An evaluation method for searching the functional relationships between property prices and influencing factors in the detected data

Pierluigi Morano

Department of Civil Engineering Sciences and Architecture,
Polytechnic University of Bari,
Via Orabona 4, Bari, 70125, Italy
Email: pierluigi.morano@poliba.it

Francesco Tajani*

Department of Architecture and Design,
Sapienza University of Rome,
Via Flaminia 359, Rome, 00196, Italy
Email: francesco.tajani@uniroma1.it
*Corresponding author

Vincenzo Del Giudice and Pierfrancesco De Paola

Department of Industrial Engineering,
University of Naples 'Federico II',
Piazzale V. Tecchio, Naples, 80125, Italy
Email: vincenzo.delgiudice@unina.it
Email: pierfrancesco.depaola@unina.it

Felicia Di Liddo

Department of Civil Engineering Sciences and Architecture,
Polytechnic University of Bari,
Via Orabona 4, Bari, 70125, Italy
Email: felicia.diliddo@poliba.it

Abstract: The economic crisis of the last decade, started from the real estate sector, has spread the awareness of the importance of the use of advanced evaluation models, as a support in the assessments and in the periodic value updates of public and private property assets. With reference to a sample of recently sold properties located in the city of Rome (Italy), an innovative automated valuation model is explained and applied. The outputs are represented by different mathematical expressions, able to interpret and to simulate the investigated phenomena (i.e. the market prices formation). The application carried out outlines, in the selection phase of the best model, the fundamental condition that the valuer must adequately know the reference market. In this way, it is possible to identify the existing patterns in the detected data in terms of mathematical expressions, according to the empirical knowledge of the economic phenomena.

Keywords: price property formation; office market; retail market; automated valuation methods; AVMs; genetic algorithm; reliable valuations.

Reference to this paper should be made as follows: Morano, P., Tajani, F., Del Giudice, V., De Paola, P. and Di Liddo, F. (2022) 'An evaluation method for searching the functional relationships between property prices and influencing factors in the detected data', *Int. J. Business Intelligence and Data Mining*, Vol. 21, No. 1, pp.23–47.

Biographical notes: Pierluigi Morano is currently a Full Professor in Real Estate Appraisal at the Department of Science of Civil Engineering and Architecture, Polytechnic University of Bari, Italy. He is the author and co-author of books, journal papers and conference papers on various topics, including the themes of the plans and investments valuation, the valuation of the cost of the public works, the analysis of the real estate market, the real estate appraisal, and the public-private negotiation in the urban planning.

Francesco Tajani is currently an Assistant Professor in Real Estate Appraisal at the Department of Architecture and Design, Sapienza University of Rome, Italy. He is the author and co-author of more than 80 published works on various topics, including the study of innovative algorithms as support to real estate appraisal, the evaluation of investments on cultural and environmental assets, the elaboration of models for the activation of regeneration projects on degraded areas and the enhancement of buildings in disuse, and the econometric analysis of the dynamics of real estate prices generated by macroeconomic variables.

Vincenzo Del Giudice is currently a Full Professor in Real Estate Appraisal at the Department of Industrial Engineering, University of Naples 'Federico II', Italy. He is the author and co-author of books, journal papers and conference papers on several topics, including the themes of the real estate valuation. He is a component and founding partner of the Italian Society of Appraisal and Valuation.

Pierfrancesco De Paola is currently an Assistant Professor in Real Estate Appraisal at the Department of Industrial Engineering, University of Naples 'Federico II', Italy. He is the author and co-author of published works on various topics, including the study of innovative algorithms as support to real estate appraisal, the econometric analysis of the dynamics of real estate prices generated by macroeconomic variables, and the decision support systems for real estate investments.

Felicia Di Liddo holds a PhD in Real Estate Appraisal and currently has a Research Grant at the Department of Science of Civil Engineering and Architecture, Polytechnic University of Bari, Italy. Her research interests are focused on the plans and investment valuation and the public-private partnerships in the urban planning.

1 Introduction

In the last years, the *advanced analytics* have been characterised by an increasing attention. This contingence is related, on the one hand, to the consistent rise of digital information data and the surprising advances in artificial intelligence (AI); on the other hand, to the widespread need of more accurate interpretative and forecasting models to be quickly implementable and easily adaptable to the unexpected variations of the starting scenario. In the USA, the National Bureau of Economic Research has pointed out the opportunities in a large amount of data and its important implications for the economics profession (Lohr, 2012; Rosasco et al., 2018). With reference to the future governance and management of smart cities, several scholars (Gillespie, 2013; Thatcher, 2014; Rabari and Storper, 2015) have highlighted the fundamental role played by 'spatial' property data for the public and private sectors.

Following the global economic crisis triggered out by the US sub-prime in 2007, in order to ensure more objective and reliable assessments and to analyse the functional relationships between the influencing factors and the property market values, the need to systematise property data from heterogeneous sources (Dachuan and Baoshan, 2012; Juan, 2013; Morano and Tajani, 2014) and to exploit the potential of AI for processing the amount of available large quantity of data (Du et al., 2014) has also spread in the real estate sector. In Italy, where the real estate market has traditionally been opaque and characterised by the scarce availability of property market data, there has been a change of the trend, thanks to the emergence of easily accessible databases [i.e., the Real Estate Observatory (OMI) of the Italian Revenue Agency], or websites dedicated to the real estate market, which allow for the detection of market data, including geo-referenced ones.

Known as automated valuation methods (AVM) and computer assisted mass appraisal (CAMA), the applications of AI on property data are numerous in the reference literature. In particular, McCluskey and Anand (1999) have elaborated a review of the main intelligent hybrid techniques for the residential properties mass appraisal, pointing out the respective strengths and weaknesses. Pagourtzi et al. (2003) have detected the main AVMs implemented in the real estate market, in order to provide for a better understanding of the assessment of locational effects on selling prices. Metzner and Kindt (2018) have outlined the main parameters to be considered in the AVM implementation for the evaluation of residential properties.

The need for mass appraisal models capable of constantly modifying in the short-term and of taking into account real-time economic indices (Bolam, 2017) has been pointed out by the outputs of various AVM applications (Potepan, 1996; Taltavull de La Paz, 2003; Kryvobokov, 2007; Gibler et al., 2014; Cao et al., 2019; Pérez-Rave et al., 2019). These works have showed the fundamental role on changes in property values played by socio-economic variables – in particular, the disposable income – with respect to the technological and locational variables. In this regard, the Art. 208 (3) (b) of the Capital Requirements Regulation (EU) No. 575/2013, that represents the main EU law aimed at decreasing the probability that banks go insolvent, highlights the importance that credit institutions "use statistical methods to monitor the value of the immovable property and to identify the immovable property that needs revaluation." Although the useful of the AVMs for urban planning processes and for contributing to the performance of urban information systems is widely shared (Stumpf González and Torres Formoso, 2006), several authors have warned about the danger of 'black boxes' or models that are not very transparent and difficult to manage from less competent users (Batty, 2012; Boyd and Crawford, 2012; Burgess and Bruns, 2012), who are forced to rely on experts, even if they do not know very often how good these experts are (Giddens, 1990).

Therefore, on the one hand, there is the need to exploit the available property data through algorithms able to process, adapt and effectively replicate the observed phenomena, on the other hand, it is mandatory to ensure the empirical reliability of the results obtained and to provide for easily interpretable and repeatable models, avoiding excessively complex tools which would inevitably lead to the automatism of the process and an increasing uncertainty in the end user (Siwei, 2009).

In the outlined context, some recent data analysis techniques aimed at discovering existing economic relations in detected data are finding great employment. In particular, approaches based on data-driven techniques are gaining considerable interest. Among these techniques artificial neural networks (ANNs) and genetic programming (GP) are probably the most known.

The first applications of ANNs date to the early 1980s. Recently, ANNs have been used to investigate different aspects of real estate market (Morano and Tajani, 2013). However, in spite of their usefulness and spread, ANNs require to identify 'a priori' the structure of a neural network (e.g., transfer functions, model inputs, number of hidden layers, etc.). Another trouble is that ANN models do not result into easily interpretable relationships, which might improve the understanding and the simulation of the economic phenomenon.

GP is a modelling technique that simulates the natural evolutionary selection, where the most suitable individuals (i.e., mathematical expressions of model) improve through successive generations. This method allows a complete exploration of the model's expression space that can use certain criteria set by the user. This strategy generate robust model search and it potentially allows the user to obtain additional information on system behaviour driving out relationships between input and output data. The GP methods are generally considered more appealing than ANNs for those contexts where the understanding of the phenomenon is not complete. The GP methods called 'symbolic regression' (Koza, 1992) is probably the most used. This technique uses the evolutionary search paradigm for developing explicit mathematical expressions of the model to fit a set of data points. However, this technique has some limitations since it produce expressions that grow in length during the evolutionary search.

2 Aim

The topic of the present work concerns the framework outlined. With reference to a case study relating to a sample of 300 properties sold in the city of Rome (Italy), for which the total selling prices and the main influence factors that contribute to the market price formation have been detected, the research has two objectives. The first one consists in the implementation of an innovative AVM to identify, from the detected data, the relationships between the property prices and the influencing factors in terms of mathematical expressions, according to the empirical knowledge of the phenomenon. The methodology employed is a hybrid data-driven technique named evolutionary polynomial regression (EPR), that combines the effectiveness of GP for developing 'transparent' and structured mathematical expression of input-output relationships, with the advantages of classical numerical regression. EPR is capable of generating, for the same case study (i.e., the same data sample), several equations characterised by different statistical performance and consequently different complexity in the functional relationships. About the applications of the EPR method to the real estate sector, the papers in the literature are very few and recent (Tajani et al., 2017a; Morano et al., 2018, 2019, 2020), although the method is characterised by interesting and unexplored potentialities.

The second objective aims at highlighting the fundamental role of the valuer in identifying and selecting the best model, among those generated by EPR, for the interpretation of the actual functional correlations among the variables involved. In particular, the research will verify if a higher statistical performance of the model returned by EPR, which is generally associated with a higher interpretative complexity of the outputs, corresponds to a better and reliable representation of the real phenomena that contribute to the property price formation.

The application, carried out on the case study articulated in the two phases of:

1 implementation of EPR

2 identification of the best model in statistical and empirical terms, defines a procedural code that should be 'mandatory' for the valuer, i.e., an 'instruction manual' of AI innovative algorithms, that constitute a support to the decision, but that cannot substitute the ability to verify the results, the sensitivity and the knowledge of the actual market phenomena of an expert user.

The paper is structured as follows. In Section 3, the case study is described, by specifying the variables considered and the respective descriptive statistics. In Section 4, the applied EPR method is explained, the calculations for the case study are developed and the results are illustrated. In Section 5, the best model – in terms of both statistical performance and consistency with the empirical phenomena – is identified. Finally, in Section 6 the conclusions of the work are discussed.

3 Case study

The case study is constituted by 300 properties located in the city of Rome (Italy), characterised by office and retail intended uses, sold in 2006–2015. The sample has been collected borrowing the data published by 'Immobilium-Nomisma', managed by 'Immobilium' (http://info@immobilium.com) for 2004–2012 and then integrated by 'Nomisma' (http://www.nomisma.it) for 2007–2015. This is a database relating to the sale of corporate properties located in the city of Rome, generally constituted by entire buildings with executive or commercial intended uses. The same database typology is also available for the city of Milan (Tajani et al., 2017b).

In particular, for each property, several factors have been reported in the mentioned database, among which the selling price and the year of sale, the intended use, the total gross floor surface, the location in terms of the property address, the year of construction and the year of the last refurbishment (in the case in which it was realised). The data have been then integrated through local surveys and via the web (http://www.google.it/maps, http://www.agenziaentrate.gov.it/) in order to analyse the main database property characters and the urban context in which each one is located.

It should be outlined that several property characteristics affect market property prices, and it is not possible to determine what may have been the certain deciding factors in the single transactions (Malpezzi et al., 1998). In fact, the market price formation mechanism is a complex problem which remains an open question. Empirical surveys have outlined that the contribution of each factor can change even for adjacent neighbours or for similar properties (Robinson, 1979; Smith et al., 1988; Lavender, 1990; Sheppard, 1999; Boyle and Kiel, 2001). The selection of the explanatory variables is always somewhat arbitrary and involves an unavoidable trade-off between bias from omitted factors and increased sampling variance associated with collinearity (Grether and Mieszkowski, 1974; Gelfand et al., 1998; McCluskey et al., 2000; Oven and Pekdemir, 2006; Selim, 2009; Ozus, 2009). There is relative agreement, however, on what constitutes major influencing factors (Janssen and Yang, 1999; Rabianski et al., 2001; Bourassa et al., 2010).

With reference to 'office' properties, several studies highlighted that better features of social sustainability and 'comfort' in the workplace, in terms of improvement of employees' well-being (Feige et al., 2013; Nappi-Choulet and Décamps, 2013; Fuerst and McAllister, 2011), increase buildings' attractiveness for occupiers and decrease risk for investors, leading to higher occupancy rate and premium on rents or asset values. Therefore, according to the studies mentioned above and taking into account Bourassa et al. (2003), who conclude that it is probably unhelpful to employ too elaborate statistical methods, underscoring the importance of the practical knowledge of real estate agents, the influencing factors have been selected through the support of the experience of appraisers and real estate agents consulted.

In the developed model, the average market value and the average market rent have been introduced among the factors that contribute to the formation of the selling price. The real estate market in the last decades has been characterised by a highly cyclical trend that does not allow considering an uniformity of the economic factors into the formation of the selling prices phenomena in a period of over two years. In order to assume the study sample constituted by 300 collected properties sufficiently consistent for mass appraisal analysis, the two 'market' variables mentioned above, allow to examine the real estate market trends in the final model and to appropriately represent the real estate cycle phase of the homogeneous market area in which each property considered in the analysis, is located. In particular, the values of two economic variables are published by the Italian Revenue Agency with reference to the 'office' and 'retail' intended uses, to the year of sale and to the OMI Micro-zone in which the property is located (http://www.agenziaentrate.gov.it/). The definition of 'micro-zone', according to the Presidential Decree No. 138/1998 and ensuing Regulation issued by the Ministry of Finance, for the Italian regulation, is a part of the urban area that must be homogeneous from an urbanistic point of view and at the same time must constitute a uniform real estate market segment. The two economic variables selected in the analysis can be considered a synthetic indication of the extrinsic characteristics relating to the specific OMI Micro-zone in which the property is located and the specific time period, i.e., the sale year. Finally, the inclusion of these two economic variables allows to apply on the study sample the logic of a static econometric analysis, and to simultaneously obtain a versatile model to the economic evolutions related to different time of assessment.

3.1 Variables e correlations

For each property considered in the analysis, the main factors that contribute to the formation of the total selling prices (P) – the dependent variable of the model – have been collected.

The influencing factors considered have been the following:

- The total *surface* (*S*) of the property, expressed in square metres of gross floor area.
- The quality of the *maintenance conditions* (*M*), considered as a qualitative variable and differentiated, through a synthetic evaluation, by the categories 'to be restructured', 'good' and 'excellent' as a dummy variable. In particular, for the definition of the quality of the *maintenance conditions* (*M*), the assessment has been carried out by comparing the information obtained from the databases consulted about the year of construction and the year of last refurbishment and surveys conducted by web and on site, i.e., through digital photographs or user comments. Each of the three categories that summarise the three possible states of maintenance denotes different quality and conditions. The 'to be restructured' condition (*Mp*) indicates properties for which substantial restructuring interventions are necessary, the 'good' state (*Mg*) indicates office or retail properties that are immediately usable and in which the maintenance conditions are acceptable, whereas the 'excellent' state (*Me*) refers to properties characterised by high aesthetic and structural values (*trophy* properties) with valuable trimmings and architectural qualities.

  Finally, with reference to the locational factors information related to the year of sale of each property, thematic maps published on websites, planning official documents, street maps, reports of the city of Rome have been consulted, in order to capture the real situation to the analysis period.
- The *distance from the nearest subway* (*A*), expressed in km it takes to walk to it.
- The *distance from the central station* of the city (*T*), expressed in km it takes to walk to it.
- The *distance from the nearest highway* (*H*), expressed in km it takes to get there by car.
- The *distance from the nearest urban park* (*G*), expressed in km it takes to walk to it.
- The *distance from the central pole* (*C*) of the city, expressed in km it takes to walk to it. As central pole of Rome the 'Altare della Patria' has been considered, that is a historical monument located in a nerve centre of the city of Rome, from which the main arterial roads of the city develop.
- The *average market value* (*MV*), expressed in euro per square metre of gross floor area, published by the OMI of the Italian Revenue Agency, relative to the 'office' or 'retail' intended use, with reference to the year of sale and to the micro-zone in which the property is located.
- The *average market rent* (*LV*), expressed in euro per square metre of gross floor area and per month, published by the OMI of the Italian Revenue Agency, relative to the 'office' or 'retail' intended use, with reference to the year of sale and to the micro-zone in which the property is located.

In order to obtain a model to be used both for interpretative and predictive purposes, it is necessary to define the functional correlations ordinarily expected by the local market operators between the dependent variable (the total selling prices) and the explanatory variables. Therefore, after having carried out a survey through interviews to the local real estate agents, the following functional relationships have been identified:

- The factor *surface* (*S*) is linked to the total selling prices through a positive correlation. This trend is verified up to very large sizes (> 25,000 m^2), beyond which there could be a change of the functional relationship, due to the reduction in the number of potential buyers able to afford significant monetary amounts for the property purchase.
- For the dummy variables relating to the building's maintenance conditions (*Mp*, *Mg*, *Me*) the variation from a worse state to a better one causes this change in the property prices: the reduction is higher in the 'to be restructured' condition; if the quality of the maintenance conditions is 'excellent', the model does not include any reduction in the selling prices, whereas the decreases occur in the event that one of the other two 'pejorative' states of maintenance conditions is verified.
- A growing trend of the economic variables [*average market value* (*MV*), *average market rent* (*LV*)] generates an increase in the total selling prices.
- For all the locational variables [*distance from the nearest subway* (*A*), *distance from the central station* of the city (*T*), *distance from the nearest highway* (*H*), *distance from the nearest urban park* (*G*), *distance from the central pole* (*C*)] the increase in the distance determines a reduction in the total selling prices.

With the exception of the variables surface (*S*) and the building's maintenance conditions (*Mp*, *Mg*, *Me*) – for which the functional correlations with the dependent variable of the total selling prices are logically expected – and the economic variables – *average market value* (*MV*) and *average market rent* (*LV*) – which constitute a summary indication of market prices relating to a specific temporal and spatial context –, for locational variables the local real estate agents survey outputs confirm the existing literature results.

The extensive literature on the economic effects of transit accessibility in terms of urban and suburban property prices, addresses various transportation modes, among which the main ones are the rapid urban transit (subway, underground, tube, etc.), the urban rail transit (railway trains, tram, etc.) and the extra-urban and urban roads for the private and public transportation facilities (fast road, highway, motorway, etc. for the transit of private and public cars and buses).

With reference to the presence of adequate public transport infrastructures in the urban systems, several authors have pointed out the importance of this factor for buyers and sellers in the negotiation phases. The main researches concern the analysis of the relationships between the possible collective transport modes and the prices of residential and commercial properties (Cervero and Kang, 2011; Trojanek and Gluszak, 2018), demonstrating a 'premium' for each increase in the service efficiency rating.

Since analysis of Dewees (1976) on subway in Toronto, and Damm et al. (1980) on subway construction in Washington, DC, the hedonic regression method represents the most applied technique to test the above-mentioned relationships. Benjamin and Sirmans (1996) have attested a decrease of –2.5% for every 0.16 km away from the metro station. Using the panel data for 35 Chinese cities for 2002–2013, Zhang et al. (2016) have demonstrated that transit facilities can significantly elevate average real estate prices.

With regards to the property value impacts of rail station proximity, it is noted that the train stations may raise the value of nearby properties by reducing commuting costs or by attracting retail activity to the neighbourhood (Bowes and Ihlanfeldt, 2001; Hass-Klau et al., 2004). Debrezion et al. (2007) have found a stronger positive impact of new rail station realisation on commercial property values at short distances and on residential values at slightly longer distances. Martínez and Viegas (2009) have attested a positive impact for the proximity to the Cascais Line in the Lisbon metropolitan area with increasing coefficients on selling prices ranging between +6.75% and +10.73%, confirming a positive influence of proximity the closer the home is to a railway infrastructure. Al-Mosaind et al. (1993) in the context of the metropolitan city of Portland (Oregon) have showed a positive capitalisation of proximity to train stations for residential properties within 500 m of actual walking distance using two distance models.

In the context of the urban projects for the highways realisation, the reference literature aims to analyse the effects of nearby streets on selling prices in terms of homeowners' utility to reach the destination as quickly as possible (Chernobai et al., 2011; Cohen and Schaffner, 2019; Allen et al., 2015). Levkovich et al. (2016) have investigated the housing price dynamics following the development of two highways in the east of the Netherlands. Their studies demonstrate the effect of the new highways has increased housing values in the surrounding residential area by approximately +2.5%–4.3%. Ossokina and Verweij (2011) have studied the economic impacts of a new highway in The Hague on the surrounding residential properties using a repeat sales approach and focusing particularly on the positive effects of the reduced traffic density. Moreover, their outputs confirm selling prices increased by an even higher rate of approximately +5% even before the highways were completed.

However, in the framework outlined numerous studies aimed to examine the relationships between the presence of transportation railway infrastructure or highway and the property prices highlight the negative externalities on nearby properties. In particular, the negative effects linked to the railway proximity are mainly due to noise, view obstruction, and the presence of neglected buildings on adjacent plots (Portnov et al., 2009), whereas the harmful impacts deriving from the new highway realisation may result from an increase in traffic noise pollution, which could be a cause of discount in the value of properties that are located along a newly developed highway (Kim et al., 2007; Nelson, 1982; Theebe, 2004; Wilhelmsson, 2000; Andersson et al., 2009; Del Giudice et al., 2017). With reference to the case study considered in this research, it should be highlighted the specificity of the city of Rome, for which the local real estate agents have underlined a significant utility of the proximity to the access to the nearest highway, especially for the workers who reach the city every day.

Among the locational variables selected in the present analysis as the most influencing characteristics in the selling formation processes, the *distance from the nearest urban park* (*G*) constitutes one of the most relevant factor discussed in the reference literature, consistent with the current policies for sustainable urban development. Over the last few decades, in fact, numerous researches corroborate the assumption that there is a significant impact of urban green areas on real estate prices (Lutzenhiser and Netusil, 2001; Netusil, 2013). In this sense, the main outputs show an increase in selling prices corresponding to a higher proximity of the property to an urban green space – +5.9% (Tajima, 2003), +60% (Fennema et al., 1996), +8/10% (Crompton, 2001). The City Parks Forum Briefing Papers (American Planning Association, 2002) highlights a +117% increase in property prices following the construction of the Centennial Olympic Park, in Atlanta. On the other hand, Troy and Grove (2008) have estimated a positive variation in prices of +5% in the situation in which a house is adjacent to an urban park, compared to another, ceteris paribus and in the same conservative state, is 1 km away from the same park. Similar results have been obtained in numerous analysis aimed at examining the impact on real estate prices of the houses direct view towards a green space – +23.1% (Jim and Chen, 2007), +18% (Damigos and Anyfantis, 2011), +7.1% (Jim and Chen, 2006), +4.9% (Tyrvainen and Miettinen, 2000).

In the analysis carried out, the adopted criterion of *distance* for the locational variables has been selected among those mostly used in the reference literature. In particular, a relevant and current debate among urban planners, traffic engineers and researches concerns the definition of 'the best' modality to objectively measure the accessibility to public spaces or facilities between the distance – expressed in km/miles it takes to walk to them and/or it takes to get there by car (Yang, et al., 2018; Andersson et al., 2010; Apparicio et al., 2007; Robitaille and Herjean, 2008; Larsen and Gilliland, 2008) and the travel time to reach them (Zahavi, 1973; Goodwin, 1978; Schwanen et al., 2002; Dijst and Vidakovic, 2000; Bates et al. 1987; Levinson and Kumar, 1994; Nishii and Kondo, 1992). With reference to the residential market of the city of Rome, the local operators generally outline a higher appreciation for the 'metric' distance modality, as the travel time does not always provide an objective indication, taking into account the different walking speed (e.g., younger people compared to older ones) and the uncertainty related to the vehicular traffic, especially in the context of metropolitan cities (e.g., rush hours compared to weekends). Furthermore, among the different typologies of *distance* that can be used, e.g., Euclidian (straight-line), Manhattan (distance along two sides of a right-angled triangle, the base of which is the Euclidian distance), the shortest network paths (Lotfi and Koohsari, 2009), in the present research the path network distance, measured through the Google Maps geolocation system (http://www.google.it/maps), has been considered. It should be highlighted, in fact, that the criterion of 'real' distance from facilities, transport, green areas or any other points of interest, in terms of the path to be taken (by walking, by car, by bus, etc.) to reach them, represents an effective *proxy* of the time to reach an interest point compared to the geometric distances, as the used geolocation tool also takes into account the real traffic situation in the connecting paths network.

In Table 1 the main descriptive statistics of the variables considered have been reported. In particular, the mean of the dependent variable of the *selling prices* is equal to 21,161,560 €, with values ranging between 261,000 € and 180,000,000 €, whereas the mean of the total *surface* is equal to 5,428 m^2, with minimum and maximum values respectively equal to 600 m^2 and 34,967 m^2. The mean values of the variables *distance from the nearest subway* and *distance from the nearest highway* are respectively equal to 1.46 km walking – value range: (0.02–11.7) – and 3.29 km by car – value range: (0.26–7). As regards to the variable *distance from the nearest urban park*, the mean value is equal to 1.47 km walking, with a wide range of values and whose extreme values are respectively equal to 0.19 km and 12.4 km. The variable *distance from the central station* also shows a wide range of values: the mean value is equal to 4.46 km walking, with the minimum distance detected equal to 0.3 km and the maximum value equal to 15.9 km walking. Analysing the variable *distance from the central pole*, the average value is equal to 4.41 km, the minimum value is equal to 300 metres and the maximum value is equal to 14.6 km. As regards to the quality of the *maintenance conditions*, most of the sample is immediately livable (46%) or constitutes a trophy asset (42%), whereas only 12% needs restructuring interventions. Finally, the economic variables *average market value* and *average market rent* are characterised by mean values respectively equal to 5,800 €/m^2 and 29.96 €/m^2 for month, with maximum values up to 11,950 €/m^2 and 72 €/m^2 for month and minimum values equal to 2,450 €/m^2 and 11,25 €/m^2 for month.

Table 1 Descriptive statistics of the variables

Variable	*Mean*	*Standard deviation*	*Levels/intervals*	*Frequency*
Selling price (€)	21,161,560	30,353,242		
Total surface (m²)	5,428	7,432		
			< 1,000	0.29
			1,000–3,000	0.23
			3,000–6,000	0.22
			6,000–20,000	0.20
			> 20,000	0.06
Distance from the nearest subway (km)	1.46	1.84		
			< 0.5	0.26
			0.5–1	0.15
			1–2	0.43
			2–4	0.11
			> 4	0.05
Distance from the nearest highway (km)	3.29	1.66		
			< 1	0.11
			1–3	0.30
			3–5	0.47
			> 5	0.12
Distance from the nearest urban park (km)	1.47	1.68		
			< 1	0.40
			1–1.5	0.24
			1.5–2	0.26
			> 2	0.10
Distance from the central station (km)	4.46	3.49		
			< 1.5	0.15
			1.5–3	0.36
			3–5	0.18
			5–10	0.24
			> 10	0.07
Distance from the central pole (km)	4.41	3.22		
			< 2	0.21
			2–3	0.25
			3–5	0.24
			5–10	0.24
			> 10	0.06
Maintenance conditions			Excellent	0.42
			Good	0.46
			To be restructured	0.12
Average market value (€/m²)	5,800	2,060		
Average market rent (€/m² · month)	29.96	14.28		

4 The method

EPR can be defined as a nonlinear global stepwise regression, providing symbolic formulas of models. It is global since the search for optimal mathematical expressions of model is based on the exploration of the entire space of formulas by leveraging a flexible coding of mathematical structures (Giustolisi and Savic, 2006).

The generic EPR expression is given as equation (1) shows:

$$Y = a_0 + \sum_{i=1}^{n} \left[a_i \cdot (X_1)^{(i,1)} \cdot \ldots \cdot (X_j)^{(i,j)} \cdot f\left((X_1)^{(i,j+1)} \cdot \ldots \cdot (X_j)^{(i,2j)} \right) \right] \tag{1}$$

where n is the number of additive terms, i.e., the length of the polynomial expression (bias excluded), a_i are numerical coefficients to be assessed, X_i are candidate explanatory variables, $(i, 1)$ – with $1 = (1, \ldots, 2j)$ – is the exponent of the l^{th} input within the i^{th} term in equation (1), f is a function selected by the user among a set of possible alternatives, including no function selection. The exponents (i, l) are chosen by the user from a range of candidate values (real numbers) which should include the value 0.

In brief, the search for model structure is performed by exploring the combinatorial space of exponents to be assigned to each candidate input of equation (1). Thus, although exponent values could be any real number, they are coded as integers during the search procedure. It is worth noting that, when an exponent is = 0, relevant input X_i is basically deselected from the resulting equation. This, in turn, reduces the complexity of final mathematical expressions.

Through the use of a genetic algorithm and the iterative implementation of the least squares method, EPR searches for statistically better expressions of functions that link the possible combinations of vectors of the explanatory variables (i.e., the influencing factors) to the dependent variable (i.e., the selling price). In particular, the algorithm of the EPR method does not require the exogenous definition of the mathematical expression and the number of parameters that fit better the data collected, since the iterative process of the genetic algorithm directly returns the best solution.

The EPR is configured as a procedure whose implementation consists of two main phases. In the first, the search is performed to identify the structure of the model by generating a set of polynomial expressions. In the second one, the classical (numerical) regression method is used in order to estimate the polynomial coefficients.

The key idea of the algorithm used is to generate an entire population of functional expressions based on the capacity of each of them to adapt to the data available. To achieve this goal, the algorithm searches both the expressions forms (i.e., the structure of the model) and the parameters values (i.e., the values of the polynomial coefficients). The technique does not require that the structure of the price function be identified ‘a priori’, i.e., that the model inputs, the numerical coefficients, the exponents, etc. are preliminarily defined by the user in first step of the technique implementation.

The statistical accuracy of each model returned following the EPR implementation is checked through its coefficient of determination (COD), that ranges between 0 and 1, defined in equation (2):

$$COD = 1 - \frac{N-1}{N} \cdot \frac{\sum_{N} \left(y_{estimated} - y_{detected} \right)^2}{\sum_{N} \left(y_{detected} - mean\left(y_{detected} \right) \right)^2} \tag{2}$$

where $y_{estimated}$ are the values of the dependent variable assessed by the method, $y_{detected}$ are the collected values of the dependent variable, N is the sample size in analysis. The model statistical accuracy is greater when the COD is close to the value 1.

A recent version of EPR (named EPR-MOGA) exploits multi-objective genetic algorithms to search those model expressions which maximise accuracy of data and parsimony of mathematical expressions simultaneously (Giustolisi and Savic, 2009). The main advantage of such approach is that EPR-MOGA returns a set of explicit expressions with different accuracy to experimental data and different degree of complexity of mathematical structure of models. The analysis of such trade-off solutions between accuracy and complexity allows the expert selecting those models which are better suited for specific applications.

In practice, the genetic algorithm underlying EPR-MOGA carries out a multi-objective optimisation strategy based on the Pareto dominance criterion. These objectives are conflictual, and aim at:

1 the maximisation of model accuracy, through the satisfaction of appropriate statistical criteria of verification of the equation

2 the maximisation of model's parsimony, through the minimisation of the number of terms (a_i) of the equation

3 the reduction of the complexity of the model, through the minimisation of the number of the explanatory variables (X_i) of the final equation.

The optimisation strategy defined above, leads to a range model solutions (i.e., the Pareto front of optimal models) for the three objectives considered, among which the user could select the most appropriate one according to the specific requirements, the available information of the phenomenon in analysis and the typology of experimental data applied.

5 Application of the method

In this research EPR-MOGA is implemented considering the generic model mathematical structure shown in equation (1) with no function *f* selected and, taking into account several studies outputs (Cassel and Mendelsohn, 1985; Tajani et al., 2016), the dependent variable is represented by the natural logarithm of the selling price (Y = ln (P)). The log-linear form of equation (1) has two attractive characteristics: it allows for the prices of one component to depend in part on the other characteristics in the real property, and it partially mitigates a common form of heteroschedasticity (Malpezzi et al., 1998).

Each additive monomial term is a combination of the selected explanatory variables raised to the proper exponents. In particular, candidate exponents belong to the range (–2; –1; –0.5; 0; 0.5; 1; 2), in order to have a wide set of solutions. Except for exponents 1 and 0 which correspond, respectively, to the case in which an independent variable participates to the price formation or not, the other exponents have been chosen on the basis of the empirical knowledge of the investigated phenomenon and therefore of the type of expected linkages between the dependent variable (i.e., the market prices) and the independent variables. For example, the exponent '–1' translates the inverse relationship that in practice is generally found between the property prices and the distance from infrastructures (e.g., the central station, the nearest motorway access, the subway, etc.) or from specific amenities (e.g., an urban park, the central pole, etc).

The maximum number n of additive terms in final equations is assumed to be eight, i.e., equal to the number of explanatory variables considered in the analysis.

The application of EPR-MOGA has generated 14 equations (Table 2) classified – from the 1st to the 14th – according to the increasing statistical accuracy of the outputs in terms of COD (Figure 1) and to the complexity of the models in relation to the number of terms, the number of selected explanatory variables and the combination of the explanatory variables that constitute each term.

Table 2 Models generated by the application of EPR-MOGA

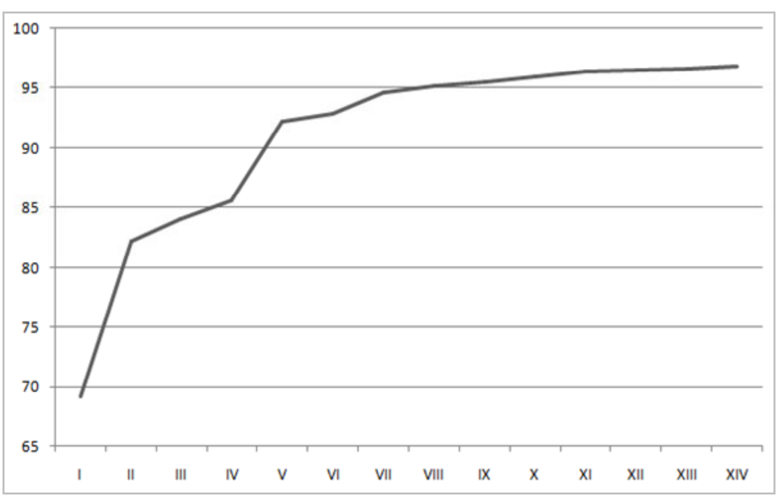
[n]	*Equation*	*COD*
(I)	$Y = 0.03112 \cdot S^{0.5} + 13.984$	69.22
(II)	$Y = 0.00049999 \cdot S^{0.5} \cdot MV^{0.5} + 13.6747$	82.16
(III)	$Y = -0.2822 \cdot T^{0.5} + 0.00050153 \cdot S^{0.5} \cdot MV^{0.5} + 14.2224$	84.08
(IV)	$Y = -0.074071 \cdot LV + 0.00063951 \cdot MV + +0.027163 \cdot S^{0.5} + 12.7337$	85.65
(V)	$Y = -0.7053 \cdot \frac{LV}{S^{0.5}} + 0.051784 \cdot MV^{0.5} + +0.022734 \cdot S^{0.5} + 11.3341$	92.17
(VI)	$Y = -0.67373 \cdot \frac{LV}{S^{0.5}} + 0.00027288 \cdot MV + -0.23159 \cdot T^{0.5} + 0.023182$ $\cdot S^{0.5} + 14.032$	92.85
(VII)	$Y = -13.2721 \cdot \frac{C^{0.5}}{S^{0.5}} - 0.50164 \cdot Mp^{0.5} - 0.029437 \cdot G^{0.5} \cdot LV$ $+ + 0.16201 \cdot G + 0.00091219 \cdot S^{0.5} \cdot LV + 16.0111$	94.62
(VIII)	$Y = -13.2721 \cdot \frac{C^{0.5}}{S^{0.5}} - 0.50164 \cdot Mp^{0.5} - 0.029437 \cdot G^{0.5} \cdot LV$ $+ + 0.16201 \cdot G + 0.00091219 \cdot S^{0.5} \cdot LV + 16.0111$	95.21
(IX)	$Y = -12.438 \cdot \frac{C^{0.5}}{S^{0.5}} - 0.42821 \cdot Mp^{0.5} - 0.51164 \cdot C^{0.5} + -0.032765$ $\cdot G^{0.5} \cdot LV + 0.066977 \cdot G^{0.5} \cdot C + +0.0008904 \cdot S^{0.5} \cdot LV + 16.973$	95.55
(X)	$Y = -6.8001 \cdot \frac{G^{0.5}}{S^{0.5}} + 0.30242 \cdot Me^{0.5} - 1{,}391.839 \cdot \frac{C^{0.5}}{MV} + +2.2312 \cdot \frac{A^{0.5}}{LV^{0.5}}$ $-0.00025939 \cdot A^{0.5} \cdot C^{0.5} \cdot LV^2 + +0.011887 \cdot S^{0.5} \cdot LV^{0.5} - 3.2935 \cdot 10^{-8}$ $\cdot S \cdot MV + 13.9319$	95.92
(XI)	$Y = -\frac{63.7145}{S} - 0.2825 \cdot Mg^{0.5} - 0.085433 \cdot Mp \cdot T - 1{,}087.8892 \cdot \frac{C^{0.5}}{MV}$ $+ - 0.36817 \cdot G^{0.5} - 0.00027038 \cdot A^{0.5} \cdot C^{0.5} \cdot LV^2 + 0.93453 \cdot \frac{A}{LV^{0.5}}$ $+ + 0.012108 \cdot S^{0.5} \cdot LV^{0.5} - 3.2875 \cdot 10^{-8} \cdot S \cdot MV + 14.5701$	96.37

Table 2 Models generated by the application of EPR-MOGA (continued)

[n]	*Equation*	*COD*
(XII)	$Y=-\frac{66.0516}{S}-0.27245\cdot Mg^{0.5}-1{,}322.3486\cdot\frac{C^{0.5}}{MV}-0.034431\cdot G\cdot LV^{0.5}$ $+-0.02429\cdot H\cdot Mp\cdot T-0.0002511\cdot A^{0.5}\cdot C^{0.5}\cdot LV^{2}+1.1555\cdot\frac{A}{LV^{0.5}}$ $++0.012039\cdot S^{0.5}\cdot LV^{0.5}-3.2682\cdot 10^{-8}\cdot S\cdot MV+14.423$	96.51
(XIII)	$Y=-0.0272857\cdot\frac{C^{0.5}\cdot MV}{S}-0.27454\cdot Mg^{0.5}-1278.414\cdot\frac{C^{0.5}}{MV}+-0.042673$ $\cdot G\cdot LV^{0.5}-0.027348\cdot H\cdot Mp\cdot T-0.00021483\cdot A^{0.5}\cdot C^{0.5}\cdot LV^{2}++0.31446$ $\cdot\frac{A\cdot T^{0.5}}{LV^{0.5}}+0.011714\cdot S^{0.5}\cdot LV^{0.5}-3.1038\cdot 10^{-8}\cdot S\cdot MV+14.5684$	96.62
(XIV)	$Y=-0.00082742\cdot\frac{C^{0.5}\cdot MV}{S^{0.5}}-0.23958\cdot Mg^{0.5}-43.1698\cdot\frac{C^{0.5}}{MV^{0.5}}$ $+0.0024846\cdot C^{0.5}\cdot T^{2}+-0.010173\cdot G\cdot LV-0.0054782\cdot H^{2}\cdot Mp\cdot T$ $-0.00014608\cdot A^{0.5}\cdot C^{0.5}\cdot LV^{2}++0.6664\cdot\frac{A}{LV^{0.5}}+0.011314\cdot S^{0.5}\cdot LV^{0.5}$ $-3.0496\cdot 10^{-8}\cdot S\cdot MV+15.259$	96.89

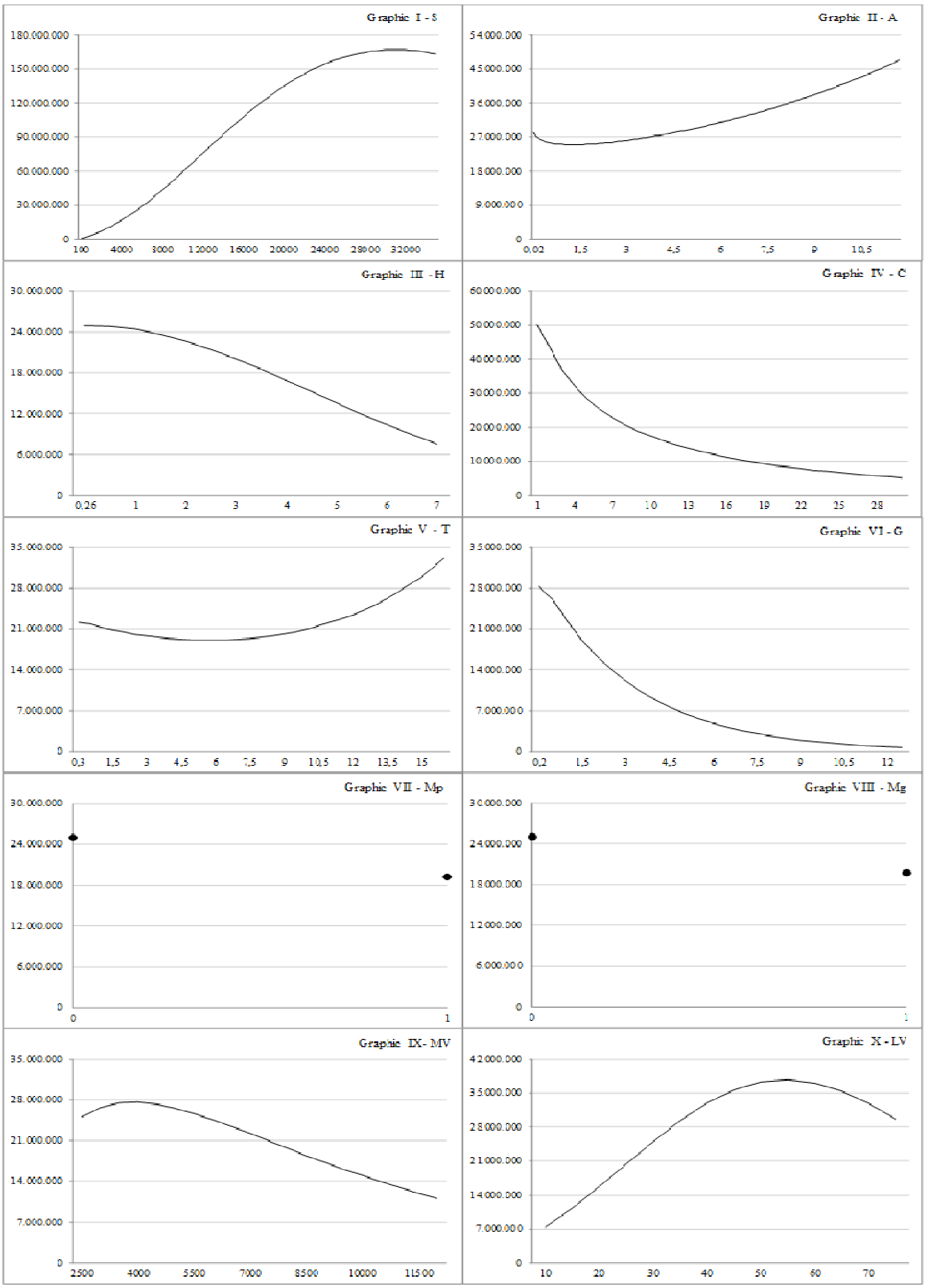
In particular it is evident that, as the COD increases from equations (I) to (XIV), the complexity of the equations also grows, and therefore the difficulty in interpreting the analysed phenomena accordingly rises. The CODs related to the models range from the minimum value of 69.22% for equation (I) to the maximum value of 96.89% for equation (XIV).

Figure 1 Trend of the CODs of the models obtained by the application of EPR-MOGA



With reference to the case study in analysis, the only use of the statistical criterion would lead to choose equation (XIV) as the model that is most capable of replicating the analysed phenomenon, as it is characterised by a COD next to unity and therefore by a very high degree of statistical reliability. This model consists of all the explanatory variables considered.

Figure 2 Functional relationships between the influencing factors and the selling prices for the model of equation (XIV)



However, the complexity of the terms of the mathematical expression does not allow an immediate interpretation of the functional relationships among the variables. Therefore, the functional links of the i^{th} independent explanatory variable with the variation of the selling prices have been explained through an exogenous simplified approach that, instead of determining the partial derivative of the dependent variable with respect to the i^{th} variable, considers the values of the other variables in the model equal to their average values of the starting database, and provides the analysis of the variations in value of the assessed changes of selling prices in correspondence of each i^{th} variable in the admissible range of its corresponding sample values.

The outputs of the elaborations carried out for the model of equation (XIV) have been represented in Figure 2. Firstly, it should be outlined that the empirical evidences detected by the interviews to the local real estate agents (see paragraph 'variables and correlations') are not verified for all the functional relationships between the selected explanatory variables and the selling prices in the model of equation (XIV). In particular, the functional relationships of the variables *distance from the nearest subway* (*A*) and *distance from the central station* (*T*) are 'parabolic' type: the correlation is negative for distances close to the considered infrastructures, beyond which it becomes positive. These correlations, therefore, do not correspond to the empirical phenomena ordinarily expected. Furthermore, the economic variables *average market value* (*MV*) and *average market rent* (*LV*) are characterised by functional relationships with the selling prices that are not always increasing, that are instead the reasonably expected trend from an empirical point of view. Therefore, although equation (XIV) is able to reproduce the data of the starting sample with the highest statistical precision, it cannot be identified as the *best* model, as it does not return an interpretation of the economic phenomenon that is consistent with what is normally observed in the local market.

Analysing the equations generated by the EPR-MOGA implementation, it can be observed that, starting from the model of equation (VII) and up to equation (XIV), the complexity of the mathematical expressions significantly changes, but the statistical performance increases in a limited way.

At this point, it is appropriate to cross-check the results in terms of statistical reliability (COD) with the empirical knowledge of phenomena, in order to identify the model among those generated by EPR-MOGA that, in addition to reproduce the investigated data, allows the optimal interpretation of the economic phenomena. For this purpose, for each of the equations in Table 2, the following selection criteria have been considered (Table 3): the number of terms (including the bias), the number of variables selected among those initially considered in the analysis, the maximum number of variables in each additive term, the statistical performance coefficient (COD), the empirical consistency expressed as the number of correlations between the dependent variable and the explanatory ones that do not verify the trend of the market phenomena outlined by the local real estate agents.

The analysis of Table 3 shows that, except for equations (I) to (VI), for which there is a less significant statistical performance, only for equations (VII) and (VIII) the empirical evidence is fully verified, i.e., all the correlations between the dependent variable and the explanatory ones are consistent with the expected trend of the phenomena (number of functional relationships for which the empirical correlation does not occur = 0). Therefore, in order to select a model capable of reproducing the economic phenomena and at the same time of interpreting it reliably, it is appropriate to choose one of these two equations.

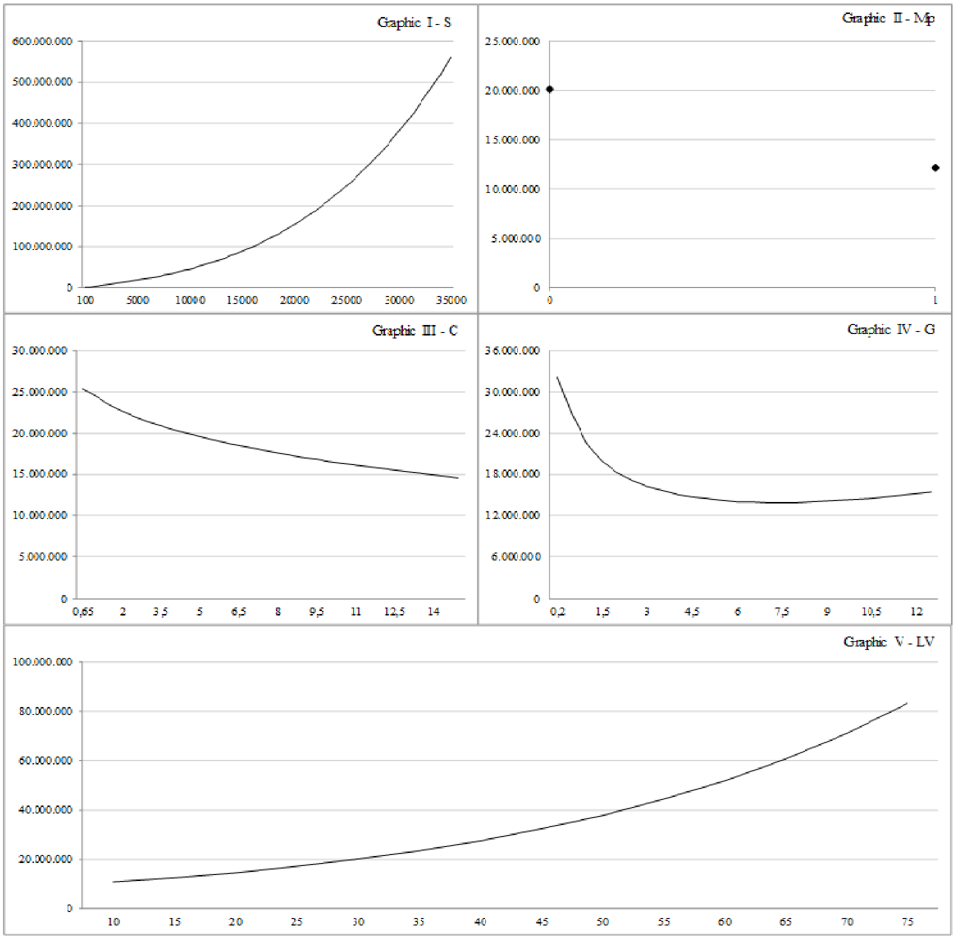
Table 3 Synoptic matrix of the values of the criteria for each model

Model	*Number of terms*	*Number of variables*	*Maximum number of variables in the terms*	*COD [%]*	*Empirical consistency not verified*
(I)	2	1	1	69.22	0
(II)	2	2	2	82.16	0
(III)	3	3	2	84.08	0
(IV)	4	3	1	85.65	1
(V)	4	3	2	92.17	1
(VI)	5	4	2	92.85	1
(VII)	6	5	2	94.62	0
(VIII)	7	5	2	95.21	0
(IX)	8	7	2	95.55	1
(X)	8	7	3	95.92	2
(XI)	10	9	3	96.37	3
(XII)	10	10	3	96.51	3
(XIII)	10	10	3	96.62	4
(XIV)	11	10	3	96.89	4

Both the equations involve only five factors among the explanatory variables considered (*C*, *S*, *G*, *Mp* and *LV*). Furthermore, the difference in the statistical performance between the two equations is really low [COD of equation (VII) = 94.62%, COD of equation (VIII) = 95.21%], whereas the most obvious difference is that equation (VIII) is slightly more complicated in mathematical terms, as it is constituted by an additional term and two square root variables. These considerations lead to the identification of equation (VII) as the *best* model, as it well satisfies the criterion of the statistical reliability, it is relevant from an empirical point of view and it allows a generalisation of the functional relationships for the examined context.

Retracing the same exogenous simplified approach previously implemented for the verification of the qualitative and quantitative correlations among the variables, in Figure 3 the functional relationships of the five influencing factors with the selling prices have been represented according to the model of equation (VII). In this case, the explicit correlations – positive for the *surface* (*S*) and the *average market rent* (*LV*), negative for the *distance from the central pole* (*C*), the *distance from the nearest urban park* (*G*) and the 'to be restructured' *maintenance conditions* (*Mp*) – are all characterised by a complete empirical reliability.

Figure 3 Functional relationships between the influencing factors and the selling prices for the model of equation (VII)



6 Conclusions

The economic crisis of the last decade generated by the real estate sector, has spread the awareness of the importance of the use of advanced models of AI also in the property valuations, as a support in the assessment and the periodic updates of the values of public and private property assets.

The risk of an excessive automation of the evaluation process is however inevitably high. Through the logic-evaluative process described in this research, articulated in the two phases of the innovative AVM implementation and the identification of the best model in statistical and empirical terms, the present work defines a behavioural code of the expert valuer for the identification of the property price function. The experimental application of the proposed AVM has highlighted the need to take into account, in selection phase the model that at best represents the phenomena of property price formation, the criteria of statistical performance, empirical evidence of the outputs, appropriate market knowledge and possibilities of generalisation of the results.

Therefore, the figure of the valuer plays a central role in the control phase of the outputs that an AVM model generates, through the interpretation of the results obtained and the verification of the generalisability of the achieved outputs, in order to escape from particular contingencies related to specific input data. The development of increasingly sophisticated and complex information systems cannot replace the need for interpretative analysis of the results, but it rather implies the peremptoriness that the valuers possess the appropriate skills in terms of mass appraisals and management of automation models (Faishal Ibrahim et al., 2005). The interconnected agendas of smart cities and available property data provide for bold and exciting opportunities for built environment professions (Dixon et al., 2018). A wider awareness of the considerable potentialities deriving from the effective management of a wide amount of data through the use of innovative tools represents an important goal for subsequent research: the willingness to adapt to its evolution will divert or deliver future opportunities.

Future insights may concern the comparison between the functional relationships derived from the stated preferences of the local market operators – obtained through appropriate elaborations on surveys, interviews and direct investigation to be carried out – and the models generated by the EPR-MOGA application and/or by the implementation of several econometric techniques, in order to point out the different outputs and to verify the consistency with the expected empirical phenomena.

Acknowledgements

The writers would like to thank the Central Observatory of the Real Estate Market and Estimative Services (Italian Revenue Agency), for his availability to provide the necessary real estate data for the analysis carried out and explained in the paper.

References

Allen, M.T., Austin, G.W. and Swaleheen, M. (2015) 'Measuring highway impacts on house prices using spatial regression', *Journal of Sustainable Real Estate*, Vol. 7, No. 1, pp.83–98.

Al-Mosaind, M.A., Dueker, K.J. and Strathman, J.G. (1993) 'Light-rail transit stations and property values: a hedonic price approach', in *Transportation Research Record 1400, TRB*, pp.90–94, National Research Council, Washington, DC.

American Planning Association (2002) *City Parks Forum Briefing Papers*.

Andersson, D.E., Shyr, O.F. and Fu, J. (2009) 'Does high-speed rail accessibility influence residential property prices? Hedonic estimates from southern Taiwan', *Journal of Transport Geography*, Vol. 18, No. 1, pp.166–174.

Andersson, H., Jonsson, L. and Ögren, M. (2010) 'Property prices and exposure to multiple noise sources: hedonic regression with road and railway noise', *Environmental and Resource Economics*, Vol. 45, No. 1, pp.73–89.

Apparicio, P., Cloutier, M.S. and Shearmur, R. (2007) 'The case of Montreal's missing food deserts: evaluation of accessibility to food supermarkets', *International Journal of Health Geographics*, Vol. 6, No. 1, p.4.

Bates, J., Dix, M. and May, T. (1987) 'Travel time variability and its effect on time of day choice for the journey to work', *Proceedings of Seminar C: Transportation Planning Methods, 15th PTRC Summer Anual Meeting*, Planning and Transport Research and Computation, London, pp.293–311.

Batty, M. (2012) 'Smart cities, big data', *Environment and Planning B*, Vol. 39, No. 2, pp.191–193.

Benjamin, J.D. and Sirmans, G.S. (1996) 'Mass transportation, apartment rent and property values', *The Journal of Real Estate Research*, Vol. 12, No. 1, pp.1–8.

Bolam, C. (2017) 'Data, data, everywhere', *Property Journal*, pp.16–17, RICS Property Journal rics.org/journals.

Bourassa, S., Cantoni, E. and Hoesli, M. (2010) 'Predicting house prices with spatial dependence: a comparison of alternative methods', *Journal of Real Estate Research*, Vol. 32, No. 2, pp.139–159.

Bourassa, S.C., Hoesli, M. and Peng, V.S. (2003) 'Do housing submarkets really matter?', *Journal of Housing Economics*, Vol. 12, No. 1, pp.12–28.

Bowes, D.R. and Ihlanfeldt, K.R. (2001) 'Identifying the impacts of rail transit stations on residential property values', *Journal of Urban Economics*, Vol. 50, No. 1, pp.1–25.

Boyd, D. and Crawford, K. (2012) 'Critical questions for big data', *Information, Communication & Society*, Vol. 15, No. 5, pp.662–679.

Boyle, M.A. and Kiel, K.A. (2001) 'A survey of house price hedonic studies of the impact of environmental externalities', *Journal of Real Estate Literature*, Vol. 9, No. 2, pp.117–144.

Burgess, J. and Bruns, A. (2012) Twitter archives and the challenges of "big social data" for media and communication research', *M/C Journal*, Vol. 15, No. 5, pp.1–7.

Cao, K., Diao, M. and Wu, B. (2019) 'A big data–based geographically weighted regression model for public housing prices: a case study in Singapore', *Annals of the American Association of Geographers*, Vol. 109, No. 1, pp.173–186.

Cassel, E. and Mendelsohn, R. (1985) 'The choice of functional forms for hedonic price equations: comment', *Journal of Urban Economics*, Vol. 18, No. 2, pp.135–142.

Cervero, R. and Kang, C.D. (2011) 'Bus rapid transit impacts on land uses and land values in Seoul, Korea', *Transport Policy*, Vol. 18, No. 1, pp.102–116.

Chernobai, E., Reibel, M. and Carney, M. (2011) 'Nonlinear spatial and temporal effects of highway construction on house prices', *The Journal of Real Estate Finance and Economics*, Vol. 42, No. 3, pp.348–370.

Cohen, J.P. and Schaffner, S. (2019) *A New Highway in Germany and the Impacts on Real Estate Prices*, Ruhr Economic Papers, No. 821.

Crompton, J.L. (2001) *Parks and Economic Development*, APA Planning Advisory Service Reports N. 502, American Planning Association, Washington.

Dachuan, C. and Baoshan, Z. (2012) 'The applications of big data in housing information system', *Journal of Information and Communication Technologies*, Vol. 5, p.4.

Damigos, D. and Anyfantis, F. (2011) The value of view through the eyes of real estate experts: a fuzzy Delphi approach', *Landscape and Urban Planning*, Vol. 101, No. 2, pp.171–178.

Damm, D., Lerman, S.R., Lam, E.L. and Young, J. (1980) 'Response of urban real-estate values in anticipation of the Washington Metro', *Journal of Transport, Economics and Policy*, Vol. 14, No. 3, pp.315–336.

Debrezion, G., Pels, E. and Rietveld, P. (2007) 'The impact of railway stations on residential and commercial property value: a meta-analysis', *The Journal of Real Estate Finance and Economics*, Vol. 35, No. 2, pp.161–180.

Del Giudice, V., De Paola, P., Manganelli, B. and Forte, F. (2017) 'The monetary valuation of environmental externalities through the analysis of real estate prices', *Sustainability*, Vol. 9, No. 2, p.229.

Dewees, D.N. (1976) 'The effect of a subway on residential property values in Toronto', *Journal of Urban Economics*, Vol. 3, No. 4, pp.357–369.

Dijst, M. and Vidakovic, V. (2000) 'Travel time ratio: the key factor of spatial reach', *Transportation*, Vol. 27, No. 2, pp.179–199.

Dixon, T., Sexton, M. and van de Wetering, J. (2018) 'Smart cities, big data and the built environment: what's required?', *Property Journal*, Vol. 19, RICS Property Journal rics.org/research.

Du, D., Li, A. and Zhang, L. (2014) 'Survey on the applications of big data in Chinese Real Estate Enterprise', *Procedia Computer Science*, Vol. 30, pp.24–33.

Faishal Ibrahim, M., Jam Cheng, F. and How Eng, K. (2005) 'Automated valuation model: an application to the public housing resale market in Singapore', *Property Management*, Vol. 23, No. 5, pp.357–373.

Feige, A., Wallbaum, H., Janser, M. and Windlinger, L. (2013) 'Impact of sustainable office buildings on occupant's comfort and productivity', *Journal of Corporate Real Estate*, Vol. 15, No. 1, pp.7–34.

Fennema, A.T., Veeneklaas, F.R. and Vreke, J. (1996) 'Meerwaarde woningen door nabijheid van groen (Surplus value of dwellings in the vicinity of green areas)', *Stedebouw en Ruimtelijke Ordening*, Vol. 77, No. 3, pp.33–35.

Fuerst, F. and McAllister, P. (2011) 'Green noise or green value? Measuring the effects of environmental certification on office values', *Real Estate Economics*, Vol. 39, No. 1, pp.45–69.

Gelfand, A.E., Ghosh, S., Knight J.R. and Sirmans, C.F. (1998) 'Spatio-temporal modeling of residential sales data', *Journal of Business & Economic Statistics*, Vol. 16, No. 3, pp.312–321.

Gibler, K.M. Tyvimaa, T. and Kananen, J. (2014) 'The relationship between the determinants of rental housing satisfaction and considering moving in Finland', *Property Management*, Vol. 32, No. 2, pp.104–124.

Giddens, A. (1990) *The Consequences of Modernity*, Polity Press, Cambridge, USA.

Gillespie, T. (2013) 'The relevance of algorithms', in Gillespie, T., Boczkowski, P. and Foot, K. (Eds.): *Media Technologies*, MIT Press, Cambridge, USA.

Giustolisi, O. and Savic, D.A. (2006) 'A symbolic data-driven technique based on evolutionary polynomial regression', *Journal of Hydroinformatics*, Vol. 8, No. 3, pp.207–222.

Giustolisi, O. and Savic, D.A. (2009) 'Advances in data-driven analyses and modelling using EPR-MOGA', *Journal of Hydroinformatics*, Vol. 11, Nos. 3–4, pp.225–236.

Goodwin, P.B. (1978) 'Travel choice and time budgets', in Hensher, D.A. and Dalvi, Q. (Eds.): *Determinants of Travel Choice*, pp.358–381, Saxon House, Farnsborough.

Grether, D.M. and Mieszkowski, P. (1974) 'Determinants of real estate values', *Journal of Urban Economics*, Vol. 1, No. 2, pp.127–145.

Hass-Klau, C., Crampton, G. and Benjari, R. (2004) 'Economic impact of light rail: the results of 15 urban areas in France, Germany, UK and North America', *Environmental and Transport Planning*, Brighton, UK.

Janssen, C. and Yang, Z. (1999) 'Estimating the market value of a proposed townhouse development', *Journal of Property Investment & Finance*, Vol. 17, No. 5, pp.501–516.

Jim, C.Y. and Chen, W.Y. (2006) 'Impacts of urban environmental elements on residential housing prices in Guangzhou (China)', *Landscape Urban Planning*, Vol. 78, No. 4, pp.422–434.

Jim, C.Y. and Chen, W.Y. (2007) 'Consumption preferences and environmental externalities: a hedonic analysis of the housing market in Guangzhou', *Geoforum*, Vol. 38, No. 2, pp.414–431.

Juan, Y. (2013) 'The precise marketing based on big data in the real estate enterprises', *Journal of Market Weekly*, Vol. 9, pp.66–67.

Kim, K.S., Park, S.J. and Kweon, Y-J. (2007) 'Highway traffic noise effects on land price in an urban area', *Transp. Res. Part D: Transp. Environ.*, Vol. 12, No. 4, pp.275–280.

Koza, J.R. (1992) *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, Vol. 1, MIT Press, Cambridge, Massachusetts.

Kryvobokov, M. (2007) 'What location attributes are the most important for market value? Extraction of attributes from regression models', *Property Management*, Vol. 25, No. 3, pp.257–286.

Larsen, K. and Gilliland, J. (2008) 'Mapping the evolution of 'food deserts' in a Canadian city: supermarket accessibility in London, Ontario, 1961–2005', *International Journal of Health Geographics*, Vol. 7, No. 1, p.16.

Lavender, S.D. (1990) *Economics for Builders and Surveyors*, Longman, Essex, UK.

Levinson, D.M. and Kumar, A. (1994) 'The rational locator: why travel times have remained stable', *APA Journal*, Summer 1994, Vol. 60, No. 3, pp.319–332

Levkovich, O., Rouwendal, J. and Van Marwijk, R. (2016) 'The effects of highway development on housing prices', *Transportation*, Vol. 43, No. 2, pp.379–405.

Lohr, S. (2012) 'The age of big data', *New York Times*, 11 February.

Lotfi, S. and Koohsari, M.J. (2009) 'Measuring objective accessibility to neighborhood facilities in the city (A case study: Zone 6 in Tehran, Iran)', *Cities*, Vol. 26, No. 3, pp.133–140.

Lutzenhiser, M. and Netusil, N.R. (2001) 'The effect of open spaces on a home's sale price', *Contemporary Economic Policy*, Vol. 19, No. 3, pp.291–298.

Malpezzi, S., Chun, G.H. and Green, R.K. (1998) 'New place-to-place housing price indexes for U.S. metropolitan areas, and their determinants', *Real Estate Economics*, Vol. 26, No. 2, pp.235–274.

Martínez, L.M. and Viegas, J.M. (2009) 'Effects of transportation accessibility on residential property values: hedonic price model in the Lisbon, Portugal, metropolitan area', *Transportation Research Record*, Vol. 2115, No. 1, pp.127–137.

McCluskey, W. and Anand, S. (1999) 'The application of intelligent hybrid techniques for the mass appraisal of residential properties', *Journal of Property Investment & Finance*, Vol. 17, No. 3, pp.218–238.

McCluskey, W.J., Deddis, W.G., Lamont, I.G. and Borst, R.A. (2000) 'The application of surface generated interpolation models for the prediction of residential property values', *Journal of Property Investment & Finance*, Vol. 18, No. 2, pp.162–176.

Metzner, S. and Kindt, A. (2018) 'Determination of the parameters of automated valuation models for the hedonic property valuation of residential properties: a literature-based approach', *International Journal of Housing Markets and Analysis*, Vol. 11, No. 1, pp.73–100.

Morano, P. and Tajani, F. (2013) 'Bare ownership evaluation. Hedonic price model vs. artificial neural network', *International Journal of Business Intelligence and Data Mining*, Vol. 8, No. 8, pp.340–362.

Morano, P. and Tajani, F. (2014) 'Least median of squares regression and minimum volume ellipsoid estimator for outliers detection in housing appraisal', *International Journal of Business Intelligence and Data Mining*, Vol. 9, No. 2, pp.91–111.

Morano, P., Tajani, F. and Locurcio, M. (2018) 'Multicriteria analysis and genetic algorithms for mass appraisals in the Italian property market', *International Journal of Housing Markets and Analysis*, Vol. 11, No. 2, pp.229–262.

Morano, P., Tajani, F., Di Liddo, F. and Anelli, D. (2020) 'A feasibility analysis of the refurbishment investments in the Italian residential market', *Sustainability*, Vol. 12, No. 6, p.2503.

Morano, P., Tajani, F., Guarini, M. R., Di Liddo, F. and Anelli, D. (2019) 'A multivariate econometric analysis for the forecasting of the interdependences between the housing prices and the socio-economic factors in the City of Barcelona (Spain)', in *International Conference on Computational Science and Its Applications*, Springer, Cham, July, pp.13–22.

Nappi-Choulet, I. and Décamps, A. (2013) 'Capitalization of energy efficiency on corporate real estate portfolio value', *Journal of Corporate Real Estate*, Vol. 15, No. 1, pp.35–52.

Nelson, J.P. (1982) 'Highway noise and property values: a survey of recent evidence', *Journal of Transport Economics and Policy*, Vol. 16, No. 2, pp.117–138.

Netusil, N.R. (2013) 'Urban environmental amenities and property values: does ownership matter?', *Land Use Policy*, Vol. 31, pp.371–377.

Nishii, K. and Kondo, K. (1992) 'Trip linkages of urban railway commuters under time-space constraints: some empirical observations', *Transportation Research*, Vol. 26B, No. 1, pp.34–44.

Ossokina, I. and Verweij, G. (2011) *Quasi-Experimental Evidence on the Effect of Traffic Externalities on Housing Prices*, CPB Netherlands Bureau for Economic Policy Analysis, The Hague.

Oven, V.A. and Pekdemir, D. (2006) 'Office rent determinants utilizing factor analysis – a case study for Istanbul', *Journal of Real Estate Finance and Economics*, Vol. 33, No. 1, pp.51–73.

Ozus, E. (2009) 'Determinants of office rents in the Istanbul metropolitan area', *European Planning Studies*, Vol. 17, No. 4, pp.621–633.

Pagourtzi, E., Assimakopoulos, V., Hatzichristos, T. and French, N. (2003) 'Real estate appraisal: a review of valuation methods', *Journal of Property Investment & Finance*, Vol. 21, No. 4, pp.383–401.

Pérez-Rave, J.I., Correa-Morales, J.C. and González-Echavarría, F. (2019) 'A machine learning approach to big data regression analysis of real estate prices for inferential and predictive purposes', *Journal of Property Research*, Vol. 36, No. 1, pp.59–96.

Portnov, B.A., Bella, G. and Barzilay, B. (2009) 'Investing a timing the effect of train proximity on apartment prices: Haifa, Israel as a case study', *Journal of Real Estate Research*, Vol. 31, No. 4, pp.371–395.

Potepan, M.J. (1996) 'Explaining intermetropolitan variation in housing prices, rents and land prices', *Real Estate Economics*, Vol. 24, No. 2, pp.219–245.

Rabari, C. and Storper, M. (2015) 'The digital skin of cities: urban theory and research in the age of the sensored and metered city, ubiquitous computing and big data', *Cambridge Journal of Regions, Economy and Society*, Vol. 8, No. 1, pp.27–42.

Rabianski, J., DeLisle, J. and Carn, N. (2001) 'Corporate real estate site selection: a community-specific information framework', *Journal of Real Estate Research*, Vol. 22, Nos. 1–2, pp.165–198.

Robinson, R. (1979) *Housing Economics and Public Policy*, MacMillan, London, UK.

Robitaille, É. and Herjean, P. (2008) 'An analysis of the accessibility of video lottery terminals: the case of Montréal', *International Journal of Health Geographics*, Vol. 7, No. 1, p.2.

Rosasco, P., Sdino, L., Torrieri, F. and Oppio, A. (2018) 'Valutazione Multicriterio e Stime Di Massa: un'applicazione ad un Patrimonio Immobili', *Laborest*, Vol. 17, pp.33–38.

Schwanen, T., Dijst, M. and Dieleman, F.M. (2002) 'A microlevel analysis of residential context and travel time', *Environment and Planning A*, Vol. 34, No. 8, pp.1487–1507.

Selim, H. (2009) 'Determinants of house prices in Turkey: hedonic regression versus artificial neural network', *Expert Systems with Applications*, Vol. 36, No. 2, pp.2843–2852.

Sheppard, S. (1999) 'Hedonic analysis of housing markets', in Cheshire, P.C. and Mills, E.S. (Eds.): *Handbook of Applied Urban Economics*, pp.1595–1635, Elsevier, New York, USA.

Siwei, C. (2009) 'The fundamentals and research methods of fictitious economy', *Journal of Management Review*, Vol. 21, No. 1, pp.3–18.

Smith, L.B., Rosen, K.T. and Fallis, G. (1988) 'Recent development in economic models of housing markets', *Journal of Economic Literature*, Vol. 26, No. 1, pp.29–64.

Stumpf González, M.A. and Torres Formoso, C. (2006) 'Mass appraisal with genetic fuzzy rule-based systems', *Property Management*, Vol. 24, No. 1, pp.20–30.

Tajani, F., Morano, P., Locurcio, M. and Torre, C.M. (2016) 'Data-driven techniques for mass appraisals. Applications to the residential market of the city of Bari (Italy)', *International Journal of Business Intelligence and Data Mining*, Vol. 11, No. 2, pp.109–129.

Tajani, F., Morano, P., Torre, C.M. and Di Liddo F. (2017a) 'An analysis of the influence of property tax on housing prices in the Apulia Region (Italy)', *Buildings*, Vol. 7, Nos. 3, 67, pp.1–15.

Tajani, F., Ntalianis, K. and Di Liddo, F. (2017b) 'An assessment model for the periodic reviews of the market values of property assets', in *International Conference on Computational Science and Its Applications*, pp.490–500, Springer, Cham.

Tajima, K. (2003) 'New estimates of the demand for urban green space: implications for valuing the environmental benefits of Boston's big dig project', *Journal of Urban Affairs*, Vol. 25, No. 5, pp.641–655.

Taltavull de La Paz, P. (2003) 'Determinants of housing prices in Spanish cities', *Journal of Property Investment & Finance*, Vol. 21, No. 2, pp.109–135.

Thatcher, J. (2014) 'Big data, big questions| Living on fumes: Digital footprints, data fumes, and the limitations of spatial big data', *International Journal of Communication*, Vol. 8, pp.1765–1783.

Theebe, M.A.J. (2004) 'Planes, trains, and automobiles: the impact of traffic noise on house prices', *J. Real Estate Financ. Econ.*, Vol. 28, Nos. 2–3, pp.209–234.

Trojanek, R. and Gluszak, M. (2018) 'Spatial and time effect of subway on property prices', *Journal of Housing and the Built Environment*, Vol. 33, No. 2, pp.359–384.

Troy, A. and Grove, J.M. (2008) 'Property values, parks, and crime: a hedonic analysis in Baltimore', *Landscape and Urban Planning*, Vol. 87, No. 3, pp.233–245.

Tyrvainen, L. and Miettinen, A. (2000) 'Property prices and urban forest amenities', *Journal of Environmental Economics and Management*, Vol. 39, No. 2, pp.205–223.

Wilhelmsson, M. (2000) 'The impact of traffic noise on the values of single-family houses', *J. Environ. Plan. Manag.*, Vol. 43, No. 6, pp.799–815.

Yang, L., Wang, B., Zhou, J. and Wang, X. (2018) 'Walking accessibility and property prices', *Transportation Research Part D: Transport and Environment*, Vol. 62, pp.551–562.

Zahavi, Y. (1973) 'The TT-relationship: a unified approach to transportation planning', *Traffic Engineering and Control*, Vol. 15, No. 5, pp.205–212.

Zhang, X., Liu, X., Hang, J., Yao, D. and Shi, G. (2016) 'Do urban rail transit facilities affect housing prices? Evidence from China', *Sustainability*, Vol. 8, No. 4, p.380.