin number at the grade threshold and is used to centre the polynomials at the distribution's cut-point. Our main explanatory variable is therefore the distance from the BER grade threshold, which can be both positive or negative, with negative values occurring on the more favourable side. Our dependent variable, y_j, represents the number of households in a given bin and is measured as a proportion of all of the households in the sample. T is a dummy variable, taking a value of 1 if the bin is to the more favourable side of the threshold, indicating that it is more energy efficient. For example, when looking at the C3/D1 threshold, T will take a value of 1 if the bin is less than 225 kWh, which is the threshold. The magnitude of the discontinuity is given by the parameter τ.

We believe a regression discontinuity design to be an appropriate measure of bunching, as the discrete threshold appears to be the main incentive to adjust assessments, which in turn leads to bunching. We therefore make the assumption that the distributions of other independent variables are relatively smooth across the distribution of BERs. As our depend-

1. Available online: http://www.seai.ie/Your_Building/BER/Your_Guide_to_Building_Energy_Rating.pdf.

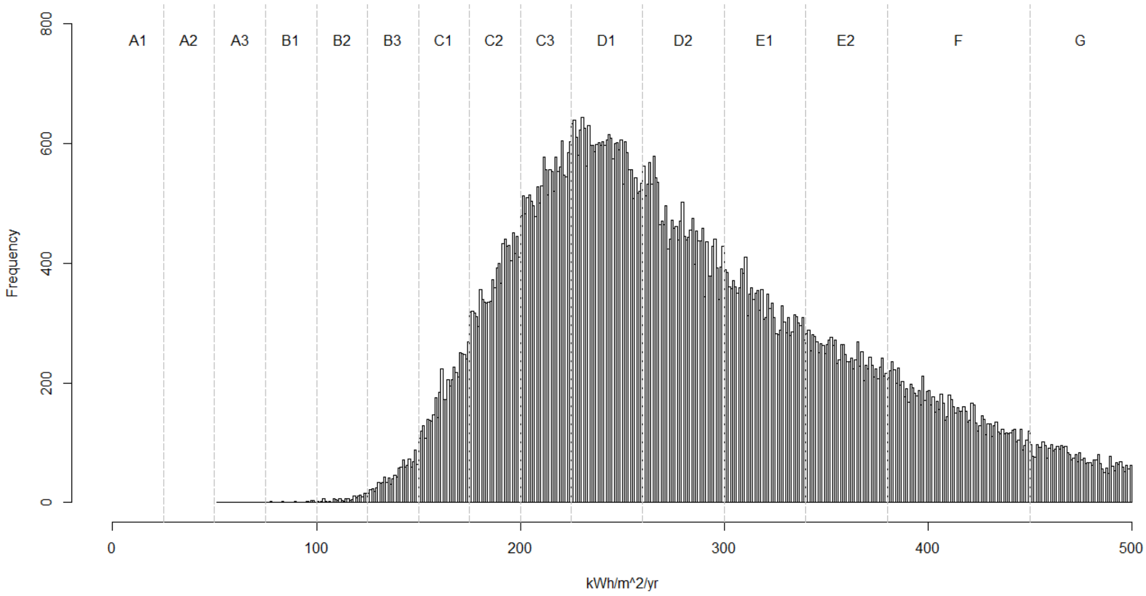

Figure 1. Distribution of pre-works Building Energy Ratings.

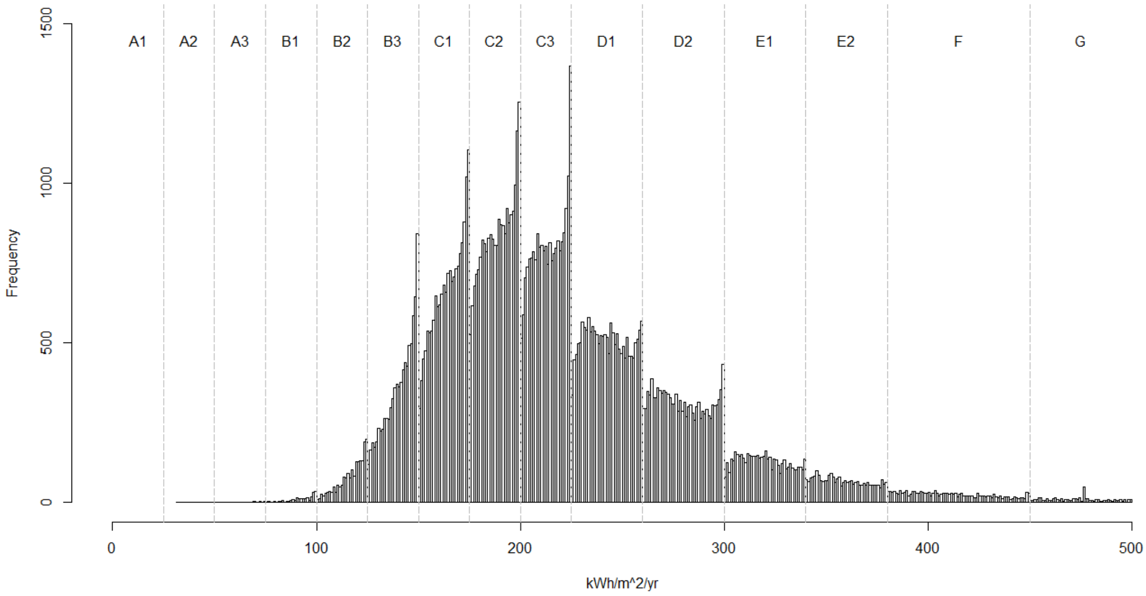

Figure 2. Distribution of post-works Building Energy Ratings.

ent variable measures the proportion of our total sample in a given bin, estimated parameters will be affected by scale. This means that our estimates will not provide any insight into the degree of relative bunching, i.e. the thresholds at which bunching is strongest.

In addition to estimating the statistical significance of bunching estimates, in order to examine more effectively the degree of bunching present in the distribution of Building Energy Ratings, we estimate the number of BER assessments which may have been adjusted. We do this by estimating a counter-factual distribution of post-works BERs. The regression used to calculate the counter-factual distribution takes the following form:

$$\hat{y}_j = \alpha + \sum_{i=1}^{p} \beta_i \left(X_j - c \middle| X_j \notin (z_u, z_l) \right)^i + \theta$$

where $\hat{y}_j$ is the number of dwellings as a proportion of all dwellings in bin j were no bunching to occur. X_j again represents the bin number, with c representing the bin number at the grade threshold and is used to centre the polynomials at the distribution's cut-point. z_l and z_u represent the lower and upper bounds, respectively, of the affected area, i.e. the area in which bunching occurs.

We compare this counter-factual distribution to both the observed distribution and the fitted distribution of assessed

BER grades to make two estimates of the number of assessments which may have been adjusted. We do this by taking the number of assessments above the counter-factual distribution on the more favourable side of the threshold and by taking the number of assessments below the distribution on the less favourable side of the threshold. The excess above the threshold and the deficit below should be roughly equal. This is done for both the actual and fitted values of the bunched distribution for those thresholds which have been shown to possess significant evidence of bunching.

Results and Discussion

SIGNIFICANCE OF BUNCHING ESTIMATES

We first estimate the significance of bunching estimates, with estimated τ coefficient presented in Table 1. We choose a bin size of 0.25 kWh and a polynomial order of 3 for the purposes of this analysis. While higher order polynomials sometimes resulted in lower Akaike information criterion (AIC) for some thresholds, improvements were found to be marginal beyond orders of three in most cases. Some higher orders also led to variable omission in estimation.

Looking first at bunching estimates in our pre-works distribution, we find evidence of negative bunching, i.e. bunching on the less efficient side of the threshold, at certain grades. Statistically significant negative bunching appears to occur at C2 and D1. Bunching at C2 is found to be significant at polynomial orders of four and five, while bunching at D1 is found to be significant at orders of one, two and five. Bunching estimates at C2 and D1 are not robust to all bin sizes. Both are robust to bin sizes of 0.5 kWh/m^2/yr and bunching at D1 is also robust to bin a bin size of 1 kWh/m^2/yr. Overall, this indicates that negative bunching may exist at C2 and D1. The recorded BER improvement achieved by engaging in an energy efficient retrofit for homes with a recorded pre-works BER of C2 has a mean of 27.18 kWh and standard deviation of 14. One potential explanation for pre-works negative bunching of ratings is that if a post-works BER assessment is in the same category as the pre-works assessment, i.e. the retrofit investment has not achieved a large enough improvement to change the BER grade assessment, the adjustment of the BER may occur on the less efficient pre-works BER grade. This might be caused by assessors feeling that a home owner may not be satisfied in not achieving a discrete BER improvement having made a significant investment in an energy efficient retrofit. Evidence of positive bunching in the post-works distribution is found at all grades from B3 to G, excluding E2. All grades where evidence of bunching is found are robust to changes in bin size. Only estimates at C2 and C3 are robust to all polynomial orders. All other significant estimates are robust to a minimum of two other polynomial orders. This provides evidence of bunching at various grade thresholds, indicating that a large number of assessments may have been unduly adjusted to place homes in more favourable BER grades.

ESTIMATING THE NUMBER OF ADJUSTED BER ASSESSMENTS

We estimate counter-factual distributions of post-works BERs around each threshold, omitting the affected area. We choose to omit homes possessing BERs within five kWh of the threshold as we assume adjusted assessments are concentrated in these areas. The fitted values of the counter-factual distributions are compared to both the observed distribution of BERs and to the fitted values of the observed distribution to provide two estimates the number of potentially adjusted assessments. We again choose a polynomial order of three to estimate our counter-factual distributions.

Table 1. Significant of bunching estimates at each grade threshold.

Threshold Grade	Threshold kWh/m^2/yr	Pre-Works τ	Post-Works τ
A1/A2	25	0.000075	0.0000068
A2/A3	50	-.000111	-0.00068
A3/B1	75	-0.000213	-0.00636
B1/B2	100	0.008659	0.00541
B2/B3	125	0.007849	0.03038
B3/C1	150	-0.066623	0.597***
C1/C2	175	-0.008273	1.151***
C2/C3	200	-0.320946*	1.7002***
C3/D1	225	-0.103184	0.915***
D1/D2	260	-0.1706542*	0.309***
D2/E1	300	-0.017400	0.395***
E1/E2	340	0.087614	0.125**
E2/F	380	-0.018429	0.015795
F/G	450	0.016260	0.0279***

*(***p<0.01, **p<0.05, *p<0.1)*

Figure 3 graphically compares the observed and fitted distribution to the counterfactual distribution estimated for the pre- and post-works BER assessments at the D2/E1 threshold. Other thresholds where significant evidence of bunching was estimated possess similar distributions. As can be seen, discontinuities exist on either side of the grade threshold. The pre-works distributions presented shows very little, if any, deviations of the counter-factual to the fitted distribution. Looking at the post-works distributions, there exists visual evidence of the variations between the fitted and counter-factual distribution. A clear spike in the distribution is evident just to the left of the threshold, with a sharp, steep drop once the threshold is reached, before rising back toward the counter-factual distribution.

For each threshold where significant evidence of bunching was found, we estimate the number of adjusted assessments on either side of each threshold, with estimated values presented in table 2. We calculate the proportion of assessments present in the spike to the more efficient side of the threshold which are not present in the counterfactual distribution and those present in the counterfactual distribution but missing from the observed distribution. The proportion of adjusted assessments varies between 3.5 and 4.5 %, which equates to a total number of adjusted assessments between 3,967 and 5,114. The number of adjusted assessments varies across grades due to the distribution of the post-works BER sample. For example, we estimate between 23 and 35 adjusted assessments at the F/G threshold and between 1,020 and 1,417 adjusted assessments at the C3/D1 threshold. As previously discussed, very few homes in our sample remain at F and G grades following energy efficient retrofit works and, as such, there are fewer assessments that can be adjusted. Estimates on the more energy efficient side of the thresholds are larger than those on the less energy efficient side.

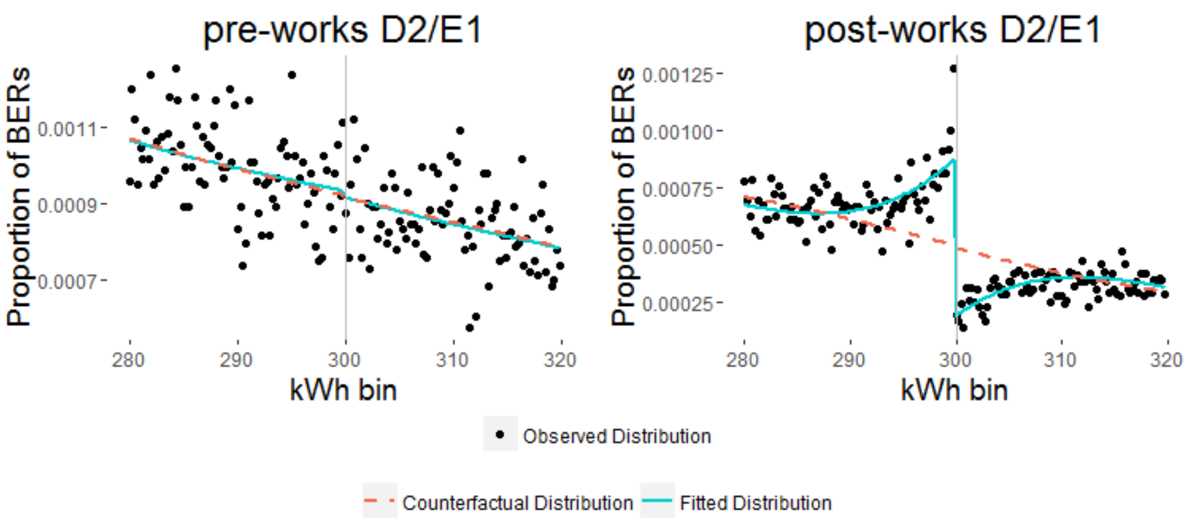

Figure 3. Pre- and post-works distributions of Building Energy Ratings at the D2 grade threshold.

Table 2. Extent of bunching at each grade threshold possessing significant evidence of a discontinuity.

Threshold		More Preferable Side of Threshold		Less Preferable Side of Threshold	
Grade	kWh/m²/yr	Proportion of Total Assessments	Number of Assessments	Proportion of Total Assessments	Number of Assessments
B3/C1	150	0.6876	774	0.4757	536
C1/C2	175	0.7443	838	0.6165	694
C2/C3	200	0.8745	985	0.7637	860
C3/D1	225	1.259	1,418	0.9065	1,021
D1/D2	260	0.3675	414	0.2796	315
D2/E1	300	0.5256	592	0.3939	444
E1/E2	340	0.0525	59	0.0744	84
F/G	450	0.0311	35	0.0212	24
		4.542 %	5,115	3.5315 %	3,978

As a measure of relative bunching, we express the estimated number of adjusted assessments as a percentage of all assessments within 12.5 kWh of each threshold. We choose 12.5 as this is the smallest distance to the mid-point of any grade. This measure of bunching is presented in figure 4. As can be seen, bunching appears to be strongest at thresholds corresponding to a change in letter grade, as the highest levels of relative bunching are seen at B3, C3, D2 and F. This indicates that home owners place, or are perceived by BER assessors to place a value on a more favourable energy rating for their home and in doing so, place a greater value on an improvement in a home's letter grade than is placed on an improvement in the home's alphanumeric grade. This is despite BERs being presented using their alphanumeric value, as opposed to just their letter grade.

EVIDENCE OF SYSTEMIC BUNCHING

Following our analysis of the extent of bunching, we aim to identify patterns in bunching across the distribution of assessments. SEAI provide us with data, including assessor ID numbers, of every recorded BER in Ireland, provided it is the latest BER assessment of a property. This allows us to examine potential adjustment across all properties, rather than only those who had a BER assessment undertaken as part of a Better Energy Homes grant. We examine potential adjustment through variations in the number of BERs within 5 kWh of a threshold on the more favourable side, taken as a percentage of total assessments within 12.5 kWh of the threshold on either side. We choose a band of ±12.5 kWh as this is the smallest distance from a threshold to the mid-point of a grade band. We refer to this area within 5 kWh of the threshold on the more preferable side as the affected area, as we believe all adjusted assessments are issued in this area, based on visual examination of the distribution of BER values.

Using the same method as above, we estimate counterfactual distributions at each threshold where significant evidence of bunching was found, this time using the population of properties which possess BERs. Comparing the observed and counterfactual distributions at each grade, we are able to estimate the number of assessments that have been adjusted at each threshold. We chose not to use the proportion of applications as estimated above as the overall population of BERs is naturally more right-skewed. This is because the sample of homes that have completed retrofits is comprised of dwellings whose energy efficiency has improved, reducing the number of E, F and G grades. We then take the proportion of home within 5 kWh of the more favourable side of the threshold at each grade within each county and compare this to the proportion that should be within this 5 kWh band, as per the national counterfactual distribution. The difference between these proportions is used as our measure of bunching.

We examine whether competition between assessors in an area is a driver of adjustment. We hypothesise that more polarised distributions of assessors, i.e. where a small number of assessors perform the majority of assessments, as opposed to many assessors performing a fewer number of assessments, possess less competitive markets for assessors. Under this assumption, it may be possible that in less polarised markets, assessors have greater incentive to adjust assessments as a means of generating repeat business and more business through word of mouth. We therefore calculate a gini coefficient pertaining to the polarisation of the distribution of assessors in each county's population of BERs. There exists a variation in this gini coefficient across counties but this variation is quite small, varying only within a range between 0.77 and 0.85. We test whether any correlation exists between this gini coefficient and potential adjustment at each grade, with results presented in table 3. We find that the correlation is quite variable across grades, with both positive and negative values across grades. In most cases this correlation is quite low with bunching at the E1/E2 threshold being the only grade to possess a correlation with the gini coefficient of greater than 0.5. We therefore can conclude that adjustment of assessments is not driven by competition between assessors. Furthermore, we examine whether any correlation exists between bunching at different grades across counties. Again, correlations between grades are found to be quite variable and quite low, for the most part. Were bunching to be based on other regional factors, we would expect bunching at each grade to be correlated across counties, perhaps between all grades, or just between grades where the letter grade also changes but this does not appear to be case. It is therefore unlikely that BERs in certain counties are more susceptible to adjustment than others.

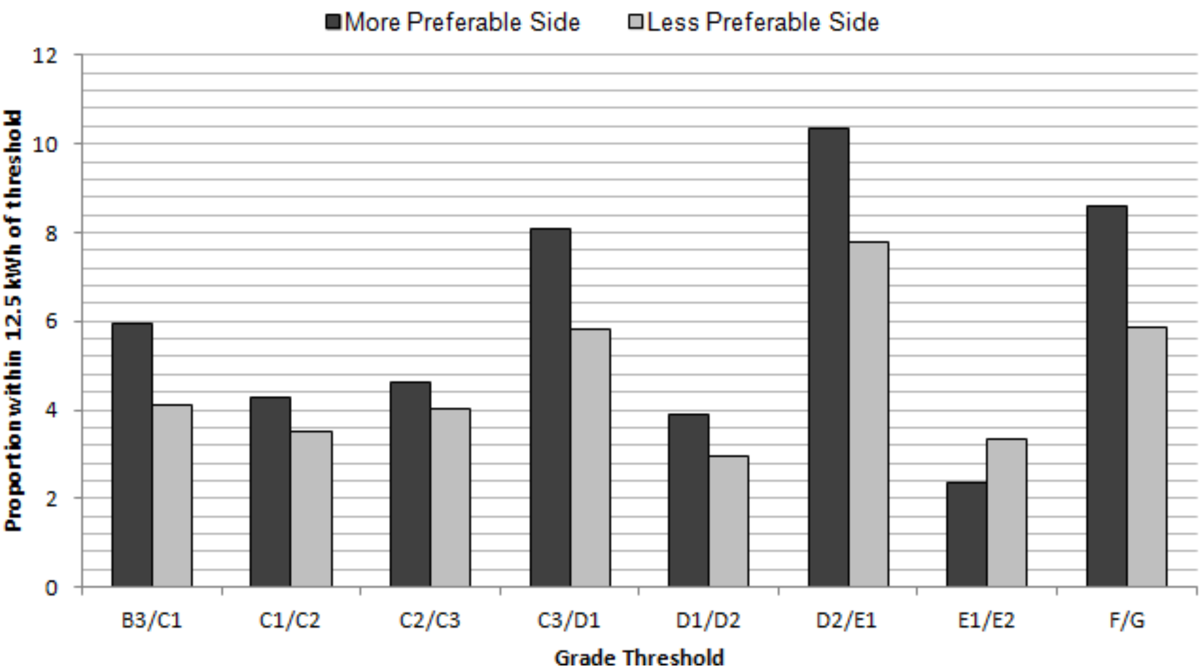

Figure 4. Relative bunching at each grade threshold possessing significant evidence of a discontinuity, as a proportion of assessments surrounding that threshold (%).

Table 3. Correlations between competition and bunching and each grade and bunching between grades.

Potential Adjustment:	B3	C1	C2	C3	D1	D2	E1	F
Gini Coefficient	0.090	0.031	-0.362	0.017	-0.075	0.095	0.534	0.332
Potential Adjustment:								
All Grades	0.603	0.334	0.587	0.850	0.549	0.384	0.430	0.126
B3	.	0.178	0.085	0.484	0.341	0.080	0.094	0.147
C1	.	.	0.039	0.176	-0.263	0.116	0.021	-0.349
C2	.	.	.	0.526	0.173	0.093	0.244	0.148
C3	.	.	.	.	0.456	0.118	0.270	0.095
D1	.	.	.	.	.	0.074	0.297	0.199
D2	.	.	.	.	.	.	0.118	-0.080
E1	.	.	.	.	.	.	.	0.115

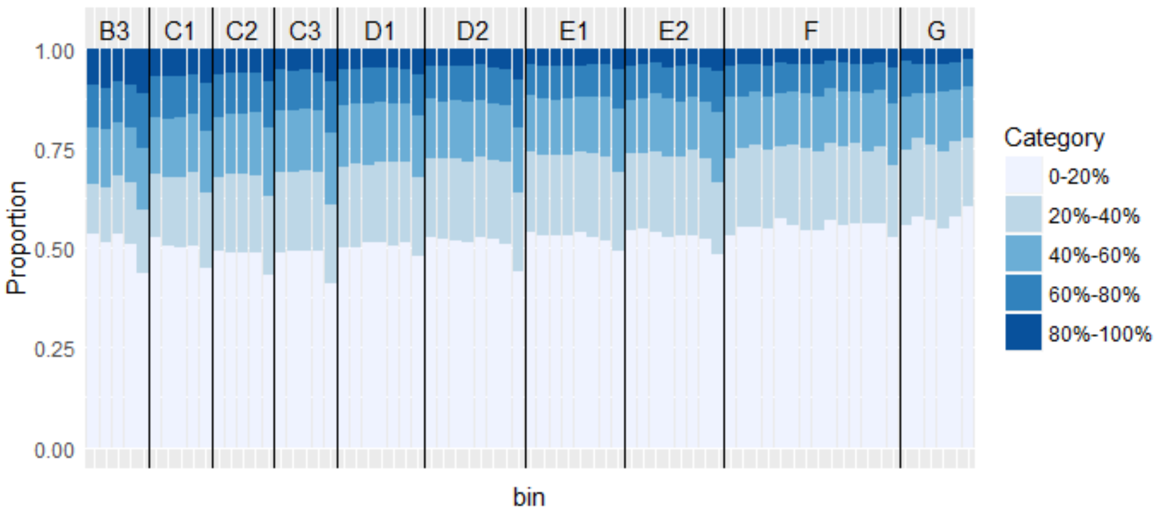

Figure 5. Discontinuities of the proportion of low energy lighting in a home.

In the absence of any drivers of adjustment, we also seek to examine how assessors may be manipulating applications. Some DEAP parameters could be seen as being relatively easy to adjust. Continuous measures, such as the percentage of the property which has been draught proofed, the percentage of lighting that can be classified as low energy lighting and the thickness of hot water store insulation vary from home to home and, as such, it would be difficult to identify adjusted values without an audit of the property's BER. This is particularly true of low energy lighting, as lighting can be changed at any point before or after a BER assessment. If an audit of a BER assessment were to find a lower level of low energy lighting than reported by an assessor, it would be difficult to prove that the home owner did not replace low energy lighting with less efficient alternatives following the assessment. Other categorical parameters might also be considered easy to adjust. Heating system control, response and efficiency categories also vary quite a lot across dwellings and would be difficult to identify as anomalies without a BER audit[2].

It is difficult to identify sources of adjustment as it is not possible, without auditing, to correctly identify which homes that are within 5 kWh of a threshold are those which have been adjusted, as the majority of these homes should in fact exist in this region of the distribution. We divide the distribution of BERs into bins of 5 kWh and plot the proportional distribution of these parameters across all bands on either side of grade threshold where significant evidence of bunching was found. The proportion of low energy lighting in a home possesses noticeable discontinuities, as properties within 5 kWh of a threshold on the more favourable side possess higher levels of low energy lighting than those in other 5 kWh bins. This is shown in figure 5. On visual inspection, other parameters do not appear to provide any systemic evidence of bunching across the distribution. While this does not provide enough evidence to conclude with certainty, it is likely that assessors are using low energy lighting to adjust assessments and assign more preferable grades to homes. This is an appealing parameter for assessors to adjust as it would be difficult for an auditor to prove adjustment, given that following assessment, a household is likely to have replaced some lighting fixtures over time and lighting that could previously have been considered low energy could be replaced by a less efficient alternative.

As a means of detecting the non-compliance in BER assessments, an auditing system currently exists with penalties of differing severity. Penalties currently apply to non-compliances resulting in a net change of 5 % or more of the BER, assessments where the sum of non-compliances results in a change of 10 % or more of the BER and assessments where non-compliance causes a change in the BER grade of a property[3]. SEAI have to date audited a sample of 57 homes within 2 kWh of grade thresholds, with low energy lighting ranking 13th by order of frequency of non-compliances among this sample, and 6th by order of frequency in all homes. Based on the findings of this research, it could be considered to increase the width of this identification band from 2 kWh to 5 kWh, as a sample of 57 homes is quite small and is therefore unlikely to identify any widespread sources of bunching. With regard to policy implications, this paper therefore fails to identify any other potential improvements to the auditing system in place for assessors. This is because the auditing system currently includes sanctions for misrepresentations of a BER of greater than 10 % of the audited BER and for misrepresentations where the submitted BER possesses a different grade to the audited BER. While the idea of not including low energy lighting in BER calculations could be considered, this is very unlikely given that lighting represents a substantial proportion of residential energy consumption.

Conclusion

Residential energy efficient retrofits contribute to reducing overall energy consumption in Ireland, helping to meet the State's energy efficiency targets. Inaccurate labels caused by adjustment of building energy efficiency assessments could lead

2. These categories were suggested as potential sources of adjustment during discussion with SEAI which helped to inform our research.

3. More information on BER auditing procedures is available at http://www.seai.ie/Your_Building/BER/DBER-Tech-June-20151.pdf.

to the misrepresentation of the energy efficiency of the residential building stock. They may also lead to unmerited windfall gains to property owners if there is a label effect associated with energy efficiency ratings (Hyland et al., 2013). A regression discontinuity methodology is used to investigate whether the adjustment of building energy ratings occurs and the extent thereof within the context of the Better Energy Homes retrofit grant scheme.

We find only marginal evidence of downward adjustment of energy efficiency assessments in the pre-works distribution but there is evidence of bunching of post-works BER ratings on the positive side of rating thresholds. Bunching in the post-works distribution is found across most sub-groups of the sample, indicating that adjustment may have been a systemic issue as opposed to being isolated to certain groups of assessments. We examine the absolute intensity of bunching, finding that between 3.5 and 4.5 % of BER assessments within the Better Energy Homes grant scheme may have been adjusted downward (i.e. more efficient). We find that bunching is relatively stronger at thresholds where the letter grade changes, e.g. the C3/D1 threshold, as opposed to thresholds where the letter grade does not change, e.g. C2/C3 or D1/D2.

This research adds to the literature on evidence of perverse incentive and the literature on bunching in energy labels in the residential sector. The implications of this research are quite clear in that between 3.1–4.7 % of assessments represent a significant proportion of the sample and in the interest of accuracy and consumer protection, the results of this analysis may be used to inform the auditing process. Analysis here suggests that adjustments are mostly achieved via the low energy lighting parameter. It is possible that assessors are adjusting this parameter in order to misrepresent a property's BER grade. It is also possible that assessors might advise a home owner that if they replace a certain proportion of their lighting with low energy alternatives, their BER will rise. An assessor might therefore record a higher level of low energy lighting on the promise that the recommended replacements are made. A visual lighting audit can be completed relatively easily and if the assessed level of low-energy lighting is not present, an assessment is likely to have been misrepresented. While it has not been assessed within this paper the mislabelling of energy ratings potentially adds a substantial premium to property owners, which will be paid by unsuspecting customers seeking to either purchase or rent such properties.

Bibliography

Alberini, A., Bareit, M., and Filippini, M. (2014). Does the Swiss car market reward fuel efficient cars? *CER-ETH Center of Economic Research at ETH Zurich Economics Working Paper Series*, 14/190. available online: http://www.cer.ethz.ch/research.

Balaras, C. A., Gaglia, A. G., Georgopoulou, E., Mirasgedis, S., Sarafidis, Y., and Lalas, D. P. (2007). European residential buildings and empirical assessment of the hellenic building stock, energy consumption, emissions and potential energy savings. *Building and Environment*, 42 (3): 1298–1314.

Bastani, S. and Selin, H. (2014). Bunching and non-bunching at kink points of the Swedish tax schedule. *Journal of Public Economics*, 109: 36–49.

Bevan, G. and Hood, C. (2006). Whats measured is what matters: targets and gaming in the English public health care system. *Public Administration,* 84 (3): 517–538.

Brounen, D. and Kok, N. (2011). On the economics of energy labels in the housing market. *Journal of Environmental Economics and Management*, 62 (2): 166–179.

Cajias, M. and Piazolo, D. (2013). Green performs better: energy efficiency and financial return on buildings. *Journal of Corporate Real Estate*, 15 (1): 53–72.

Camacho, A. and Conover, E. (2011). Manipulation of social program eligibility. *American Economic Journal*: *Economic Policy*, pages 41–65.

Chetty, R., Guren, A., Manoli, D., and Weber, A. (2011). Are micro and macro labor supply elasticities consistent? a review of evidence on the intensive and extensive margins. *The American Economic Review*, 101 (3): 471–475.

CSO (2015) Domestic Building Energy Ratings Release, *Central Statistics* Office, December 2015, Table 15.

DCENR (2014). *National Energy Efficiency Action Plan 2014*. Department of Communications, Energy and Natural Resources.

DEHL (2006). *European Communities (Energy Performance of Buildings) Regulations 2006, volume S.I. No. 666 of 2006*. Department of the Environment, Heritage and Local Government.

European Commission (2011). *Communication from the commission to the European parliament, the council, the European economic and social committee and the committee of the regions: Energy efficiency plan 2011*. European Union.

European Union (2002). *Directive 2002/91/EC of the European Parliament and of the Council*. European Parliament and the Council of the European Union.

European Union (2012). *Directive 2012/27/EU of the European Parliament and of the Council of 25 October 2012 on energy efficiency, amending directives 2009/125/EC and 2010/30/EU and repealing directives 2004/8/EC and 2006/32/EC*. European Parliament and the Council of the European Union.

Fuerst, F., McAllister, P., Nanda, A., and Wyatt, P. (2015). Does energy efficiency matter to home-buyers? An investigation of EPC ratings and transaction prices in England. *Energy Economics*, 48: 145–156.

Fuerst, F., McAllister, P., Nanda, A., and Wyatt, P. (2016). Energy performance ratings and house prices in Wales: An empirical study. Energy Policy, 92: 20–33.

Gourio, F. and Roys, N. (2014). Size-dependent regulations, firm size distribution, and reallocation. *Quantitative Economics.* 5 (2): 377–416.

Hyland, M., Lyons, R. C., and Lyons, S. (2013). The value of domestic building energy efficiency: evidence from Ireland. *Energy Economics*, 40: 943–952.

Kleven, H. J. (2016). Bunching. *Annual Review of Economics*, 8 (1).

Kleven, H. J. and Waseem, M. (2013). Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from Pakistan. *The Quarterly Journal of Economics*, 128 (2): 669–723.

Lee, D. S. and Lemieux, T. (2009). Regression discontinuity designs in economics. *NBER Working Paper,* 14723. available online: http://www.nber.org/papers/w14723.

Pierce, L. and Snyder, J. A. (2012). Discretion and manipulation by experts: Evidence from a vehicle emissions policy change. *The BE Journal of Economic Analysis & Policy*, 12 (3).

Saez, E. (2010). Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy*, 2 (3): 180–212.

Sallee, J. M. and Slemrod, J. (2012). Car notches: Strategic automaker responses to fuel economy policy. *Journal of Public Economics*, 96 (11): 981–999.

SEAI (2013). *Dwelling Energy Assessment Procedure*. Available online: http://www.seai.ie/Your_Building/BER/BER_Assessors/Technical/DEAP/Introduction_to_DEAP_for_Professionals.pdf.

SEAI (2013). *Dwelling Energy Assessment Procedure: Introduction to DEAP for Professionals*. Sustainable Energy Authority of Ireland. available online: http://www.seai.ie/Your_Building/BER/BER_Assessors/Technical/DEAP/Introduction_to_DEAP_for_Professionals.pdf.

Acknowledgements

We acknowledge the Sustainable Energy Authority of Ireland for access to the anonymous dataset of Better Energy Homes scheme applications and thank contributors at SEAI, at the Italian Association of Energy Economists Energy Symposium and others for helpful comments and suggestions. This research has been financially supported by Sustainable Energy Authority of Ireland and the Energy Policy Research Centre.