Spatial Automated Valuation Model (sAVM) – From the Notion of Space to the Design of an Evaluation Tool

Spatial Automated Valuation Model (sAVM) – From the Notion of Space to the Design of an Evaluation Tool

João Lourenço Marques[1](✉), Paulo Batista[1], Eduardo Anselmo Castro[1], and Arnab Bhattacharjee[2]

[1] GOVCOPP, Department of Social, Political and Territorial Sciences, University of Aveiro, Campus Universitário de Santiago, Aveiro, Portugal
jjmarques@ua.pt

[2] Heriot-Watt University, Edinburgh, UK

Abstract. Assuming that it is not possible to detach a dwelling from its location, this article highlights the relevance of space in the context of housing market analysis and the challenge of capturing the key elements of spatial structure in an automated valuation model: location attributes, heterogeneity, dependence and scale. Thus, the aim is to present a spatial automated valuation model (sAVM) prototype, which uses spatial econometric models to determine the value of a residential property, based on identification of eight housing characteristics (seven are physical attributes of a dwelling, and one is its location; once this spatial data is known, dozens of new variables are automatically associated with the model, producing new and valuable information to estimate the price of a housing unit). This prototype was developed in a successful cooperation between an academic institution (University of Aveiro) and a business company (PrimeYield SA), resulting the Prime AVM & Analytics product/service. This collaboration has provided an opportunity to materialize some of fundamental knowledge and research produced in the field of spatial econometric models over the last 15 years into decision support tools.

Keywords: Spatial automated valuation model · Housing market analysis · Spatial econometric models

1 Introduction

Housing can be understood in the strict sense as "… *the stock of houses, apartments, and other shelters that provide the usual residences of persons, families, and households*" [1], but could be more than that. More generally it is: i) a physical facility unit, used and appropriate by an individual or household; ii) a social or collective good that works as a public policy mechanism of social inclusion; iii) a package of services, since one of the fundamental characteristics of a dwelling is its location and in turn all the urban amenities that are located in the neighbourhood; and iv) an economic good, subject to market mechanisms in which there is supply and demand [2]. For these reasons, housing,

O. Gervasi et al. (Eds.): ICCSA 2021, LNCS 12952, pp. 75–90, 2021.
https://doi.org/10.1007/978-3-030-86973-1_6

or more broadly the housing market, plays a central role in modern societies, being closely related to the socio-economic system, to the quality of life of each individual and to the structure of the territory [3, 4]. Housing is a commodity in the usual sense but has some distinguishing features that separate it from other commodity markets [1, 2, 5]. It is characterized by: i) being heterogeneous, in terms of the typology of construction, infrastructure and accessibility; ii) being rigid, since it is a fixed asset in space and has long durability; iii) providing shelter, security and well-being; iv) being an instrument of social distinction, associated with the status image; and v) involving massive collective and private investments [6, 7].

The multiplicity of agents involved in the process of urban transformation, and more precisely in the housing market, with different and contradictory objectives and needs, explains the complexity associated with market operations [8]. This complexity is also compounded by the volatility of exogenous factors that the housing market depends on and by the lack of transparency which results from scarce and asymmetrically distributed information among agents, as well as the inability to use it; this fact leads to a biased price formation mechanism and it is somewhat difficult to explain it rationally and explicitly. In short, it is necessary to provide more information, but most of all it must be better organized to be incorporated into different decision support models.

In order to overcome some of these challenges, particularly the lack of transparency in real estate valuation, automated valuation models (AVM) have been developed and used to more accurately assess the real value of a property. These tools produce objective estimates of market value, based on database and statistical methods, and are designed to perform the same function as manual appraisals (see [9, 10]. Despite some limitations stressed by several authors, essentially related to the unavailability of information, these computational tools are currently popular systems in property analyses [11, 12]. Since the spatial dimension is a determinant component in the estimation of the price of a dwelling (as will be argued in Sect. 3) and captures significant intangible information housing preferences, the aim of this article is to present a spatial automated valuation model (sAVM), which uses spatial econometric models and dozens of housing attributes to determine the value of a residential property. Thus, this article presents a sAVM prototype where the value of a housing unit is estimated (high level of accuracy) using a reduced number of attributes (eight).

In addition to the introduction, this article is organized in four more sections: the second section presents the challenges of having reliable information in the context of housing market analysis; in the third section the relevance and challenges in modelling the spatial component in the real estate market are emphasized; in section four the sAVM and the indicators (housing attributes) behind the spatial model of this prototype are presented; and section five provides conclusions.

2 Data and Information Required

An important consideration highlighted in the previous section is that housing is a heterogenous good that can be defined by a bundle of characteristics, both locational and physical, and that vary in space and also in time. Thus, the availability of reliable and relevant housing data which allows an appropriate analysis on the complexities of the urban housing theme is a central question to monitoring and forecasting housing market dynamics.

In addition to the physical characteristics (intrinsic), the location and all the attributes associated to the spatial environmental are crucial dimensions in housing market analysis [8]. Each individual has specific interests, preferences and economic capabilities, and these thus tend be organized into a complex territorial pattern (heterogeneity) and relations (interactions and spillovers), which are sometimes not easily explained by simple geometric measures (see [7, 8, 13]).

Thus, one of the critical aspects for the success of housing market analysis is the capability to integrate location attributes that can influence the price of a dwelling directly (distances to any specific urban amenity) and indirectly (spatial heterogeneity and dependence). Advances in statistics and spatial econometrics provide valuable contributions to the analysis and understanding of spatial phenomena; however, the methods developed in theses scientific domains depend on the consistency and completeness of the initial information. If a close connection is not ensured between data source, methods and objectives, the results produced may be unreliable [14].

An empirical study of housing can include a multitude of aspects and, depending on the goals of the research, each set of housing attributes can assume different levels of importance; thus, a relevant and corresponding database should be used.

Depending on the purpose of the analysis, various types of data can be used, which can range over different level of spatial disaggregation. However, to take advantage of the potential of spatial econometric techniques, the level of detail of the geographical information must be adequate to capture the particularities of the territory in terms of housing preferences. At the micro scale level, in the particular case of Portugal, three different sources of housing data can be used. First, and the most commonly used, is statistical information provided by the National Institute of Statistics (INE) and the Local Tax on Real Estate database (IMI – Imposto Municipal sobre Imóveis). The former makes a wide range of indicators available on construction and housing in Portugal, while the latter is a database of the Ministry of Finance, in which detailed information of property attributes for all housing that has been traded on the market since 2004 is available. Despite the rich information provided, these data are confidential and the values of the prices shared have a tendency to be significantly below market – both buyer and seller have incentives to declare the minimum value for the transaction to avoid paying higher taxes. These disparities are adjusted by local expert committees, which assess the real price of the dwellings in loco. However, the quality of these adjustments is questionable because the homogeneity of appreciation is not guaranteed.

Real estate agencies are other possible data sources. These private institutions (individually or organized in a group) collect housing market data resulting from their business activity. One of the problems identified in these data sources is the level of territorial coverage. Since each agency has its own niche market, the information is partial and often unrepresentative of the market as a whole, both in sold or rented dwellings.

Another alternative is the data gathered by loan associations and mortgage banks. Some limitations are also to be found in this type of data: one is that housing values are truncated at maximum price, because of the upper limit on conventional mortgage lending amounts; and also properties transacted that are not subject to a mortgage are not included.

This non-exhaustive systematization of data sources that can be used in the context of housing market analysis to illustrate the difficulty in producing data on the real estate market; however, its importance to monitoring the market dynamics is absolutely crucial.

3 Spatial Analysis and Modelling

As mentioned before, the critical challenge for the success of housing market analysis is the ability to incorporate a set of location attributes able which are to explain housing price and capture the (in)tangible spatial elements (in)directly related to a dwelling.

Based on the seminal work of [16] and on revealed preference theory, the use hedonic models is a common approach to decomposing a heterogenous good, based on the idea that a specific good derives from its properties. In the context of housing, the properties are the characteristics, both physical (buildings and typological characteristics of the lot) or location (proximity effect to surrounding neighbourhoods and accessibility to goods and services) that can influence the price of a dwelling. As an dependent variable, we have the dwelling unit values (or proxies such as price or rents) that are regressed on a bundle of housing characteristics (independent variables) which are considered most relevant in the explanation of the house price value.

Despite the challenges in coupling territorial attributes in these models [17], as highlighted before, additional efforts are needed to incorporate this dimension into housing market analysis.

Hedonic pricing models, proposed by [18], allow the price of an item, in this case a house, to be decomposed into separate components that determine its unit values (P), or proxies such as price or rents. In the specific application of housing, the price of a dwelling is regressed on a bundle of characteristics (H) with their respective shadow prices (v), which are normally unknown. The following equation represents the hedonic model in a reduced form [19, 20]:

$$ln\mathrm{p} = f(\mathrm{H}, v) + \varepsilon \tag{1}$$

Housing attributes (H) can typically be organized in two groups: i) structural attributes or intrinsic characteristics and ii) location, neighbourhood attributes or extrinsic characteristics; ε is the vector of regression errors. As the spatial dimension is the focus of this article, we assume that the physical attributes are given by the real estate agencies and are representative of the main intrinsic characteristics.

The integration of spatial attributes faces important challenges [21, 22]. Apart from the choice of explanatory variables and functional specification of the model, the validity of the econometric estimation depends on restrictive assumptions that do not necessarily hold in a housing market context. Spatial homogeneity is a strong assumption in the hedonic housing price context, and if not analysed appropriately' it can be a potential source of specification errors (3).

From urban studies disciplines, a clear statement emerges that the assumption of (bi-dimensional) Euclidean geometric space is quite limited in capturing the above-mentioned spatial complexity [3, 13]. Thus, the notion of space evolved from a concept of absolute space, defined by traditional notions of physical or geographical distances (exogenously given), to a relational space, in which space is socially produced by people – multi-dimensional non-Euclidean space (endogenously produced and spatially organized as a social product). While the former is a reductionist perspective, in which space is socially structured in terms of classes accurately defined (labour and capital), the former is a non-reductionist view of space, in which boundaries and interactions that occur in such space cannot be precisely defined or captured by multiple geometries [8, 23, 24].

Spatial analysis (and consequently, the analysis of the housing market), has been a fast development, and has made considerable efforts to contribute to understanding permanent urban transformations. However, there is still an unsatisfactory connection between spatial analysis and spatial theory, i.e. the conceptual assumptions to support the analytical mechanisms [25]. Theoretical contributions developed in geography, sociology or other social sciences (the "soft sciences"), reject statistical approaches on the grounds that they are incapable of coping with the complexity of the real world and provide misleading generalizations that do not reflect the uniqueness of places and interaction processes. On the other hand, spatial analysis/econometrics (the "hard sciences") are used to include the spatial components as a way to increase the efficiency of their models or to not violate any "sacrosanct" model assumption. Considerable effort has been expended in the spatial econometric field to overcome several statistical problems associated with non-stationarity, normality and homoscedasticity of the errors; however, rather less attention seems to have been paid to the role of these approaches in the interpretation of spatial structures [8]. A brief contextualization of the understanding of space in the urban studies literature (geography, economy and sociology) can be found in [3, 25].

In general, analyses of spatial structures (should) involve four distinctive but interdependent aspects: i) the definition and calculation of location attributes (influences and distances to different territorial amenities); ii) spatial heterogeneity (related to the structural differences between spatial units, regarding their characteristics), iii) spatial dependence (representing the spatial interactions across spatial units) and spatial scale (the territorial level where these phenomena occur); and iv) spatial scale (the level of granularity/disaggregation at which phenomena are described – type of the lens and zoom).

i) *(Proximity to) Location attributes*

Different geographical distances are identified in the literature to capture the explanatory variables of location (environmental and neighbourhood attributes; proximity to urban amenities; and public service characteristics [20, 26], which can contribute to a better capture of the housing value. For instance, [27] suggest three types of geographical distances: i) *global distances*, measured by counting the number of abstract units of length between two objects (kilometres); ii) *effort distances*, related to the effort that someone expends when moving from one point in space to another (travel time, monetary costs, or stress caused by traffic congestion, speed limits, or road quality); and iii) *metaphorical distances*, regarding the process of social cognition – meaning that two objects which are geographically close can be considered to be very distant.

For the location attributes, they can be defined combining the location of each house and the influence of different urban amenities. These geographic influences can be computed according to the classification and principles illustrated in Fig. 1: i) the intensity of use – which can vary from occasional to intensive; ii) the scale and level of influence – which can be local, resulting from the precise location of a dwelling, parishes or any other administrative boundary, or global, involving the influence of external factors; and iii) type of measurements – which can be nominal, measuring if it is under certain influences; ordinal, measuring the level of proximity; or scalar, measuring the proximity of the objects.

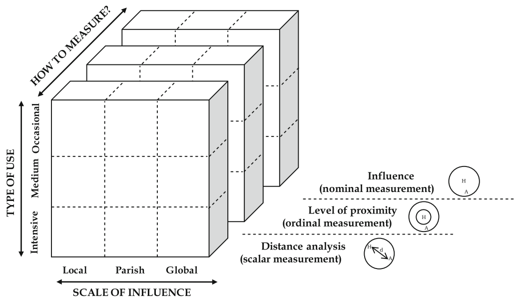

Fig. 1. Three dimensions to build location attributes H_L

It is possible to associate these perspectives, respectively, with Harvey's tripartite notion of absolute, relative and relational space [23], as shown in Table 1.

Table 1. Notions of space and types of distances

Notions of space	Type of distance (and space)
Bi-dimensional Euclidean space	Global distances (absolute space)
Multi-dimensional non-Euclidean space	Effort distances (relative space)
	Metaphorical distances (relational space)

ii) Locality – Spatial Heterogeneity

The relevance and the difficulty in defining submarkets (heterogeneity) is broadly discussed in the literature, as are methods to analyse this aspect of spatial structure [28–31]. Housing prices and attributes, the interests and the preferences of households are not expected to be constant over space. Thus, if not analysed appropriately, it may be a source of specification errors.

The definition of housing sub-market areas has also proved a difficult problem. Nevertheless, some common definitions of a housing sub-market can be found in the literature linked to the concept of substitutability. It comprises dwellings in which, independent of their location, share a set of characteristics that can be considered as substitutes for the potential purchasers. In line with this idea, [28, 29] pointed out that sub-markets are distinctive because houses within them are viewed (more or less) as perfect substitutes by the households. More recently, the same economic concept of substitution has been followed by many works [30, 32].

The early empirical works on sub-markets tended to be segmented into two perspectives, considering that a set of socio-economic and physical characteristics of the territory are acceptable criteria to define housing submarkets; those studies adopt: i) a supply side determinant, including structural and neighbourhood characteristics of dwellings [33], and ii) a focus on demand side determinants, based on household incomes or other demographic and socioeconomic characteristics [34, 35].

Conceptually, submarkets can be defined based on three major perspectives [30]: i) similarity in hedonic housing characteristics', when the submarket is characterized as a collection of locations, or housing units located therein, that have a similar bundle quality; ii) similarity in hedonic prices, when submarkets are characterized as locations where hedonic prices are homogeneous; iii) substitutability of housing units, when submarkets are defined by similarity in house prices. In which circumstance are these approaches equivalent? A comprehensive discussion of this issue is beyond the scope of this article, but it is possible to envisage situations where homogeneity in hedonic characteristics does not imply a close substitutability, for example, two locations with similar houses inhabited by two different social groups (however, different tastes and different responses to fashion are expected to generate a local branding effect in the medium term). Nevertheless, two locations with similar houses and hedonic prices must be good substitutes and it is very difficult to make a distinction between them. Therefore, similarity in hedonic prices and characteristics is a sufficient condition for substitutability, but it is not a necessary condition – for example, two houses with very different hedonic characteristics can be good substitutes: a flat in the centre or a house on the periphery.

There is substantial literature that presents appropriate methods for defining housing markets, following: i) an inductive perspective, in which pre-existing geographic administrative boundaries [36] or a subjective knowledge and expertise of the real estate agents [37] are assumed as a proxy of homogeneous units and used to define submarkets – these procedures are based on a priori judgement, and are subject to a posteriori validation; ii) a deductive and analytical perspective in which several statistical methods are applied, for instance, geographically weighted regression (GWR) [36] and functional data analysis (FDA) [30]. Examples of the use of the hedonic approach to identify space heterogeneity can be found in [39] (see Table 2).

Table 2. Notions of space and sub-markets

Notions of space	Approaches to define submarkets
Bi-dimensional Euclidean space	Inductive perspective (administrative boundaries and expert knowledge) and Deductive and perspective (statistical methods)
Multi-dimensional non-Euclidean space	

iii) Interaction – Spatial dependence

Spatial dependence (or interaction, or spillover, or externalities or spatial autocorrelation) is related to the level of mutual connection of two objects in space, i.e. it occurs when observations at a specific location *i* depend on other observations at a different location *j*. A major challenge here is how to describe the interaction between spatial units, or in other words, to know how individuals or sets of individuals interact. These principles of spatial dependence are closely related to Waldo Tobler's first law of geography, which states "everything is related to everything else, but nearer things are more related than distant things" [40]. But what is the meaning of "near" and what is the notion of space (absolute, relative or relational) on which we base our concept of neighbourhood?

Spatial econometric literature makes a major contribution in this context, describing the interactions between spatial units through a spatial weights matrix (W) [41]. Accordingly, spatial weights are closely related to distances and the challenge lies in the measurement of such distances. There are two distinctive philosophical perspectives in which spatial interaction can be modelled.

One of them largely follows the traditional assumption of an ad hoc predefined matrix W, using geometric distances, both spatial contiguity and geographic, to measure the weights (a deductive approach). Based on this assumed W, global tests for spatial autocorrelation (Moran's index and LISA indicator) and more specific tests of spatial autocorrelation, such as spatial error dependence (SED) and spatial lag dependence (SLD) are used [41].

An inductivist alternative is to move away from the usual practice of an ex-ante definition of spatial interactions, considering that the spatial weights matrix is unknown but can be estimated under some structural constraint (refs). These assumptions are necessary to fully identify W. From this inductive approach, a third challenge emerges, concerning the theoretical background which sustains the assumptions required for identification. Two main alternatives can be formulated:

i) the symmetrical (anti-symmetrical) assumption, considering that $I_{ij} = I_{ji}$ or $I_{ij} = -I_{ji}$, assuming that the phenomenon under analysis describes fluxes between pairs of places. The results are typically represented by topological pattern (e.g. neighbourhoods, graphs – such as communication networks, or space syntax, where the notion of symmetry (or anti-symmetry) can be assumed [3, 13].
ii) a hierarchical diffusion, where the focus is on understanding the spatial diffusion of a shock (e.g. spread of ideas, innovations, trends). The results are usually described

by the amount of “information” spread in the system. In this context, social networks principles suggest a system of hierarchical triangular relations as a plausible structural assumption [15] (see Table 3).

Table 3. Notions of space and sub-markets

Notions of space	Approaches to define spatial dependence
Bi-dimensional Euclidean space	A priori assumption of geographical distances (known weighting matrices)
Multi-dimensional non-Euclidean space	Distances determined by socio-economic relations (unknown weighting matrices)

iv) Scale

Definition of (housing) submarkets and spatial interactions is important and challenging at both conceptual and empirical levels. One of the difficulties in accurately capturing these two spatial features is the definition of the “right zoom” in which these phenomena occur, and consequently the availability of disaggregated data to analyse them [3, 13].

In line with some consideration of the previous theoretical background, the next section presents the model, both the specification and the dimensions, used in the context of the prototype sAVM PrimeYield – UA Tool.

4 sAVM PrimeYield – UA

The sAVM PRIMEYIELD-UA tool was developed to provide real estate property valuation using a spatial econometric model and a restricted number of housing attributes. This tool uses three different data sources: i) data accumulated by PrimeYield (www.prime-yield.com), derived from its real estate evaluation services; ii) statistical data from the National Statistical Institute of Portugal (Instituto Nacional de Estatística - https://www.ine.pt/); and the European Environmental Agency – Corine Land Cover (https://www.eea.europa.eu). More than 120 variables representing the main housing characteristics are used, divided into three main categories: i) physical/intrinsic (such as, dimension, preservation); ii) location (accessibility to several urban amenities) and iii) neighbourhood (socioeconomic and urban characteristics of the territory).

The spatial econometric models were developed considering a database with 625 cases which covered the main urban areas of the municipality of Sintra, located in the metropolitan area of Lisbon – Portugal (see Fig. 2), for a time-period of seven years.

The value of each property, used in these econometric models, refers to the evaluation value given by expert appraisers in their evaluation process. In this context, the results represent the estimations of the possible evaluation value, simulating the (recent past) accumulated behaviour and knowledge of PrimeYield’s expert appraisers.

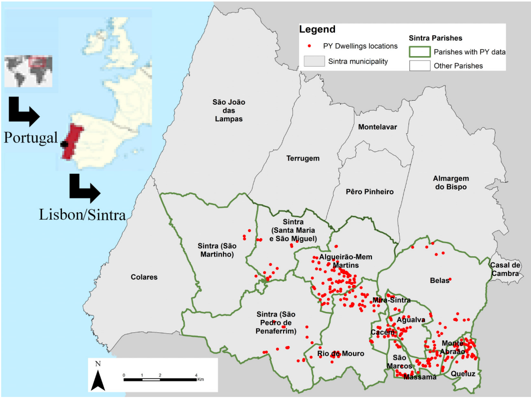

Fig. 2. Location of housing in the municipality of Sintra – Portugal

4.1 sAVM Dimensions and Attributes

The tool AVM-PRIMEYIELD-UA requires an initial identification of eight housing characteristics: 1. Type (flats or single-family houses); 2. Level of conservation (new or used); 3. Net area (m^2); 4. Land area[1] (m^2); 5. Typology (number of rooms). 6. Quality of the property[2] (NA/From 1 to 3); 7; Year of construction; 8. Location (GPS coordinates of the property).

These data inputs are automatically processed by the tool, providing the set of housing variables necessary for correctly estimating the value of a house. It should be noted that, for example, since the precise location of a dwelling is introduced, more than 40 housing dimensions are automatically associated and considered in the model. The detailed information considered in the model is described below.

The specification of the sAVM PRIMEYIELD-UA spatial econometric model follows Eq. 1, defined previously:

$$P = f(v, F; L; T; \varepsilon) \tag{2}$$

where: P is the housing value (in euros with logarithmic transformation); F is the intrinsic (physical) characteristics of housing; L is the extrinsic characteristics (location and neighbourhood) of housing; T is the temporal dimension; υ is the regression coefficients of the model (weights of each attribute in the value of the dwelling); and finally, ε is the stochastic component of the model.

i) Specification of dimension F [Intrinsic (physical) characteristics of housing]

$$\mathrm{H} = \mathrm{f}(T; C; S; A; Q) \tag{3}$$

where: **T** is the typology of housing (T = 1,…,N, with 1 corresponding to T0 and so on); **C** is the level of conservation (new dwellings; dwellings used up to 15 years; dwellings

[1] Only for single-family houses.

[2] The variable related to the level of conservation refers to a qualitative classification provided by PrimeYield's expert appraisers. Based on the available data, only two classifications were incorporated – good and fair; for new properties this classification is not applicable.

used between 15 and 30 years); **S** is the size of a dwelling (small size flats [until 95 m^2], large size flats [more than 95 m^2], small single-family houses [up to 190 m^2], medium-size single family houses [between 190 m^2 and 323 m^2], large single family houses [more than 323 m^2], **A** is the property area (total and land [lnA_i (m^2); lnA_{at} (m^2)]; and **Q** is the quality of the dwelling[3] (1 – Fair, 3 – Good).

In the definition of the categorical variables related to the physical dimension of flats and single-family houses, multivariate analysis statistical techniques – principal component analysis and cluster analysis techniques – were used. This approach is justified by the need to avoid multicollinearity and to capture market segmentation effects (spatial heterogeneity), emphasized in Sect. 3, both in terms of the value and of the size of dwellings.

ii) Specification of dimension L [Extrinsic characteristics (location and neighbourhood) of housing]

$$L = f(IDU; ISE; IAP; IIP, IEP; ITI; IAI; IAN; IAS, D_{zi}) \tag{4}$$

where:

IDU –Indicator for urban density
This includes % of buildings with 1 or 2 floors, % of buildings with five or more floors, % of dwellings with an area between 30 m^2 and 100 m^2, % urban area of continuous urban fabric.
ISE – Indicator for socioeconomic characteristics (qualifications and housing characteristics)
This includes % of buildings built in the 70s, % of buildings built between 2000 and 2005, % of dwellings with an area between 50 m^2 and 100 m^2, % of dwellings with an area less than 200 m^2, % of rented dwellings, % of resident individuals with the first cycle of basic education, % of resident individuals with the second cycle of basic education and % of resident individuals with higher education.
IAP – Indicator for professional activity
This includes % of residents aged between 25 and 65 years, % of residents with the third cycle of basic education, % of residents with secondary education, % of residents employed in the tertiary sector.
IIP – Indicator for the typo morphology and age of the housing stock
This includes % of building blocks, % of buildings built before 1920, % of buildings built between 1920 and 1945, % of buildings built between 1945 and 1960, % of buildings built in the 70s, % of rented dwellings.
IEP – Indicator for the buildings and population age
This includes % of residents aged over 65, % of buildings built between 1945 and 1960, % of buildings built in the 60 s, % of buildings built between 1995 and 2000.
ITI – Indicator for the typo-morphology and age of population

[3] Anticipating possible data collection problems related to this variable, the tool provides an alternative econometric model, which does not use this attribute. As this specification alternative proved to be consistent with the general specification (only a slight decrease in explanatory capacity is noted), it was decided to omit references to this alternative model throughout the following sections.

This includes % of buildings built between 2000 and 2005, % of buildings built between 2005 and 2010, % of residents aged up to four years, % of residents aged between four and five years and % of urban area of medium-continuous urban fabric.
IAI – Indicator for internal global accessibility of the urban area

This includes a measure of accessibility to a set of urban amenities, such as: train stations, schools (kindergarten, primary school (first to third cycles, secondary school), health care facilities (health centre, hospital, pharmacies and diagnostic centres), other shops and services (shopping centres, hypermarkets, grocery stores, markets, ATMs & post offices, restaurants, entertainment facilities and sports facilities).
IAN – Indicator to measure accessibility of municipalities located in the north to the municipality of Sintra.
IAS – Indicator to measure the accessibility of municipalities located in east and south to the municipality of Sintra.
D_{zi} – Dummy to identify zones (Zi = 1,…, 31)
These zones correspond to a spatial disaggregation lower than the administrative delimitation of the parishes. When the zones had a reduced number of cases they were grouped, using a combination of georeferenced information methods and multivariate analysis.

iii) Specification of dimension T

$$T = f(D_{ti}) \tag{5}$$

D_{t1} -Dummy for the year of time i = 1,…,7

4.2 sAVMs Assessment

As discussed above, the adequacy (and the efficiency of the evaluation model) is conditioned by the information provided by the initial data set. Thus, the validation of the estimates is reported under a qualitative index of reliability. The notion of reliability used here refers to a verification of the characteristics introduced for a property and the characteristics (sample) of the supporting data of similar properties. In this version, the verification procedure is applied for two characteristics: the location of the property and its dimensional characteristics, which is based on the following criteria:

Property with parameters within sampling limits
Property with parameters partially out of the sampling limits
Property out of the geographic sample

For the two last classifications it is possible to identify which are the variables where the sample limits are exceeded: **A** - Land area of the property is out of sample; **B** - Housing net area is out of the sample; **C** - Flat net area is out of the sample; **D** – The number of properties of the sample in this zone is reduced (less than 9 records); **E** – In this zone there are no new properties in the sample; **F** – In this zone there are no used in the sample.

The tool implemented a statistical routine for the estimation of confidence intervals. In this context, the confidence intervals aim at providing a concrete indication of the reasonableness of the point estimate – given the unexplained variability embedded in the sample data.

As the use of parametric statistical techniques may be problematic in this context (requiring more demanding statistical assumptions), a non-parametric mechanism of estimation of value ranges predicted by the econometric model was used. Thus, the routine implements a successive sampling scheme, named bootstrapping. In this routine, 500 estimates of the housing value are generated, which are obtained by defining a specific econometric model, generated for a set of 500 subsamples (obtained through replacement based on the original sample data set).

Figure 3 presents the layout of the AVM PRIMEYIELD-UA PROTOTYPE | Version 1.0.

Fig. 3. Interfaces of AVM PRIMEYIELD-UA prototype

5 Conclusions

The main purpose of this article was to explore spatial analysis techniques to be incorporated into a property valuation tool. A spatial automated valuation model (sAVM) was presented, in which the role of space was assumed as a fundamental aspect. Despite the complexity and difficulties of analysing space, four key elements were emphasized: i) location housing attributes, ii) spatial heterogeneity (structural differences between housing markets or housings) and iii) spatial interaction (spatial interactions across submarkets or housings). The scale is a fourth crucial element that makes the process of understanding the relative location (influence of the neighbourhoods), absolute location, the spatial patterns and the spillovers more complex.

The results show different strategies for dealing with spatial heterogeneity across housing submarkets, and independent of the methodology used, the shadow prices and willingness-to-pay for different housing characteristics are different in the selected housing submarkets. In this study no major development of spatial interaction analysis has been presented. However, it can be assumed that an adequate treatment of spatial heterogeneity could considerably reduce the presence of spatial dependence effects, even though the two problems are theoretically distinct [42].

This work was an opportunity to apply some fundamental research, developed in an academic context, in the business world. This was a good example of cooperation, contributing to a more transparent and objective analysis of the real estate market.

Acknowledgements. The authors are grateful to the two reviewers for many helpful comments and suggestions which helped us improve upon the paper. The usual disclaimer applies. This work is an output of the research project DRIVIT-UP - DRIVIng forces of urban Transformation: assessing pUblic Policies, Grant/Award Number: POCI-01–0145-FEDER-031905; Research Unit on Governance, Competitiveness and Public Policy (GOVCOPP), Grant/Award Number: UID/CPO/04058/2019.

References

1. Adams, J.S.: Housing America in the 1980s. Russell Sage Foundation (1988)
2. Bourne, L.S.: The Geography of Housing. Edward Arnold, London (1981)
3. Marques, J.L.: The notion of space in urban housing markets, Doctoral dissertation, Universidade de Aveiro, Portugal (2012)
4. Leung, C.K.Y., Ng, C.Y.J.: Macroeconomic aspects of housing. In: Oxford Research Encyclopedia of Economics and Finance (2019)
5. Bramley, G., Bartlett, W., Lambert, C.: Planning, the Market and Private Housebuilding. UCL Press Ltd, London (1995)
6. Adams, J.S.: The meaning of housing in America. Ann. Assoc. Am. Geogr. **74**(4), 515–526 (1984)
7. Berridge, G.R., Lloyd, L.: In: Barder, R.B., Pope, L.E., Rana, K.S. (eds.) The Palgrave Macmillan Dictionary of Diplomacy, pp. 309–329. Palgrave Macmillan UK, London (2012). https://doi.org/10.1057/9781137017611_18
8. Marques, J.L., Batista, P., Castro, E.A., Bhattacharjee, A.: Housing consumption: a territorial analysis of housing market drivers. In: Anthropological Approaches to Understanding Consumption Patterns and Consumer Behavior, pp. 265–285. IGI Global (2020)
9. d'Amato, M., Kauko, T.: Advances in Automated Valuation Modeling. Springer International Publishing AG, 10, 978-983 (2017)
10. Metzner, S., Kindt, A.: Determination of the parameters of automated valuation models for the hedonic property valuation of residential properties: a literature-based approach. Int. J. Hous. Markets Anal. (2018)
11. Renigier-Bilozor, M., Janowski, A., Walacik, M.: Geoscience methods in real estate market analyses subjectivity decrease. Geosciences **9**(3), 130 (2019)
12. European AVM Alliance: Standards for statistical valuation methods for residential properties in Europe. London (UK). http://www.europeanavmalliance.org/. Accessed 28 Mar 2021 (2017)
13. Bhattacharjee, A., Castro, E., Marques, J.L.: Spatial interactions in hedonic pricing models: the urban housing market of Aveiro, Portugal. Spat. Econ. Anal. **7**(1), 133–167 (2012)
14. Pollakowski, H.O.: Data sources for measuring house price changes. J. Hous. Res., 377–387 (1995)
15. Batista, P.R.L.: The Interaction Structure of E-Territorial Systems: Territory, Housing Market and Spatial Econometrics, Doctoral dissertation, Universidade de Aveiro, Portugal (2019)
16. Lancaster, K.J.: A new approach to consumer theory. J. Polit. Econ. **74**(2), 132–157 (1966)

17. Hasanzadeh, K., Kyttä, M., Brown, G.: Beyond housing preferences: urban structure and actualisation of residential area preferences. Urban Sci. **3**(1), 21 (2019)
18. Rosen, S.: Hedonic prices and implicit markets: product differentiation in pure competition. J. Polit. Econ. **82**(1), 35–55 (1974)
19. Maclennan, D.: Some thoughts on the nature and purpose of house price studies. Urban Stud. **14**(1), 59–71 (1977)
20. Malpezzi, S.: Hedonic pricing model: a selective and applied review. In: O'Sullivan, T., Gibb, K. (eds.) Housing Economics and Public Policy. Blackwell Science Ltd (2003)
21. Straszheim, M.: An econometric analyis of the urban housing market (N. B. of E. Research (Ed.)). Cambridge MA (1975)
22. Freeman, A.M.: The hedonic price approach to measuring demand for neighborhood characteristics. In: Segal, D. (ed.) The economics of neighborhood, pp. 191–217. Elsevier (1979)
23. Harvey, D.: Space as a keyword. In: Castree, N., Gregory, D. (eds.) David Harvey: A Critical Reader, pp. 270–293. Blackwell Publishing Ltd (2006)
24. Lefebvre, H.: The Production of Space. John Wiley & Sons Ltd (1991)
25. Marques, J.L., Batista, P., Castro, E.A.: A noção de espaço em ciência regional. In: Marques, J.L., Carballo, F. (eds.) Ciência Regional em Perspetiva, pp. 11–46. Almedina (2021)
26. Stull, W.J.: Community environment, zoning, and the market value of single-family homes. J. Law Econ. **18**(2), 535–557 (1975)
27. Falk, T., Abler, R.: Intercommunications, distance, and geographical theory. Geografiska Annaler: Series B, Human Geography **62**(2), 59–67 (1980)
28. Grigsby, W.: Housing Markets and Public Policy. University of Pennsylvania Press, Philadelphia (1963)
29. Rapkin, C., Winnick, L., Blank, D.: Housing market analysis. A study of theory and methods. Housing and Home Finance Agency, Division of Housing Research.
30. Bhattacharjee, A., Castro, E., Maiti, T., Marques, J.L.: Endogenous spatial regression and delineation of submarkets: a new framework with application to housing markets. J. Appl. Economet. **31**(1), 32–57 (2016)
31. Marques, J.L., Castro, E., Bhattacharjee, A.: Os padrões espaciais do mercado da habitação: uma noção multidimensional não euclideana do espaço. Revista Portuguesa de Estudos Regionais **32**, 5–23 (2013)
32. Pryce, G.: Housing submarkets and the lattice of substitution. Urban Stud. **50**(13), 2682–2699 (2013)
33. Galster, G.C.: Residential segregation and interracial economic disparities: a simultaneous-equations approach. J. Urban Econ. **21**(1), 22–44 (1987)
34. Adair, A.S., Berry, J.N., McGreal, W.S.: Hedonic modelling, housing submarkets and residential valuation. J. Prop. Res. **13**(1), 67–83 (1996)
35. Maclennan, D., Tu, Y.: Economic perspectives on the structure of local housing systems. Hous. Stud. **11**(3), 387–406 (1996)
36. Straszheim, M.: Hedonic estimation of housing market prices: a further comment. Rev. Econ. Stat., 404–406 (1974)
37. Palm, R.: Spatial segmentation of the urban housing market. Econ. Geogr. **54**(3), 210–221 (1978)
38. Manganelli, B., Pontrandolfi, P., Azzato, A., Murgante, B.: Using geographically weighted regression for housing market segmentation. Int. J. Bus. Intell. Data Min. 13 **9**(2), 161–177 (2014)
39. Goodman, A.C.: Housing submarkets within urban areas: definitions and evidence. J. Reg. Sci. **21**(2), 175–185 (1981)
40. Tobler, W.R.: A computer movie simulating urban growth in the Detroit region. Econ. Geogr. **46**(sup1), 234–240 (1970)

41. Anselin, L.: Spatial Econometrics: Methods and Models. Kluwer Academic Publishers, Dordrecht, The Netherlands (1988)
42. Se Can, A., Megbolugbe, I.: Spatial dependence and house price index construction. J. Real Estate Finan. Econ. **14**(1), 203–222 (1997)