Energy Economics 86 (2020) 104699

Contents lists available at ScienceDirect

Energy Economics

journal homepage: www.elsevier.com/locate/eneeco

On the use of hedonic regression models to measure the effect of energy efficiency on residential property transaction prices: Evidence for Portugal and selected data issues

Rui Evangelista [a], Esmeralda A. Ramalho [b,*], João Andrade e Silva [b]

[a] *Instituto Nacional de Estatística, Portugal*
[b] *ISEG – Universidade de Lisboa, Department of Mathematics, REM – Research in Economics and Mathematics, CEMAPRE, Portugal*

ARTICLE INFO

Article history:
Received 28 February 2019
Received in revised form 23 January 2020
Accepted 30 January 2020
Available online 4 February 2020

Keywords:
Energy efficiency
Residential property market
Hedonic price models
Cross-country comparisons
Quantile regression

ABSTRACT

Using a unique dataset of 256,000 residential property sales for Portugal, this paper reveals a clear sales premium for energy efficiency, which is more pronounced for apartments (13%) than for houses (5 to 6%). Price premiums tend to increase from 2009 to 2013, a period in which the Portuguese housing market was depressed. Quantile regression reveals that, when compared with the rest of the sample, the group of the most energy efficient properties receives a statistically relevant price premium, that is mostly stable across the entire spectrum of the conditional price distribution. Cross-country comparisons suggest that energy efficiency price premiums are higher than those found for central and northern European markets. Illustrations on the effect of data issues in hedonic regression models are provided. They show how the use of appraisal prices and explanatory variables with measurement errors may seriously bias energy efficiency partial effect estimates. In contrast, the omission of variables associated with the quality of the properties has not produced relevant distortions. Finally, it became apparent that the use of smaller datasets would have produced similar results, as no significance inflation was produced by the Portuguese large scale dataset.

1. Introduction

In the European Union (EU), buildings account for about 36% of all of the carbon dioxide (CO_2) emissions production and for 80% of the energy consumed for heating and cooling, together amounting to around 50% of the EU's total energy consumption (Directive 2018/844/EU). Given the relevance of the residential sector to the total building stock, the implementation of policies aimed at increasing its energy performance, such as the introduction of energy efficiency labelling schemes, rank among the most feasible ways of reducing CO_2 emissions and mitigating a country's energy dependency. In Portugal, the importance of residential buildings is correspondingly high as they are responsible for 17% of the country's total energy usage and for 27.7% of total national electricity consumption (ADENE, 2018: 16).

Energy labelling has been applied across Europe for many years with household appliances being the earliest general area of application. Energy labels for buildings were first implemented in Denmark in the 1990s (Jensen et al., 2016). More recently, with the introduction of the Energy Performance of Buildings Directive (EPBD) in 2002, which was later recast into Directive 2010/31/EU, EU member states were required to develop and implement an Energy Performance Certificate (EPC) system essentially able to convey the energy performance of property units according to an energy efficiency scale. In the particular case of Portugal, EPCs started to be issued for all new residential buildings in which the total area of all units in the building had more than 1000 square meters from July 2007 onwards, and began to be mandatory in all residential transactions (of both new and existing properties) since the beginning of 2009, and in all properties listed for sale or rent since December 2013.

One core objective of this paper involves assessing whether energy efficiency, as measured by the EPC rating system, generates any impact on the transaction prices of residential properties in Portugal. The relationship between energy efficiency and residential house prices has typically been defined within the framework of hedonic price models (Rosen, 1974), which require the availability of information on dwelling transaction prices as well as on energy efficiency attributes and other price determinants such as size, age, and location quality. This methodology has been applied in a variety of settings, including those by Hyland et al. (2013), Fuerst et al. (2015), Fesselmeyer (2018) and Fuerst and Warren-Myers (2018), who provide results for Ireland, England, Singapore and Australia, respectively. Although the majority of these papers point to the conclusion that increased energy efficiency entails a market price premium at the time of sale (e.g., Fuerst et al., 2016), the relationship between these two variables is far from straightforward. Due to factors such as the anticipation of higher future costs

---

* Corresponding author.
E-mail address: eramalho@iseg.ulisboa.pt (E.A. Ramalho).

https://doi.org/10.1016/j.eneco.2020.104699

from maintaining energy efficiency technology, price premiums may get reduced or even take the form of price discounts (Yoshida and Sugiura, 2015).

The present paper exploits a dataset that includes more than 256,000 property transactions for Portugal completed between 2009 and 2013 with the prices and dwelling characteristics sourced from the EPC information system, transfer and property tax records. The taxation sources cover all property transactions carried out in Portugal, both of new and existing properties as it is impossible to complete a property transaction without proof of transfer tax payment. This leads to a situation in which the EPC directive compliance rate is virtually 100% and hence avoiding any sample selection problems and/or any consequent need for additional corrections to the estimators, typically present in papers addressing the impact of energy efficiency on housing prices; see, for example, Hyland et al. (2013).

This paper offers two key contributions to the literature. First, it constitutes the first large-scale study for a southern European country in which the effect of EPC labels on residential transaction prices is assessed. Although some evidence is given for Spain in Ayala et al. (2016), their results are based on a small sample and on the owner appraisal property valuation instead of effective transaction prices. A detailed analysis is provided based not only on conditional mean hedonic models, including an investigation on how energy efficiency varies across time and regions but also on quantile regression models, which allow for estimating the effects of interest along the distribution of dwelling prices. This paper also provides the results of a cross-country comparability study that enables an assessment of whether energy efficiency is valued more or less greatly in Portugal than in central and northern European countries. Second, the availability of an unusually large and rich dataset allows for the clarification of important data issues associated with the application of the hedonic price model, which has been employed as the lead workhorse in this field of research. In particular, given their practical importance to estimating the partial effects of energy efficiency, three data issues were considered: (i) the influence of large samples on the potential inflation of significance levels of relevant parameters, (ii) the impact of using error-prone measures of the price and property characteristics, and (iii) the importance of the omission of new or rarely used potential price determinants in the hedonic model.

This paper is organized as follows. Section 2 briefly reviews the literature on the impact of energy efficiency on residential property prices, focusing on data issues related to the measurement of variables crucial to specifying hedonic models. Section 3 describes the sources, variables and information available for research and presents some descriptive data statistics. Section 4 sets out the estimations of the effects of energy efficiency on property prices in Portugal based both on the conditional mean and the quantile regression models. This section also includes the results of robustness checks and the cross-country comparison exercise. In turn, Section 5 studies the effects of some of the measurement issues often present in hedonic models designed to capture the effects of EPC rating systems. Finally, the last section provides a summary and a discussion of the main results.

2. Energy efficiency, hedonic models and housing prices

The link between energy efficiency and housing prices encapsulates the concept that markets are able to internalize the benefits of lower energy consumption patterns. For instance, as more energy efficient properties experience lower future utility bills, it is expected that they should display a market transaction premium when compared to their less efficient counterparts; Dinan and Mirarowski (1989). In reality, however, the relationship between energy efficiency and prices is far from straightforward. For example, as Yoshida and Sugiura (2015) point out, in markets with prevailing perceptions of high energy efficiency standards, extra efficiency gains may be regarded as imposing additional technological maintenance costs and result in property market price discounts. Therefore, the relationship between property prices and energy efficiency may not only be insignificant but also take the form of either market price premiums or discounts.

2.1. Measuring the impact of energy efficiency through hedonic price models

The hedonic price model has been widely applied to measuring the effect of property characteristics on their prices. Nelson (1982), who summarises nine studies estimating the relationship between traffic noise and property values, represents one early example. Chin and Chau (2003) and Malpezzi (2003) provide two excellent reviews of the application of the hedonic price model to housing. On the specific case of the effects of energy efficiency on housing values, Fesselmeyer (2018) and Fuerst and Warren-Myers (2018) constitute two recent examples.

The existence of a functional relationship between prices and attributes is central to the hedonic price model. This may be expressed as follows:

$$p^* = \beta_0 + \beta_1 \cdot E + \sum_{k=2}^{K} \beta_k \cdot x_k^* + u. \tag{1}$$

In Eq. (1), p^* and E represent the price and the energy efficiency of the dwelling, respectively. Moreover, x^* corresponds to the remaining housing attributes, * signals how the attribute may be subject to transformation while u is a term for representing additional random factors, which are not measured by the k variables included in the model. In the housing context, typical examples of x include the property's location (Kiel and Zabel, 2008), its area and floor space (Colwell, 1993) and the age of the residential structure (Goodman and Thibodeau, 1995). In contrast, less exploited explanatory factors, which are available in our dataset, are the scenic quality of the location, building technology and the construction quality measured at the individual property level.[1]

This paper focuses mainly on assessing the statistical significance, sign and magnitude of the partial effect of E, as measured by the EPC rating system, on property transaction prices. As noted elsewhere (see, inter alia, Cropper et al., 1988), theory sheds little light on the selection of the appropriate functional form of Eq. (1) and the derivation of the hedonic function is essentially approached as an empirical matter. In the literature dealing with the impact of energy efficiency on prices, p^* typically assumes the form of a logarithmic transformation of p and the explanatory variable of interest E results from the transposition of a discrete measurement scale into one or more binary variables. This paper also follows this approach. In this situation, the relative effect of E can be measured by $r = exp\ (\beta_1) - 1$, an estimator proposed by Halvorsen and Palmquist (1980).[2] In some studies, however, this effect is grasped simply by $\hat{\beta}_1$. Although this provides a reasonable approximation of r when E is included in the hedonic model as a continuous variable (see Megerdichian, 2018), it is not adequate for dealing with the categorical energy efficiency rating scales typically adopted for energy efficiency labelling schemes.

According to the EPC certification system first adopted in Portugal, the energy performance of a property is expressed by a nine-level categorical scale, ranging from A^+, the most efficient level, to G, the least efficient level.[3] The levels attributed reflect a ratio between annual primary energy needs and a reference limit value resulting from

---

[1] The scenic quality of the location refers to situations in which a building or part of the building gains panoramic views of the sea, rivers, mountains or other visual elements susceptible to impact on its market value.

[2] For a more detailed discussion of r estimators, see Kennedy (1981) and van Garderen and Shah (2002).

[3] The G level was eliminated from the EPC certification system following the transposition of the recast EPBD into Portuguese Law by Decree Law no. 118/2013 in December 2013.

calculations made for a property of similar characteristics. When a property obtains a ratio score of less than or equal to 0.25, meaning that it consumes 25% or less energy than the stipulated reference, it receives the A^+ rate. Fig. 1 depicts the adopted scale as presented to dwelling owners in the EPC certificate. As one moves from A^+ to G, the energy needs increase up to a level where G, the least efficient label, corresponds to a situation in which energy consumption is 300% higher than the standard reference rate.

Another issue conditioning the definition of the EPC label scale of measurement stems from the fact that, despite the existence of a common European energy performance framework, the methodology underlying the implementation of the EPC labelling scheme is tailored to the national contexts prevailing in each country and correspondingly preventing direct comparisons of the magnitude of different energy efficiency estimates. For an overview of the different EPC schemes within the EU, see Atanasiu and Constantinescu (2011). However, by introducing some changes to the hedonic price models used in this paper, it is feasible to boost the degree of comparability between studies and present a qualitative cross-country assessment of the energy efficiency impact on dwelling prices; see Section 4.3.

2.2. Data issues

While the application of the hedonic price model for estimating the relationship between market transaction prices and energy efficiency is well established in the literature, there remain important empirical issues that are seldom, if ever, assessed. This situation at least partially derives from the fact that researchers are often confronted with the data they have available and are not able to conduct experiments involving different data contexts. Taking advantage of the quality of our database, we are able to discuss some of these questions.

The first issue this paper addresses are the consequences of using exceptionally large datasets. Since parameters' standard errors decrease as the size of the sample increases, significance levels of energy efficiency and other key variables may be inflated to a point in which the standard *t* and other statistical tests become artificially relevant; see Ziliak and McCloskey (2004). Apart from some notable exceptions (e.g., Lin et al., 2013), this topic has not received much attention. However, this is nevertheless an important matter as, following the dissemination of energy labels, we may expect any problems stemming from the utilisation of large datasets in this area to become increasingly relevant; see, for example, Fuerst et al. (2015), who base their conclusions on a sample of over 330,000 observations.

A second data issue arises from the sensitivity of energy efficiency partial effect estimates to replacing key variables in the hedonic specification by variables that necessarily display some sort of measurement error. A clear example is the use of surrogate, instead of actual, transaction prices in energy efficiency partial effect estimates; see Ayala et al. (2016) where appraisal prices are applied as a proxy for market prices. This usually constitutes a reasonable approach as the dependent variable measurement error is often innocuous in ordinary least squares (OLS) coefficient estimations (see Wooldridge, 2013: 318–20).

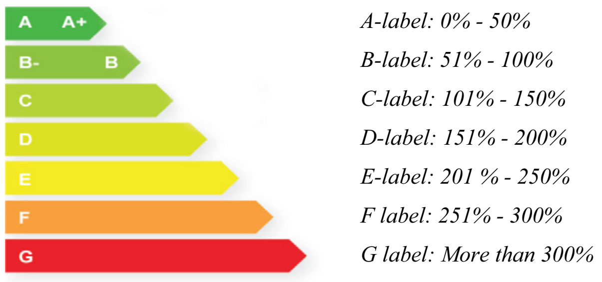

Fig. 1. EPC label scale. Note: % of annual energy consumption needs relative to reference values.

However, when the differences between the proxy and the true transaction prices correlate with some of the covariates included in the model (e.g., dwelling dimension), the OLS estimator of energy efficient partial effects becomes biased and inconsistent. In fact, the application of inappropriate estimators may help in explaining the existence of different energy efficiency price impact results in the literature. As the database used in the present paper includes both the appraisal values of properties for taxation purposes and their transactions prices, we are able to illustrate differences in the coefficient estimates stemming from the application of proxy and actual price measures in the hedonic regression model. On the other hand, the consequences of incorporating erroneously measured property attributes, which are expected to cause the inconsistency of the coefficient estimators, as measurement error of the explanatory variables has in general severe destructive effects, are also illustrated.

One final important data issue revolves around the sensitivity of energy efficiency partial effects to the omission of variables for measuring the transacted property quality. While the omission of relevant covariates appears as the elephant in the room problem in hedonic regression applications, the literature in this field is not particularly prolific in tackling this topic. Stanley et al. (2016) highlight the importance of including controls for dwelling age. Further examples include the omission of locational attributes and dwelling quality, reported by Fuerst et al. (2015) and the non-inclusion of hard-to-measure factors, such as the buyer's predisposition to environmental issues (Brounen and Kok, 2011) and the developer's reputation (Zheng et al., 2012). Despite the fact that variable omission leads only to inconsistencies in the parameter estimators in those cases where the variable omitted correlates with the included covariates (see Wooldridge, 2013: 172), housing attributes, which are themselves hard to measure and often not available in dataset, for example the location's visual attractiveness (e.g. when located on a site with sea views), may correlate with both the EPC label and the other covariates included in the hedonic model (note that properties with outstanding views often possess features, such as panoramic windows, that increase energy consumption and, consequently, negatively correlate with the dwelling's energy performance). By exploring the dataset available by regression analysis, this paper assesses the sensitivity of energy efficiency partial effect estimates to the omission of key and hard to measure variables.

3. Data and descriptive analysis

The dataset analysed in this paper combines energy efficiency data sourced from the national supervision body responsible for implementing and running the European EPC system in Portugal (ADENE) with transaction prices and dwelling attributes obtained from the Portuguese Tax and Customs Authority. The transaction prices were obtained from real estate transfer tax (IMT) records and property attributes were taken from local property tax (IMI) data.[4] The IMT source, which is available from 2009 onwards, provides the overall residential property sale population as it is not possible to carry out a transaction without proof of payment of this tax. Similarly, the energy efficiency data does provide very good coverage of the population of all dwelling transactions taking place from 2009 onwards due to Portugal being one of the first European countries to implement a sound system of practical EPC label enforcement. In contrast with other countries (e.g., the Netherlands; see Brounen and Kok, 2011), the EPC system was not adopted on a voluntary basis but with the mandatory status of energy certificates imposed by the Portuguese legal system, firstly for new buildings (as of July 2007 onwards) and then for existing buildings from the beginning of 2009 (Dispach no. 461/2007). The EPC system and the methodology underlying the energy rating

---

[4] The real estate transfer tax is designated as *Imposto Municipal sobre a Transmissão Onerosa de Imóveis* or simply as IMT. The local property tax is entitled *Imposto Municipal sobre Imóveis* (IMI).

measurement underwent a revision, which came into effect in December 2013 (Decree Law no. 118/2013). To avoid any possible adverse effects stemming from this change in measurement methodology, this paper focuses on the 2009 to 2013 period, a timeframe overlapping with recession in the Portuguese residential property market; Lourenço and Rodrigues (2017).

The matching of the IMT, IMI and ADENE data sources took place through recourse to the property register identification number, a unique code attributed to each property unit, as the key linking variable. Due to the existence of incomplete (or inaccurate) information in the ADENE location variables, essential to constructing the property register identification number, it was only possible to match 60% of the transactions available in the taxation records. Despite this drawback, the end-product of this merging process resulted in a dataset containing information on the prices, energy performance and other dwelling characteristics of 256,145 residential property transactions; see the description of the data matching process in the supplementary appendix. This constitutes one of the largest datasets ever applied for estimating the impact of energy efficiency on transaction prices. The IMT and IMI data currently serve for the compilation of the residential and commercial property price indexes for Portugal (INE, 2017a, 2017b). A subset of the latter data was used in Ramalho et al. (2017) in an empirical exercise involving the construction of hedonic price indexes with aggregate information which, however, did not include the EPC rating as a price determinant. Table 1 provides the summary statistics for the dataset group of selected variables. In addition to these totals, this also presents descriptive statistics for four market segments: existing apartments, new apartments, existing houses and new houses.

The data reveals clear differences among the different dwelling types. As expected, new is generally more expensive than existing and houses are more expensive than apartments. For instance, while existing apartment prices present an average value of nearly €97,000, new houses register a much higher average of approximately €180,000. Existing apartments stand out as the most commonly sold property in the dataset (accounting for 59% of all observations) and with new houses as the least frequently purchased property type (5% of all transactions). Naturally, the price dispersion is also much higher for houses than for apartments, suggesting that the former property type is more heterogeneous than the latter and that the hedonic model specification may be more complex for houses than for apartments. The data also demonstrate how the appraisals carried out for fiscal purposes generally define lower prices than the transaction prices with the average value of the former value being 21% to 34% lower than the latter.

The most common energy rating in the transactions data is C (36.3%). However, while new apartments most frequently receive the B label rating (39.3%), in the case of houses, the most common is the less energy efficient D rating (25–33% of all observations). Although higher for new properties, the percentage of A^+ or A rates is relatively low (4.9% of total transactions). When grouped with the residential unit transactions bearing a B or B^- label, the percentage of transacted dwellings rises to 37.9%. On the other hand, low rated properties labelled E, F or G, are rarely transacted, especially for new apartments.

Based on these descriptive results and in line with other studies in this field (see, inter alia, Brounen and Kok, 2011), it was decided to grasp the information on energy efficiency combining two main specification approaches. The first portrays energy efficiency by dummy variables for all nine levels of the energy efficiency scale considered in Portugal (see Fig. 1). The C-rate, which identifies energy consumption needs 1% to 50% above the standard, is withdrawn from the regression models to serve as the reference. The second approach condenses the information on energy efficiency into a single dummy variable, which assumes the value of 1 when the property sold was attributed one of the four top rates in the EPC scale (i.e. A^+, A, B and B^-). By aggregating the A and B labels into a single dummy variable, designated as AB, we are able to focus the analysis on all residential units signalled as having annual consumption energy needs that are either the same or are lower than the standard reference consumption values. Moreover, this approach is flexible enough to provide a good summary of the effects of energy efficiency across the distribution of prices and the consequences of some sampling issues evaluated in simulation experiments without having to make the models too burdensome due to the inclusion of an excessively large number of terms of interaction.

The differences across market segments are also made explicit by other variables. In terms of dimension, the gross floor area is larger for houses than for apartments and, interestingly, there is a clear difference between new and existing dwellings, with the latter property type smaller than the former. Moreover, the summary statistics for age demonstrate that existing houses display an average age of 29 years while the set of transacted new houses show an average of 2 years. It should be noted that, while older (in age) properties are not expected to be classified as new, there may be cases of properties classified as new with several years of existence (e.g., newly built homes that, due to the market recession, remained on the market for a substantial period before being sold). Our data analysis also reveals a relatively high percentage of new houses completed in or before 1991, a figure explained by the existence of major improvements and renovations, which accounts for a significant proportion of this dwelling category (35.9% of

Table 1
Summary statistics of a group of selected variables.

Variable description	All (N = 256,145)		Existing apartments (N = 149,920)		New apartments (N = 59,410)		Existing houses (N = 33,282)		New houses (N = 13,533)	
	Mean	Stdev	Mean	Stdev	Mean	Stdev	Mean	Stdev	Mean	Stdev
Transaction value (€)	119,888	98,131	97,695	68,876	149,007	93,688	143,783	150,717	179,155	145,671
Fiscal Appraisal value (€)	89,634	61,116	76,752	46,097	111,038	57,445	95,137	85,986	124,852	96,557
Energy label A^+	0.007	0.083	0.002	0.043	0.020	0.141	0.005	0.068	0.009	0.096
Energy label A	0.042	0.200	0.014	0.117	0.118	0.323	0.021	0.145	0.061	0.239
Energy label B	0.202	0.402	0.159	0.366	0.393	0.489	0.073	0.259	0.157	0.364
Energy label B^-	0.128	0.334	0.120	0.325	0.157	0.364	0.110	0.313	0.142	0.349
Energy label C	0.363	0.481	0.468	0.499	0.228	0.419	0.185	0.388	0.222	0.415
Energy label D	0.148	0.355	0.132	0.338	0.062	0.241	0.330	0.470	0.248	0.432
Energy label E	0.086	0.28	0.095	0.293	0.018	0.133	0.160	0.366	0.099	0.298
Energy label F	0.021	0.142	0.009	0.096	0.002	0.045	0.093	0.291	0.048	0.214
Energy label G	0.005	0.067	0.001	0.032	0.0004	0.021	0.024	0.152	0.014	0.117
Gross floor area (m2)	110.6	50.7	96.0	34.0	113.1	38.0	148.3	79.7	168.6	67.4
Dependent floor area (m2)	31.1	39.1	18.6	21.1	36.1	26.3	60.2	66.7	75.4	64.4
Uncovered land area (m2)	78.2	375.0	2.9	15.8	4.9	21.6	441.0	797.3	415.6	788.0
Number of bedrooms (#)	2.5	1.2	2.3	1.0	2.3	0.9	3.5	1.8	3.3	1.1
Age of property (years)	16.1	18.9	20.1	17.5	2.0	2.1	29.3	25.9	1.5	2.1
Improv. or renewed property	0.054	0.226	0.015	0.120	0.089	0.285	0.045	0.207	0.359	0.480

the total, as detailed in the last row of Table 1). In these cases, the previously old property goes onto the market listed as a new property and is considered as such in the database. Given the lesser variation in the age of new dwellings, it may be argued that this variable thus holds less importance to explaining the formation of price for this property type than it does for existing dwellings. However, as renovated properties may display vintage effects, it was decided to include information on the year of their completion in all the hedonic models.

4. Hedonic models for property prices in Portugal

The key question approached by in this paper, within the framework of the hedonic price model, is whether or not energy efficiency impacts on transaction prices. Given the evidence provided for other countries (see, inter alia, Fuerst et al., 2016), the relationship between these two variables is expected to be positive.

The modelling approach builds on the market segments identified in the descriptive analysis of Section 3. Partly due to the lack of data, many hedonic housing studies focus only on a particular market segment (e.g., the housing market segment of a national capital) and do not have to address such issues. On the other hand, this involved taking special care over including in the models all likely price-determining variables suitable to capturing the impact of energy efficiency on residential property prices; see Chin and Chau (2003) for a comprehensive list of the housing attributes applied in hedonic price models. The complete list of all the explanatory variables in our regressions is provided in the supplementary material.

The results discussed in this section derive mainly from the conditional mean hedonic models, estimated for the four market segments under analysis. First, estimates for the impact of energy efficiency are discussed for both a disaggregated 9-level scale and the single dummy case where properties more and less efficient than the standard are distinguished. Then, a time and space analysis of the impact of interest is developed in Section 4.2. Section 4.3 presents both some robustness analysis and cross-country comparisons. Finally, Section 4.4 presents quantile regression analysis results for the full range of property prices.

4.1. Linear regression results

Table 2 presents the OLS energy efficiency coefficient results for the different EPC labels featured in Fig. 1. In addition to information on the sign, magnitude and statistical significance of the coefficient estimates of interest, the table also provides statistics attesting to the quality of the chosen models. For the sake of space, the tables provided in this paper focus on the energy efficient partial effect results with the complete regression results made available in the supplementary material. Additionally, Fig. 2 displays the confidence intervals for the coefficients of interest.

In general, the expected positive and negative impact is observed for the labels, respectively, above (A and B-type) and below (D-G-type) the reference consumption value. However, both the statistical significance and the magnitude of the label impacts differs between apartments and houses. In terms of the A and B-type label price premiums, while for apartments the statistical significance is clear, the pattern is more complex for houses. For existing houses only the B^- label revels a significant premium but still modest in magnitude, while for new houses the top three labels A^+, A, and B report significant rewards, which, however, are 50% less than those observed for apartments. Despite these differences, apartments and houses share a similar pattern of impact across the significant coefficients: the reward for A^+ and A rates is similar, then the decline triggered by dropping from an A to a B rating ranges from 27.6% to 37.2% and with the drop in the impact resulting from going from B to B^- being greater than 50%. This suggests a clear pattern of increasing premium declines as the energy efficiency falls in comparison with the standard value.

In contrast, the impact of the D-G low energy scores significantly differs from the C reference for both types of housing but remains insignificant for new apartments classified as F and G and only attains significance at the 10% level for existing apartments issued E and G certificates. We would also note that for the D-G rates, the confidence intervals for the impact of different labels often overlap, at least partially, indicating that these rates, denoting the lowest efficiency levels, often do not differently affect the price. Finally, note that the wide confidence intervals for the G label, especially for new apartments, are certainly a consequence of the reduced number of transactions in this category (only 24); see Table 1. Therefore, despite the fact that increasing discounts emerge as the level of efficiency drops off, their magnitude is similar across property types and does not greatly differ across energy labels. This suggests that, especially for apartments, but also on a smaller scale for new houses, the market reaction to good energy performance is of a higher magnitude than to low performance standards. Note that a similar pattern of absence of statistical significance in differences in

Table 2
Estimation results for the relationship between energy rating and price.

	Existing apartments	New apartments	Existing houses	New houses
Energy efficiency score:				
A^+	0.236^{***}	0.229^{***}	0.020	0.088^{**}
	(0.0177)	(0.0080)	(0.0419)	(0.0317)
A	0.207^{***}	0.204^{***}	−0.008	0.077^{***}
	(0.0064)	(0.0038)	(0.0204)	(0.0130)
B	0.150^{***}	0.135^{***}	0.010	0.048^{***}
	(0.0024)	(0.0028)	(0.0101)	(0.0086)
B^-	0.075^{***}	0.057^{***}	0.022^{**}	0.008
	(0025)	(0.0033)	(0.0086)	(0.0091)
D	-0.014^{***}	0.014^{*}	-0.033^{***}	−0.019
	(0.0026)	(0.0056)	(0.0065)	(0.0081)
E	−0.006	-0.036^{***}	-0.041^{***}	-0.050^{***}
	(0.0032)	(0.0103)	(0.0081)	(0.0124)
F	-0.028^{**}	−0.001	-0.059^{***}	-0.046^{**}
	(0.0095)	(0.0357)	(0.0099)	(0.0174)
G	−0.064	0.067	-0.076^{***}	-0.124^{***}
	(0.0374)	(0.1053)	(0.0206)	(0.0328)
Root MSE	0.303	0.246	0.413	0.325
R^2-adj	0.678	0.740	0.671	0.754
Reset, P-value	0.495	0.366	0.093	0.044
Sample size	149,920	59,410	33,282	13,533

Notes: The omitted energy class is C. *** significance at the 0.001 level, ** significance at the 0.01 level, * significance at the 0.05 level. Robust (to the presence of heteroskedasticity) standard errors in brackets. The Ramsey's (1969) Reset specification test results were obtained using a robust to the presence of heteroskedasticity procedure, which is based on Lagrange-Multiplier (LM) statistics, developed by Wooldridge (1991). Test hypothesis H0: Correct specification.

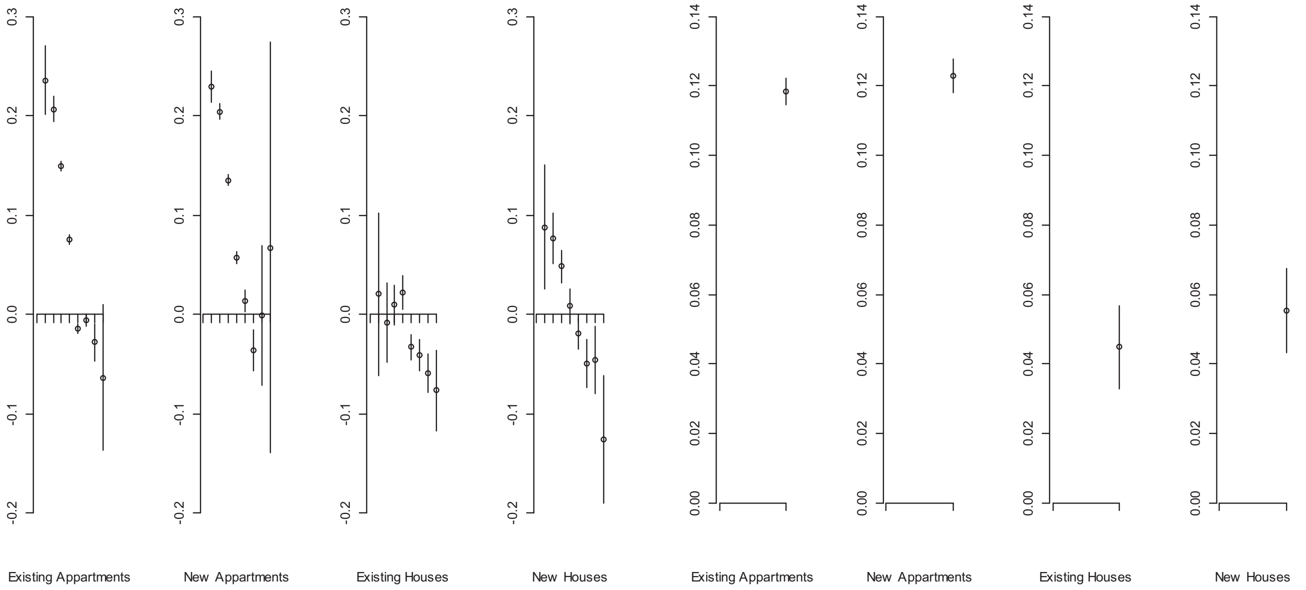

Fig. 2. Point estimates and 95% confidence intervals for the impact of energy efficiency.

the effect of EPC labels below the comparison rate was found by Fuerst et al. (2016), for the Helsinki metropolitan area of Finland. According to these authors, this is justified by the buyer's predisposition to pay more for more energy efficient properties and disregarding performances below the average energy efficiency reference, a feature designated as a green clientele effect.[5]

One possible explanation for the observation of higher rewards being attributed to apartments over houses may rest on the idea that land, which does not affect energy efficiency, represents a larger proportion of the house property value than in the case of apartments. Were we able to separate land prices from total transaction prices, then the differences across apartments and houses might become more comparable. However, as a property is a bundled good comprising both of land and structural assets, the separation of the two components is challenging and generally difficult to achieve with plausible results (for an excellent consideration of this issue, see Diewert et al., 2015). Moreover, this needs to take into account that, as houses are more expensive on average than apartments, the gap between the two property types is not as great as the percentage difference suggests. Finally, because houses include appealing features (for example, the presence of gardens or swimming pools), the importance buyers attribute to energy efficiency may be attenuated when compared with that of apartments.[6]

To conclude this section, the aggregated version of the energy scale that contrasts A-B and for D-G labels is considered. This simple baseline specification is important for coping with (i) the introduction of terms for time and space interactions in Section 4.2, (ii) the scarcity of observations for some labels in extreme price quantiles in Section 4.4, and (iii) the simulation designs of Section 5. In these models, the price effect of receiving an EPC label above or at the reference energy consumption standard (A or B-type) is measured relative to the reference of being below that limit, reflecting the information available for home buyers. While for apartments the results for the 9-level scale reflect this cutoff, for houses the separation is not so clear. In this last situation, A^+, A and B rates do not differ from C for existing properties and B^- rate does not differ from C for new properties. However, some empirical support for the AB/CDFEF split is provided as C differs from the adjacent B^- label in the former case, for which the coefficient is positive and, in the latter case, C is similar not only to B^- but also to D (the transition between positive and negative impacts occurs somewhere in B^-, C, and D). Table 3 and the right-hand side of Fig. 2 present these results. It should be noted that, at a 5% significance level, none of these models is rejected by the RESET specification test and the specification applying a 9-level scale is only rejected for the new houses subpopulation.

Naturally, the coefficient estimates for the most energy efficient properties reflect the previously identified patterns - they are positive and statistically significant. The magnitude of price premiums again differs for apartments and houses, with apartments returning a market premium of around 13% and houses displaying a considerably lower premium of 5% to 6%. Fig. 2 sets out how this difference is statistically significant with the 95% confidence interval for apartments never overlapping that found for houses.

4.2. Analysis across time and space

To study the impact of energy efficiency across time and space, four additional sets of regression models were run in which the interactions between energy efficiency dummy variables and dummy variables signalling either years or regions were added to the hedonic model. The

Table 3
Estimation results when energy rating is defined as a binary scale.

	Existing apartments	New apartments	Existing houses	New houses
Energy efficiency score				
AB	0.118***	0.123***	0.045***	0.055***
	(0.0019)	(0.0024)	(0.0061)	(0.0062)
Root MSE	0.303	0.248	0.413	0.325
R^2-adj	0.676	0.733	0.670	0.753
Reset, P-value	0.567	0.098	0.083	0.052
Sample size	149,920	59,410	33,282	13,533

Notes: Energy classes below the C score are omitted. *** significance at the 0.001 level, ** significance at the 0.01 level, * significance at the 0.05 level. Robust (to the presence of heteroskedasticity) standard errors in brackets. Reset's test hypothesis H0: Correct specification.

---

[5] It should be noted that some results may be influenced by the extremely small number of observations displaying a particular energy label. This is the case for the G rate for new houses with high value (−0.125) based on 189 transactions, representing 0.07% of the complete sample available for regression analysis.

[6] We owe these arguments to two anonymous referees. We would additionally note that Fuerst et al. (2015) also report different price premiums across dwelling types in the British market. In their paper, terraced houses, a type of property with shared walls that is uncommon in Portugal, display the highest premium, while semi-detached, detached (dense and sparse) houses, which are similar to the housing concept prevailing in Portugal, do not display significant price effects for the top energy ratings (AB; C) or, at best, offer lower price premiums than apartments, as in our case.

first two sets of models are estimated for the four market segments and rely on the single dummy approach for AB properties. The second two sets of experiments focus on the existing apartment market using the 9-level scale. This approach was limited to the segment which included the largest number of observations, since it involved the estimation of models containing 71 and 87 parameters, for the year and the region case, respectively. Figs. 3, 4 and 5 depict the resulting estimates and confidence intervals for the impact of EPC labels, with the complete results for the regression coefficients displayed in the supplement (Tables A.4 and A.5) for the single dummy case and available upon request for the longest models. The definition of the regions followed the official nomenclature of seven regional territorial units in effect for Portugal. The only exceptions were the North of Portugal, in which the Porto metropolitan area was isolated and included in the model as a separate region (because of its relevance in transaction terms) and the Azores and Madeira archipelago, which were grouped into a single category (because of the scarcity of observations for these two regions).

Analysis of the energy efficiency impact across time reveals that the differences from the comparison year (2009) are statistically relevant for most years. The exceptions occur in 2010, for existing apartments and new houses, and in 2011 for new houses. These differences are typically positive and, with the exception for new apartments in 2012 and for new houses in 2013, reveal an increasing price premium pattern over the years. In general, despite the existence of differences in the evolution and magnitude of the impacts in the four subpopulation of transactions, the results support the view that the impact of energy efficiency on sales prices increases over time.

For existing apartments, for instance, the price premium placed on A and B rated properties rises from 9.6% in 2009 (coefficient of 0.092) to 17% in 2013 (coefficient of 0.092, plus an interaction term of 0.065). This outcome is consistent with Hyland et al. (2013) where, for the Irish housing market, the effects of the energy EPC rating were reported to be higher when sale conditions were worse. This market characterization broadly corresponds to the period under study, which was strongly marked by a severe contraction in mortgage credits and a worsening sales market. Another possible explanation for the increasing impact of energy efficiency over time rests on the idea that, while the benefits of the EPC label may not have been so evident in the eyes of the market in 2009 (i.e. when energy certificates first became mandatory in dwelling transactions), awareness about their benefits may have gradually consolidated over subsequent years. Additional analysis for the existing apartments segment based on the 9-level scale results of Fig. 4 reveals that properties issued A-B labels display significant price premiums, that increase as the energy label is higher (note that premiums of A$^+$ and A almost doubles those of B and B$^-$). On the other hand, D-G properties display a static behaviour through time with statistically insignificant coefficients, which certainly motivates the low awareness pattern for these labels observed previously in the results of Table 2.

As regards regional analysis (Fig. 3), the results provide only a few statistically relevant differences from the Lisbon metropolitan area (the basis of comparison). This is particularly clear for houses, where only one third of the interactions have statistical relevance. However, the statistically relevant coefficient estimates and the respective 95% confidence intervals, suggest that energy efficiency is clearly rewarded in the North, Porto and Centre regions for apartments and in Porto, Centre and Lisbon for existing houses. For new houses, Fig. 3 reports a heavy energy efficiency premium in the Alentejo region, which registers extreme weather conditions (heat in summer and cold in winter). A detailed analysis for the existing apartments market based on the 9-level scale based on Fig. 5 reveals again that property buyers are insensitive to the lowest energy labels D-G (significant results for F and G for Alentejo and Algarve are due to a very low number of observations in these categories). For the above the reference labels A and B it is apparent that the relevant premiums higher than the Metropolitan area of Lisbon, observed in the single dummy analysis, North, Porto, Centre and Lisbon, are very similar in A$^+$ and B$^-$ types, due to the overlapping of the confidence intervals, get slightly different for B labels, and differ substantially for A labels. In general, greater awareness of the value of the EPC label is apparent in the most urbanized areas of Portugal, in the North, and in a region (Alentejo) with extreme weather conditions.

4.3. Robustness analysis and cross-country comparisons

To test the robustness of the results presented in the previous section, several models were run with different energy efficiency measurements. Table 4 presents the results for two of such measurements. The first applies the nominal annual primary energy needs for heating, cooling and preparing hot water, which is estimated at the property level. This continuous variable is one of the two that define the individual energy rating (the other, the maximum estimated primary energy needs, is not available in the dataset) and measured in kilograms of oil equivalent (kgoe). As it is included in the regressions in its logarithm form, the coefficient estimates associated with this variable can be

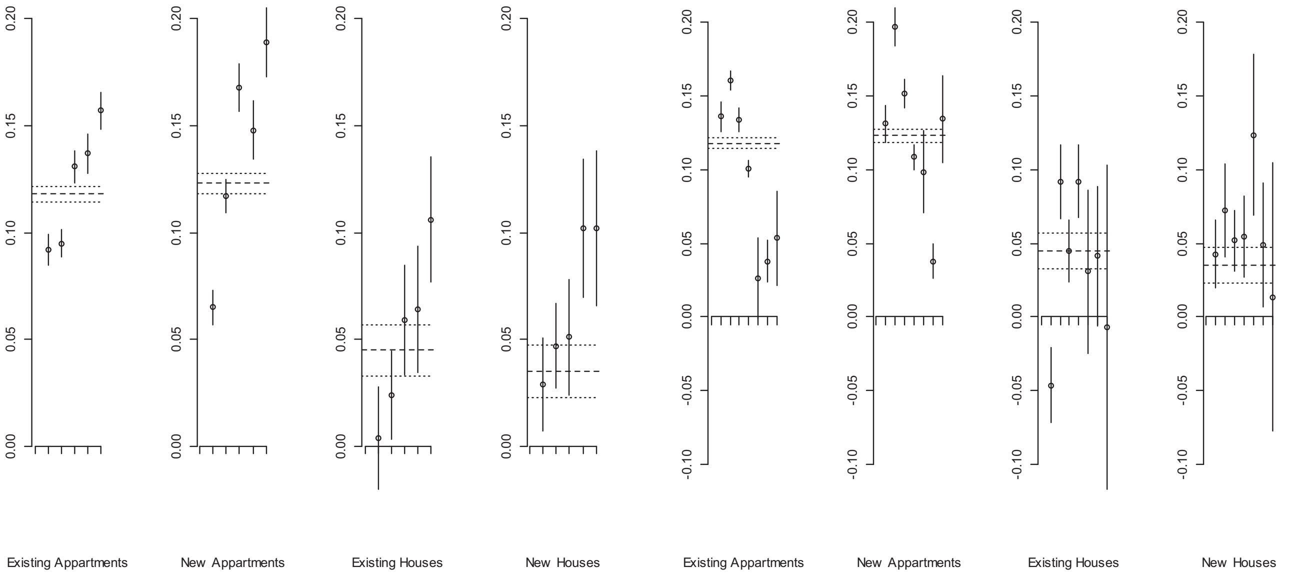

Fig. 3. Point estimates and 95% confidence intervals of the impact of energy efficiency across time and regions. Note: the dashed lines represent the point estimates and the confidence interval for the baseline model.

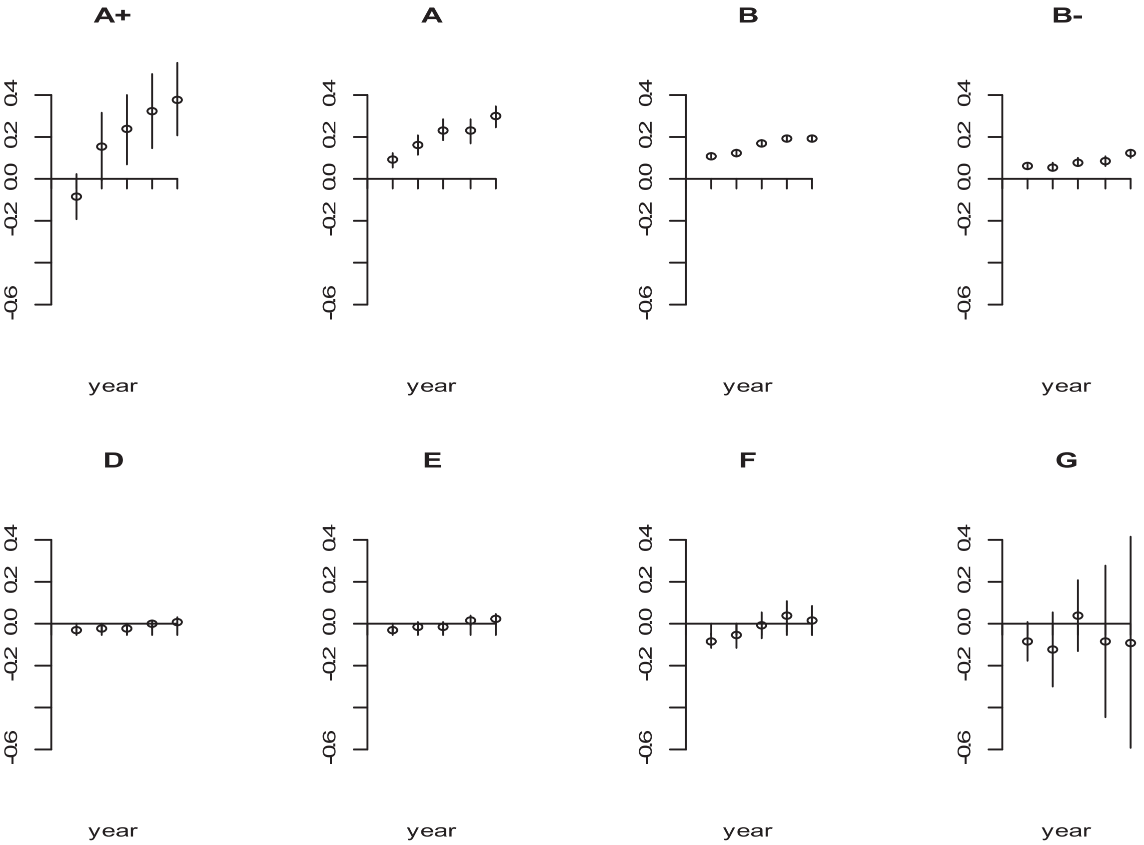

Fig. 4. Point estimates and 95% confidence intervals of the impact of energy efficiency across time for existing apartments. Left to right: years 2009 to 2013.

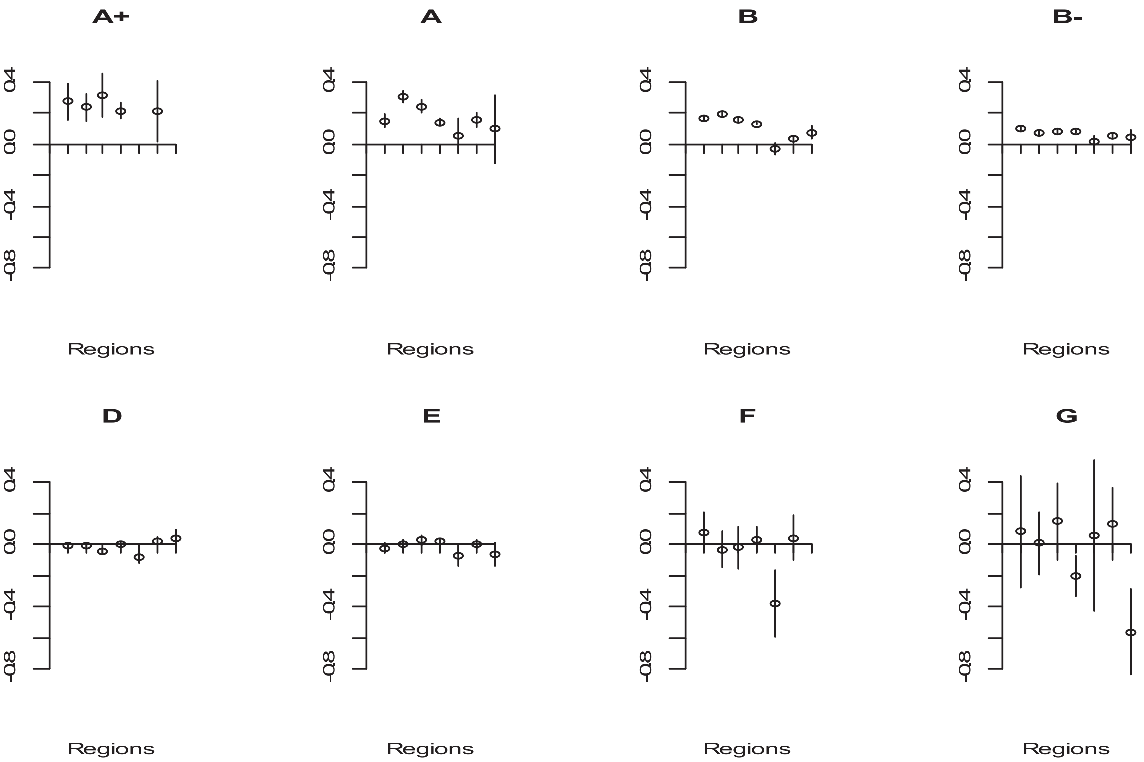

Fig. 5. Point estimates and 95% confidence intervals of the impact of energy efficiency across time for existing apartments. Left to right: North, Porto, Centre, Lisbon, Alentejo, Algarve, Madeira and Azores islands.

Table 4
Estimation results for continuous and alternative discrete energy rating scales.

	Existing apartments		New apartments		Existing houses		New houses	
	LnNTc	CEE	LnNTc	CEE	LnNTc	CEE	LnNTc	CEE
	-0.064^{***}	-0.043^{***}	-0.137^{***}	-0.054^{***}	-0.052^{***}	-0.052^{***}	-0.061^{***}	-0.022^{***}
	(0.0015)	(0.0008)	(0.0024)	(0.0010)	(0.0035)	(0.0035)	(0.0061)	(0.0021)
Root MSE	0.305	0.304	0.247	0.247	0.411	0.413	0.324	0.325
R^2-adj	0.673	0.675	0.739	0.738	0.673	0.670	0.755	0.754
Reset, P-value	0.013	0.0004	0.109	0.031	0.110	0.119	0.095	0.0497
Sample size	149,920	149,920	59,410	59,410	33,278	33,282	13,532	13,533

Notes: *** significance at the 0.001 level, ** significance at the 0.01 level, * significance at the 0.05 level. Robust (to the presence of heteroskedasticity) standard errors in brackets. The LnNTc is the natural logarithm of estimated annual energy primary needs. The CEE variable converted the A+ into 1, the A into 2, the B into 3, …, the F into 8, and finally the G into 9.

interpreted as the price elasticity of an increase in energy efficiency. The second alternative measure of energy efficiency converts the EPC rating scheme into a single variable in which the change from one rate to another always yields a one point difference; see Lyons et al. (2013) for a similar approach of dealing with EPC rating scales. This discrete energy efficiency measure, denoted as CEE, thereby converted A^+ into 1, A into 2, B into 3, …, F into 8, and finally G into 9 in order to enable the quantification of the effect of a one-letter improvement in the energy efficiency scale.

The results demonstrate how energy efficiency, as measured by lnNTc and CEE, is statistically relevant for all property types. The signs of the estimated coefficients are also in accordance with expectations correspondingly reflecting a negative relationship between the sales price and an increase in either the amount of estimated primary energy needs or a change from a higher to a lower energy efficiency rating. Price elasticity is higher for new apartments (−13.7%) and lower for existing houses (−6.4%) while the difference between price elasticities for new and existing properties only holds significance for apartments. Table 4 also sets out how the effect of a one-letter improvement in the EPC scale varies from 5.4% (new apartments) to 1.5% (existing houses).

In order to grasp the impact of energy efficiency on residential property prices in Portugal *vis-à-vis* those generated in other countries, the models were then rerun according to specifications for the energy efficiency scales that approximated those used in studies for Ireland (Hyland et al., 2013), Finland (Fuerst et al., 2016) and the Netherlands (Brounen and Kok, 2011); for a general approach where, using datasets of several European countries at either the country level or for specific cities, energy efficiency is measured by the conciliation of the different EPC scales using a discrete energy measure of the type of CEE, see Lyons et al. (2013).

We would note that, naturally, the three studies in analysis use data with features that differ from our own. In these three cases, the dependent variable is the logged form of list prices, transaction prices, and transaction prices per square meter, respectively. The re-estimation of our hedonic regression models applies transaction prices in the two first cases and transaction prices per square meter for the Brounen and Kok (2011) design; note that list prices are considered an accurate barometer of transaction prices; see Lyons (2018). As regards the hedonic model property characteristics, while Hyland et al. (2013) consider the attributes of the properties (for example, dwelling type and size measured by the number of bedrooms and bathrooms), Fuerst et al. (2016) and Brounen and Kok (2011) considered additional housing characteristics and several neighbourhood variables (for example, the former considers maximum floor space, sauna facilities, mean income per capita and the prevailing unemployment rate while the latter addresses features such as the existence of central heating, the level of external maintenance, housing density, and percentage of green vote). In terms of the properties attributes, our data does not only offer good coverage of most of the previously considered explanatory variables but also has the advantage of including difficult to measure characteristics and that are often absent from hedonic studies. Examples of the latter variables include measures for the availability of public goods and the scenic quality of the location, among others. Table 5 details these results.

Interestingly, the results suggest that energy efficiency is rewarded with higher price premiums in Portugal than in Ireland, Finland and the Netherlands. This emerges particularly clearly for top-rated energy efficiency properties with the A, B and C ratings receiving price premiums generally higher than those reported in the aforementioned three studies. One possible reason for these higher price premiums may be linked to the strong presence of the EPC label in Portugal and a greater awareness of its benefits. In fact, the Buildings Performance Institute Europe (2010) classified Portugal as one of the three countries, out of the twelve analysed, achieving the best performance in the promotion, administration, and compliance/enforcement of EPC labels. Another possible explanation may derive from household energy prices that rank among the highest in the European Union. According to Eurostat (2019), Portugal accounted for the fourth and second most expensive household electricity and gas prices respectively, out of the total of 31 (electricity) and 25 (gas) European countries compared in 2018. For gas, only Sweden levied higher prices and, for electricity, only Germany, Denmark and Belgium.

Table 5 highlights how the overall fit of the regressions aligns with those reported by similar studies. While the regressions using the price level logarithm as the dependent variable are not rejected by the Ramsey (1969) test at either the 5% or 1% significance levels, the specification with the price per square meter logarithm does get rejected. This result reinforces the idea that the choice of the price model instead of the price per square meter model constituted the most suitable means of assessing the impact of energy efficiency on the prices prevailing in the Portuguese residential property sales market. Moreover, the specifications in which the data is pooled together into the same model do not pass the RESET test (see rightmost column of Table 5), reinforcing the idea that models should be specified separately to take into account market specificities. Additionally, the RESET test results indicate that, at the 5% significance level, no alternative EPC label disaggregation leads to the absence of rejections observed in the single AB dummy model presented in Table 3.

4.4. Quantile regression results

According to Koenker and Basset's (1978) seminal work, conditional quantile regression serves to characterize the impact of energy efficiency over the distribution of residential property prices. By allowing buyers of higher-priced properties to respond differently to purchasers of lower-priced properties, the results of quantile regression provide a more comprehensive picture of the true relationship between energy efficiency and residential prices. Although already applied in the real estate context (see, inter alia, Zhang and Yi, 2017) this is, at least to our knowledge, the first time this technique is used to study the impact of the energy efficiency attribute. The quantile regression results (point

Table 5
Robustness and cross-country analysis.

		Exist. apart.	New apart.	Exist. houses	New houses	All dwellings
Hyland et al. (2013)$^{(+)}$						
A	0.093**	0.211***	0.188***	0.030	0.095***	0.182***
	(0.015)	(0.0065)	(0.0060)	(0.0184)	(0.0129)	(0.0037)
B	0.052**	0.127***	0. 095***	0.050***	0.47***	0.104***
	(0.008)	(0.0029)	(0.0054)	(0.0069)	(0.0079)	(0.0023)
C	0.017**	0.015***	−0.014**	0.033***	0.018*	−0.002
	(0.006)	(0.0026)	(0.0050)	(0.0065)	(0.0081)	(0.0022)
E	−0.004	0.009 *	−0.050***	−0.008	−0.031*	−0.018***
	(0.009)	(0.0037)	(0.0112)	(0.0071)	(0.0121)	(0.0031)
F/G	−0.106**	−0.018	−0.004	−0.030***	−0.044**	−0.011
	(0.009)	(0.0095)	(0.035)	(0.0087)	(0.0157)	(0.0059)
R^2	0.65	0.680	0.740	0.670	0.750	0.720
Reset, P-value	–	.800	.219	.095	.043	.000
Sample size	15,060	149,920	59,410	33,282	13,533	256,145
Fuerst et al. (2016)$^{(+)}$						
ABC	0.0130*	0.053***	0.067***	0.040***	0.040***	0.046***
	(0.0062)	(0.0026)	(0.0053)	(0.0055)	(0.0070)	(0.0021)
E	0.000	0.010**	−0.048**	−0.008	−0.031*	−0.016***
	(0.0057)	(0.0037)	(0.0113)	(0.0071)	(0.0121)	(0.0031)
FG	0.0002	−0.019*	−0.005*	−0.030***	−0.046**	−0.012*
	(0.0052)	(0.0095)	(0.0349)	(0.0087)	(0.0156)	(0.0059)
R^2	0.93	0.670	0.72	0.670	0.750	0.710
Reset, P-value	–	.030	0.066	.092	.041	.000
Sample size	6194	149,920	59,410	33,282	13,533	256,145
Brounen and Kok (2011)$^{(-)}$						
ABC	0.037**	0.044*	0.058***	0.046***	0.055***	0.047***
	(0.003)	(0.0021)	(0.0049)	(0.0050)	(0.0066)	(0.0018)
R^2	0.53	0.550	0.610	0.480	0.580	0.570
Reset, P-value	–	.000	.000	.000	.000	.000
Sample size	31,993	149,920	59,410	33,282	13,533	256,145

Notes: (+) The omitted energy class is D. (−) Energy classes D,E,F, and G omitted. The dependent variable of the model is the natural logarithm of price per square meter. *** significance at the 0.001 level, ** significance at the 0.01 level, * significance at the 0.05 level. Robust (to the presence of heteroskedasticity) standard errors in brackets.

estimates and 95% confidence intervals) for the 0.1th, 0.3th, 0.5th, 0.7th, and 0.9th price quantiles are listed in Fig. 6.

These empirical results suggest that energy efficiency is rewarded with a price premium across the entire spectrum of the conditional price distribution. This is an important finding as it could be the case that a price discount, rather than a price premium, could be associated with a particular group in the conditional distribution (e.g. higher-priced properties). Moreover, the impact of energy efficiency seems

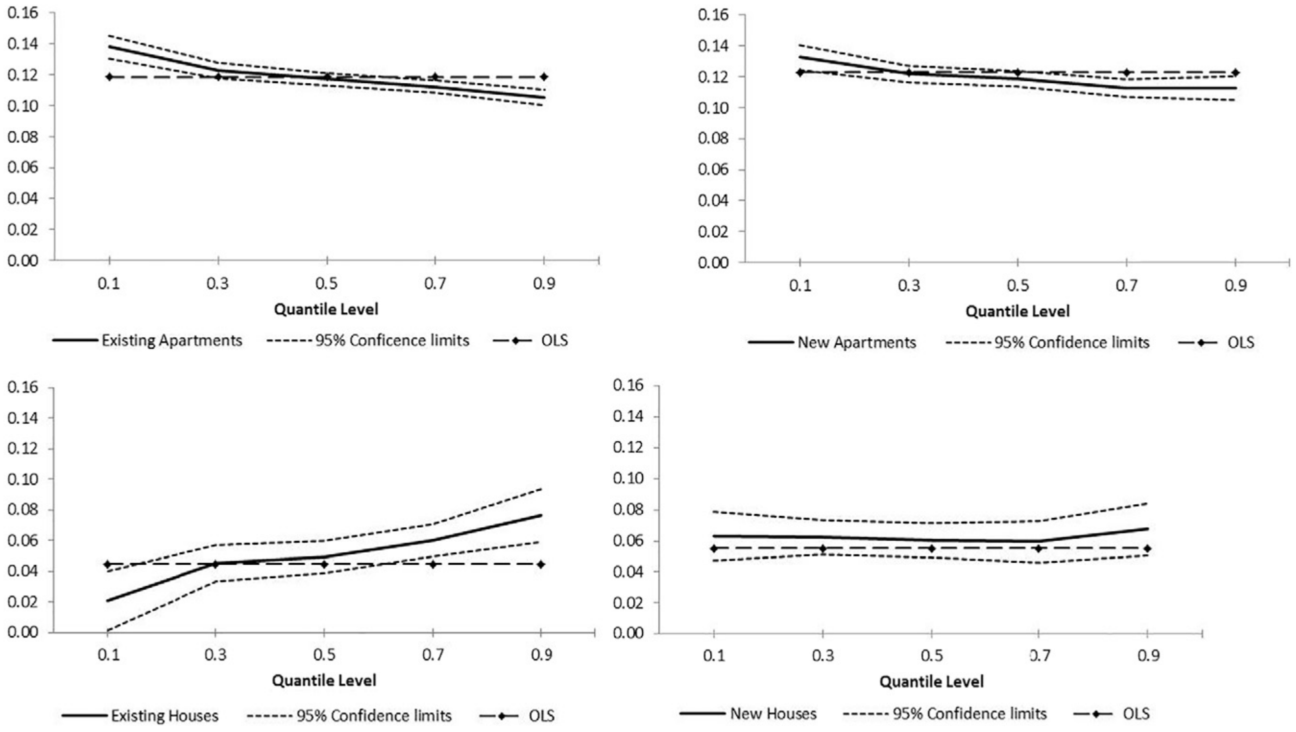

Fig. 6. Parameter estimate by quantile level for energy efficiency.

quite stable across the quantiles. This is particularly evident for new properties and at the median, 0.3th, and 0.7th quantiles of all property types, where the differences among the different quantile energy efficiency partial effect estimates generally do not attain significance. For instance, although energy efficiency price premiums for new apartments drop from 0.122 to 0.112 when moving from the 0.3th to the 0.7th price quantile, their 95% confidence intervals overlap (thus not excluding that they might not differ from one another). Note that, as the price gets smaller, especially for existing properties, premiums display an increasing (decreasing) trend for apartments (houses). This could be essentially driven by demand-side factors, with lower-income house purchasers attributing, when compared with high-income buyers, less importance to energy efficiency than to other dwelling characteristics that are not available in apartments (e.g., size of available plot).

The 0.5th quantile of the energy efficiency partial effect estimates additionally provides an important indication as to the robustness of the OLS findings. As compared to the least squares methods, the estimates at this quantile, which correspond to least absolute deviations (LAD), are less sensitive to the presence of outliers. The results clearly indicate that OLS point estimates are essentially the same as those for the 0.5th quantile. For existing apartments, for instance, the OLS and LAD estimates are 0.118 and 0.117, respectively. The largest difference is obtained for new houses where the OLS estimate is 0.055 and the median regression estimate 0.063. However, even for this case, the difference is not significant with the 95% confidence intervals for the mean and median estimates overlapping each other.

5. Selected data issues

This section addresses some data issues often encountered when measuring energy efficiency by hedonic models. The first subsection considers smaller subsamples of the large scale dataset employed in this paper to check whether the individual statistical significance of the EPC dummy is inflated by the sample size. The remaining two sections assess the sensitivity of energy efficiency partial effects to either the introduction of measurement errors in the dependent and explanatory variables or the omission of relevant information in the hedonic model outputs.

5.1. Large samples

The impact of using different sample sizes on the quality of the regression results was investigated through experiments in which the hedonic regression models were rerun for a number of samples with different sizes. In particular, the energy efficiency coefficients were correspondingly calculated on the basis of 1000 samples with sizes of 500, 1000, 2500, 5000 and 10,000 observations drawn for existing apartments, new apartments, existing houses and new houses.

The averages of energy efficiency parameter estimates over the 1000 replications for each sample size are depicted in Table 6. The standard errors are provided in brackets underneath each average value along with the counts of statistically significant positive coefficients, provided in squared brackets.

With the exception of existing houses for the three smallest sample sizes, there is a substantial number of statistically significant energy

Table 6
Energy efficiency partial effects, different experiments.

	Sub-market			
	Existing apartments	New apartments	Existing houses	New houses
Benchmark estimate	0.118**	0.123**	0.045**	0.055**
	(0.0019)	(0.0024)	(0.0061)	(0.0062)
Parameter results, averages over 1000 replications$^{(+)}$				
n = 500	0.121	0.134	0.047	0.067
	(0.034)	(0.028)	(0.052)	(0.033)
	[947]	[996]	[169] $^{(*)}$	[551]
n = 1000	0.119	0.134	0.049	0.066
	(0.025)	(0.019)	(0.035)	(0.024)
	[996]	[1000]	[272]	[814]
n = 2500	0.120	0.134	0.048	0.065
	(0.015)	(0.012)	(0.023)	(0.015)
	[1000]	[1000]	[572]	[993]
n = 5000	0.120	0.134	0.048	0.065
	(0.011)	(0.009)	(0.016)	(0.010)
	[1000]	[1000]	[849]	[1000]
n = 10,000	0.120	0.135	0.048	0.065
	(0.008)	(0.006)	(0.011)	(0.007)
	[1000]	[1000]	[982]	[1000]
Parameter estimates, measurement error scenario $^{(.)}$				
Fiscal appraisal as dependent variable	0.0386**	0.0485**	−0.0085*	−0.0010
	(0.00108)	(0.00163)	(0.00360)	(0.00418)
No. of bedrooms as size measurement	0.1481**	0.1479**	0.0940**	0.0842**
	(0.00215)	(0.00281)	(0.00702)	(0.00706)
Parameter estimates, omitted variables scenario $^{(.)}$				
Central heating and/or air cond.	0.125**	0.128**	0.051**	0.061**
	(0.00190)	(0.00244)	(0.00608)	(0.00624)
Visual quality	0.118**	0.123**	0.048**	0.056**
	(0.00193)	(0.00245)	(0.00616)	(0.00625)
Location quality	0.131**	0.135**	0.055**	0.066**
	(0.00201)	(0.00255)	(0.0064)	(0.00641)
Construction quality	0.120**	0.127**	0.045**	0.055**
	(0.00192)	(0.00245)	(0.00613)	(0.00624)
All omitted	0.145**	0.153**	0.070**	0.073**
	(0.00201)	(0.00259)	(0.00642)	(0.00645)

Notes: *** significance at the 0.001 level, ** significance at the 0.01 level, * significance at the 0.05 level.
$^{(+)}$ The point estimates refer to the averages over the 1000 simulations; standard deviation provided in parenthesis. The number of statistically significant positive coefficients is shown between square brackets. $^{(*)}$ This experiment yielded 4 statistically significant negative coefficient estimates. $^{(.)}$ Robust standard errors in parenthesis.

efficiency coefficients, even in cases where the sample size is sharply reduced, that display an important increase in the coefficients variability. For example, for new houses, the spread of the estimates based on samples with 10,000 observations (0.007) stands at approximately one fifth of that obtained for samples with 500 observations (0.033). This suggests that the statistical relevance of the energy efficiency results does not get inflated by sample size. The results also demonstrate how the incidence of statistically significant coefficients with conflicting signs is extremely rare, occurring only four times for existing houses and for the smallest sample size ($n = 500$). In addition, this also identifies how the average energy efficiency coefficients remain very stable across the different sample sizes. For instance, the benchmark estimate for existing apartments is 0.118, which is similar to the 0.121 average found for the 1000 rounds of samples with 500 observations, which represent less than 0.4% of the total transactions in this market segment.

Overall, it emerges that the results drawn from smaller sample sizes are essentially the ones obtained from the available large scale dataset. This is a reassuring conclusion, as most of the previous research on the effect of energy efficiency on property prices is based on small datasets.

5.2. Measurement errors

This section investigates the extent to which the energy efficiency partial effects are sensitive to the introduction of either dependent or explanatory variables with measurement errors. To illustrate the former case, the transaction price logarithm is replaced by the fiscal appraisal value logarithm, which is also available to the dataset. The use of fiscal appraisal logarithm enables the study of the effects of including a variable that, although returning a high correlation with transaction prices, is generally set below sales prices (see both Table 1 and the scatter plots in Fig. A.2 provided in the Supplement). In Portugal, fiscal appraisals are derived through the usage of a formula that includes six coefficients that are used as building blocks for calculating the value of any given property. This formula is legally stipulated (article 38 in Decree Law no. 287/2003) and covers important price determinant factors such as the property age, its area, and location. However, energy efficiency is only at best partially incorporated into this formula. This stems from how the dwelling characteristics influencing the respective energy performance are included in only one of its coefficients, which measures the property's quality and comfort. More precisely, out the total of over 20 items covered by this coefficient, two may be correlated with energy efficiency. The first identifies the existence of central systems for the cooling and heating of the property and the second assesses the property's construction quality including, among other factors, its thermal and acoustic insulation. On the other hand, illustrating the impact of applying an erroneously measured explanatory variable on energy efficiency partial effects is provided by re-estimating the hedonic models with the number of bedrooms replacing the area variables. The former variable is a poor measurement of property scale but which researchers may make recourse to when having access to no better measurements of size. The summary statistics for the EPC regression coefficients are displayed in the middle of Table 6 and with the 95% confidence intervals presented on the left-hand side of Fig. 7.

The damaging effects of applying either the fiscal appraisal logarithm as a dependent variable or the number of bedrooms as a measurement for size become evident. In the former case, the energy efficiency coefficients are substantially smaller than those reported by the benchmark scenario. This is particularly evident both for existing houses, where the energy efficiency coefficient exhibits a small price discount (−0.0085) and for new homes, where the parameter is statistically insignificant. The smaller magnitude and the insignificance of the estimates certainly reflects how the fiscal evaluation does not depend directly on energy efficiency parameters while the statistical significance documented in some cases may stem from the dependence of the appraisal on features such as the existence of central cooling and heating systems which are likely to correlate with the property's respective energy efficiency level. To grasp the magnitude of the destructive effects of applying this appraisal price in Portugal, see how far the confidence intervals differ from the reference values in Fig. 7. In general, fiscal appraisals following the formula used in Portugal appear as an unsuitable transaction price replacement variable in the measurement of the effects of energy efficiency. Moreover, as EPC labels appear in this paper as a major determinant of the property price, a possible policy recommendation, is their inclusion in the formula that delivers the market value of a property.

When the area variable is replaced by the number of bedrooms, the distortion in the energy efficient partial effects also becomes clear. The energy efficiency coefficient for existing houses more than doubles that of the reference, changing from 0.0448 to 0.0940, which represents a price premium increase from 4.6% to 9.9%. In the remaining segments, there are lesser distortions in the coefficients but even so larger than 20%, in such a way that the confidence intervals clearly lie above those of the reference case for each of the four properties types.

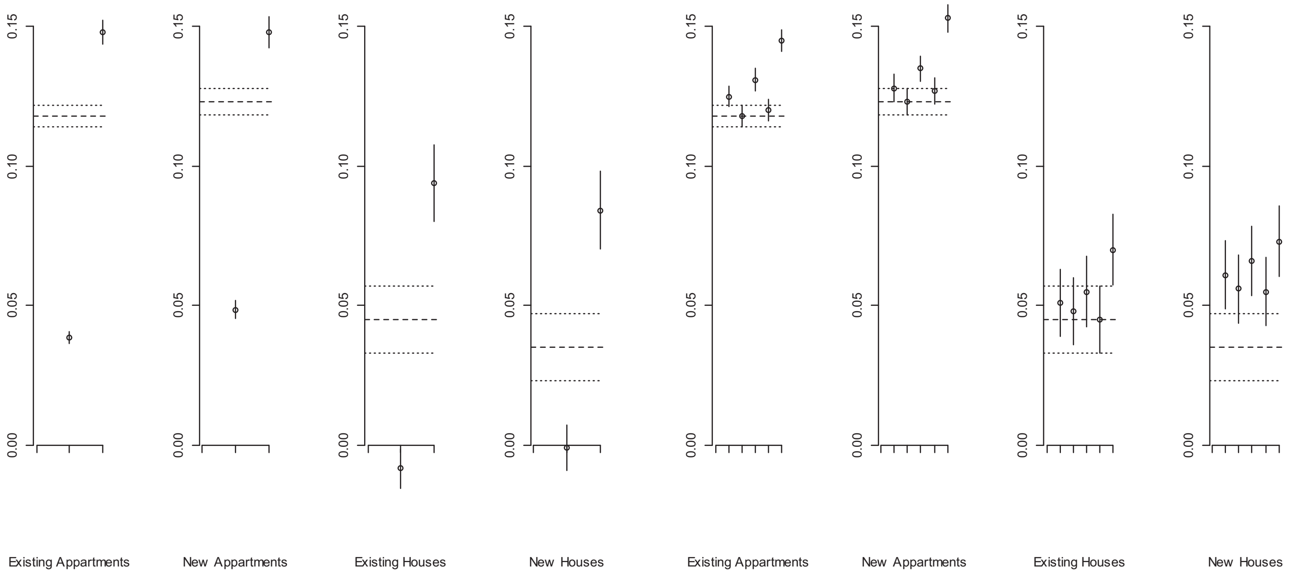

Fig. 7. Point estimates and 95% confidence intervals for measurement error and omitted variables scenario. Note: The dashed lines represent the point estimates and 95% confidence interval for the baseline model.

5.3. Omitted variables

To illustrate the sensitivity of energy efficiency partial effect estimates to the omission of relevant variables, the regression hedonic models were re-estimated without a selected group of dwelling characteristics. The choice of the variables for omission rested on those quality attributes that were a priori deemed to bear a reasonable correlation with energy efficiency (i.e., the existence of central heating and/or air conditioning) and not often available in hedonic regression studies in this research field (i.e., the scenic value of the location, the location and the construction quality of the residential properties sold, all measured at the individual property level). The correlations of these variables with the AB label dummy (provided in the supplement) is clearly higher for the former variable. A total of five omitted variable scenarios was thereby considered with the first four omitting merely a single variable and a fifth calculation omitting all the selected variables for this experiment; see both the bottom of Table 6 and the 95% confidence intervals provided on the right-hand side of Fig. 7.

As expected, the largest deviations from the benchmark situation were obtained for the all-variables omitted scenario. For new apartments, this scenario yields a 0.153 (16.5%) point estimate that compares with the 0.123 (13.1%) coefficient returned by the benchmark model. In terms of price premiums, this represents a substantial 3.4 percentage point difference, the largest obtained by any scenario. The omission of a single dwelling characteristic also produced some differences between the benchmark and the omitted variable results. Such is the case when the two dummy variables measuring the location quality were omitted, which produced an upward shift in the coefficients estimated and a consequent non-overlapping of the respective confidence intervals with that of the reference in all cases except for existing houses. For example, this omission for new houses leads to boosting the coefficient from 0.055 to 0.066, which drives a non-negligible price premium increase from 5.7% to 6.8%. Moreover, the exclusion of the dummy variable controlling for the existence of central heating and/or air conditioning systems, the omitted variable with the highest correlation with the EPC dummy, produced a noticeable upward shift in the energy efficiency coefficient estimates for both existing apartments and new houses (notice the absence of any overlapping between these confidence intervals and those of the benchmark). For example, in the former case, the energy efficiency valuation rose from 0.118 to 0.125, a result implying an energy efficiency price premium rise from 12.5% to 13.3%. In contrast, excluding the variables measuring the visual prominence of the location and the property's construction quality do not greatly impact on the estimated energy efficiency partial effects. Similarly, in global terms, the results for existing houses display a robust behaviour in relation to variable omission when compared with the three other property types. Therefore, these experiments generate mixed evidence in suggesting single variable omission effects may not be severe in the framework of this dataset.

6. Conclusions and discussion

The results provide support to the perspective that the most energy efficient properties are positively rewarded by the Portuguese residential market with the four top A and B energy efficiency scores displaying price premiums in relation to the reference C rate. For other EPC rates, the evidence suggests that the sales market does not differentiate between levels of energy inefficiency as the magnitudes of the price discounts estimated for the D to G rates are often statistically similar across dwelling types and energy levels.

Moreover, the output of hedonic regression models provides evidence of statistically significant differences between the way energy efficiency is rewarded for apartments and houses, with the former dwelling category yielding higher price premiums than the latter. When compared with less efficient properties, A and B rated new and existing apartments receive sales price premiums of 13.1% and 12.5%, respectively. Houses obtain smaller price premiums with new and existing houses receiving 5.7% and 4.6% sales premiums over the same period. This information is important to policy makers. As houses represent the majority of the dwelling stock in Portugal,[7] this result may play a crucial role in defining policy measures designed to increase energy efficiency standards in a cost-effective way.

The analysis of energy efficiency over time also discloses the existence of an upward trend in energy efficiency price premiums. This result is consistent with the findings of Hyland et al. (2013), which suggest that for the case of Ireland the value of certification rose when the prevailing market conditions were worse. As the time period under analysis broadly overlaps with a period when the housing market suffered from illiquidity, uncertainty and credit constraints, buyers may have seen the most efficient energy labels as an additional guarantee of value. In this period, on a year-on-year basis, the number of properties sold decreased 28% in 2011 and prices shrank 7.1% in 2012 (the year when the real estate market hit the bottom); from 2013 onwards, the sales market recovered with prices increasing quite substantially (e.g. in 2017 and 2018, prices rose respectively by 9.3% and 10.5% year-on-year). In general, the increasing annualised price premiums over the 2009–2013 period indicates that energy efficiency is likely to remain positively rewarded in the Portuguese market. Certainly, a future avenue of research involves the production of a similar dataset for more recent years. On the other hand, qualitative comparison across different studies and markets suggests the existence of higher price premiums in the Portuguese market than those reported by studies for the Netherlands, Ireland and Finland. Greater EPC label awareness and the existence of higher energy costs in Portugal provide possible explanations for the existence of these greater price premiums.

Quantile regression analysis further reinforces this idea of a clear market premium existing for top-rated energy efficient residential units. When compared with the rest of the sample, the properties group awarded A or B energy efficiency ratings receives a price premium relevant not only at the mean but also across the entire spectrum of the conditional price distribution. Moreover, the energy efficiency impact was found to be relatively homogeneous over the full extent of the conditional distribution of prices.

Finally, the empirical simulation exercises provide valuable information. Firstly, they demonstrate that, at least in our framework, the use of a large dataset does not artificially inflate the significance of the effects of the EPC rating level on property prices and show that the exploration of significantly smaller datasets essentially lead to the same conclusions. Secondly, it clearly illustrates how the application of error-prone measures, either the transaction price or a control variable for measuring size, consubstantiated by the fiscal appraisal price and a discrete attribute measuring the respective property's number of rooms, both lead to sizable distortions in the magnitude of the EPC partial effects. Conversely, the isolated omission of variables for attributes that are usually unavailable (e.g. indicating the presence of central heating or location quality measures) left unaffected most of the energy efficiency partial effects of interest. Nevertheless, we would note that the existence of a mix of omitted variables, which is likely to occur in practice, did still emerge as a potentially disturbing factor in all the statistical experiments.

In general, this paper builds on a particularly rich dataset for Portugal exploiting both conditional mean and quantile regression-based hedonic models together with some simulation experiments that exploit certain unique features of the dataset, specifically the existence of (i) a measure for transaction price together with an error-prone fiscal appraisal value; (ii) a nine-level EPC scale and one of the underlying continuous measures employed in its construction; (iii) an exceptionally rich set of control variables, including rarely measured

---

[7] The Portuguese housing stock amounts to 5,859,540 classic residential dwellings (INE, 2012). Of these, 52% are single family residential (detached, semi-detached and row) homes (calculations by the authors based on Census data).

attributes, such as visual or location quality and alongside different measurements of housing features, such as the gross floor area in square meters and the number of bedrooms to capture property size. Measuring the effects of energy efficiency on property prices correspondingly emerges as a complex and therefore interesting issue.

CRediT authorship contribution statement

Rui Evangelista: Data curation, Software, Validation, Writing - original draft. **Esmeralda A. Ramalho:** Supervision, Conceptualization, Writing - review & editing. **João Andrade e Silva:** Investigation, Methodology, Writing - review & editing, Funding acquisition.

Acknowledgments

The authors would like to thank the Instituto Nacional de Estatística (INE) and Agência para a Energia (ADENE), who have made possible the access to the data used in this paper at the premises of INE. The authors would also like to thank the two anonymous referees for their insightful comments on this paper. This research was partially supported by the Project CEMAPRE-REM UIDB/05069/2020 - financed by FCT/MCTES, through national funds.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.eneco.2020.104699.

References

Agência para a Energia (ADENE), 2018. Energy Efficiency Trends and Policies in Portugal. July 2018. ADENE, Lisboa.

Atanasiu, B., Constantinescu, T., 2011. A comparative analysis of the energy performance certificates schemes within the European Union: implementing options and policy recommendations. ECEEE Summer Study 2013 Proceedings. European Council for an Energy-Efficient Economy (ECEEE).

Ayala, A., Galarraga, I., Spadaro, J.V., 2016. The price of energy efficiency in the Spanish housing market. Energy Policy 94, 16–24. https://doi.org/10.1016/j.enpol.2016.03.032.

Brounen, D., Kok, N., 2011. On the economics of energy labels in the housing market. J. Environ. Econ. Manag. 62 (2), 166–179. https://doi.org/10.1016/j.jeem.2010.11.006.

Buildings Performance Institute Europe, 2010. Energy performance certificates across Europe: from design to implementation. Retrieved from. http://bpie.eu/publication/energy-performance-certificates-from-design-to-implementation/.

Chin, T., Chau, K., 2003. A critical review of literature on the hedonic price model. International Journal for Housing and its Applications. 27 (2), 145–165.

Colwell, P.F., 1993. Comment: Semiparametric estimates of the marginal price of floorspace. J. Real Estate Financ. Econ. 7 (1), 73–75. https://doi.org/10.1007/BF01096937.

Cropper, M.L., Deck, L.B., McConnell, K.E., 1988. On the choice of functional form for hedonic price functions. Rev. Econ. Stat. 70 (4), 668–675. https://doi.org/10.2307/1935831.

Decreto-Lei (Decree Law) n.° 118/2013 de 20 de agosto. *Diário da República n.° 159*/2013, Série I. Ministério da Economia e do Emprego. Lisboa.

Decreto-Lei (Decree Law) n.° 287/2003 de 12 de novembro. *Diário da República n.° 262*/2003, Série I-A. Ministério das Finanças. Lisboa.

Diewert, W.E., de Haan, J., Hendriks, R., 2015. Hedonic regressions and the decomposition of a house price index into land and structure components. Econ. Rev. 34 (1–2), 106–126. https://doi.org/10.1080/07474938.2014.944791.

Dinan, T.M., Mirarowski, J.A., 1989. Estimating the implicit price of energy efficiency improvements in the residential housing market: a hedonic approach. J. Urban Econ. 25 (1), 52–67. https://doi.org/10.1016/0094-1190(89)90043-0.

Directive 2010/31/EU. *Official Journal of the European Union no L 153/1.* Strasbourg: European Parliament and of the Council.

Directive 2018/844/EU. *Official Journal of the European Union no L 156/75.* Strasbourg: European Parliament and of the Council.

Eurostat, 2019. Energy prices in 2018. [Press Release]. Retrieved from. https://ec.europa.eu/eurostat/documents/2995521/9802432/8-21052019-AP-EN.pdf/7f42181d-d795-4ce1-9dde-ba93fe247166.

Fesselmeyer, E., 2018. The value of green certification in the Singapore housing market. Econ. Lett. 163, 36–39. https://doi.org/10.2139/ssrn.3017774.

Fuerst, F., Warren-Myers, G., 2018. Does voluntary disclosure create a green lemon problem? Energy-efficiency ratings and house prices. Energy Econ. 74, 1–12. https://doi.org/10.1016/j.eneco.2018.04.041.

Fuerst, F., McAllister, P., Nanda, A., Wyatt, P., 2015. Does energy efficiency matter to home-buyers? An investigation of EPC ratings and transaction prices in England. Energy Econ. 48, 145–156. https://doi.org/10.1016/j.eneco.2014.12.012.

Fuerst, F., Oikarinen, E., Harjunen, O., 2016. Green signalling effects in the market for energy-efficient residential buildings. Appl. Energy 180, 560–571. https://doi.org/10.1016/j.apenergy.2016.07.076.

Goodman, A.C., Thibodeau, T.G., 1995. Age-related heteroskedasticity in hedonic house price equations. J. Hous. Res. 6 (1), 25–42.

Halvorsen, R., Palmquist, R., 1980. The interpretation of dummy variables in semilogarithmic equations. Am. Econ. Rev. 70 (3), 474–475.

Hyland, M., Lyons, R.C., Seán, L., 2013. The value of domestic building energy efficiency: evidence from Ireland. Energy Econ. 40, 943–952. https://doi.org/10.1016/j.eneco.2013.07.020.

Instituto Nacional de Estatística (INE), 2012. Censos 2011, XV recenseamento geral da população, V recenseamento geral da habitação, resultados definitivos. Instituto Nacional de Estatística, Lisboa.

Instituto Nacional de Estatística (INE) (2017a). *Documento metodológico, Índice de Preços da Habitação*. Versão 2.0.: Instituto Nacional de Estatística.

Instituto Nacional de Estatística (INE), 2017b. *Índice de Preços das Propriedades Comerciais. Documento Metodológico*. Versão 1.0. Instituto Nacional de Estatística, Lisboa.

Jensen, M.O., Hansen, A.R., Kragh, J., 2016. Market response to the public display of energy performance rating at property sales. Energy Policy 93, 229–235. https://doi.org/10.1016/j.enpol.2016.02.029.

Kennedy, P.E., 1981. Estimation with correctly interpreted dummy variables in semilogarithmic equations. Am. Econ. Rev. 71 (4), 801.

Kiel, K.A., Zabel, J.E., 2008. Location, location, location: the 3L approach to house price determination. J. Hous. Econ. 17 (2), 175–190. https://doi.org/10.1016/j.jhe.2007.12.002.

Koenker, R., Basset, G., 1978. Regression quantiles. Econometrica 46 (1), 33–50.

Lin, M., Lucas, H.C., Shmueli, G., 2013. Too big to fail: large samples and the p-value problem. Inf. Syst. Res. 24 (4), 906–917. https://doi.org/10.1287/isre.2013.0480.

Lourenço, R.F., Rodrigues, P.M.M., 2017. Preços da habitação em Portugal - uma análise pós-crise. Revista de Estudos Económicos 3 (4), 45–62 Banco de Portugal.

Lyons, R.C., 2018. Can list prices accurately capture housing price trends? Insights from extreme market conditions. Financ. Res. Lett. 30, 228–232. https://doi.org/10.1016/j.frl.2018.10.004.

Lyons, L., Cohen, F., Lyons, R., Fedrigo-Fazio, D., 2013. Energy Performance Certificates in Buildings and Their Impact on Transaction Prices and Rents in selected EU Countries. Final Report Prepared for the European Commission (DG Energy).

Malpezzi, S., 2003. Hedonic pricing models: a selective and applied review. In: O'Sullivan, T., Gibb, K. (Eds.), Housing Economics and Public Policy: Essays in Honour of Duncan Maclenn. Blackwell, pp. 67–90.

Megerdichian, A., 2018. Further results on interpreting coefficients in regressions with a logarithmic dependent variable. Journal of Econometric Methods 7 (1), 1–5. https://doi.org/10.1515/jem-2016-0015.

Nelson, J.P., 1982. Highway noise and property values: a survey of recent evidence. Journal of Transport Economics and Policy 16 (2), 117–138.

Portaria (Dispach) n.° 461/2007 de 5 de junho. *Diário da República n.° 108, Série II.* Ministérios da Administração Interna, do Ambiente, do Ordenamento do Território e doo Desenvolvimento Regional, da Economia e da Inovação e das Obras Públicas, Transportes e Comunicações. Lisboa.

Ramalho, E.A., Ramalho, J.J.S., Evangelista, R., 2017. Combining micro and macro data in hedonic price indexes. JISS 26 (2), 317–332. https://doi.org/10.1007/s10260-016-0367-6.

Ramsey, J.B., 1969. Tests for specification errors in classical linear least-squares regression. J. R. Stat. Soc. 31 (2), 350–371.

Rosen, S., 1974. Hedonic prices and implicit markets: product differentiation in pure competition. J. Polit. Econ. 82 (1), 34–55. https://doi.org/10.1086/260169.

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

van Garderen, K.J., Shah, C., 2002. Exact interpretation of dummy variables in semilogarithmic equations. Econometrics Journal 5 (1), 149–159. https://doi.org/10.1111/1368-423X.00078.

Wooldridge, J.M., 1991. On the application of robust, regression-based diagnostics to models of conditional means and conditional variances. J. Econ. 47 (1), 5–46. https://doi.org/10.1016/0304-4076(91)90076-P.

Wooldridge, J.M., 2013. Introductory Econometrics: A Modern Approach. 5th ed. South-Western Cengage Learning, Mason, Ohio.

Yoshida, J., Sugiura, A., 2015. The effects of multiple green factors on condominium prices. J. Real Estate Financ. Econ. 50 (3), 412–437. https://doi.org/10.1007/s11146-014-9462-3.

Zhang, L., Yi, Y., 2017. Quantile house price indices in Beijing. Reg. Sci. Urban Econ. 63, 85–96. https://doi.org/10.1016/j.regsciurbeco.2017.01.002.

Zheng, S., Jing, W., Kahn, M.E., Deng, Y., 2012. The nascent market for “green” real estate in Beijing. Eur. Econ. Rev. 56 (5), 974–984. https://doi.org/10.1016/j.euroecorev.2012.02.012.

Ziliak, S.T., McCloskey, D.N., 2004. Size matters: the standard error of regressions in the American economic review. Journal of Socio-Economics 33 (5), 527–546. https://doi.org/10.1016/j.socec.2004.09.024.