THE MARGINAL PRICE OF HOUSING ENERGY-EFFICIENCY IN METROPOLITAN BARCELONA: ISSUES OF SAMPLE SELECTION BIASES

EL PRECIO MARGINAL DE LA EFICIENCIA ENERGÉTICA DE LA VIVIENDA EN LA BARCELONA METROPOLITANA: PROBLEMAS DE SESGO DE SELECCIÓN DE MUESTRAS

CHEN, Ai
Department of Architectural Technology (TA), Centre of Land Policy and Valuations (CPSV)
Technical University of Catalonia (UPC)
PhD student
Av. Diagonal, 649, 4th floor, Barcelona, C.P. 08028, Spain
E-mail: ai.chen@upc.edu
Telephone: +34 934054385

MARMOLEJO-DUARTE, Carlos
Department of Technology of Architecture (TA), Center for Land Policy and Valuations (CPSV)
Polytechnic University of Catalonia (UPC)
Associated Professor
Av. Diagonal, 649, 4th floor, Barcelona, C.P. 08028, Spain
E-mail: carlos.marmolejo@upc.edu
Telephone: +34 934054385

Key words: sample selection biases; energy-efficiency price; EPCs; Metropolitan Barcelona

Palabras clave: sesgos de selección de muestra; precio de la eficiencia energética; EPCs; Barcelona Metropolitana

Abstract

Purpose

Housing energy-efficiency has become a hot issue in the residential sector along with the mandatory requirement by EPBD to exhibit an energy performance certificates (EPC) when transacting real estate. Numerous studies have focused on energy-efficient marginal price using hedonic price models. Nevertheless, in some markets such as the Spanish one a vital proportion of properties to be let or sold do not exhibit the EPCs in the real estate advertisement. By not considering this issue the impact of EPCs on housing prices may result biased. In other words, those cases without EPC labels that are not considered, when analyzing impacts of energy label on housing prices, do actually matter to them. This ignorance of sample selection bias may reduce the accuracy of results, or even give an adverse estimation. In this case, we aim to explore the presence of sample selection bias and correcting these biases for better the following studies.

Methodology

A collected selling listing prices from Habitaclia, one of the leading web-based real estate listings in Catalonia is the main source of information and Heckman model is used to identify the likelihood of selection bias in metropolitan Barcelona by the two-step method, including a Selection model and a Hedonic Price model. After tested robustness and quantized the bias

Citación: CHEN, A. & MARMOLEJO-DUARTE, C. *The marginal price of housing energy-efficiency in Metropolitan Barcelona: issues of sample selection biases.* En: Libro de proceedings, CTV 2018. XII Congreso Internacional Ciudad y Territorio Virtual. "Ciudades y Territorios Inteligentes". UNCuyo, Mendoza, 5-7 septiembre 2018. Barcelona: CPSV, 2018, p. 247-262.

from those non-EPC-labeled properties, an energy-efficient premium will be revised and compared with the traditional OLS estimate results.

Conclusion

The estimation results suggest that the sample selection indeed exist and does matter to energy-efficient premium in Barcelona Metropolitan. This premium increases from 9.6% to 12.6% when houses improve energy ranking from G to A, or from 0.9% to 2% with every ranking increasing after correcting those sample selection bias. At the same time, we found that the effect of sample selection bias is stronger where properties are higher with medium-high floor area size.

Resumen

Propósito

La eficiencia energética de la vivienda se ha convertido en un problema desde la obligatoriedad impuesta por la EPBD de exhibir un certificado de eficiencia energético (EPC) al realizar transacciones de bienes raíces. Numerosos estudios se han centrado en precios marginales de la eficiencia energética utilizando modelos de precios hedónicos. Sin embargo, en algunos mercados como el español, una importante proporción de propiedades en alquiler o venta no exhiben los EPC en el anuncio inmobiliario. Al no considerar este tema, el impacto de las EPC en los precios de la vivienda puede resultar sesgado. En otras palabras, estos casos sin etiquetas EPC no pueden ser considerados. Este desconocimiento del sesgo de la selección de la muestra puede reducir la precisión de los resultados, o incluso dar una estimación adversa. En este caso, el objetivo de este trabajo es explorar la presencia del sesgo en la selección de la muestra y corregirlo, para mejorar los siguientes estudios.

Metodología

Se utilizan, como principal fuente de información, los listados de propiedades inmobiliarias de *Habitaclia*, empresa líder en Internet en Cataluña y se aplica el modelo de Heckman para identificar la probabilidad de sesgo de la selección en la Barcelona Metropolitana, mediante el método de dos pasos, que incluye el modelo de Selección y el modelo de Precios Hedónicos. Después de probar la robustez y cuantificar el sesgo de las propiedades no etiquetadas con EPC, se revisará eficiencia energética Premium y se comparará con los resultados de la estimación tradicional de OLS.

Conclusión

En la Barcelona Metropolitana, los resultados de la estimación sugieren que la selección de la muestra es efectiva y que si importa la eficiencia energética Premium. Esta prima aumenta del 9.6% al 12.6% en el caso de viviendas que mejoran su ranquing energético desde G a A, o desde 0,9% a 2%, con cada aumento de clasificación, después de corregir el sesgo de selección de muestra. Al mismo tiempo, encontramos el efecto que el efecto del sesgo de selección de la muestra es más fuerte en las propiedades son más grandes con una superficie media-alta.

Citación: CHEN, A. & MARMOLEJO-DUARTE, C. *The marginal price of housing energy-efficiency in Metropolitan Barcelona: issues of sample selection biases.* En: Libro de proceedings, CTV 2018. XII Congreso Internacional Ciudad y Territorio Virtual. “Ciudades y Territorios Inteligentes”. UNCuyo, Mendoza, 5-7 septiembre 2018. Barcelona: CPSV, 2018, p. 247-262.

1. Introduction

Energy efficiency in the housing sector has become a hot issue along with Energy Performance Certificates (EPCs) introduced by Energy Performance of Buildings Directives (EPBD) in 2002. Numerous studies have concluded the EPC impacts on housing prices by hedonic models. Brounen and Kok (2011) indicated that there is an energy-efficiency premium 3.6% with energy ranking increase in Netherland. Fuest et al. (2015) found in England and Wales, an 11.8% housing premium increases when dwellings improved from ranking G to ranking A. Likewise, Hyland et al. (2013) found the same trend of the increase premium is higher when selling in Ireland. Bottero and Bravi (2014) indicated the detailed 26.44 euros per square meter increase with energy ranking in Turin. De Ayala et al. (2016) suggested in Spanish 5 cities, there is housing prices premium after making a survey to ask for the opinion value from households. Marmolejo (2016) concluded there is a 0.85% increase on housing prices in Metropolitan Barcelona Area while in 2019, the premium increase to 1.4% with energy ranking (Marmolejo and Chen, 2019). However, there are still studies out of conspicuous premium or total inverse penalty on housing prices (Bio intelligence et al. 2013; Fregonara et al. 2017).

Regarding sample selection biases, a number of studies has indicated that selection bias do matter to housing prices and residential analysis (Jud and Seaks, 1994; Gatzlaff and Haurin, 1998; Hill, 2011; Hedman and Van-Ham, 2012). They proposed that a necessary selection biased correction should implement before any hedonic price models and calculations. They indicated the missing test for sample selection biases might have an inverse impact on estimation results or the conclusion. For this reason, Heckman two-step method was put forward by Heckman (1976) and developed by following relative studies (Heckman, 1977, 1986, 1990; Puhani, 2000). They suggested that the biases can be estimated by a procedure where a proxy variable could be produced and the Heckman two-step model is the best choice to solve the selection biases. Gordon and Winkler (2016) applied a corrected-biased model to explore the impacts of the price percentage discount in housing prices in North Alabama. They found a discount impact 2.98% was made after correcting sample selection biases. The same conclusions were suggested using the Heckman two-step model by Seko and Sumita (2007) and Gracias and Enriques (2008). They indicated that the impact of the tenure choice is negative when properties were transacted. However, just a few studies show attention to the sample selection biases when analyzing the relationship between EPC and housing prices. Brounen and Kok (2011) found that homes with a “green” label sell at a premium of 3.6% relative to otherwise comparable dwellings with non-green labels using Heckman two-step method. In such case, this paper is to explore the presence of selection biases and to correct these biases by the Heckman two-step model, as an initial analysis of hedonic housing prices.

The remainder of this paper is organized as follows: first, a general introduction to the methodology and models in detail; next, a description of the scope of the study and data statistics; followed by the results and discussion; and finally as a conclusion and acknowledge.

2. Methodology

After having delimited the case study, the method has consisted in 4 steps:

1) First, a sample depuration procedure will be made by eliminating cases which prices was +/- standard deviation above or below average price and using Mahalanobis distance.

Citación: CHEN, A. & MARMOLEJO-DUARTE, C. *The marginal price of housing energy-efficiency in Metropolitan Barcelona: issues of sample selection biases.* En: Libro de proceedings, CTV 2018. XII Congreso Internacional Ciudad y Territorio Virtual. “Ciudades y Territorios Inteligentes”. UNCuyo, Mendoza, 5-7 septiembre 2018. Barcelona: CPSV, 2018, p. 247-262.

2) Second, a Probit model will be elaborated which can be regarded as the selection equation model of Heckman two-step model. In this model, dependent variable is a binary one where the energy-labeled dwellings is equal to 1 and otherwise is 0. Subsequently, a new variable - “Inverse Mills Ratio” (IMR) will be produced which represents the existence of sample selection biases if the P-value of IMR is less than 0.05 (confidence level =95%).

3) Third, a four-equation OLS hedonic price model will be built into 2 groups where the difference is the expressive forms of energy label in dwellings. Noted the IMR variable will be applied in these two groups to correct impacts of sample selection biases.

4) Finally, estimation results from the former four equations will be analyzed to identify the corrected impacts of sample selection biases, and a coefficient-estimated distribution of energy label and related variables also will be made as maps by ArcGIS.

2.1 Heckman two-step Model

Often, dwellings without energy-labels, according to previous literature, fail to estimate in the study to explore impacts of energy label on housing prices. However, such dwellings have influence on the local housing prices and housing prices of energy-label equipped dwellings, in turn, will be affected by the condition of local real estate markets. That is to say, those cases we used are non-random ones and this ignorance may lead to bias in our estimation.

In order to identify and eliminate this bias, an econometric model called Heckman two-step model was made by Heckman (1976). He pointed that the maximum likelihood estimation of a nonlinear model (e.g. Probit model) produced consistence, asymptotically normal estimator and the usual standard error and test statistics are valid if the selection is entirely a function of the exogenous variables. Heckman two-step model is made of 2 equations:

2.1.1 Selection equation - Probit model

Using all n cases, estimate a probit model of a series related buildings and economic characteristics and factors on the presence of energy label for a dwelling. Then IMR is produced to identify the existence of sample selection biases.

$$Dum\_EPC_i = \beta_i + \sum_{s=1}^{n} \beta_{is} SD_{is} + \sum_{k=1}^{n} \beta_{ik} SB_{ik} + \sum_{m=1}^{n} \beta_{im} A_{im} + \sum_{f=1}^{n} \beta_{if} E_{if} + \sum_{a=1}^{n} \beta_{a} S_{ia} + \varepsilon_i$$

In equation (I), the existence of EPC of an apartment i depends on a set of variables related to SD structural attributes of dwellings; SB structural attributes of buildings; A accessibility indicators; E environmental quality indicator; S socioeconomic hierarchy indicator while e is a vector representing the random error.

In the SD and SB dimension, there are covariates and factors related to physical structural features (such as dwelling’s and building’s quality) and facilities (such as lift, heating as well as air conditioner). It is worth saying, heating and air conditioner as well as the presence of reform of dwellings is correlated to energy efficiency, since in Spanish regulation and law of energy efficiency in buildings EPC is made of some items related to such facilities. This dimension also includes the EPC ranks that are mandatory to be noted in the advertisement of properties as it has been sold.

Citación: CHEN, A. & MARMOLEJO-DUARTE, C. *The marginal price of housing energy-efficiency in Metropolitan Barcelona: issues of sample selection biases.* En: Libro de proceedings, CTV 2018. XII Congreso Internacional Ciudad y Territorio Virtual. “Ciudades y Territorios Inteligentes”. UNCuyo, Mendoza, 5-7 septiembre 2018. Barcelona: CPSV, 2018, p. 247-262.

The A dimension includes accessibility indicators, such as centrality index, average time to work. It is worth saying that centrality index is an integrated variable which includes information of time-density, density of activities, distance travelled by people making activities in a given zone by using DP2 methodology (Pena, 1977; Zarzosa, 2009).

The E dimension includes perception of the presence of green areas and percentage of different functional facilities (e.g. health facility, social services, cultural premises). It is supposed that higher proportion of such facilities proportion in a city or in local districts will contribute to a higher housing price premium due providing to a satisfactory living environment.
In the S dimension, education and income level and are key factors. It includes the percentage of residents holding a university degree living around each of the analyzed apartments. In order to depict a wider picture of the socioeconomic structure of the city a Principal Component Analysis (PCA) has been computed departing from the professional categories (e.g. managers, clerks, blue-collar workers, etc.) of employed people living around each of the apartments. The resulting PC represents proxies for high and low-income population. Socioeconomic indicators are relevant for price formation and EPC rank market premium since income and education are correlated with purchasing power, social prestige and environmental concerns (Banfi et al., 2008; Himmelberg et al., 2005).

Noted that in this model, a new variable, IMR, is produced by the model calculation. It is the ratio of the probability density of fucntion over the cumuative distribution function of a distribution. This is usually applied to explore the presence of sample selection bias. The coefficient of inverse mills ration in probit model can explain the presence of selection bias if the P value is less than 0.05 (based on confidence level 95%)

2.1.2 OLS hedonic price equation

Hedonic price model is made by Rosen (1974). This method assumes that the price paid for the asset from housing buyers is equal to the total utility they extract from it, being this a composite utility coming from the marginal attribute of the dwelling (e.g. area, quality, location, etc.) It is possible to calculate such marginal utility expressed in monetary terms by a regression model.
In the literature little advice can be found on the functional form that hedonic modes shall adopt (Can, 1992; Sheppard, 1999; Malpezzi, 2003; Epple et al. 2014).

Nonetheless, the semi-log function has been intensively used in the context of real estate price analysis. Marmolejo and Gonzalez (2009) summarized advantages of semi-log function:

i) It helps to normalize the price and residual distributions which is fundamental for OLS regression analysis;

ii) Coefficients can be read as semi-elasticity (i.e. coefficients express marginal price variation in percent terms for each unit of change), making it possible to directly compare the importance of the attributes with the results of other studies.

Four models are established by using the samples equipped with EPC label information as following:

Citación: CHEN, A. & MARMOLEJO-DUARTE, C. *The marginal price of housing energy-efficiency in Metropolitan Barcelona: issues of sample selection biases.* En: Libro de proceedings, CTV 2018. XII Congreso Internacional Ciudad y Territorio Virtual. “Ciudades y Territorios Inteligentes”. UNCuyo, Mendoza, 5-7 septiembre 2018. Barcelona: CPSV, 2018, p. 247-262.

$$MOD1: ln(\mathrm{P})_1 = \beta_{i1} + \sum_{s=1}^{n} \beta_{is} SD_{is} + \sum_{k=1}^{n} \beta_{ik} SB_{ik} + \sum_{m=1}^{n} \beta_{im} A_{im} + \sum_{f=1}^{n} \beta_{if} E_{if} + \sum_{a=1}^{n} \beta_{a} S_{ia} + \beta_{n1} EPC_{in} + \varepsilon_i$$

$$MOD2: ln(\mathrm{P})_2 = \beta_{i2} + \sum_{s=1}^{n} \beta_{is} SD_{is} + \sum_{k=1}^{n} \beta_{ik} SB_{ik} + \sum_{m=1}^{n} \beta_{im} A_{im} + \sum_{f=1}^{n} \beta_{if} E_{if} + \sum_{a=1}^{n} \beta_{a} S_{ia} + \beta_{n2} EPC_{in} + IMR + \varepsilon_i$$

$$MOD3: ln(\mathrm{P})_3 = \beta_{i3} + \sum_{s=1}^{n} \beta_{is} SD_{is} + \sum_{k=1}^{n} \beta_{ik} SB_{ik} + \sum_{m=1}^{n} \beta_{im} A_{im} + \sum_{f=1}^{n} \beta_{if} E_{if} + \sum_{a=1}^{n} \beta_{a} S_{ia} + \beta_{n3} EPC_{io} + \varepsilon_i$$

$$MOD4: ln(\mathrm{P})_4 = \beta_{i4} + \sum_{s=1}^{n} \beta_{is} SD_{is} + \sum_{k=1}^{n} \beta_{ik} SB_{ik} + \sum_{m=1}^{n} \beta_{im} A_{im} + \sum_{f=1}^{n} \beta_{if} E_{if} + \sum_{a=1}^{n} \beta_{a} S_{ia} + \beta_{n4} EPC_{io} + IMR + \varepsilon_i$$

Where:

EPC_{in} indicates the nominal EPC level in an apartment i (seven variables assigned 1 if it is in existence)

EPC_{io} indicates the ordinal EPC level in an apartment i (variable assigned as A=7, B=6, C=5, D=4, E=3, F=2, G=1)

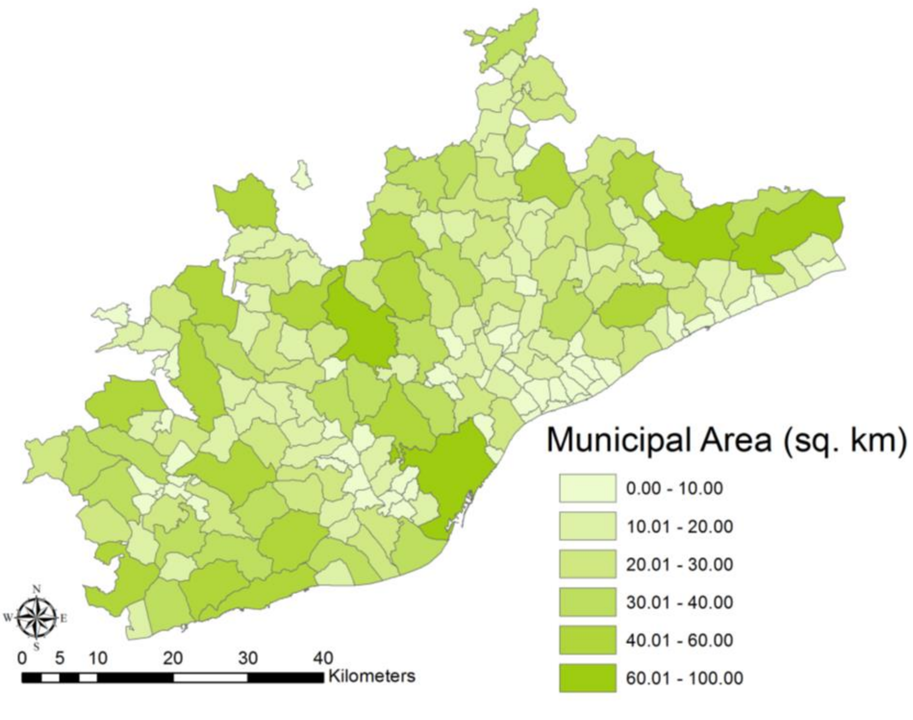
IMR means the Inverse Mills Ratio, the corrected variable of selection biases where it is come from the previous probit model.

2.2 Case study and data

2.2.1 Case study

Metropolitan Barcelona Area (MBA) is selected as case study. In order to identify the limits of this agglomeration the travel-to-work method based on interaction value of Roca et al. (2009) has been used, such approach allows also to detect centralities, which in turns is relevant for this study since accessibility to centers and sub-centres might influence residential prices. As a result, a selected-functional AMB is formed by 189 municipalities in 3,810 sq. km. comprising a population of 5.22 million people.

Figure 1. **Delimitation of Barcelona Metropolitan Area**



Source: own elaboration

Citación: CHEN, A. & MARMOLEJO-DUARTE, C. *The marginal price of housing energy-efficiency in Metropolitan Barcelona: issues of sample selection biases.* En: Libro de proceedings, CTV 2018. XII Congreso Internacional Ciudad y Territorio Virtual. "Ciudades y Territorios Inteligentes". UNCuyo, Mendoza, 5-7 septiembre 2018. Barcelona: CPSV, 2018, p. 247-262.

2.2.2 Data sources

Selling listing prices for apartments coming from Habitaclia is the main source of information. Habitaclia is one of the leading web-based real estate listings in Catalonia. The original dataset comprises 35,116 flats and includes architectonic structural attributes as well as geo-locations. Data refers to November 2014, it is to say, almost 1 years after the RD 235/2013 has made it mandatory to include EPC label information in real estate advertising. Nonetheless such obligation in the sample only 15% of the offers do include energy information. It is worth saying, that autonomous community Catalonia is one of the regions in Spain with the higher proportion of certified houses.

In order to control all the location attributes that might influence apartments' listing price (i.e. environmental quality, accessibility and socioeconomic structure of neighborhoods) a comprehensive GIS has been built departing from the following complementary sources of information:

- *Dwelling and population census INE (2001):* It includes socioeconomic information of resident population as well as perception of noise annoyance at census track level and employment information and journey to work flows at municipal level. Data from the last 2011 census has been discarded since it is based in a survey that is not representative in statistical terms at census track level.
- *Metropolitan Transport of Barcelona (2005):* Street cartography has been used to identify the main transport axis as well as train and metro stations that have been conveniently digitalized. Departing from such information, the precise distance among census tracts has been calculated using TransCAD.
- *Cadastral database (2013):* The information of built-up density and area allocated from a selection of land use has been retrieved at census tract level.

2.2.3 Data description

All the contextual information has been incorporated into each of the analyzed flats using a spatial query departing from a buffer of 300 meters of radius around each dwelling. In order to eliminate extreme cases a twofold approach has been used: 1) first all the cases with price values located beyond +/- Std. Dev from the average valued have been removed, 2) second, the remaining cases have been depurated using the Mahalanobis Distance.

This latter procedure allows to remove the cases whose price is not explained by the covariates but rather by other unmeasured aspects, such as landscaping or specific insulation against noise pollution (Li, 2005). After filtering invalid cases, an effective sample with 4,248 labeled dwellings has been made.

Table 1 shows the statistical description of attributes for the 4,248 cases database. According to such data the average selling price for apartments is 211,396 Euro (implying a unitary price of 2,197 Euro/sq. m.), the area of an average aparment is 89 sq. m, and has 1.36 bathrooms. Regarding the facilities of condominium, 6% of apartments are equipped with swimming pool and 48% have lift; 33% of the listed apartments have air conditioners and 46% heating systems. The area of terraces and balconies in very dense and hot Mediterranean cities is pretty well appreciated by housing demand.

Regarding EPC rank the average class is 2.72, where the most efficient class in Spain is A=7 and the worst is G=1, only 15.77% of the sample is ranked as class A, B or C. All in all, it depicts a housing stock where thermal energy efficiency has a large room for improvement.

Citación: CHEN, A. & MARMOLEJO-DUARTE, C. *The marginal price of housing energy-efficiency in Metropolitan Barcelona: issues of sample selection biases.* En: Libro de proceedings, CTV 2018. XII Congreso Internacional Ciudad y Territorio Virtual. "Ciudades y Territorios Inteligentes". UNCuyo, Mendoza, 5-7 septiembre 2018. Barcelona: CPSV, 2018, p. 247-262.

ISBN: 978-84-8157-661-0

Table 1. **Descriptive statistics for depurated sample**

Dimensions	Variables	N	Minimum	Maximum	Mean	Std. Deviation
Structual Characteristics of Dwelling	Price (Euro)	4,248	22,800	8,000,000	211,396	251,925
	Unit price (Euro/sq.m)	4,248	304	15,385	2,197	1,352
	Area (sq.m)	4,248	25	600	89	39
	Number of bathrooms	4,248	0	6	1.36	0.60
	Number of rooms	4,248	0	15	2.95	0.96
	Ratio bathrooms/rooms	4,248	0	3	0.49	0.23
	Energy Rating (ordinal)	4,248	1	7	2.72	1.29
	Level of the apartment in the building	4,248	0	18	2.26	1.83
	Balcony or terrace areas (sq.m)	4,248	0	256	10.77	16.67
	Living room area (sq.m)	4,248	0	100	12.61	11.13
	Air conditioner (dummy)	4,248	0	1	0.33	0.47
	Heating (dummy)	4,248	0	1	0.46	0.50
	Quality/retrofit (dummy)	4,248	0	1	0.11	0.31
	Penthouse (dummy)	4,248	0	1	0.04	0.20
	Duplex/triplex (dummy)	4,248	0	1	0.05	0.22
Structual Characteristics of Building	Communal swimming pool (dummy)	4,248	0	1	0.06	0.24
	Communal garden (dummy)	4,248	0	1	0.10	0.30
	Elevator (dummy)	4,248	0	1	0.48	0.50
Accessibility Indicators	Built density (area floor ratio)	4,248	0.19	5.90	2.08	1.37
	Time-density	4,248	324	1,154,882	136,251	171,947
	Centrality index	4,248	2.52	20.53	11.59	2.54
	Land use diversity (of the context)	4,248	0.35	1.64	1.04	0.22
	Diversity of activities (of the context)	4,248	0.00	1.92	1.32	0.27
	Average time to work (minutes)	4,248	7.95	37.01	23.31	4.48
	Land use diversity at street level	4,248	0.00	90.10	12.93	14.16
Environmental Quality indicators	Average age of buildings (of the context)	4,248	21.17	124.35	55.65	16.29
	Perception of the presence of green areas	4,248	12.45	97.89	64.00	14.00
	% Health facilities (of the context)	4,248	0.00	41.88	2.08	2.96
	% Educational premises (of the context)	4,248	0.00	93.00	2.17	3.08
	% Social services premises (of the context)	4,248	0.00	68.47	1.84	4.30
	% Cultural premises (of the context)	4,248	0.00	95.15	1.64	3.87
	% Premises for trade (of the context)	4,248	0.00	89.93	40.75	13.55
	% Premises for offices (of the context)	4,248	0.00	100.00	16.52	14.12
	% Industrial premises (of the context)	4,248	0.00	97.01	8.88	11.26
Indicators of Social Hierarchy	% People holding university degree (of the context)	4,248	2.34	68.73	21.78	14.38
	% buildings with porter services (of the context)	4,248	0.00	84.67	8.34	10.59
	CP low socioeconomic level	4,248	-1.97	7.42	0.03	0.96
	CP high socioeconomic level	4,248	-3.26	7.16	-0.21	0.85

Source: own elaboration

Citación: CHEN, A. & MARMOLEJO-DUARTE, C. *The marginal price of housing energy-efficiency in Metropolitan Barcelona: issues of sample selection biases.* En: Libro de proceedings, CTV 2018. XII Congreso Internacional Ciudad y Territorio Virtual. "Ciudades y Territorios Inteligentes". UNCuyo, Mendoza, 5-7 septiembre 2018. Barcelona: CPSV, 2018, p. 247-262.

3. Results and discussion

3.1 The presence of sample selection biases

Table 2 shows the estimation results of the selection model where the dependent variable is the presence of EPC information when transacting in the market. It is a dummy variable where dwellings equipped EPC label is equal to 1, otherwise 0.

In Table 2, the appliances (e.g. air conditioning and heating) and facilities in buildings (e.g. lift and public swimming pool) do matter to the presence of EPC but their impacts are negative. We deduce that the insulation function in energy-efficient dwellings is better than those unequipped ones, especially considering Mediterranean climate in Barcelona Metropolitan. For a better energy–efficient dwelling, that is to say, it is likely to resist the presence of the air conditionings and heatings.

Noted Here the p-value of IMR is close to 0.000, indicating selection biases in this sample indeed exist. Subsequently, this corrected variable, IMR, will be introduced into the following hedonic models to detect and revise those selection biases.

Table 2. **Estimation Results of Selection Model (Probit Model)**

	Coef.	Std. Err.	z	P>z	[95% Conf.Interval]	
Dependent Variable:Dum_EPC						
Constant	-1.12	0.094	-11.850	0.000	-1.304	-0.934
Unit price (Euro/sq.m)	0.00	0.000	2.520	0.012	0.000	0.000
Area (sq.m)	0.00	0.000	0.780	0.433	0.000	0.001
Level of the apartment in the building	0.03	0.005	5.390	0.000	0.016	0.034
Balcony or terrace areas (sq.m)	0.00	0.000	-0.950	0.341	-0.001	0.000
Living room area (sq.m)	0.00	0.001	-3.530	0.000	-0.004	-0.001
Air conditioner (dummy)	-0.03	0.022	-1.590	0.112	-0.078	0.008
Heating (dummy)	-0.28	0.023	-12.380	0.000	-0.326	-0.237
Quality/retrofit (dummy)	-0.04	0.028	-1.320	0.186	-0.091	0.018
gran terrace	0.00	0.000	0.200	0.843	-0.001	0.001
Communal swimming pool (dummy)	-0.11	0.043	-2.500	0.012	-0.192	-0.023
Communal garden (dummy)	0.02	0.034	0.570	0.567	-0.048	0.087
Elevator (dummy)	-0.18	0.021	-8.540	0.000	-0.224	-0.140
Built density (area floor ratio)	-0.02	0.011	-1.850	0.064	-0.041	0.001
Centrality index	0.00	0.005	0.520	0.602	-0.008	0.013
Perception of the presence of green areas	0.00	0.001	0.860	0.392	-0.001	0.002
% People holding university degree (of the context)	0.01	0.002	3.830	0.000	0.003	0.010
% buildings with porter services (of the context)	-0.01	0.001	-4.400	0.000	-0.009	-0.003
CP low socioeconomic level	0.01	0.019	0.580	0.559	-0.026	0.048
CP high socioeconomic level	-0.16	0.035	-4.540	0.000	-0.226	-0.090
	Coef.	**Std. Err.**	**z**	**P>z**	**[95% Conf.Interval]**	
IMR	-1.19	0.151	-7.900	0.000	-1.489	-0.897
rho	-1.00					
sigma	1.19					

Note: Dependent variables is the dummy of EPC in dwellings. Coefficients (Coef.), Standard Error (Std.Err.), Confidence (Conf.). The grey variables mean they could not represent the effect of variables on the presence of EPC.
Source: own elaboration

Citación: CHEN, A. & MARMOLEJO-DUARTE, C. *The marginal price of housing energy-efficiency in Metropolitan Barcelona: issues of sample selection biases.* En: Libro de proceedings, CTV 2018. XII Congreso Internacional Ciudad y Territorio Virtual. "Ciudades y Territorios Inteligentes". UNCuyo, Mendoza, 5-7 septiembre 2018. Barcelona: CPSV, 2018, p. 247-262.

3.2 Corrected samples selection biases on housing prices

Table 3 shows the estimation results of various hedonic models where column 1 (MOD1) and column 3 (MOD3) are the ordinary least squares (OLS) models separated by the nominal and ordinal EPC variables. The other two columns are the results of the Heckman two-step model by IMR variables corrected the samples selection biases. Noted that variables show significance at confidence of 95% and ranking G is the control group.

Table 3. **Estimation Results of Hedonic Models**

		MOD1 **(OLS Model)**	MOD2 **(Heckman two-step Model)**	MOD3 **(OLS Model)**	MOD4 **(Heckman two-step Model)**
	R square	0.654	0.721	0.653	0.721
	R square adjusted	0.652	0.720	0.651	0.720
	Sigma	0.2859	0.3661	0.2862	0.3660
	(Constant)	10.236*** (0.05)	10.861*** (0.151)	10.229*** (0.05)	10.840*** (0.152)
	IMR		-0.408*** (0.094)		-0.410*** (0.094)
Structural characteristics of dwellings	**Area (sq.m)**	0.018*** (0.001)	0.011*** (0.000)	0.018*** (0.001)	0.011*** (0.000)
	Air conditioner	0.101*** (0.013)	0.146*** (0.017)	0.101*** (0.013)	0.146*** (0.017)
	Number of bathrooms	0.064*** (0.012)	0.128*** (0.013)	0.062*** (0.012)	0.129*** (0.013)
	Heating	0.044*** (0.013)	0.182*** (0.031)	0.046*** (0.013)	0.184*** (0.031)
	Quality/retrofit indicator	0.042** (0.017)	0.066*** (0.021)	0.043** (0.017)	0.066** (0.021)
	Area^2	0.000*** (0.000)	0.000*** (0.000)	0.002*** (0.000)	0.003*** (0.000)
Structural characteristics of buildings	**Lift*floor level**	0.012*** (0.002)	0.022*** (0.003)	0.013*** (0.002)	0.022*** (0.003)
	Communal swimming pool	0.134*** (0.026)	0.293*** (0.029)	0.136*** (0.026)	0.294*** (0.029)
Accessibility	**Floor/area ratio**	0.038*** (0.006)	0.052*** (0.007)	0.038*** (0.006)	0.052*** (0.007)
	Centrality indicator	0.01*** (0.003)	0.025*** (0.004)	0.01*** (0.003)	0.025*** (0.004)
Socio hierarchy	**% people holding university**	0.005*** (0.001)	-0.007*** (0.001)	0.005*** (0.001)	0.007*** (0.001)
	CP high socioeconomic level	0.061*** (0.014)	0.101*** (0.019)	0.061*** (0.014)	0.101*** (0.019)
	% buildings with porter services	0.004*** (0.001)	0.003*** (0.001)	0.005*** (0.001)	0.003*** (0.001)
Energy rating	**A**	0.096*** (0.034)	0.126*** (0.037)		
	C	-0.027 (0.026)	0.071** (0.029)		
	D	0.039* (0.019)	0.058*** (0.022)		
	E	0.022 (0.013)	0.036** (0.015)		
	F	0.011 (0.017)	0.007 (0.020)		
	Ord_EPCs			0.009* (0.004)	0.020*** (0.005)

Notes: Dependent variable is ln (total price); *** significance at 99%, ** significance at 95%, *significance at 90%; The grey variables mean they could not represent the effect of variables on the presence of EPC.

Citación: CHEN, A. & MARMOLEJO-DUARTE, C. *The marginal price of housing energy-efficiency in Metropolitan Barcelona: issues of sample selection biases.* En: Libro de proceedings, CTV 2018. XII Congreso Internacional Ciudad y Territorio Virtual. "Ciudades y Territorios Inteligentes". UNCuyo, Mendoza, 5-7 septiembre 2018. Barcelona: CPSV, 2018, p. 247-262.

After correcting sample selection biases by IMR, the R square increase from 0.65 to 0.72. That is to say, the model with the same variables can explain more than 7% cases, which can strengthen the persuasion and results' accuracy. Noted that IMR shows a negative impact on housing prices. The less selection biases, that is to say, the higher housing prices premium.

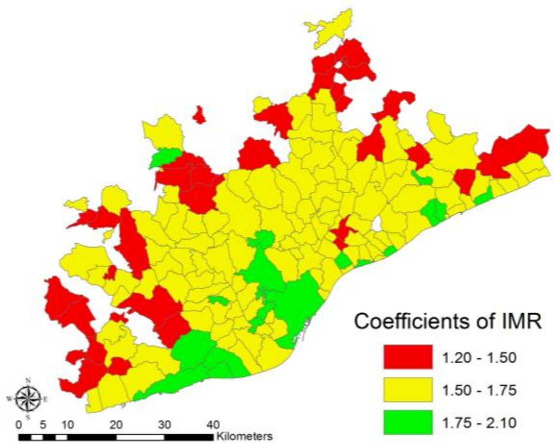
Majority variables show an increase premium on housing prices after biases corrected, especially the impact of the presence of heating and public swimming pool on housing prices, around 15% premium growth. The same conclusion we have concluded from the previous selection model where appliances and facilities in buildings contributed to the presence of EPC.

Regarding energy efficiency information, an energy-efficient premium on housing prices increases from 9.6% to 12.6% when dwellings are improved from ranking G to ranking A or from 0.9% to 2% with energy ranking increase after corrected sample selection biases. Noted that more nominal EPC variables show the significant impacts on housing prices (e.g. ranking C and ranking E). It is to say that sample selection biases may not only influence on estimation results but also on model specification.

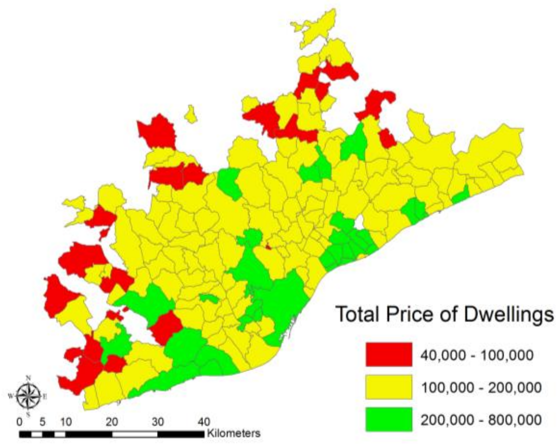
3.3 Selection biases impacts across urban

As previous stated, IMR shows the impact of selected biases in the whole sample: the larger number the higher impacts. Figure 2 (a) shows that distribution of IMR and housing prices on unity price and total prices. The sample selected biases are higher along the coastline, such as Sitges, Barcelona and Maresme zones. In figure 2 (b), we can find these zones affected by selection biases are the place where housing prices are higher. That is to say, it is likely that selection biases happened in the place with high housing prices. The same distribution is to other factors, such as floor area of dwellings and the zone of the proportion of people holding university degree (see figure 2 (c) and figure 2 (d)). Generally, selection biases are more likely happened to dwellings with high prices and medium size floor area, surrounding by a higher proportion university education neighborhood.

Figure 2: **(a) Coeffcients of IMR; (b) Totoal Price of Dwellings; (c) Superficie of Dwellings; (d) Porportion of People Holding University Degree**

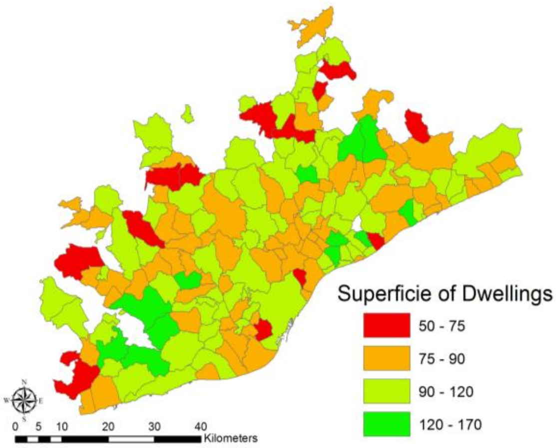


(a)

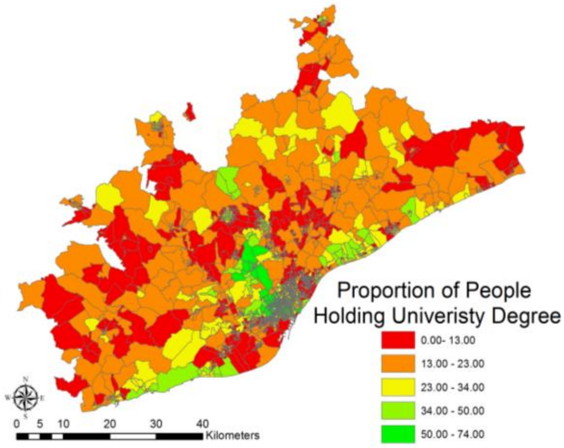


(b)

Citación: CHEN, A. & MARMOLEJO-DUARTE, C. *The marginal price of housing energy-efficiency in Metropolitan Barcelona: issues of sample selection biases.* En: Libro de proceedings, CTV 2018. XII Congreso Internacional Ciudad y Territorio Virtual. "Ciudades y Territorios Inteligentes". UNCuyo, Mendoza, 5-7 septiembre 2018. Barcelona: CPSV, 2018, p. 247-262.



(c)



(d)

Source: own elaboration

4. Conclusions

The process of Energy Performance Certificates has made a great achievement after it is introduced by EPBD in 2002. In order to enhance the public awareness on energy efficiency and promote EPCs process in the residential market, it is mandatory to offer EPCs information when transacting in real estate market from 2010.

Therefore, numerous studies on housing prices impacted by EPCs are investigated but a few studies concerning the selection biases when taking into consideration. In such case, we applied Heckman two-step method to detect the presence of selection biases and corrected these biases in the Hedonic model using IMR.

Our results suggest that selection biases indeed exist and have impact on housing prices regarding energy efficient label. This premium increases from 9.6% to 12.6% when houses improve energy ranking from G to A, or from 0.9% to 2% with every ranking increasing. That is to say, correcting the impact of selection biases brings a 3% increase on housing prices from G to A or 1.1% with energy ranking.

Simultaneously, we find that selection biases are more likely happened to dwellings with high prices and medium size floor area, surrounding by a higher proportion university education neighborhood.

References

BANFI, S.; FARSI, M.; FILIPPINI, M. & JAKOB, M. *Willingness to pay for energy-saving measures in residential buildings*. In: Energy economics, 2008, 30 (2): 503-516. [On line] Available at: https://www.sciencedirect.com/science/article/pii/S0140988306000764 DOI: https://doi.org/10.1016/j.eneco.2006.06.001

Citación: CHEN, A. & MARMOLEJO-DUARTE, C. *The marginal price of housing energy-efficiency in Metropolitan Barcelona: issues of sample selection biases.* En: Libro de proceedings, CTV 2018. XII Congreso Internacional Ciudad y Territorio Virtual. “Ciudades y Territorios Inteligentes”. UNCuyo, Mendoza, 5-7 septiembre 2018. Barcelona: CPSV, 2018, p. 247-262.

BERGSTRÖM, L. & VAN HAM, M. *Understanding neighbourhood effects: Selection bias and residential mobility.* In: IZA Discussion Paper, No. 5193: 77-99. 2012. [On line] Available at SRRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1682714 & https://ssrn.com/abstract=1682714

BIO INTELLINGE SERVICE; MUDGAL, S.; LYONS, L. & COCHEN, F. *Energy Performance Certificates in Buildings and Their Impact on Transaction Prices and Rents in Selected EU Countries*, Bio Intelligence ServiceWorking Paper, April 2013. [On line] Available at: https://ec.europa.eu/energy/sites/ener/files/documents/20130619-energy_performance_certificates_in_buildings.pdf

BOTTERO, M & BRAVI, M. *Valutaziones dei benefici conessi al riasparmio energetico degli edifice: un approccio econométrico*. En: GEAM. GEOINGEGNERIA AMBIENTALE E MINERARIA, 2014, 15-24. [On line] Available at: https://iris.polito.it/retrieve/handle/11583/2595754/80875/articolo%20GEAM.pdf

BROUNEN, D. & KOK, N. *On the economics of energy labels in the housing market*. In: Journal of Environmental Economics and Management, 2011, 62 (2): 166-179. [On line] Available at: https://www.sciencedirect.com/science/article/abs/pii/S0095069611000337?via%3Dihub DOI: https://doi.org/10.1016/j.jeem.2010.11.006

CAN, A. *Specification and estimation of hedonic housing price models*. In: Regional science and urban economics, 1992, 22 (3): 453-474. [On line] Available at: https://www.sciencedirect.com/science/article/pii/0166046292900394?via%3Dihub DOI: https://doi.org/10.1016/0166-0462(92)90039-4

DE AYALA, A.; GALARRAGA, I. & SPARDO, J. *The Price of Energy Efficiency in the Spanish Housing Market*, Energy Policy, 2016, 94, 16-24. [On line] Available at: https://www.sciencedirect.com/science/article/pii/S0301421516301355 DOI: https://doi.org/10.1016/j.enpol.2016.03.032

EPPLE, D.; QUINTERO, L. & SIEG, H. *A new approach to estimating hedonic pricing functions for metropolitan housing markets*. In: FRB of Cleveland, No. 1516. 2014.

FREGONARA, E; ROLANDO, D. & SEMERARO, P. *Energy Performance Certificates in the Turin real estate market.* In: Journal of European Real Estate Research, 2017, 10 (2): 149-169. [On line] Available at: https://www.emeraldinsight.com/doi/pdfplus/10.1108/JERER-05-2016-0022 DOI: https://doi.org/10.1108/JERER-05-2016-0022

FUERST, F.; MCALLISTER, P.; NANDA, A. & WYATT, P. *Does energy efficiency matter to home-buyers? An investigation of EPC ratings and transaction prices in England*. In: Energy Economy, March 2015, 48, 145-156. [On line] Available at: https://www.sciencedirect.com/science/article/pii/S0140988314003296 DOI: https://doi.org/10.1016/j.eneco.2014.12.012

GARCÍA, J. A. B. & HERNÁNDEZ, J. E. R. *Housing demand in Spain according to dwelling type: Microeconometric evidence*. In: Regional science and urban economics, 2008, 38 (4): 363-377. [On line] DOI: https://doi.org/10.1016/j.regsciurbeco.2008.02.002 Available at: https://www.sciencedirect.com/science/article/pii/S0166046208000471

Citación: CHEN, A. & MARMOLEJO-DUARTE, C. *The marginal price of housing energy-efficiency in Metropolitan Barcelona: issues of sample selection biases.* En: Libro de proceedings, CTV 2018. XII Congreso Internacional Ciudad y Territorio Virtual. "Ciudades y Territorios Inteligentes". UNCuyo, Mendoza, 5-7 septiembre 2018. Barcelona: CPSV, 2018, p. 247-262.

GATZLAFF, D. H. & HAURIN, D. R. *Sample selection and biases in local house value indices*. In: Journal of Urban Economics, March 1998, 43 (2): 199-222. [On line] Available at: https://www.sciencedirect.com/science/article/pii/S0094119097920453 DOI: https://doi.org/10.1006/juec.1997.2045

GORDON, B. L. & WINKLER, D. T. *The effect of listing price changes on the selling price of single-family residential homes*. In: Journal of Real Estate Finance and Economics, 2017, 55 (2): 185-215. [On line] Available at: https://link.springer.com/article/10.1007/s11146-016-9558-z DOI: https://doi.org/10.1007/s11146-016-9558-z

HECKMAN, J. J. *The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models*. In: BERG, V. (Editor), NBER, Annals of Economic and Social Measurement, 1976, 5 (4): 475-492. [On line] Available at: https://www.nber.org/chapters/c10491

HECKMAN, J. J. *Sample selection bias as a specification error (with an application to the estimation of labor supply functions*). In: NBER, National Bureau of Economic Research, March 1977, Working Paper, 172. Cambridge, Mass., USA. [On line] Available at: https://www.nber.org/papers/w0172.pdf DOI: https://doi.org/10.3386/w0172

HECKMAN, J. J. & ROBB, R. Alternative methods for solving the problem of selection bias in evaluating the impact of treatments on outcomes. In: WAINER, H. (eds) Drawing Inferences from Self-Selected Samples. Springer, New York, NY, pp. 63-107. Springer, 1986. [On line] Available at: https://link.springer.com/chapter/10.1007/978-1-4612-4976-4_7#citeas DOI: https://doi.org/10.1007/978-1-4612-4976-4_7

HECKMAN, J. J. *Selectivity bias: new developments.* In: American Economic Review, 1990, 80 (2): 313-318.

HECKMAN, J. J. *Selection bias and self-selection*. In: Durlauf S.N., Blume L.E. (eds) Microeconometrics. The New Palgrave Economics Collection. Palgrave Macmillan, London. Springer. pp. 201-224, 1990. [On line] DOI: https://doi.org/10.1057/9780230280816_29 Available at: https://link.springer.com/chapter/10.1057/9780230280816_29

HILL, R. *Hedonic price indexes for housing*. In: OECD Statistics Working Papers, 2011/01, OECD Publishing, Paris, 2011. [On line] DOI: http://dx.doi.org/10.1787/5kghzxpt6g6f-en

HIMMELBERG C., M. C. & SINAI, T. *Assessing high houses prices: Bubbles, fundamentals and misperceptions*. In: Journal of Economic Perspectives, American Economic Association, 2005, 19 (4): 67-92. [On line] DOI: http://dx.doi.org/10.1257/089533005775196769 Available at: https://www.aeaweb.org/articles?id=10.1257/089533005775196769

HYLAND, M.; LYONS, R. & LYONS, S. *The Value of Domestic Building Energy Efficiency: Evidence from Ireland.* In: Energy Economics, November 2013, 40, 943-952. [On line] Available at: https://www.esri.ie/system/files/media/file-uploads/2015-07/JACB201377.pdf DOI: https://doi.org/10.1016/j.eneco.2013.07.020

Citación: CHEN, A. & MARMOLEJO-DUARTE, C. *The marginal price of housing energy-efficiency in Metropolitan Barcelona: issues of sample selection biases.* En: Libro de proceedings, CTV 2018. XII Congreso Internacional Ciudad y Territorio Virtual. "Ciudades y Territorios Inteligentes". UNCuyo, Mendoza, 5-7 septiembre 2018. Barcelona: CPSV, 2018, p. 247-262.

JUD, G. D. & SEAKS, T G. *Sample selection bias in estimating housing sales prices*. In: Journal of Real Estate Research, Summer 1994, 9 (3): 289-298. [On line] Available at: https://www.jstor.org/stable/44095499

LI, F.; NORRBIN, S.; RASMUSSEN, D. & UELAND, J. *Hedonic regression models when unmeasured quality differences are present.* Working paper, Florida State University, Tallahassee, FL. 2005.

MALPEZZI, S. *Hedonic pricing models: a selective and applied review*. In: O'SULLIVAN, T. & GIBB, K. (Editors) Housing economics and public policy, February 2008, 67-89. [On line] Available at: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470690680.ch5 DOI: https://doi.org/10.1002/9780470690680.ch5

MARMOLEJO-DUARTE, C. *The incidence of the energy rating on residential values: an analysis for the multifamily market in Barcelona*. In: Informes de la Construccion, 2016, 68 (543): e-156. [On line] DOI: http://dx.doi.org/10.3989/ic.16.053 Available at: http://informesdelaconstruccion.revistas.csic.es/index.php/informesdelaconstruccion/article/view/5659/6533

MARMOLEJO-DUARTE, C. & CHEN, A. *The Uneven Price Impact of Energy Efficiency Ratings on Housing Segments. Implications for Public Policy and Private Markets*. In: Sustainability, 2019, 11 (2): 372. [On line] Available at: https://www.mdpi.com/2071-1050/11/2/372 DOI: https://doi.org/10.3390/su11020372

MARMOLEJO-DUARTE, C. & GONZÁLEZ-TAMEZ, C. *Does noise have a stationary impact on residential values?* In: Journal of European Real Estate Research, 2009, 2, 259-279. [On line] Available at: https://www.emeraldinsight.com/doi/abs/10.1108/17539260910999992 DOI: https://doi.org/10.1108/17539260910999992

PENA-TRAPERO, J. B. & PENA-TRAPERO, X. *Problemas de la medición del bienestar y conceptos afines: una aplicación al caso español.* Madrid: España: Instituto Nacional de Estadística. 1977. 218 p.

PUHANI, P. *The Heckman correction for sample selection and its critique*. In: Journal of economic surveys, 2000, 14 (1): 53-68. [On line] DOI: https://doi.org/10.1111/1467-6419.00104 Available at: https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-6419.00104

ROCA-CLADERA, J.; MARMOLEJO-DUARTE, C. & MOIX, M. *Urban structure and polycentrism: Towards a redefinition of the sub-centre concept*. In: Urban Studies, 2009, 46: 2841-2868. [On line] DOI: https://doi.org/10.1177/0042098009346329 Available at: https://journals.sagepub.com/doi/abs/10.1177/0042098009346329

ROSEN, S. *Hedonic prices and implicit markets: product differentiation in pure competition*. In: Journal of Political Economy, February 1974, 82 (1): 34-55. [On line] Available at: https://www.jstor.org/stable/1830899

Citación: CHEN, A. & MARMOLEJO-DUARTE, C. *The marginal price of housing energy-efficiency in Metropolitan Barcelona: issues of sample selection biases.* En: Libro de proceedings, CTV 2018. XII Congreso Internacional Ciudad y Territorio Virtual. "Ciudades y Territorios Inteligentes". UNCuyo, Mendoza, 5-7 septiembre 2018. Barcelona: CPSV, 2018, p. 247-262.

SEKO, M. & SUMITA, K. *Japanese housing tenure choice and welfare implications after the revision of the tenant protection law*. In: Journal of Real Estate Finance and Economics, 2007, 35 (3): 357-383. [On line] Available at: http://hdl.handle.net/10.1007/s11146-007-9040-z DOI: https://doi.org/10.1007/s11146-007-9040-z

SHEPPARD, S. *Hedonic analysis of housing markets*. In: Handbook of regional and urban economics, 1999, 3: 1595-1635. [On line] DOI: https://doi.org/10.1016/S1574-0080(99)80010-8 Available at: https://www.sciencedirect.com/science/article/pii/S1574008099800108

ZARZOSA-ESPINA, P. *Estimación de la pobreza en las comunidades autónomas españolas, mediante la distancia DP2 de Pena*. In: Estudios de economía aplicada, (2009, 27 (2): 397-416. [On line] Available at: https://dialnet.unirioja.es/descarga/articulo/3056904.pdf

Citación: CHEN, A. & MARMOLEJO-DUARTE, C. *The marginal price of housing energy-efficiency in Metropolitan Barcelona: issues of sample selection biases.* En: Libro de proceedings, CTV 2018. XII Congreso Internacional Ciudad y Territorio Virtual. "Ciudades y Territorios Inteligentes". UNCuyo, Mendoza, 5-7 septiembre 2018. Barcelona: CPSV, 2018, p. 247-262.