ESRI Working Paper No. 733

October 2022

How well do building energy performance certificates predict heat loss?

Tensay Hadush Meles[a,b*], Niall Farrell[a,b] & John Curtis [a,b,c]

a) Economic and Social Research Institute, Dublin, Ireland
b) Department of Economics, Trinity College Dublin, Dublin, Ireland
c) MaREI Centre, Environmental Research Institute, University College Cork, Ireland

*Corresponding Author:
Dr Tensay Hadush Meles
Economic and Social Research Institute,
Whitaker Square, Sir John Rogerson's Quay,
Dublin, Ireland
Email: tensay.hadushmeles@esri.ie

Preprint submitted to Elsevier: September 23, 2022

Abstract

This paper evaluates the predictive power of building energy performance certificates on ex-post home heat loss. Improving the insulative capacity of residential properties is a policy priority in many markets, with building energy performance certificates providing the indicative benchmark. We exploit a rich panel dataset of high-frequency thermostat readings, coupled with data detailing weather and buildings characteristics, to identify an ex-post metric of heat loss. Our results show a significant effect of building energy performance rating on indoor temperature, a proxy for home heat loss. However, we do not find evidence of a distinct gradient in performance between building energy rating categories, as suggested by ex-ante estimates of home heat loss.

JEL Codes: C55; D12; Q4; Q55

Keywords: Energy-efficiency; Energy performance gap; Home heating; High-frequency data; Smart thermostat

1. Introduction

This paper evaluates the predictive power of building energy performance certificates on ex-post home heat loss. We find evidence that building energy performance certificates affect indoor temperature, a proxy for home heat loss. However, we do not find evidence to support the distinct gradient along the building energy performance scales as suggested by ex-ante estimates of home heat loss. Our results highlight the presence of much within-classification variance in energy performance.

The EU Energy Performance of Buildings Directive (EPBD) aims to improve building energy efficiency in member states and legislates for the use of Energy Performance Certificates (EPC) to improve information for buyers and sellers on the indicative energy performance of buildings, and to provide guidance on possible energy efficiency improvements (European Union, 2018). EPC ratings, such as those under the EU's 'Energy Performance of Building Directive' and the 'Energy Star Certified Homes' in the United States have been a central element of energy policies to promote investment in energy efficiency and to meet targets of greenhouse gas emissions reduction.[1]

EPCs are broadly used as a policy metric within the residential sector in the context of ambitious climate targets (UNEP, 2021). EPCs by their nature are unable to capture the full nuance of energy performance observed in a property and there is potential for attenuation between expected and observed energy performance (Coyne and Denny, 2021; Cozza et al., 2020; Zou et al., 2018; van den Brom et al., 2018; Gram-Hanssen and Georg, 2018; Majcen et al., 2013; De Wilde, 2014). This difference is known commonly as the Energy Performance Gap.

Many of the empirical studies comparing ex-ante projected energy performance with actual energy consumption are unable to fully disentangle building fabric performance from the intensity of oc-

[1]For details of the European Union (EU) Energy Performance of Buildings Directive: See, https://energy.ec.europa.eu/topics/energy-efficiency/energy-efficient-buildings/energy-performance-buildings-directive.
For different kinds of energy certifications in the USA (including buildings): See, https://www.energystar.gov/about.

cupants' behavioural effects (Coyne and Denny, 2021; Cozza et al., 2020; Zou et al., 2018; van den Brom et al., 2018; Gram-Hanssen and Georg, 2018; Majcen et al., 2013; De Wilde, 2014). For instance, energy consumption data will reflect occupants' preferences for ambient internal temperature, or hot water demand, which are distinct from building fabric performance. Understanding the potential magnitude of such attenuation is important to validate the projected estimates of climate impact measures. This paper provides this contribution, adding to the energy performance gap literature by attempting to isolate building fabric performance from occupant behavioural effects to examine the relationship between building fabric performance and EPC ratings. This is relevant as some recent ex-post evaluations have cautioned policymakers relying on theoretical EPCs' energy use as a mechanism to deliver real energy savings or cast doubt on the projected benefits of an energy efficiency investment (e.g., Levinson, 2016; Fowlie et al., 2018; Davis et al., 2020; Coyne and Denny, 2021).

The present research entails an ex-post evaluation of the effects of EPCs on home heat loss, using indoor temperature as its proxy, in existing residential buildings. To isolate from occupant behavioural impacts, the analysis focuses on data from the early morning hours when the heating system is confirmed as being turned off and behavioural impacts, such as secondary heating, are less likely to arise. We exploit a high-frequency panel dataset of household temperature and heating system operation over a 2-year period. This is in contrast to many studies that rely on metered energy consumption data. These data are matched with information on weather and property energy performance, as measured by EPCs. This allows us to clearly evaluate the impact of building fabric on temperature change within the dwelling, a proxy for heat loss and the insulatve performance of the building fabric.

The remainder of this paper is structured as follows. Section 2 presents the institutional setting. Section 3 outlines the data employed in this analysis. Section 4 provides the empirical strategy. Section 5 presents the results. Section 6 discusses the results. Finally, section 7 concludes the paper.

2. Institutional setting

The EU Energy Performance of Buildings Directive (EPBD) was first introduced in 2002 and recast in 2010 and 2018, with the aim of improving the energy performance of buildings within the European Union (European Union, 2018). Among other measures, the EPBD requires EU Member States to provide information on a building's energy performance through the use of Energy Performance Certificates. The rationale behind this requirement is simple; salient information regarding a dwelling's energy performance can help guide individual decision-making towards the achievement of EPBD energy efficiency goals.

Energy performance certificates provide information to consumers on buildings they plan to purchase or rent. It includes an energy performance rating and recommendations for cost-effective improvements. Certificates must be included in all advertisements in commercial media when a building is put up for sale or rent. This must also be shown to prospective tenants or buyers when a building is being constructed, sold, or rented. Following the EU EPBD, Ireland adopted a mandatory energy performance certificate program. This program began on the first of January 2009 and the certificate is known as the Building Energy Rating (BER). By law, all new homes and homes for sale or rent are obliged to have a BER certificate for the purpose of providing information in advance to prospective tenants and purchasers of the home (SEAI, 2022a).

For each building, the BER certificate provides an estimation of energy use associated with lighting, ventilation, space heating and water heating (SEAI, 2022b). It does not include electricity used for cooking, refrigeration, laundry, and entertainment. The energy performance of a building is expressed in terms of primary energy use per squared metre of floor area per year (kWh/m^2/yr) on a 15-scale from A1 to G and the associated carbon dioxide (CO_2) emissions in kgCO_2/m^2/yr. Figure 1 demonstrates how the 15 BER scales (A1–G) map to the BER in kWh/m^2/yr. The rat-

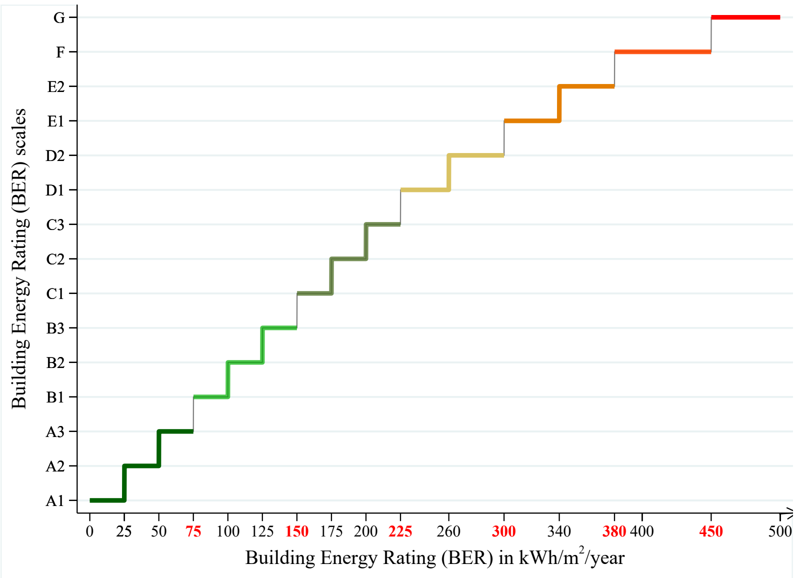

Figure 1: Mapping the BER scales (A1–G) across BER in kWh/m^2/year

ing scale is similar to the EU energy labelling for products subject to energy labelling regulation (EC, 2017). A1-rated properties, with an energy performance rating of 25 kWh/m^2/year or less, are the most energy efficient. On the other end of the scale, G-rated properties, with an energy performance rating of more than 450 kWh/m^2/year, are the least energy efficient (SEAI, 2022b).

The Irish BER certificate is administered by the Sustainable Energy Authority of Ireland (SEAI). The assessments are carried out by SEAI-registered BER assessors and the certificate is valid for up to 10 years. A BER certificate becomes invalid if there are modifications that could significantly affect energy performance (SEAI, 2022a). The BER assessment follows a standardised Dwelling Energy Assessment Procedure (DEAP) where property fabric and heating systems are inspected (DEAP, 2022). The DEAP accounts for factors such as property dimensions (size and geometry); construction material; thermal insulation of building fabric; ventilation (air infiltration due to openings and air tightness of the structure); characteristics of space and water heating systems; solar gains through glazed openings; property thermal storage (mass) capacity; fuel used for heating; and renewable and alternative energy generation technologies.

Data on BER-assessed properties is freely available on the SEAI website.[2] In addition to the BER rating in kWh/m^2/year and corresponding scales, the database contains information on the size and type of property, year of construction, fuels used by a main space heating system, and the thermal transmittance of building fabrics and associated area of exposed and semi-exposed parts of the buildings. As of the beginning of February 2022, BER assessments have been completed on more than 960,000 properties. This corresponds to around 52% of the total number of occupied houses recorded in the 2022 Irish census (CSO, 2022).

3. Data

3.1. Data sources

We wish to analyse the relationship between the recorded Building Energy Rating and the insulative performance of a property, as revealed ex-post by observed temperature change. For this analysis, we use smart thermostat data which provides information on indoor temperature and heating system operation by property. This is matched to two datasets. First, each property is assigned a BER value using the online public search facility. Secondly, the concurrent outdoor temperature and weather conditions are matched using data from the Irish Meteorological Service. Each data source will now be outlined in turn.

High-frequency data detailing household temperature and heating system operation for the main living space[3] of each sample dwelling are sourced from a Hub Controller, an Automatic Energy Manager device with smart thermostat functionality, hereinafter referred to as the 'smart thermostat'.[4] The smart thermostat unit reports this information at regular intervals averaging every three minutes. The gross dataset received comprises 10,000 Irish homes for 24 months: October 01,

[2]Freely available Irish BER database: https://ndber.seai.ie/BERResearchTool/ber/search.aspx

[3]Through personal communication with Hub Controls Ltd., we have learned that each smart thermostat is installed in the main living area of the household.

[4]Additional variables in this dataset include humidity of the living space, thermostat set-points, and whether an operational boiler (gas or oil) is in heating or boost mood

2019–September 30, 2021.[5]

These data are matched with energy performance certificate data from each household's Irish Building Energy Rating (BER) certificate. The energy performance certificate provides information on the household's BER rating, both in terms of primary energy use per squared metre of floor area per year (kWh/m^2/yr) and on a 15-scale from A1 to G. The BER certificate also contains information on dwelling floor area and estimated carbon dioxide (CO_2) emissions in kgCO_2/m^2/yr, alongside information on the reason for obtaining the BER certificate.

The final data source employed in this analysis is local weather data from Ireland's National Meteorological Service, Met Éireann.[6] The weather data consists of hourly air temperature (°C), relative humidity (%), wind speed (knots), sunshine duration (% per hour), and precipitation (mm). This weather data is then matched with the high-frequency thermostat data set at an hourly level, after collapsing the smart thermostat high-frequency data to an hourly level. Properties in the smart thermostat dataset are located in the Greater Dublin area. Consequently, we use data from the Dublin Airport weather station. These data were matched to the smart thermostat data of each property at hourly level.

3.2. Data Processing

We process the data by limiting the analysis to time periods where changes in temperature are plausibly influenced by the observed variables of ambient temperature and BER rating alone. To do so, we restrict the time period of analysis to the main winter heating months in Ireland: December to February. To abstract from occupant behaviour, we limit the data to the hours of 00:00 to 05:59am inclusive, conditional on the heating system being turned off. This is carried out for the following reason; it is plausible that there is no secondary energy input during this time, such as an open fire, and therefore the rate of temperature change is a reflection of the insulative

[5] See https://thehubcontroller.com/ for further information on Hub Controls Ltd.
[6] Source: www.met.ie

capacity of the building. If the heating system is turned on prior to 06:00am, we exclude all subsequent data points from that analysis window. If a heating system is switched off for a lengthy period prior 00:00, it may be difficult to capture how heat loss is associated with BER rating, as heat loss has already occurred. Consequently, we limit the analysis to properties that were heated in 12 hours or less prior to midnight.

The smart thermostat data is recorded in 3-minute intervals, on average. Observations pertaining to the same one-hour interval (i.e. 00.00; 01:00; 02:00; 03:00; 04:00; 05:00) within the 00.00–05.59 observation window are constructed from this raw data. We find the closest recording on or after the required interval (+/- 3 minutes, as the thermostat frequency is 3 minutes). To illustrate, a recording for 00:00 may be required but a smart thermostat may not have recorded the temperature at this exact time. Thus, we extract the date-time stamp and associated indoor temperature of the first reading closest to 00:03 and then retrieve the subsequent temperature readings at intervals of 1 hour. Figure 2 illustrates how the indoor temperature data is extracted from the high-frequency smart thermostat data for an example starting at 00:03 and the subsequent 5 hourly data points (at 01:03; 02:03; 03:03; 04:03; 05:03). While analysis at a sub-hourly frequency is possible, the resource intensity for some of the statistical methods subsequently employed increases non-linearly. Hence, the analysis was undertaken at an hourly frequency without any loss in information pertinent to the analysis. Upon completion of this data processing, 703 properties remain in the dataset with a total of 356,318 hourly observations for analysis.

3.3. Descriptive statistics

This section provides insight into the distribution of the assembled data. First, though not intended to be representative of the national housing stock, the degree to which the matched dataset matches the national distribution is explored. Table 1 compares the distribution of the 703 matched observations to the 967,608 residential properties with a BER assessment as of February 2022. Column (2) in Table 1 shows that about 90% of the sample have a “C1” rated property or lower compared

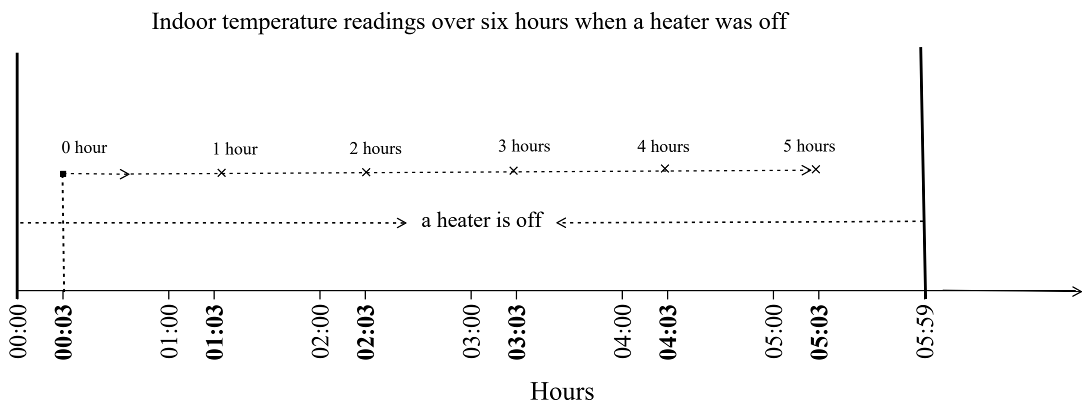

Figure 2: Illustration of data used in the analysis

to approximately 80% of properties in the BER database. The mean BER rating is about 242 kWh/m^2/year for both the smart thermostat sample properties and entire BER database. On average, the 703 properties are older and smaller in terms of property floor area and living room area compared to the national BER database. The average number of years since a BER assessment is similar at approximately six years.

Second, we explore the distribution of indoor temperature and outdoor weather conditions during the sample period. Table 2 provides summary statistics of indoor temperature (°C), outdoor temperature (°C), relative humidity (%), and wind speed (knots) at an hourly level for the 703 sample properties when the heating system was off during the interval 00:00–05:59. The mean indoor temperature in the 703 properties was 16.58°C in the 6-hour period 00:00–05:59 across the three months (December – February) over two years. Table 2 also reports corresponding values for properties by BER rating. There is a high level of variability of indoor temperatures across properties with minimum and maximum values of 6°C and 34°C. The distribution of temperatures

Table 1: Descriptive statistics of sample properties

	Smart thermostat properties	BER database
Variables	Mean/percent	Mean/percent
BER ($kWh/m^2/year$)	243.02	241.95
BER scales (percent):		
A1	0	0.1
A2	0	3.6
A3	2.6	5
B1	0.6	1.4
B2	1.1	3
B3	6.7	7.6
C1	10.1	11.4
C2	14.8	12.6
C3	15.1	12
D1	13.7	11.6
D2	14.2	10
E1	7.4	5.8
E2	6.1	4.6
F	4.8	4.7
G	2.8	6.8
Dwelling type (percent):		
Detached house	7.5	28.7
Semi-detached house	45.7	27.2
End of terrace house	16.2	7.7
Mid-terrace house	27.3	14.1
House (general)	0.1	3.5
Maisonette	0.3	1.1
Ground-floor apartment	1.8	5.4
Mid-floor apartment	0.3	6.5
Top-floor apartment	0.7	5.6
Building age in 2022 (years)	45.9	39.5
Total floor area of a dwelling (m^2)	98.7	111.3
Area of a living room (m^2)	18.6	21.7
Main space heating fuel (percent):		
Gas	78.7	38.3
Oil	18.1	36.5
Electricity	3	17.8
Others (including solid fuels)	0.3	7.4
Years since the BER assessment (in 2022)	5.9	6.6
Purpose of BER (percent):		
Grant support	34.4	21
Sale	41.8	38
Other purposes (including letting)	23.8	41
Number of observations (properties)	703	967,608

are plotted in Figure 3. The solid red line depicts the density of the indoor temperature readings at 00:03 across the 703 sample properties when a heating system was off. The solid black line shows the density of indoor temperature 5 hours after the initial readings at 00:03. The mean indoor temperature declines from 17.62°C at 00:03 to a mean of 15.70°C after five hours. This is an average of 2°C drop in indoor temperature over five hours while a heating system was off throughout.

We further breakdown the average indoor temperature by hours across BER scales. Table 3 shows the average indoor temperature and its difference over hours across BER scales. In the first hour, the overall average drop in indoor temperature is about 0.54°C and it continues to decline and get closer to zero (a steady state point), with a small variations across the BER scales. The decline in temperature after a heating system is turned off is anticipated. The research question is to what extent the decline in temperature systematically varies by BER rating of properties. When comparing the mean temperature values in Table 2 there is a slightly greater decline in temperature among lower energy efficiency rated properties. The next section outlines a more systematic approach to investigate this question.

Table 2: Summary statistics of indoor temperature and weather variables for the 703 sample properties

Variables	Obs.	Mean	Std. Dev.	Min	Max
Indoor temperature (^{0}C) at hour h	356,318	16.58	2.82	6.04	33.88
A3–B3	35,175	17.35	2.31	6.21	29.55
C1	37,526	16.72	2.81	6.29	25.91
C2	53,383	16.76	2.79	6.04	26.82
C3	55,181	16.75	2.73	6.56	28.41
D1	49,152	16.30	2.91	6.17	31.15
D2	52,346	16.55	2.82	6.27	28.18
E1 or E2	43,386	15.75	3.07	6.08	33.88
F or G	30,169	16.51	2.77	6.29	24.92
Average outdoor temperature (^{0}C) at hour h	356,318	4.87	3.40	-5.4	12.3
Average relative humidity (%) at hour h	356,318	86.98	7.55	52	100
Average wind speed (knot) at hour h	356,318	10.02	4.61	1	31

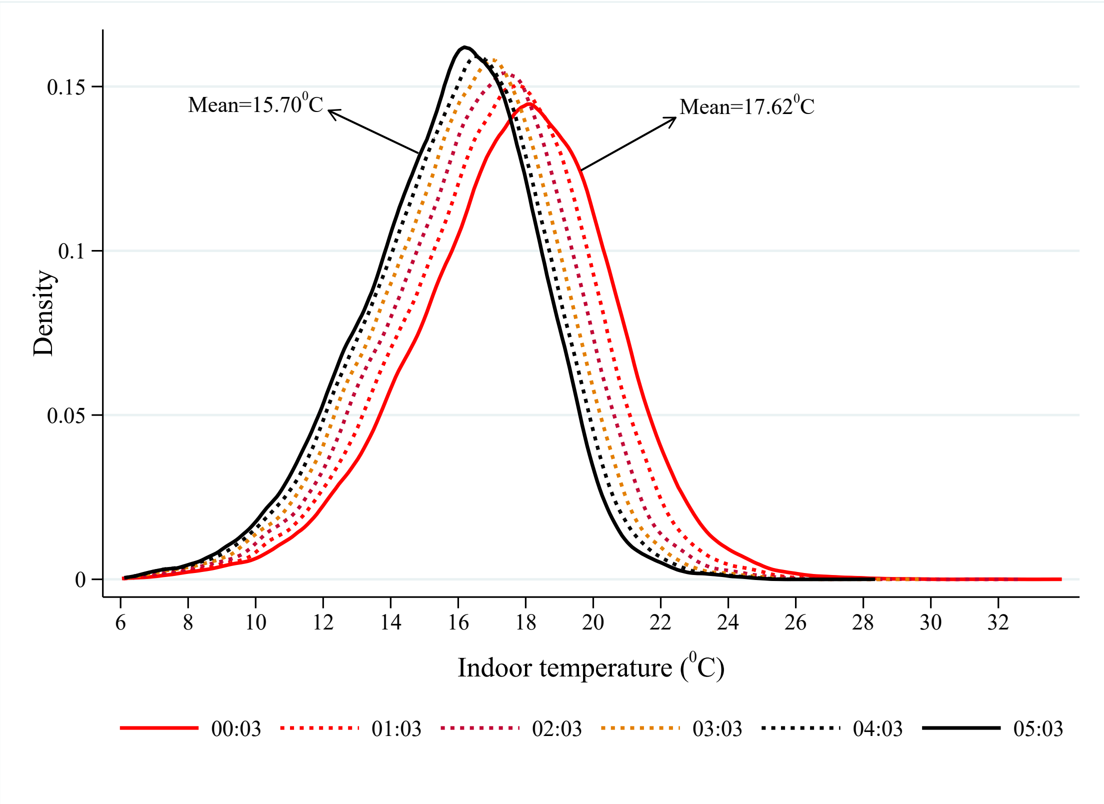

Figure 3: Density of indoor temperature over six hours when a heating system was off

Table 3: Average indoor temperature by hours across BER scales for the 703 sample properties

Hours	All BER	A3 – B3	C1	C2	C3	D1	D2	E1 or E2	F or G
	Panel A: Average indoor temperature by hours across BER scales								
00:03	17.62	18.33	17.64	17.75	17.82	17.45	17.67	16.79	17.57
01:03	17.08	17.83	17.19	17.24	17.29	16.84	17.1	16.23	17.05
02:03	16.62	17.42	16.78	16.78	16.81	16.34	16.61	15.77	16.59
03:03	16.24	17.07	16.45	16.42	16.41	15.93	16.2	15.42	16.2
04:03	15.94	16.78	16.19	16.18	16.08	15.56	15.84	15.11	15.84
05:03	15.70	16.54	15.98	15.97	15.84	15.33	15.57	14.85	15.56
	Panel B: Difference of the average indoor temperature over hours								
01:03 - 00:03	-0.54	-0.50	-0.45	-0.51	-0.53	-0.61	-0.57	-0.56	-0.52
02:03 - 01:03	-0.46	-0.41	-0.41	-0.46	-0.48	-0.50	-0.49	-0.46	-0.46
03:03 - 02:03	-0.38	-0.35	-0.33	-0.36	-0.40	-0.41	-0.41	-0.35	-0.39
04:03 - 03:03	-0.30	-0.29	-0.26	-0.24	-0.33	-0.37	-0.36	-0.31	-0.36
05:03 - 04:03	-0.24	-0.24	-0.21	-0.21	-0.24	-0.23	-0.27	-0.26	-0.28
Observations	356,318	35,175	37,526	53,383	55,181	49,152	52,346	43,386	30,169
Sample properties	703	77	71	104	106	96	100	95	54

4. Methodology

We seek to understand the extent with which building energy performance certificates capture the insulative capacity of the home. We use changes in indoor temperature as a proxy for unobserved heat loss. There are a number of confounding factors relating to energy use and behaviour that must be incorporated into the analysis. For example, households with large heating demand could self-select into A or B-rated buildings. Energy-efficient households may also self-select into better-rated buildings and/or may have different preferences for indoor temperature. To address these and other effects relating to occupants' behaviour, we limit our analysis to early morning hours when a heating system is off. The occupants' behavioural impact is anticipated to be minimal during this time, as potential secondary heating sources (e.g., open fire, portable heaters) are less likely to be operational. In following this approach, we isolate the effects of building fabric from consumers' behaviour on indoor temperature.

The underlying premise of our analytical approach is that temperature within the property in the early morning hours, isolated from occupant behaviour, is a function of three factors. The first and potentially greatest impact relates to temperature inertia. If a property had a high temperature reading one hour ago, its current temperature reading is also likely to be relatively high. This autoregressive approach for modelling heating is widely used (Massana et al., 2017; Fazeli et al., 2016; Fang and Lahdelma, 2016; Powell et al., 2014). To fully exploit this inertia, we limit the period of analysis to the winter months (December–February) when a heating system is likely to be operational. The second factor relates to the insulative capacity of building fabric as measured by BER. This is our subject of interest. The third factor is ambient weather. Internal temperature is affected by external temperature, humidity, and wind conditions.

Modelling temperature as a function of lagged temperature values introduces a potential source of endogeneity as the lagged dependent variable is likely to be correlated with the error term (Anderson and Hsiao, 1981). A common solution is to adopt dynamic panel models using a generalised method of moments (GMM) estimator (e.g., Arellano and Bond, 1991; Arellano and Bover, 1995; Blundell and Bond, 1998; Roodman, 2009). This is a panel dataset with many time periods which presents some modelling challenges. A standard panel comprises N units of analysis (e.g., properties) across T time intervals (e.g., hours or years). There is an excess of 600 time intervals in the current dataset for some properties. Dynamic panel estimators (e.g., Arellano and Bond, 1991; Arellano and Bover, 1995) are designed for situations with small T, as the number of instruments increases quadratically in the number of time periods making estimation of large T models resource intensive and practically difficult. We follow three estimation strategies to address the challenge, which in practice return broadly similar results.

4.1. Standard panel data estimator

In the first we specify a panel data model where the time dimension is the hourly smart thermostat data frequency, while the panel dimension is residential property. The model is estimated using a standard random-effects panel estimator. Such an approach does not address potential for biased coefficients associated with dynamic panels, termed "dynamic panel bias" (Nickell, 1981), however with large T the bias is small.[7] The model is outlined in equation (1):

$$Temp_{ihdmy} = \alpha + \beta Temp_{ih-1dmy} + \gamma Efficiency_i + \delta Weather_{hdmy} + \epsilon_{ihdmy} \tag{1}$$

Where $Temp_{ihdmy}$ is indoor temperature (°C) in property i, at hour h, day d, month m and year y. The indoor temperatures are those recorded by the smart thermostat at hourly intervals in the early morning hours. $Efficiency_i$ is a measure of the energy efficiency rate of property i, of which we utilise several specifications. In the first instance, we use a property's BER assessment, specified as a categorical scales (A–G) or in $kWh/m^2/year$. In alternative model specifications, the sum of the thermal transmittance (U-value denoted in in W/m^{2}K) of the different components of a building: wall, roof, floor, window, and door dimensions multiplied by the corresponding area (m^2) of the building fabrics is utilised. Lower U-values (thermal transmittance) are associated with greater levels of energy efficiency (i.e., highly insulated). External weather variables ($Weather_{hdmy}$) include mean hourly outdoor temperature (°C), mean hourly outdoor relative humidity (%) and wind speed (knot). ϵ_{ihdmy} is the stochastic disturbance term and α, β, γ and δ are parameters to be estimated. γ is our main parameter of interest that captures the effect of building energy ratings on indoor temperature, a proxy for home heat loss.

4.2. Arellano-Bond type dynamic panel estimator

The second estimation strategy is to follow the common approach for estimating panel data with lagged dependent variables, which explicitly address dynamic panel bias (Arellano and Bond, 1991; Arellano and Bover, 1995; Blundell and Bond, 1998; Roodman, 2009). However, as noted earlier such models are designed for situations with small T and estimation with large T datasets is resource intensive. To counter the estimation issues associated with large T in such estimators,

[7]Nickell (1981) show that $\frac{-(1+\beta)}{(T-1)}$ provides an approximation of the bias, where β is the coefficient on the lagged dependent variable as in equation (1). For $\beta = [0.5, 1]$ the associated downward bias is less than 2% for $T > 150$ and therefore in practice almost negligible.

we restructure the data in the following manner. For each property, we use the mean temperature values at each hour for every month and year, as specified in equation (2).

$$\overline{Temp}_{imyh} = \frac{1}{|D|}\sum_{d=1}^{D} Temp_{ihdmy} \quad \forall h, m, y \tag{2}$$

We then specify the time dimension solely as the hour index (h), representing the early morning hours ($h \leq 6$). The panel dimension is represented by an index of property-month-year ($imy >$ 3100). The goal is to estimate the effects of building energy performance but the variable of interest is time-invariant. One strategy to address this problem is to conduct a panel analysis with a two-stage GMM procedure (Kripfganz and Schwarz, 2019). In the first stage, we use the GMM approach to estimate the time variant variables. The model estimated is equation (3), where subscripts i, m, and y from equation (2) are subsumed as a single index representing the property-month-year, though we still use imy as a subscript for clarity. The estimated parameters include α, β, δ, with the $\gamma Efficiency_{imy}$ term dropping out when first differences are taken during GMM estimation. Note that the variable $Efficiency_i$ in equation (1) is equivalent to variable $Efficiency_{imy}$ in equation (3).

$$\overline{Temp}_{imyh} = \alpha + \beta\overline{Temp}_{imy(h-1)} + \gamma Efficiency_{imy} + \delta\overline{Weather}_{myh} + \lambda h + \theta_{imy} + \nu_{imyh} \tag{3}$$

where h is a set of hour dummies, which accounts for correlations across unit of analysis (Roodman, 2009).[8] θ_{imy} is unobserved property specific effects, while $Efficiency_i$ is observed time-invariant. ν_{imyh} is the error term. The description of the variables is similar to equation (1) except the values are the monthly means at each hour.

[8](Roodman, 2009) suggests the inclusion of time dummies is to control for correlations across our unit of analysis as the autocorrelation test and the robust estimates of the coefficient standard errors assume no correlation across units in the idiosyncratic disturbances

The second stage entails estimation of the time invariant parameters to retrieve the γ parameter from equation (3). To do so, we regress the composite residuals from the first stage, $\hat{u}_{imyh}$, on the observed time-invariant variables (i.e., $Efficiency_{imy}$), as illustrated in equation (4). Since we are looking at the effect of the physical building, by excluding occupants' behavioural effects, we assume that $Efficiency_i$ is uncorrelated with unobserved property specific effects, θ_{imy}, or the error term for the second stage estimation, ω_{imyh}.

$$\hat{u}_{imyh} = \overline{Temp}_{imyh} - \hat{\alpha} - \hat{\beta}\overline{Temp}_{imy(h-1)} - \hat{\delta}\overline{Weather}_{myh} - \hat{\lambda}h = \gamma Efficiency_{imy} + \theta_{imy} + \omega_{imyh} \quad (4)$$

In the Difference GMM approach, lagged levels are weak instruments if the coefficient on the lagged variable is close to one (Arellano and Bond, 1991), which is the case in this empirical application. Hence, we implement System GMM with a two-step estimator. Pooled ordinary least squares (OLS) and panel fixed effects specifications are commonly estimated for comparison. While both these estimators are biased and inconsistent due to the correlation between the composite error terms and lagged indoor temperature, their estimates bound the true value. In the OLS regression the lagged temperature is positively correlated with the disturbance terms and provides a coefficient that is biased upward whereas in the fixed effects regression the is biased downward due to the negative sign on the transformed error.

4.3. Individual property level estimates

Our third estimation strategy entails estimating temperature equations at individual property level, as specified in equation (5). The objective in this approach is to illustrate the heterogenity of building performance across the BER scales in contrast to the point estimates from the prior two approaches. With the building fabric constant within individual properties, our focus moves to temperature inertia. Within a single property, $\hat{\beta}$ provides an estimate of how much heat, using temperature as a proxy, is retained within the building fabric after one hour's time while the heating system is turned off. In a property that is not being actively heated, one would anticipate $\hat{\beta} < 1$, with estimated values declining as energy efficiency declines. We utilise the same dynamic panel estimator as previously, (i.e., Roodman, 2009) with the hour index (h) as the time dimension (i.e., early morning hours, $1 \leq T \leq 5$) and the panel dimension represented by the number of days over which data is available ($1 \leq N \leq 181$). $\hat{\beta}$ therefore represents an estimate of the average temperature inertia within a property. We plot kernal densities of $\hat{\beta}$ associated with each BER scale to illustrate both the heterogenity of temperature inertia for a given BER rating and how the densities differ across BER scales. Kolmogorov–Smirnov tests are utilised to test equality of the estimated distributions.

$$Temp_{hdmy} = \alpha + \beta Temp_{(h-1)dmy} + \delta Weather_{hdmy} + \upsilon_{hdmy} \quad \forall i \tag{5}$$

5. Results

5.1. Standard panel data estimates

5.1.1. Main results

We begin presenting the estimates for a standard random-effects panel model. Table 4 presents the parameter estimates for equation (1), with several alternative specifications included. The main model specification is reported in Column (1).[9] The coefficients on the BER scales have a negative sign, indicating a decline in temperature relative to the reference category of A3–B3 rated properties. The absolute value of the coefficients is broadly increasing in magnitude as the BER

[9] Due to a small number of properties for some BER categories (see Table 1), we have amalgamated some scales during model estimation (e.g., E1 and E2, F and G). The reference category includes properties in BER categories A3, B1, B2, and B3, which are the most efficient properties (A3–B3). Note that while the number of hourly observations reported in Table 2 is 356,318, the inclusion of the lagged indoor temperature variable in the regression reduces the number of observations for estimation to 287,211 across the 703 smart thermostat properties. The R-square for the estimated models exceeds 0.92, indicating how well the model explains changes in temperature within each of the households over time.

scale value moves from A to G, with the exception of F or G rated properties. Only for properties rated C3 and below are the coefficients statistically different than zero. The magnitude of temperature decline is greater among the least energy efficient properties, as one would anticipate. However, the gradient of performance decline is much less than one would anticipate. For instance, the decline in indoor temperature for D1-rated property relative to A3–B3 properties is 0.12°C, while point estimate detailing the decline in indoor temperature for E-rated properties is only marginally greater, at 0.15°C. BER categories of C1 to C3 tend to have either insignificantly different degrees of performance or significant differences of relatively small magnitude. BER categories of D1 or greater tend to have significant differences of a relatively greater magnitude.

Contrary to expectation, the magnitude of the coefficient on the F and G-rated properties is not the greatest in absolute value. F and G-rated properties have the lowest assessed level of energy efficiency. This result is potentially a reflection of the composition of the F and G-rated properties in our sample. Over 83% of F and G rated properties had their BER assessment completed in 2014 or earlier. Also, two-thirds completed their BER assessment for the purpose of selling the property. Given the length of the intervening period and the likelihood that the properties were renovated subsequent to sale, there is a strong possibility that the BER ratings of some properties in this category are no longer valid. Given the overall number of F or G rated properties in the sample, at just 54, it is likely that any renovated properties will have a substantial impact on the coefficient estimate. Irrespective of the point estimate for F and G rated properties, the conclusion from the regression estimate remains that a significant difference remains, relative to an A3-B3 -rated property, of a magnitude that is similar to D1-E2 -rated properties.

While the energy efficiency parameter estimates are of primary interest, the coefficient estimate on lagged temperature (β in equation 1) is also noteworthy. The estimate at 0.91 indicates that in the absence of heating, the indoor temperature at any hour will be approximately 90% of the temperature level an hour earlier with other factors such as thermal efficiency and external weather accounting for the balance. As wind speed increases temperature drops more rapidly, whereas the opposite is the case for external temperature and humidity.

Table 4: Estimates from a standard panel data model

Variables	(1)	(2)	(3)	(4))
	Dependent variable: Indoor temperature (^{0}C) at hour h			
	Full sample	Excl. Recent BER	Excl. Grant	Excl. Sale
BER scales (Reference: A3–B3):				
C1	-0.001	-0.057	-0.019	0.007
	(0.036)	(0.057)	(0.046)	(0.040)
C2	-0.048	-0.146***	-0.089**	-0.019
	(0.033)	(0.054)	(0.043)	(0.038)
C3	-0.066**	-0.137**	-0.116***	0.008
	(0.033)	(0.054)	(0.043)	(0.038)
D1	-0.122***	-0.215***	-0.174***	-0.052
	(0.034)	(0.053)	(0.043)	(0.043)
D2	-0.116***	-0.200***	-0.152***	-0.108**
	(0.035)	(0.054)	(0.045)	(0.047)
E1 or E2	-0.152***	-0.246***	-0.189***	-0.143***
	(0.037)	(0.056)	(0.046)	(0.046)
F or G	-0.135***	-0.217***	-0.177***	-0.125**
	(0.039)	(0.055)	(0.046)	(0.060)
Indoor temperature (^{0}C) at hour $h-1$	0.914***	0.915***	0.913***	0.912***
	(0.002)	(0.003)	(0.003)	(0.004)
Average outdoor temperature (^{0}C) at hour h	0.035***	0.035***	0.036***	0.033***
	(0.001)	(0.001)	(0.001)	(0.001)
Average relative humidity (%) at hour h	0.003***	0.003***	0.003***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)
Average wind speed (knot) at hour h	-0.005***	-0.005***	-0.005***	-0.005***
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.744***	0.845***	0.801***	0.792***
	(0.047)	(0.066)	(0.054)	(0.066)
R-squared (within)	0.980	0.979	0.980	0.979
Observations	287,211	207,421	254,617	162,552
Sample properties	703	492	605	409

Robust standard errors clustered at the property level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

5.1.2. Sensitivity analysis

To investigate the robustness of the model estimates reported in Column (1), the same model specification is re-estimated for various sub-sample categories and reported in Columns 2–4 of Table 4. The pattern described above remains broadly the same: BER categories of C1 to C3[10]

[10]or D1 in column 4

tend to have either insignificantly different degrees of performance or significant differences of relatively small magnitude. BER categories of D1[11] or greater tend to have significant differences of a relatively greater magnitude.

These sensitivities were chosen to rule out any possible confounding factors influencing our analysis. In Column (2), properties where the BER assessment was completed after December 2019, which is the start point for the smart thermostat data in our analysis, are excluded. The rationale for this is that recently assessed homes may have had an energy efficiency renovation during the period of the smart thermostat data collection. Excluding these observations precludes this situation. In this instance the sample drops to 492 properties. Broadly, the estimates are similar to those in Column (1) though the coefficients on the BER variables have roughly doubled in magnitude. The largest coefficient, on E-rated properties, is -0.25 relative to -0.15 in Column (1). The pattern observed in Column (1) prevails: there is a statistically significant drop in temperature across the BER scales relative to the reference category, with the difference growing as rated energy efficiency declines subject to the same caveat for F and G rated properties. The differences among grades C3 or lesser are less than the differences among grades E1 or greater, although the distinction is less clear in this specification.

The results in Column (3) exclude properties where the BER assessment was for retrofit grant support from December 2019 onward. BER assessments for grant applications occur after renovation works are completed. In the case where the BER assessment occurred from December 2019 forward, it is possible that the smart thermostat data could cover both before and after the retrofit work. The coefficient estimates on the BER scales in absolute value are somewhat greater than those in Column (1) but less than those in Column (2). The pattern from Column (1) emerges once again: BER categories of C1 to C3 tend to have either insignificantly different degrees of performance or significant differences of relatively small magnitude. BER categories of D1 or greater

[11] or D2 in column 4

tend to have significant differences of a relatively greater magnitude.

The purpose of some BER assessments is for the sale of the property. New property owners often undertake renovation works, some of which could change the energy efficiency status of the property. For instance, the likelihood of fuel system upgrades is much higher when occupancy changes (Curtis and Grilli, 2021). It is feasible that renovation works were completed but an updated BER assessment was not undertaken and registered. In such circumstances, the BER rating linked to the smart thermostat data might not reflect the true BER status of the property. The results presented in column (4) exclude all properties where the BER was undertaken for the purpose of selling the property. The coefficient estimates on the BER scales in this instance are broadly similar to those in Column (1), though only three of the BER coefficients are statistically different than the reference category. Nevertheless, the pattern observed in Column (1) prevails once more, with the lesser performing group extending to include D1-rated dwellings. We see in Column (4) that BER categories of C1 to D1 tend to have either insignificantly different degrees of performance or significant differences of relatively small magnitude. BER categories of D2 or greater tend to have significant differences of a relatively greater magnitude.

While there are some small differences in the coefficient estimates across Columns 1–4, they are broadly similar. Focusing on the BER coefficient estimates, those in Column (2) are the largest in magnitude but still the hourly drop in temperature is less than 0.25°C irrespective of BER rating relative to the most energy efficient A3–B3 rated properties within the sample. It is feasible that secondary heating sources (e.g., open fire, plugged electric heaters) operate for some time after the main heating system (gas or oil boiler) are turned off. To account for this, we re-run the same model specifications but restrict our analysis to hours after 2:00am when secondary heating sources are less likely to operate. Results are reported in the Appendix Table A1 and are broadly the same as those in Table 4. Several other models were estimated based on various sub-samples, for example, weekend or weekdays, and excluding cases of high (>25°C) or low (<15°C) temperatures, with parameter estimates broadly similar to those reported here. The robustness of the estimates across the different samples highlights that neither retrofits undertaken within the analysed period nor properties with a typical heating profiles have disproportionate impact on the estimates.

5.2. Arellano-Bond type dynamic panel estimates

Column (3) in Table 5 presents the first stage System GMM estimates, with the OLS and fixed effects estimates provided for comparison as noted earlier. Also reported in Table 5 are tests that determine validity of the GMM models, including a first and second order serial correlation tests and a Hansen test of over-identifying restrictions. The AR(1) test indicates the presence of first order correlation in the residuals, supporting the argument that the error terms contain unobserved property specific effects. The AR(2) tests fail to reject the null hypothesis that the difference errors in period 'h' and 'h-2' are uncorrelated, indicating that a second lagged value is a valid instrument. Also, the Hansen's test statistic indicates the validity of the instruments.

Table 6 presents the results of the second stage regressions for GMM residuals. The relative patterns of temperature decline across the BER scales broadly matches that of the standard random-effects estimations in Table 4. The negative estimated coefficients on the dummies for BER scales indicate temperature declines relative to the A3–B3 (reference category). BER categories of C1 to C3 tend to have either insignificantly different degrees of performance or significant differences of relatively small magnitude. BER categories of D1 or greater tend to have significant differences of a relatively greater magnitude. Indeed, the difference between relatively high (i.e. C1–C3) and relatively low-performing (i.e. D1–G) BER categories is more pronounced when assessed using the Arellano Bond-type estimator. Columns 2–4 comprise estimates based on different sub-samples of our data, similar to those discussed earlier in the sensitivity analysis in section 5.1.2. In terms of magnitude, the estimated parameters from the GMM residuals are larger than those from standard random-effects, with E-rated properties showing a relatively large decline (a mean of 0.62°C drop per hour relative to the default category).

Table 5: First stage regressions for GMM models

Variables	(1)	(2)	(3)
	OLS	Fixed effects	System GMM
Indoor temperature (^{0}C) at hour $h-1$	0.976***	0.722***	0.736***
	(0.001)	(0.005)	(0.017)
Average outdoor temperature (^{0}C) at hour h	0.055***	0.167***	0.115***
	(0.002)	(0.004)	(0.015)
Average relative humidity (%) at hour h	0.010***	0.027***	0.016***
	(0.001)	(0.001)	(0.006)
Average wind speed (knot) at hour h	-0.002	-0.033***	-0.012
	(0.002)	(0.002)	(0.009)
Hour dummies	Yes	Yes	Yes
Constant	-1.233***	1.562***	2.190***
	(0.096)	(0.133)	(0.617)
Hansen test of overid. restrictions			0.229
Arellano-Bond test for AR(1)			0.00
Arellano-Bond test for AR(2)			0.629
Observations	15,282	15,282	15,282
Number of groups	3,136	3,136	3,136
Sample properties	703	703	703

Robust standard errors in parentheses. p-values are reported for the Arellano-Bond test for serial correlation and Hansen test of the validity of overidentification restriction. The null hypothesis of the Hansen test is H_0: overidentifying restrictions are valid. The null hypothesis of the Arellano-Bond test for serial correlation is H_0: no autocorrelation. *** p<0.01, ** p<0.05, * p<0.1.

5.3. Individual property level estimates

The estimates at individual property level are presented graphically in Figure 4. The distribution of the coefficients on the lagged indoor temperature shown with separate density plots associated with each BER category. This provides insight into the heterogeneity of performance within BER categories.

Figure 4 clearly demonstrates that there is greater within-BER heterogeneity than between-BER heterogeneity. While the greater majority of properties have estimated coefficients in the range 0.8–0.95, there are many properties with estimated coefficients below 0.8. Ex-ante one would have anticipated a clearer difference in the mean performance between property types. However, there is no distinct pattern when observing these plots, further emphasising the findings of the preceding analyses. In addition, Kolmogorov-Smirnov tests fail to reject equality of distributions

Table 6: Second stage regressions for GMM residuals

Variables	(1)	(2)	(3)	(4))
	Dependent variable: GMM residuals			
	Full sample	Excl. Recent BER	Excl. Grant	Excl. Sale
BER scales (Reference: A3–B3):				
C1	-0.101	-0.113	-0.137	-0.082
	(0.095)	(0.149)	(0.124)	(0.102)
C2	-0.179**	-0.314**	-0.281**	-0.102
	(0.087)	(0.146)	(0.113)	(0.107)
C3	-0.183**	-0.276**	-0.304***	-0.013
	(0.083)	(0.140)	(0.110)	(0.094)
D1	-0.315***	-0.443***	-0.445***	-0.076
	(0.089)	(0.142)	(0.113)	(0.109)
D2	-0.274***	-0.410***	-0.364***	-0.202*
	(0.088)	(0.138)	(0.113)	(0.118)
E1 or E2	-0.475***	-0.619***	-0.574***	-0.449***
	(0.093)	(0.140)	(0.116)	(0.114)
F or G	-0.288***	-0.406***	-0.394***	-0.226
	(0.099)	(0.145)	(0.120)	(0.152)
Constant	0.234***	0.377***	0.340***	0.232***
	(0.058)	(0.118)	(0.090)	(0.062)
Adjusted R-squared	0.0388	0.0537	0.0454	0.0348
Observations	15,282	11,055	13,521	8,579
Sample properties	703	492	605	409

Robust standard errors clustered at the home level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

between each of the BER scales. Similar results arise when the samples are restricted in a similar way to those discussed in the sensitivity analysis in section 5.1.2.

6. Discussion

Two striking results emerge from the preceding analysis; (1) we find little between-BER heterogeneity relative to within-BER heterogeneity and (2) we observe a lesser than expected gradient of performance between BER categories.

Table 3 and Figure 4 show that the mean performance of properties across BER scales are broadly similar and are overshadowed by within-category variance. This suggests that factors other than BER have an overwhelming influence of building fabric performance. Our analysis considers temperature change in the main living space, which may vary considerably between dwellings.

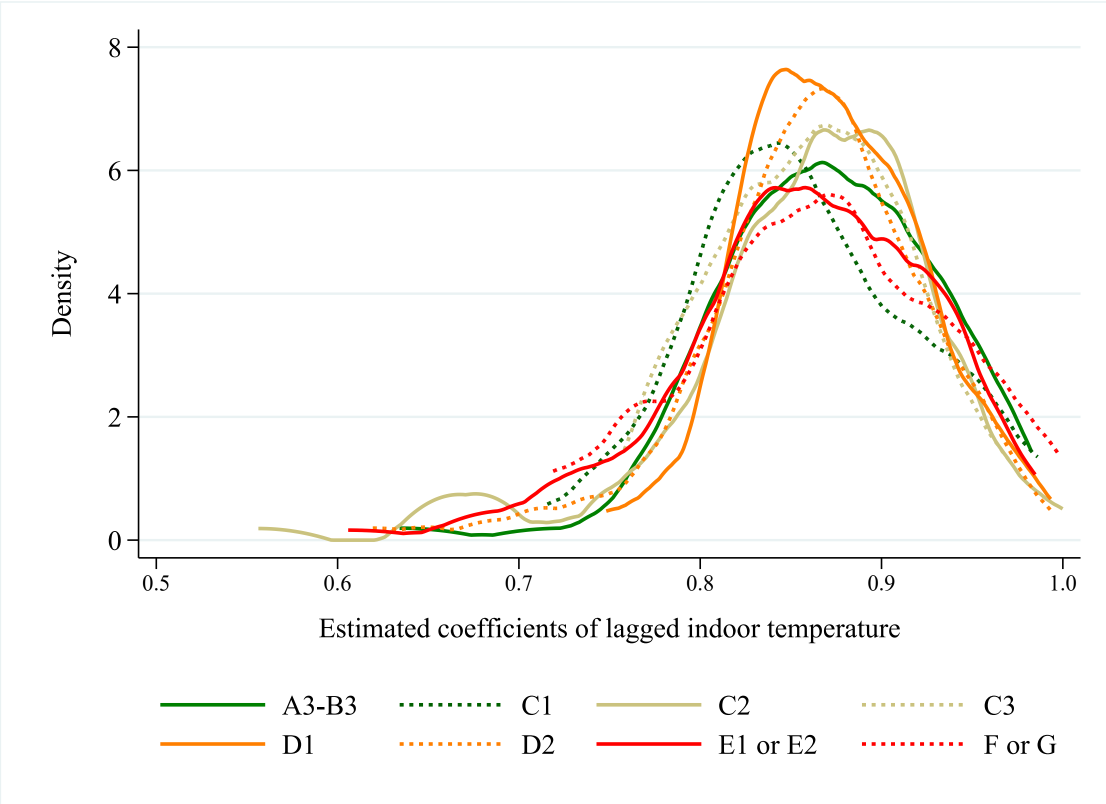

Figure 4: Density of estimates of lagged temperature coefficients at individual property level

However, there is no reason to believe that there is a systematic difference in the distribution of these factors across BER categorisations and a difference in mean performance should still prevail. A substantial share of properties across all BER ratings perform relatively strongly in terms of temperature inertia, while another substantial share of properties across all BER ratings perform relatively poorly.

While BER may be a good standardised approach to measure potential performance across properties, these results suggest that there are additional factors to be considered when evaluating energy performance in the home. From national energy statistics we know that fossil fuel use per household has declined by more than 28% since 2002. This presumably can be attributed to extensive program of residential energy retrofits plus higher building standards. Results from this and similar papers in the literature (e.g. Coyne and Denny, 2021) provides evidence to suggest that relying on theoretical energy performance certificate data may lead to a misspecification of true energy performance in the home. This insight was achieved through the use of ex-post data analysis, both in the case of this paper and that of Coyne and Denny (2021), motivating the incorporation of such data into a more comprehensive energy performance evaluation going forward.

This paper also finds that the performance gradient between BER categories is less than expected ex-ante. While previous research has demonstrated how energy retrofits within the Irish housing stock leads to a reduction in energy consumption (Beagon et al., 2018; Coyne et al., 2018; Rau et al., 2020), none of these studies examines the gradient of performance between BER scales. Broadly consistent with the results here, Coyne and Denny (2021) find a lack of variation in average metered energy use across BER categories among 10,000 Irish properties and conclude that energy demand is unresponsive to the energy efficiency rating of properties. The Irish building energy performance standard (BER/DEAP) is consistent with EU guidance and similar differences between theoretical and actual residential energy performance have been identified elsewhere (Majcen et al., 2013; van den Brom et al., 2018; Cozza et al., 2020).

The research presented in this paper, as well as the findings of Coyne and Denny (2021), suggest that achieving a policy target of retrofitting 500,000 properties to a B2 BER standard (CAP, 2021) may not necessarily lead to the same degree of energy savings as predicted ex-ante. This has important implications for the efficient allocation of public funds, with €8billion earmarked for residential energy retrofits in Ireland (CAP, 2021). The findings of this paper and others in the literature suggests that there are considerable deficiencies in the design of energy performance certificates, with scope for greater emissions reduction per unit of funds spent through a more representative measure of energy performance. Further information is required, however, to understand the missing information.

It is likely to be practically and administratively difficult to design and implement a subsidy scheme that is directly linked to improved energy and emissions performance. What is more feasible is the development of national surveys with appropriate samples and statistical analysis to understand the relationships between energy efficiency standards; energy retrofits; energy use, and; occupant use and behaviours. With more comprehensive information, retrofit grant schemes can be regularly reviewed to ensure the most efficient use of public funds.

7. Conclusion and policy implications

Energy performance certificates (EPCs) are widely used as a benchmark of performance against which residential investment in energy efficiency is measured. Indeed, EPCs form the basis of national programmes of energy efficiency in order to meet climate targets. While energy performance certificates do not purport to be a projection of occupants' actual energy usage, they are used as the basis for public policy. This research, supporting earlier findings by Coyne and Denny (2021), does not find a distinct gradient in performance between BER ratings, lending evidence to suggest that BER is not as strong an indicator of building fabric performance as one would expect ex-ante. In addition, we find that there is a wide heterogeneity of building fabric performance within BER grades, to extent that this is far greater than between-BER heterogeneity.

Two key policy implications follow from this research. Firstly, more research is required to improve our understanding of the relationship between energy efficiency standards, energy use and occupant behaviour. Using national surveys with appropriate samples combined with data from smart meters, data loggers, and other devices controlling heating systems, a substantially better understanding of energy use is feasible.

Secondly, many national policies frame energy efficiency objectives relative to a particular energy performance standard, as measured by energy performance certificates. While energy efficiency retrofits will invariably reduce residential energy use, this research finds that the Irish energy performance certificate captures a relatively small degree of total heterogeneity in energy use; there are many other factors unaccounted for by this metric. Given this finding, directly linking policy targets to a given EPC standard may lead to an outcome substantially different than envisaged.

Acknowledgements

The authors would like to thank Hub Controls Ltd. for access to the smart thermostat data, and to Oliver Hynes, Brendan McGrath, Dan Bowyer, Sonja Zaric and Jim Devlin for helpful discussions. We also acknowledge the funding and support of the Economic and Social Research Institute's Energy Policy Research Centre (EPRC).

References

Anderson, T.W., Hsiao, C., 1981. Estimation of dynamic models with error components. Journal of the American statistical Association 76, 598–606. http://dx.doi.org/10.1080/01621459.1981.10477691.

Arellano, M., Bond, S., 1991. Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. The review of economic studies 58, 277–297. https://doi.org/10.2307/2297968.

Arellano, M., Bover, O., 1995. Another look at the instrumental variable estimation of error-components models. Journal of econometrics 68, 29–51. https://doi.org/10.1016/0304-4076(94)01642-D.

Beagon, P., Boland, F., O'Donnell, J., 2018. Quantitative evaluation of deep retrofitted social housing using metered gas data. Energy and Buildings 170, 242–256. https://doi.org/10.1016/j.enbuild.2018.04.022.

Table A1: Effects of BER certificates on indoor temperature after 2am

Variables	(1)	(2)	(3)	(4))
	Dependent variable: Indoor temperature (^{0}C) at hour h			
	Full sample	Excl. Recent BER	Excl. Grant	Excl. Sale
BER scales (Reference: A3–B3):				
C1	0.007	-0.020	0.008	0.015
	(0.026)	(0.043)	(0.035)	(0.028)
C2	-0.037	-0.097**	-0.053	-0.022
	(0.024)	(0.041)	(0.032)	(0.026)
C3	-0.047**	-0.092**	-0.070**	-0.004
	(0.024)	(0.041)	(0.032)	(0.027)
D1	-0.099***	-0.155***	-0.122***	-0.045
	(0.025)	(0.040)	(0.032)	(0.029)
D2	-0.092***	-0.146***	-0.104***	-0.092***
	(0.026)	(0.042)	(0.033)	(0.033)
E1 or E2	-0.096***	-0.156***	-0.113***	-0.078**
	(0.026)	(0.042)	(0.034)	(0.032)
F or G	-0.100***	-0.152***	-0.116***	-0.094**
	(0.027)	(0.041)	(0.034)	(0.039)
Indoor temperature (^{0}C) at hour $h-1$	0.949***	0.949***	0.947***	0.951***
	(0.002)	(0.002)	(0.002)	(0.002)
Average outdoor temperature (^{0}C) at hour h	0.027***	0.027***	0.028***	0.025***
	(0.001)	(0.001)	(0.001)	(0.001)
Average relative humidity (%) at hour h	0.002***	0.002***	0.002***	0.002***
	(0.000)	(0.000)	(0.000)	(0.000)
Average wind speed (knot) at hour h	-0.003***	-0.003***	-0.004***	-0.004***
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	0.253***	0.326***	0.295***	0.253***
	(0.034)	(0.050)	(0.040)	(0.043)
R-squared (within)	0.988	0.988	0.988	0.987
Observations	166,012	119,636	146,916	94,633
Sample properties	698	488	601	406

Robust standard errors clustered at the home level are in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Blundell, R., Bond, S., 1998. Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics 87, 115–143.

van den Brom, P., Meijer, A., Visscher, H., 2018. Performance gaps in energy consumption: Household groups and building characteristics. Building Research & Information 46, 54–70. https://doi.org/10.1080/09613218.2017.1312897.

CAP, 2021. Climate Action Plan 2021: Securing our Future. Government of Ireland. https://www.gov.ie/en/publication/6223e-climate-action-plan-2021/.

Coyne, B., Denny, E., 2021. Mind the energy performance gap: testing the accuracy of building energy performance certificates in Ireland. Energy Efficiency 14, 1–28. https://doi.org/10.1007/s12053-021-09960-1.

Coyne, B., Lyons, S., McCoy, D., 2018. The effects of home energy efficiency upgrades on social housing tenants:Evidence from Ireland. Energy Efficiency 11, 2077–2100. https://doi.org/10.1007/s12053-018-9688-7.

Cozza, S., Chambers, J., Patel, M.K., 2020. Measuring the thermal energy performance gap of labelled residential buildings in Switzerland. Energy Policy 137, 111085. https://doi.org/10.1016/j.enpol.2019.111085.

CSO, 2022. Census of Population 2022 - Preliminary results. Central Statistics Office. https://www.cso.ie/en/releasesandpublications/ep/p-cpr/censusofpopulation2022-preliminaryresults/housing/. Accessed July 20, 2022.

Curtis, J., Grilli, G., 2021. Does moving home affect residential heating decisions? Exploring heating fuel switching in Ireland. Energy and Buildings 241. https://doi.org/10.1016/j.enbuild.2021.110918.

Davis, L.W., Martinez, S., Taboada, B., 2020. How effective is energy-efficient housing? Evidence from a field trial in Mexico. Journal of Development Economics 143, 102390.

De Wilde, P., 2014. The gap between predicted and measured energy performance of buildings: A framework for investigation. Automation in construction 41, 40–49. https://doi.org/10.1016/j.autcon.2014.02.009.

DEAP, 2022. Dwelling Energy Assessment Procedure (DEAP) software and methodology. Sustainable Energy Authority of Ireland. https://www.seai.ie/home-energy/building-energy-rating-ber/support-for-ber-assessors/software/deap/. Accessed May 20, 2022.

EC, 2017. Regulation (eu) 2017/1369 of the European Parliament and of the Council of 4 july 2017 setting a framework for energy labelling and repealing directive 2010/30/EU. Official Journal of the European Union https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017R1369&from=EN.

European Union, 2018. Directive (EU) 2018/844 of the European Parliament and of the Council of 30 May 2018 amending Directive 2010/31/EU on the Energy Performance of Buildings and Directive 2012/27/EU on energy efficiency). European Parliament and the Council of the European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32018L0844&from=EN.

Fang, T., Lahdelma, R., 2016. Evaluation of a multiple linear regression model and SARIMA model in forecasting heat demand for district heating system. Applied Energy 179, 544–552. http://dx.doi.org/10.1016/j.apenergy.2016.06.133.

Fazeli, R., Davidsdottir, B., Hallgrimsson, J.H., 2016. Residential energy demand for space heating in the Nordic countries: Accounting for interfuel substitution. Renewable and Sustainable Energy Reviews 57, 1210–1226. http://dx.doi.org/10.1016/j.rser.2015.12.184.

Fowlie, M., Greenstone, M., Wolfram, C., 2018. Do energy efficiency investments deliver? Evidence from the Weatherization Assistance Program. The Quarterly Journal of Economics 133, 1597–1644.

Gram-Hanssen, K., Georg, S., 2018. Energy performance gaps: promises, people, practices. Building Research & Information 46, 1–9. https://doi.org/10.1080/09613218.2017.1356127.

Kripfganz, S., Schwarz, C., 2019. Estimation of linear dynamic panel data models with time-invariant regressors. Journal of Applied Econometrics 34, 526–546.

Levinson, A., 2016. How much energy do building energy codes save? Evidence from California houses. American Economic Review 106, 2867–94.

Majcen, D., Itard, L., Visscher, H., 2013. Theoretical vs. actual energy consumption of labelled dwellings in the Netherlands: Discrepancies and policy implications. Energy policy 54, 125–136. https://doi.org/10.1016/j.enpol.2012.11.008.

Massana, J., Pous, C., Burgas, L., Melendez, J., Colomer, J., 2017. Identifying services for short-term load forecasting using data driven models in a smart city platform. Sustainable cities and society 28, 108–117. http://dx.doi.org/10.1016/j.scs.2016.09.001.

Nickell, S., 1981. Biases in dynamic models with fixed effects. Econometrica 49, 1417–1426. https://doi.org/10.2307/1911408.

Powell, K.M., Sriprasad, A., Cole, W.J., Edgar, T.F., 2014. Heating, cooling, and electrical load forecasting for a large-scale district energy system. Energy 74, 877–885. http://dx.doi.org/10.1016/j.energy.2014.07.064.

Rau, H., Moran, P., Manton, R., Goggins, J., 2020. Changing energy cultures? Household energy use before and after a building energy efficiency retrofit. Sustainable Cities and Society 54, 101983. https://doi.org/10.1016/j.scs.2019.101983.

Roodman, D., 2009. How to do xtabond2: An introduction to difference and system gmm in stata. The Stata Journal 9, 86–136. https://doi.org/10.1177/1536867X0900900106.

SEAI, 2022a. Building energy rating certificate (BER). https://www.seai.ie/home-energy/building-energy-rating-ber/.

SEAI, 2022b. Understand a BER. https://www.seai.ie/home-energy/building-energy-rating-ber/

understand-a-ber-rating/. Accessed July 20, 2022.

UNEP, 2021. 2021 Global status report for buildings and construction: Towards a zero-emission, efficient and resilient buildings and construction sector. United Nations Environment Programme, Nairobi. https://www.globalabc.org. Accessed April 29, 2022.

Zou, P.X., Xu, X., Sanjayan, J., Wang, J., 2018. Review of 10 years research on building energy performance gap: Life-cycle and stakeholder perspectives. Energy and Buildings 178, 165–181. https://doi.org/10.1016/j.enbuild.2018.08.040.