Advances in mass appraisal methods – an international perspective

Editors: Tom Kauko, Maurizio d’Amato

Contents:
1. Introduction: Suitability issues in mass appraisal methodology (Kauko & d’Amato)
2. Mass-appraisal practice and recommendations (Borst & Thompson, Siniak)
3. Orthodox methods (des Rosiers, Francke, Renigier)
4. Heretic methods (Kauko, d’Amato, Stumpf Gonzales)
5. Comparison of tools using a set of specific criteria (all authors)
6. Conclusions
1. Introduction: Suitability issues in mass appraisal methodology (Kauko & d'Amato)

In this chapter we explore the possibilities to develop mass appraisal methods, following two different arguments: one, that the performance and feasibility of appraisal methods may be compared and evaluated with regard to a set of technical criteria; and two, that differences in the suitability of methods also has to do with the particular context where application takes place.

Mass appraisal may be defined as a systematic appraisal of groups of properties using standardized procedures. Mass appraisal methodologies normally refer to large groups of properties rather than to a single property. The accurate assessment of the value of a predefined set of properties, or one particular property, indirectly, using a model, for a given practical purpose, is the main target of these methodologies (e.g. McCluskey et al. 1997; Gonzales et al. 2002a,b). Several contributions have addressed the importance of mass appraisal, exploring the relationship between the property value, the property characteristics and urban social and economic problems. The market behavior is influenced by the property prices, the high durability of the property asset and by the fixed geographic location (Robinson 1979; Harvey 1996). Hedonic price modeling has been proposed to define an econometric relationship between the price and the property characteristics, particularly in a residential context. Arguably, the standard multiple regression analysis (MRA) based hedonic price models are not suitable for capturing all the necessary information involved in the value formation process, and the literature on how to develop the value modeling tools further is (or at least should be) evolving. Although the problems are highlighted, MRA remains at the moment the most important theoretical framework in mass appraisal.

The quantitative, MRA-based methodology may be referred to as the ‘orthodox’ approach to mass appraisal valuation. Several methodologies have been applied in the last decade defining a new approach to property mass appraisal valuation. In this work these methodologies are dubbed as ‘heretic’ because of their different theoretical basis from the MRA, the dominant approach to mass appraisal. Model-free estimation techniques such as neural networks and fuzzy logic have been introduced to bring some flexibility to the property value calculations, without neglecting the mathematical rigor. In doing so, the value model becomes more powerful than its formal regression based and completely crisp counterpart. Pattern recognition is yet another relatively untried approach within this realm. Indeed a number of contributions here offer ingenious and pragmatic, if not totally transparent, modeling methodology (see e.g. Jenkins et al 1999; McCluskey and Anand 1999).

1.1. Orthodox approaches to mass appraisal valuation

Two related modeling traditions today exist both of which deploy MRA for estimation: the model driven hedonic approach, and the data driven statistical approach. Hedonic price models comprise the most frequently applied models in the valuation practice as well as in monitoring the housing market. In these models the variables are usually of two basic types: internal physical (i.e. house and plot specific, structural) and external locational. On top of that there may be additional variables, most notably some type of inflation control. (e.g. Miller, 1982.) The purpose of the development of the hedonic price model was to make possible econometric analysis of large databases of price and other recorded information describing the nature of the property and its vicinity and possibly some specific (other) circumstances of the transaction. A more practical or atheoretical statistical, especially regression analysis -based value/price-modeling tradition has been applied in order to provide tools for valuation conducted by the public and private sectors in many countries with
convenient land information infrastructure (i.e. readily available digital register information with possibility of multiple spatial aggregation).

There are several MRA-based house price studies that are made either with or without the formal hedonic price theory underpinnings. As shown by a multitude of studies (see e.g. reviews by Ball 1973; and Lentz & Wang 1998) the measures of success have been the model fit (indicated by the total correlation coefficient, the R-squared statistic), the model significance (the F-test at a given level of significance), whether each independent variable has the anticipated sign of price association (indicated by the partial correlation coefficients) and whether each independent variable is statistically significant (indicated by the t-statistic and a given risk-% level). The standard error of the model and various test statistics are also applied as such formal criteria of modeling performance. Recently, hedonic modeling has been applied successfully for constant-quality price indices (Hoesli et al. 1997a); determining rental values (Hoesli et al. 1997b); estimating the disturbance effect of traffic (Wilhelmsson 2000); and estimation of implicit prices (Laakso 1997).

In the empirical hedonic modeling literature locational proxy variables may be defined in various ways (cf. surveys by Ball 1973; Miller 1982; Laakso 1997; Lentz and Wang 1998; see also Wyatt, 1999, and Thériault et al., 2005, for sophisticated accessibility measures). Kang and Reichert (1991) constructed a locational quality dummy based on levels of price per sq.m. living-space. Similarly, McCluskey and Anand (1999) used a solution, where the location was captured with a seven-valued categorical ‘ward’-variable, with values based on mean transaction prices for that area.

Adding location to the already complex analysis of residential differentiation has become easier with the help of a modern geographic information system (GIS). There obviously are operational problems with the hedonic model, such as lack of suitable variables and data, and the issue of spatial resolution, that is choosing a wrong level of spatial aggregation. This brings us towards GIS-aided analysis, with the benefit of data visualization and storage, the possibility to construct more efficient accessibility measures and spatial analysis. An additional point is that the GIS -technology makes value-modeling applications more user-friendly. For GIS-demonstrations within property valuation, see Orford (1999); Rodriguez et al. (1995); Wyatt (1995, 1996a,b, 1997, 1999a); Lake et al. (1998); and Ding et al. (2000). See also Bible & Hsieh (1996) for analysis of apartment rents Furthermore, a GIS also helps correcting the spatial discontinuity problem when the spatial dimension is incorporated to the analysis more explicitly than in the standard hedonic model. (Orford 1999, 63-67; see also Wyatt, 1995, 1996a,b).

In a GIS only a few spatial statistical methods can be integrated. When the spatial analysis system is 'loosely coupled’, the statistical analysis are performed separately from the GIS interface. Both packages thus have their own function in the system: the raw data is transported to a GIS for storage, visualization and generation of variables for analysis of locational externalities, most notably proximity indicators (e.g. Glascock et al., 1997). Then, the latter may be exported to a separate statistical package anew, where for instance standard regression analysis or more sophisticated multi-level regression analysis (see below) may be undertaken. Finally, the analyzed data may be imported to the GIS again, for storing and visualizing. Then, the output of a GIS becomes a surface indicating the spatial effect of a given variable on price. Studies by Orford (1999); Lake et al. (1998), Rodriguez et al. (1995), Bible & Hsieh (1996), and Des Rosiers et al. (2000) manage to show that in one way or another, the importance of location is great (see also Wyatt, 1996a,b,1997). In general, the
GIS aided property value analysis is feasibly linked to broader register information (Federal land cadastre of Russia, 2001).

Today a variety of advanced spatial techniques enhance the possibilities of handling location in the hedonic based house price analysis. Namely, the consideration of submarkets and spatial drift may improve results substantially (see Orford, 1999). This requires an appropriate routine of handling the non-linearity and dynamics prevailing across space. State of the art methods include multi-level specifications and spatial expansion models.

In multi-level specifications, each externality effect is measured at an appropriate level. In order to add some efficiency into the (hedonic) value model, the variation in house prices is decomposed between different spatial scales. In the case of property valuation applications the appropriate levels may be neighborhood, street and property levels. A major advantage of this specification is the ability to differentiate between compositional and contextual effects of location on house prices, in other words of the place in itself and spatial variations in the housing stock. (Orford, 1999, 2002)

In the spatial expansion model the contribution of a housing characteristic to the price is allowed to change over space. This reflects a series of interrelated submarkets with sliding boundaries. Many applications use such a specification where parameters vary in order to cope with the spatial heterogeneity of a housing market. (E.g. Geoghegan et al., 1997; see also Wilhelmsson, 2002).

We can note related contributions: Gillen et al. (2001) discuss autocorrelation when it is either a function of distance separating properties in space (isotropic) or function of distance and direction (anisotropic). Riddel (2001) improves the hedonic approach by allowing for interaction between space and time in a dynamic modeling framework, as environmental quality may induce market disequilibria. Hierarchical trend modeling with hedonic underpinnings using a Kalman filter has been proposed by Francke and Vos (2004).

In many situations exogenous factors or lack of information constrain individuals to participate in segments of a larger market (Michaels & Smith, 1990). However standard hedonic analysis has not ignored market segmentation completely. In principle a hedonic regression cannot detect zonal boundaries, only the significance of the direction and coefficient of the effect of the value factors as well as the accuracy and explanatory power within the total sample of observations. One way of clarifying the issue is to use dummy variables. (Laakso, 1997.) Another solution is to calculate hedonic quality ranks for each observation (Rothenberg et al. 1991: p. 380-385). A third option is to construct separate models for separate subsets of the data, with each subset, usually comprising all transactions within a region, having its specific hedonic equation (i.e. the partitioning approach). Hence the data is split into different segments, which are either a priori predefined or synthesized somehow. Also if segmentation in a theoretical sense is ignored, the partitioning approach may be justified (see Needham et al., 1998).

Demand side segmentation, that is collective preferences according to membership to an a priori defined ethnic or socio-economic group, is often studied with the specified two stage procedures of hedonic modeling. If the functional form is curvilinear we may derive WTP/WTA estimates based on the shadow prices and specified demand side data. (If the equation is linear, the second step in the two-step hedonic approach cannot be taken.) The
target then is the marginal WTP estimates (i.e. demand functions) for each relevant characteristic by groups. (E.g. Bökemann & Feilmayr, 1997)

An integration between the adjustment grid methods and regression analysis was developed by Colwell, Cannaday and Wu (1983), in order to integrate the OLS estimation of adjustment factors to the standard method. Here the multiplicative percentage grid adjustment method is considered a particularly promising option (see Kang and Reichert 1991).

Another example of earlier innovations is the Stein rule (Knight et al. 1993). Here the idea is to improve the estimation accuracy by the following procedure: first add all variables to the model and estimate the partial correlations (i.e. beta coefficient, marginal adjustment factors) with OLS. Then, remove the variables, for which it is believed that beta equals zero, in other words the insignificant effects on price, and estimate a new beta with a restricted regression. After that the estimates are compared. According to the Stein rule, if the coefficients are close to each other, then the beta estimated with the restricted regression obtains a greater weight, and the more the results differ, the smaller the weight for the restricted beta estimates. The result is then a product of both the data and the relevant domain knowledge of the appraiser.

The repeat sales method is based on the idea of selling the same property twice, in order to derive the change in value (e.g. Gatzlaff & Haurin, 1997; see also Wilhelmsson, 2000; and Meese & Wallace, 1997).

Logistic regression (i.e. regression, where the dependent variable is a dummy; discrete choice) is not so common within the value-modeling field (Feenberg & Mills, 1980: p. 110). In a rare study, Bolen et al. (1999) used logistic regression to estimate probabilities of land value increases and residential rent increases related to certain characteristics in Istanbul. They justified the use of the method with the irregularity of the urban system.

1.2. In between orthodoxy and heresy - model-free regression

Flexible (i.e. model-free) regression methods are considered an interesting research methodology within the formal modeling paradigm too. More generally, Verkooijen (1996) used the term ‘flexible regression’ rather than ‘non’- (or ‘semi’) parametric regression, for

- local approximations \( E[Y \mid X=x] \) (e.g. locally weighted regression such as splines or kernel)
- low dimensional expansions \( f(x) = \sum \phi_i z_i \) (e.g. additive models, with both parametric and non-parametric components)
- adaptive computation (e.g. neural networks, see next section).

In the context of estimating house prices or property values, flexible regression has been discussed and encouraged by Meese and Wallace (1991); Pace (1995); Mason and Quigley (1996); Verkooijen (1996); and Kyllönen and Räty (2000) to name some recent studies. Coleman and Larsen (1991) in turn remain more critical towards alternative estimation techniques. Furthermore, a number of contributions apply this approach for the explicitly spatial dimension. Pavlov’s (2000) space-varying coefficients not only combine the two main spatial methods (the error method and drift/lagged method), but also manage to incorporate location as a non-parametric influence and without any inferential theory underpinnings. The spline approach was used by Dubin and Sung (1987), who used a linear, an exponential and a quadratic approximation for the rent gradient, measured through ‘rays’ drawn from the city centre, each of them capturing the price-distance relationship in one significant direction.
According to Mason and Quigley (1996), theory does not provide guidance about the choice of functional form; thus this is purely an empirical matter. For this goal, non-parametric (and semi-parametric) methods offer advantages in relation to confirming non-linear relationships. The generalized additive model (GAM) proposed by Mason and Quigley is based on an arbitrary smoother of the curvature. It is a compromise between the generality of methods such as Kernel or local regression, and the comprehensibility of the parametric MRA. Its benefit is the tractability of the results, but it comes at the cost of suppression of complexity.

In much similar vein, Colwell (1998) and subsequently Colwell and Munneke (2003) apply an interesting approach for estimating price surfaces of urban land within a non-parametric regression approach. The shape of the spatial price surface cannot be assumed to be of any particular form – hence, non-parametric estimation. The particular method, referred to as *piecewise parabolic multiple regression analysis* (PPMRA), assumes straight lines between valleys and peaks, and full continuity. The space is divided into discrete, but contiguous square spatial units. The resulting function is a continuous, but less than totally smooth (i.e. not everywhere differentiable) spline function. As independent variables Colwell and Munneke apply barycentric coordinates – a standardization and weighted average procedure that transforms the absolute values into relative ones, to add to the efficiency of the method. The regression coefficients of these coordinates reflect the vertex relative to the constant term. They also note a tradeoff between a gain of continuity and a loss of degrees of freedom. On balance, their work is innovative and offers an adequately valid tool for general property market analysis, as well as mass appraisal purposes (feasibility criteria notwithstanding). A particular advantage is the capability to avoid the problem of too smoothed point estimates that mar the analysis of most semi-parametric methods. This shows the advantage of the piecewise method: it generates a true surface output, and not just a matrix of points. A balancing aspect in the comparison is however the problem that the lines between the vertices in piecewise regression are straight lines (thus, the curve is not differentiable, although continuous), which in turn suggests that the method is not so flexible in this respect.

Finally, a few words on the assumptions they use: they note that (1) market discontinuities caused by topography and government policy are destroyed by arbitrage, which we really consider true only for time series not space, and that (2) if such discontinuity persists, this can easily be modeled. This bold ‘guarantee’ is of course still within the belief in neoclassical modeling rationality and tractability; personally we really do not believe in this as a categorical assumption, as land markets are not continuous in space. Thus, we have spotted a problem with the method – it uses after all fairly rigid orthodox underpinnings. This point is worth reiterating: while standard housing models and urban/land economics models do not allow for discontinuities and nonlinearities very well, it seems reasonably to assume that time has a linear price association, whereas space/segmentation has a nonlinear relation (for example, due to regulation, perceptions of individuals or topography).

Paez et al. (2002) deal with geographically weighted kernel models that allow for locational heterogeneity. Geographically Weighted Regression (GWR, Fotheringham) is growing in popularity among geographers and real estate market modellers. It is considered a locally weighted regression method that operates by assigning weights to all observations depending on their distances to a geographical focal point (Páez et al., 2002). The idea is to run a separate regression for each observation point, and, using X, Y identification of the observations, produce a surface of response functions. Thus, GWR generates different coefficients for each point in the data set, and subsequently the responses for each X, Y point.
can then be computed based on the input (either the actual indicator values of the observations or their mean values).

Flexible estimation is carried out in order to estimate the distribution itself from the data. While flexible estimation is an alternative to a fixed parametric one, an inevitable tradeoff makes the decision of modeling choice less straightforward: flexible regression is less efficient than fixed parametric regression, but avoids model specification problems – the problem of parametric methods.

1.3. Heresy in mass appraisal and valuation – a variety of approaches
A methodology is here defined heretic because of the contrast with the dominant framework of multiple regression analysis. These techniques have been developed in several fields. Their application represents an experimental and state-of-the-art benchmark in research activity on mass appraisal. While the evolution of the paradigm is different, and to some extent the philosophical position and professional jargon too, we note a partial overlap with the approaches discussed in previous section.

(Artificial) neural networks (ANN) are a sort of flexible, model-free regression (cf. Verkooijen 1996; Pace 1995; see also McCluskey and Anand 1999). The nature of the neural network is a ‘black box’, which means that there is no clear functional relationship between the input and output values. The algorithm learns by training. The basic elements in a neural network are called neurons or nodes. The connections between them are determined by weights. Together the neurons process direct inputs, or inputs from other neurons, to give an output value. The inputs must be numerical values. Then the output values are corrected iteratively, until the system has achieved desired accuracy. There are three basic types of algorithms, the feed forward, feedback and competitive networks, feed forward one being the most common one. The multi-layer perceptron (MLP) -feed forward network bears resemblance with multiple regression, except it has no explicit mathematical relationship between the input and output (but see White, 1989). In general, ANNs have proved rather successful for classification in finance and economics, in a similar sense as more established statistical methods such as logistic regression and linear discriminant analysis (Verkooijen 1996, 97-98). However, during the 1990s, several authors found serious problems with the ANN (e.g. Worzala et al, 1995), which lead to abandon this technique from mass appraisal.

The self-organizing map (SOM, Kohonen Map), a particular type of ANN, produces a ‘feature map’ of clusters, each of which is represented as a specific characteristic combination of attribute levels. As a result of the analysis, the researcher obtains a surface where the areas with similar combinations of variables can be looked at as a whole, and on the other hand, compared with different combinations of variables. It is also possible to interpret a ‘typical value’ of each node, for a given feature. The SOM, being a particularly visual approach, has the advantage that it allows for some qualitative analysis, on top of a quantitative one.

Plenty of similarities can be noted with the SOM and the flexible regression methods described above – for example, Colwell and Munneke (2003) compare their PPMRA method with kernel estimation, another flexible regression technique, which is closely related to the SOM. Thus, these methods are also subject to much of the same criticism. However, a few differences are worth noting. One notable difference is that, whereas the kernel (and the SOM) imposes a priori grid-iron structure on the surface, it inevitably ‘smooths’ particularly high and low values (i.e. irons the valleys and peaks), the piecewise parabolic method uses the peaks and valleys as vertices of the grid. Obviously, with the SOM this can be corrected by
enlarging the dimensions of the map, so that a sufficiently large resolution allows the peaks and valleys to be fitted with the response.

Several papers have been written on this topic since the early 1990s (e.g. Tay and Ho, 1992; Borst, 1995; Worzala et al., 1995; and McGreal et al., 1998). Some are for and others against the use of ANN for valuation – mainly due to the ‘lack of robustness’-argument. Nguyen and Cripps (2001) add to the long list of performance comparisons between MLP (backpropagation-ANN) and MRA. For accuracy indicator they use the mean absolute percentage error of the model (i.e. the internal accuracy of the estimated model), and the forecasting error when thresholds of 5% and 15% are used (i.e. how large a % of the test cases are predicted within p-% of the actual sale), when measured with validation samples. They conclude that the performance of the MRA model improves, when the functional specification improves, whereas the performance of the MLP model improves, when the dataset used for training increases. Thus, the MRA performs better, when small samples are used, but if sufficient data size, and appropriate network parameters are found, the MLP performs better. When we compare with a similar comparison between the SOM and MRA (Kauko and Peltomaa, 1998; Kauko, 2002), the conclusions were the opposite to these: the ANN-based model performed better than the MRA models for small data sizes in recognising a hypothesised impact between externality and price. On the other hand, Gonzalez et al. (2002a, 2005) found out that for the accuracy performance, ANN is slightly better than MRA and other models. Some of these problems are the same for both ANN-types, the SOM and the MLP: how to set the field ranges of the input-variables, when we do not want one factor to dominate too much; when to stop the run based on trial-and-error, and how to know, whether we at that stage have found the optimum; how many variables is optimal to include; and how to explain the ‘black box’?

Some further thoughts concerning the similarities between the GWR and SOM techniques can now be put forward. Both are based on models where proximity and intensity determines the response, and that the output is heterogeneous as it applies for any point in the data (GWR) or any cluster of points in the data (SOM). The crucial difference is that GWR models this influence in X, Y space and the SOM in the n-dimensional space defined by the input variables. In principle this means that, if the SOM includes X and Y among the input, and these two variables are given infinite weight, thus in practice sufficiently strong weight compared to the other variables, the output looks like the output of GWR. Therefore, if the SOM generates an output where each point corresponds with one single observation then it should correspond with GWR. However, whether the outcome is exactly the same depends on the definitions of response function and in the case of SOM, also on the initial values.

The genetic algorithm is another recent semi-parametric tool from the machine-learning paradigm. According to some experts (e.g. J.H. Holland and D.E. Goldberg, cited in Cooley et al. 1994, 182), genetic algorithms have proved successful in applications involving the efficient investigation of large search spaces. The genetic algorithm performs an artificial ‘breeding’ of a replacement population from a randomly generated population of previous encoding, from ‘parents’ to ‘offspring’ (see also Cooley et al. 1994).

McCluskey and Anand (1999) highlighted a number of issues within the mass appraisal realm related to functionality and accuracy of value modeling approaches. They also built an intelligent hybrid system, where the neural networks and genetic algorithms were used in association with a nearest neighbor algorithm, a technique applied in classification and pattern recognition. The application involved a distance metrics (Euclidean or significant mean) for
determining the similarity between the comparables for nine attributes. Soiberman and González (2002a,b, 2005, 2006) maintain that data-mining analysis – using various techniques – is helpful in its search for unknown patterns. The most important thing however is to note that these techniques are only as good as the data you feed them.

Some discussion has been devoted to the issue, whether a more qualitative technique of computerized data processing would after all be better suited to dealing with a complex field such as property valuation (e.g. O’Roarty et al., 1997; McCluskey and Anand, 1999). Rule-based expert systems emerged as a counter paradigm to neural networks and other numerical techniques. This technology models human judgment and decision making by explicit rules, as opposed to learning automatically from historic information. Therein also lies their biggest drawback: intensive expert interviewing takes much time and money (Verkooijen 1996, 97-98). The problem is that the model designer has to formalize everything, which is rather clumsy, and, if the optimal decision has to be determined based on a data mapping, even misleading. In other words, they do not offer robust solutions. Nevertheless, expert systems have their applicability. Scott and Gronow (1990) discussed the components of valuation expertise within the mortgage valuation domain and further explored the different levels at which this expertise is exhibited. Finally, they suggested the production of an expert system which would reproduce the expertise of the human valuer. McCluskey and Anand (1999) in turn considered knowledge elicitation and simulation of expertise as strengths, but the lack of robustness and rigidity of models as weaknesses.

Case-based reasoning deals with retrieval of past cases similar to the one to be assessed. Within property valuation this means being based on historical transactions. It does not encapsulate heuristic knowledge, while the selection of variables relies on the data. The method utilises a case library, formed from comparable evidence, and the most useful of this information held can be retrieved to form an opinion about the value. (see O’Roarty et al. 1997; Gonzalez and Laureano-Ortiz 1992; Bonissone et al. 1998.) O’Roarty et al (1997) stressed, among other things, the flexibility of the system in retrieving cases, rather than to merely search for cases that exactly replicate the input problem cases. Case-based reasoning is, according to Wyatt (1999) the most promising of the new techniques/methods, because it avoids problems in knowledge acquisition. Reasoning from past cases, objectivity and explainability are the main strengths of this method. It negates many of the problems associated with expert systems. However, it requires considerable data (McCluskey and Anand 1999; see also Pacharavanich et al., 2000).

Fuzzy logic deals with “imprecision of the present” as opposed to probability which deals with “uncertainty of the future”. The theory of fuzzy logic is formally specified as a measure for the degree of membership for an element’s belonging to a set. These gradations can sometimes be used in property valuation as well. (See Bagnoli and Smith 1998). This is about computing with words. The fuzziness reflects the way of evaluating alternatives and making decisions. Lee et al. (2003) argue that fuzzy quantification theory can reduce the subjectivity caused by the appraiser, and also allow for a flexible adjustment of effecting factors. For preparation of the more discrete variables they propose a fuzzy linguistic method. By replacing crisp memberships, that is absolute values of predictor variables, with membership functions that allow for imprecision, the basis for the valuation is improved, they argue. Sui (1992) noted that the problem with conventional crisp methods is the loss of information, when faced with ambiguity and imprecision. At the extremes, however, both the fuzzy and crisp type of applications, yield the same results. According to Sui, the approach based on
fuzzy set theory too has its problems: the difficulty is how to define the correct membership function for spatial purposes, as the membership function is derived *ad hoc*.

*Rough set theory* (RST) seems promising based on early evidence (see d’Amato, 2002, 2003). The method derives Boolean-rules from actual market data, and not expert knowledge. The method derived from RST is robust, but on the other hand, may be too subjective due to the external adjustments (same problem with ANNs); specifically, the k-input, which determines the choice of the appropriate rule. Furthermore, RST is surely more transparent than the ANN. It is also more qualitative than any ANNs (which we find a strength), which of course does not contradict with the fact that having more observations to model the rules with would be better here too.

Hybrid systems may be created based on various kinds of AI. Pagourtzi and Assimakopoulos (2003) utilized several of the methods discussed. Gonzales et al. (2002b) built a mass appraisal model, where fuzzy rules were extracted from the ANN. The aim of the rules was to explain the computations in the valuation process. These would be understood as kind of hedonic coefficients. They concluded that while more testing was needed, the approach was characterized as promising (see also d’Amato and Siniak, 2003).

1.4. Methods based on interviews and surveys

Some recent comments freely encourage the use of interview survey methodology for residential valuation. According to Lentz and Wang (1998), to simply ask from individuals about their willingness to pay for certain property characteristics, for instance aesthetic value, is an intuitively appealing technique. However, Lentz and Wang (1998) noted that a successful application of the survey method depends on the existence of an informed populace with market experience regarding the attributes in question. In the following, we discuss some of the approaches that use judgmental data, beginning from the most common ones to the most controversial (but state-of-the-art) ones.

The *contingent valuation* -method (CV) is based on a formal questionnaire about the respondent’s willingness to pay (WTP) or – in reverse situations – his/her willingness to accept (WTA) a given sum of money (see Brefjle et al., 1998; Ruokolainen and Tempelmans Plat, 1998; Ruokolainen, 1999). Contingent valuation is the most widely used method of monetary evaluation of environmental benefit (Mäntymaa, 1993). Estimates generated by CV and hedonic modeling have been compared in several contexts (see e.g. Willis & Garrod, 1993; Wilhelmsson, 2000). Usually the hedonic method is considered more reliable, because the analysis is based on actual rather than hypothetical data. The sensitivity to the rate of discount might prove another problem with contingency valuation if monthly and total expenditures have to be compared (e.g. Vainio, 1995). However, in some cases the prices paid do not reflect all the possible externalities, as they become familiar only with time. For instance, in Vainio’s (1995) comparison of hedonic pricing and CV the questionnaire was sent three years after the transaction, in what time the buyer had perceived the full extent of a disturbance effect from the noise of a nearby motorway. In this case the hedonic models underestimated the effect.

Gartner et al. (1996) noted that in situations were a property possessing certain attributes is not frequently traded at an open market, the owners’ own estimates of value provide more useful estimates of economic benefits than those derived from sales transactions. Ready et al. (1997) asserted that when non-use values (e.g. altruism toward current residents and preservation of cultural heritage) are large, contingency valuation may be preferred to hedonic
methods (see also Magat et al., 2000). The main weaknesses of conventional survey tools are (1) the possibility to manipulate the outcome by predetermining the nature of response mode; (2) they are incapable of accommodating explicitly the multi-attribute nature of tradeoffs between alternative choices. We might need very context sensitive insight into how various multidimensional values are being perceived by the individual. Then, a pure competitive market approach loses validity. Indeed, CV is a rigorous option, but if one requires estimation of other than monetary benefits, we need another approach.

The multi-attribute value tree (MAVT, in Anglo-American literature often: utility tree) provides a formal way of thinking through multidimensional eliciting of peoples’ weighted objectives in the context of their expressed values and their selected project alternatives (e.g. Gregory et al. 1997). Tools such as the MAVT are suitable for evaluation of other than monetary values when they are mixed with or linked with monetary ones (e.g. Miettinen & Hämäläinen, 1996). This approach include techniques such as the analytic hierarchy process (AHP), the self-explicated utility method and conjoint analysis. The first two are hierarchical models and thus apply the value tree concept, whereas the last one is based on choice profiles. All three are aimed at making choices according to preferences in a multi-attribute problem setting, in contrast to the purely economic WTP-setting of revealed preferences and CV. (e.g. Pöyhönen 1998; Miettinen & Hämäläinen 1996.) All of these techniques contain an assumption about deterministic preferences of the interviewed subjects. In the built environment context, they could be understood as different perceptions of experts from a flexible, problem-specific point of view (e.g. Laakso et al. 1995; Nevalainen et al. 1990).

In these methods the weighting of the preferences becomes a question of elicitation (Ruokolainen & Tempelmans Plat, 1998; Pöyhönen, 1998a). The AHP uses a pair-wise matrix comparison of preferences, especially, when no price-information is available. The AHP has been applied in several ways within this particular research area. Like Kauko (2002), Fischer (2003) too sees a more qualitative approach as an improvement, the difference between these two specific methods being that Fischer mixes also price criteria with the other criteria already in the model structure, whereas Kauko arrives at a pure quality rank – possibly to compare with actual prices at a later stage. Fischer (2003), like Kauko (2002), concludes that the problem for this approach if it is being adopted for practical applications may be the very low time- and cost-efficiency. Ghyoot (2001) uses the AHP for site selection, together with the repertory grid (RG) – a more qualitative method, the aim of which is to assist in finding the best choice on the basis of in-depth interviews. The combination of weighted attributes obtained could also be used to construct a quality-constant geoindex included in the hedonic model (see Laakso et al. 1995, Bender et al. 1997, 1999, Din et al. 2001). With a quality model based on pair-wise comparisons with the AHP one can compare the elicitation of different interest groups for different type of areas or houses. (E.g. Nevalainen et al., 1990; see also Ruokolainen & Tempelmans Plat, 1998; Pöyhönen, 1998; Gregory, 2000). Yet another strand of the behavioral real estate literature is inspired by the seminal work of psychologists Tversky & Kahneman on heuristic problem solving (see Diaz, 1998; Goodman and Ittner, 1992; Adair et al., 1996; Daly et. al, 2003).

1.5. Comparison of methods
Based on our reading of the literature, it seems evident that value modeling is a growing field. Despite the lack of consensus with regard to a single, dominant framework, the great opportunities offered by computer assisted mass appraisal allow researchers to apply innovative procedures. They may be grouped into the following five areas:
• Strict quantitative method based on parametric hedonic regression (or more generally, economic equilibrium) models – either with or without an explicitly spatial extension.
• Flexible (non-/semi-parametric) quantitative methods; these are still formally accommodated within mainstream economic modeling.
• Flexible quantitative method based on machine learning; these are ‘intelligent’ (i.e. machine learning) methods from the computer science discipline.
• Rule-based expert systems of valuation, using either transaction price or expert interviews data; may be based on machine learning.
• Methods, where interviews and surveys such as multi-criteria decision methods replace market data calculation; usually both quantitative in the sense that they are mathematical, and qualitative in the sense that they use judgmental data. These stated choice/preferences methods serve a purpose in certain well-specified problem settings.

We highlight the following criteria (in descending order) to establish a protocol:
1. **Accuracy of independent valuations** (see e.g. Thibodeau, 2003)
2. **Conceptual soundness**
3. **Analysis of valuation variation** where more than one mass appraisal methodology is applied (Brown et al. 1998)
4. **Internal consistency of the model**
5. **Nature of the adjustment**. Is the adjustment of the model predominantly spatial/structural or temporal in nature? (Only in the latter case a smooth, linear model is valid)
6. **Reliability and robustness of the model** (i.e. sensitivity of result to external alterations in input parameters)
7. **Feasibility** (i.e. cost and time efficiency).

To give some examples, the multi-attribute approach is steadily gaining support. It seems flexible and is promising. Arguably, allowing for (consumer) behavior and quality improves the conceptual soundness of the value model (as suggested by Daly et al, 2003); therefore, MAVT probably is the most conceptually sound approach to valuation as it explicitly deals with such elements. It is, however, difficult to imagine a possible application of these works to mass appraisal. Besides, there are uncertainties yet to be solved when applying methods based on hypothetical data. The obvious one is the feasibility of the study and its cost. However, the most important concern is how to perform an objective comparison between estimates based on market data and those based on hypothetical values.

For the spatial aspect, the interview-based methods discussed above come particularly handy, as environmental info is not easy to record (see e.g. Langdon, 1978). However, the preferences and choices evoke under hypothetical conditions that do not necessary represent the actual choices made in the real world. Also temporal and spatial stability is hard to incorporate to these methods.

As each approach has its limits, the question is always about what particular aspect needs to be reflected upon. Is it accuracy, feasibility or some non-technical aspect that may require more thorough analysis such as the adjustment structure? We may find that, regardless of criteria used, the MRA does remain a credible method. Finally, we will introduce yet two ‘institutional criteria’ in the application of mass appraisal methodology: one, **suitability to the property market context**; two, **path-dependence** – is the institutional practice favorable towards introducing new tools? Particularly in developing countries, data quality is the main problem to overcome. For example, Pacharavanich and Rossini (2001) conclude that there is little
point in developing the methodology, if the data in general and the transaction prices in particular, are not good (cf. Kryvobokov, 2004).

1.6. Summary and conclusions
When reviewing the set of mass appraisal methods, the criteria are partly methodological and partly institutional. Not only do we forward claims about what constitutes an adequate and appropriate method, but also about the suitability of the method for the institutional context of use. The most common methodological framework for mass-appraisal is the hedonic regression model, which recently has been extended into various flexible and spatial regression models. This tradition may be divided into two main types: parametric and flexible (i.e. semi- and non-parametric) methods. The parametric variant is more efficient but prone to high specification error. This type is model-driven, which means that in-sample estimation is sufficient to determine the accuracy (but see Fletcher et al., 2003, who suggest that, even in parametric regression, out-of-sample testing is required). The flexible variant is inefficient but has a low specification error. This method is data driven, which means that out-of-sample testing is necessary to evaluate its performance.

Interview based methods merely deal with hypothetical data, and generate only hypothetical results about property value (which, paradoxically, is a hypothetical concept). If no good market data exist this approach becomes crucial to be able to obtain any results. Also, if market data is considered invalid for the specific appraisal task, in particular, if we need to ascertain the behavioral nuances involved, this category of tools become relevant.

Finally, the institutional dimensions should not be overlooked. Correlating valuation method with the market context may prove explanatory, when considering differing results achieved from different times or places. For example, a small or middle-sized city typically has less location factors influencing prices, and a more general functional relationship between such influences and the price, than a big city, where price influences are of several kinds, and often intangible and interrelated. Thus, there would be added value with a more complex model in the case of a larger city only.

When looking at the current trends of valuation modeling research, a likely scenario would be that price research is going towards spatial tools on one hand, and on the other hand towards pragmatic tools. Perhaps there is also a turn to more qualitative methods, but that depends on whether a qualitative method can be adapted cost-efficiently. Here two different arguments about how to advance the field can be forwarded:

- Following the orthodox view: to remain within equilibrium modeling and sophisticated econometric techniques in order to keep up with the academic tradition.
- Following a more heretic/heterodox view: to only consider the practical aspects of accuracy and feasibility; then, we can look beyond hedonic modeling extensions towards computer simulation, artificial intelligence and machine learning paradigms; the best result we get when combining two or more techniques.

One crucial requirement within the mass-appraisal pertains to incorporate the influence of the location into the model. A further issue is whether we need automation, which has to do with cost-savings and benefits generated by economics of scale.

The arguments discussed may be summarised in three crucial aspects, each of which poses potential problems for the estimation of value:
(1) The technical aspect: variables, transformations etc – this is traditional valuation territory.
(2) The institutional aspect: market efficiency, data availability – this is an area yet to be explored on both sides of ‘the pond’.
(3) How to adapt further data collection to progress or degradation in the surrounding environment or the standards of application? This issue is about capturing unknown influences and about robustness in relation to the model/criteria of the system of valuation.

The parts to follow deal with mass-appraisal practice (2.), orthodox methods (3.) heretic methods (4.), comparison of tools using a set of specific criteria (5.), and conclusions (6.).

References


