Property Valuations in Times of Crisis. Artificial Neural Networks and Evolutionary Algorithms in Comparison

Francesco Tajani[1], Pierluigi Morano[1(✉)], Marco Locurcio[2], and Nicola D'Addabbo[2]

[1] Department of Science of Civil Engineering and Architecture, Polytechnic of Bari, Bari, Italy
francescotajani@yahoo.it, pierluigi.morano@poliba.it
[2] Department of Architecture and Design, University "Sapienza", Rome, Italy
marco.locurcio@uniroma1.it, nda.archt@gmail.com

Abstract. In the current economic situation, characterized by a high uncertainty in the appraisal of property values, the need of "slender" models able to operate even on limited data, to automatically capture the causal relations between explanatory variables and selling prices and to predict property values in the short term, is increasingly widespread. In addition to Artificial Neural Networks (ANN), that satisfy these prerogatives, recently, in some fields of Civil Engineering an hybrid data-driven technique has been implemented, called Evolutionary Polynomial Regression (EPR), that combines the effectiveness of Genetic Programming with the advantage of classical numerical regression. In the present paper, ANN methods and the EPR procedure are compared for the construction of estimation models of real estate market values. With reference to a sample of residential apartments recently sold in a district of the city of Bari (Italy), two estimation models of market value are implemented, one based on ANN and another using EPR, in order to test the respective performance. The analysis has highlighted the preferability of the EPR model in terms of statistical accuracy, empirical verification of results obtained and reduction of the complexity of the mathematical expression.

Keywords: Property valuations · Artificial neural networks · Evolutionary polynomial regression · Genetic algorithms · Estimative analysis · Market value

1 Introduction

In current economic situation, the use of tools for the evaluation of real estate values has become essential for sector operators (buyers, sellers, institutions, insurance companies, banks, etc.) [6, 17]. The continuous change of the boundary conditions [4, 31, 32, 33] causes that it is necessary to use, rather than models characterized by a strong theoretical and methodological basis, "slender" models, able to operate even on limited data and to automatically capture the causal relations between explanatory variables and prices, as well as to predict property values in the short term [30].

The work must be attributed in equal parts to the authors.

O. Gervasi et al. (Eds.): ICCSA 2015, Part III, LNCS 9157, pp. 194–209, 2015.
DOI: 10.1007/978-3-319-21470-2_14

The artificial neural networks (ANN) satisfy these prerogatives. Many studies [3, 12, 27, 28, 29, 34, 37] have highlighted that: ANN provide very good performance in forecasting market values, even when the data are limited; they avoid the econometric problems linked to the multicollinearity, the heteroskedasticity and the spatial autocorrelation, that are typical of other models (e.g. hedonic prices); they are more robust to model misspecification regarding how explanatory variables are measured.

However, ANN models have several weaknesses [9, 14, 21]. First of all, they provide that the structure of the neural network (e.g. model inputs, transfer functions, number of hidden layers, etc.) is exogenously defined. Furthermore, over-fitting problems are frequent in parameter estimation. Another disadvantage is the inability to incorporate known economic laws into the learning process.

On these aspects, some steps have been made through the Genetic Programming (GP), that is an approach of artificial intelligence capable of generating a structured representation of the system model. The most common method of GP, named symbolic regression, allows to obtain mathematical expressions to fit a set of data points using operations analogous to the evolutionary processes that occur in nature [23].

Several authors have borrowed the logic of genetic algorithms to improve the application of mathematical procedures to the real estate market. Among them, Kròl et al. [24] have developed a fuzzy rule-based system to assist the real estate appraisal, employing an evolutionary algorithm to generate the rule base. Wang [36] has elaborated a decision support system that, through a data envelopment analysis, converts numerical data into information that can be used to evaluate real estate investments. Dzeng e Lee [10] have proposed a model to optimize the development schedule of resort projects using a polyploidy genetic algorithm.

Recently, in some fields of Civil Engineering an hybrid data-driven technique has been implemented, called Evolutionary Polynomial Regression (EPR), that combines the effectiveness of Genetic Programming with the advantage of classical numerical regression [2, 22, 25]. An advanced version of this method, called EPR-MOGA, uses Multi-Objective Genetic Algorithms to search those model expressions that simultaneously maximize accuracy of data and parsimony of mathematical functions. This approach generates a set of explicit expressions with different accuracy to experimental data and different degree of complexity of the structural models. The analysis of the expressions generated allows to select the solution which corresponds to the best compromise in terms of accuracy and complexity and better suited for specific applications.

In the present paper, ANN methods and the evolutionary approach based on EPR-MOGA are compared for the construction of estimation models of real estate market values. On the same database and with reference to the same explanatory variables of the unit price of residential apartments, two estimation models are implemented, one based on ANN and another using EPR, in order to test the respective performance.

The paper is structured as follows. In section 2, notes on ANN and on EPR-MOGA theory are reported. In section 3 the case study is introduced: it is relative to a sample of residential apartments recently sold in a segment of the real estate market of the city of Bari (Italy). In section 4 the ANN model and the EPR-MOGA model are specified, the calculations are carried out and the results are compared and then illustrated. In section 5 the conclusions of the work are discussed.

2 Outline of ANN Models and EPR Models

ANN are complex systems [26], formed by a set of elementary unit, the neurons, combined in an opportune way in a netting structure made of *layers* presenting an elevated degree of interconnection.

The complexity of the structure of a neural network depends on the number of neurons and the number of existing connections. Neurons can be classified on the basis of the level they occupy in the network. The first level, called the *input layer*, is formed by neurons which contain the exogenous information, translated in terms of the pulse for the neurons of the upper level. Opposite to the input layer is the *output layer*, whose neurons return the result generated by the operation of the network. Between these two levels it is possible to predict one or more intermediate levels of hidden neurons or *hidden layers*, which are entrusted with the task of developing the information coming from the input layer and translating them in the output.

In ANN the training is the responsibility of the only variable part of the structure, i.e. the weights of the connections. By altering the weights of the connections through a learning rule, the neural network is able to learn a distinctive function from couples of examples input/output (*training set*) that are repeatedly presented to it.

The transmission of information between the neurons takes place through an activation function, associated to each neuron and almost always common to all the nodes of the model.

For connections between the neurons is generally adopted the hierarchical structure, in which the connections are present only between neurons of two successive levels, and the pulses of the neurons are direct (one way) from the input layer to the output layer. In this model, called *feed-forward*, the input of the neurons that are located at a level higher than the first is given by the set of all the signals from the neurons of the lower level. Indicating with w_{ij} the weight associated to the connection between the neuron i and the neuron j, the input I_j of the neuron j can be defined as follows:

$$I_j = \sum_{i=1}^{N} O_i \cdot w_{ij} \tag{1}$$

where N is the number of neurons present in the level lower than the level in which the neuron j is located, O_i is the output of neuron i, which, through the activation function f, can be expressed as:

$$O_i = f(I_i) \tag{2}$$

So that the neural network is able to learn a particular task assigned, it is necessary to transfer to the system a learning technique, that is a rule by which it is possible to appropriately update the weights of the connections of the network.

There are many learning techniques, each of them suited to achieving the objective. The learning rule more frequently implemented is the *back-propagation* method. This rule allows to modify iteratively the weights of the connections of the network in function of the error, i.e. the difference from time to time found between the actual output of the network and the target value.

The method of EPR can be considered as a generalization of the original stepwise regression, that is linear with respect to regression parameters, but it is non-linear in the model structures. The following equation summarizes a generic non-linear model structure that can be implemented in EPR:

$$Y = a_0 + \sum_{i=1}^{n} [a_i \cdot (X_1)^{(i,1)} \cdot \ldots \cdot (X_j)^{(i,j)} \cdot f((X_1)^{(i,j+1)} \cdot \ldots \cdot (X_j)^{(i,2j)})] \qquad (3)$$

where n is the number of additive terms, a_i are numerical parameters to be valued, X_i are candidate explanatory variables, *(i, l)* - with $l = (1, ..., 2j)$ - is the exponent of the l-th input within the i-th term in Eq. (3), f is a function selected by the user among a set of possible mathematical expressions. The exponents *(i, l)* are also selected by the user from a set of candidate values (real numbers).

The iterative investigation of model mathematical structures, implemented by exploring the combinations of exponents to be attributed to each candidate input of Eq. (3), is performed through a population based strategy that employs a Genetic Algorithm, whose individuals are constituted by the sets of exponents in Eq. (3) and chosen by the user.

The algorithm underlying EPR does not require the exogenous definition of the mathematical expression and the number of parameters that fit better the data collected, since it is the iterative process of the genetic algorithm that returns the best solution.

The accuracy of each equation returned by EPR is checked through its Coefficient of Determination (COD), defined as:

$$COD = 1 - \frac{N-1}{N} \cdot \frac{\sum_N (y_{EPR} - y_{detected})^2}{\sum_N (y_{detected} - mean(y_{detected}))^2}, \qquad (4)$$

where y_{EPR} are the values of the dependent variable estimated by the EPR model, $y_{detected}$ are the collected values of the dependent variable, N is the sample size in analysis [15]. The fitting of a model is greater when the COD is close to the unit value.

A recent version of EPR, called EPR-MOGA [16], reproduces an evolutionary multi-objective genetic algorithm, as optimization strategy based on the Pareto dominance criterion. These objectives are conflictual, and aim at i) the maximization of model accuracy, through the satisfaction of appropriate statistical criteria of verification of the equation; ii) the maximization of model's parsimony, through the minimization of the number of terms (a_i) of the equation; iii) the reduction of the complexity of the model, through the minimization of the number of the explanatory variables (X_i) of the final equation. Through the use of a Microsoft Office Excel add-in function (EPR MOGA-XL v.1, freely available at site: *www.hydroinformatics.it*), the optimization strategy defined above, based on the Pareto dominance criteria, allows to obtain, at the end of the modeling phase, a set of model solutions (i.e. the Pareto front of optimal models) for the three objectives considered. In this way, a range of solutions is offered to the operator, among which it is possible to select the most appropriate solution according to the specific needs, the knowledge of the phenomenon in analysis and the type of experimental data used.

3 The Case Study

With the help of estate agents operating on site, an estimative sample of 90 residential properties sold in 2013-2014 in the Madonnella district of the city of Bari (Italy) has been collected. The boundary of the district has been defined so as to coincide with the relative "Microzone". The Italian real estate agents consider a geographical segmentation of the market in Microzones, defined according to the Presidential Decree 138/1998 and ensuing Regulation issued by the Ministry of Finance. For the Italian regulation, the “Microzone” is a part of the urban area that must be urbanistically homogeneous and at the same time must constitute a homogeneous real estate market segment. A Microzone, in other words, is an area of the real estate market in which extrinsic factors (accessibility, presence of services, building characteristics, green areas, pedestrian zones, etc.), involved in the formation of real estate values, evolve in a substantially uniform manner [11].

The Madonnella district is a central area of Bari, characterized by numerous Liberty style buildings realized in the late nineteenth century and several public buildings of cultural value made during the Fascist era. It is predominantly a residential district, with a population of about 18,000 inhabitants. The district is next to the Nazario Sauro waterfront and near the historical centre of the city (San Nicola district). It is quite accessible thanks to several bus lines.

With the help of estate agents operating in the district, the following information have been obtained for each housing unit: the *unit selling price* (*PRZ*), in euro per square meter of floor area of the property; the *floor* (*F*) the apartment is on; the *number of bathrooms* (*B*) of the property; the *panoramic view* (*P*) of the apartment, taken as a qualitative variable and differentiated, with a synthetic evaluation, by the categories “none”, “enough” and “good”. In the model, for this explanatory variable two dummies have been considered, respectively for the state “enough” (P_e) and “good” (P_g); the presence of *independent heating* (*H*), expressed with a dichotomous criterion, with 1 if the heating is independent in the apartment and 0 if the heating is centralized; the distance from the *center* (*C*) of the city, expressed in minutes it takes to walk to it; the *rental situation* of the apartment (*R*), expressed through a dichotomous criterion, with 1 if the apartment is rented and 0 otherwise; the *surface* (*S*) of the apartment, expressed in square meters of floor area of the property.

Other common intrinsic characteristics have been excluded from the analysis due to the relative mechanism of appreciation of the market value in the segment investigated, identifying equal conditions between the building units of the sample and not contributing to the explanation of the market value.

Table 1 provides, for the sample collected in Madonnella district, statistics for quantitative variables: continuous (unit selling price, distance from center, surface), discrete (floor the apartment is on, number of bathrooms) and dummies (panoramic view, presence of independent heating, rental situation).

Both in the ANN model as in the EPR-MOGA developed in this work, the logarithm of the unit selling price (*PRZ*) identifies the dependent variable, whereas the other parameters collected (F, B, P_e, P_g, H, C, R, S) are the explanatory variables.

3.1 The ANN Model

To specify the ANN model it is necessary to define the network topology (number of hidden layers and the neurons in each of them), the propagation rule, the activation function and the learning rule.

Here the Multi-Layer Perceptron network is employed, with one hidden layer which includes thirteen nodes, an input layer with the eight exogenous variables defined, and an output layer with the natural logarithmic of real unit prices.

A fully connected feed forward network is assumed, which means activation travels in a direction from the input layer to the output layer, and the units in one layer are connected to every other unit in the next layer up.

In the input layer and in the output layer an activation function of sigmoidal type is associated to each neuron, which is modeled continuously between 0 and 1 using the following analytical expression:

$$f(x) = \frac{1}{1+e^{-kx}} \tag{5}$$

where x is the input and k is the slope of the tangent to the curve at the inflection point.

Several alternative topologies have been tried, with two and three hidden layers, and different number of neurons and activation functions. However the best results have been obtained with the ANN model that will be presented.

The model is implemented through the software "BKP – Neural Network Simulator" [1], that employs the algorithm of Back-Propagation (BKP) to adjust iteratively the connection weights.

In accordance to standard analytical practice, the estimative sample has been divided in a random basis into two sets, the "training set" and the "test set". The training set includes 80% of the sample, corresponding to 72 transactions, leaving the remaining 18 cases as the test set.

In the software employed, a random starting point has been used, with the following values for the main training parameters: slope k term = 1; learning rate = 0.65; momentum term = 0.1; maximum number of iteration = 25,000.

The Root Mean Squared Error (RMSE) has been the error function selected. A value of RMSE equal to 1.2623 is related to the model defined. The determination index (R^2) is equal to 0.9932. The Mean Absolute Percentage Errors (MAPE), that is the average percentage error between the prices of the original sample and the values estimated with ANN, is equal to 3.9155. The Maximum Absolute Percentage Errors (MaxAPE), that is the maximum percentage error between the prices of the original sample and the values estimated with the ANN model, is 10.0744.

Table 1. Sample descriptive statistics

Variable	Mean	Standard Deviation	Levels/Intervals	Frequency
Unit selling price [€/m^2]	2,764.30	433.80		
Floor [n.]	2.97	1.23		
			0	0.02
			1	0.12
			2	0.14
			3	0.39
			4	0.23
			5	0.10
Number of bathrooms [n.]	1.66	0.56		
			1	0.38
			2	0.57
			3	0.05
Panoramic view			none	0.22
			enough	0.23
			good	0.55
Presence of independent heating (1-independent, 0-centralized)	0.42	0.51		
			0	0.58
			1	0.42
Distance from the center [min. walking]	9.36	4.15		
			<5	0.11
			6-10	0.55
			11-15	0.07
			16-20	0.14
Rental situation (1-rented, 0-available)	0.16	0.36		
			0	0.85
			1	0.15
Floor surface [m^2]	88.92	34.87		
			<50	0.11
			51-70	0.23
			71-90	0.29
			91-110	0.08
			>110	0.29

Sensitivity analysis (Table 2) allows the evaluation of the influence of each exogenous variable using its error ratio, obtained as the RMSE of the model without the explanatory variables in analysis compared to the RMSE of the model including all the variables.

Table 2. Sensitivity analysis for the ANN model

VARIABLE	*RATIO*	*ORDER OF IMPORTANCE*
C	2.016161	1
H	1.596292	2
F	1.509742	3
P_g	1.487650	4
R	1.354432	5
S	1.349758	6
P_e	1.298589	7
B	1.222877	8

It is noted that with ANN the importance of the variable, in descending order, is as follows: the *distance from the center* (C), the presence of the *independent heating* (H), the *floor* (F), a "good" *panoramic view* (P_g), a "sufficient" *panoramic view* (P_e), the *rental situation* (R), the *surface* (S) of the apartment and finally the *number of bathrooms* (B) of the property.

3.2 The EPR Model

The base model structure reported in Eq. (3) is used with no function f selected, whereas each additive monomial term is assumed to be a combination of the inputs (i.e. the explanatory variables) raised to the proper exponents., Candidate exponents belong to the set (0; -1; 1), in order to facilitate the interpretation of the results generated and to represent the direct/inverse relationship between candidate inputs and the dependent variable. The maximum number n of additive terms in final expressions is assumed to be 8, that is equal to the number of explanatory variables considered.

With these conditions, the run of the EPR-MOGA software has returned the twelve models (M_i) reported in Table 3, characterized by different number of additive terms and explanatory variables, as well as different accuracy in terms of COD (Figure 1).

The examination of the models shows that starting from the model *M4*, the mathematical expressions generated by the software are complicated, since the same explanatory variables appear in more terms of the model, also combined with other variables. This circumstance, if it leads to functions that can reproduce better prices, on the other hand generates complicated expressions that make it difficult to interpret the phenomenon.

Table 3. Equations obtained by the implementation of EPR-MOGA

M1	$ln(PRZ) = 8.2664 - 0.0379\,C$	(6)
M2	$ln(PRZ) = 8.2046 - 0.0339C + 0.0574H$	(7)
M3	$ln(PRZ) = 8.1536 - 0.03159C + 0.0563H + 0.01F$	(8)
M4	$ln(PRZ) = 8.1905 - 0.0325C + 0.3561\dfrac{H}{C}$	(9)
M5	$ln(PRZ) = 8.1641 - 0.0295\,C + 0.0711\,H + 0.4403\,R - 0.0284\,C\,R$	(10)
M6	$ln(PRZ) = 8.1139 - 0.0243\,C + 0.4705\,R - 0.0322\,C\,R + 0.5132\,\dfrac{H}{C}$	(11)
M7	$\ln(PRZ) = 8.0293 - 0.0258\,C + 0.3784\,R + 0.0344\,B - 0.0247\,C\,R + 0.5147\,\dfrac{H}{C} + 0.0022\,\dfrac{F\,C}{B}$	(12)
M8	$\ln(PRZ) = 8.1398 - 0.0266\,C + 0.5637\,R + 0.1025\,\dfrac{1}{B} - 0.0314\,C\,R + 0.545\,\dfrac{H}{C} + 0.0034\,\dfrac{F\,C}{B} - 0.0508\,F\,R$	(13)
M9	$\ln(PRZ) = 8.096 - 0.0299\,C - 0.3860\,R + 0.0232\,B + 0.5324\,\dfrac{H}{C} + 0.0039\,\dfrac{F\,C}{B} + 5.9505\,\dfrac{R}{C} - 0.02\,\dfrac{F}{B}$	(14)
M10	$\ln(PRZ) = 8.0916 - 0.0234\,C + 0.5066\,R + 0.0456\,H - 0.0341\,C\,R + 0.2252\,\dfrac{B\,H}{C} - 3.7734\,\dfrac{P_g\,H}{S} + 1.2658\,\dfrac{F\,P_g}{B\,S}$	(15)
M11	$\ln(PRZ) = 8.1135 - 0.0249\,C + 0.4646\,R - 0.0312\,C\,R + 0.2671\,\dfrac{B\,H}{C} - 2.4002\,\dfrac{P_g\,H}{S} + 1.1754\,\dfrac{F\,P_g}{B\,S}$	(16)
M12	$\ln(PRZ) = 8.1041 - 0.0244\,C + 0.475\,R - 0.0318\,C\,R + 0.3073\,\dfrac{B\,H}{C} + 1.5295\,\dfrac{F\,P_g}{B\,S} - 0.9643\,\dfrac{F\,P_g\,H}{S}$	(17)

The COD relative to the models obtained through EPR-MOGA is always close to unity, ranging from the minimum value of 91.43%, corresponding to the model *M1*, to the maximum value of 96.29%, which is obtained for the model *M10* (Fig. 1). However, after the model *M3*, characterized by a COD equal to 93.69%, the increase of accuracy becomes relatively significant. In particular, for the models *M4*, *M9* and *M11*, in addition to the reduction in COD, the evident increase of the complexity of the mathematical function, compared to the model that immediately precedes them, should be highlighted.

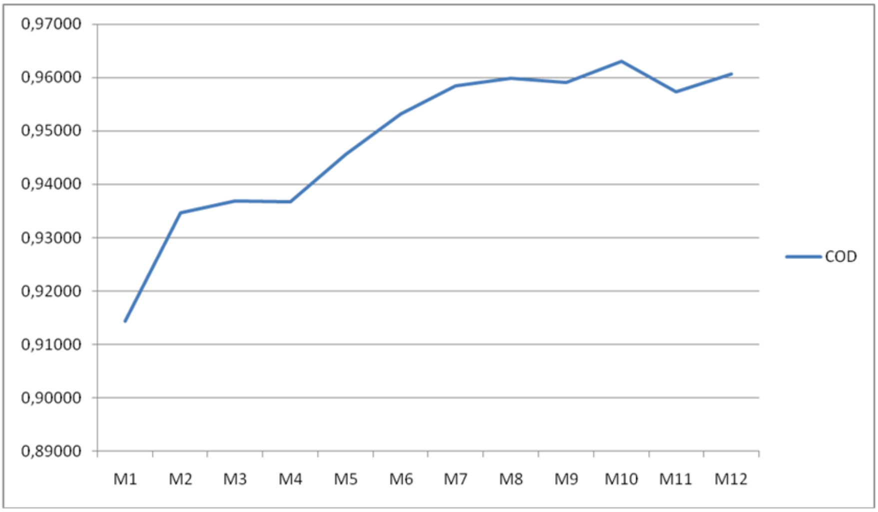

Fig. 1. Accuracy of the models (M_i) in terms of Coefficient of Determination (COD)

These considerations suggest to focus only on the first three models generated by the software.

The evolution of the mathematical equations obtained from the model *M1* to the model *M3*, provides some important information: the first concerns the importance of the main exogenous variables in the explanation of the observed prices, which can be deduced by the sequence in the appearance of the variables in the first three models: the *distance from the center* (*C*), the presence of the *independent heating* (*H*), the *floor* (*F*); the second information is related to the empirical coherence of the signs of the coefficients of the independent variables, which make explicit the inverse proportionality between the unit price (*PRZ*) and the variable *distance from the center* (*C*) and, vice versa, the direct proportionality with the variables presence of the *independent heating* (*H*) and *floor* (*F*); the third information is that the combination of the variables *distance from the center* (*C*), presence of the *independent heating* (*H*) and *floor* (*F*) already provides a good explanation of the prices. The inclusion of the other variables in the equations of the models subsequent to *M3* would generate too complex expressions, unusable for the interpretation of the phenomenon.

Therefore, *M3* is the model that is effectively able to explain and reproduce the mechanism of formation of the prices of the sample analyzed. For this model, a value of RMSE equal to 1.0723 is obtained; the determination index (R^2) is equal to 0.9998; the MAPE is equal to 3.3172; the MaxAPE is 8.8689. It should be added that the model *M3* allows to calculate the *hedonic prices* of the explanatory variables (Table 4). In fact, in a log-linear model for a non-dichotomous variable the coefficient multiplied by 100 directly provides the percentage change of the dependent variable caused by the change in the explanatory variable; whereas, for dichotomous variables, the percentage effect on the dependent variable can be obtained by the relation $\Delta p_{boj} = 100 \cdot (e^{\beta j} - 1)$, where βj is the coefficient of the *j*-th dummy variable [13, 20].

Table 4. Percentage effects of the explanatory variables selected by the model *M3* on the unit prices (*PRZ*)

Variable	Coefficient	Percentage effect [%]
distance from the center (C)	-0.03159	-3.159
indipendent heating (H)	0.0563	5.792
floor (F)	0.01	1.00

Table 4 shows the following results: each additional minute that it takes to walk to the *center* determines a decrease in the unit price equal to -3.159%. The incidence of this reduction ranges from 12.636% to 63.18%, respectively for travel time walk in about 4 minutes and about 20 minutes; the presence of the *independent heating* produces an increase of the unit price equal to 5.792%; every additional *floor* level originates an increase of the price equal to 1.00%. The incidence of this raise ranges from 0% to 5.00%, respectively for the ground floor and the fifth floor level.

3.3 Comparison Between the ANN and the EPR-MOGA Models and Interpretation of Results

The performance of ANN and EPR-MOGA models, computed in terms of R^2, RMSE, MAPE and MaxAPE, are summarized in Table 5 and Fig. 2. The indicators selected and the Graph in Figure 3 show the best accuracy of EPR-MOGA *M3* model for the evaluation of property prices relative to Madonnella district of Bari.

The comparison between the 90 detected prices and the corresponding values determined by the ANN model and the EPR-MOGA *M3* model is graphically shown in Fig. 3. The correspondence is excellent in both cases.

It is interesting to note that the EPR-MOGA model selected (*M3*) confirms the order returned by the ANN model related to the importance of the explanatory variables, at least for the positions of the “podium” (Table 2). The analysis confirms that in the Madonnella district the location (variable *C*) is the feature with the highest influence on the appreciation of a residential property. In this regard, it should be noted that the importance of this variable is enhanced by the absence of fast connections in the Madonnella district (e.g. subway or light railway) with the center of Bari.

The presence of the independent heating (*H*) obtains the second place. The relevance of this variable expresses, on the one hand, the appreciation of buyers for more home comfort determined by the independent heating; on the other hand, the possibility to avoid disagreements that often arise in apartment blocks about the heating: indeed, although the Italian Condominium Act (L. 220/2012) has simplified the procedures for the separation from the central heating system - provided that the intervention does not entail any additional expenses and/or inconvenience to the operation of the central heating system - other regulations (D.P.R. 59/2009; D.L. 192/2005) require that the owner of the apartment who wants to perform the separation from the central heating system must verify the energy saving capacity of the independent heating to be realized. The legal disputes that may arise from the failure to comply these regulations are often a disincentive to the conversion of the central heating system in an independent heating of an apartment.

The significant impact of the floor (F) the apartment is on, reflects the appreciation that is normally attributed to this feature when the building in which the apartment is located has a lift, which happens to all the apartments of the sample collected.

The other variables of the initial database (P_g, R, S, P_e, B) participate in the mechanism of price formation in the ANN model, but they are not included in the EPR-MOGA *M3* model.

Furthermore, both models allow to overcome the effects of collinearity that might occur with other procedures (e.g. hedonic pricing method), that, for example, can involve the variables *floor* and *panoramic view*. In particular, the EPR-MOGA recognizes the *floor* level as a "proxy" variable of the *panoramic view* in all models generated: indeed, only the variable "good" *panoramic view* (P_g) appears in the last three models (*M10*, *M11* and *M12*), characterized by mathematical expressions considerably complex, in which the variables F and P_g end up to being combined in a unique additive term.

Table 5. Comparison of the performance of ANN and EPR-MOGA M_3 models

Performance measures	*ANN model*	*EPR model*
R^2	0.9932	0.9998
RMSE	1.2623	1.0723
MAPE	3.9155	3.3172
MaxAPE	10.0744	8.8689

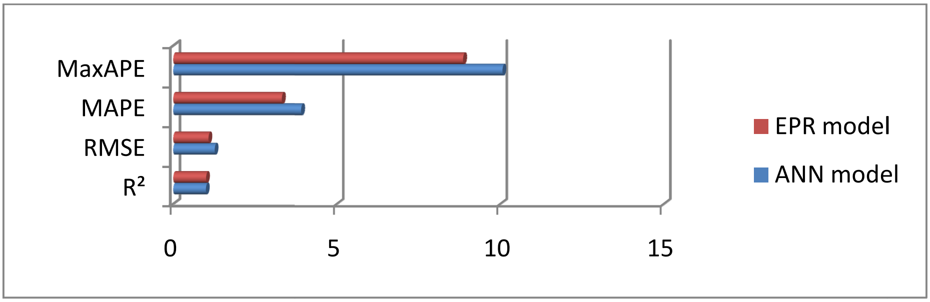

Fig. 2. Graphical comparison of the performance of ANN and EPR-MOGA M_3 models

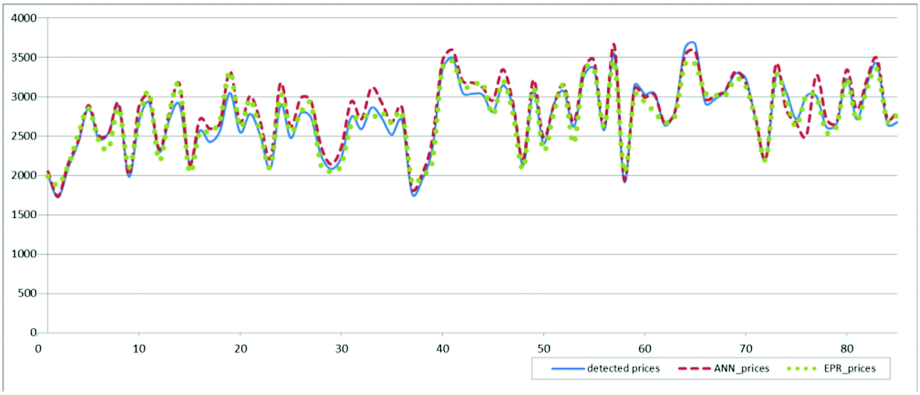

Fig. 3. Comparison between the detected unit prices (continuous line), the unit prices estimated with the ANN model (broken line) and the EPR-MOGA M_3 model (dashed line)

4 Conclusions

In the phase of uncertainty that is characterizing the Italian real estate sector, the use of innovative assessment tools may allow market operators to formulate more reliable estimates, as well as to effectively monitor the evolution of property values [5, 7, 8, 18, 19, 35].

In this paper, two methods have been implemented on the same database and with reference to the same explanatory variables of market prices of residential apartments recently sold in a district of the city of Bari (Italy): one based on ANN theory, the other one using EPR-MOGA procedure and selecting the most appropriate equation in terms of accuracy and interpretability of the results.

Both the models tested have shown excellent performance, but the EPR-MOGA M_3 model allows to obtain simultaneously the best statistical accuracy in the prediction of the market prices, a quick check of the empirical consistency of results obtained as well as the overcoming of the main limitations of ANN models. In fact, ANN are “black boxes”, i.e. it does not allow to generate a straightforward functional relationship between the input and the output values nor punctually investigate and reproduce the mechanisms of the prices formation. Furthermore, it can happen that the results obtained through ANN could not be stable, but could improve with increasing sample size, as well as results from models prepared with the same data but generated by different software packages could be different. The EPR model, instead, overcomes these shortcomings: the transparency of the mathematical expression obtained allows to verify (and quantify) the significance of the explanatory variables in the formation of property prices.

The EPR-MOGA *M3* model selected in this work is configured as a semi-logarithmic equation obtainable from the application of hedonic pricing method, considering as independent variables the only three characteristics identified by the EPR-MOGA *M3* model as the most representative of the property prices. The advantage of the implementation of the genetic algorithm underlying the logic of

EPR-MOGA is the ability to automatically and quickly select the optimal regressive functions in terms of accuracy and understanding. This result can be hardly obtained through the use of traditional hedonic pricing method. Therefore, this is an important added value in the real estate appraisal, in which it is essential to have forecasting functions characterized by high performance and at the same time easy to interpret.

The EPR-MOGA model identified in this work could have interesting applications: indeed, expressing property prices as a function of only three explanatory variables (the distance from the center, the presence of the independent heating and the floor level) and of a constant additive term, it is likely to be a simple and effective tool in the reform of Italian Cadastre (L. 23/2014) for the assignment of a mass appraisal function to properties located in the same territorial microzone and whose, in this case, the location and two intrinsic characteristics, easily acquirable, should be taken over.

Acknowledgements. This study has been developed within research activities being carried out by Real Estate Valuation Center of the MITO-LAB (LABoratory of Multimedia Information for Territorial Objects) of the Polytechnic of Bari, Italy (web site: http://mitolab.poliba.it).

References

1. Barile, S., Magna, L., Marsella, M., Miranda, S.: A marketing decision problem solved by application of neural networks. In: International Conference on Computational Intelligence and Multimedia Applications (1999)
2. Berardi, L., Kapelan, Z., Giustolisi, O., Savic, D.: Development of pipe deterioration models for water distribution systems using EPR. Journal of Hydroinformatics **10**(2), 113–126 (2008)
3. Brunson, A.L., Buttimer, R.J., Rutherford, R.C.: Neural Networks, Nonlinear Specification and Industrial Property Values. Working Paper Series, pp. 94–102. University of Texas at Arlington (1994)
4. Calabrò, F., Della Spina, L.: The cultural and environmental resources for sustainable development of rural areas in economically disadvantaged contexts. Economic-appraisals issues of a model of management for the valorisation of public assets. Advanced Materials Research **869–870**, 43–48 (2014)
5. Calabrò, F., Della Spina, L.D.: The public-private partnerships in buildings regeneration: A model appraisal of the benefits and for land value capture. Advanced Material Research **931–932**, 555–559 (2014)
6. D'Alpaos, C., Canesi, R.: Risks assessment in real estate investments in times of global crisis. WSEAS Transactions on Business and Economics **11**, 369–379 (2014)
7. Del Giudice, V., De Paola, P.: Geoadditive models for property market. Applied Mechanics and Materials **584**, 2505–2509 (2014)
8. Del Giudice, V., De Paola, P.: The effects of noise pollution produced by road traffic of Naples Beltway on residential real estate values. Applied Mechanics and Materials **587**, 2176–2182 (2014)
9. Do, A.Q., Grudnitski, G.: A neural network analysis of the effect of age on housing values. Journal of Real Estate Research, American Real Estate Society **8**(2), 253–264 (1993)

10. Dzeng, R.J., Lee, H.Y.: Optimizing the development schedule of resort projects by integrating simulation and genetic algorithm. International Journal of Project Management **25**(5), 506–516 (2007)
11. Forte, F.: Costs of noise and Italian urban policies. In: 36th International Congress and Exhibition on Noise Control Engineering, vol. 6, pp. 3991–3999. Istanbul, Turkey (2007)
12. Gallego, J.: La inteligencia artificial aplicada a la valoraciòn de inmuebles. Un ejemplo para valorar Madrid. Revista CT/Catastro **50**, 51–67 (2004)
13. Giles, D.E.: Interpreting Dummy Variables in Semi-logarithmic Regression Models: Exact Distributional Results. Econometrics Working Paper EWP1101, University of Victoria, Canada (2011)
14. Giustolisi, O., Laucelli, D.: Increasing generalisation of input-output artificial neural networks in rainfall-runoff modelling. Hydrological Sciences Journal **3**(50), 439–457 (2005)
15. Giustolisi, O., Savic, D.: A symbolic data-driven technique based on evolutionary polynomial regression. Journal of Hydroinformatics **8**(3), 207–222 (2006)
16. Giustolisi, O., Savic, D.: Advances in data-driven analyses and modelling using EPR-MOGA. Journal of Hydroinformatics **11**(3–4), 225–236 (2009)
17. Guarini, M.R., Battisti, F.: Social Housing and Redevelopment of Building Complexes on Brownfield Sites: The Financial Sustainability of Residential Projects for Vulnerable Social Groups. Advanced Materials Research **869–870**, 3–13 (2014)
18. Guarnaccia, C., Quartieri, J., Mastorakis, N.E., Tepedino, C.: Development and Application of a Time Series Predictive Model to Acoustical Noise Levels. WSEAS Transactions on Systems **13**, 745–756 (2014)
19. Guarnaccia, C.: Advanced Tools for Traffic Noise Modelling and Prediction. WSEAS Transactions on Systems **12**(2), 121–130 (2013)
20. Halvorsen, R., Palmquist, R.: The interpretation of dummy variables in semilogarithmic regressions. American Economic Review **70**, 474–475 (1980)
21. Islam, K.S., Asam, Y.: Housing market segmentation: a review. Review of Urban & Regional Development Studies **21**(2–3), 93–109 (2009)
22. Javadi, A.A., Rezania, M.: Applications of artificial intelligence and data mining techniques in soil modeling. Geomechanics and Engineering **1**(1), 53–74 (2009)
23. Koza, J.R.: Genetic programming: on the programming of computers by natural selection. MIT Press, Cambridge (1992)
24. Król, D., Lasota, T., Trawiński, B., Trawiński, K.: Investigation of evolutionary optimization methods of TSK fuzzy model for real estate appraisal. International Journal of Hybrid Intelligent Systems **5**(3), 111–128 (2008)
25. Laucelli, D., Giustolisi, O.: Scour depth modelling by a multi-objective evolutionary paradigm. Environmental Modelling & Software **26**(4), 498–509 (2011)
26. Limsombunchai, V., Gan, C., Lee, M.: House price prediction: hedonic price model vs. artificial neural network. American Journal of Applied Sciences **1**(3), 193–201 (2004)
27. Liu, J.-G., Zhang, X.-L., Wu, W.-P.: Application of fuzzy neural network for real estate prediction. In: Wang, J., Yi, Z., Żurada, J.M., Lu, B.-L., Yin, H. (eds.) ISNN 2006. LNCS, vol. 3973, pp. 1187–1191. Springer, Heidelberg (2006)
28. McCluskey, W.J., Dyson, K., McFall, D., Anand, S.: The mass appraisal of residential property in Northern Ireland. In: Computer Assisted Mass Appraisal: An International review, pp. 59–77. Ashgate Publishing Limited, England (1997)
29. Morano, P., Tajani, F.: Bare ownership evaluation. Hedonic price model vs. artificial neural network. International Journal of Business Intelligence and Data Mining **8**(4), 340–362 (2013)

30. Morano, P., Tajani, F.: Least median of squares regression and minimum volume ellipsoid estimator for outliers detection in housing appraisal. International Journal of Business Intelligence and Data Mining **9**(2), 91–111 (2014)
31. Oppio, A., Corsi, S., Mattia, S., Tosini, A.: Exploring the relationship among local conflicts and territorial vulnerability: The case study of Lombardy Region. Land Use Policy **43**, 239–247 (2015)
32. Scorza, F., Casas, G.L., Murgante, B.: Overcoming interoperability weaknesses in e-government processes: organizing and sharing knowledge in regional development programs using ontologies. In: Lytras, M.D., Ordonez de Pablos, P., Ziderman, A., Roulstone, A., Maurer, H., Imber, J.B. (eds.) WSKS 2010. CCIS, vol. 112, pp. 243–253. Springer, Heidelberg (2010)
33. Scorza, F., Casas, G.B., Murgante, B.: That's ReDO: ontologies and regional development planning. In: Murgante, B., Gervasi, O., Misra, S., Nedjah, N., Rocha, A.M.A., Taniar, D., Apduhan, B.O. (eds.) ICCSA 2012, Part II. LNCS, vol. 7334, pp. 640–652. Springer, Heidelberg (2012)
34. Selim, H.: Determinants of house prices in Turkey: hedonic regression versus artificial neural network. Expert Systems with Applications **36**(2), 2843–2852 (2009)
35. Torre, C.M., Mariano, C.: Analysis of fuzzyness in spatial variation of real estate market: Some Italian case studies. Smart Innovation, System and Technologies **4**, 269–277 (2010)
36. Wang, W.K.: A knowledge-based decision support system for measuring the performance of government real estate investment. Expert Systems with Applications **29**(4), 901–912 (2005)
37. Wong, K.C., Albert, P.S., Hung, Y.C.: Neural network vs. hedonic price model: appraisal of high-density condominiums. Real Estate Valuation Theory **8**(2), 181–198 (2002)