Artificial intelligence in property valuations
An application of artificial neural networks to housing appraisal

PIERLUIGI MORANO, FRANCESCO TAJANI, CARMELO MARIA TORRE
Department of Science of Civil Engineering and Architecture
Polytechnic of Bari
Via Orabona 4, 70125 Bari
ITALY
pierluigi.morano@poliba.it, francescotajani@yahoo.it, cartorre@yahoo.it,

Abstract: In recent years, social, economic and fiscal factors have produced strong modifications of the Italian real estate market, that currently appears as a complex system characterized by continuous transformation. In this context, for real estate operators the use of "slender" tools, able i) to operate even on limited data, ii) to automatically capture from data the causal relations between explanatory variables and prices, iii) to appraisal the market values that will reasonably occur in the short term, has become essential. Among artificial intelligence models, artificial neural networks (ANN) meet these prerogatives. In this paper, ANN is applied to the evaluation of market values of residential properties starting from a sample of apartments recently sold in a neighborhood of the city of Bari (Italy). The excellent results confirm the effectiveness of ANN in property valuations. The work must be attributed in equal parts to the authors.

Key-Words: - housing appraisal, artificial neural network, estimative analysis, market value.

1 Introduction

It is evident that the crisis of the Italian real estate sector is structural, covering all the fields of real estate, and is destined to last longer than previous cycles. In fact, in recent years, social, economic and fiscal factors have produced deep modifications of the Italian real estate market, that currently appears as a complex system characterized by continuous transformation. Among these factors, the negative trend of the main macroeconomic indicators (GDP reduction, increase in public debt and in unemployment, banking credit crunch), which led the operators to a cautious behavior, must be certainly mentioned. Although the European Central Bank has made significant efforts to revive the real estate sector, through policies of *quantitative easing* and the reduction of the official reference rate, set to 0.05% from September 2014, the attitude of the banks remains cautious and is characterized by financial products that favor borrowers with appropriate liquidity and strong guarantees, and provide loans to value below 80%. The majority of potential buyers, therefore, is excluded from the possibility of accessing to a funding.

Moreover, taxes on Italian real estate have reached in a few years unprecedented levels, with continuous changes of names and rates. The tax uncertainty that has arisen and the fear of further tax increases, related to the current passage from a Cadastre of incomes to a Cadastre of values (L. 23/2014), are causing further mistrust of operators.

Furthermore, the sale of public real estate, considered as the solution for reducing the national debt and satisfying the constraints of the European Stability Pact, added to the sale of constructions started before the crisis, is creating an excess supply which can be absorbed only in the long term.

The consequence is the contraction in sales, the lengthening of the sale time and the formation of anomalous prices [1]: each transaction, rather than the outcome of uniform behaviors, ends up constituting an event which depends on the contractual capacity and the liquidity of the parties, as well as on their interest and urgency to conclude the transaction.

In this context, the use of tools for the evaluation of real estate values has become essential for sector operators (buyers, sellers, institutions, insurance companies, banks, etc.). The continuous change of the boundary conditions 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 estimate property values in the short term.

ISBN: 978-1-61804-280-4

The artificial neural networks (ANN) satisfy these prerogatives. Many studies [2, 3, 4, 5] 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.

In real estate sector ANN models have been applied primarily to the property prices forecast [6, 7, 8] and to the segmentation of the real estate market [9, 10]. The aim of this paper is to deepen the issue and to test ANN in the construction of models for the evaluation of real estate market values.

The paper is structured as follows. In section 2, notes on neural networks theory are reported. In section 3 the case study is introduced: it is relative to a sample of apartments recently sold within the real estate market of the city of Bari (Italy). In section 4 the ANN model is specified and applied, and the results are illustrated. In section 5 the conclusions of the work are discussed.

2 Outline of Artificial Neural Networks models

ANN are complex systems [11], 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. Diametrically 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:

$$I_j = \sum_{i=1}^{N} O_i w_{ij} \quad (1)$$

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:

$$O_i = f(I_i) \quad (2)$$

The diagram of a neuron base is described in Figure 1, in which the neuron is divided into two parts: in the first part, on the left, the pulses received are added; in the second part, the output is determined through the activation function.

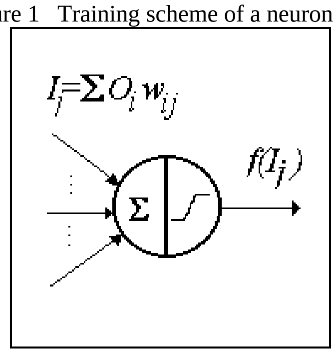
So that the neural network is able to learn a particular task assigned, it is necessary to transfer to the system a learning technique, that is a rule by which it is possible to appropriately update the weights of the connections of the network.

There are many learning techniques, each of them suited to achieving the objective. In general, it is possible to distinguish two categories of learning rules: the *supervised learning* and *unsupervised learning*.

Through the *supervised learning*, it is necessary to provide the network, as examples, with the possible inputs and the corresponding output (*targets*). In this way, the network learns by adjusting the weights of the connections so that, for any input given, the output is as close as possible to the target.

The *unsupervised learning* instead requires that the outputs are not provided, but only the inputs, i.e. the points of the space of possible inputs (*clusters*). In this case, the neural network tends to be configured in order to create a correspondence between sets of similar clusters and neurons of the output level.

Figure 1   Training scheme of a neuron base



The learning rule frequently implemented is the *back-propagation* method, which is one of the techniques of supervised learning. 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 value desired (target).

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.

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* (*FLO*) the apartment is on; the *number of bathrooms* (*BATH*) of the property; the *panoramic view* (*PAN*) 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” (PAN_{enough}) and “good” (PAN_{good}); the presence of *independent heating* (*HEAT*), 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* (H_{center}) of the city, expressed in minutes it takes to walk to it; the *rental situation* of the apartment (*RENT*), expressed through a dichotomous criterion, with 1 if the apartment is rented and 0 otherwise; the *surface* (*SURF*) of the apartment, expressed in square meters of floor area of the property.

Other common intrinsic characteristics, such as the presence of elevator or building quality, 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).

4.1 The ANN model

In the model developed in this work, the unit selling price (*PRZ*) identifies the dependent variable, whereas the other parameters collected (*FLO*, *BATH*, PAN_{enough}, PAN_{good}, *HEAT*, H_{center}, *RENT*, *SURF*) are the explanatory variables.

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.

ISBN: 978-1-61804-280-4

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

Table 1 – Sample descriptive statistics

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 (Figure 2) is associated to each neuron, which is modeled continuously between 0 and 1 using the following analytical expression:

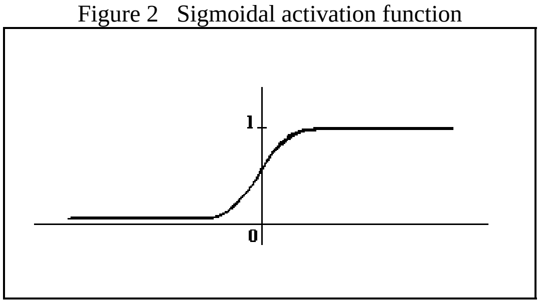
$$f(x) = \frac{1}{1+e^{-kx}} \tag{3}$$

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.

Figure 2 Sigmoidal activation function



The model is implemented through the software "BKP – Neural Network Simulator" [12], that employs the algorithm of Back-Propagation (BKP) to adjust iteratively the connection weights.

ISBN: 978-1-61804-280-4

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.

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 one of the explanatory variables compared to the RMSE of the model including all the variables.

It is noted that with ANN the importance of the variable, in descending order, is as follows: the *distance from the center* (H_{center}), the presence of the *independent heating* (*HEAT*), the *floor* (*FLO*), a “good” *panoramic view* (PAN_{good}), the *rental situation* (*RENT*), the *surface* (*SURF*) of the apartment, a “sufficient” *panoramic view* (PAN_{enough}) and finally the *number of bathrooms* (*BATH*) of the property.

The ANN model compares the output obtained, estimated on the entire sample of 90 transactions, to the detected prices. The graphical comparison is reported in Figure 3.

The graph shows a good correspondence between detected prices and estimated prices with the ANN model.

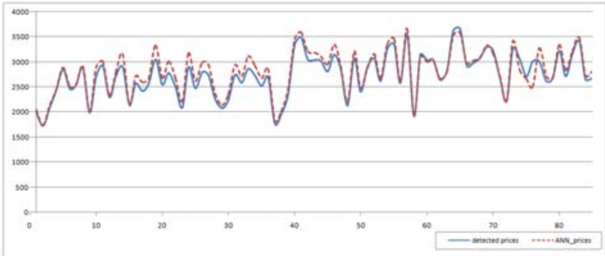
4.2 Interpretation of results obtained

The analysis of results shown in Table 2 confirms that in the Madonnella district the location is the feature characterized by the highest influence on the appreciation of a residential property. In this regard, it should be noted that the importance of this variable rather than the other exogenous factors considered is enhanced by the absence of fast connections in the Madonnella district (e.g. subway or light railway) with the center of Bari.

VARIABLE	*RATIO*	*ORDER OF IMPORTANCE*
FLO	1.509742	3
BATH	1.222877	8
PAN_{enough}	1.298589	7
PAN_{good}	1.487650	4
HEAT	1.596292	2
H_{center}	2.016161	1
RENT	1.354432	5
SURF	1.349758	6

Table 2 Sensitivity analysis

Figure 3 Comparison between the detected unit prices (continuous line) and the unit prices estimated with the ANN model (broken line)



The presence of the independent heating 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.

ISBN: 978-1-61804-280-4

The significant impact of the floor 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.

A similar weight to the floor level is obtained for the variable "good" panoramic view, relative to the apartments of the sample that are overlooking the sea. It should be noted that the use of ANN allows to overcome the effects of collinearity that might occur with other procedures (e.g. hedonic prices), because of the correlation that generally binds the variable "floor" and "panoramic view". Obviously, a minor appreciation corresponds to a lower panoramic view: indeed, the feature "enough" panoramic view stands at the penultimate place in terms of impact on property prices of the sample.

The rental situation and the surface of the apartment have comparable impacts. For the rental situation, the modest weight of this variable compared to what happens in other national contexts should be related to the excellent annual gross rental yields (+4.2%), reported for the city of Bari in the decade 2002-2012 [13]; this situation leads to a strong interest in buying for investment, especially for apartments located in areas - as the Madonnella district - adjacent to the center. In the case of the surface, the surveys carried out [14] have outlined that in the city of Bari major requirements concern two-room (32.6%) and three-room apartments (40.8%), whereas for higher sizes the demand falls below 10%, because of the fiscal pressure that discourages the purchase of large properties.

Finally, the weight of the variable "number of bathrooms", lower than the other explanatory variables, on the one hand expresses the less need of having more bathrooms, given the decrease in the number of people that constitute, on average, a family unit, equal to 2.4 [15]; on the other hand, it shows the modest importance attributed to this parameter when the buyer already provides the renovation of the apartment.

5 Conclusions

In the phase of uncertainty that is characterizing the Italian real estate sector, the use of "advanced" assessment tools may allow market operators to formulate more reliable estimates, as well as to effectively monitor the evolution of property values [16, 17, 18, 19, 20].

The ANN model defined in this work highlights excellent performance of the results, allowing to value other properties with the use of little information normally available on the market.

To conclude the work, two considerations should be illustrated, according to it has been already outlined by some authors. The first is that ANN are "black boxes": in fact, with ANN it is not possible to generate a straightforward functional relationship between the input and the output values nor punctually investigate and reproduce the mechanisms of the prices formation.

The other consideration is the possibility 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.

Interesting perspectives for developing future research and experiments are related to these issues.

6 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.

References:

[1] P. Morano, F. Tajani, 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, 2014, pp. 91-111.

[2] A.L. Brunson, R.J. Buttimer, R.C. Rutherford, *Neural Networks, Nonlinear Specification and Industrial Property Values*, University of Texas at Arlington, Working Paper Series, 1994, pp. 94-102.

[3] W.J. McCluskey, K. Dyson, D. McFall, S. Anand, The mass appraisal of residential property in Northern Ireland, in *Computer Assisted Mass Appraisal Systems*, Gower Publisher, 1997, pp. 59-77.

[4] K.C. Wong, P.S. Albert, Y.C. Hung, Neural network vs. hedonic price model: appraisal of high-density condominiums, *Real Estate Valuation Theory*, Vol. 8, No. 2, 2002, pp. 181-198.

[5] H. Selim, Determinants of house prices in Turkey: hedonic regression versus artificial

neural network, *Expert Systems with Applications*, Vol. 36, No. 2, 2009, pp. 2843-2852.

[6] J. Gallego, La inteligencia artificial aplicada a la valoraciòn de inmuebles. Un ejemplo para valorar Madrid, *Revista CT/Catastro*, No. 50, 2004, pp. 51-67.

[7] J. Liu, X. Zhang, W. Wu, Application of fuzzy neural network for real estate prediction, *LNCS*, Vol. 3973, 2006, pp. 1187-1191.

[8] P. Morano, F. Tajani, Bare ownership evaluation. Hedonic price model vs. artificial neural network, *International Journal of Business Intelligence and Data Mining*, Vol. 8, No. 4, 2013, pp. 340-362.

[9] A.Q. Do, G. Grudnitski, A neural network analysis of the effect of age on housing values, *Journal of Real Estate Research*, *American Real Estate Society*, Vol. 8, No. 2, 1993, pp. 253-264.

[10] K.S. Islam, Y. Asam, Housing market segmentation: a review, *Review of Urban & Regional Development Studies*, Vol. 21, Nos. 2-3, 2009, pp. 93-109.

[11] V. Limsombunchai, C. Gan, M. Lee, House price prediction: hedonic price model vs. artifical neural network, *American Journal of Applied Sciences*, Vol. 1, No. 3, 2004, pp. 193-201.

[12] S. Barile, L. Magna, M. Marsella, S. Miranda, A marketing decision problem solved by application of neural networks, *International Conference on Computational Intelligence and Multimedia Applications (ICIIMA)*, 1999.

[13] Tecnocasa, *Rendimenti Immobiliari 2012*, 2013, available online.

[14] Tecnocasa, *Mercato Immobiliare Puglia 2013*, 2014, available online.

[15] ISTAT, *15° Censimento Generale della popolazione e delle abitazioni 2011*, 2012, available online.

[16] C. Guarnaccia, Advanced Tools for Traffic Noise Modelling and Prediction, *Transactions on Systems*, Vol. 12, No. 2, 2013, pp. 121-130.

[17] V. Del Giudice, P. De Paola, Geoadditive models for property market, *Applied Mechanic and Materials*, Vol. 584, 2014, pp. 2505-2509.

[18] V. Del Giudice, P. De Paola, The effects of noise pollution produced by road traffic of Naples Beltway on residential real estate values, *Applied Mechanics and Materials*, Vol. 587, 2014, pp. 2176-2182.

[19] F. Calabrò, L.D. Della Spina, The public-private partnerships in buildings regeneration: A model appraisal of the benefits and for land value capture, *Advanced Material Research*, Vols. 931-932, 2014, pp. 555-559.

[20] C.M. Torre, C. Mariano, Analysis of fuzzyness in spatial variation of real estate market: Some Italian case studies, *Smart Innovation, System and Technologies*, Vol.4, 2010, pp. 269-277.