Rough Set Theory as Automated Valuation Methodology: The Whole Story

Maurizio d'Amato
Associate Professor
1st Faculty of Engineering
Technical University Politecnico di Bari - Italy
madamato@interfree.it

What information consumes is rather obvious: it consumes the attention of its recipients.
Hence a wealth of information creates a poverty of attention, and
a need to allocate that attention efficiently among the overabundance of information sources that might consume it.
(*H.Simon,(1971) Computers, Communications and the Public Interest, pages 40-41, Martin Greenberger, ed., The Johns Hopkins Press*)

Introduction

Mass appraisal may be defined as a systematic valuation of groups of properties using standardized procedures. Mass appraisal methodologies refer to large groups of properties rather than to a single property. The accurate valuation of a predefined group of properties, or one particular property, belonging to a group, using a model, for a given practical purpose, is the issue of these methodologies (e.g. McCluskey et al. 1997; Gonzales et al. 2002a,b). Among the Heretic approaches to mass appraisal there is Rough Set Theory. (Kauko and d'Amato,2004)

This work represents a summary of the most important evolution from a methodological and empirical points of view of this valuation methodology. RST is a mass appraisal method recently proposed by the author of this contribution. It has been applied for the first time to a small sample of real property transactions in the residential real estate market of Bari (d'Amato,2002). Therefore was applied again to a larger sample of transactions of residential properties in Bari (d'Amato, 2004a) and later to a sample of 100 observations in Amsterdam (d'Amato,2004b).The last application has been to 600 observations in the downtown of Helsinki (d'Amato,2007).

The empirical applications showed interesting results, quite closer to MRA. After five years of empirical testing it is possible to observe two different applications of RST for mass appraisal purposes. The former has been explored in the first article on RST for this reason this approach will be described here as "article 2002[1]". In this case the estimated price is a class instead of a crisp value. The latter is based on the integration between RST and Value Tolerance Relation. This integration allows the appraiser to obtain a crisp value. The use of Value Tolerance relation has been explored in three articles from 2003 till now. As the last article will be published in the next year the approach will be indicated as "articles from 2003-2007". In this work a different integration will be proposed between RST and Value Tolerance relation for valuation purposes.

The paper is organized as follows: the first section will show why we need "flexible models" (even if heretic) of mass appraising such as RST. Therefore the theoretical basis of Rough Set Theory and its integration with Value Tolerance Relation will be addressed, some trivial example will help the reader to get used to this method; in the second section a comparison between MRA and RST will be carried out using a sample of 7000 observations referred to single family residential properties located in Catawba County in North Carolina (US). Final remarks and future directions of research will be indicated at the end.

1.1 An Introduction to Rough Set Theory and Value Tolerance Relation

Before introducing RST as automated valuation method it is important to stress the origin of the choice to investigate alternative and heretic methods alternative to MRA. In order to make clear the theoretical premise of RST application to real estate valuation it is important to consider that a valuer normally tries to foresee property market price. Those are originated by a decision making process which involve both sellers and buyers.
Economic analysis of decision processes is dominated by two theoretical approaches. The former is neoclassical one whose mathematical foundations were originally proposed by Arrow and Debreu (1954), and the one started by Simon's contributions on bounded rationality and problem-solving (Simon,1957; 1979;1981).

---

[1] It should be defined article 2001 because the article was based on a paper submitted to PRRES 2001 meeting in Christchurch.

In Arrow-Debreu world rationality is identified with a methodological strategy that consists of solving maximization problems. Simon has strongly criticised this approach by pointing out the binding and unrealistic hypothesis. In particular, it is calibrated on a representative agent with unlimited computational and predictional abilities, while the human mind, has serious structural limitations to their "power". Human mind performances cannot fit the standards of such "olympic" perfection and consequently the economic agent, when faced with a problem, will find solutions commonly not optimal but "satisficing" according to his subjective and modifiable aspiration level using a "bounded rationality".

The application of hedonic price theory (Griliches 1971; Rosen 1974) is based on assumptions of general equilibrium, and the driving logic is based on *homo oeconomicus* behaviour in a static framework. Therein lies the weakness of this approach. Its underlying assumptions, notably the smooth, continuous and linear relationships between the variables under study, and (in the economic sense) rational behaviour of the buyer and seller, may be not always realistic, as the market operates within a variety of constrains, and the individual market actors suffer from inconsistency and idiosyncracy, as well as information and power imbalances.

The empirical case will be discussed later of Catwba County property transactions will consider the causal relationship between the property market price and a set of property features composed by XX variables. An important question is the following: did the buyers and the sellers take really into account all these variables when they bought and they sold those properties? How many buyers and sellers were really concerned about the number, the quality of technical features of their properties? Do we live in a world where people drive the car without knowing neither the name nor how the parts of their car works. The computer that is helping me to write this paper has some parts that I do not know and it would be difficult to me to explain exactly how they work. If this is the real world we live, probably we should consider the eventuality that the human behaviour on which is based the relationship between price and property features, may be driven by a "bounded rationality" instead of an "olympic rationality". If this can be consider correct, although heretic, the contribution of approaches different from MRA may be consider helpful. This is the necessary premise of this paper that tries to show the evolution in the application of RST as automated valuation method .

1.1.1 Rough Set Theory: The original idea (article 2002)

Looking for a different causal (less strong) relationship between property price and characteristics I had the opportunity to read some earlier works written by a polish mathematician (Pawlak,1982; Pawlak,1991). This method has been applied in several fields and define causal relationship between attributes of a object through if then rules (boolean algebra). In the RST vision "The information about a decision is usually vague because of uncertainty and imprecision coming from many sources…Vagueness may be caused by granularity of representation of the information. Granularity may introduce an ambiguity to explanation or prescription based on vague information,,," (Pawlak and Slowinski,1993)

The vagueness and imprecision problems are even present in the information that is going around in the real estate market whereas its information, in same case, maybe unprecise or perceived as unprecise and often asymmetric. In the real estate market the single element or object is the single real estate transaction which is part of a universe of real estate transactions which compose the property market.

In the real estate market a piece of information is based on a real estate transaction and has several attributes, such as the price, the technical characteristics, the tenant characteristic, etc.. All these attributes may be referred to the property on which is based the transaction. In term of RST all these characteristics are defined attributes. If a real estate transaction is considered as an element, the only available information is the specific characteristics (attributes) related to the property and the price.

These characteristics (attributes) may be owned or not by an object set. The relation between an object and its characteristics can be described by three regions of knowledge "Certainly, Possibly and Certainly not". For instance, in a property market there are a several property transactions (universe), a valuer observed within a property transactions group (universe object set) that a property (object) has not the elevator (attribute). The relation between object and its attribute can be defined as "Certainly not". If the property has the elevator then the relationship between the object and the attribute will be "Certainly". The ranking of a triple A tenant may be defined a "Possible" attribute.

Rough Set Theory's application to property valuation do not need a model or assumptions. The valuation process is directly obtained from the observations, in this case the valuation process relies on an internal knowledge only. If in a considered group of property transactions, the properties have the same attributes, then they can be considered ***indiscernible*** at a certain level of information. An indiscernible element is defined as an "elementary set" as it is not possible to confuse it with any other element. In the case of real estate property market two residential properties with the highest level of comparability can be considered indiscernible. This happen even if two identical properties can not exist. Indiscernibility relation is the foundation of RST.

In the first stage the study of the relations between the object and its attributes or between the property transactions and the property characteristics is carried out methodologically through a "informative table". In the line of this table there are the objects. In the columns there are different attributes that may belong to the objects. For instance a line can contain a property transaction whose features are listed in its columns. In each cell there is the quantitative or qualitative description of the relationship between an object and its attribute. In the first application the presence or the absence of a parking (attribute) in a property (object) will be indicated in the informative table with a dummy variable, while the commercial area (attribute) of a propriety (object) will be expressed by square metres. The informative table S can expressed by the following formula (1):

$$S = \langle U, Q, V_q, f \rangle \text{ (1)}$$

Where U is the universe or a finite element set. For our applications this set in composed by propriety transactions. Q is a finite set of attributes related to the sold properties of the universe. V_q is the attribute q domain which may be (0,1) for a dummy variable. At the end *f* is the information function (Pawlak,1991). This function describes the relationship between the object belonging to the universe U and the attribute belonging to the Q set, which varies inside V_q domain. This concept can be expressed in the formula (2):

$$f : U \times Q \rightarrow V \ \ and \ \ f(x,q) \ \in \ V_q \ \ \forall q \in Q \ and \ x \in \ U \text{ (2)}$$

Each object $x \in U$ is described by a line (vector). Each element of this vector represents the value given to the relative attribute with reference to the x object, and which can be defined as $Des_Q(x)$. Among the universe objects some relationships can be marked out. There is an indiscernibleness or equivalency relationship[i] between two objects that belong to the universe U when the respective attributes are identical. For example two real estate properties which have both a 100 sqm area are indiscernible as regards this attribute. The relationship is confirmed even when it is set up for more than one attribute. Considering a non-empty subset N of the Q-attribute set for $N \subseteq Q$ The formula (c) below will describe the indiscernibility relation:

$$I_N = \{(x, y) \in U \times U : f_q(y), q \in N\} \quad (3)$$

The couple (x, I_N) is an approximative space. If $(x, y) \in I_N$, then it will be possible to say that x and y are N-indiscernible. Moreover if N = Q, the Q elementary sets are called atoms. In this case all the elements are indiscernible.If all the X set units of the U universe are analised acccording to the N attribute set and if they must result to be similar to each other (for example all the real estate properties have a 100 sqm area and are near the centre) then they will be indescernible. So two important concepts related to the indescernibleness relationship can be defined. In describing an object by its attributes some difficulties often arise as the observation of applicative cases presents a certain situation variability. For instance two real estate properties may have only one difference but a relevant difference in price. In order to define accurately an object two important sets must be defined. If we consider U as the object universe, X as a universe object set (real estate properties of whom the price is known), Q as the attribute set (that is under the above said universe), and N as an attribute subset. This with reference to the relationship between the objects X and the N-attribute subset for which they are analysed. Therefore the following set indicated in the formula (4) can be defined as the Lower Approximation:

$$N_{-}(X) = \{x \in U : N(x) \subseteq X\} \quad (4)$$

and the following set as the Upper Approximation:

$$N^{-}(X) = \{x \in U : N(x) \cap X \neq 0\} \quad (5)$$

In general terms the lower approximation is a description of the domain objects which are known with certainty to the belonging to the subset of interest, whereas the upper approximations is called a rough set. (Ziarko,1993). The Upper Approximation is defined by all that elementary systems which show a non-empty intersection with X. That is to say that if a real property undoubtly has an attribute, then it will be part of its positive or lower region. On the contrary if there are some set elements that have it and others do not, then the attribute will be described by the upper approximation.
The RST will value the phenomenon through these approximations. The difference between the upper or lower regions will be represented by a "*boundary region*" of rough sets. Comparing the information carried out by the rough sets to the consumption of an orange, the eatable part of the orange is defined through the difference between an inner part that is eatable in all its points - its contents - and an outer region where the fruit is not eatable at all. The yellow coloured intermediate content is a boundary region that is partially eatable, depending on the taste. The boundary region is expressed by the formula:

$$BN_N(X) = N^{-}(X) - N_{-}(X) \quad (6)$$

The three described regions are useful to define an information whose nature is "granular" as we stated in the premises. An information that appears to the buyers and the sellers as the pixel of a TV. The difference between the two regions contributes to qualify the object, describing in what cases the attributes, which qualify it, are always inside the lower element set and in what other cases they are sometimes inside (upper region). The objects that have to be analysed through some attributes and that belong to the same category are not distinguishable. This means that their membership status related to a arbitrary subset of the domain cannot be clearly definable. Therefore it drives to the definition of a subset with a lower or upper approximation. To describe an object it is possible to use both qualitative and quantitave attributes. If the boundary is not empty, the one represented is a rough set that can be defined through an upper union and a lower union. In this way a non-perfect reality can be defined. Many dimensions influence the granular nature of information: the attribute characteristics, the attribute numbers and each attribute domain. Obviously this procedure is strongly depending on the information quality, on the capacity to classify the information, on the ability to single out the attributes apt to describe them and on the confidence level and problem knowledge.

For this reason a first applicative phase of this methodology is defined through a classification in an "informative table". As already said, in its columns all the attributes, which can contribute signicantly to offer a problem view through the upper and lower approximations, will be pointed out. While in its lines all the elements of the considered universe will be taken into account. Considering the RST application to the real estate valuation, in the columns we list all the attributes (panoramic quality, maintenance need, area) each of them measured in a different domain. The lines will then contain the single components of the universe that are all the real properties considered in the transactions. Obviously the first stage is very delicate. If an heterogeneous set is considered, the consequences will clearly have operative problems. In a following stage the informative table can be turned into a decisional table, by dividing the attributes in: conditional (C set) and decisional (D set).
In fact, the attributes permit to define the inquiry object. The presence or the absence of these attributes, with different conditions in the object group, can contribute to clarify the lower and upper approximations above mentioned. But the division between the conditional and decisional attributes permit to establish a causal relation between the attributes. Defining the price as a decisional variable and the attibutes as a conditional variable, it is simply requested that the procedure analyses and evaluate the object determining a lower and an upper approximation based on the relationships between the set of elements containing the price (decisional attribute) and the ones containing other attributes through which the price behaviour is investigated (conditional attributes). The conditional attributes can be those that normally are used in the definition of the regression analysis. The result of this second stage is the processing of a "decisional table" in which, on the same table previously carried out, the decisional attributes are distinguished from the conditional ones.

The methodology returns the answers of the input received at a qualitative level with no distinction between large and small samples. The third stage – the last one – will analyse the relationships between the conditional and the decisional attributes. Those relationships will be analysed by taking into account the lower or the upper approximations between the decisional set D "of the price attribute" and the set (C) of the attributes that have been selected as conditional. As the following example will show, the origin of the "if…then" rules will allow to define the causal relationship between decisional attributes and conditional ones.

There are two general kinds of decisional rules. The first is the "exact decisional rule", named also deterministic, where the decisional set (the price) contains the conditional attributes (area or other features). The second is the "approximative decisional rule" in which only some conditional attributes (area or other features) are included in the decisional set (price). Needless to say the pertinence of the deterministic rules only regarding the real estate valuation problems. As a matter of fact in this case the causal relationships between the property features and its value are appraised without any uncertainty. The logical prepositions "if…then" allow the valuer to create a preferential system based on the property market data. The "granularity" of the system, its uncertainty can be increased in case the information is based on few observations. An example (d'Amato,2007) of rule that may be used for valuation purposes is indicated below (7):

$$\textbf{IF SQM = 36} \wedge \textbf{ ROOMS =1} \wedge \textbf{ DATE =41} \wedge \textbf{ YEARS = 17} \rightarrow \textbf{ PRICE=70.632,00 €} \quad (7)$$

As a consequence all the properties that have these attributes must have the following price or an interval of prices. It is evident a casual relationship between the attributes and the price but there is no econometric modelling that support the valuation process. The values are originated from the data without any inference. Having a great amount of observations there will probably several possible rules. It is possible to define a quantitative measure of the quality of the rule. Assuming S that in the universe U a number of conditional attributes C are casually related to a decisional variable D, therefore it will be possible to write the following decision table (Pawlak,2002):

$$S = (U, C, D) \quad (8)$$

The number supp_x (C,D) will be the support of a decision rule and will be indicated in the formula below:

$$\sup p_x(CD) = |C(x) \cap D(x)| \quad (9)$$

The following ratio will be also defined as the strength of decision rule

$$\sigma_x(C,D) = \frac{\sup p_x(C,D)}{U} \quad (10)$$

Another important indicator will be the coverage factor which can be expressed as follows:

$$\mathrm{cov}_x(C,D) = \frac{|C(x) \cap D(x)|}{D(x)} = \frac{\sigma_x(C,D)}{D(x)} \quad (11)$$

The strength and the coverage of the rule are important indicators to distinguish among several rules the right one for the object. Remain the problem that if a property differs of one only attributes the rule can not be applied. After the following trivial example the solution arrived in the following articles from 2003.

1.2 Integrating RST with Value Tolerance Relation (articles from 2003 to 2007):a crisp value

The original interest in RST as valuation methodology had another limitation. Although the if then rules allowed the valuer to appraise a property using a relation less strong than econometrical modelling[2], there was a problem: the valuer needs a crisp value, while RST in the previous version offered only intervals of value. For this reason starting from the 2003 the works integrated the RST with a functional extension named Value (or Valued) Tolerance Relation (d'Amato,2004;2004b;2007)

VTR can be considered a more flexible way to deal with the indiscernibility relation.
Classical Rough Set Theory relies on the crucial concept of indiscernibility relation as a *crisp* equivalence relation. Two properties may be indiscernible only if they have similar attributes. In

[2] That is based on an "olympic rationality"

property markets this is a strong assumption, the value tolerance relation allows the appraiser to develop upper or lower approximation with different "degrees" of indiscernibility relation. The formal relation is indicated in the formula (12) below:

$$R_j(x, y) = \frac{\max(0, \min(c_j(x), c_j(y)) + k - \max(c_j(x), c_j(y)))}{k} \quad (12)$$

The relation R_j may assume continuous values included in the interval 0-1. It is a variation ratio based on sets where membership function may have values included in the interval [0,1] (they are also called fuzzy sets). As a consequence the Value Tolerance Relation brings flexibility to traditional Rough Set Theory. In this context the choice of the minimum in the membership function represents the intersection between two sets, while the maximum in membership function results in the union between the two sets. Two objects x and y may have different levels of indiscernibility depending on a discriminant threshold k which measures the attributes c_j.

This functional extension may be defined the basement for a general application of RST as Automated Valuation Methodology and as appraisal method in the future. The k threshold can be applied to different measures of these attributes for all objects. For example, the indiscernibility relation between two objects (properties A and B) considering a k threshold of 10 sqm whose sqm are 120 and 190 may be calculated as in the formula (13) below:

$$R(c_a; c_b) = \frac{\max(0;120 + 10 - 190)}{10} = \frac{\max(0;-60)}{10} = \frac{0}{10} = 0 \quad (13)$$

The two objects cannot be considered similar respect to a k threshold of 10 sqm. The result of the application of a value tolerance relation with the same k to two objects as property transactions whose sqm. area are 120 and 129 is indicated below in the formula (14)

$$R(c_a; c_b) = \frac{\max(0;120 + 10 - 129)}{10} = \frac{\max(0;1)}{10} = \frac{1}{10} = 0{,}1 \quad (l4)$$

Using VTR the measure of indiscernibility relation is not crisp, but may have different degrees. If the value of R_j equals 1, according to a specific k threshold, the two objects are highly similar. This happen, with a degree of 0.5 in the formula i. Otherwise if the R_j is equal to 0 as in the formula h, therefore the two objects are completely different. This mathematical formula can also be used for the relationship between the object of a universe (properties) and a R_j set of rules developed for valuation purposes where the characteristics of the object (property transaction) are compared with the conditional part of the rule considered and indicated in the following formula as $c_j(\rho)$. Therefore it will be possible to write the formula 15:

$$R(x, \rho) = \frac{\max(0; \min(c_j(x), c_j(\rho)) + k - \max(c_j(x), c_j(\rho)))}{k} \quad (15)$$

In the formula there is a level of indiscernibility relation between the object and the rule assuming a k level of threshold for the measure of the attribute. In the first work in which was applied VTR (d'Amato,2004) the measure of k-threshold was found to be subjective due to the preferences and characteristics of the specific property market. In the forthcoming work (d'Amato, 2007) an objective measure of k-threshold is proposed as the standard deviation of each attribute contained in the group of properties to be estimated. If rules concern properties with similar characteristics then the thresholds (standard deviation) is low. The threshold is high, on the other hand, when the rules refer to a sample of elements containing properties with different features.

The value tolerance relation greatly improve the flexibility of RST . In this case the indiscernibility relation have different degree. The valuation process will not be a rigid application of a rule to an object, but it is possible to compare the object with all the conditional part of the selected rules defining which one is closer to it.

The relationship among all the attributes of an object and the conditional part of the "rules" is calculated assuming the "intersection" of all sets. The intersection is obtained comparing object with rule. As a consequence it is possible to obtain several Rjs according to the n attributes of the property and the conditional part of the rule. The select Rj will be the minimum Rj among n comparisons between conditional part of the rule and the attribute of the object indicated in the formula 16:

$$R_j(x,\rho) = \min_{j=1}^{n}(R_j(x,\rho)) \quad (16)$$

Where *Rj* is the value tolerance relation, *x* is a attribute of the property considered, ρ is the attribute belonging to the conditional part of the rule developed and n is the number of attributes of a property and the conditional part of the rule. The R (x,ρ) gives a flexible measure of this relationship. As an object may have more than one attribute, the appraiser has to take into account the minimum R_j among all attributes, as indicated in formula m. In the real world properties are compared referring to a high number of variables that can be considered attributes.

The formula m gives a rank to the comparison between each rule and a object. The right rule will be selected following three different criteria that will be indicated in the forthcoming article (d'Amato,2007). The first criteria will be indicated in the following formula (17)

$$R_j(x,\rho)_{1^{st}\ criteria} = \max_{j=1}^{m}(R_j(x,\rho)) \quad (17)$$

The higher the Rj, the greater is the similarity among single object and rules. Applying the above critaria to the property markets of Bari and Amsterdam, it was found that more than one rule had the same minimum R_j. In this case the appraiser considers as ***second criteria*** the rule with the highest sum of R_j calculated in comparison between property and the single rule (absolute maximum). The formula (18) is indicated below:

$$R_j(x,\rho)_{2^{nd}\ criteria} = \max_{j=1}^{m}\left(\sum_{j=1}^{n} R_j(x,\rho)\right) \quad (18)$$

In fact a property with a greater sum presents a higher Rj than other objects. It may happen that the highest sum of R_j indicated as second criteria (absolute maximum) does not match the first criteria, but the application of the criteria must follow the order indicated .Only if the first criteria does not fit the second criteria can be applied. In the case neither the first nor the second criteria can be applied
The right rule is given as the highest sum among those rules satisfying the first criteria (relative maximum). These criteria must be considered fundamental to choose the right rule for mass appraisal purposes. By applying these rules a comparison between MRA and RST on a large sample has been possible in Helsinki (600 observations) (d'Amato,2007) and in Catawba county (7000 observations)

that is the empirical case of this paper. In particular the last case has been developed using AVAMERST a software that the Real Estate Market Observatory of the 1st Faculty of Engineering developed for Automated Valuation purposes.

1.3 A Brief comparison between MRA and RST

Some differences between RST and MRA must, however, be highlighted. The greater difference is in the final output. Multiple Regression analysis allows the appraiser to define the price of each property characteristic considered in the model, while Rough Set Theory does not give information about hedonic – marginal prices.

In MRA the final issue is an econometric model while in RST the valuation is based on a boolean product and the valuer arrives at the final value estimate looking for the right if then rule suitable for the object. As MRA is based on an econometric modelling a set of assumption is fundamental. In the application of MRA a set of assumptions on errors and on the model if they are violated the model will be unreliable. In the MRA application several software may be used, while in the application of RST a software is on going[3]. While MRA has a limitation in the number of observations that are required which should be at least 30, RST can work also with small sample. In RST no such assumption is made; control indexes are restricted to the two main indexes, "accuracy" and "coverage" of rules.

The two valuation procedures are similar in other respects. As one can see both the application of RST and MRA are based on cross sectional process. The valuation process starts with the definition of "attributes" in Rough Set Theory and independent variables in Multiple Regression Analysis. In fact, a cause effect relationship is assumed in both Multiple Regression Analysis and in RST. With MRA output is a mathematical model while in RST the output is a boolean sum, or an if then rule. Both valuation procedures give the same results starting from the same sample and the same group of attributes.

There are no risks of different results coming from different "algorithms". Application of RST may be recommended for mass appraisal in those markets where the property market is not transparent such as in European Eastern Countries.

2. The case of Catawba County in North Carolina

The comparison between MRA and RST carried out on a sample of 7107 observation located in the Catawba County in North Carolina. The data referred to single family residential house sold in an interval of time between 2000 and 2005. Both MRA and RST are referred to 01.09.2006. The price have been included in an interval with a minimum of 22.000 $ and a maximum 1.800.000 $ and a medium prize of 159.941 $.

2.1 Automated Valuation Methodologies with MRA and RST in Catawba County

The regression model was runned on a sample of 3469 observations from the original "in sample" group of 3500 observations. For MRA and RST application the variable considered are listed in the table 1 below:

VARIABLE	MEANING
PRICE	Sale Price
DATE	Sale Date
TOTACS	Total Acres
TOTROOM	Total Rooms
BEDROOM	Number of Bedrooms
FBATH	Number of Full Baths
HBATH	Number of Half Baths
ADDFIX	Additional Plumbing Fixtures
BASEMENT	Basement Type
HEATYPT	Fuel Type
SYSTEM	System Type
ATTIC	Attic Code
OFP-LO	Open Frame Porch Lower Level
OMP-LO	Open Masonry Porch Lower Level
EFP-LO	Enclosed Frame Porch Lower Level
EMP-LO	Enclosed Masonry Porch Lower Level
EMP-UP	Enclosed Masonry Porch Upper Level
WDU	Wood Deck Lower
WDL	Wood Deck Upper
FGST	Flagstone/Tile Patio
ATTFR	Attached Frame Garage Area
ATTMA	Attached Masonry Garage Area
RECRO	Rec room width, used to compute area (SF)
CPEN	Open Carport Area
EXT WALL	Exterior Wall Material
CANOPY	Canopy Area
FRAME	Frame Utilità Building
MASONR	Masonry Utilità Building
MS-STO	Ms – stoop- Terrace
MT FP	Metal Fireplaces

Table 1 – List of Variables considered for MRA and RST application

The econometric additive function is indicated in the following formula 19 :

$$\begin{aligned} PRICE = & -27606{,}681 - 7{,}787 DATE - 303{,}625 AGE + 7011{,}361 TOTACS + 10648{,}773 TOTROOM + \\ & - 5774{,}044 BEDROOM + 28469{,}117 FBATH + 17946{,}618 HBATH - 2505{,}452 ADDFIX + 5437{,}450 BASEMENT + \\ & + 10046{,}023 HEATYPE - 1653{,}942 SYSTEM + 5808{,}039 ATTIC + 36{,}231 OFPLO + 193{,}961 OMPLO + \\ & + 24{,}605 EFPLO + 0{,}865 EMPLO - 509{,}324 EMPUP + 46{,}321 WDL - 33{,}203 WDU + 36{,}382 FGST + \\ & + 20{,}261 ATTFR + 65{,}478 ATTMA - 11{,}224 RECRO + 9{,}282 CPEN + 112{,}732 EXTWALL - 21{,}782 CANOPY + \\ & + 21{,}252 FRAME + 28{,}926 MASONR + 31{,}634 MSSTO + 14088{,}91 MTFP + \varepsilon \end{aligned} \tag{19}$$

Regression runned having with a R-squared of 0.72 and a susbantial validity of the model and of all the parameter taken into account and explained in the appendix 1. Then the differences between the actual and the estimated price have been carried out. In the following table n.1 it is possible to observe the internal validity or the in sample difference between the actual and estimated price with MRA . The

Mean Absolute Percentage Error and the forececasting error for proportion of errors have been calculated and indicated in the table 2 below

Proportion of Errors – In Sample				
0-10%	10-20%	20-30%	more than 30%	Mean Absolute Percentage Error
40,74%	29,50%	16,91%	12,85%	0,158169

Table 2 – MRA Internal Validity

The same difference was calculated for the out of sample in the table 3 measuring the valuation accuracy of the MRA model.

Proportion of Errors - Out of Sample				
0-10%	10-20%	20-30%	more than 30%	Mean Absolute Percentage Error
31,41%	26,06%	16,85%	25,68%	0,205299

Table 3 – MRA Valuation Accuracy

For RST application, the case of Catawba County has an important innovation, respect the case of Helsinki (d'Amato,2007). While in Helsinki case the rules were developed on a small sample in this case as the sample was composed by 7000 observations, it was divided in two parts the former in-sample part was used completely to develop the rules, while the latter out of sample part was used to test the rules. In this case, having 3500 observations the rule with the highest coverage factor (rhs) were selected. The calculation of the coverage of the rule was made using the software ROSETTA[4].
Using these rules the objects were compared with the conditional parts of the rules. The comparison followed the three criteria indicated in the paragraph 1.2. The results of insample internal validity of RST are indicated in the following table:

Proportion of Errors - In of Sample				
0-10%	10-20%	20-35%	more than 30%	Mean Absolute Percentage Error
33,89%	21,77%	14,86%	29,49%	26,73%

Table 4 – MRA Internal Validity

The results of out of sample valuation accuracy of RST are indicated in the table n. 5 divided for proportions of errors

Proportion of Errors - Out of Sample				
0-10%	10-20%	20-30%	more than 30%	Mean Absolute Percentage Error
17,66%	16,86%	20,95%	44,53%	35,69%

Table 5 – RST Valuation Accuracy

The empirical results confirmed the superiority of MRA on RST. The proportion of errors of MRA including in the interval 0-10% are double compared to RST .

[4] http://www.idi.ntnu.no/~aleks/rosetta/

Final Remarks, Future Directions of Heretic Research

The work showed an enhancement in the application of RST as AVM. RST was applied to a sample with a greater number of variables and observations than in the previous works. In this case the rules have been generated with the contribution of ROSETTA software. Only the rule with coverage of decisional attribute (RHS) equal to 1 have been selected. After selecting the rules each in-sample and out-of-sample object have been compared with the set of selected rules. The growth of conditional attributes create problems to RST application.

The results of this works highlighted that RST works better with sample with a low number of attributes such as in the previous works. Although MRA demonstrates its superiority, the relationship between the object and the rule in the application of RST for mass appraisal purposes is an important direction of research. In particular a control system to eliminate outliers in RST may contribute to help empirical application of RST for large sample with a great number of variables.

References

Arrow K.J. e Debreu G. (1954), Existence of an Equilibrium for a Competitive Economy, Econometrica, 22, p.265.
d'Amato M. (2002) Appraising properties with Rough Set Theory, Journal of Property Investment and Finance, Vol.20 n.4, pp. 406-418
d'Amato M. (2004a) A Comparison Between MRA and Rough Set Theory for Mass Appraisal. A Case in Bari, International Journal of Strategic Property Management, vol.8 n.4 pp.205-217
d'Amato M. (2004b) Un'applicazione della RST per Mass Appraisal :il caso di Amsterdam, Rivista del Consulente Tecnico,n.2 Ed. Maggioli Bologna
d'Amato (2007) Comparing RST with MRA as Automated Valuation Methodologies, International Real Estate Review, n.2 (to appear)
Griliches Z. (ed.) (1971*)* Price Indices and Quality Change: Studies in New Methods of Measurement. Cambridge, Mass.: Harvard University Press.
Kauko T and d'Amato M. (2004) Mass Appraisal Valuation Methodologies:Between Orthodoxy and Heresy, Paper unpublished presented in Milan 2nd -5th June
Pawlak Z.(1982), Rough Sets, International Journal of Information and Computer Sciences, 11,341-356
Pawlak Z.(1991), Rough Sets. Theoretical Aspects of Reasoning about Data, Kluwer Academic Publisher, Dordecht
Pawlak Z and Slowinski,R. (1993) Rough Set Approach to Multi-Attribute Decision Analysis.ICS Research Report 36,Warsaw University of Technology
Pawlak Z.(2002), Rough Set Theory and its Applications, Journal of Telecommunications and Information Technology, n. 3 pp.7-10
Rosen S. (1974) Hedonic prices and implicit markets: product differentiation in pure competition. Journal of Political Economy, 82, pp. 34-55.
Simon H. (1957), Models of Man, Wiley, New York.
Simon H. (1979), Models of Thought, Yale University Press, New Heaven.
Simon H.(1981), Models of Bounded Rationality: Behavioral Economics and Business Organization, vol.II, MIT Press, Cambridge MA.
Ziarko W. (1993), A Brief Introduction to Rough Sets, The First International Workshop on Rough Sets: States of Art and Perspectives, University of Regina Saskatchewan