Linear Programming in a Multi-Criteria Model for Real Estate Appraisal

Benedetto Manganelli[1(✉)], Pierfrancesco De Paola[2],
and Vincenzo Del Giudice[2]

[1] University of Basilicata, Viale dell'Ateneo Lucano, 85100 Potenza, Italy
benedetto.manganelli@unibas.it
[2] University of Naples "Federico II", Piazzale Vincenzo Tecchio,
80125 Naples, Italy
pfdepaola@libero.it, vincenzo.delgiudice@unina.it

Abstract. In real estate appraisal, research has long been addressed to the experimentation of multi-parametric models able to reduce the margin of error of the estimate and to overcome or to limit, as far as possible, the problems and difficulties that the use of these models often involves. On the one hand, researchers are trying to overcome the essentially deductive approach that has characterized the traditional discipline, and on the other, to minimize the problems arising from a merely inductive approach. The real estate market is characterized by an inelastic supply and by properties whose complexity and differentiation often involve, also and especially on the demand side, subjective and psychological elements that could distort the results of an inductive investigation. This problem can be overcome by increasing the size of the survey sample, and by using statistical analysis. Statistical analyses, however, are often based on very strong assumptions. A multi-criteria valuation model that uses linear programming is applied to the real estate market. The model, integrated with the inductive and deductive approach, exceeds many of the assumptions of the best known statistical approaches.

Keywords: Linear programming · Multi-criteria valuation model · Real estate market

1 Introduction

The multi-parametric models in real estate appraisal that use quantitative data analysis, can be divided into two groups: (a) those based on statistical techniques, such as multiple regression analysis [1–5], neural network [6–11], genetic algorithms [12, 13], and (b) those using only mathematical processing, such as structural equation systems [14, 15] and rough set theory [16], UTA [17, 18]. The statistical and mathematical approaches are different, of course, not so much for the content but for the way of thinking of Statistics and the way of thinking of Mathematics. The fundamental nature that distinguishes the two approaches is that Statistics is an inductive discipline, while Mathematics is rather deductive.

The model applied in this study is of the second type [19]; it uses linear programming techniques [20–22] and derives its theoretical basis from tools developed as part of decision theory and operational research.

The decision-making process, approached with multi-criteria analysis, formally can be summarized in an evaluation matrix. The columns of the matrix represent the alternatives, while the rows describe the evaluation criteria. The values in each cell of the matrix are the attributes, namely the qualitative or quantitative level reached by each alternative for each criterion. In turn, the appraisal process involves the comparison of the property to be estimated with a sample of which the selling prices are known. This comparison is done by measuring the difference between the property characteristics taken into consideration among those that contribute to the formation of the value. The estimate may thus derive from the solution of a system of equations that has the following expression:

$$s = D - 1 \cdot p,$$

where s is the vector of unknowns, namely, the property value and marginal prices of the characteristics, p is the price vector and D is the matrix of the differences between the characteristics. Therefore, it is evident the analogy with a decision-making process where, in this case, the alternatives are the property of the sample and the criteria are the characteristics used for the comparison.

The search for a model able to reproduce the actual decision-making of market participants (supply and demand) as best as possible, has led many researchers to propose the application, in the field of real estate appraisal, of procedures borrowed from decision theory. These analysis techniques approximate the utility functions of the players in the housing market. The utility function must be understood as that which describes the marginal price of the property characteristic.

Researchers, in an attempt to retrace the decision-making process in a real way, have tried to increase the flexibility of the models using non-linear functions. On the other hand, efforts have been directed to the development of models able to limit the negative impact on the results due to the presence of strongly correlated explanatory variables.

This model interprets the process of the price's formation in the same way as a multi-criteria choice [23], with a multi-objective approach, where the features of the properties that the market considers to be essential represent the selection criteria.

It is therefore a multi-equation model, but unlike other models of the same type, the equations have no endogenous variables except for those that each equation tries to explain. This is why no problem arises for the identification and simultaneous estimation of the parameters. The model equations describe the contributions of the features taken into account in the estimation process of the market value.

The contributions are integrated into a single function-price additive [24]. Because the price variable is expressed as the sum of univariate functions, one for each feature of the property, the model is able to obviate the typical statistical approach problem, relative to the size of the sample data, which is a function of the number of variables taken into consideration. The problem, known as curse of dimensionality requires, for example, that in a multivariate regression model, the amount of data required to maintain the same statistical accuracy, grows more rapidly than the number of variables

taken into consideration. Hence, compared with multiple linear regression models or other models using multivariate analysis, its basic assumptions appear much weaker. Moreover, the model allows to impose limitations on the functions that describe the individual contributions. The operator can deductively impose whether the contribution is positive or negative.

In deductive logic, it is also possible to assign piecewise-defined functions. In this way, preserving the simplicity of linear forms, the marginal contribution of individual features can better adapt to the economic logic or to the very special conditions of the housing market. You can thus take into account some economic principles as the law of diminishing marginal utility (with the increase of the internal floor area of the units, a reduction in the marginal price is expected); or you can adapt the model to the actual trend of some observable phenomena (for example the change in the sign function of the price based on the floor level, given that the intermediate levels generally have the highest values).

This paper is structured as follow. The following section provides the model description, focusing specially on the way in which to set the constraints, able to impose the shape of utility functions that describe marginal prices. Section 3 presents the case study, the real estate properties used for comparison and the constraints imposed. The last section illustrates and discusses the results also comparing them to those obtained using other approaches.

2 The Mathematical Formalism of the Model

In the mathematical formalism of the model $A = \{i, 1 \leq i \leq m\}$ is the set of m units of the sample, $C = \{j, 1 \leq j \leq n\}$ describes the n criteria (features) that identify the units, chosen from among those that the market considers to be most significant. Given these two sets, for each criterion j, V_{ij} is the score, that is, the numerical value of the generic element of the set A (housing unit). A prerequisite for the development of the analytical model is for the scores to always be greater than zero ($V_{ij} > 0$). Scores assigned to the units V_{ij}, for each of the selected criteria j, contribute to the formation of the sale price of a property. This contribution is represented under the symbol U_{ji}; it can be positive or negative and is expressed as a linear piecewise-defined function that binds it to V_{ij} score based on criterion j. For this purpose, the range of non-null values of the j criterion should be split into subintervals T_j (integer), constructed so that the elements of each are not present in another and that the set of elements of all the subintervals contains all measured values of the j criterion. Once the subintervals have been defined, for each of those relative to the j-th criterion, the upper D_{tj}^+ limit and the lower D_{tj}^- limit shall be indicated; where t is the generic subinterval between all the subintervals T_j. The symbols α e W indicate, respectively, the constant and the angular coefficient of a generic linear function. The marginal evaluation function of the j criterion then assumes the following expression:

$$U_{ij}^t = f_j^t(V_{ij}) = \begin{cases} 0, & \text{if } V_{ij} = 0 \\ \alpha_{tj} + V_{ij}W_{tj} & \text{if } V_{ij} \neq 0 \text{ and } D_{tj}^- \leq V_{ij} \leq D_{tj}^+ \end{cases}$$

The piecewise linear representation of the function expressing the contribution of the j criterion, provides an approximation of the probable non-linear function that could possibly represent the relationship between price and scoring of different characteristics. The subdivision into sub-intervals or sections therefore depends on the nature of the criterion adopted, and should be an expression of real elasticity of market prices towards changes in the values of the measured scores for the j-th criterion. The indispensable condition for the division into sections is that for every defined section there correspond a number of observations, sufficient to provide a representation of the function. If due to the peculiarities of the survey sample, the number of discrete values of the scores relative to a given criterion is poor, the problem can be resolved by considering each value as a section. The price function of the i-th property is constructed as an additive sum of the individual contributions which are obtained with respect to each j criterion (property feature). Its analytical form is:

$$f\left(U_{ij}, \ldots \ldots, U_{in}\right) = U_o + \sum_{j=1}^{n} U_{ij}$$

The d_i^- and d_i^+ symbols respectively indicate the negative and positive residual. One of them of course will be null. These residues are expressed in absolute value as the difference between the observed price pi and the estimated value for the i-th property:

$$Pr_i - \left[U_o + \sum_{j=1}^{n} U_{ij}\right] = d_i = \begin{cases} d_i^+ & \text{if } d_i \geq 0 \\ d_i^- & \text{if } d_i < 0 \end{cases}$$

DA is the sum of the residues, weighted on relative prices observed.

$$\sum_{i \in A} \frac{1}{Pr_i} \left(d_i^- + d_i^+\right)$$

The model is developed on the minimum calculation of DA function and respecting the constraints imposed.

$$Min(DA)$$

with the following constraints

$$U_o + \sum_{j=1}^{n} U_{ij} - d_i^+ + d_i^- = Pr_i, \quad d_i^+ \geq 0, \quad d_i^- \geq 0 \quad \text{with } i \in A$$

$$f\left(D_{tj}^+\right) \leq f\left(D_{t+1,j}^-\right) \quad \text{for} \quad 1 \leq t \leq T_j - 1, \quad j \in C^+$$

$$W_{tj} \geq 0 \text{for} \quad \text{for} \quad 1 \leq t \leq T_j, \quad j \in C^+$$

$$f\left(D_{tj}^+\right) \geq f\left(D_{t+1,j}^-\right) \quad \text{for} \quad 1 \leq t \leq T_j - 1, \quad j \in C^-$$

$$W_{tj} \leq 0 \text{for} \quad \text{for} \quad 1 \leq t \leq T_j, \quad j \in C^-$$

3 The Case Study

The data used for the study are those already elaborated in a previous research [25], specifically 148 sales of residential property units located in a central district of a city of the Campania region, i.e. in a homogeneous market area with identical extrinsic characteristics, over a period of eight years (Tables 1, 2).

Table 1. Variable description

Variable	Description
Price observed (pr)	expressed in thousands of Euros
Age of the property (age)	expressed retrospectively in no. of years
Date of sale (date)	expressed retrospectively in no. of months
Internal area (int)	expressed in sqm
Balconies area (balc)	expressed in sqm
Connected area (conn)	expressed in sqm (lofts, cellars, etc.)
Number of services (serv)	no. of services in residential unit
Number of views (views)	no. of views on the street
Maintenance (main)	expressed with dichotomous scale(1 or 0, respectively, for the presence or absence of optimal maintenancestate)
Floor level (f_lev)	no. of floor levels of residential unit

Table 2. Statistical description of variables

Variable	Std. Dev.	Median	Mean	Min	Max
Pr	32.134,11	114.585	113.502	41.316	250.000
age	7,40	24	23,51	10	35
date	23,77	33	37,30	2	96
int	25,88	117	118,80	48	210
balc	10,19	16	17,56	0	59
conn	15,78	0	9,53	0	62
serv	0,52	2	1,61	1	3
views	0,60	2	2,27	1	4
man	0,33	0	0,12	0	1
f_lev	1,63	3	3,16	1	7

Table 3. Sections of variation of the scores

criterion	number of sections	Range of variation
age	3	[10, 20]; [21, 30]; [31,35];
date	5	[0, 12]; [13, 36]; [37, 60]; [61, 74]; [75, 96]
int	5	[48, 60]; [61, 100]; [101, 130; [131, 170]; [171, 210]
bal	3	[0,15]; [16, 30]; [31, 60];
conn	3	[0, 20]; [21, 40]; [41, 62];
serv	3	[1,1]; [2, 2]; [3, 3]
views	3	[1,1]; [2, 2]; [3, 3]
man	2	[0,1];
f_level	3	[1, 2]; [3, 5]; [6, 7];

Table 3 describes the sub-intervals of variation of the scores assigned to each criterion.

The empirical knowledge of the likely contribution to the price of the selected features has enabled the following choices: for the age variable the function is constrained to be decreasing; for the variables related to the surfaces, the facilities, the

Table 4. Coefficients of the piecewise functions

Uo =	-12.339,38					
section (t)	1		2		3	
U = a + VW	a	W	a	W	a	W
age	16221,39	-811,07	0,00	0,00	0,00	0,00
date	31294,90	0,00	39826,11	-656,25	19300,76	-83,77
int	-35436,66	738,26	-45779,67	823,57	43214,95	0,00
bal	3525,93	895,61	16960,02	0,00	-3825,78	670,51
conn	929,83	775,20	15443,66	51,14	4886,67	307,38
serv	0,00	0,00	1770,63	0,00	1770,63	0,00
views	0,00	0,00	2762,04	0,00	5330,64	0,00
man	0,00	0,00	2402,50	0,00		
f_level	30455,22	1173,27	26901,20	1475,65	46107,88	-2750,55

section (t)	4		5	
U = a + VW	a	W	a	W
date	56297,18	-771,19	6781,24	0,00
int	-53335,69	791,25	104973,30	0,00

views and the maintenance, the functions are assumed to increase, while the functions related to the floor level and the date of sale are free from constraints.

4   The Results

The Lp_solve software is used to analyse the real estate data. Lp_solve is a free Mixed Integer Linear Programming (MILP) solver.

Table 4 shows the coefficients defining the piecewise functions of the individual criteria.

The residue analysis indicates that the model has good predictive ability. The average percentage error is 7.13 %. The predictive power of the model is therefore higher than the multiple regression analysis (MRA) 7.84 % but slightly lower compared to a semi-parametric regression method based on Penalized Spline Smoothing, 6.47 %. The comparison between the residuals of these three models, however, shows that the one proposed in this study has the best performance on a percentage of 85 % of the sample in comparison with the MRA and on a percentage of 68 % compared to the non-linear regression model (Fig. 1).
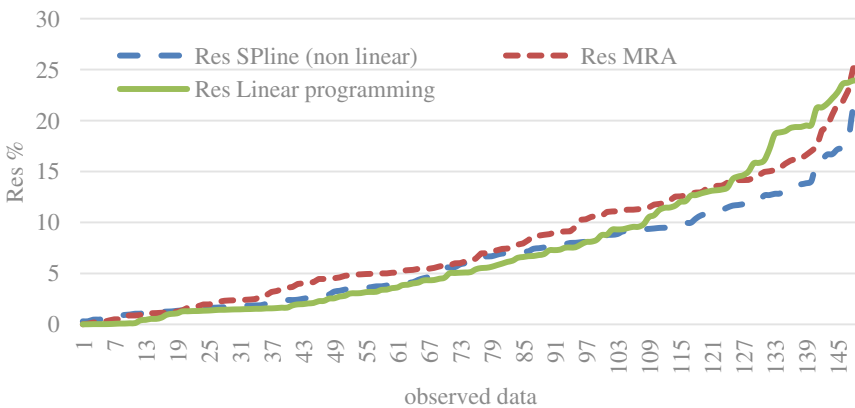

Fig. 1. Residue analysis (Color figure online)

The graphs in Figs. 2, 3, 4 and 5 show the marginal value piecewise functions of some features.

For a better understanding of the graphs, when the piecewise functions have the steps, the moving average is calculated.

The results are consistent with the expectations, especially with regard to those parameters that are not constrained. The law of diminishing marginal utility is respected (Figs. 2, 4), and the variation of the marginal contribution of the floor level is consistent, first increasing then decreasing (Fig. 5). Even the marginal contribution on the date of sale reflects the dynamics of the market observed by official Observatories of the real estate market (Fig. 3).
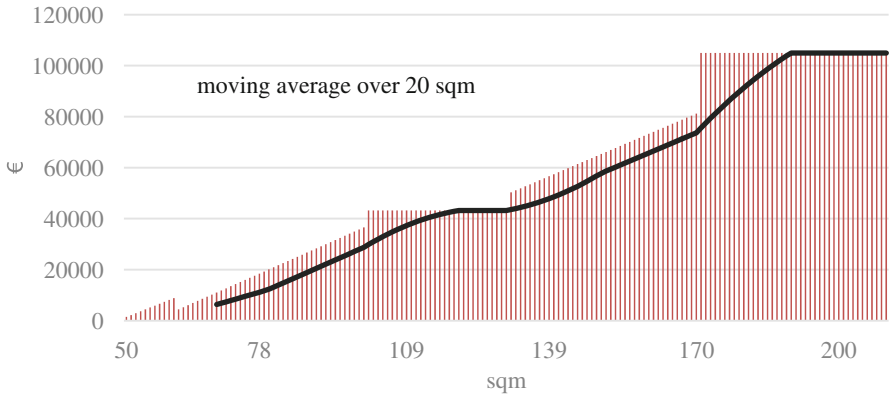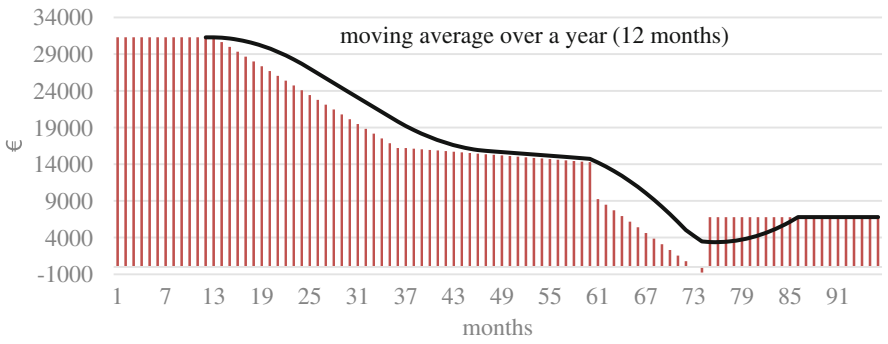
Fig. 2. Value Piecewise functions of *internal area*
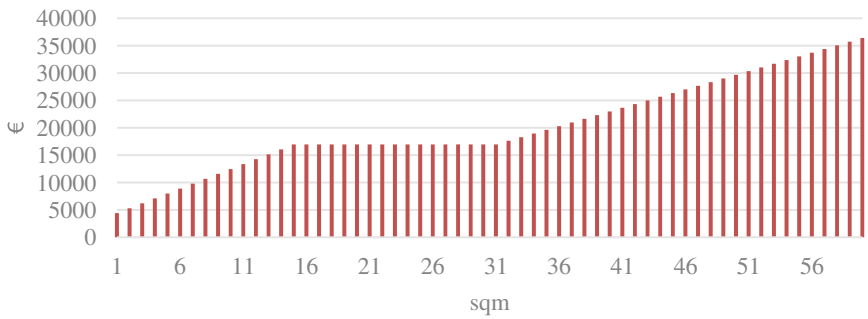

Fig. 3. Value Piecewise functions of *date of sale*

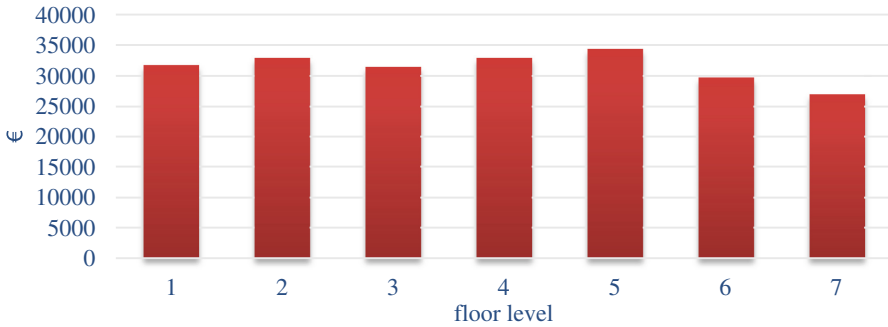**Fig. 4.** Value Piecewise functions of *balconies area*

Fig. 5. Value Piecewise functions of *floor level*

5 Conclusion

The model applied in this study, is an effective real estate estimation tool. It shows greater confidence in the results compared to the linear statistical analysis models. On the one hand, it retains the advantages of a linear approach; on the other hand, it adds some operating advantages resulting from the reduction of the basic assumptions. The improved reliability comes from its ability to interpret the investigated phenomenon, due to an approach in which the deductive component plays a decisive role.

A priori knowledge of the phenomenon and the evaluator's experience with adequate perception of the market mechanism, allows to set some constraints that lead the inductive analysis results within a predetermined track. One of the strengths of the model lies precisely in the possibility, not necessarily conceived as an exercisable option, of connecting the estimate to the indications that the deductive analysis provides about the shape and/or direction of the individual functions of the marginal contributions.

The essential feature in inductive reasoning is its generalizability, however, generalization is effective when the sample is very large and representative of the population. The complexity of the real estate market makes the construction of this type of sample difficult. The model proposed by integrating the inductive analysis with a deductive approach overcomes the limitations of the statistical analysis.

The marginal prices that easily derived from the piecewise functions may be used as benchmarks even in the estimates, very frequent in practice, which are based on a reduced number of comparable units.

References

1. Isakson, H.R.: Using multiple regression analysis in real estate appraisal. Appraisal J. **69**(4), 424–430 (2001)
2. Manganelli, B., Pontrandolfi, P., Azzato, A., Murgante, B.: Using geographically weighted regression for housing market segmentation. Int. J. Bus. Intell. Data Min. **9**(2), 161–177 (2014)

3. Manganelli, B., Pontrandolfi, P., Azzato, A., Murgante, B.: Urban residential land value analysis: the case of Potenza. In: Murgante, B., Misra, S., Carlini, M., Torre, C.M., Nguyen, H.-Q., Taniar, D., Apduhan, B.O., Gervasi, O. (eds.) ICCSA 2013, Part IV. LNCS, vol. 7974, pp. 304–314. Springer, Heidelberg (2013)
4. Newsome, B.A., Zietz, J.: Adjusting comparable sales using multiple regression analysis - the need for segmentation. Appraisal J. **60**, 129–135 (1992)
5. Simonotti, M., Salvo, F.: Ciuna, M,: Multilevel methodology approach for the construction of real estate monthly index numbers. J. Real Estate Lit. **22**(2), 281–302 (2015)
6. Do, Q., Grudnitski, G.: A neural network approach to residential property appraisal. The Real Estate Appraiser, pp. 38–45, December 1992
7. Evans, A., James, H., Collins, A.: Artificial neural networks: an application to residential valuation in the UK. J. Property Valuation Investment **11**(2), 195–203 (1992)
8. Tay, D., Ho, D.: Artificial intelligence and the mass appraisal of residential apartments. J. Property Valuation Investment **10**(2), 525–540 (1993)
9. Worzala, E.M., Lenk, M.M., Silva, A.: An exploration of neural networks and its application to real estate valuation. J. Real Estate Res. **10**(2), 185–201 (1995)
10. Peterson, S., Flanagan, A.: Neural network hedonic pricing models in mass real estate appraisal. J. Real Estate Res. **31**(2), 147–164 (2009)
11. Tajani, F., Morano, P., Locurcio, M., D'Addabbo, N.: Property valuations in times of crisis: artificial neural networks and evolutionary algorithms in comparison. In: Gervasi, O., Murgante, B., Misra, S., Gavrilova, M.L., Rocha, A.M.A.C., Torre, C., Taniar, D., Apduhan, B.O. (eds.) ICCSA 2015. LNCS, vol. 9157, pp. 194–209. Springer, Heidelberg (2015)
12. Morano, P., Tajani, F., Locurcio, M.: Land use, economic welfare and property values: an analysis of the interdependencies of the real estate market with zonal and socio-economic variables in the municipalities of the Region of Puglia. Int. J. Agric. Environ. Inf. Syst. **6**(4), 16–39 (2015)
13. Manganelli, B., De Mare, G., Nesticò, A.: Using genetic algorithms in the housing market analysis. In: Gervasi, O., Murgante, B., Misra, S., Gavrilova, M.L., Rocha, A.M.A.C., Torre, C., Taniar, D., Apduhan, B.O. (eds.) ICCSA 2015. LNCS, vol. 9157, pp. 36–45. Springer, Heidelberg (2015)
14. Manganelli, B.: Un sistema di equazioni strutturali per la stima di masse di immobili. Genio Rurale, 2 (2001)
15. Bravi, M., Fregonara, E.: Structural equations models in real estate appraisal. In: International Real Estate Conference - American Real Estate and Urban Economics Association, Orlando, 23–25 May 1996
16. D'Amato, M.: Comparing rough set theory with multiple regression analysis as automated valuation methodologies. Int. Real Estate Rev. **10**(2), 42–65 (2007)
17. Jacquet-Lagreze, E., Siskos, Y.: Assessing a set of additive utility functions for multicriteria decision-making, the UTA Method. Eur. J. Oper. Res. **10**(2), 151–164 (1982)
18. Siskos, Y., Yannacopoulos, D.: UTASTAR: an ordinal regression method for building additive value functions. Investigacao Operacional **5**(1), 39–53 (1985)
19. Kettani, O., Oral, M., Siskos, Y.: A multiple criteria analysis model for real estate evaluation. J. Global Optim. **12**(2), 197–214 (1998)
20. Oral, M., Kettani, O.: A mathematical programming model for market share prediction. Int. J. Forecast. **5**(1), 59–68 (1989)
21. Pekelman, D., Subrata, K.S.: Mathematical programming models for the determination of the attribute weights. Manage. Sci. **20**, 1217–1229 (1974)
22. Sivrinisan, V., Shocker, A.D.: Linear programming techniques for multidimensional analysis of preferences. Psychometrica **38**(3), 337–369 (1973)

23. Oral, M., Kettani, O.: Modeling the process of multiattribute choice. J. Oper. Res. Soc. **40** (3), 281–291 (1989)
24. Pace, K.: Appraisal using generalized additive models. J. Real Estate Res. **15**(1), 77–99 (1998)
25. Del Giudice, V., Manganelli, B., De Paola, P.: Spline smoothing for estimating hedonic housing price models. In: Gervasi, O., Murgante, B., Misra, S., Gavrilova, M.L., Rocha, A. M.A.C., Torre, C., Taniar, D., Apduhan, B.O. (eds.) ICCSA 2015. LNCS, vol. 9157, pp. 210–219. Springer, Heidelberg (2015)