Buildings energy performance and real estate market value: An application of the spatial auto regressive (SAR) model

Buildings Energy Performance and Real Estate Market Value: An Application of the Spatial Auto Regressive (SAR) Model

Marta Bottero, Marina Bravi, Giulio Mondini and Antonio Talarico

Abstract This work explores the role of the buildings energy performance in defining the real estate market value, taking in consideration the presence of spatial autocorrelation. At this regard, it is necessary to put in evidence that a great heterogeneity exists on the Italian territory with reference to buildings energy performance; for this reason, being able to identify a class of most performing estimation models, suitable to separate the spatial effects from the influence of the building components —including the energy rating—on the value, seems to be an interesting goal. In particular, this work illustrates an experiment based on the Spatial Auto Regressive (SAR) model implemented on a sample of residential units located in the city of Turin and represents a first step of a more wide research program.

Keywords Energy requalification · Hedonic pricing model · Residential buildings · Estimation · Post-carbon city

1 Introduction

In recent years the European debate regarding targets and methodologies related to energy policies and building energy performance has greatly intensified. As a matter of fact, in Europe the building sector is responsible for more than 40% of the

M. Bottero (✉) · M. Bravi · G. Mondini
Department of Regional and Urban Studies and Planning,
Politecnico di Torino, Turin, Italy
e-mail: marta.bottero@polito.it

M. Bravi
e-mail: marina.bravi@polito.it

G. Mondini
e-mail: giulio.mondini@polito.it

A. Talarico
Compagnia di San Paolo—Sistema Torino, Turin, Italy
e-mail: antonio.talarico@polito.it

S. Stanghellini et al. (eds.), *Appraisal: From Theory to Practice*,
Green Energy and Technology, DOI 10.1007/978-3-319-49676-4_16

total energy consumption and for 36% of the CO_2 emissions (Blest et al. 2010; Klessman et al. 2011). To avoid a further increase of these values, the European Union decided to issue several Directivesin order to encourage the reduction of energy consumption and to promote the use of renewable energy sources. In this sense, the European Union has committed itself reducing greenhouse gas emissions by 20% compared to 1990 levels, increasing the share of renewable sources in the energy mix to 20%, and achieving the 20% energy efficiency target by 2020. To reach those targets, European legislation set out a cross-sectional framework of ambitious targets for achieving high energy performances in buildings. Key part of this European regulatory framework is the Energy Performance of Buildings Directive 2002/91/EC—EPBD (European Commission 2002; Becchio et al. 2015).

International and national researches have produced, in recent years, a number of studies on how the buildings energy performance affects the real estate market value (Fuerst and McAllister 2011; Eichholtz et al. 2013; Morri and Soffietti 2013; Högberg 2013; Hyland et al. 2013) Some conclusions about the Italian housing market (the largest sector of the real estate) appear to be shared, in particular:

- the existence of an appreciation of the users/buyers for the higher energy efficiency classes (A and B) also considered the strong segmentation of housing markets and demand (Bottero and Bravi 2014);
- a non-proportional relationship between energy rating and the market value, as well as a lower explanatory power of the intermediate energy classes (Fregonara et al. 2014);
- a strong level of spatial heterogeneity in the incidence of the energy rating on the market value, given the characteristics of the built environment.

In this respect, the paper explores the role of the buildings energy performance in the definition of the real estate market value, considering the problem of the presence of spatial autocorrelation and focusing, in particular, on the use of the spatial econometrics.

As it is well known and following the hedonic prices theory (Malpezzi 2003), real estate properties are composite goods where different attributes are affecting the value. For example, historical buildings normally have very low energy performance but very high value due to the central location in the city; on the contrary, new social housing projects, with suburban location, can be very performing from the point of view of energy efficiency.

Effectively, another very important aspect of the real estate markets is related to their complex stratification and segmentation. In particular, they are characterized by a very rigid supply and by a demand affected by cyclical fluctuations (caused by the possibility of getting credit, the households' income or the supply of new constructions in the market). In this sense, the substitution of the buildings characterized by low energy performance with more innovative buildings is conditioned by the trends of the real estate sector, by the availability of technological solutions and by the presence of incentives to investments in this industry (Bottero and Bravi 2014). In this respect, the territorial distribution of the real estate prices is, at the same time, expression of the market dynamic and of stratification, over time, of different levels of building quality. In other words, energy rating is another way to define the quality of a property; to the same extent, it is a way to identify a sub-market and a demand segment. What about the spatial structure in all of this? Since the raising of the building energy rating is achieved, or by new constructions or by a renovation, at least in theory, we expect that this phenomenon has a coherent spatial structure (Fregonara et al. 2012). These implications are very important for local taxation and policies devoted to saving energy. Consistently, the estimation, at the local scale, of costs/benefits of the improvement of the buildings energy performance can profit from an approach of this type.

In the light of the aforementioned considerations, it is necessary to analyze and to map the relationships between buildings energy performance and territorial driving forces that determine the value in order to recognize meaningful patterns and to distinguish different areas where proper policies could operate at local scale.

The present work illustrates a methodological proposal that is based, on the one hand, on the theory of hedonic prices and, on the other, on spatial econometrics (Krause and Bitter 2012), providing an application on a sample of residential units located in the city of Turin. From the methodological point of view, the analysis seems to be innovative because it faces the problem of spatial heterogeneity and inelasticity of the market supply. In fact, the application of Hedonic Pricing Model through the analysis of the spatial dependences involves the problem of spatial autocorrelation to be taken into account (Brasington and Hite 2005). Finally, the identification of spatial patterns of homogeneity/non-homogeneity allows creating a large data base that can be useful for the definition of sectorial policies.

2 Research Methodology

With the purpose of supporting decision processes in the context of energy requalification operations, the research aims at investigating innovative evaluation approaches in attempting to measure the economic value of buildings energy performances. In particular, the study is based on the application of the Spatial Auto Regressive (SAR) model that belongs to the general family of spatial econometrics. These approaches explicitly incorporate space or geography in the analysis and they are rapidly gaining attention in different fields such as urban economics, demand analysis, environmental economics and others. The famous proposition of Tobler (1970) that refers to the first law of geography, remembers us: «*Everything is related to everything else, but closer things more so*». Following Anselin (2010) it is possible to recall that spatial econometrics was originated in the early 1970s as a new investigation field in Europe because of the need to deal with sub-country data in regional econometric models; in recent years, this approach is getting more and more important, especially in the domain of social sciences.

Although this method was developed in other disciplines than real estate appraisal (Paelinck and Klaassen 1979; Anselin 1998), in the last decade, it was successfully applied in this field (Krause and Bitter 2012; Osland 2010). Considering that the *asking prices* are often correlated with nearby property prices, and that the appraisers determine values for financing purposes based on nearby comparable sales, the spatial dependence plays actually a big role in estimation process.

In spatial econometrics, spatial effects pertain to two categories of specifications, namely *spatial autocorrelation* and *spatial heterogeneity*. Spatial heterogeneity concerns structural instability of the relationships between variables in regression models; it represents a special case of observed or unobserved heterogeneity, a familiar problem in standard econometrics. Spatial autocorrelation takes instead into consideration the relative position (distance, spatial arrangements) of the observations in the geographic space. The present study refers to the latter.

Spatial autocorrelation is something like the temporal one, but a little bit different and surely more complicated. In the case of temporal autocorrelation, it is only possible to go one way: what happens at one time can be influenced only by what has happened in the past. In the case of spatial autocorrelation, we can potentially go in any direction. For this reason, models of temporal autocorrelation cannot simply shifted in the geography field (Viton 2010).

It should be remembered that, in standard linear regression models, spatial dependence can be incorporated in two distinct ways: as an additional independent factor, in the form of a spatially lagged variable, or in the error structure. The first is referred as Spatial Auto Regressive (SAR) or *spatial lag* model and it is appropriate when the focus of interest is the estimation of the strength of spatial interactions. Formally, a SAR model can be expressed as:

$$y = +\rho Wy + X\beta + \varepsilon \quad \text{with } \varepsilon \sim N(0, \sigma^2 I)$$

where y is the dependent random variable, α is the constant term, ρ is a spatial autoregressive coefficient, Wy is the spatial weights matrix and ε is the vector of error term IID. Note that the spatial weights matrix Wy (which is usually row-standardized) is correlated with the error term because the disturbances of a neighbour point (or area) depends, in turn, from its neighbours and so progressively. Since Ordinary Least Square (OLS) models would be biased and inconsistent, due to the simultaneity bias, a proper estimation method, which can take in account for this endogeneity, must be employed. Maximum Likelihood Estimation (MLE) approach can answer this need but it is still based on the assumption of error term normality.

At this regard, a very important item of the definition of spatial autocorrelation is the spatial weights matrix that, for each location in the system, specifies which of the other locations is affecting the value. As LeSage and Pace (2009) pointed out, there is an endless number of ways to construct such a matrix. For example, we can consider linear contiguity (two regions with common borders) or a distance-based approach.

As previously mentioned, in this experiment, spatial econometrics is employed in order to investigate the contribution of the energy performance in the definition of the market value starting from a question: does the values exhibit spatial autocorrelation or not? If not, it is clearly possible to employ the standard non-spatial models. But if the spatial autocorrelation does matter, a different path has to be followed that leads to the modeling of spatial relationships. In particular, the method was implemented on the real estate market of the city of Turin and the research was organized according to subsequent steps, some of which are still in progress:

1. Collection and geo-referencing of the data.
   Data related to *asking prices* of residential units in the Turin metropolitan area were collected making use of direct sources (real estate agencies). The records included the address of the apartment, the price, the energy rating and other relevant characteristics affecting the value, such as surface, floor, etc. The single addresses were also geo-referred by means of geographic coordinates (latitude and longitude).
2. Calibration of the models at a small scale (pilot experiment).
   The analysis was implemented on sub-samples of the full data-base with the aim to define the criteria to be used in the construction of the spatial weights matrix and to test the true presence of spatial autocorrelation.
3. Implementation of the models at a large scale.
   Once verified the efficiency of the estimation model at a small scale, the study will try to test the model at a large scale, considering the overall area under investigation. The final goal is to estimate, on a data-base of almost 3,000 cases, the amount of benefits generated by an energy requalification scenario on the metropolitan area of Turin.

3 Pilot Experiment

Following the research methodology illustrated in Sect. 2, a first pilot experiment was developed. It covers the steps number 1 and 2 and concerns a sample of 500 residential units pertaining to the real estate market in the city of Turin.

3.1 Construction of the Sample

The *asking prices* were collected through the Web portal "immobiliare.it" for a period of six months and were geo-referred by means of the open source tool Batchgeo (https://batchgeo.com/). It is important to highlight that the model makes use of *asking prices* or *list prices* that cannot be considered as final prices, but rather as *most probable selling price* affected by error. This is a well-known problem in real estate literature starting from Haurin's (1988) investigation. In Italian real estate market, for example, Fregonara et al. (2014) tried to test whether the appraisers are usual to take in account the location in defining the *asking prices* that definitively represent the first values in a negotiation between supply and demand. In an experimental prospective, the *asking price* is considered a good signal of the value creation process, a clear evidence of the role that space (location) performs in this respect. So, the goals of the pilot experiment were:

– to demonstrate that the energy rating explains the market value;
– to understand if the inclusion of spatial effects provides better results in terms of reduction of model error;
– to prepare and test a methodology whose purpose goes beyond these first steps (economic valuation of the benefits can be produced from an improvement in building energy performances).

The first step consists in the creation of a geo-referred database. For each case, five variables were entered: residential unit address, surface, floor, energy rating and *asking price* plus the latitude and longitude related to every single address. Actually, the available variables were more, but, as first approach to the investigation problem, it was preferred to only test the most important of them.

3.2 Estimation Results

Two estimation models were implemented on a sample of 500 residential units: a traditional linear one, based on the Ordinary Least Square (OLS) algorithm (Table 1) and a SAR model, based on the Maximum Likelihood Estimator (MLE) (Table 2). In both cases, the SpaceStat software, ver. 3.8 (2013), a special purpose package handling both estimation and specification testing of spatial regression models, was used.

As for the most delicate aspect, namely the choice of the spatial weights matrix, it must be specified that point geography spatial weights, and not polygons weights (as neighbourhoods or census tracts), were employed. In this direction, the five points (prices) closer to the point concerned (ego) were considered, defining the weight values for these neighbors. Note that spatial weights sets do not have a weight between ego and itself. The weights calculus was based on the inverse distance rank[1].

The exam of the results of Tables 1 and 2 puts in evidence that they are consistent with similar findings in the literature. In particular, following Osland (2010), the SAR model performs better compared to the OLS model; in fact, the value of R-squared is higher in the former. Moreover, when spatial regression models are

[1]This option assigns weights to nearest neighbors equal to the inverse of their rank. So the nearest neighbor's weight is 1, the second nearest neighbor weight is 1/2, the third nearest neighbor weight is 1/3, etc. For a specified nearest neighbor count, this weight value is essentially standardized because all objects in the data-set have the same count.

Table 1 OLS (linear) estimation model—dependent variable: total asking price

Variable	Coefficient	Std error	Z value	Probability
Intercept	−50297.8	16783.2	−2.99692	0.003
Surface	2970.126	93.32386	31.82601	0.000
Energy rating	−9526.12	2364.248	−4.02924	0.000
Floor	4867.794	1725.151	2.821663	0.005
R-squared			0.795126	
Adjusted R-squared			0.793077	
Log likelihood			−3795.38	
F value			388.1046	
P value			0.0	

Table 2 MLE (autoregressive) estimation model—dependent variable: total asking price

Variable	Coefficient	Std error	Z value	Probability
Rho	0.296535	0.03611	8.21195	0.000
Intercept	−101693	15985.51	−6.36161	0.000
Surface	2748.833	87.80264	31.30694	0.000
Energy rating	−6280.49	2136.156	−2.94009	0.003
Floor	4657.691	1541.389	3.021749	0.002
R-squared			0.829185	
Adjusted R-squared			0.829185	
Square correlation			0.834817	
LogLikely			-3765.06	
AIC			7540.126	
SIC			7558.711	
Sig-Sq (ML)			3.298E 9	

estimated by MLE, inference on the spatial autoregressive coefficients may be based on a Wald or asymptotic *t*-test (from the asymptotic variance matrix) or on a likelihood ratio test (Anselin and Bera 1998). In this case and in both models, the intercept, surface, energy rate and floor pass the *t*-test and they show the validity of the first hypothesis about the role of the building energy performances in determining the asking price. It should also be noted that the energy rating, in the absence of other explanatory variables, such as the status of maintenance and the construction time, tends to incorporate these features in itself as a proxy variable. Moreover, the energy rating has negative sign because the measurement order goes from the worst (G) to the best (A) class. In MLE model, the ρ and the other coefficients are also significant at 0.00 level. The estimates are quite stable in the amounts and the signs are correct.

Table 3 Spatial dependence diagnostics

Test	MI/DF	Value	Probability
Moran's I (error)	0.57188	13.34455	0.000
Lagrange multiplier (error)	1	164.8943	0.000
Lagrange multiplier (lag)	1	182.1001	0.000

Starting from the results of the model comparison, the further analysis of the spatial dependence diagnostics allows validating the initial assumptions. The tests are based on Anselin (1988a, b) (Table 3).

The most commonly used specification test for spatial autocorrelation is derived from a statistic developed by Moran in 1948 as the two-dimensional analog of a test for univariate time series correlation. The Moran's index is normally used in geographic analysis to identify clusters of observations with similar values located in the space. General formal conditions for the asymptotic normality of Moran's I in a wide range of regression models are given in Pinkse (1998). The results of the Moran's I (Table 3) confirm the presence of spatial autocorrelation effects in the case under investigation. In particular, the value of probability equal to 0 means that we have to accept the hypothesis that a spatial dependence exists in the error terms. Moreover, the Lagrange Multiplier verifies the presence of spatial heterogeneity and dependence of the error and of the spatial lag variable.

4 Conclusions

The scant considerations reported here try to testify the first steps of a research still in progress, whose purposes go well beyond what has been shown. Theses short findings emphasize that the estimation of real estate values based on the buildings energy performance goes through a new approach, where the geography becomes an endogenous variable of the observed phenomenon.

In this respect, the feasible uses of a database of this nature appear numerous: from the revision of cadastral values to local taxation; but especially it could be employed for the identification of specific forms of economic incentives, so important for the achievement of the national energy savings.

Another research perspective is to overcome the concept of *estimation sample*, or comparable units, so familiar in the real estate appraisal field, taking into account the spatial and temporal heterogeneity, pivoting on geo-referred and dynamics data-bases (i.e. continuously implemented); these could exceed, at the same time, the problem of little samples representativeness, since the ultimate goal is the almost total coverage of a geographical space.

The big family of models developed by many generations of scholars could be experimented also for this purpose. For example, it will be possible to incorporate the temporal dimension, when observations are available across space as well as over time. Four types of models can be distinguished: *pure space-recursive*, in which the dependence pertains to neighboring locations in a different period; *time–space recursive*, in which the dependence relates to the same location as well as the neighbors locations in another period; *time–space simultaneous*, with both a time-wise and a spatially lagged dependent variable and *time–space dynamic*, with all forms of dependence are present (Anselin 2003).

A lot of work remains to be done and, in this direction, these few lines have tried to mention only the main research goals and the methodology used up to this point.

References

Anselin L (1988a) Spatial econometrics: methods and models. Kluwer, Dordrect

Anselin L (1988b) Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity. Geogr Anal 20:1–17

Anselin L, Bera A (1998) Spatial dependence in linear regression models with an introduction to spatial econometrics. In: Ullah A, Giles DEA (eds) Handbook of applied economic statistics. Marcel Dekker, New York, pp 237–289

Anselin L (2003) Spatial econometrics. In: Baltagi BH (ed) A companion to theoretical econometrics. Blackwell Publishing Ltd, Oxford

Anselin L (2010) Thirty years of spatial econometrics. Pap Reg Sci 89(1):3–25

Becchio C, Bottero M, Corgnati S, Ghiglione C (2015) nZEB design: challenging between energy and economic targets. Energy Procedia 78:2070–2075

BioMedware (2013) SpaceStat. Software for the visualization, analysis, modeling and interactive exploration of spatiotemporal data. User Manual ver. 3.8. BioMedware©

Blesl M, Kober T, Bruchof D, Kuder R (2010) Effects of climate and energy policy related measures and targets on the future structure of the European energy system in 2020 and beyond. Energy Policy 38:6278–6292

Bottero M, Bravi M (2014) Valuing benefits from buildings energy savings: an econometric approach. GEAM. Geoingegneria Ambientale e Mineraria 143:15–24

Brasington D, Hite D (2005) Demand for environmental quality: a spatial hedonic analysis. Reg Sci Urban Econ 35:57–82

Eichholtz P, Kok N, Quigley JM (2013) The economics of green buildings. Rev Econ Stat 95 (1):50–63

European Commission (2002) Directive 2002/91/EC of the European Parliament and of the Council of 16 December 2002 on the energy performance of buildings

Fregonara E, Rolando D, Semeraro P, Vella M (2012) The value of spatial component in real estate market: the Turin case study. Aestimum 60:85–113

Fregonara E, Rolando D, Semeraro P, Vella M (2014) The impact of energy performance certificate level on house listing prices. First evidence from Italian real estate. Aestimum 65:143–163

Fuerst F, McAllister P (2011) The impact of energy performance certificates on the rental and capital values of commercial property assets. Energy Policy 39(10):6608–6614

Haurin DR (1988) The duration of marketing time of residential housing. J Am Real Estate Urban Econ Assoc 16:396–410

Högberg L (2013) The impact of energy performance on single-family home selling prices in Sweden. J Eur Real Estate Res 6(3):242–261

Hyland M, Lyons RC, Lyons S (2013) The value of domestic building energy efficiency. Evidence from Ireland. Energy Econ 40:943–952

Klessmann C, Held A, Rathmann M, Ragwitz M (2011) Status and perspectives of renewable energy policy and deployment in the European Union—what is need to reach the 2020 targets? Energy Policy 39:7637–7657

Krause AL, Bitter C (2012) Spatial econometrics, land values and sustainability: trends in real estate valuation research. Cities 29:19–25

LeSage JP, Pace RK (2009) Introduction to spatial econometrics. Taylor & Francis, Abingdon

Malpezzi S (2003) Hedonic pricing models: a selective and applied research. In: O'Sullivan T, Gibb K (eds) Housing economics and public policy. Blackwell Science, Oxford

Morri G, Soffietti F (2013) Greenbuilding sustainability and market premiums in Italy. J Eur Real Estate Res 6(3):303–332

Osland L (2010) An application of spatial econometrics in relation to hedonic house price modeling. J Real Estate Res 32(3):289–320

Paelinck J, Klaassen L (1979) Spatial econometrics. Saxon House, Farnborough

Pinkse J (1998) Asymptotic properties of the Moran and related tests and a test for spatial correlation in probit models. Working Paper. Department of Economics, University of British Columbia, Vancouver, BC

Tobler WR (1970) A computer movie simulating urban growth in the Detroit region. Econ Geogr 46:234–240

Viton AP (2010) Notes on spatial econometric models. City Reg Plann 870(03):1–23