sustainability

MDPI

Article

Real Estate Appraisals with Bayesian Approach and Markov Chain Hybrid Monte Carlo Method: An Application to a Central Urban Area of Naples

Vincenzo Del Giudice [1], Pierfrancesco De Paola [1,*], Fabiana Forte [2] and Benedetto Manganelli [3]

[1] Department of Industrial Engineering, University of Naples "Federico II", 80138 Napoli, Italy; vincenzo.delgiudice@unina.it
[2] Department of Architecture and Industrial Design, University of Campania "Luigi Vanvitelli", 81100 Caserta, Italy; fabiana.forte@unicampania.it
[3] School of Engineering, University of Basilicata, 85100 Potenza, Italy; benedetto.manganelli@unibas.it
* Correspondence: pierfrancesco.depaola@unina.it

Received: 15 August 2017; Accepted: 15 November 2017; Published: 21 November 2017

Abstract: This paper experiments an artificial neural networks model with Bayesian approach on a small real estate sample. The output distribution has been calculated operating a numerical integration on the weights space with the Markov Chain Hybrid Monte Carlo Method (MCHMCM). On the same real estate sample, MCHMCM has been compared with a neural networks model (NNs), traditional multiple regression analysis (MRA) and the Penalized Spline Semiparametric Method (PSSM). All four methods have been developed for testing the forecasting capacity and reliability of MCHMCM in the real estate field. The Markov Chain Hybrid Monte Carlo Method has proved to be the best model with an absolute average percentage error of 6.61%.

Keywords: real estate appraisals; hedonic price model; artificial neural networks; Bayesian approach; Markov Chain Hybrid Monte Carlo Method; multiple regression analysis; Penalized Spline Semiparametric Method

1. Introduction

In general, the appraisal theory is based upon the hypothesis that a real estate sample is randomly drawn by a normal population with real estate sales characterized by known prices [1]. These hypotheses are rarely mirrored in practice because [2]: (a) the subjects that treat the sampling, rarely follow criteria based on random logic; (b) the atypicality of real estate properties systematically invalidates the normal distribution of the population.

In the field of real estate appraisals often there is low transparency with regard to market information, together with stationary conditions, and these aspects force the analysts to work with single and small datasets for the implementation of hedonic pricing models.

In the last decades, many deterministic and probabilistic models for hedonic analysis of housing sales prices were developed.

Hedonic price models assume that the values of real estate properties are influenced by their characteristics, with predicted values strongly influenced by dimension and quality of available real estate data. In particular, is it understood that the minimum dimension of real estate sample necessary to implement a statistical inference model is correlated to the number of independent variables explaining real estate characteristics [3–18].

Traditional statistical models used to determine the econometric function of the real estate prices assume that the price distribution (depend variable) is representative of a deterministic function of the real estate characteristics considered (independent variables). The noisiness of statistical sample is

generally modelled hypothesizing that the outputs are affected by a Gaussian noise with a zero mean and constant variance, which does not depend on input variables. Then, the interpolating function is chosen using the maximum likelihood criterion. In this sense, it has been demonstrated that in the case of an output noise independent from the input variables, the maximum likelihood criterion provides distorted results because it systematically underestimates the variance of the noise itself [19–22].

The possibility to characterize the interpolating function in probabilistic terms and, thus, to obtain as output a distribution of sales prices instead of deterministic values, jointly with the capability to work with small samples, are aspects that mainly make us lean in favour of the use of the Bayesian approach to artificial neural networks (ANNs or NNs) for real estate appraisals.

ANNs are techniques programmed to function as human neural networks, with learning capacity about complex function approximations from input/output. ANNs can be combined with Bayesian approaches in order to avoid some problems of stochastic optimization. Bayesian learning is particularly suited to real estate appraisals, because Bayesian inference techniques lead to reliable results, especially with small and noisy samples within the framework of probabilistic inference carried out with ANNs. Taking into account these aspects, a neural networks model with Bayesian learning has been experimented for an urban central area of Naples (Vomero neighbourhood).

2. Target and Research Design

This paper experiments a neural networks model (NNs) with a Bayesian approach on a small real estate sample. The output distribution is calculated by operating a numerical integration on the weights space with the Markov Chain Hybrid Monte Carlo Method (MCHMCM).

The basic goal of the paper is to test the forecasting potentialities and the reliability of MCHMCM in the real estate field.

On the same real estate sample, MCHMCM is compared with a NNs model, traditional multiple regression analysis (MRA) and the Penalized Spline Semiparametric Method (PSSM). All four methods are developed to compare the forecasting capacity of MCHMCM, with all of these methods sharing the same application fields.

More precisely, the NNs model was chosen to understand the difference compared to NNs with Bayesian learning. MRA has been chosen for comparison purposes since it has become a well-established methodology over time that provides robust and reliable results in real estate appraisals. Finally, PSSM has been chosen because it is a more flexible method than MRA and, potentially, it is in able to provide better results than MRA in real estate appraisals.

The paper consists of four parts. The first part contains a literature review on NNs starting from the hedonic price model. In the second part, the Bayesian approach for NNs is dealt with in analytic terms. In the third part, a hedonic analysis of housing sales prices is developed for the four methods mentioned, also describing real estate data and the main aspects of MCHMCM. At the end of the third part a results comparison is presented. Finally, in the fourth part, some concluding remarks are given.

3. Background

In the 1960s, Lancaster [23] defined the utility as function of the characteristics of goods. However, Lancaster aimed his studies on the fundamental aspects of consumers; only with Rosen [24] a model of market equilibrium was formulated, taking into account consumers and producers and, thus, all interactions between supply and demand.

The basic model of the real estate market equilibrium [25] assumes that real estate involves complex goods, with sales prices (P_i) which depend on location-specific environmental characteristics (q_i), structural characteristics (S_i), social and neighbourhood characteristics (N_i), and locational characteristics (L_i):

$$P_i = P(q_i, S_i, N_i, L_i)$$

Obviously, the above hedonic equilibrium equation applies to all real estate goods falling in the same market, but this aspect creates problems when different submarkets are present in the same urban area (with consequent spatial heterogeneity) [26]. Large-scale implementation of linear or nonlinear hedonic models can be found in the literature, but their usefulness depends on how models minimize pricing errors due to the large samples from which dataset are extrapolate. Specification error is almost always unavoidable because values does not map linearly onto real estate characteristics, and so errors are due to neglected nonlinearities [27].

In general, nonlinear models would be the preferred choice for real estate appraisals, but the exact nonlinear form is neither apparent nor are there necessarily practical steps one could take to find the correct form.

Artificial neural networks (ANNs) provide a practical alternative to conventional least squares forms (including nonlinear least squares) that is easily implementable and which efficiently models nonlinearities in the underlying relationships (including the parameters).

There are conflicting views on the superiority of ANNs compared to hedonic pricing models based on multiple regression analysis (MRA). Tsukuda and Baba [28], Do and Grudnitski [29], Tay and Ho [30], and Huang et al. [31] state that ANNs are superior to MRA. Our study supports these conclusions. Allen and Zumalt [32] and Worzala et al. [33] suggested otherwise. Guan et al. [34] combine fuzzy set theory in ANNs architecture to assess property values, and also argue that ANNs better replicate the heuristic thought processes and the imprecision in their decision calculus. Nguyen and Cripps [35] found that ANNs outperformed MRA with large datasets. Worzala, Lenk, and Silva [33] compared two ANNs to MRA with a small set of real estate transactions, concluding that ANNs were not superior to MRA in housing real estate appraisals, and warned appraisers who wish to use ANNs to do so with caution, citing inconsistent results between different software packages. It is indeed true that ANNs can be over-trained, resulting in good training runs but poor out-of-sample performance, that there is some sensitivity to outliers in training, and that results may be inconsistent in small samples. Results may also appear inconsistent for the simple reason that network weights are often randomly initialized; thus, gradient descent algorithms produce different solutions for the network's weight vectors simply because they begin iterating from different points on the loss surface.

More recently, in literature others studies have been proposed to confirm the superiority of ANNs or to show possible integrations with others techniques. Most relevant and recent studies have been proposed by Lam et al. [36], McCluskey et al. [37,38], Zurada et al. [39], Morano et al. [40,41], Tajani et al. [42], Ding et al. [43]. Critics of the ANNs also cite the relative ease of interpretation of hedonic MRA with respect to ANNs; in particular, the partial differentiation of linear models easily isolates each explanatory variable's contribution to value. Although the differentiation of ANNs is more difficult given variable interdependencies, it is relatively straightforward to uncover individual variable attributions [44,45].

4. Bayesian Approach for Neural Networks

Neural networks (NNs) had wide interest due to empirical achievements on a wide range of learning issues. NNs are highly expressive models that can learn complex function approximations from input/output examples [46], with a particular ability to train them on massive data sets with stochastic optimization [47] and the backpropagation algorithm [48]. These aspects have resulted in successes for NNs in many fields of applied science and the economy.

The Bayesian approach to NNs can potentially avoid some of the problems of stochastic optimization [49,50]. In fact, Bayesian methods consider the uncertainty in parameter estimates and can extend and transform this uncertainty into predictions. Often, Bayesian approaches are more robust to overfitting, since they average over parameter values instead of choosing a single point estimate. Different approaches have been proposed in literature for Bayesian learning of NNs, based on, for example, the Laplace approximation [49,50], Hamiltonian Monte Carlo [51–53], expectation propagation [54], and variational inference [55]. These approaches have not seen widespread adoption

due to their lack of scalability for network architecture and data size. An exception is the scalable variational inference approach of Graves [56], but this method seems to perform poorly in practice due to noise from Monte Carlo approximations within the stochastic gradient computations. A different scalable solution based on expectation propagation was proposed by Soudry et al. [57], although its extension to continuous weights is unsatisfying as it does not produce estimates of posterior variance.

The central goal in network training is to model the underlying generator of the data, so that the best possible predictions for the output vector t can be made for the input vector x.

The most general and complete description of the generator of the data is in terms of the probability density $p(x, t)$. For associative prediction problems it is convenient to decompose the joint probability density into the product of the conditional density of the target data, conditioned on the input data, and the unconditional density of input data, so that:

$$p(x,t) = p(t|x)p(x) \tag{1}$$

where $p(t \mid x)$ indicates the probability density of t given that x takes a particular value, and $p(x)$ represents the unconditional density of x. For making predictions of t for values of x, it is the conditional density $p(t \mid x)$ which we need to model.

Often, the error function is based on the principle of maximum likelihood.

If $\{x^n, t^n\}$ is a set of training data the likelihood L can be written as:

$$L = \prod_n p(t^n|x^n) = \prod_n p(t^n|x^n)p(x^n) \tag{2}$$

where we have assumed that each data point (x^n, t^n) is drawn independently from the same distribution, and hence we can multiply the probabilities. For computational reasons, instead of maximizing the likelihood, it is more convenient to minimize the negative logarithm of the likelihood, so that the error function (E) is given by:

$$E = -\ln L = -\sum_n \ln p(t^n|x^n) - \sum_n \ln p(x^n) \tag{3}$$

A feed-forward neural network can be regarded as a framework for modelling the conditional probability density $p(t \mid x)$, then the second term in Equation (3) does not depend on the network parameters, and so represents an additive constant which can be dropped from the error function:

$$E = -\ln L = -\sum_n \ln p(t^n|x^n) \tag{4}$$

Consider the case of c target variables t_k where $k = (1, ..., c)$, and suppose that the distributions of the different target variables are independent:

$$p(tx) = \prod_{k=1}^{c} p(t_k|x) \tag{5}$$

If the distribution of the target data is Gaussian, or the target variable t_k is given by some deterministic function of x added Gaussian noise ε, then:

$$t_k = h_k(x) + \varepsilon_k \tag{6}$$

We can assume that the errors ε_k have a normal distribution with zero mean, and a standard deviation σ which does not depend on x or k. In these cases, the distribution of ε_k is given by:

$$\rho(\varepsilon_k) = \frac{1}{(2\pi\sigma^2)^{1/2}} \cdot exp\left(-\frac{\varepsilon_k^2}{2\sigma^2}\right) \tag{7}$$

The $h_k(x)$ function can be chosen with parametric or nonparametric approach. In the first case, it needs to define the functional form that better than anyone else nears to dataset observations. In the second case, as in the neural networks, the input-output mapping is connected to the realization of pattern learning phase. In both cases, models parameters must be selected (coefficients that define in univocal way the interpolating function for parametric approaches, or synaptic weights matrix for neural networks).

The $h_k(x)$ function can be modelled with a neural network with outputs $y_k(x, w)$, where w is the set of weight parameters governing the neural network mapping. Using (6) and (7), the probability distribution of target variables is given by:

$$p(t_k|x) = \frac{1}{(2\pi\sigma^2)^{1/2}} \cdot exp\left(-\frac{\{y_k(x,w) - t_k\}^2}{2\sigma^2}\right) \tag{8}$$

Together with Equations (4) and (5), we have the following expression for the error function:

$$E = \frac{1}{2\sigma^2} \sum_{n=1}^{N} \sum_{k=1}^{c} \{y_k(x^n, w) - t_k^n\}^2 + Nc \ln \sigma + \frac{Nc}{2} \ln(2\pi) \tag{9}$$

The second and third terms, as also the overall factor $1/\sigma^2$, do not depend from the weights w and they can be omitted:

$$E = \frac{1}{2\sigma^2} \sum_{n=1}^{N} \sum_{k=1}^{c} \{y_k(x^n, w) - t_k^n\}^2 = \frac{1}{2} \sum_{n=1}^{N} \|y(x^n, w) - t^n\|^2 \tag{10}$$

The expression of the error function E has been derived in the hypothesis that the target data distribution is Gaussian. In case it is not reasonable to assume a similar hypothesis, for example, when the distribution is multimodal, the previously approach can lead to distorted results that very difficult to interpret.

The real estate sample may be also subject to noise not always distributed uniformly and independently by the input vectors. Bayesian approaches may provide reliable results with small and noisy samples in the field of probabilistic inference using neural network models [58].

A Bayesian approach makes it possible to express the uncertainty in the weights value through a probability distribution $p(w \mid D)$ conditioned on the dataset. In this way the neural network is no longer represented by a single vector of the synaptic weights obtained by applying the maximum likelihood criterion, but by a density probability on the weights space. The main operative consequence is that the model output will have probabilistic features and can be represented by any distribution, calculating the most probable values and confidence interval of each single prevision.

Also, Bayesian approaches make it possible to select the regularization coefficients and, thus, the model complexity, only using a training set without a validation set. This latter circumstance is particularly important when there is a limited dataset, as in real estate appraisals often occurs.

Bayesian methods make it possible to avoid overfitting problems which are typical of conventional approaches used for the network training and, at least on principle, they do not limit the model complexity [53].

The issue of determination of the value distribution of outputs can be solved using the integration on multidimensional space of the synaptic weights. With the rule of probability marginalization it is possible to express the output distribution for a given input vector x as follows:

$$p(t|x, D) = \int p(t|x, w) p(w|D) dw \tag{11}$$

where $p(w \mid D)$ is the posterior weights distribution on the target data $D \equiv (t^1, ..., t^N)$, that can be expressed starting from a prior distribution $p(w)$ using the Bayes theorem:

$$p(w|D) = \frac{p(D|w)p(w)}{p(D)} \tag{12}$$

where in (12) the denominator of arithmetical ratio is a normalization factor.

The integration over the weights space is analytically untreatable, so that the main approximation methods are two. The first leads to an analytical integration in the restrictive hypothesis that the posterior weights distribution is expressed with a Gaussian distribution centred on the most probable weights vector [49,50]. The second applies a numerical integration using the Monte Carlo technique considering the high dimensionality of the problem [51–53].

5. Hedonic Analysis of Housing Sales Prices

5.1. Data Description

The application has been made with a real estate sample relative to a urban central area of Naples (Vomero neighbourhood).

The sample consists of 65 housing sales occurred in the last 12 months. The housing units present the same building typology and are situated in a homogeneous urban area under the profile of the distribution of public and private services. The age of buildings where real estate units are located dates back to the 1960s.

Only non-homogeneous real estate characteristics have been detected for each sampled unit and, in particular (see Table 1): the sale price (PRC) is expressed in €/total square metres (×1000); the number of bathrooms (BATH) of housing unit; the outer surface (OUT) is expressed in square metres; floor level (FLOOR) of the housing unit; maintenance and preservation status (MAIN) is expressed with a score scale (1 = mediocre; 3 = good; 5 = very good); panoramic quality of housing unit (PAN) is expressed with a score scale (1 = without panoramic views; 3 = half of views of the housing unit are panoramic; 5 = all views of the housing unit are panoramic); the occupancy status of the property (LOC) is expressed with a dichotomous measure (1 = not rented property; 0 = rented property); the distance from the funicular station (STAT) is expressed with a dichotomous measure (1 = if the distance between the housing unit and the station is lower than 500 m; 0 = if otherwise).

Table 1. Statistical description of variables. Sale price (PRC); number of bathrooms (BATH); outer surface (OUT); floor level (FLOOR); maintenance status and preservation status (MAIN); panoramic quality of housing unit (PAN); occupancy status of the property (LOC); distance from the funicular station (STAT).

Variable	Std. Dev.	Median	Mean	Min	Max
PRC	1.06	3.44	3.40	1.18	6.11
BATH	0.50	1.00	1.45	1.00	2.00
OUT	11.75	12.00	13.89	0.00	40.00
FLOOR	1.45	3.00	2.65	0.50	7.00
MAIN	1.46	3.00	2.94	1.00	5.00
PAN	1.28	1.00	1.89	1.00	5.00
LOC	0.49	1.00	0.62	0.00	1.00
STAT	0.50	0.00	0.46	0.00	1.00

The property characteristics and the relative amounts have been considered in the same way in each technique applied (MCHMCM, NNs, MRA, PSSM).

Other real estate characteristics (like noise level, accessibility, etc.), manifesting with the same modalities in all the sampled units, have been excluded from the analysis.

The selection of the most significant real estate characteristics has been verified on the basis of direct knowledge of authors about the real estate market of Naples and following a statistical pre-processing (verification of normal data distribution, correlation test on variables) connected to the need to determine the intrinsic dimensionality of the data set used [59]. This has been done to reduce the effects of correlation among the variables and to obtain a better coverage of the characteristics space. The reduction of dimensionality has been conducted to enable the identification of the optimal balance between loss of information (e.g., elimination of a variable) and the need to conserve relevant information contained in the dataset [60].

For small real estate sample sizes [15], Green [61] used statistical power analysis in order to compare the performance of several rules of thumb for how many subjects were required for multiple regression analysis. Marks et al. [62] specifies a minimum of 200 subjects for any regression analysis. Tabachnick and Fidell [63] suggested that although 20 subjects per variable would be preferable, the minimum required subjects should be five. Harris [64] argued that the number of subjects should exceed the sum of 50 and number of predictor variables. Schmidt [65] determined that the minimum number of subjects per variables lies in the range of 15 to 20, while Harrell [66] suggested that 10 subjects per variable was the minimum required sample size for linear regression models to ensure accurate prediction in subsequent subjects. Finally, Del Giudice [67] argued that the required sample size minimum is equal to four to five subjects per independent variable.

As detailed above, also considering the opacity of Italian real estate markets conditions [68,69], the small real estate sample used in this work should not prejudice the operational usability of statistical techniques.

5.2. Markov Chain Hybrid Monte Carlo Method

The type of proposed Bayesian model is Multi Layer Perceptron (MLP) with a single hidden layer. It consists of 30 housing units and it is trained using just 35 housing sales drawn from the real estate sample. This is done in order to verify the approach capacity of working with small dimension samples also related to the input vector dimensionality. The remaining part of the sample (30 housing sales) serves to test the network generalization capacity.

In the first stage there is the choose of weights prior distribution which, in some way, should reflect a possible prior knowledge that the analyst may have on the input-output mapping form. Various are the possible functions for the prior distribution of weights, but generally the form used is exponential [53,70]:

$$p(w) = \frac{1}{Z_w(\alpha)} exp(-\alpha E_w) \tag{13}$$

where $Z_w(\alpha)$ is a normalization factor, α is called hyperparameter inasmuch it controls the distribution of other parameters (weights and biases) and is determined during the training phase, while:

$$E_w = \frac{1}{2}\|w\|^2 \tag{14}$$

E_w is a function that favours small values for the network weights in order to encourage the network smooth mapping. To achieve a good balance between bias and variance in order to avoid overfitting phenomena by training data, it is an advisable choice for a function smooth network [19].

In the same way, for a $p(D \mid w)$ distribution an exponential form may be chosen so that:

$$p(D|w) = \frac{1}{Z_D(\beta)} exp(-\beta E_D) \tag{15}$$

where $Z_D(\beta)$ is a normalization factor, β is a hyperparameter determined during the training phase that controls the noise variance, and E_D is the classical square function derived from the maximum likelihood.

Using the Bayes theorem (12) the posterior weights conditioned on the training set is given by the following relation:

$$p(w|D) = \frac{1}{Z_S} exp(-\beta E_D - \alpha E_w) \quad (16)$$

where Z_S is a normalization factor. It can be verified that, with α and β fixed, increasing the training set dimension, the first term represents the maximum likelihood while the second term represents the prior knowledge that decreases, so that larger is the set and better the maximum likelihood approximates the most probable weights distribution [53].

The output distribution (11) is calculated operating a numerical integration on the weights space with the help of the Markov Chain Hybrid Monte Carlo Method (MCHMCM) [51,52]. The basic idea is to select a random sample of points w_i, in the weights space paying attention to generate such sample so that it may be representative of the distribution $p(w \mid D)$. Then it is necessary to make sure that the selected weights vectors are in the region where $p(w \mid D)$ is reasonably large. This can be done considering the vector sequence, where each following vector depends on the previous, adding a stochastic component so that:

$$w_{new} = w_{old} + \varepsilon \quad (17)$$

where ε is a small random vector chosen, for example, by a spherical Gaussian distribution with small variance parameter. In order to avoid w_{new} orientating towards the direction where the posterior probability is low, it is possible to use the information about the gradient which can be obtained using a standard back-propagation algorithm. This is the so-called Hybrid Monte Carlo method [71]. With the Hybrid Monte Carlo Algorithm it is possible to generate a sample w_i and, in correspondence to each x input vector, we can calculate the network prevision $y(x_i, w_i)$ so we can assess the distribution $p(t \mid x, D)$ as sum of the single $y(x_i, w_i)$. At the same time, during the training phase, the hyperparameters can be assessed with the Gibbs sampling technique (the elaborations have been carried out with Netlab free software, Aston University, Birmingham, UK).

In Table 2 the most probable values, the confidence interval at 95% and the percentage error, are given for each housing unit of the test set (the test set contains housing units not used to assess the model and, thus, completely unknown to it). The model shows good performance in terms of forecasting capacity with an absolute average percentage error of 6.61%. Also, the confidence intervals are less wide where the given input density is larger, this is because in the region with a high data density the contribution to the output variance due to the weights uncertainty has a trend with an inverse relation to the probability density (and, thus, the variance on the output is mostly due to the noise term) [72].

Table 2. Model Performances of Markov Chain Hybrid Monte Carlo Algorithm.

No.	BATH	OUT	FLOOR	MAIN	PAN	LOC	STAT	PRC	Lower Bound	Most Probable Value	Upper Bound	% Error
1	2	0	2	5	1	1	0	5.00	4.34	4.63	4.95	−7.40
2	2	10	5	3	1	1	1	4.65	4.06	4.37	4.67	−6.10
3	1	0	0.5	1	1	1	1	4.00	3.80	4.14	4.44	3.55
4	2	0	3	5	1	0	1	2.24	2.22	2.42	2.62	8.04
5	2	20	1	3	1	1	1	4.00	4.13	4.40	4.67	10.00
6	1	10	2	5	3	1	1	4.32	3.60	3.86	4.12	−10.63
7	1	10	3	3	1	0	1	3.04	2.57	2.72	2.87	−10.53
8	2	20	5	1	1	1	0	4.61	4.18	4.42	4.68	−4.12
9	2	20	1	3	5	1	1	3.14	2.69	2.99	3.29	−4.76
10	1	15	1	1	3	1	1	3.66	3.31	3.61	3.91	−1.34
11	1	30	3	3	3	1	0	4.07	3.86	4.12	4.38	1.20
12	1	0	1	3	3	0	0	2.22	2.16	2.40	2.84	8.11
13	2	25	3	3	1	1	1	6.11	4.99	5.22	5.45	−14.57
14	2	10	2	1	3	0	0	2.42	2.35	2.53	2.72	4.55
15	2	15	2	3	1	0	1	2.36	2.27	2.54	2.81	7.63
16	2	20	4	3	1	0	0	2.89	2.99	2.86	3.14	−1.00

Table 2. *Cont.*

No.	BATH	OUT	FLOOR	MAIN	PAN	LOC	STAT	PRC	Lower Bound	Most Probable Value	Upper Bound	% Error
17	2	20	4	1	1	1	1	4.39	3.66	3.93	4.20	−10.47
18	1	0	2	3	1	1	1	5.15	4.16	4.46	4.76	−13.40
19	2	15	4	3	1	1	0	4.23	3.85	4.10	4.35	3.07
20	2	0	0.5	5	1	1	0	2.89	2.35	2.68	2.97	−7.34
21	1	15	4	5	1	1	0	4.34	3.71	3.98	4.25	−8.23
22	2	10	3	1	3	0	0	4.00	3.48	3.72	3.96	−7.00
23	2	20	3	3	5	1	0	3.91	3.74	4.01	4.27	2.43
24	1	20	1	1	1	0	0	1.22	0.88	1.18	1.19	−2.96
25	1	10	4	3	3	1	0	3.20	3.26	3.52	3.78	10.00
26	1	30	2	3	1	1	0	2.13	1.97	2.27	2.57	6.57
27	2	15	5	5	3	1	0	2.67	2.23	2.50	2.74	−6.37
28	2	15	6	1	1	1	0	2.27	2.00	2.40	2.72	5.73
29	1	10	4	3	3	1	0	4.19	3.56	3.82	4.03	−8.83
30	1	12	3	5	1	1	0	3.44	3.37	3.52	3.72	2.33

5.3. Neural Network Model

On the same real estate sample, a neural networks model has been developed to compare the forecasting capacity of the Markov Chain Hybrid Monte Carlo Method.

In Table 3, the dataset is divided into a learning set, a validation set and a test set.

Table 3. Real estate sample for neural network model: learning set, validation set, and test set.

No.	BATH	OUT	FLOOR	MAIN	PAN	LOC	STAT	PRC	Description
1	2	0	2	5	1	1	0	5.00	Learning
2	2	0	3	1	1	1	0	4.00	Test
3	1	10	1	5	1	1	0	4.06	Validation
4	1	0	0.5	3	3	0	0	2.17	Learning
5	2	10	5	3	1	1	1	4.65	Learning
6	1	0	0.5	1	1	1	1	4.00	Learning
7	1	0	3	1	1	1	0	5.00	Test
8	2	30	3	5	1	0	1	2.61	Validation
9	2	0	3	5	1	0	1	2.24	Test
10	1	10	3	3	1	0	1	3.04	Test
11	2	20	1	3	1	1	1	4.00	Learning
12	1	30	0.5	1	5	0	1	1.18	Validation
13	2	35	4	5	1	0	1	3.82	Learning
14	1	10	3	5	1	0	1	2.77	Learning
15	2	25	4	1	1	0	0	2.87	Test
16	1	10	2	5	3	1	1	4.32	Learning
17	1	10	2	1	1	1	1	3.02	Validation
18	1	10	0.5	3	1	1	0	2.56	Test
19	1	10	4	3	1	0	1	4.44	Test
20	2	20	5	1	1	1	0	4.61	Learning
21	2	20	3	3	3	1	1	4.17	Learning
22	1	0	1	3	1	1	0	3.79	Validation
23	2	20	1	3	5	1	1	3.14	Learning
24	1	15	1	1	3	1	1	3.66	Learning
25	1	0	0.5	5	1	1	0	2.40	Learning
26	2	0	3	3	3	1	1	4.20	Validation
27	1	30	3	3	3	1	0	4.07	Learning
28	1	0	1	3	3	0	0	2.22	Test
29	1	0	1	1	3	0	0	2.72	Learning
30	1	0	4	3	3	1	1	3.62	Test
31	2	20	2	5	3	1	1	3.33	Validation
32	2	25	3	3	1	1	1	6.11	Learning
33	2	15	3	5	1	1	0	4.12	Learning

Table 3. *Cont.*

No.	BATH	OUT	FLOOR	MAIN	PAN	LOC	STAT	PRC	Description
34	1	40	3	5	1	0	0	2.25	Test
35	1	0	2	3	1	0	1	4.07	Learning
36	1	0	3	1	1	0	1	2.78	Learning
37	2	10	2	1	3	0	0	2.42	Test
38	1	20	3	3	1	0	0	2.81	Learning
39	2	15	2	3	1	0	1	2.36	Test
40	2	20	4	3	1	0	0	2.89	Test
41	2	40	3	1	3	0	1	2.87	Test
42	1	40	2	1	5	0	0	3.47	Validation
43	2	20	4	1	1	1	1	4.39	Learning
44	2	20	2	5	3	1	1	4.16	Learning
45	1	0	2	3	1	1	1	5.15	Learning
46	2	15	4	3	1	1	0	4.23	Learning
47	1	30	0.5	3	1	1	1	5.58	Validation
48	2	0	0.5	5	1	1	0	2.89	Learning
49	1	15	4	5	1	1	0	4.34	Learning
50	2	10	3	1	3	0	0	4.00	Learning
51	2	20	3	3	5	1	0	3.91	Learning
52	1	0	4	3	3	0	0	3.17	Test
53	1	20	1	1	1	0	0	1.22	Validation
54	1	0	0.5	3	1	1	0	2.00	Test
55	1	10	4	3	3	1	0	3.20	Learning
56	1	40	2	3	1	0	1	1.21	Validation
57	1	30	2	3	1	1	0	2.13	Learning
58	2	15	5	5	3	1	0	2.67	Test
59	2	10	3	3	1	0	1	1.69	Learning
60	2	15	6	1	1	1	0	2.27	Validation
61	1	15	3	1	3	1	1	4.39	Test
62	1	10	4	3	3	1	0	4.19	Test
63	2	16	7	3	5	1	0	2.95	Learning
64	1	12	3	5	1	1	0	3.44	Learning
65	1	10	4	3	1	1	0	3.67	Validation

For the selection of subdata it is necessary to guarantee a high level of homogeneity, thereby eliminating outliers data that may compromise the performance of the neural networks model. The test set contains 19 housing sales (about 25% of real estate data), and the validation set contains 12 housing sales obtained with random extraction (about 20% of real estate data). Consequently, the learning set contains 33 housing sales (about 55% of real estate data).

Homogeneity among the above subdata has been checked through some analysis of descriptive statistics (variance analysis, cluster analysis) with some modification of the test set and learning set. Statistical analysis becomes necessary when a random extraction is carried out on the basis of a small dimension sample, which does not guarantee a high homogeneity among subdata.

The dataset has been normalized dividing each variable with its maximum amount in order to obtain matrixes and vectors with number values between zero and one, and to avoid the possible phenomenon of saturation of the activation functions.

The normalized data have been used to train a neural net with seven inputs (locational and physical parameters), with two hidden layers each made up of seven neurons and with single output neuron representing the unitary price. The net has been trained over 6000 cycles with activation function of logsig type for the neurons of the hidden layers and linear for the output neuron. The learning has been carried out through backpropagation algorithm. The validation set has been used for the tuning of parameters and the control of overfitting phenomenon, interrupting the learning phase in correspondence with the error level representative of threshold, beyond which

the error function on the validation patterns starts to grow, while the error function on the training patterns keeps decreasing [48].

The forecasting capacity of neural networks model has been tested with the help of test set made up with patterns completely unknown to the net. Results are reported in Table 4 where the last column contains the percentage error between sale price and predicted price.

The absolute average percentage error is 7.33%, and this value is under the acceptability threshold of 10% for real estate appraisals.

Table 4. Results of neural networks model.

No.	Price	Predicted Value	% Error
2	4.00	3.64	−9.00%
7	5.00	4.58	−8.40%
9	2.24	2.10	−6.25%
10	3.04	3.06	0.66%
15	2.87	2.66	−7.32%
18	2.56	2.54	−0.78%
19	4.44	4.24	−4.50%
28	2.22	2.02	−9.01%
30	3.62	3.87	6.91%
34	2.25	2.38	5.78%
37	2.42	2.53	4.55%
39	2.36	2.60	10.17%
40	2.89	2.81	−2.77%
41	2.87	3.06	6.62%
52	3.17	3.51	10.73%
54	2.00	2.26	13.00%
58	2.67	3.03	13.48%
61	4.39	3.98	−9.34%
62	4.19	3.79	−9.55%

In order to further confirm the results, some nonparametric statistical tests with a validation set have been carried out. In particular, Spearman's rank correlation coefficient or Spearman's Rho has been calculated (r_s = 0.780) and compared with r_s critical (0.746) with a $\alpha_{0.05}$. This comparison shows a positive association between ranked housing prices and predicted housing values at the 1% confidence level. The calculus of the Spearman coefficient has been carried out as follows:

$$r_s = 1 - 6 \times \frac{\sum_0^n a^2}{n(n^2 - 1)} \tag{18}$$

where a is the difference between the ranked housing prices and predicted housing values, and n is the number of observations for validation set [73].

Spearman's rank correlation coefficient is a nonparametric measure of rank correlation (statistical dependence between the ranking of two variables). It assesses how well the relationship between two variables can be described using a monotonic function. If there are no repeated data values, a perfect Spearman correlation of +1 or −1 occurs when each of the variables is a perfect monotone function of the other. Intuitively, the Spearman correlation between two variables will be high when observations have a similar (or identical for a correlation of 1) rank (i.e., relative position label of the observations within the variable: 1st, 2nd, 3rd, etc.) between the two variables, and low when observations have a dissimilar (or fully opposed for a correlation of −1) rank between the two variables. Spearman's coefficient is appropriate for continuous and discrete variables, including ordinal variables.

5.4. Multiple Regression Analysis

With the aim to show the forecasting potentialities of the Markov Chain Hybrid Monte Carlo Method (MCHMCM), a traditional multiple regression analysis (MRA) with and without intercept (MRA_{int}, $MRA_{no.int.}$) on the same real estate sample (all 65 housing units) has been implemented. From a general point of view MCHMCM and MRA have the same potentialities and application fields, thus, the comparison between the two techniques may be performed in terms of residue analysis.

With respect to MCHMCM, the MRA also provides the marginal prices for each real estate characteristic.

The output data of MRA_{int} (with intercept) show a determination index (R^2) equal to 0.367, a determination index corrected ($R^2{}_C$) equal to 0.289, and a standard error equal to 0.890 and a negative *F*-test (under 95% of significance level). The absolute average percentage error is 23.16%; this value is over the acceptability threshold of 10% for real estate appraisals.

Output data of $MRA_{no.int.}$ (without intercept) show a determination index (R^2) equal to 0.921, a determination index corrected ($R^2{}_C$) equal to 0.895, a standard error equal to 1.060, and a positive *F*-test (over 95% of significance level). The absolute average percentage error is 26.22%, and the acceptability threshold of 10% is also overcome in this case.

Table 5 shows a detailed comparison between the residuals of the Markov Chain Hybrid Monte Carlo Method, the neural networks model, multiple regression analysis and the Penalized Spline Semiparametric Method.

5.5. Penalized Spline Semiparametric Method

The relationship between sales price and explanatory variables can be examined with a semi-parametric additive model, characterized by the combination of a generalized additive model, which expresses the relationship between non-linear response of model and explanatory variables, and a linear mixed effects model, which expresses the spatial correlation of observed values [8,9,11,12,16]:

$$P = y_j = \beta_0 + \beta_1 x_1 + \ldots.. + \beta_k x_k + f_{k+1}(x_{k+1}) + \ldots.. + f_n(x_n) + \varepsilon \quad (19)$$

In expression (19) the additive component, the mixed effects and the erratic component (ε), are independent. In order to obtain a function estimated using the procedures relating to models mixed effects, it is considered a version of low rank both for the additive component both for the mixed effects.

For non-linear components of the model (outer surface, OUT; floor level of the housing unit, FLOOR) are used in the Penalized Spline functions qualified by the following general expression:

$$f(x) = \alpha_0 + \alpha_1 x + \ldots.. + \alpha_p x^p + \sum_{z=1}^{Z} \alpha_{pz} (x - \kappa_z)^p{}_+ \quad (20)$$

where the generic function $(x - \kappa_z)^p{}_+$ has $(p - 1)$ continuous derivatives.

For $p > 0$ the expression that can be used to determine the fitted values is the following:

$$\hat{y} = X(X^T X + \lambda^{2p} D)^{-1} X^T y \quad (21)$$

where:

$$X = \begin{bmatrix} 1 & x_1 & .. & x_1{}^p & (x_1 - \kappa_1)^p{}_+ & .... & (x_1 - \kappa_z)^p{}_+ \\ .. & .. & .. & .... & .... & .... & \\ .. & .. & .. & .... & .... & .... & \\ .. & .. & .. & .... & .... & .... & \\ 1 & x_p & .. & x_n{}^p & (x_n - \kappa_1)^p{}_+ & .... & (x_1 - \kappa_z)^p{}_+ \end{bmatrix}$$

$$D = diag(0_{p+1}, 1_Z).$$

The smoothing parameter λ intervenes for determination of freedom degrees in the nonlinear component of model and makes it possible to control the trade-off between fitting model to the observed values (smoothing parameter near to zero value) and the smoothness of the same model (high values of smoothing parameters).

Selection the smoothing parameter, for a spline function of *p-degree*, occurs with the restricted maximum likelihood condition.

The tool used for the analysis of real estate data is R-project software (SemiPar package) [74].

Like MRA, PSSM is also able to provide the marginal prices for each real estate characteristic. However, given that the main aim of this paper is to show the forecasting potentialities of the MCHMC method for real estate appraisals, for brevity of discussion the results of the Penalized Spline Semiparametric Method (PSSM) are synthetically shown only graphically (see Figure 1), and the comparison between MCHMCM and PSSM has been performed in terms of residue analysis (see Table 5).

From the statistical point of view, for PSSM the determination index is equal to 0.830, the corrected determination index is equal to 0.819, and the standard error is equal to 0.326 and a positive *F*-test (over 95% of significance level). On the same real estate sample PSSM provides an absolute average percentage error equal to 20.92% (this value is over the acceptability threshold of 10% for real estate appraisals).

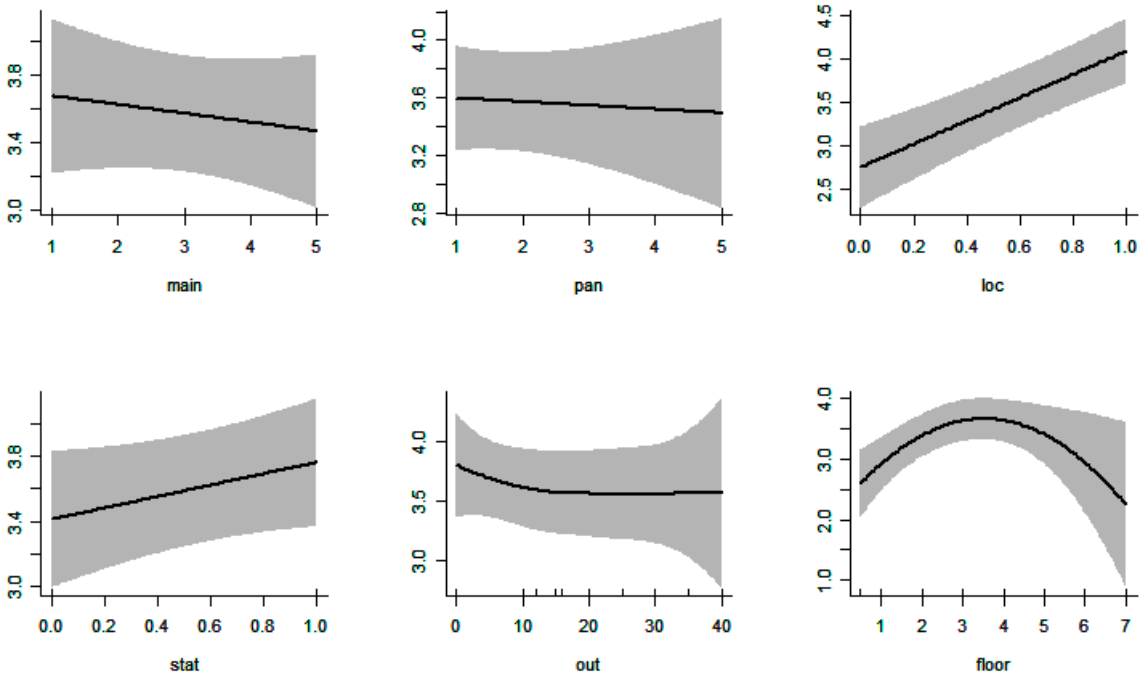

Figure 1. Effects of linear and nonlinear components on sales prices.

5.6. Results Comparison

In general, all four methods compared have the same forecasting potentialities and application fields, but unlike those of the Markov Chain Hybrid Monte Carlo Method (MCHMCM) and the neural networks model (NNs), multiple regression analysis (MRA) and the Penalized Spline Semiparametric Method (PSSM) are able to provide the marginal price for each real estate characteristic.

From a statistical point of view, even if PSSM shows better results than MRA, these two methods show obvious predictive limits due to the small sample size, and their predictive results are unacceptable.

MCHMCM and NNs show much better performances with respect to PSSM (20.92%) and MRA (better result 23.16% with intercept), in terms of forecasting capacity with an absolute average percentage error of 6.61% and 7.33%, respectively.

The analysis has shown an excellent prevision capacity of MCHMCM and NNs, even if the first method is better than the second (about 9.82% more precise), also considering that the learning set is not very numerous (35 housing sales). This is because Bayesian learning, by its nature, does not need a set for model validation and overfitting checking. For MCHMCM the confidence intervals have a good appraising coherence, showing themselves to be larger where the input data set density is smaller.

MCHMCM is the best model for maximum overestimation (10.00%), absolute average percentage error (6.61%), percentage of overestimation in the sample (0.00%) and total percentage of overestimation/underestimation (16.67%). NNs is the best model for maximum underestimation (9.55%), and the percentage of underestimation in the sample (0.00%). Among the four methods compared, MRA has shown the worst results.

Table 5. Comparison between the Markov Chain Hybrid Monte Carlo Method (MCHMCM), neural networks (NN), multiple regression analysis (MRA) and the Penalized Spline Semiparametric Method (PSSM).

Description	MCHMCM	NN	MRA	PSSM
Max overestimation (%)	10.00%	13.48%	40.02%	51.22%
Max underestimation (%)	14.67%	9.55%	129.33%	126.45%
Absolute average percentage error (%)	6.61%	7.33%	20.92%	23.16%
% of overestimation in the sample (>10%)	0.00%	21.05%	30.77%	26.15%
% of underestimation in the sample ($\leq$10%)	16.67%	0.00%	32.31%	32.31%
% of overestimation/underestimation in the sample (Total)	16.67%	21.05%	63.08%	58.46%

6. Conclusions

Real estate appraisals carried out with the Bayesian approach follow a probabilistic model that, not undergoing the traditional theoretical hypothesis of statistical inference (Gaussian noise with zero mean and constant standard deviation not dependent from input vectors), makes it possible to characterize the appraisals according to the real phenomena of the real estate market analysed.

With the aim of showing the forecasting potentialities of the Markov Chain Hybrid Monte Carlo Method (MCHMCM), it has been compared, using the same real estate sample, with the neural networks model (NNs), traditional multiple regression analysis (MRA) and the Penalized Spline Semiparametric Method (PSSM). From a general point of view all four methods have the same potentialities and application fields, but considering the small real estate sample, only MCHMCM and NN have shown good forecasting capacity with an absolute average percentage error of 6.61% and 7.33%, respectively.

MCHMCM has proved to be the best model. For this method the possibility to associate each prevision the corresponding confidence interval endogenously with the model represents an important appraisal advancement towards the improvement in the reliability of the real estate appraisals. From the point of view of model selection, it should be noted that there is the possibility to use a high number of regularization coefficients in order to obtain smoother interpolating functions with reduced computational costs if compared with traditional parametric approaches.

Further developments may concern the possibility to obtain, during the model assessment phase, information indicating the need to introduce further data in order to improve the prevision capability of the model itself (active learning). Besides, the possibility to consider the relative importance of several input vectors using different values of the hyperparameters for each input (automatic relevance determination) appears very interesting.

Author Contributions: This paper is to be attributed in equal parts to the authors.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Del Giudice, V.; De Paola, P.; Forte, F. Bayesian neural network models in the appraisal of real estate properties. In *ICCSA 2017, Part III, Lecture Notes in Computer Science*; Springer: Berlin, Germany, 2017; Volume 10406, pp. 478–489.
2. Antoniucci, V.; Marella, G. Immigrants and the City: The Relevance of Immigration on Housing Price Gradient. *Buildings* **2017**, *7*, 91. [CrossRef]
3. Brotman, B.A. Linear and Non Linear Appraisals Models. Available online: http://digitalcommons.kennesaw.edu/cgi/viewcontent.cgi?article=2293&context=facpubs (accessed on 12 November 2017).
4. Del Giudice, V.; De Paola, P. Spatial analysis of residential real estate rental market. In *Advances in Automated Valuation Modeling, Studies in System, Decision and Control*; d'Amato, M., Kauko, T., Eds.; Springer: Berlin, Germany, 2017; Volume 86, pp. 9455–9459.
5. Del Giudice, V.; De Paola, P.; Forte, F. The appraisal of office towers in bilateral monopoly's market: Evidence from application of Newton's physical laws to the Directional Centre of Naples. *Int. J. Appl. Eng. Res.* **2016**, *11*, 9455–9459.
6. Saaty, T.L.; De Paola, P. Rethinking Design and Urban Planning for the Cities of the Future. *Buildings* **2017**, *7*, 76. [CrossRef]
7. Del Giudice, V.; Manganelli, B.; De Paola, P. *Depreciation Methods for Firm's Assets, ICCSA 2016, Part III, Lecture Notes in Computer Science*; Springer: Berlin, Germany, 2016; pp. 214–227.
8. Manganelli, B.; Del Giudice, V.; De Paola, P. *Linear Programming in a Multi-Criteria Model for Real Estate Appraisal, ICCSA 2016, Part I, Lecture Notes in Computer Science*; Springer: Berlin, Germany, 2016; pp. 182–192.
9. Del Giudice, V.; De Paola, P.; Manganelli, B. *Spline Smoothing for Estimating Hedonic Housing Price Models, ICCSA 2015, Part III, Lecture Notes in Computer Science*; Springer: Berlin, Germany, 2015; pp. 210–219.
10. Del Giudice, V.; De Paola, P. Geoadditive Models for Property Market. *Appl. Mech. Mater.* **2014**, *584–586*, 2505–2509. [CrossRef]
11. Del Giudice, V.; de Paola, P. The Effects of Noise Pollution Produced by Road Traffic of Naples Beltway on Residential Real Estate Values. *Appl. Mech. Mater.* **2014**, *587–589*, 2176–2182. [CrossRef]
12. Del Giudice, V.; De Paola, P.; Manganelli, B.; Forte, F. The Monetary Valuation of Environmental Externalities through the Analysis of Real Estate Prices. *Sustainability* **2017**, *9*, 229. [CrossRef]
13. Del Giudice, V.; De Paola, P. Rough Set Theory for Real Estate Appraisals: An Application to Directional District of Naples. *Buildings* **2017**, *7*, 12. [CrossRef]
14. Del Giudice, V.; De Paola, P.; Cantisani, G.B. Valuation of real estate investments through Fuzzy Logic. *Buildings* **2017**, *7*, 26. [CrossRef]
15. Del Giudice, V.; De Paola, P.; Forte, F. Using Genetic Algorithms for Real Estate Appraisal. *Buildings* **2013**, *7*, 31. [CrossRef]
16. Del Giudice, V.; Manganelli, B.; De Paola, P. Hedonic analysis of housing sales prices with semiparametric methods. *Int. J. Agric. Environ. Inf. Syst.* **2017**, *8*, 65–77. [CrossRef]
17. Child, P.D.; Ott, S.H.; Riddiough, T.J. *Incomplete Information, Exercise Policy and Valuation of Claims on Noisy Real Assets*; MIT Real Estate Center: Cambridge, MA, USA, 1997.
18. Napoli, G.; Giuffrida, S.; Trovato, M.R.; Valenti, A. Cap Rate as the Interpretative Variable of the Urban Real Estate Capital Asset: A Comparison of Different Sub-Market Definitions in Palermo, Italy. *Buildings* **2017**, *7*, 80. [CrossRef]
19. Bishop, C.M. *Bayesian Methods for Neural Networks*; Technical Report; Department of Computer Science and Applied Mathematics, Aston University: Birmingham, UK, 1995.
20. Bishop, C.M. *Neural Network for Pattern Recognition*, 2nd ed.; Oxford University Press: New York, NY, USA, 1996.
21. Bishop, C.M. Neural Network: A Pattern Recognition Perspective. In *Handbook of Neural Computation*; Fiesler, E., Beale, R., Eds.; Oxford University Press: New York, NY, USA, 1996.
22. Bishop, C.M.; Qazac, C.S. Regression with input-dependent noise: A bayesian treatment. In *Advances in Neural Information Processing Systems 9*; MIT Press: Cambridge, MA, USA, 1997.
23. Lancaster, K.J. A new approach to consumer theory. *J. Political Econ.* **1966**, *74*, 132–156. [CrossRef]

24. Rosen, S. Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *J. Political Econ.* **1974**, *82*, 34–55. [CrossRef]
25. Freeman, A.M.I. *The Measurement of Environmental and Resource Values*; Resources for the Future: Washington, DC, USA, 1999.
26. Lozano-Gracia, N.; Anselin, L. Is the price right? Assessing estimates of cadastral values for Bogotá, Colombia. *Reg. Sci. Policy Pract.* **2012**, *4*, 495–508. [CrossRef]
27. Peterson, S.; Flanagan, A.B. Neural network hedonic pricing models in mass appraisal. *J. Real Estate Res.* **2009**, *31*, 147–164.
28. Tsukuda, J.; Baba, S. Predicting Japanese corporate bankruptcy in terms of financial data using neural network. *Comput. Ind. Eng.* **1994**, *27*, 445–448. [CrossRef]
29. Do, A.Q.; Grudnitski, G. A Neural Network Approach to Residential Property Appraisal. *Real Estate Apprais.* **1992**, *58*, 38–45.
30. Tay, D.P.H.; Ho, D.K.H. Artificial Intelligence and the Mass Appraisal of Residential Apartments. *J. Prop. Valuat. Invest.* **1992**, *10*, 525–540. [CrossRef]
31. Huang, C.-S.; Dorsey, R.E.; Boose, M.A. Life Insurer Financial Distress Prediction: Neural Network Model. *J. Insur. Regul.* **1994**, *13*, 131–167.
32. Allen, W.C.; Zumwalt, J.K. *Neural Networks: A Word of Caution*; Unpublished Working Paper; Colorado State University: Fort Collins, CO, USA, 1994.
33. Worzala, E.; Lenk, M.; Silva, A. An Exploration of Neural Networks and Its Application to Real Estate Valuation. *J. Real Estate Res.* **1995**, *10*, 185–201.
34. Guan, J.; Zurada, J.; Levitan, A.S. An Adaptive Neuro-Fuzzy Inference System Based Approach to Real Estate Property Assessment. *J. Real Estate Res.* **2008**, *30*, 395–422.
35. Nguyen, N.; Cripps, A. Predicting Housing Value: A Comparison of Multiple Regression Analysis and Artificial Neural Networks. *J. Real Estate Res.* **2001**, *22*, 313–336.
36. Lam, K.C.; Yu, C.Y.; Lam, K.Y. An Artificial Neural Network and Entropy Model for Residential Property Price Forecasting in Hong Kong. *J. Prop. Res.* **2008**, *25*, 321–342. [CrossRef]
37. McCord, J.; McCord, M.J.; McCluskey, W.; Davis, P.; McIhatton, D.; Haran, M. Belfast's iron(ic) curtain. *J. Eur. Real Estate Res.* **2013**, *6*, 333–358. [CrossRef]
38. McCord, J.; McCord, M.J.; McCluskey, W.; Davis, P.; McIhatton, D.; Haran, M. Effect of public green space on residential property values in Belfast metropolitan area. *J. Financ. Manag. Prop. Constr.* **2014**, *19*, 117–137. [CrossRef]
39. Zurada, J.; Levitan, A.; Guan, J. A comparison of regression and artificial intelligence methods in a mass appraisal context. *J. Real Estate Res.* **2011**, *33*, 349–387.
40. Tajani, F.; Morano, P.; Locurcio, M.; D'Addabbo, N. Property Valuations in Times of Crisis: Artificial Neural Networks and Evolutionary Algorithms in Comparison. *Lect. Notes Comput. Sci.* **2015**, 194–209. [CrossRef]
41. Morano, P.; Tajani, F. Bare ownership evaluation. Hedonic price model vs. artificial neural network. *Int. J. Bus. Intell. Data Min.* **2013**, *8*, 340. [CrossRef]
42. Antoniucci, V.; Marella, G. Small town resilience: Housing market crisis and urban density in Italy. *Land Use Policy* **2016**, *59*, 580–588. [CrossRef]
43. Ding, L. Process Improvement Study of BP Neural Networks Model for Financial Crisis Prediction Based on Chinese Real Estate Listed Enterprises. *J. Inf. Comput. Sci.* **2013**, *10*, 5773–5786. [CrossRef]
44. Garson, G.D. Interpreting Neural-Network Connection Weights. *Artif. Intell. Expert* **1991**, *6*, 47–51.
45. Intrator, O.; Intrator, N. Interpreting Neural-Network Results: A Simulation Study. *Comput. Stat. Data Anal.* **2001**, *37*, 373–393. [CrossRef]
46. Hornik, K.; Stinchcombe, M.; White, H. Multilayer feedforward networks are universal approximators. *Neural Netw.* **1989**, *2*, 359–366. [CrossRef]
47. Bottou, L. Large-scale machine learning with stochastic gradient descent. In Proceedings of the COMPSTAT'2010, Paris, France, 22–27 August 2010; pp. 177–186.
48. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, *323*, 533–536. [CrossRef]
49. MacKay, D.J.C. Bayesian interpolation. *Neural Comput.* **1992**, *4*, 415–447. [CrossRef]
50. MacKay, D.J.C. The evidence framework applied to classification networks. *Neural Comput.* **1992**, *4*, 720–736. [CrossRef]

51. Neal, R.M. *Bayesian Training of Back-Propagation Networks by the Hybrid Monte Carlo Method*; Technical Report; Department of Computer Science, University of Toronto: Toronto, ON, Canada, 1993.
52. Neal, R.M. *Probabilistic Inference Using Markov Chain Monte Carlo Method*; Technical Report; Department of Computer Science, University of Toronto: Toronto, ON, Canada, 1993.
53. Neal, R.M. Bayesian Learning for Neural Network. Ph.D. Thesis, University of Toronto, Toronto, ON, Canada, 1994.
54. Jylanki, P.; Nummenmaa, A.; Vehtari, A. Expectation propagation for neural networks with sparsity-promoting priors. *J. Mach. Learn. Res.* **2014**, *15*, 1849–1901.
55. Hinton, G.E.; van Camp, D. Keeping the neural networks simple by minimizing the description length of the weights. In Proceedings of the Sixth Annual Conference on Computational Learning Theory (COLT'93), Santa Cruz, CA, USA, 26–28 July 1993.
56. Graves, A. Practical variational inference for neural networks. *Adv. Neural Inf. Process. Syst.* **2011**, *24*, 2348–2356.
57. Soudry, D.; Hubara, I.; Meir, R. Expectation backpropagation: Parameter-free training of multilayer neural networks with continuous or discrete weights. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 963–971.
58. Jaynes, E.T. Bayesian Methods: General Background. In *Maximum Entropy and Bayesian Methods in Applied Statistics*; Justice, H.D., Ed.; Cambridge University Press: Cambridge, UK, 1986.
59. Fukunaga, K. *Introduction to Statistical Pattern Recognition*, 2nd ed.; Academic Press: San Diego, CA, USA, 1990.
60. Jordan, M.L.; Bishop, C.M. Neural Networks. *Handbook of Computer Science*; Tucker, A., Ed.; CRC Press: Boca Raton, FL, USA, 1996.
61. Green, S.B. How Many Subjects Does It Take to Do a Regression Analysis. *Multivar. Behav. Res.* **1991**, *26*, 499–510. [CrossRef] [PubMed]
62. Marks, M.R. Two kinds of regression weights that are better than betas in crossed samples. In Proceedings of the American Psychological Association, Toronto, ON, Canada, 8–11 August 1996.
63. Tabachnick, B.G.; Fidell, L.S. *Using Multivariate Statistics*, 6th ed.; Pearson: Boston, MA, USA, 2012.
64. Harris, R.J.; Quade, D. The Minimally Important Difference Significant Criterion for Sample Size. *J. Educ. Behav. Stat.* **1992**, *17*, 27–49. [CrossRef]
65. Schmidt, F.L. Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers. *Psychol. Methods* **1996**, *1*, 115–129. [CrossRef]
66. Harrell, F.E., Jr. *Regression Modeling Strategies*; Springer-Verlag: New York, NY, USA, 2001.
67. Del Giudice, V. *Estimo e Valutazione Economica dei Progetti*; Paolo Loffredo Editore: Napoli, Italy, 2015.
68. Ciuna, M.; Milazzo, L.; Salvo, F. A Mass Appraisal Model Based on Market Segment Parameters. *Buildings* **2017**, *7*, 34. [CrossRef]
69. Napoli, G. Housing Affordability in Metropolitan Areas. The Application of a Combination of the Ratio Income and Residual Income Approaches to Two Case Studies in Sicily, Italy. *Buildings* **2017**, *7*, 95. [CrossRef]
70. Buntine, W.L.; Weigend, A.S. Bayesian back-propagation. *Complex Syst.* **1991**, *5*, 603–643.
71. Duane, S.; Kennedy, A.D.; Pendeleton, B.J.; Roweth, D. Hybrid Monte Carlo. *Phys. Lett. B* **1987**, *195*, 216–222. [CrossRef]
72. William, C.K.I.; Qazac, C.; Bishop, C.M.; Zhu, H. On the relationship between bayesian error bars and the input data density. In Proceedings of the 4th International Conference on Artificial Neural Networks, Cambridge, UK, 26–28 June 1995.
73. Evans, A.; James, H.; Collins, A. Artificial Neural Networks: An application to residential valuation in the UK. *J. Prop. Valuat. Invest.* **1992**, *11*, 195–204.
74. The R Project for Statistical Computing. Available online: http://www.R-project.org (accessed on 12 November 2017).