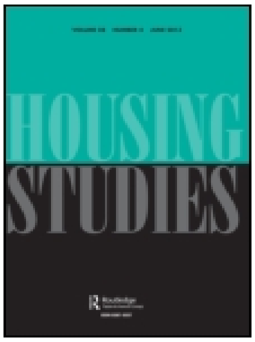


Housing Studies

ISSN: 0267-3037 (Print) 1466-1810 (Online) Journal homepage: http://www.tandfonline.com/loi/chos20

Capturing Housing Market Segmentation: An Alternative Approach based on Neural Network Modelling

TOM KAUKO , PIETER HOOIMEIJER & JACCO HAKFOORT

To cite this article: TOM KAUKO , PIETER HOOIMEIJER & JACCO HAKFOORT (2002) Capturing Housing Market Segmentation: An Alternative Approach based on Neural Network Modelling, Housing Studies, 17:6, 875-894, DOI: 10.1080/02673030215999

To link to this article: http://dx.doi.org/10.1080/02673030215999

Housing Studies, Vol. 17, No. 6, 875–894, 2002

Carfax Publishing
Taylor & Francis Group

Capturing Housing Market Segmentation: An Alternative Approach based on Neural Network Modelling

TOM KAUKO[1], PIETER HOOIMEIJER[2] & JACCO HAKFOORT[3]

[1]*OTB Research Institute for Housing, Urban and Mobility Studies, TU Delft, The Netherlands;* [2]*Faculty of Geographical Sciences, Utrecht University, Utrecht, The Netherlands;* [3]*Ministry of Economic Affairs, The Hague, The Netherlands*

[Paper first received 27 November 2000; in final form 7 May 2002]

ABSTRACT *Various location specific attributes cause segmentation of the housing market into submarkets. The question is, whether the most relevant partitioning criteria are directly related to the transaction price or to other, socio-economic and physical, features of the location. On the empirical side, several methods have been proposed that might be able to capture this influence. This paper examines one of these methods: neural network modelling with an application to the housing market of Helsinki, Finland. The exercise shows how it is possible to identify various dimensions of housing submarket formation by uncovering patterns in the dataset, and also shows the classification abilities of two neural network techniques: the self-organising map (SOM) and the learning vector quantisation (LVQ). In Helsinki, submarket formation clearly depends on two factors: relative location and house type. Price-level clearly has a smaller role in this respect.*

KEY WORDS: housing market segmentation, neural network modelling, submarket classification.

Introduction

Already in the 1950s and 1960s a group of housing economists at Columbia University (among others Grigsby, 1963) suggested that, apart from the continuous price effect caused by the quality of the location, also a more discontinuous effect can be observed, reflecting market segmentation into submarkets (cited in Jones *et al*., 1999). Housing market segmentation refers to the differentiation of housing due to income and preferences of the residents and administrative circumstances. Segmentation is real, when there are criteria that separate mutually distinctive submarkets from each other. Submarkets are defined as groups of dwellings, which are reasonably close substitutes (i.e. compensating goods) for each other, but at the same time bad substitutes for dwellings in other groups (Bourassa *et al*., 1997; Grigsby *et al*., 1987; Rothenberg *et al*., 1991; Whitehead, 1999).

0267-3037 Print/1466-1810 On-line/02/060875–20
DOI: 10.1080/0267303022000059515

Housing market segmentation has been an issue in recent work conducted at several universities and research institutes around the world (Adair *et al*., 1996; Bourassa *et al*., 1997, 1999; Jones *et al*., 1999; Maclennan & Tu, 1996; Morrison & McMurray, 1999; Sharkawy & Chotipanich, 1998). Much of the work approaches the segmentation process either along the lines of Schnare & Struyk (1976, cited in Leishman, 2001) or of Rothenberg *et al*. (1991, ch. 3). In the approach by Schnare & Struyk, housing submarkets are assumed to arise due to insufficient competition in the spatial housing market, a standpoint that implies a rejection of the ideal situation assumed by neo-classical economic theory. In the approach by Rothenberg *et al*. the view is that submarkets represent different price levels of housing that need to be adjusted for quality with a hedonic regression model, a standpoint more in line with economic theory. Whereas the first approach tests for non-price based segmentation and against spatial arbitrage, the second approach accommodates the segmentation aspect within housing market analysis in a more orthodox economic sense, by recognising the heterogeneity of the housing price formation. In short, any change in the submarket formation is assumed internally induced by Schnare & Struyk, but externally induced by Rothenberg *et al*.

In the non-price segmentation approach above, segmentation has been observed empirically to be based on six factors: (1) tenure or lease; (2) house type (or building format); (3) number of rooms, and qualitative dwelling characteristics; (4) source of financing; (5) age of building stock; and (6) location, both macro and micro. This paper emphasises the importance of location, even if the five other factors may be indirectly associated with certain compositions of location attributes and consequently with certain relative house price levels. The focus is on the methodology to uncover the most relevant partitioning criteria. The question is to what extent two neural network techniques: the self-organising map (SOM) and the learning vector quantisation (LVQ), are effective in identifying housing submarkets.

The use of the SOM as a property valuation method has already been demonstrated in a number of studies (Carlson, 1991, 1992; James *et al*. 1994; Jenkins *et al*., 1999; Kauko, 1997; Kauko & Peltomaa, 1998; Lam, 1994). As the housing market segmentation aspect can be argued to go hand in hand with the residential valuation aspect (e.g. Kauko, 1999; see also Jenkins *et al*., 1999), it is logical to extend the applicability of the SOM-based method of price analysis towards modelling spatial housing market dynamics in general and towards classification of housing submarkets in particular.

The remainder of this paper will begin with a brief positioning of the research among the theoretical approaches to segmentation, followed by a review of the methodology in general and the methodology based on neural network techniques in particular. The next section analyses housing market data from Helsinki with the two specific neural network techniques. The final section gives a summary and discussion about the findings of the study.

Theoretical Explanations of Segmentation

Theoretical approaches to explain segmentation may be classified as neo-classical economic, behavioural-cultural, or somewhere in between these two extremes. Although there is not a one-to-one connection of these approaches to the three types of principal criteria for empirical observation (i.e. price, other 'objective' criteria, or 'more behavioural' criteria), there is some correspondence of these criteria with extended equilibrium economic, localised disequilibrium, and socio-cultural theory perspectives. The main difference between these perspectives lies in the assumed effect of preferences, behaviour and institutions: in the purely economic approaches they are given exogenously, whereas in the more behavioural and institutional approaches they are allowed to vary more explicitly.

Neo-classical Economic Models

Even though being primarily an equilibrium model of land and property price formation accommodated within an urban land use setting, the trade off -theory of residential location (Alonso, Muth, Mills) applies for segmentation as well. Even if other substantial value factors do not exist, an urban area may be segmented, if the preferences and/or income of the households are different with respect to space and accessibility. In addition, the later land use/environmental preferences approach (Evans, Richardson, Wheaton) can be seen as an explanation for the occurrence of segmentation. Besides, segmentation within an urban area may be based on additional factors, such as the dominating type of building, plot efficiency or even the internal attributes of the dominating type of apartment (see Bourassa *et al*., 1997; Grigsby *et al*., 1987; Laakso, 1997).

Localised Disequilibrium

In this perspective, arguments that stress the fragmented nature of the housing market are taken more seriously (see e.g. Grigsby *et al*., 1987; Maclennan & Tu, 1996). In reality different buyers have different housing preferences and face a variety of dwelling alternatives which may not comprise a single market. Therefore Maclennan & Tu (1996) suggest a non- or partly co-ordinated view as opposed to the dominating 'unitary equilibrium' view, claiming that there is no point to model housing markets within an instantaneous equilibrium model. In such a framework the focus would be in processes of adjustment rather than in what they call 'standard outcome' data. Hence, the assertion for 'persistent localised disequilibrium' caused by both spatial and sectoral factors in either supply or demand side diversification.

Behavioural and Cultural Models

Recent theoretical extensions have come from socio-cultural and actor-led institutionalist positions. As an addition to dominantly economic theories of segmentation and segregation a variety of theories have been developed stemming from a socio-cultural background. These emphasise the diversified consumption of the household. According to the so-called constructionist perspective segregation is socially produced. According to this interpretation, there are no facts, but they have to be elaborated instead. According to the theories of ways of life and everyday practices segregation is a result of the socio-cultural choices of the consumers. (e.g. Ilmonen, 1997, pp. 20–23.) The theories in question complement one another and no sharp boundaries can be drawn between them. What they have in common is that they emerged as a criticism of the assumptions on rationality in the neo-classical models and became widespread during the 1980s.

Positioning the Research

Studies on housing market segmentation conducted in various housing related disciplines differ with respect to the definitions and methods used. The essential question is whether the reasons behind segmentation are 'objective' criteria pertaining to assumptions of uniform preferences among individual residents, or socio-cultural factors and human behaviour.

This paper uses a market data-based approach, and consequently restricts the analysis to determining 'objective' criteria as opposed to looking for more behavioural reasons behind submarket formation in an urban context, as some of the theories above suggest. The neo-classical approach is chosen as a basis. However, it is fair to say that, at least partially, segmentation is dependent on both institutional and socio-cultural factors: the submarkets emerge on the one hand because of constraints set by policy makers, and on the other hand because of different ways of perceiving quality. Therefore, the most fruitful perspective for the study on segmentation would suggest recognising these factors also in the operational model (Rothenberg *et al.*, 1991, ch. 3).

Existing Techniques to Detect Submarkets

Viable tools for empirical research on housing market segmentation include hedonic models, projecting and clustering techniques and spatial statistics. (This is not to deny that also technically simpler methods such as interviews and descriptive statistics are being frequently used for the same purpose.) These tools will be discussed below to create a point of reference for the neural network technique, the choice for the empirical investigation to follow.

The hedonic model of housing markets may be understood as the multi-dimensional extension of the Alonso-Muth-Mills model. Location and the other factors of segmentation mentioned above may be included in straightforward hedonic modelling as well. However, 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. The issue can however be clarified by using dummy variables. Additionally, and not surprisingly, an isolated price effect, such as view, may be found context dependent by using interaction variables (see Benson *et al.*, 1998; Wolverton, 1997). Essentially, the hedonic model does not attempt to explain absolute prices, but only the variation in prices within one market, where the behavioural and institutional factors are assumed constant.

A pragmatic way of dealing with multiple submarkets within this approach is to partition the data. The data is split into different segments, which are either *a priori* predefined or synthesised somehow. If hedonic models estimated for partial datasets then differ from each other significantly, it indicates a segmentation (e.g. Laakso, 1997). According to Maclennan & Tu (1996), the grouping of dwelling units should start from their observable characteristics (including location) rather than in relation to *ad hoc* sectoral or areal aggregations.

In the widely applied method proposed by Rothenberg *et al.* (1991, pp. 380–385) a hedonic index is estimated in order to calculate hedonic values for each house within the sample. The hedonic values are then ranked into classes based on their quality levels, regarding the characteristics of the house (number of rooms, age, plumbing facilities, condition and tenure). These classes are thus referring to ranked clusters with internal substitutability and can be used as a basis for partitioning the total market into submarkets. Technically this method builds on competitive market assumptions, where submarkets are solely determined by price-related criteria. In addition, the 'hierarchy of price groups' method deployed by Costello (2001) is in line with such assumptions. The aim of this technique, taken from the market efficiency literature, is to capture price changes for each group, but only for the middle part of the market.

A more open variant in the broad hedonic approach to managing the segmentation of data is to chain different statistical methods to one another, pre-processing the data before the hedonic model is applied. The summarising of multi-dimensional transaction data into less dimensions is made by the tools of factor analysis, which include projection methods such as principal component analysis (PCA), or multi-dimensional scaling (MDS). Then, the reduced dimension data are divided into submarkets by cluster analysis: either discriminant, hierarchical or partitional techniques. Finally, the intrinsic estimation of price is performed with respect to each segment by hedonic regression.

Bourassa *et al*. (1997, 1999) used this method of combining three different statistical techniques to analyse residential submarkets in Sydney and Melbourne, Australia: (1) PCA in order to extract a set of factors from the original variables; (2) cluster analysis on the principal components in order to determine the most appropriate number and composition of segments; and (3) multiple regression analysis in order to estimate the hedonic price equations for all the submarkets separately and each city as a whole. They also used *a priori* classified submarkets. Finally, the weighted mean square errors from the equations were compared to determine the most appropriate classification of submarkets. (Another method is to study the variation in coefficients between subsamples, see e.g. Leishman, 2001.) However, they noted that the optimal number of submarkets still remained difficult to determine based on the cluster analysis literature. (See also Leishman, 2001 and Ball & Kirwan, 1977, for applications based on the statistical methods above and house price data.)

To explore the relationship between housing submarkets by examining *household mobility patterns* is an idea originally developed by Grigsby in the 1960s. Such an analysis is based on data on intra-metropolitan household migration and open market transactions. For instance, in a study on Glasgow, UK, by Jones *et al*. (1999), six potential submarkets were chosen *a priori*, and later these were aggregated to three submarkets by hedonic price modelling. When cross-tabulation was undertaken to determine the mobility percentages between submarkets, a rather strict containment was observed in each submarket, which confirmed the argument about segmented markets and neighbourhood attachment.

To present a last 'conventional' technique, demand functions, where collective preferences of the households are related to the membership of an *a priori* defined ethnic or socio-economic group, are sometimes estimated with the specified two stage procedure of hedonic modelling. In this technique willingness to pay (WTP) -estimates for distinct house price characteristics among social groups are derived based on the coefficients of the hedonic price equation and demographic, socio-economic and other household characteristics (e.g. Bökemann & Feilmayr, 1997; Laakso, 1997).

As the sample of studies shows, market segmentation is usually analysed with various methods loosely tied to the hedonic approach: multiple regression analysis is applied either as a final procedure (partitioning approach) or as a preliminary procedure (WTP, mobility patterns). Instead of just heuristics, the analysis involves a mathematical rigour, thus satisfying the need for more 'scientific' analysis (see e.g. Adair *et al.*, 1996). Furthermore, the criterion chosen and the particular method used varies a lot, which broadens the perspective of analysis a great deal. The underpinnings of the hedonic models are also fairly easy to demonstrate to the end user.

Recently the spatial statistics family of methods has begun to make important contributions in detecting housing market segmentation (see e.g. Dubin, 1992; Dubin *et al.*, 1999). The idea of the kriging-technique is to utilise the dispersal of residual errors to construct a distance decay type of function, which subsequently can be used to improve the accuracy and efficiency of the model. The further the observations are situated from the target observation, the less they contribute to the value effect of the latter. For example, Dubin *et al.* (1999) have emphasised the importance of nearby properties, when the house price estimate is a function of proximity and degree of spatial dependence. Other newcomers include semi- and non-parametric (i.e. flexible, model-free) regression techniques. In one study using this technique, Kyllönen & Räty (2000) undertook a hedonic modelling of the housing market(s) of Joensuu, Finland with a partial spline function-extension, thus combining both parametric and non-parametric components into an additive, semi-parametric model.

It is obvious that the various partitioning, spatial and flexible modelling techniques bring more options to study the housing market structure of an urban area. This should improve the results, given that submarkets exist in a given spatial and temporal context, and that they can be measured and related to theory. Interestingly, some of the new tools are based on the paradigm of machine-learning, using either symbolic or numerical information processing, such as the neural network approach or the genetic algorithm. In fact, the neural network modelling technique is also categorised as a semi-parametric regression technique.

A successful method/technique must be able to manage the market segmentation on an aggregate level, where the various anomalies caused by institutional and physical constraints are discernable. One of the issues closely related to this is how to capture outliers that nevertheless may be important in the models. The encouraging findings obtained by Kauko (1997) indicate that this is possible within a SOM-based approach (see next section), as it performs data projecting, clustering and estimation, much like the combined approach by Bourassa *et al.* presented above. In fact, the SOM is more apt for detecting rather than estimating purposes (e.g. James *et al.*, 1994). This enables performance comparisons between the SOM and more conventional statistical methods (see e.g. Kaski, 1997). Later, the study will compare the partitioning based on the SOM with *k*-means clustering.

The SOM-LVQ Classifier

Neurocomputing or the (artificial) neural network comprises an emerging category of numeric, learning or 'intelligent' techniques, and a sort of flexible, model-free regression (see e.g. McCluskey & Anand, 1999; Pace, 1995; Verkooijen, 1996). The nature of the neural network is a black box, which means that there is no clear functional relationship between the input and output

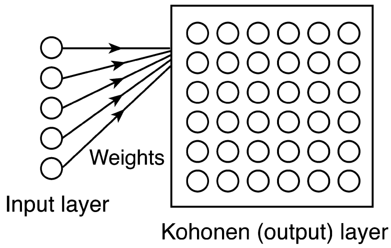


Figure 1. The principle of the competitive network architecture. *Source:* modified illustration based on James *et al.*, 1994.

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 a numerical signal coming from outside the network in such a manner that a connection, between input and output information is developed. The connection is referred to as the intelligence of the network.

The self-organising map (Kohonen, 1995) is a type of neural network with a competitive network architecture (see Figure. 1). In the current application the input is a set of individual transaction data that list the attributes of the dwellings (including location attributes). The SOM reduces this input to a limited, *a priori* defined, number of nodes that are ordered into a two-dimensional lattice (Kohonen *et al.*, 1996a). The output therefore looks like a map. Each node represents a characteristic combination of attribute levels (or categories in case of nominal attributes), and a 'typical' value for every attribute is calculated for each node. Each node with its attribute levels may also be displayed using geographical identifiers (e.g. postcodes or municipal codes).

The starting point for using the SOM is to initialise the map by generating random values for each node. The training procedure of the algorithm then proceeds in three stages: first, select a training vector x; then find the best matching neuron, node c, that is closest to x; finally, adjust the node c and its neighbours towards the observation x (e.g. Koikkalainen, 1994). Usually the matching is determined by the smallest Euclidean distance between node c and vector x (e.g. Kohonen *et al.*, 1996a).

The technique is based on the principle of unsupervised competitive learning, which could be described as 'the winner takes all'. Thus, the winner is the node with shortest distance to the observation vector, and its weights are adapted towards the observation (see Figure. 1). This continues until all desired observations are used for training, usually more than once. Neighbouring nodes on the map are being similarly adapted towards the observation, but the extent of this depends on the selected parameters. In a SOM-context the 'neighbourhood' concept is defined as the 'winner' with adjacent neurons.

Stated differently, the original observations are uniquely linked to the node that 'won' them. At the same time 'empty' nodes can also occur, as the dataset might not contain relevant cases for that part of the structure. The typical values however are sophisticated averages, in the sense that every observation on the map affects the values. This produces relatively smooth surfaces in the visual inspection of the map in which outliers can be identified as 'patches' shown up in this surface. By comparing the typical values to the ones from the original dataset a measure of goodness of fit can be derived. Typical values can also be used to estimate values for items and even cases that are missing in the data and hence also used for price estimation purposes of transactions that have not (yet) occurred.

When using the SOM, the dimensions of the map and the learning parameters (network parameters) which have an impact on the outcome, have to be chosen beforehand. These are determined much in an *ad hoc* nature, but Kohonen and others give some guidelines. The relative contribution of each network parameter is defined as follows (see also Kohonen *et al*., 1996a):

- *the dimensions of the map* (i.e. the number of nodes); a larger map leads to a better resolution, but to a less parsimonious model (e.g. Tulkki, 1996, p. 14); furthermore, due to the nature of the algorithm, the organisation is more stable if the dimensions of the map are chosen as different sizes (Kohonen, 1995; Kauko, 1997);
- *the running length* (i.e. the number of learning steps in training); there is a trade-off between accuracy (possibly validity) and time-saving (feasibility); Kohonen (1995) gives the advise that the number of steps during the fine-tuning stage of the run should be 500 times the number of nodes on the map;
- *the initial learning rate parameter* (α); determines the length of each step and eventually decreases to zero during training; the only advice is to choose it small but not too small;
- *the initial radius of the training area*; decreases to one during training; should be chosen smaller than the size of the diagonal of the map, but for the basic run not much smaller than it;
- *the neighbourhood function*; can be either 'bubble', when only the neighbourhood (i.e. the immediate neighbouring nodes) affects the node to be trained or 'Gaussian, when all the nodes take part in the organisation process, but the weaker, the further they are situated from it; in Kauko (1997) bubble generated a more accurate result than Gaussian.

The learning vector quantisation (LVQ), is an extension of the SOM. It is suitable for testing and also for improving the classification provided by the feature map (Kohonen, 1995). The LVQ is based on the principle of supervised competitive learning. The principal idea of this algorithm is that the observations are now approximated into various classes of the input vector x. A variety of labelling options may be tried; simple *a priori* classifications, or *a posteriori* classifications based on the resulting patterns in the feature map. Using these labels of the observations the feature map may then be calibrated in such a way that each node on the map gets a corresponding label based on the resemblance to a certain class of observations (i.e. the smallest Euclidean distance between the response and the observation). Finally, the accuracy of the classification is determined, preferably with a set aside sample. The classification performance is evaluated by the recognition accuracy, the ratio (percentage) of successful hits on average over all classes. (Kohonen *et al*., 1996b)

The LVQ clearly adds credibility to the SOM-based submarket identification. The unsupervised SOM does not give us a full verification of the meaningfulness of a certain partitioning, given the context of study. To be able to determine the success of various clustering solutions generated by the SOM over the whole sample more rigorously than the visual interpretation allows, two additional techniques are needed: first, one that compares the response of the neurons with the original sample, and second, one that improves this correspondence between input and output towards an ideal situation, in which the risk for misclassification is minimised.

The optimisation routine of the SOM is to minimise the difference between x and c. The result is dependent on the initial values of the map and on exogenously administered parameters, such as the number of iterations. The longer the run, the less sensitive the map becomes to random input shocks. The result is expected to converge towards a global optimum, but at the moment of stopping the run, one cannot know for sure whether that is reached.

For the LVQ in turn the optimisation routine is to maximise the recognition accuracy, ideally to 100 per cent. Additionally, the length of the run may determine the result. The right moment to stop the training has to be determined by comparing the accuracy results obtained with the training and test samples after each run. Over-training occurs when the network instead of learning from the training sample begins to memorise it, in which case the accuracy of the results with a test sample begins to decline. This is the stopping criterion for the training procedure in algorithms based on supervised learning (see e.g. Borst, 1995; Worzala *et al.*, 1995).

To summarise, neural networks arrive at results through an iterative process, where the input is linked with the output and the linkage is adjusted by weights. The results are strongly dependent on the data. Unfortunately the lack of a straightforward functional relationship between the input and output creates a problem of 'explainability', that is the classic black box argument. Nevertheless, the ability to visualise patterns and partition the sample with the SOM-LVQ-method suggests some potential in overcoming shortcomings of the present methods in capturing market segmentation. This method could be used partly as a substitute for the commonplace techniques and partly as a complement. The following general conclusions of the usefulness of the SOM-LVQ-method may now be drawn:

- patterns are formed based on the input variables, where similar categories of observations, which are clearly different from others, form clusters on the map surface (see Figure 2 in the next section);
- some of the patterns may be unanticipated, especially if there are correlations among the most important variables affecting the organisation of the map;
- some of the patterns reflect the segmentation (location or other criteria);
- the choice of input variables allows socio-economic and environmental comparisons between different locations, as well as comparisons between different dwelling types at the same location;
- with the SOM alone, only a rough classification is obtained (hence a further processing with the LVQ);
- within one submarket or locality the price dimension of the SOM output (i.e. typical values) may be used as an indicator of attractiveness for one category of observations compared to others, which allows linking the results to the price dependent criteria for segmentation.

Submarket Classifications in Metropolitan Helsinki

The Study Area and the Data

Based on casual observing and previous research we know that the housing markets of 1990s Helsinki are segmented (e.g. Kortteinen & Vaattovara, 1999). However, it is more unclear what the relevant criteria are for segmentation, price level or any other 'objective' characteristics of the housing or environmental bundle. When attempting to investigate the matter empirically, the neural network approach to classification of market segments outlined in the previous section will be used. Following Kauko (1997), where the method itself was explained, the focus is on the results of submarket structure obtained with the SOM-LVQ neural network approach.

Helsinki metropolitan area (henceforth Helsinki) consists of four municipalities: Helsinki, Espoo, Vantaa and Kauniainen. The population of the whole area is about 950 000, with the City of Helsinki consisting of 60 per cent of it (2000). Helsinki metropolitan area is the central part of the greater Helsinki region, which is by far the largest agglomeration in the country, with approximately one-fifth of the Finnish population.

In Helsinki there are 400 000 dwellings, of which approximately 60 per cent are owner occupied (Laakso & Loikkanen, 1995; see also www.stat.fi). In 1993 the Finnish housing market was still in a recession. At the time of writing, almost a decade later, the price levels are substantially higher in most places. Nevertheless, the price levels reflect structural differences which depend on the differences in attractiveness between areas, most of which prevail independently from cyclical price fluctuations. The city centre of Helsinki, the suburbs on the coastline of western Helsinki and southern Espoo (west of Helsinki) as well as some suburbs on the eastern coast of Helsinki, the tiny municipality of Kauniainen situated within the borders of Espoo and some of the low density areas in northern Helsinki include a clear price premium, while others accommodate some cheap housing in areas distinguished by large housing estates built in the late 1960s and 1970s.

If the price per square metre is considered instead of the total transaction price, as will be the case in this study, the picture changes somewhat. There is a remarkably strong association between high prices per square metre floor space and a good accessibility to the CBD of Helsinki. Thus, the role of the land price gradient is substantial in explaining (per square metre) house prices of Helsinki. However, in this study the input variables chosen are direct measures of the physical and social compositions of different locations, and only possible indirect measures of the accessibility factor. Furthermore, some of the locational differences are qualitative and discontinuous by nature, more than gradual changes in consumer purchase power and distance decay.

The most important findings from recent urban housing modelling studies on Helsinki by Laakso (1997), Lankinen (1997), Maury (1997), and Vaattovaara (1998) largely agree on at least three distinctive prevailing housing submarkets: (1) the inner part of Helsinki (if necessary, this may be split into two further submarkets dependent on the status of micro-location); (2) multi-storey housing in suburban districts, the most common type of residential area in urban Finland; and (3) terraced, detached and semi-detached houses in suburban districts. Casual observing also supports this segmentation. All this information

Table 1. Description of the variables

Micro-level variables:
(1) price of the dwelling per square metre (1000 FIM)
(2) age of the building (10 yrs)
(3) dwelling format: (semi-)detached 1, terraced 2, multi-storey apartment 3, else 5
(4) number of rooms

Statistical subarea-level variables:
(5) number of commercial services in the sub-area/10
(6) number of public services in the sub-area
(7) amount of undeveloped land in the vicinity within a two kilometre range
(8) 'status', compounded by: proportion of the population with higher educational degree, average income of the working population, proportion of owner occupied dwellings, the unemployment rate
(9) level of negative social externalities, compounded by: unemployment rate, proportion of ARAVA-tenancies (the publicly financed housing-sector) of all rented dwellings, proportion of foreigners, crime rate
(10) inverse indicator of urbanisation, compounded by: median year of construction, proportion of detached or semi-detached housing, average density

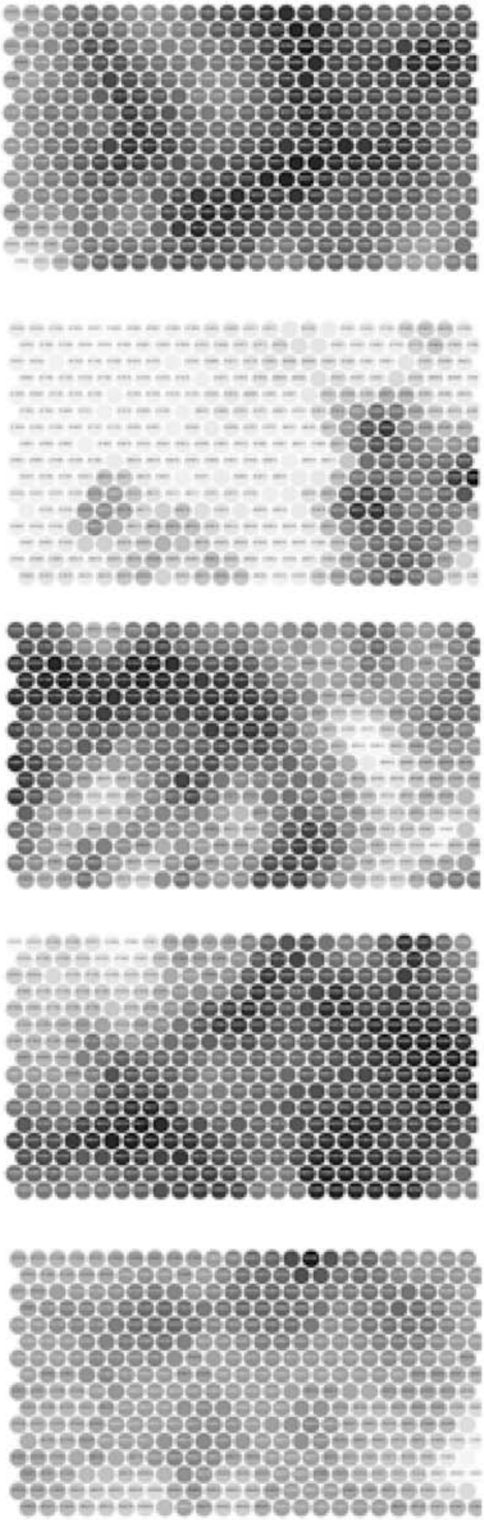
provides a convenient benchmark for empirical analysis to be performed with the neural network method.

The exploration of house prices in Helsinki is based on dwelling transactions during one year (1993) in the Helsinki Metropolitan Area, with locational attributes aggregated on the statistical subarea-level for each transaction. Table 1 describes the variables for the dataset. There are four structural and six locational variables available in the dataset (see Table 1). The data is taken from the sources of Statistics Finland. Three of the locational attributes show the amount of commercial and public services, and the share of open space in the surrounding area. The remaining three locational variables (status, social externalities and urbanisation) were constructed by principle component analysis (by Seppo Laakso). For the visual interpretation and especially for classification with the LVQ, various labelling options, most notably the location, were added to the observations in both datasets. The labels did not affect the calculations of the SOM-algorithm.

The Visualisation of the SOM Analysis

As discussed earlier, the SOM-algorithm learns by training, and produces a feature map of neurons, which are represented as different characteristic combinations of attributes. Furthermore, it is possible to interpret a typical value of each node, for a given feature, as an indicator for the apartment or, given a suitable labelling, location. To enable visual examination of the feature maps, differences in this value estimate across the map layer is depicted through the shade of grey: the lighter the shade, the larger the value of the variable in question.

As already mentioned, in this application of the SOM the results in a quantitative sense depend on the operational choice of a number of network parameters. In this run, among others, 24*16 map dimensions, a bubble neighbourhood function, and a running length in line with Kohonen's recommendations were chosen.



Feature map illustrating price levels in Helsinki subareas (dark colour = cheap area; light colour = expensive area).

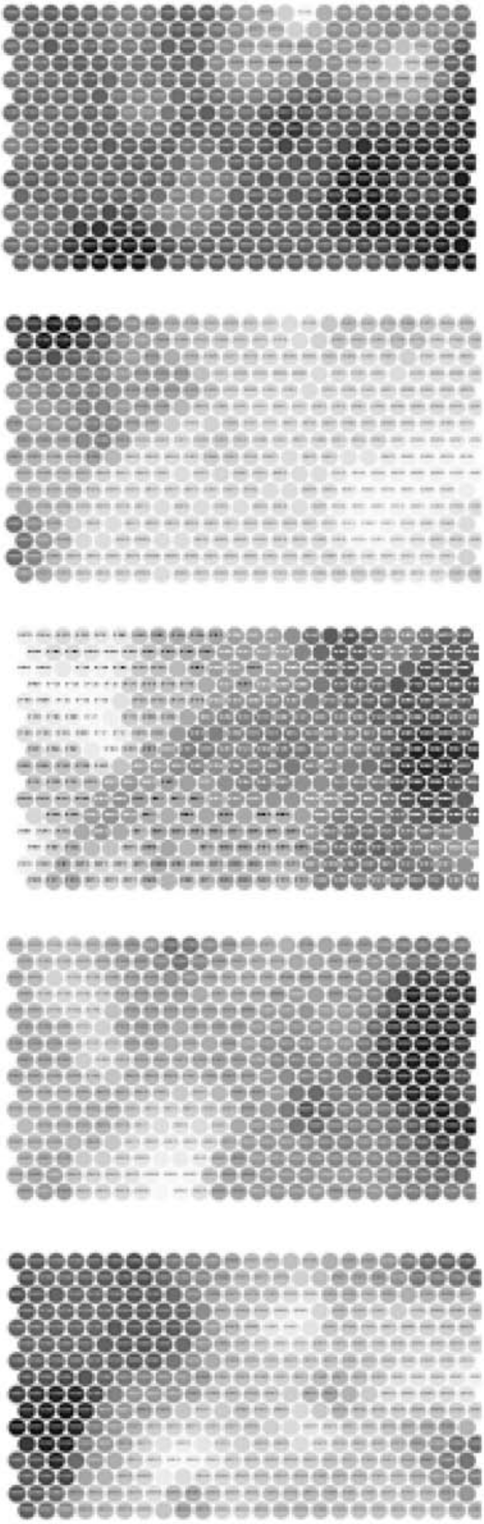
The dwelling format (dark colour = single-family houses; light colour = multi-storey houses).

The number of rooms (dark colour = 1–2 rooms; light colour = 3+ rooms).

Age (dark colour = new buildings; light colour = old buildings).

Status (dark colour = low status; light colour = high status).

Figure 2. Feature map layers of



Negative social externalties
(dark colour = low levels;
light colour = high levels).

The 'urban' indicator
(dark colour = most urban areas;
light colour = least urban areas).

Commercial services
(dark colour = low levels of services;
light colour = high levels of services).

Public services
(dark colour = low levels of services;
light colour = high levels of services).

The 'open space' indicator
(dark colour = least undeveloped
land in the area; light colour = most
undeveloped land in the area).

the Helsinki housing market.

The feature map illustrated layer by layer in Figure 2 revealed the specific nature of the spatial housing market structure:

- single-family housing forms two separate homogeneous clusters: (1) a larger group comprising areas of mixed nature from all three main municipalities, which suggests a physically homogeneous space across municipal boundaries, and (2) another, much smaller group comprising an area in southern Espoo;
- the most expensive areas (in the left corners of the map) are the most urban, with the least open space and the best commercial services, mostly older neighbourhoods;
- the cheapest areas are positioned along the upper middle part and right corner, on the edge of the map, most notably the outlier neighbourhood Jakomäki (9th neuron from the upper right corner in the uppermost row of neurons), a symbol for poverty and social externalities; these are also areas with low status and high social externalities;
- the newest building stock is positioned on the right-hand side of the map;
- average, less interesting cases are situated in the centre of the map (as the SOM is able to capture the interesting types of observations and outliers);
- four or five groups of varying nature are comprising larger dwellings.

The visual SOM analysis obtains the same three or four submarkets as expected based on the *a priori* knowledge of Helsinki housing markets explained above:

(1) locations in the inner city and the nearest old suburbs:
    (a) absolute top location, high price areas;
    (b) older, low status working-class areas, still relatively high price;
(2) other locations, dominated by multi-storey housing; mostly low status and price levels;
(3) detached and terraced housing.

The possibility to compare the results from the SOM-clustering with *k*-means clustering was mentioned earlier. Given Kaski's (1997) comparison of techniques, an interesting question arises: how similar or different are the results with the SOM to the results with cluster analysis from the same area? To answer the question, the feature maps were compared with the results from a *k*-means clustering run with the same data. (Obviously also the assumption of the *k*-means run had to meet the same number of submarkets as with in the SOM.)

If four segments are defined as a basis for the clustering, a *k*-means cluster analysis obtains the same submarkets as the SOM. The interesting thing is that if only three segments are defined, then the clustering is different from (1)–(3) above. (a) is one segment, (b) is another, and the categories (2) and (3) become the same large segment (actually two-thirds of the observations).The overall outcome of the comparison was that the difference (in Euclidean distance) between the measured average vectors of (2) and (3) is in reality much bigger than the difference between (a) and (b), and this is not, given the variables used, captured by the *k*-means algorithm, but to some extent by the SOM.

The good qualitative results notwithstanding, there are a variety of problems regarding the technical assumptions of the analysis, that one needs to be aware of. The first question is how to pre-process the data, especially how to determine the optimal field-range of a given variable (scaling, cf. assignment of attribute weights, McCluskey & Anand, 1999), as SOM variables with a wide field-range are more dominant in the structuring of the map. The second question (in this particular SOM-application) is how to select optimal network parameters, which also might have a substantial effect on the outcome (e.g. Kohonen *et al.*, 1996a). As already explained, these include the dimensions of the map, the number of steps in training, the initial learning rates and the initial radius of the training. One suggestion for using the SOM is to conduct several runs with different parameters, visually select a well-structured feature map, and compare values relatively within it. The final questions concern the size of the dataset and the repeatability of the results. Due to the non-parametric nature of the method, an enlargement of data-set results in more complex multidimensional patterns, but on a rough level the clustering prevails (Kauko, 1997).

The Results of the LVQ Post-processing

The meaningfulness of the clustering was tested using two independent samples, and efforts were made to improve it with the LVQ-network. Each observation was given a label based on its attributes (location or combination of characteristics) and after a calibration procedure also the output, the feature map obtained this labelling. The idea was to apply the LVQ to compare the *a priori* labelled data with the corresponding neurons on the feature map, each of which was representing a particular category of observations. As already mentioned, the classification (or labelling) accuracy means the percentage of successful matches between the observations and the response, based on the given labelling.

Table 2 provides a verification of the segmentation based on classification results for various labelling criteria. It shows how a variety of specific classification criteria are compared with respect to the recognition accuracy, keeping the sample, map size and network parameters constant across the runs. The classification is based either on an *a priori* chosen criterion or an *a posteriori* chosen clustering. In principle, the less the number of labels, the easier the task for the algorithm, and the better the expected classification result. The first observation is that when only a few labels are used the result, the captured market segments, are logical and coherent.

The very high levels of classification accuracy imply segmentation within the dataset. Perhaps surprisingly, the best classification result is obtained with the dichotomous open space indicator. As expected, the urbanisation indicator, which proxies CBD accessibility, gives a good result as well. The house type also matters, as expected. The age of the building serves as a proxy for location (possibly also an independent effect as a proxy for aesthetic values attached to the architecture or design), and is as such very important. Both types of services are important and also the macro-location matters: either the municipality (Helsinki, Vantaa or Espoo) or a more specific grouping of areas based on the SOM clustering, where two groups comprising dwellings in northern, eastern and north western Helsinki, Espoo and Vantaa form separate groups from the rest of the data. These areas have poor services and low density, and are located relatively far away from the centre of Helsinki. The majority of dwellings in these areas have an average or low price per square metre, are situated in average or new building stock, and comprise three or more rooms. The conclusion is that a segmentation based on house type, location and other factors (age?) seems more appropriate in Helsinki than a segmentation based on price-levels.

Table 2. LVQ-classification of Helsinki market segments

No. of labels and criterion	Exact definition of labelling criterion	Classification accuracy, validation sample (training sample in brackets)	Success of supervised training with the LVQ: the best map before overtraining occurs; test sample
2 open space indicator	Amount of undeveloped land in the vicinity within a 2km range: 0–4.99 km^2/5.00 + km^2	99.32% (99.37%)	(what is the added value with trying to improve this accuracy?)
2 location in relation to CBD	Area urbanisation indicator (good proxy for accessibility): $< -2/> -2$	97.98% (98.30%)	—
2 age	0–49 yrs/50 + yrs	96.71% (96.50%)	—
2 location combined with factors	*a posteriori* clustering based on the organised maps: certain suburbs/the rest of the data	95.64% (95.17%)	—
2 public services	No. of public services in the area: 0–49/50 +	94.95% (95.35%)	—
2 commercial services	No. of commercial services in the area: 0–39/40 +	91.38% (91.76%)	—
2 location	Municipality: Helsinki/else	88.50% (90.33%)	Improvement = > 91.58%, but no clear overtraining
2 house type	Dwelling format: multi-storey apartment/other	88.25% (89.26%)	Improvement = > 93.11%, but no clear overtraining
2 negative social externalities	Area sos.ext-indicator: positive/negative	87.51% (88.21%)	Improvement = > 91.95%,
3 location combined with other factors	*A posteriori* clustering based on the organised maps: 2 separate groups and rest of the data	87.47% (93.75%)	marginal improvement = > 87.62%, but no clear overtraining
2 price per sq.m.	FIM 7369 or less/FIM 7370 +	87.32% (86.82%)	Improvement = > 94.22%, but no clear overtraining
2 status	Area status-indicator: positive/negative	86.43% (88.05%)	Improvement = > 95.14%, but no clear overtraining
3 (4) location	Municipality: Helsinki/ Espoo/Vantaa (/Kauniainen minor segment)	85.39% (87.43%)	Improvement = > 91.86%, but no clear overtraining
3 price per sq.m.	FIM 4869 or less/FIM 4870 — 9869/FIM 9870 +	81.54% (83.20%)	Improvement = > 90.30%
3 age	0–24 yrs/25–74 yrs/75 + yrs	81.27% (83.61%)	Improvement = > 91.67%, but no clear overtraining
2 size (rooms)	1–2 rooms/3 + rooms	70.47% (72.08%)	Improvement = > 88.21%
3 size (rooms)	1 room/2 rooms/3 + rooms	54.92% (57.96%)	Improvement = > 73.65%
4 size (rooms) ~ price/sq.m.	1 room/2 rooms/3–4 rooms/ /5 + rooms	46.50% (47.58%)	marginal improvement = > 48.61%
~ 400 micro-location	subareas	30.56% (35.55%)	too difficult

Finally, the supervised training properties of the LVQ-algorithm (i.e. reallocation of the nodes according to an expected match between observation and codebook vectors) may be used to improve the classification accuracy of the feature maps. This way the input and output are being trained based on the matching of the labelled observation and codebook vectors. The training is done in a predefined number of iterations. In a LVQ context an iteration means comparing the labels of all observations in the sample with the corresponding response once. Then the result in terms of classification accuracy is checked with an independent sample, and if the result is not satisfying, the training continues with a new run of iterations. In addition, the results of the supervised runs were rather promising: for most labelling alternatives an improvement in the classification accuracy has been achieved, and for many criteria the classification accuracy of the trained map exceeded 90 per cent with a set aside sample.

Summary and Discussion

Can housing market segmentation be captured with a neural network modelling and classification approach? Some findings from Helsinki suggest a positive albeit modest answer: the neural network must be applied with care. The exercise has shown that by visualising realistic clusters and patterns on the map, it is first found where potential market segments are situated, and second what their determinants are. These determinants are not only input variables but also other features based on interactions among variables. However, these results are difficult to relate to any of the economic or socio-cultural theories of housing market segmentation.

The analysis could also have been conducted through using and linking more conventional statistical methods (factor, cluster and regression analysis). However, another interesting property of the neural network method is that a quantitative input possibly leads to a qualitative output. Furthermore, as an alternative to the standard methods reviewed in the third section, by post-processing the SOM output with the LVQ, the relevance of various measurable factors that potentially cause segmentation may be ranked. Based on these rankings it may then be concluded, whether the predominant criteria for submarket formation is the transaction price, as argued in the more orthodox economic -based modelling literature, or other factors that describe the nature of the dwelling and its location, as argued in the other main body of literature. In the case of 1990s Helsinki it was shown that segmentation clearly depends on two factors: relative location and house type.

The purpose of the study was not to present a detailed evaluation of the method and there is still plenty of ground to explore within this realm. A number of methodological issues were raised in this paper but were not elaborated on, namely the sensitivity of the SOM to: the data-definition (scaling of the field-range in particular), the selection of network parameters; and the size of the dataset. Further testing of the robustness of the model will have to be undertaken. A related, more substantive issue is the stationarity of the results over time and space. Repeating the analyses in a different period or a different market will generate other results. Due to the black box nature of the method it will remain unclear which functional relations account for the variation in results. The interpretation of results relies heavily on the expert knowledge of the researcher. It is clear that the neural network approach is not apt in testing predefined hypotheses and that even linking outcomes to existing theory is problematic. Yet in particular in the study of housing market segmentation, the exploratory nature of the neural network approach will be helpful in generating ideas on the processes that shape submarkets.

Acknowledgements

An earlier draft of the paper was presented by Tom Kauko at the ENHR Conference workshops in Gävle 26–30 June 2000. The authors are indebted to helpful comments from many persons, notably participants of the above mentioned conference, Seppo Laakso from Helsinki (also for the use of his dataset) and the staff at the Urban Research Centre Utrecht. We also note the software packages used: *SOM_PAK* and *LVQ_PAK*, in *MS/DOS*, produced by the Laboratory of Computer and Information Science at the Helsinki University of Technology. Finally, thanks are due to the three evaluators of the draft manuscript, namely Moira Munro, and two anonymous referees for their very helpful suggestions that undoubtedly helped in improving the paper to its final format.

Correspondence

Tom Kauko, OTB/TUD, Postbus 5030, NL-2600 GA Delft, The Netherlands. Email: kauko@otb.tudelft.nl

References

Adair, A. S., Berry, J. N. & McGreal, W. S. (1996) Hedonic modelling, housing submarkets and residential valuation, *Journal of Property Research*, 13, pp. 67–83.

Ball, M. J. & Kirwan, R. M. (1977) Accessibility and supply constraints in the urban housing market, *Urban Studies*, 14, pp. 11–32.

Benson, E. D., Hansen, J. L., Schwartz, A. L., Jr. & Smersh, G. T. (1998) Pricing residential amenities: the value of a view, *Journal of Real Estate Finance and Economics*, 16, pp. 55–73.

Borst, R. A. (1995) Artificial neural networks in mass appraisal, *Journal of Property Tax Assessment and Administration*, 1, pp. 5–15.

Bourassa, S. C., Hamelink, F., Hoesli, M. & MacGregor, B. D. (1997) *Defining residential submarkets: evidence from Sydney and Melbourne*. Real Estate Research Unit Working Paper Series No. 3 (Auckland, University of Auckland, Department of Property).

Bourassa, S. C., Hamelink, F., Hoesli, M. & MacGregor, B. D. (1999) Defining housing submarkets, *Journal of Housing Economics*, 8, pp. 160–183.

Bökemann, D. & Feilmayr, W. (1997) *Differences among social groups in their willingness to pay for distinct real estate properties*. Paper at the American Real Estate and Urban Economics Association 6th International Conference, Berkeley, 31 May–2 June.

Carlson, E. (1991) Self-organizing feature maps for appraisal of land value of shore parcels. *ICANN-9 Proceedings*, pp. 1309–1312 (North-Holland, Elsevier Science Publishers BV).

Carlson, E. (1992) *Neural thematic maps and neural GIS*. Paper presented at the 4th Scandinavian Research Conference on GIS, Espoo, 22–25 November.

Costello, G. (2001) *A spatial approach to price segmentation in housing markets*. Paper presented at the 8th European Real Estate Society Conference, Alicante, Spain, 27–29 June.

Dubin, R. A. (1992) Spatial autocorrelation and neighborhood quality, *Regional Science and Urban Economics*, 22, pp. 433–452.

Dubin, R., Pace, K. R. & Thibodeau, T. G. (1999) Spatial autoregression techniques for real estate data, *Journal of Real Estate Literature*, 7, pp. 79–95.

Grigsby, W., Baratz, M., Galster, G. & Maclennan, D. (1987) The dynamics of neighbourhood change and decline, *Progress in Planning*, 28, pp. 1–76.

Ilmonen, M. (1997) Viisi näkökulmaa alueiden erilaistumiseen (in Finnish), in: M. Ilmonen, M. Lankinen, H. Lehtonen, J. Maury & J. Päivänen *Mitä osoite osoittaa? Asuinalueiden erilaistuminen*

Helsingin seudulla. (How do you address yourself? Spatial differentiation in the Helsinki Region) Pääkaupunkiseudun julkaisusarja B, 1997, 2, pp. 13–27.

James, H., Collins, A. & Lam, E. (1994) *The principles and practice of artificial neural networks in property valuation studies*. Discussion paper (Portsmouth, University of Portsmouth, Department of Land and Construction Management).

Jenkins, D. H., Lewis, O. M., Almond, N., Gronow, S. A. & Ware, J. A. (1999) Towards an intelligent residential appraisal model, *Journal of Property Research*, 16, pp. 67–90.

Jones, C., Leishman, C. & Watkins, C. (1999) *Exploring linkages between submarkets: theory and evidence*. Paper at Cutting Edge Conference, Cambridge, 5–7 September.

Kaski, S. (1997) Data exploration using self-organizing maps. *Acta Polytechnica Scandinavica, Mathematics,* Computing and Management in Engineering Series, Nr. 82 (Espoo).

Kauko, T. (1997) Asuinalueiden attraktiivisuuserojen havainnollistaminen neuroverkon avulla (Visualizing differences in attractiveness between localities with a neural network, in Finnish), *Maanmittaus*, 1–2/72, pp. 37–70.

Kauko, T. (1999) The spatial dimension of residential valuation, in: K. Artimo H. Haggren (Eds) *Spatial Statistics in GIS Applications, Helsinki University of Technology, Carthography and Geoinformatics,* pp. 85–110 (Espoo).

Kauko, T. & Peltomaa, H. (1998) A comparison between neural network and multiple regression approaches: effects of transmission line on residential property values, *Surveying Science in Finland*, 16, pp. 52–87.

Kohonen, T. (1995) *Self-Organizing Maps*. Springer Series in Information Sciences (Springer-Verlag).

Kohonen, T., Hynninen, J., Kangas, J. & Laaksonen, J. (1996a) *SOM_PAK: The Self-Organizing Map Program Package*. Report A31 (Helsinki University of Technology, Faculty of Information Technology, Laboratory of Computer and Information Science).

Kohonen, T., Hynninen, J., Kangas, J., Laaksonen J. & Torkkola K. (1996b) *LVQ_PAK: The Learning Vector Quantisation Program Package*. Report A30 (Helsinki University of Technology, Faculty of Information Technology, Laboratory of Computer and Information Science).

Koikkalainen, P. (1994) Progress with the tree-structured self-organizing map, *Proceedings 11th European Conference on Artificial Intelligence, 8–12 August*, pp. 209–215 (John Wiley and Sons).

Kortteinen, M. & Vaattovaara, M. (1999) Huono-osaisuus pääkaupunkiseudulla 1980 ja 1990 luvuilla -käännekohta kaupunkiseudun kehityksessä? (Distribution of the underprivileged in the metropolitan areas of Helsinki in the 1980s and 1990s – turning point in the development? (in Finnish)) *Terra*, 111, pp. 133–145.

Kyllönen, L. & Räty, T. (2000) Asuntojen hinta-laatusuhde Joensuussa, semiparametrinen estimointi (in Finnish). *VATT-tutkimuksia*, 67.

Laakso, S. (1997) *Urban Housing Prices and the Demand for Housing Characteristics*, A 27 (Helsinki, The Research Institute of the Finnish Economy (ETLA).

Laakso, S. & Loikkanen, H. (1995) Finnish homes – through passages or traps? *Real Estate Economics*, 23, pp. 475–495.

Lam, E. (1994) *An artificial neural network approach to residential valuation: the self-organising map*. Working Paper Nr. 1 in Adaptive Algorithms (Portsmouth, University of Portsmouth, Department of Land and Construction Management).

Lankinen, M. (1997) Asuinpaikan valinta: status, etäisyys vai totunnaisuus (in Finnish)? in: M. Ilmonen, M. Lankinen, H. Lehtonen, J. Maury & J. Päivänen *Mitä osoite osoittaa? Asuinalueiden erilaistuminen Helsingin seudulla (How do you address yourself? Spatial differentiation in the Helsinki Region)* Pääkaupunkiseudun julkaisusarja B, 1997, 2, pp. 28–51.

Leishman, C. (2001) House building and product differentiation: an hedonic price approach, *Journal of Housing and the Built Environment*, 16, pp. 131–152.

Maclennan, D. & Tu, Y. (1996) Economic perspectives on the structure of local housing systems, *Housing Studies*, 11, pp. 387–406.

Maury, J. (1997) Ilmakuva pääkaupunkiseudun asuinalueiden sosiaalisista eroista 1990-luvun alussa. (An aerial view of the social differences in the Helsinki region, in Finnish). in: M. Ilmonen, M. Lankinen, H. Lehtonen, J. Maury & J. Päivänen *Mitä osoite osoittaa? Asuinalueiden erilaistuminen Helsingin seudulla (How do you address yourself? Spatial differentiation in the Helsinki Region)* Pääkaupunkiseudun julkaisusarja B, 199, 2, pp. 55–87.

McCluskey, W. J. & Anand, S. (1999) The application of intelligent hybrid techniques for the mass appraisal of residential properties, *Journal of Property Investment and Finance*, 17, pp. 218–238.

Morrison, P. S. & McMurray, S. (1999) The inner-city apartments versus the suburb: housing sub-markets in a New Zealand City, *Urban Studies*, 36, pp. 377–397.

Pace, R. K. (1995) Parametric, semiparametric, and nonparametric estimation of characteristic values within mass assessment and hedonic pricing models, *Journal of Real Estate Finance and Economics*, 11, pp. 195–217.

Rothenberg, J., Galster, G., Butler, R. & Pitkin, J. (1991) *The Maze of Urban Housing Markets: Theory, Evidence and Policy* (Chicago, University of Chicago Press).

Sharkawy, M. A. & Chotipanich, S. (1998) Housing segmentation in developing countries in transition: a recent case study of residential development in Bangkok, *Journal of Real Estate Literature*, 6, pp. 29–42.

Tulkki, A. (1996) Itseorganisoivan kartan soveltuvuus asuntojen hintojen arviointiin (in Finnish) B73, Publications of Real Estate and Law (Espoo, Helsinki University of Technology).

Vaattovaara, M. (1998) *Pääkaupunkiseudun sosiaalinen erilaistuminen – ympäristö ja alueellisuus. (Residential differentiation within the metropolitan area of Helsinki, Finland – environment and spatiality*.) Doctoral thesis (in Finnish). City of Helsinki Urban Facts Research Series, 1998:7.

Verkooijen, W. J. H. (1996) *Neural networks in economic modelling*. Doctoral dissertation (Tilburg University, Center for Economic Research).

Whitehead, C. M. E. (1999) Urban housing markets: theory and policy, in: E. S. Mills & P. Cheshire (Eds) *Handbook of Regional and Urban Economics*, pp. 1559–1594 (North-Holland, Elsevier Science BV).

Wolverton, M. L. (1997) Empirical study of the relationship between residential lot price, size and view, *Journal of Property Valuation and Investment*, 15, pp. 48–57.

Worzala, E., Lenk, M. & Silva, A. (1995) An exploration of neural networks and its application to real estate valuation, *The Journal of Real Estate Research*, 10, pp. 185–201.

Internet:

www.stat.fi. Homepage of Statistics Finland (Tilastokeskus)