Bunching of residential building energy performance certificates at threshold values

Matthew Collins[a,b], John Curtis[a,c,*]

[a]*Economic and Social Research Institute, Sir John Rogerson's Quay, Dublin, Ireland*
[b]*Sustainable Energy Authority of Ireland, Wilton Park House, Wilton Place, Dublin, Ireland*
[c]*Trinity College Dublin, Dublin, Ireland*

Abstract

Energy performance certificates (EPC) provide a measure of and raise awareness of the energy efficiency of homes. The Irish system of energy performance certificates comprises fifteen alpha-numeric grades, which imposes fourteen discrete notches in an otherwise continuous scale. Visual evidence of bunching occurs in the distribution of EPC certificates on the favourable side of EPC grades among homes that have completed energy efficiency retrofits but not in the equivalent distribution of EPC certificates prior to completion of retrofits. Using a regression discontinuity design and estimation of a counter-factual distribution, evidence is found of bunching in the post-works distribution but not the pre-works distribution. We analyse whether statistical evidence exists of drivers of this bunching and whether sources of bunching can be identified. We find that bunching is widespread but not systemic and find discontinuities in the distribution of certain parameters of assessment (i.e. low energy lighting) which suggests their contribution toward bunching. With comparable EPC schemes across Europe the results have policy implications for other countries. Where continuous EPC scales are used, additional energy efficiency improvements may be achieved if a multi-point scale is considered. In countries where multi-point EPC scales are used, bunching of assessments is also likely to occur and quality control measures may be necessary to ensure that the bunching is due to genuine energy efficiency improvements and not a reflection of illicit activity.

Keywords: Energy efficiency, energy performance certificates, bunching, regression discontinuity design

1. Introduction

The Energy Performance of Buildings Directive (European Union, 2002), transposed into Irish law in 2006 (DEHL, 2006), established a methodological framework for calculating energy performance and required the development of Energy Performance Certificates (EPC) to communicate the energy performance of homes using a standardised ranking. Ireland's EPC, Building Energy Rating (BER), uses a 15-point scale ranging from A1 to G, with A1 as the most energy efficient grade. The assignment of alpha-numeric grade follows an on-site assessment of more than 130 property characteristics, yielding a building energy performance measurement in units $kWh/m^2/yr$, which is transposed into the 15-point alpha-numeric scale. EPCs provide various market benefits, particularly the reduction of information asymmetry. With an energy efficiency rating system, agents seeking to buy or rent are aware of the energy performance of buildings which would otherwise be unknown. Provided consumers value energy efficiency, for comfort gains, for monetary savings through reduced energy usage, for environmental concerns or otherwise, this will be reflected in property prices. A body of research exists to show that properties with higher energy efficiency secure higher prices, for example in Ireland (Hyland et al., 2013; Stanley et al., 2016), England (Fuerst et al., 2015), Wales (Fuerst et al., 2016), Germany (Cajias and Piazolo, 2013), and the Netherlands (Brounen and Kok, 2011). EPCs also provide knowledge of the energy efficiency status of the building stock, which allows policy makers to identify

---

*Corresponding author
Email address:* `john.curtis@esri.ie` (John Curtis)

where policy implementation may need adjustment. For example, certain categories of the building stock might require greater investment than others. For these benefits to be most effectively translated to the market, performance ratings must be accurate and dwellings appropriately labelled.

This paper provides evidence of bunching of BER assessments on the favourable side of the thresholds on the 15-point BER scale in Ireland. Bunching in this instance is not inherently a bad outcome, as it indicates home-owners are striving to improve their properties' energy efficiency. But there is a financial incentive from having an assessment on the favourable side of the thresholds within the 15-point scale. Each 1-point improvement along the scale, on average, is associated with a 1% increase in the property's price (Stanley et al., 2016). Understanding whether the bunching is real and reflects accurate assessments of properties' energy performance or the result of illicit activity is important. If the bunching reflects reality then dividing the continuous BER scale, denoted in units $kWh/m^2/yr$, into a 15-point scale has achieved a beneficial outcome of encouraging home-owners to upgrade their homes beyond the next attainable threshold. This has policy implications for other EU countries that use continuous rather than multi-point EPC scales, e.g. Germany and Belgium. If the bunching represents illicit activity the associated price premiums constitute a fraud on buyers or renters and has relevance for the integrity of the BER scheme. While the analysis in this paper is specific to the Irish EPC scheme it has much wider policy implications. EPC schemes across Europe have their origins in the Energy Performance of Buildings Directive introduced in 2002 and revised in 2010 (European Union, 2002, 2010). Any successes or challenges in implementing the Directive in one country are useful lessons for authorities in other European countries implementing the same legislation. Understanding the drivers of bunching in the Irish BER scale, including what retrofit measures are most closely associated with the bunching, has relevance for policy-makers and those involved in improving the energy efficiency of the residential building stock across Europe. Furthermore, the level of accessibility to EPC databases for research purposes varies widely across Europe and Ireland stands out as having one of the most well-established databases and provides opportunities for research not feasible elsewhere (Arcipowska et al., 2014). Access to property level statistical data from the BER scheme is freely available for "personal, research or education purposes".[1]

An energy labelling system with discrete performance thresholds may give rise to perverse incentive, i.e. a policy incentive may also lead to unintended and undesirable outcomes. Bunching may be a symptom of this perverse incentive, as the introduction of an incentive to improve a building's energy efficiency may have also enticed misrepresentation of energy efficiency ratings. Bevan and Hood (2006) define three types of perverse incentive, one of which is potentially relevant in this instance. 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. The outcome of a BER assessment is visible to an assessor throughout the assessment procedure and parameters can be adjusted during the assessment process. Combined with how the property market values energy efficiency, there is an incentive to marginally falsify BER assessments, i.e. to manipulate the assessment in such a manner to move a rating from the less desirable side of a BER threshold to the more desirable side. On the other hand, an ongoing awareness by assessors of the assessment score during the assessment procedure also allows assessors to advise home-owners of measures that could be easily undertaken to improve their BER, which would have a positive impact on the energy efficiency of the building stock. Understanding which is the source of the bunching is important both for the integrity of the scheme but also from a policy perspective to learn which retrofit measures are being used to nudge BER assessments to the favourable side of the threshold.

Bunching analysis was initially developed to estimate behavioural responses in the public finance literature (Saez, 2010; Chetty et al., 2011; Kleven and Waseem, 2013), thought there are now applications in many settings (Kleven, 2016). It investigates whether discontinuities in incentives elicit behavioural responses. For example, Saez (2010); Chetty et al. (2011); Kleven and Waseem (2013) find that discontinuities, i.e. thresholds, in the application of income taxes lead people to organise their individual affairs in a such a way that collectively there is bunching on one side of the threshold. It is the threshold, or the discontinuity in the incentive, that is driving the bunching behaviour. Whether in taxation or elsewhere incentive discontinuities have existed in perpetuity but bunching analysis is a relatively recent

---

[1]The National BER Research Tool is available at `https://ndber.seai.ie/BERResearchTool/Register/Register.aspx`

approach, closely linked to the upsurge in research using administrative data. As bunching usually occurs in close proximity to specific points, large datasets are required to observe the phenomenon. Bunching is rarely observed in survey data due to small sample sizes and measurement error (Kleven, 2016). In large administrative datasets simply plotting the raw data can often reveal evidence of bunching, as is evident for the BER data in Figure 3. In addition to public finance, applications of bunching analysis cover pensions (Brown, 2013), health insurance (Einav et al., 2015), labour market (Gourio and Roys, 2014) and electricity prices (Ito, 2014), amongst others. There are also a number of applications related to EPCs. There is evidence of bunching in EPC assessments in residential properties in the United Kingdom (Comerford et al., 2016) and in commercial properties both in the United States and United Kingdom (Atasoy, 2016). Within the energy efficiency literature Pierce and Snyder (2012) argue that bunching is a sign that cheating is taking place. They investigate vehicle emissions testing in New York finding evidence of bunching in the distribution of test scores at the passing thresholds for five out of six emissions tests. Sallee and Slemrod (2012) argue that bunching reduces welfare because it provides discontinuous incentives and induces actions that have negative net social benefits related to air emissions and health impacts. Similarly, Alberini et al. (2016) show that in the presence of bunching at arbitrary thresholds, that consumers are willing to pay more for otherwise identical goods. Within specific buildings any welfare impact associated with bunching is likely to be confined to private benefits, specifically to participants in property transactions, i.e. affecting the sale or rental price. Both Matisoff et al. (2014) and Shewmake and Viscusi (2015) argue that the motivation for bunching in EPCs relates to the price premium associated with EPC labels and that builders or owners strategically incorporate energy efficient features to achieve higher ratings rather than attempt to cheat.

Hyland et al. (2016) already show evidence of bunching in Irish BERs, finding that the bunching is particularly strong at the lower end of the BER scale. They suggest that property vendors, for which a BER is mandatory to complete the sale, have a strong preference for receiving a high BER rating, motivated by a strong 'asking' or 'offer' price premium (Hyland et al., 2013, 2016; Stanley et al., 2016). However, they suggest that there is no incentive for BER assessors to cheat, as their services are remunerated by fixed fee and not directly linked to the sale price of a property. The research by Hyland et al. (2016) is consistent with the arguments of Matisoff et al. (2014) and Shewmake and Viscusi (2015) suggesting that a price premium is motivating owners to achieve higher BER grades, which are achieved by genuine means rather through cheating on the BER assessment. The existence or perception of a price premium is central to this thesis on the bunching of BER grades. While BERs are mandatory to conclude a property transaction in Ireland, not all BER assessments are undertaken in anticipation of a property transaction. Many BER assessments are undertaken as part of a grant application for a residential energy efficiency retrofit without any intention to subsequently sell or rent the property. Bunching of BER grades also occurs for such properties, which makes the price premium motivation less credible, or at least not the only motivation for bunching. There are a number of alternative explanations for bunching. Though Hyland et al. (2016) suggest that a fixed fee removes the incentive for BER assessors to cheat, the incentive to cheat can be incorporated within a fixed fee. So perverse actions by BER assessors remains an alternative explanation for bunching. Second, instead of a price premium motive, home-owners, particularly those that are upgrading their properties, could be motivated by improving the energy performance of their home. Thirdly, the observed bunching could be an artefact of the BER assessment software. Hyland et al. (2016) already provide evidence that this latter alternative is not the case.

The analysis here focuses on a sub-set of the properties considered by Hyland et al. (2016), those that undertook a grant-supported energy efficiency retrofit. For such grant supported properties we have BER assessments of their homes both pre- and post- retrofit. Bunching occurs in the distribution of post-works assessments but not in the equivalent pre-works distribution. We employ a regression discontinuity design to provide statistical evidence of bunching and estimate a counter-factual distribution at each notch to estimate the extent to which bunching occurs. We believe that the bunching is due to grade-targeting by home-owners. Based on a number of analyses we find no conclusive evidence of illicit behaviour. This research is important for energy conservation in the residential sector. Firstly, it shows in the Irish case that an EPC rating with discrete performance thresholds has encouraged home-owners to upgrade their homes beyond the next attainable threshold, which will have a positive outcome for energy conservation in such properties. Secondly, there are lessons for other countries implementing EPC schemes. An EPC scale with discrete thresholds has the potential to deliver higher levels of energy conservation than EPC schemes using a continuous scale. The EPC assessment process is also an opportunity to educate home-owners on energy efficiency. The ongo-

ing awareness by Irish BER assessors of the BER score during the assessment procedure allows assessors to advise home-owners of measures that could be easily undertaken to improve the energy performance of their homes. It is potentially this activity that is the origin of bunching on the favourable side of threshold in the distribution of Irish BERs.

The remainder of the paper is organised as follows: Section 2 describes the data used, including a more complete description of the Building Energy Rating framework. Section 3 contains a discussion of modelling and estimation issues. This is followed by the presentation and discussion of the estimation results in Section 4, while Section 5 concludes.

Figure 1: Example domestic BER certificate
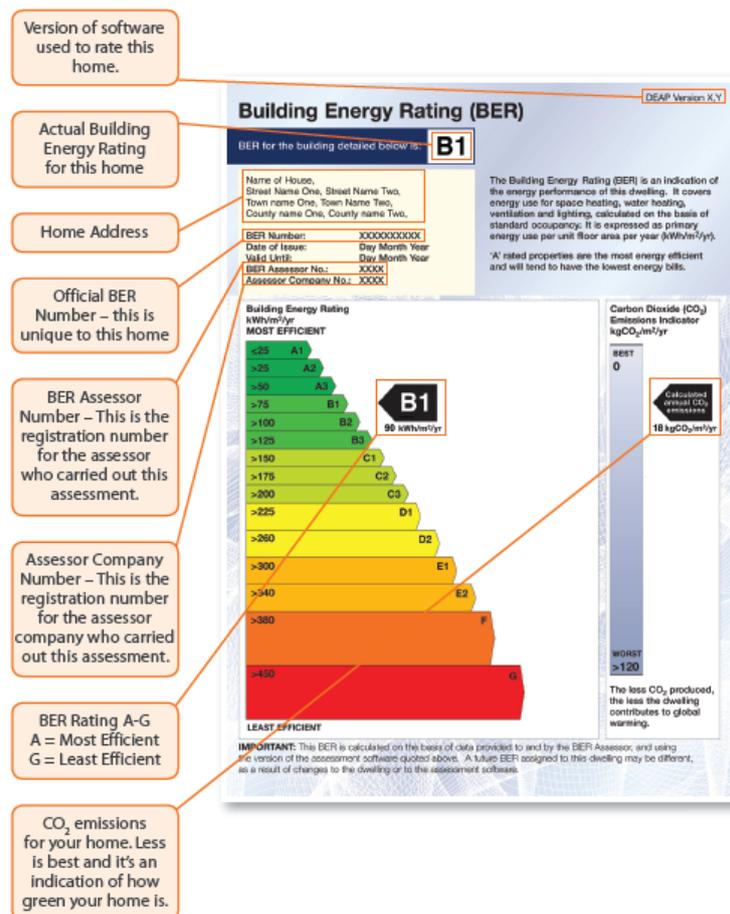
2. Data

2.1. Building Energy Rating (BER)

Ireland's EPC, Building Energy Rating (BER), uses a 15-point scale ranging from A1 to G, with A1 being the most energy efficient. A BER assesses energy requirements for space heating, ventilation, water heating and lighting, less savings from energy generation technologies based on a standardised occupancy, with living areas heated to 21°C and other rooms to 18°C (SEAI, 2013). The assessment incorporates a property's dimensions, orientation, insulation and space and water heating system efficiencies but excludes energy use of electrical appliances such as cookers, washing machines, etc. Figure 1 shows an example BER certificate. BER assessments are undertaken by accredited and registered assessors using a bespoke software, Dwelling Energy Assessment Procedure (DEAP). Throughout

the assessment three values are visible to the assessor: BER alpha-numeric grade; the continuous BER measure in $kWh/m^2/yr$; and CO_2 emissions. Assessors may revise data entered into the DEAP software and instantaneously see the adjustment in BER rating. This feature of the software potentially enables either grade-targeting retrofit measures or cheating, as an assessor can revise data inputs until the desired outcome is achieved.

2.2. The Better Energy Homes scheme

The Sustainable Energy Authority of Ireland (SEAI) administer the Better Energy Homes (BEH) scheme, which provides grants to households to undertake retrofit works to improve the energy efficiency of their property. Grants are available for attic insulation, one of three types of wall insulation (cavity insulation, external wall insulation or internal dry-lining), three types of heating system upgrade (oil boiler or gas boiler with heating controls upgrade or heating controls upgrade only) and solar heating installation, with bonus payments available for multiple measures. Since June 2010, all participants to the scheme must undertake a BER assessment following completion of retrofitting works. For the purposes of the BEH scheme, home-owners first apply for grant aid. A contractor registered with SEAI then undertakes the relevant works. Following completion of the works, an independent and qualified BER assessor who is not the contractor then examines the property to calculate its BER. Following the BER assessment, relevant paperwork is submitted to SEAI before grant aid is administered. The BEH scheme requires pre- and post-works BER assessments. While assessing a property's post-works BER, a pre-works BER is estimated using information on the pre-works characteristics of those aspects of the home altered during retrofitting works. This pre-works BER is submitted to SEAI in addition to the final BER assessment.

In the BEH scheme default levels of grant aid are awarded for each retrofit measure, subject to some terms and conditions. The grant aid does not vary except by property type in the case of solid wall insulation (dry-lining or external wall insulation). The level of grant aid awarded is not in any way related to the technical energy efficiency of a dwelling before works are undertaken or the level of improvement accrued from engaging in retrofit works. The grant scheme therefore does not provide any incentive for misrepresentation of either the pre-works or post-works BER.

Our dataset comprises all BERs for homes that participated in the BEH scheme. This allows for comparison of the distributions of household BERs before and after retrofit works. The dataset only contains the latest BER assessment so properties that were subsequently assessed for other purposes (sale, letting, etc.) were excluded from the analysis. Figures 2 and 3 show the distribution of pre- and post-works BERs, respectively, detailing the number of homes in each 1 kWh bin in the distribution. The pre-works distribution is reasonably smooth and is right-skewed, with more homes possessing E, F and G grades than A or B grades. There is no obvious 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 on the more efficient side of BER grade thresholds in the distribution. The bunching coincides with density holes on the less efficient side of the BER grade thresholds. This is particularly noticeable at grades B3 to D2, which possess larger 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.

3. Methodology

3.1. Regression discontinuity design

To identify statistical evidence of bunching, we take a similar approach to Pierce and Snyder (2012) by using a regression discontinuity design to estimate the significance of the discontinuity at each notch (i.e. BER threshold value) in the distribution of BERs. The discontinuity in this case refers to the distance between the end of the bunching point and density hole at the notch. This is undertaken for both the pre- and post-works distributions. The distribution of building energy ratings surrounding each notch is divided into equal-sized bins and the proportion of the overall distribution present in each bin is taken as a function of the bin number. We choose the proportion of the overall distribution which is present in each bin as our outcome variable, rather than the absolute number, because, as will be

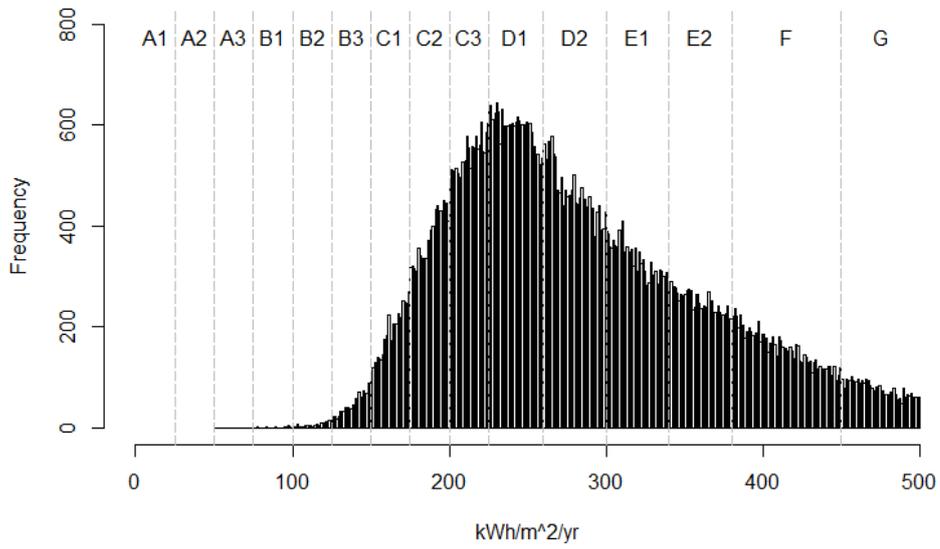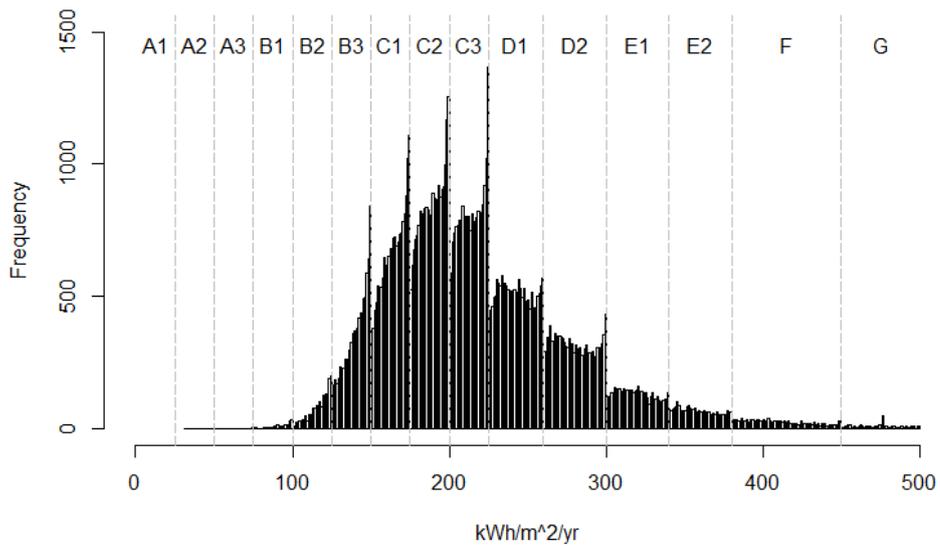Figure 2: Distribution of Pre-works Building Energy Ratings

Figure 3: Distribution of Post-works Building Energy Ratings

discussed in section 4.3, we want to be able to compare bunching across sub-samples of the distribution. A pooled polynomial regression, as described by Lee and Lemieuxa (2010), is modelled as follows:

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

X_j represents the bin number along the distribution as a whole and y_j represents the number of households in a given bin, measured as a proportion of all of the households in the sample. The model is estimated on a notch by notch basis, with c representing the bin number at the notch, i.e. the threshold where each BER grades change. c is used to centre the polynomials at the notch as our main explanatory variable is the distance from the notch, which can be both positive or negative, with negative values occurring on the more favourable side of the notch point. 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 number is less than 225 $kWh/m^2/yr$, which is the notch point. The magnitude of the discontinuity is given by the parameter τ. In order to check the robustness of our estimates, we estimate our model with varying polynomial orders and bin sizes.

We believe a regression discontinuity design to be an appropriate measure of bunching, as the discrete threshold provides a point at which a discontinuity can occur in the distribution, 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 dependent variable measures the proportion of our total sample in a given bin, estimated parameters will be affected by scale within the distribution. 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. For example, model coefficients estimated for D-grades will exceed those of B-grades due to the proportion of the sample present in that area of the distribution. This analysis is concerned with the statistical significance of these estimates. After identifying where evidence of bunching exists, the number of homes in each bunching point, and those missing from each density hole, are calculated through the estimation of a counter-factual distribution in the next section.

3.2. Estimating a counterfactual distribution

In order to examine the degree of bunching present in the distribution of BERs, we estimate the proportion of BER assessments that may have been adjusted. We do this by estimating a counter-factual distribution of post-works BERs, which is then compared with the observed distribution discussed in section 3.1. The regression used to calculate the counter-factual distribution takes the following form:

$$\hat{y}_j = \alpha + \Sigma_{i=1}^{p}\beta_i(X_j - c|X_j \notin [z_u, z_l])^i + \theta \tag{2}$$

where $\hat{y}$ 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 of the affected area, i.e. the area in which bunching occurs. The area of the distribution between these two bounds is excluded from our data when estimating the counter-factual distribution.

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 notch point, i.e. the number of assessments in the bunching area. Similarly, the number of assessments absent in the density holes to the less favourable side of the notch point are estimated for comparison. The number of assessments in the bunching area is calculated using equation (3), where M_g represents an estimate of the proportion of dwellings affected by bunching at notch g, Y_i represents the observed proportion of dwellings in each bin, $\hat{y}_i$ represents the counter-factual value of Y_i, as calculated using equation (2), all measured as a proportion of total dwellings.

$$M_g = \Sigma_{i=x_l}^{c}(Y_i - \hat{y}_i) \tag{3}$$

4. Results and discussion

4.1. Evidence of bunching

To estimate the significance of bunching estimates we choose a bin size of 0.25 *kWh* and a polynomial order of 3. Gelman and Imbens (2014) recommend the use of linear or quadratic functions in regression discontinuity designs. The introduction of bunching regions and density holes around grade thresholds necessitates a second turning point in the function and hence a polynomial of at least order three is necessary. While higher order polynomials sometimes resulted in lower Akaike Information Criterion (AIC) values for some notches, improvements were found to be marginal beyond orders of three in most cases. Some higher orders also led to variable omission in estimation. For details of recorded AIC statistics for each threshold and polynomial order, see Appendix A. The estimated τ statistics for these regressions of varying polynomial order were used to assess the robustness of our results, alongside those of regressions estimated using varying bin sizes.

Table 1: Estimated discontinuities at each BER threshold

Threshold Grade	$kWh/m^2/yr$	Pre-Works[1] τ	Post-Works[1] τ
A1/A2	25	0.00	0.00
A2/A3	50	0.00	0.00
A3/B1	75	0.00	-0.01
B1/B2	100	0.01	0.01
B2/B3	125	0.01	0.03
B3/C1	150	-0.07*	0.60***
C1/C2	175	-0.01	1.15***
C2/C3	200	-0.32**	1.70***
C3/D1	225	-0.10	0.92***
D1/D2	260	-0.17**	0.31***
D2/E1	300	-0.02	0.40***
E1/E2	340	0.09	0.13***
E2/F	380	-0.02	0.02*
F/G	450	0.02	0.03***

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

[1] Full estimation results are presented in Appendix B.

Estimated discontinuities at each BER threshold based on a polynomial order of 3 are presented in Table 1, with the complete model estimates reported in Appendix B. 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 three grade thresholds. Statistically significant negative bunching appears to occur at B3/C1, C2/C3 and D1/D2. Bunching at each of these thresholds is significant at polynomial orders of one, three, and five but not all cases at orders of two and four. However, the bunching estimates at B3/C1, C2/C3 and D1/D2 are not robust to all bin sizes. For instance, while all three are robust to bin sizes of 0.5 $kWh/m^2/yr$ only the bunching at D1/D2 is robust to a 0.75 $kWh/m^2/yr$ size. Overall, this indicates that negative bunching may exist at C2 and D1.

Evidence of positive bunching, i.e. on the more efficient side of the threshold, in the post-works distribution is found at all grades from B3/C1 to F/G. Bunching at grades C2/C3 and C3/D1 are robust to all polynomial orders, as shown in Appendix C. Bunching at the other grades between B3/C1 and F/G are robust to a minimum of two other polynomial orders. All grades where evidence of bunching is found are robust to changes in bin size, as shown in Appendix D. This provides evidence of bunching at various grade thresholds. A significant number of properties have BER assessments just on the favourable side of the threshold than one might ordinarily expect, and there is a absence of properties just on the unfavourable side of the threshold. The absence of bunching in the pre-works distribution

provides evidence that the introduction of benefits or perceived benefits associated with energy efficiency labelling can lead to the occurrence of bunching.

4.2. Estimating the number of 'adjusted' BER assessments

As discussed in section 3.2, we estimate counter-factual distributions of post-works BERs around each threshold, omitting the affected area. We choose to omit homes possessing BERs within 5 *kWh* of the threshold as we assume that bunching of assessments is concentrated in that area. 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 assessments that potentially move from the unfavourable to the favourable side of the threshold. The counter-factual distributions are estimated with a polynomial order of 3.

Figures 4 and 5 graphically compare the observed and fitted distribution to the counterfactual distribution estimated for the pre- and post-works BER assessments at each grade threshold where significant evidence of bunching was estimated. As can be seen, discontinuities exist on either side of most grade thresholds within the affected areas. The pre-works distributions presented show very little, if any, deviations of the counter-factual from the fitted distribution. In section 4.1 we noted that evidence of bunching in pre-works BERs occurred only at the B3/C1, C2/C3 and D1/D2 thresholds. This is also noticeable in the fitted distributions at these grades but not so in the observed distributions at these thresholds in Figures 4 and 5.

Looking at the post-works distributions, the thresholds B3/C1 to F/G all show visual evidence of the variations between the fitted and counter-factual distributions. Clear spikes in the distribution are 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. These are less noticeable in the observed distributions at E1/E2 and F/G grades due to sample size. As can be seen, relative to the pre-works distributions of these grades, the post-works distributions have lower samples sizes.

Table 2: Estimated proportion of assessments in bunching areas

| Threshold | | Estimate | 95% Confidence Intervals | |
Grade	$kWh/m^2/yr$		Lower bound	Upper bound
B3/C1	150	0.00688	[0.00610	0.00765]
C1/C2	175	0.00744	[0.00648	0.00840]
C2/C3	200	0.00874	[0.00780	0.00969]
C3/D1	225	0.01259	[0.01178	0.01340]
D1/D2	260	0.00367	[0.00325	0.00410]
D2/E1	300	0.00526	[0.00495	0.00557]
E1/E2	340	0.00052	[0.00033	0.00072]
F/G	450	0.00031	[0.00024	0.00038]
		0.04542	[0.04092	0.04992]

For each threshold where significant evidence of bunching was found, we estimate the number of assessments occurring in each bunching area. The proportion of assessments in each bunching area are reported in Table 2, along with 95% confidence intervals of the estimates. As discussed in section 3.2, the estimates are calculated as the distance between the observed distribution and the counter-factual distribution. We estimate that 4.5% of assessments occur in these bunching areas, with lower and upper bounds of 4.09% and 4.99%, respectively. This equates to 5,106 assessments of homes who participated in the BEH scheme. 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 shortest distance to the mid-point of any grade. This measure of bunching is presented in Figure 6. 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/C1, C3/D1, D2/E1 and F/G. This suggests that greater emphasis is placed on a property achieving a more favourable letter grade assessment than achieving a more favourable alpha-numeric

Figure 4: Actual, fitted and counter-factual distribution surrounding B- and C-grade thresholds at which significant evidence of bunching was found in the post-works BER distribution
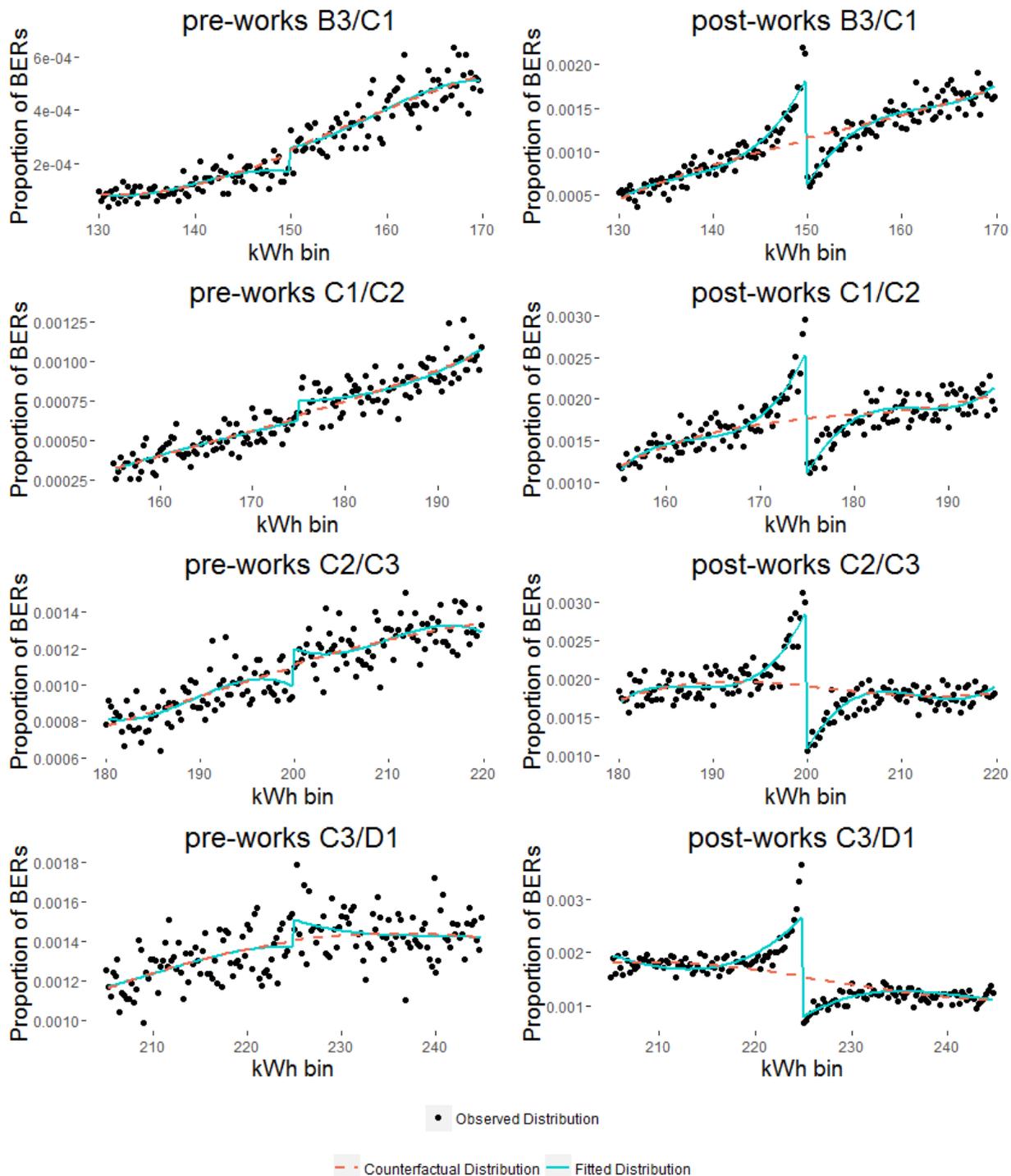
Figure 5: Actual, fitted and counter-factual distribution surrounding D-, E- and F-grade thresholds at which significant evidence of bunching was found in the post-works BER distribution
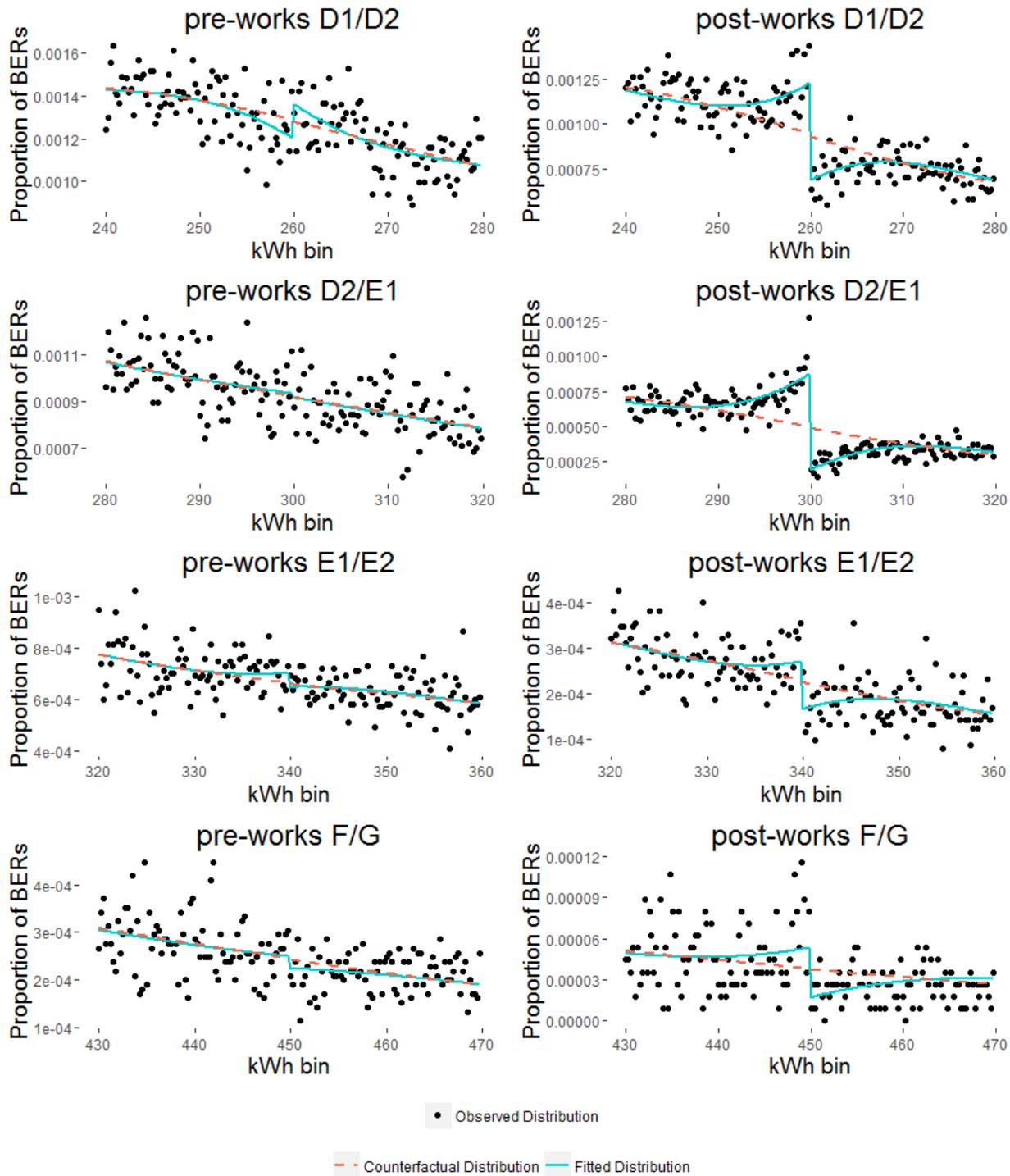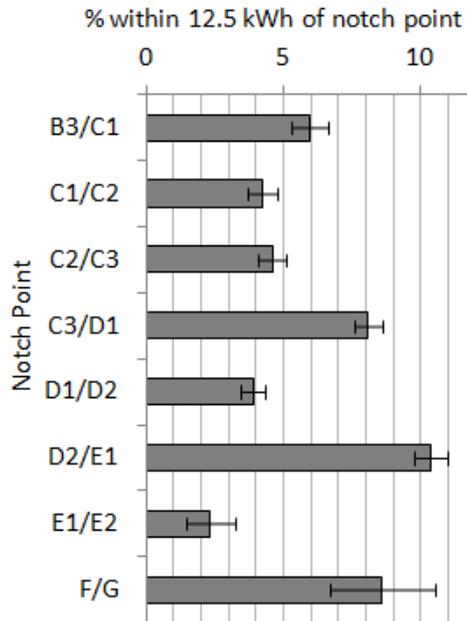
Figure 6: Relative bunching: Bunching area as a proportion of assessments surrounding each notch point

grade within a specific letter grading, e.g. C2 versus C3.

4.3. Drivers of bunching

After finding evidence of bunching in the distribution of BERs it is of interest to identify systemic drivers of bunching. We hypothesise that one potential explanation for bunching is perverse activity by assessors. Such activity may be driven by financial gain or alternatively could be explained by less sinister motives. For instance, some assessors may empathise with home-owners that narrowly miss a more favourable grade and accordingly make a minor adjustment to the assessment.

Our dataset contains unique and anonymous identifiers for all BER assessors, which enables us to examine whether there is any systematic evidence of bunching associated with specific assessors. First, we examine whether certain assessors are more likely to provide assessments in the region of the distribution where bunching occurs, i.e. within $5\ kWh$ of the BER threshold. Examining all assessors with 300 or more assessments we find no evidence that any specific assessors are more likely to have assigned properties in these areas of the BER distribution.

We also examine the data for any regional trends in bunching. Using the same method as in section 3.2, we estimate, for each county, counter-factual distributions at each threshold where significant evidence of bunching was found. The cumulative size of the bunching area in between this counter-factual distribution and observed distribution is recorded for each county. There are no noticeable regional trends with the level of bunching in most counties between 4% to 6% of applications.

As noted above, a competitive market for BER assessor services may encourage assessors to offer more favourable grades to homes in an attempt to generate market share. We use a Gini coefficient to measure competition in the market within each county, assuming that assessors generally operate within their own county. The Gini coefficient measures the polarisation of the distribution of assessors in each county's population of BERs. We calculate the correlation between the Gini coefficient and bunching across counties, with results presented in Table 3. There is a relatively low and variable correlation between competition and bunching. Bunching at the E1/E2 threshold is the only grade to possess a correlation with the Gini coefficient of greater than 0.5. We therefore conclude that adjustment of assessments

12

Table 3: Correlations between assessor polarisation and bunching at each grade

	B3/C1	C1/C2	C2/C3	C3/D1	D1/D2	D2/E1	E1/E2	F/G
Gini Coefficient	0.090	0.031	-0.362	0.017	-0.075	0.095	0.534	0.332
Bunching:								
Across all Grades	0.603	0.334	0.587	0.850	0.549	0.384	0.430	0.126
B3/C1	.	0.178	0.085	0.484	0.341	0.080	0.094	0.147
C1/C2	.	.	0.039	0.176	-0.263	0.116	0.021	-0.349
C2/C3	.	.	.	0.526	0.173	0.093	0.244	0.148
C3/D1	.	.	.	.	0.456	0.118	0.270	0.095
D1/D2	.	.	.	.	.	0.074	0.297	0.199
D2/E1	.	.	.	.	.	.	0.118	-0.080
E1/E2	.	.	.	.	.	.	.	0.115

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 generally low and variable. 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 the case. It is therefore unlikely that BERs in certain counties are more susceptible to bunching than others.

4.4. Sources of Bunching

In the absence of any quantitatively identifiable drivers of adjustment, we seek to examine parameters within BER assessments that are more often associated with bunching. Within the DEAP assessment software some parameters may be more easily adjusted to manipulate an assessment. For example, continuous measures, such as the percentage of the property that has been draught proofed, the percentage of low energy lighting and the thickness of hot water store insulation, can be easily adjusted without enforcement detection in the absence of an audit of the property. Heating system control, response and efficiency categories also vary substantially across dwellings and would be difficult to identify as anomalies without a BER audit.[2] Alternatively, home-owners may undertake particular energy efficiency measures to improve their BER assessment. For example, install low energy lighting. Without home audits it is impossible to distinguish whether bunching associated with a particular activity (e.g. low energy lighting) is due to a perverse adjustment by BER assessors or by home-owners attempting to improve the building's energy performance. Nonetheless, it is of interest to identify if there are particular parameter values within the DEAP software that are more likely to be associated with BERs bunched on the favourable side of a BER threshold.

Using the population of BERs 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 possesses noticeable discontinuities at the BER thresholds. Properties within 5 *kWh* of a threshold on the favourable side possess higher levels of low energy lighting than those in other 5 *kWh* bins, as illustrated in Figure 7. For none of the other parameters was there a such a noticeable discontinuity at the BER thresholds, as shown in Appendix E. While this does not provide enough evidence to conclude with certainty, it is likely that low energy lighting is the parameter most highly associated with bunching at the favourable side of BER thresholds. Low energy lighting is an obvious candidate for home-owners to consider to improve their energy performance and in many instances it is also easy to install. While it is also conceivable that BER assessors make perverse adjustments on this parameter, it is less credible that perverse activity by BER assessors is focused primarily on this one parameter. As noted earlier, without home audits it is impossible to distinguish which is a true explanation but at present this is not convincing evidence of perverse activity.

---

[2]These categories were identified during discussions with SEAI staff.

4.5. Policy Implications

While the analysis in this paper is specific to the Irish EPC scheme, it has implications for the implementation of EPC schemes across Europe and beyond. With bunching evident in the Irish EPC scheme it is likely that bunching is also occurring in EPCs for other EU countries (e.g. Comerford et al., 2016), especially as EPC schemes across Europe have their origins in the same legislation. The presence of bunching suggests that thresholds within an EPC scale compared to a continuous scale provides an incentive to home-owners to surpass the thresholds when they invest in home retrofits. Where continuous EPC scales are used, such as in Germany and the Flanders region of Belgium, additional energy efficiency improvements may be achieved if a multi-point scale is considered. Because energy efficiency price premia are easily associated with multi-point scales (e.g. Brounen and Kok, 2011; Fuerst et al., 2015; Stanley et al., 2016) countries where multi-point EPC scales are used need to be cognisant of the incentive to cheat on EPC assessments and accordingly devise sufficient quality control measures to avoid its occurrence.

Figure 7: Proportion of Low Energy Lighting by 5 *kWh* bin
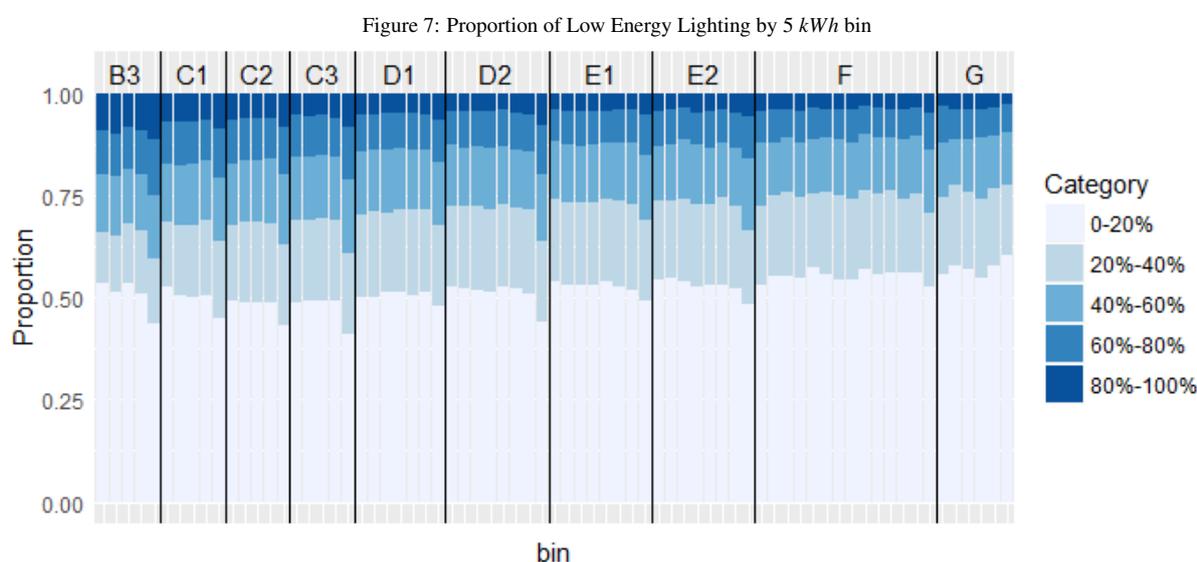
5. Conclusion

BERs, or EPCs as they are more widely known, provide authorities with an overview of the energy efficiency of the building stock, including information on whether policy measures are required to improve residential energy efficiency. BERs are also beneficial to housing market participants, reducing information asymmetry on energy performance and hence on household energy costs and thermal comfort. These benefits are dependent on performance ratings being accurate and properties appropriately labelled, without which confidence in the BER scheme will erode. There is clear evidence of bunching on the favourable side of grade thresholds in the distribution of properties that received grant aid to complete an energy efficiency upgrade. As outlined earlier, three hypotheses were offered to explain why bunching might occur. Bunching could be a result of perverse actions by BER assessors, motivated by property owners seeking to secure property price premia linked to BER grades. Alternatively, home-owners, particularly those that are upgrading their properties, could be motivated to improve the energy performance of their homes. In such circumstances, bunching could be interpreted as evidence that the energy efficiency retrofit grant scheme is nudging home-owners to achieve assessments on the favourable side of the threshold and consequently achieving the grant scheme's aims to increase the energy performance of residential buildings. The third option was that bunching is an artefact of the BER assessment software, though Hyland et al. (2016) already discount this explanation. The

results of the analysis undertaken are not conclusive but do provide useful insight. Without audits of previously assessed properties we do not have irrefutable evidence of which explanation is more applicable. Equally, we were unable to establish any convincing evidence that assessors are engaged in perverse activity. Based on an analysis of a substantial number of energy efficiency measures, the only retrofit measure with a high incidence rate of properties that are 'bunched' on the favourable side of the BER threshold compared to properties outside of the 'bunching' area is low energy lighting. Low energy lighting is easily installed and an obvious candidate for home-owners to consider to improve their property's energy performance. That is the simplest explanation of the bunching in BER assessments.

Between 4–5% of BER assessments within the BEH grant scheme occur in bunching areas. 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. Of the 4–5% of BER assessments in the bunching areas, only a proportion could be considered as 'adjusted' from the unfavourable to the favourable side of the threshold, whether illicitly or otherwise. Assuming that half of 4–5% of BER assessments in the bunching area were 'adjusted', and further assume that a proportion of the balance are accurate assessments, then a relatively small proportion of BER assessments are potentially subject to perverse activity, if it exists. Irrespective of scale, inaccurate BER assessments undermine the BER scheme and the benefits it affords to the property market, while buyers or renters of properties with inflated BER assessments suffer financial losses, due to the price premia associated with BER grades. Preserving the integrity if the BER scheme is of utmost importance, consequently, there should be increased quality control measures focusing on assessments on the favourable side of grade thresholds, including on-site reassessment of properties.

Increased auditing and enforcement of the BER scheme will ensure that bunching, when it occurs, is not due to BER assessors' perverse activity but reflects home-owners efforts to improve energy efficiency. As described in section 2.1, the bespoke BER assessment software displays and continuously updates BER values both by alpha-numeric grade and $kWh/m^2/yr$ plus permits repeatedly revising input values. This software design feature could arguably facilitate cheating by BER assessors, though we have found no evidence of such behaviour. It also provides assessors the opportunity to explore with their home-owner clients, potential options for efficiency gains by comparing between a variety of retrofit improvements. For example, investigate if installation of (minor) retrofit options such as draught proofing or low energy lighting would nudge a BER assessment across a grade threshold. As noted above it is this latter activity that is potentially the source and most reasonable explanation of bunching in BER assessments. Nonetheless, the software design feature that potentially aids assessors and home-owners achieve marginal improvements in energy efficiency could also facilitate cheating. Close monitoring, possibly by the software itself, should ensure that this design feature continues to be used for achieving real energy efficiency improvements.

Acknowledgements

References

Alberini, A., Bareit, M., and Filippini, M. (2016). What is the effect of fuel efficiency information on car prices? evidence from Switzerland. *Energy Journal*, 37(3). http://dx.doi.org/10.5547/01956574.37.3.aalb.

Arcipowska, A., Anagnostopoulos, F., Mariottini, F., and Kunkel, S. (2014). *Energy performance certificates across the EU: A mapping of national approaches*. Buildings Performance Institute Europe (BPIE). http://bpie.eu/uploads/lib/document/attachment/81/BPIE_Energy_Performance_Certificates_EU_mapping_-_2014.pdf.

Atasoy, A. T. (2016). Behavioral responses of green builders to discontinuous certification schemes. *Institute for Future Energy Consumer Needs and Behavior (FCN) Working Paper*, 18/2016:1–35. https://www.fcn.eonerc.rwth-aachen.de/global/show_document.asp?id=aaaaaaaaaauvkwk.

Bevan, G. and Hood, C. (2006). What's measured is what matters: targets and gaming in the English public health care system. *Public Administration*, 84(3):517–538. http://dx.doi.org/10.1111/j.1467-9299.2006.00600.x.

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. http://dx.doi.org/10.1016/j.jeem.2010.11.006.

Brown, K. M. (2013). The link between pensions and retirement timing: Lessons from california teachers. *Journal of Public Economics*, 98:1–14. https://doi.org/10.1016/j.jpubeco.2012.10.007.

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. http://dx.doi.org/10.1108/JCRE-12-2012-0031.

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. http://www.jstor.org/stable/29783791.

Comerford, D., Lange, I., and Moro, M. (2016). The supply-side effects of energy efficiency labels. *Colorado School of Mines Working Paper*, 2016-01.

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.

Einav, L., Finkelstein, A., and Schrimpf, P. (2015). The response of drug expenditure to nonlinear contract design: evidence from medicare part d. *The quarterly journal of economics*, 130(2):841–899. https://doi.org/10.1093/qje/qjv005.

European Union (2002). *Directive 2002/91/EC of the European Parliament and of the Council of 16 December 2002 on the energy performance of buildings*. European Parliament and the Council of the European Union. http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32002L0091&from=EN.

European Union (2010). *Directive 2010/31/EU of the European Parliament and of the Council of 19 May 2010 on the energy performance of buildings (recast)*. European Parliament and the Council of the European Union. http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32010L0031&from=en.

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. http://dx.doi.org/10.1016/j.eneco.2014.12.012.

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. http://dx.doi.org/10.1016/j.enpol.2016.01.024.

Gelman, A. and Imbens, G. (2014). Which high-order polynomials should not be used in regression discontinuity designs. *National Bureau of Economic Research*, (w20405). http://www.nber.org/papers/w20405.

Gourio, F. and Roys, N. (2014). Size-dependent regulations, firm size distribution, and reallocation. *Quantitative Economics*, 5(2):377–416. http://dx.doi.org/10.3982/QE338.

Hyland, M., Alberini, A., and Lyons, S. (2016). The effect of energy efficiency labeling: Bunching and prices in the Irish residential property market. *Trinity Economics Papers, TEP Working Paper*, 516. https://www.tcd.ie/Economics/assets/pdf/tep0516.pdf.

Hyland, M., Lyons, R. C., and Lyons, S. (2013). The value of domestic building energy efficiency: evidence from Ireland. *Energy Economics*, 40:943–952. http://dx.doi.org/10.1016/j.eneco.2013.07.020.

Ito, K. (2014). Do consumers respond to marginal or average price? evidence from nonlinear electricity pricing. *The American Economic Review*, 104(2):537–563. http://www.jstor.org/stable/42920707.

Kleven, H. J. (2016). Bunching. *Annual Review of Economics*, 8(1):435–464. http://dx.doi.org/10.1146/annurev-economics-080315-015234.

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. http://dx.doi.org/10.1093/qje/qjt004.

Lee, D. S. and Lemieuxa, T. (2010). Regression discontinuity designs in economics. *Journal of economic literature*, 48(2):281–355. http://www.jstor.org/stable/20778728.

Matisoff, D. C., Noonan, D. S., and Mazzolini, A. M. (2014). Performance or marketing benefits? the case of LEED certification. *Environmental science & technology*, 48(3):2001–2007. http://dx.doi.org/10.1021/es4042447.

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). http://dx.doi.org/10.1515/1935-1682.3246.

Saez, E. (2010). Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy*, 3(2):180–212. http://dx.doi.org/10.1257/pol.2.3.180.

Sallee, J. M. and Slemrod, J. (2012). Car notches: Strategic automaker responses to fuel economy policy. *Journal of Public Economics*, 96(11):981–999. https://doi.org/10.1016/j.jpubeco.2012.06.005.

SEAI (2013). *Dwelling Energy Assessment Procedure: Introduction to DEAP for Professionals*. Sustainable Energy Authority of Ireland. https://www.seai.ie/resources/publications/Introduction_to_DEAP_for_Professionals.pdf.

Shewmake, S. and Viscusi, W. K. (2015). Producer and consumer responses to green housing labels. *Economic Inquiry*, 53(1):681–699. http://dx.doi.org/10.1111/ecin.12140.

Stanley, S., Lyons, R. C., and Lyons, S. (2016). The price effect of building energy ratings in the Dublin residential market. *Energy Efficiency*, 9(4):875–885. http://dx.doi.org/10.1007/s12053-015-9396-5.

Appendix A. Estimated Akaike information criterion

Table A.4: Estimated Akaike information criterion for each threshold at varying polynomial orders

Grade	Order=1	Order=2	Order=3	Order=4	Order=5
A1/A2	-3939.19	-3937.62	-3935.64	-3936.88	-3938.07
A2/A3	-3521.68	-3522.23	-3519.70	-3520.65	-3522.38
A3/B1	-3177.63	-3179.10	-3179.23	-3179.51	-3179.51
B1/B2	-2820.42	-2833.54	-2835.10	-2835.68	-2835.68
B2/B3	-2588.57	-2597.44	-2597.84	-2598.80	-2598.86
B3/C1	-2330.69	-2394.64	-2421.28	-2427.48	-2426.17
C1/C2	-2281.59	-2318.34	-2359.74	-2366.72	-2365.85
C2/C3	-2254.60	-2309.17	-2357.86	-2357.86	-2358.79
C3/D1	-2800.84	-2889.38	-2914.21	-2914.21	-2979.80
D1/D2	-4033.98	-4047.66	-4068.06	-4067.25	-4067.38
D2/E1	-4424.38	-4506.42	-4541.69	-4547.97	-4548.92
E1/E2	-4772.20	-4772.75	-4778.86	-4778.86	-4780.68
E2/F	-7166.80	-7173.89	-7174.30	-7174.30	-7172.61
F/G	-9568.67	-9571.03	-9581.27	-9583.19	-9580.87

Appendix B. Complete estimation results

Table B.5: Complete estimation of pre-works regression discontinuity design

Threshold $kWh/m^2/yr$	A1/A2 25	A2/A3 50	A3/B1 75	B1/B2 100	B2/B3 125	B3/C1 150	C1/C2 175	C2/C3 200	C3/D1 225	D1/D2 260	D2/E1 300	E1/E2 340	E2/F 380	F/G 450
Constant	0.00	0.00	0.00	-0.01	-0.02	0.21**	-0.05	1.15**	0.31	0.44**	0.09	-0.25	0.05	-0.04
	(0.00)	(0.00)	(0.00)	(0.01)	(0.03)	(0.10)	(0.27)	(0.47)	(0.25)	(0.21)	(0.24)	(0.27)	(0.05)	(0.05)
T	0.00	0.00	0.00	0.01	0.01	-0.07*	-0.01	-0.32**	-0.10	-0.17**	-0.02	0.09	-0.02	0.02
	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.04)	(0.08)	(0.13)	(0.08)	(0.07)	(0.07)	(0.07)	(0.02)	(0.02)
$(X-c)$	0.00	0.00	0.00	0.00	0.00	0.00**	0.00	-0.02**	0.00	0.00**	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$(X-c)^2$	0.00	0.00	0.00	0.00	0.00	0.00**	0.00	0.00**	0.00	0.00**	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$(X-c)^3$	0.00	0.00	0.00	0.00	0.00	0.00**	0.00	0.00**	0.00	0.00**	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$T*(X-c)$	0.00	0.00	0.00	0.00	0.00	0.00*	0.00	0.00**	0.00	0.00**	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$T*(X-c)^2$	0.00	0.00	0.00	0.00	0.00	0.00*	0.00	0.00**	0.00	0.00**	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$T*(X-c)^3$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)
Observations	159	159	159	159	159	159	159	159	199	259	279	279	399	519
R-squared	0.08	0.06	0.05	0.31	0.83	0.91	0.87	0.77	0.36	0.69	0.68	0.62	0.75	0.80

Standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1)

Table B.6: Complete estimation of post-works regression discontinuity design

Threshold $kWh/m^2/yr$	A1/A2 25	A2/A3 50	A3/B1 75	B1/B2 100	B2/B3 125	B3/C1 150	C1/C2 175	C2/C3 200	C3/D1 225	D1/D2 260	D2/E1 300	E1/E2 340	E2/F 380	F/G 450
Constant	0.00	0.00	0.00	-0.03	-0.05	-1.42***	-3.33***	-5.38***	-2.05***	-0.87***	-1.08***	-0.45***	-0.02	-0.07***
	(0.00)	(0.00)	(0.00)	(0.03)	(0.09)	(0.26)	(0.48)	(0.71)	(0.38)	(0.18)	(0.17)	(0.16)	(0.03)	(0.02)
T	0.00	0.00	-0.01	0.01	0.03	0.60***	1.15***	1.70***	0.92***	0.31***	0.40***	0.13***	0.02*	0.03***
	(0.00)	(0.00)	(0.01)	(0.04)	(0.04)	(0.09)	(0.15)	(0.19)	(0.11)	(0.06)	(0.05)	(0.04)	(0.01)	(0.01)
$(X-c)$	0.00	0.00	0.00	0.00	0.00	0.03***	0.05***	0.08***	0.03***	0.01***	0.01***	0.00***	0.00	0.00***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$(X-c)^2$	0.00	0.00	0.00	0.00	0.00	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00	0.00***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$(X-c)^3$	0.00	0.00	0.00	0.00	0.00	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00	0.00***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$T*(X-c)$	0.00	0.00	0.00	0.00	0.00	-0.01***	-0.01***	-0.02***	-0.01***	0.00***	0.00***	0.00***	0.00*	0.00***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$T*(X-c)^2$	0.00	0.00	0.00	0.00	0.00	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00*	0.00***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$T*(X-c)^3$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)	(.)
Observations	159	159	159	159	159	159	159	159	199	259	279	279	399	519
R-squared	0.02	0.13	0.48	0.85	0.95	0.92	0.80	0.79	0.88	0.87	0.90	0.73	0.71	0.44

Standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1)

Appendix C. Varying polynomial order

Table C.7: τ statistics for regressions of varying polynomial order

Threshold		Order=1		Order=2		Order=3		Order=4		Order==5	
Grade	$kWh/m^2/yr$	τ (pre)	τ (post)	τ (pre)	τ (post)	τ (pre)	τ (post)	τ (pre)	τ (post)	τ (pre)	τ (post)
A1/A2	25	0.000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
A2/A3	50	0.00	0.00	0.00*	0.00	0.00	0.00	0.00	-0.01	0.01	0.04
A3/B1	75	0.00	0.00***	0.00	0.00	0.00	-0.01	-0.01	-0.02	-0.01	-0.02
B1/B2	100	0.00**	0.00***	0.00	0.00	0.01	0.01	0.02	-0.14	0.02	-0.14
B2/B3	125	0.00***	0.00***	0.00	0.01	0.01	0.03	0.01	0.03	0.28	-0.13
B3/C1	150	0.00***	0.00*	0.01	0.13***	-0.07*	0.60***	-0.14	-0.22	-0.07*	0.60***
C1/C2	175	0.00	0.00***	-0.03**	0.17***	-0.01	1.15***	-0.32	-0.49	-0.01	1.16***
C2/C3	200	0.00*	0.00**	-0.01	0.29***	-0.32**	1.70***	-0.32**	1.70***	-0.33***	1.71***
C3/D1	225	0.00***	-0.01***	-0.01	0.34***	-0.10	0.92***	-0.10	0.92***	0.09	2.39***
D1/D2	260	0.00**	0.00	-0.03***	0.03***	-0.17**	0.31***	-0.11	0.04	-0.18**	0.35***
D2/E1	300	0.00	0.00	0.01	0.08***	-0.02	0.40***	0.75**	-0.24	-0.02	0.40***
E1/E2	340	0.00**	0.00**	0.02	0.01	0.09	0.13***	0.09	0.13***	0.09	0.13***
E2/F	380	0.00	0.00***	0.00	0.01***	-0.02	0.02*	-0.02	0.02*	-0.71*	-0.02
F/G	450	0.00***	0.00***	0.00	0.00*	0.02	0.03***	0.05	-0.01	0.02	0.03***

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

Appendix D. Varying bin size

Table D.8: τ statistics for regressions of varying bin size

Threshold		Bin = 0.25 kWh		Bin = 0.5 kWh		Bin = 0.75 kWh		Bin = 1 kWh	
Grade	$kWh/m^2/yr$	τ (pre)	τ (post)	τ (pre)	τ (post)	τ (pre)	τ (post)	τ (pre)	τ (post)
A1/A2	25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
A2/A3	50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
A3/B1	75	0.00	-0.01	0.00	-0.01	0.00	-0.02	0.00	-0.03*
B1/B2	100	0.01	0.01	0.02	0.02	0.03	0.04	0.03	0.07
B2/B3	125	0.01	0.03	0.01	0.05	0.01	0.03	0.04	0.11
B3/C1	150	-0.07*	0.60***	-0.14*	1.23***	-0.15	1.77***	-0.28	2.49***
C1/C2	175	-0.01	1.15***	-0.05	2.33***	0.13	1.77***	-0.11	4.62***
C2/C3	200	-0.32**	1.70***	-0.57**	3.39***	-0.60	3.78***	-1.06*	6.47***
C3/D1	225	-0.10	0.92***	-0.20	1.83***	-0.30	2.71***	-0.40	3.62***
D1/D2	260	-0.17**	0.31***	-0.34**	0.62***	-0.38*	0.74***	-0.65**	1.24***
D2/E1	300	-0.02	0.40***	-0.03	0.77***	-0.03	1.15***	0.08	1.51***
E1/E2	340	0.09	0.13***	0.15	0.25***	0.25	0.30**	0.28	0.53***
E2/F	380	-0.02	0.02*	-0.03	0.03	-0.05	0.04	-0.07	0.05
F/G	450	0.02	0.03***	0.04	0.05***	0.05	0.08***	0.09	0.11***

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

Appendix E. Parameters that may be susceptible to adjustment

Figure E.8: Proportion of Heating System Control Categories by 5 *kWh* bin
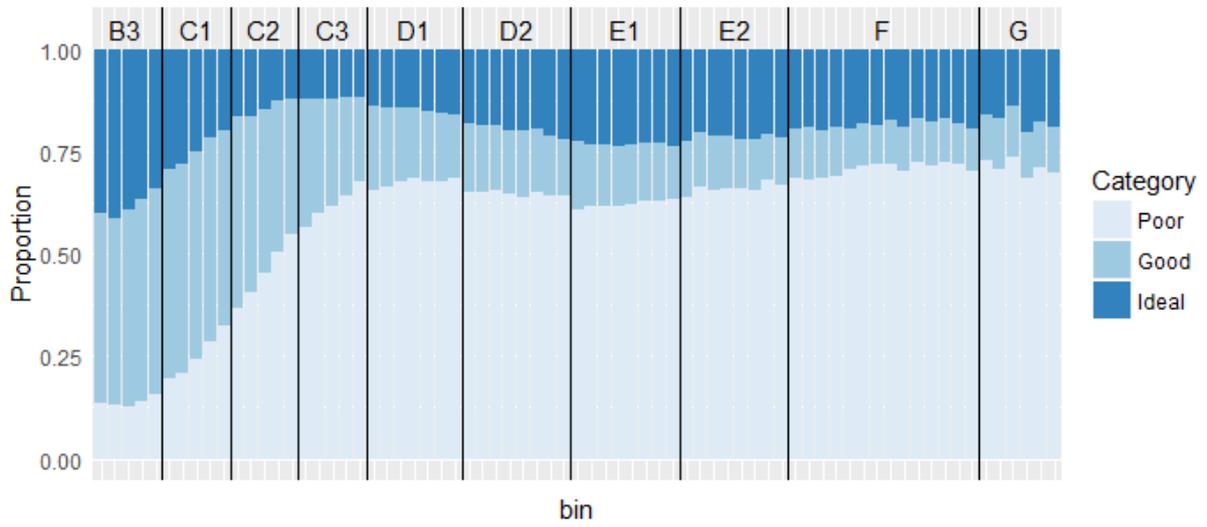

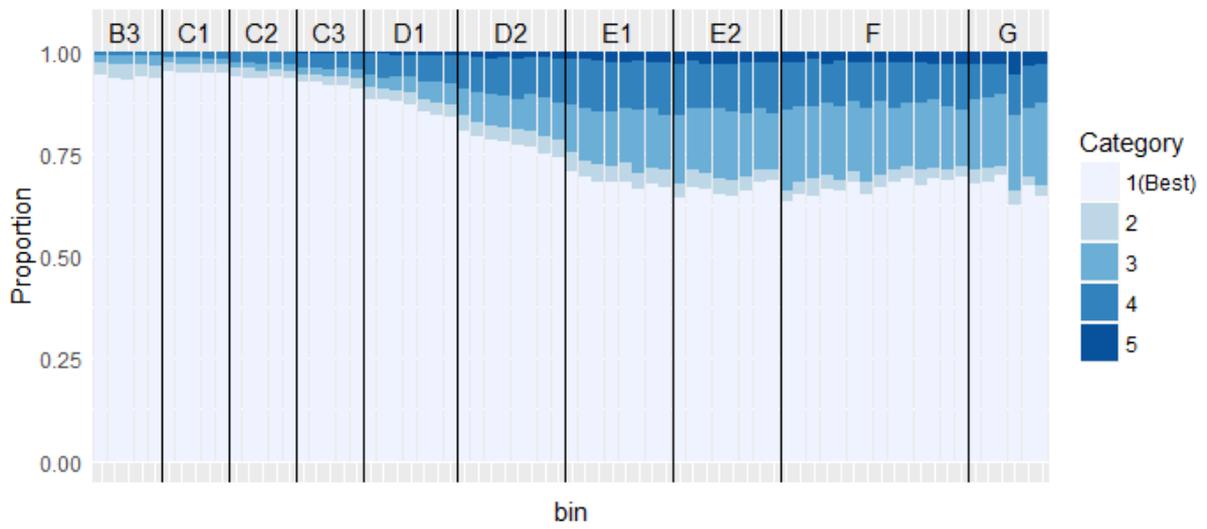Figure E.9: Proportion of Heating System Response Categories by 5 *kWh* bin

Figure E.10: Main Space Heating System Efficiency by 5 *kWh* bin
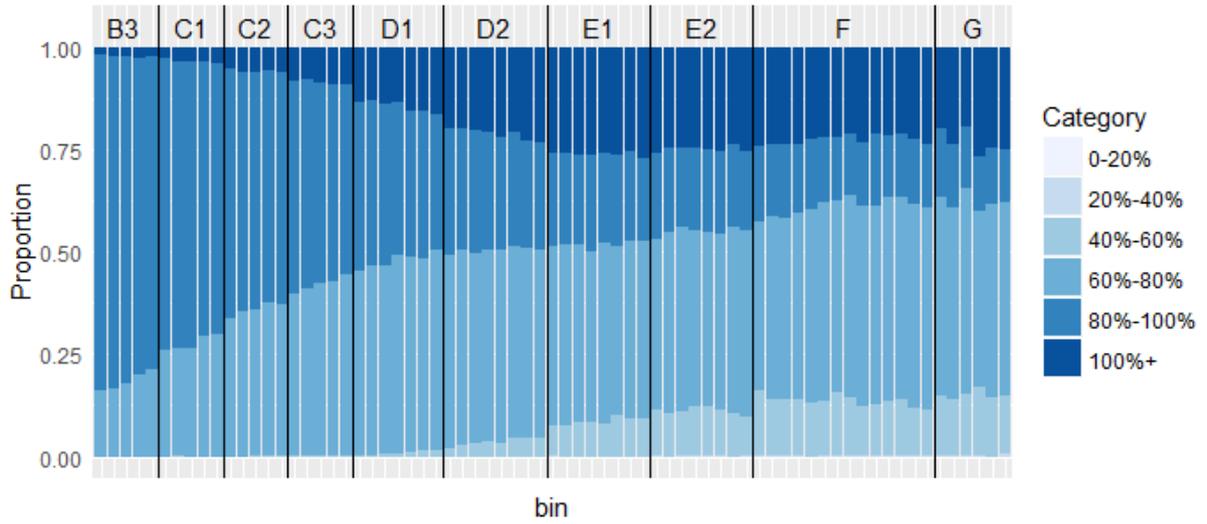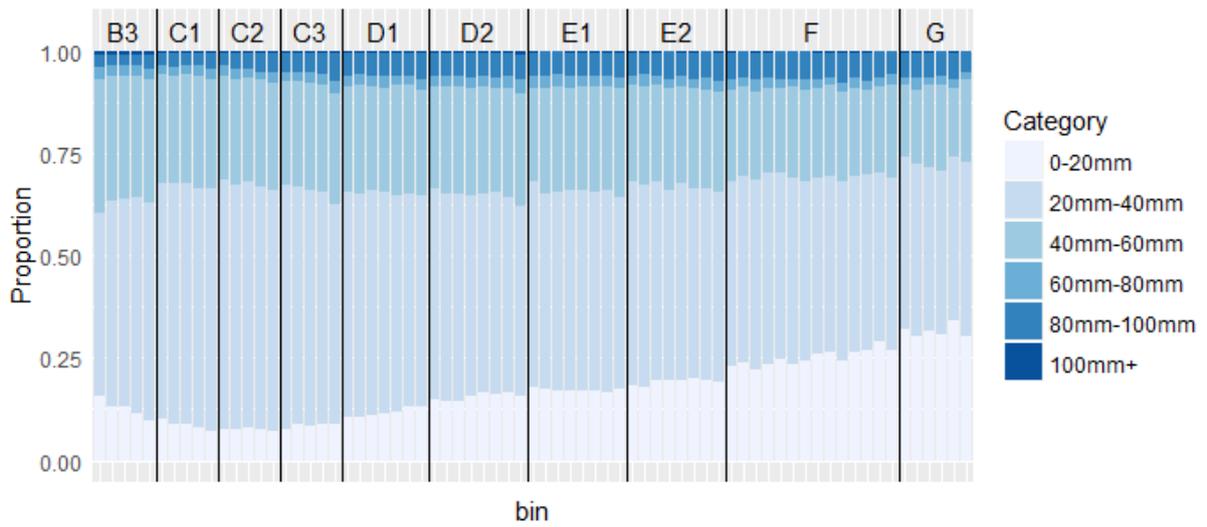

Figure E.11: Hot Water Store Insulation Thickness by 5 *kWh* bin

Figure E.12: *% Draught Proofing by 5 kWh bin*