
Using the self-organising map to identify regularities across country-specific housing-market contexts

Tom Kauko

OTB Research Institute for Housing, Urban and Mobility Studies, Jaffalaan 9, PO Box 5030, NL-2600 GA Delft, The Netherlands; e-mail: kauko@otb.tudelft.nl

Received 9 August 2003; in revised form 30 July 2004

Abstract. The aim of exploring and monitoring housing-market fundamentals (prices, dwelling features, area density, residents, and so on) on a macrolocational level relates to both public and private sector policymaking. Housing market segmentation (that is, the emergence of housing submarkets), a concept with increasing relevance, is defined as the differentiation of housing in terms of the income and preferences of the residents and in terms of administrative circumstances. In order to capture such segmentation empirically, the author applies a fairly new and emerging technique known as the 'self-organising' map (SOM), or 'Kohonen map'. The SOM is a type of (artificial) neural network—a nonlinear and flexible (that is, nonparametric or semiparametric) regression and 'machine learning' technique. By utilising the ability of the SOM to visualise patterns, one can analyse various dimensions within the variation of the dataset. Segmentation may then be detected depending on the resulting patterns across the map layers, each of which represents the data variation for one input variable. Utilising an inductive modelling strategy, the author runs cross-sectional and nationwide data on the owner-occupied housing markets of Finland (documentation presented elsewhere), the Netherlands, and Hungary with the SOM technique. On the basis of the resulting configurations certain regularities (similarities and differences) across the three national contexts are identified. In all three cases the segments are determined by physical and institutional differences between the housing bundles and localities. The exercise demonstrates how the inductive SOM-based approach is well-suited for illustrating the contextual factors that determine housing market structure.

1 Introduction

The aim to explore and monitor housing-market fundamentals (prices, dwelling features, area density, residents, etc) at a macrolocational level is motivated by various policy-related and business-related goals (for example, see Tu, 2000). Housing-market segmentation (that is, the emergence of housing submarkets), a concept with increasing relevance, relates to the differentiation of housing through the income and preferences of residents and in terms of administrative circumstances. The causes of segmentation processes pertain to a multitude of impacts related to geographical location (including topographical constraints), structural (that is, house-specific and site-specific) attributes, prices, consumer preferences and incomes, government regulation, and asymmetric information and power positions between market participants (for example, see Watkins, 2001). The monitoring of segmentation has various practical implications when a comparative evaluation of attractiveness is required.

However, the research problem of determining segments and tracking the processes behind segmentation has also received attention as a purely academic endeavour within housing economics. According to MacLennan and Tu (1996), submarkets are the product of market imperfections in relation to local-level supply-and-demand structures. Thus, differentiation of preferences and product groups, trade friction and search processes lead to a complexity and variety in the housing commodity; furthermore, any equilibrium tends to be unstable, in the short term at least. Although depending on the exact definitions used, in any one housing-market context segments are likely to occur over time and space, which implies that the verification of segmentation is predominantly an empirical issue. Although a *national* housing market is

hardly a valid concept, it may well be that a certain house type is more common in one city than in another; for instance, in a predominantly middle-class city, environmental variables tend to be especially important (Ball, 1973). By comparing the residential patterns in different time periods and places, it may be possible to formulate new hypotheses about the relative importance of various determinants of housing-market dynamics (for example, Tu, 2000). Exactly how the specific locality context matters is determined by path dependence, that is to say, by institutional and behavioural factors.

In this contribution I explore subnational housing-market structures by means of Kohonen's early 1980s 'self-organising' map (SOM)—a type of neural network modelling technique—by utilising patterns in the data. There is reason to believe that this type of exploratory and inductive approach will become more popular in the future, when its real potential becomes more widely known in the community of social scientists (for example, see Openshaw, 1998). The methods used are valid alternatives to the more orthodox method of modelling [for a thorough review of national, regional, and urban housing market modelling studies based on the methodology of mainstream economics, including their policy implications, see Meen (2001)]. Instead of formal submarket tests (see Watkins, 2001), I apply a more qualitative approach, based on a visual examination of the SOM output matrix in relation to the most pertinent housing-market features. In relation to the practical as well as the scientific goals discussed above, the added value of the neural network and other flexible regression methods of house-price modelling is argued to lie in three aspects that are overlooked in the mainstream approaches: the diversification of demand and supply; manifestations of fuzzy and nonlinear processes; and location as a residual, idiosyncratic element.

From these arguments concerning objectives and methodology, a number of research questions are developed. First, is the SOM a feasible tool for analysing housing market structure on a nationwide level? Second, is it possible to distinguish discriminating features that are related either to supply-side factors such as level of urbanisation and density, or to sociodemographic demand-side factors? Third, if such criteria are to be identified, what are their associations with property values, given that the house-price level is frequently treated as a generic indicator of attractiveness? Moreover, in this study I aspire to undertake a comparative exploration of hypotheses regarding subnational market structures derived from largely distinct market contexts. Thus, fourth, in a cross-country comparison of such contexts, what are the similarities and the differences in terms of certain key dimensions related to location, price, supply, and demand? Fifth, although direct behavioural factors are unlikely to be captured at this general level, does the method allow one to distinguish any institutional aspects?

Thus, the idea of the methodology is a SOM-based hypothesis-generation method evaluated through comparative analysis. The hypotheses generated concern the spatial and sectoral housing-market structure in relation to physical features, sociodemographic features, and attractiveness. The comparative element, in turn, may be useful in adapting a broader view of the relative significance of a number of key dimensions—that is, visually observable features, and the associations between them, found in the nation-specific analyses.

The remainder of the paper is structured as follows. In section 2 I briefly review the literature and present an innovative method based on the SOM concept. In sections 3 and 4 I present the results of the neural-network-based housing-market analysis for three national contexts using nationwide datasets from Finland [documentation presented elsewhere (see Kauko, 2002)], the Netherlands, and Hungary. Finally, in section 5 I summarise the findings of the study, and—although this is not the primary goal of this study—I also outline the policy relevance of the method.

2 The Kohonen map as an alternative technique for submarket detection

As already indicated, the premise of this research is that the identification of housing submarkets at various spatial scales is of importance, and that such an exercise is worth conducting with the specific artificial neural network (ANN) technique in question, the Kohonen map. In the following the method is discussed in more detail. However, in order to save space, the analytical presentation will not concern the SOM technique itself but rather how the SOM technique is applied as a tool for the identification of intraurban and interurban housing-market segments, as presented in sections 3–5. First, the topic of the study is presented (in section 2.1), after which the validity of the proposed method is highlighted (section 2.2), before, finally, the actual research design for the empirical study is set out (section 2.3).

2.1 Geographically extended housing-market analysis

According to Maclennan and Tu (1996), to recognise that the real dimensions of housing systems may distort market operation is important for the study of housing-market structure—this relating to a local context, where submarkets have different degrees of internal stability in relation to time as well as different levels of substitutability in relation to various sectors and space. Furthermore, Maclennan and Tu argue that the submarket perspective is relevant both theoretically and empirically.

Elsewhere, I have demonstrated the use of the SOM-based method, with one city as an exemplar (see Kauko, 2002; Kauko et al, 2002). Here, the aim is to apply nationwide datasets. Although not confined to the urban or local context, the idea is nonetheless to retain the microapproach in order to evaluate to what extent space serves as a friction and a constraint, as postulated by Maclennan and Tu (1996). Furthermore, at least two contexts are needed to incorporate the institutional aspect. The choice of the three countries was driven by data concerns and not by theory—with the possibility that theory (or at least hypotheses) could be elaborated based on the empirical findings. The three countries chosen are expected to provide interesting insights into the structure of the spatial housing market at a subnational (macrolocal) level. When comparing different contexts, the determinants of house-price levels may either be completely different factors or be the same factors, only with differences in magnitude.

At the nationwide level, the situation in terms of differences in price levels and market segments is ostensibly much sharper than in the intraurban case. Although most of the literature on submarkets concerns one urban area only, models of regional, and even nationwide, housing markets offer an interesting option for exploration. Not unlike the case for urban–local level analysis, the aim is to look for differences or similarities with respect to space or other, physical, sociodemographic, or, possibly, institutional features. A few sophisticated attempts to identify subnational segmentation may be noted in various research traditions, such as intermetropolitan hedonic regression modelling within urban economics (Izraeli, 1987; Potepan, 1996), econometric modelling techniques (Tu, 2000), spatial interpolation and statistics for regional-level analysis of housing-market structure and dynamics (Meen, 2001), and factor-ecology-based area analysis (Siikanen, 1992; Wong, 2001). When used for empirical housing-market analysis, the SOM approach also helps us gain insights into the dataset, as will be shown below.

2.2 Developing a method based on the Kohonen map

Neural networks are a category of machine-learning methods that have recently become widely used in various problems involving prediction, classification, and pattern recognition. The basic principle of the neural network is the ability to ‘learn’ from a complex set of input data through a training process, where a signal coming from outside the system is elaborated and transmitted by a set of ‘neurons’ or ‘nodes’ in such a manner that a stimulus–response connection between input and output is formed. Learning in

neural networks is suboptimal: the nature of the connection is partly random and partly dependent on the externally managed parameters, most of which are determined in a rather ad hoc way. The SOM (or Kohonen map) technique belongs to the competitive category of neural network techniques. The SOM technique is best defined as a mapping from a high-dimensional data space onto a (usually) two-dimensional lattice of points (Kohonen et al, 1996). In other words, disordered information is profiled and analysed to give visual patterns, thus forming a landscape of the phenomenon described by the dataset (see Kohonen, 1995).

The training of the SOM proceeds as follows: first, the map has to be initialised by generating random values for each node, where the structure of the node is a vector. Then, the training procedure of the algorithm begins. In this procedure the correspondence between the vectors for each observation and for each node are calculated iteratively as follows: first, a training vector x is selected (randomly); then, the best matching neuron c , that is closest to x , is found; and, finally, the values of the node c and its neighbours are adjusted towards the observation x (for example, Koikkalainen, 1994). Usually, the *matching* is determined by the smallest Euclidean distance between c and x . The technique is based on the principle of *unsupervised competitive learning*. The ‘winner’ is the node with the shortest distance to the observation vector, and its weights are adapted towards the observation. This goes on until all desired observations are used for training—usually more than once. Adjacent nodes on the map are similarly adapted towards the observation, but the extent of this depends on the selected parameters, the two most important of which are the initial learning rate (that is, the speed of training), which decreases linearly with time to zero during training, and the initial radius of the training area (that is, the area of influence), which decreases linearly with time to one during training. The decisions concerning the design of the map (that is, the choice of network parameters) are not of importance for understanding the qualitative way the technique is used in the explorations that follow [but for a discussion within this problem field, see Kauko (2002)]. In order to run the algorithms of the SOM technique there are plenty of examples of software from which to choose; here, the SOM_PAK (developed by the Helsinki University of Technology) is used (Kohonen et al, 1996).

It is convenient to make an analogy to statistical cluster analysis (for example, Kaski, 1997). The crucial difference from the k -means classifier is the ‘neighbourhood’ concept—that is, the concept of a ‘winner’ node with adjacent neurons (at the beginning of the learning process, the exact extent of the hexagonal or rectangular neighbourhood depends on the chosen network parameters; at the end of the learning process, the neighbourhood comprises only the closest neighbouring nodes). This concept is incorporated into the SOM algorithm but is missing from the k -means algorithm (for example, see Openshaw et al, 1994). Any comparisons between SOM and cluster analysis are, however, rare. In one such comparison—perhaps the first made—Kaski (1997) advocated the SOM method as a viable alternative to more established multivariate data-analysis methods. What is important here is to grasp the fundamental idea of the proposed method through four stages:

- (1) to generate the output of the SOM by means of numerical, identifiable, input from the phenomenon under study;
- (2) to interpret the output and look for associations across and within the variation of the variables;
- (3) to repeat the analysis with a different dataset, possibly from a very different context; and
- (4) to discover whether any potentially interesting similarities or differences exist between the datasets.

2.3 Some advantages and drawbacks when applying the SOM for spatial and cross-national housing market analysis

Within the problem area of preprocessing the data for property valuation, the SOM approach has been used specifically for clustering and detecting outlier objects, with promising results (for example, see James et al, 1994; Jenkins et al, 1999). The SOM approach seems to offer plenty of possibilities, for example, it could be understood as a spatial method of house price analysis. As a result of the analysis, one obtains a surface where areas with similar combinations of variables can be looked at as a whole and compared with areas of a different nature. In this way, potential market segments can be illustrated. A similar and perhaps even safer point of reference for this study can be found in the geodemographic classification work by Openshaw et al (1994), which uses the SOM approach with British census data (compare Kaski and Kohonen, 1995). Most recently, Hatzichristos (2004) carried out a demographic classification of Athens, Greece, using the SOM algorithm in combination with fuzzy logic, much in the way of the pioneering work by Openshaw and others in the 1990s.

When using the SOM approach within this problem field, the starting point is the generation of visually observable groups of neurons ('clusters') on the map surface (that is, the output matrix of neurons defined by the user). These are reciprocally similar observations that are different from other observations with respect to one or more dimensions, measured in Euclidean space. From this it follows that the larger these differences and the greater the number of dimensions along which such differences exist the further away from each other the observations or clusters are on the map surface. The other useful property of the SOM technique is the identification of output according to its membership in a predefined category of input ('labelling'). The neuron is labelled based on the majority of the labels it has 'won'. In this way one may obtain a classification of similar types of observations that are significantly different from other types of observations.

It is true that analyses based on the SOM technique are simple. However, multiple equilibria and city-suburb dynamics are complex targets for analysis and therefore require simplification in terms of a reduction of dimensions, the isolation of plausible mechanisms, and the detection of certain key relationships. What exactly is new here is the detection of an institutional element: informal and formal factors that affect the organisation of the data, which is not common in economic modelling research. Although the method does not fit into any specific research tradition, it may be understood as an interface between three lines of research: property value modelling and area classification; housing economics; and comparative housing research. The aim is to make new theory based on the actual outcome (that is, through an inductive approach), which consists of feature maps and local knowledge gained by experience and existing surveys.

The purpose of the method is the classification of datasets into groups—not prediction with new data (for that, see Kauko, 2002). The basic principle of this inductive research strategy is to look for any similarities between the resulting configurations. As Daly et al (2003) emphasise a method of 'inductive enquiry' and the comparative dimension of research, they propose the utilisation of "all available and pertinent data concerning the preconditions of a specific outcome" and play down the importance of statistical criteria. The analyses in this paper, however, are very general and explorative. The feature maps generated by the SOM method show the underlying data structures in a fuzzy sort of way. Each input variable is represented in the feature map as its own specific map layer, from which patterns and clusters can be detected. Sometimes, the clusters may be determined based on 'new features'—indirect combinations of input variables, as opposed to any input variable in itself. Indeed, the aim of a housing-market

study should be to reduce the number of variables, instead of increasing it, as might occur in a SOM-based exploration. However, factors other than the input variables themselves might also be relevant, because of interactions between variables. This implies an added value of the SOM approach: one leaves out certain variables from the analysis yet still manages to detect them as extra elements. These types of factors form context boundaries based either on physical (for example, a seashore) or on institutional (for example, a local land-use planning regime) constraints (Kauko, 2002).

Within academia, a great deal of criticism can, of course, be levied at the mechanical element of the SOM technique—this being the opinion of some ‘traditional’ social scientists. However, as the visual inspection of the resulting surface requires, in fact, strong support from local expertise to obtain a possible explanation for the phenomenon under study, such comments can be considered too myopic. The SOM approach should, in this context, be seen as one more tool to help us increase our knowledge of specific housing markets with regard to prices, sociodemographic variables, physical features, and institutions. In this way the resulting SOM models are used inductively, as in the spirit of Leontief (see Leontief, 1976; see also Amsden, 1995). In this application of the model to housing-market analysis, I first explore the housing markets in two relatively similar countries, Finland and the Netherlands; then, based on the regularities discovered, I abstract the discussion to a general level; finally, I apply the models to Hungary so as to obtain preliminary results with regard to similarities and differences between two or three specific contexts, represented by datasets, that depict similar targets.

The nature of the analysis is cumulative. The idea is that in the course of the analysis certain general patterns may emerge, and these may be used subsequently if not as hypotheses then at least as expectations when a new country case is being examined. In this way, the SOM is used as ‘idea generator’ for exploring spatial and topological relations within cross-sections of nationwide housing-market data. To reiterate, spatial and sectoral segmentation may be detected depending on the resulting patterns across the map layers, each of which represents the data variation for one particular input variable. After an inductive modelling strategy, cross-sectional and nationwide data of the owner-occupied housing markets of three different country contexts are run with the SOM. On the basis of the resulting configurations certain regularities (similarities and differences) across the various contexts may be identified in relation to physical and institutional aspects that form visual clusters. Such an exercise demonstrates how the inductive SOM-based approach is well suited to illustrating the contextual factors that determine housing-market structure.

In order to save space, sections 3 and 4 cover the Dutch and the Hungarian analyses only, as the Finnish analysis has already been presented in earlier work (see Kauko, 2002; Kauko et al, 2002). Each section begins with an outline of the spatial housing market structure of urban areas in the specific country context in order to provide some background information, followed by an examination of the SOM output. The evaluation concerns only the validity of the approach (for reliability and robustness issues, see Kauko, 2002): more specifically, it concerns the ability to capture housing-market circumstances about which we have some prior knowledge. The main results of the Finnish analysis are summarised before the Dutch results are evaluated, and in the case of Hungary further evaluation takes place in relation to the findings from Finland and the Netherlands.

As the method operates on secondary datasets, issues related to data collection are also noted. It was not possible to look at exactly the same variables for all three cases. Individual house-price data [which are always preferable to data based on areal aggregations (sums, percentages, and means) could be obtained from Finland and

Hungary, but from the Netherlands only an aggregated (but comprehensive and detailed) data source was available for this project. To compensate for this lack of variation, the number of variables was greater in the Netherlands dataset (20) than in the two other cases (16).

3 Analysis of the Netherlands with the SOM method

3.1 The study area and the data

It may be observed that a change towards a polycentric urban structure, sometimes with a cross-commuting feature, is taking place in the Netherlands. This is particularly notable in the network of cities in the western part of the country known as the Randstad, the most densely populated region in Europe (see figure 1). Another thing to note is that the physical size of the country is small: the land surface is only 10% or 15% of the size of Finland (depending on whether or not the polders are counted). These two circumstances—polycentricity and land scarcity—have strong implications for residential mobility, commuting, and the housing market as a whole.

The Netherlands is also characterised as ‘the most planned country in Europe’ (Dutt and Costa, 1985; cited in Bontje, 2001, page 135). However, since the early 1990s a variety of changes have taken place towards a reduction in government regulation. In the ‘compact city’ approach, decisions over where to build were increasingly left to private investors, and planners had limited opportunities to interfere. At present, there is also a growing mismatch between housing supply and demand in the Netherlands, especially at the level of specific residential environments (Bontje, 2001, pages 38–57).

When Bootsma (1998) ranked the level of urbanisation of Dutch cities, based on factor scores in a principal components analysis (PCA), twenty-five municipalities were classified as cities: Amsterdam, The Hague (Den Haag), Haarlem, Leiden, Utrecht,



Figure 1. The Netherlands.

Maastricht, Delft, Rotterdam, Alkmaar, Schiedam, Groningen, Den Bosch, Leeuwarden, Breda, Deventer, Nijmegen, Amersfoort, Arnhem, Eindhoven, Tilburg, Venlo, Zwolle, Hilversum, Dordrecht, and Enschede (see figure 1). However, it is not possible to identify a 'prime city'. Amsterdam used to be a prime city until the mid-19th century, when the industrialisation of the country started, and Rotterdam, The Hague, Utrecht, and other cities took over some of this role from Amsterdam. The Dutch municipalities may be categorised into the following types: rural, urbanised rural, suburban, small cities (up to 100 000 inhabitants), medium-sized cities (100 000–200 000 inhabitants), large cities (more than 200 000 inhabitants), and designated locations for urban development, called 'growth centres' (Bontje, 2001, pages 101–105).

During the 1990s house prices rose substantially in the Netherlands and, especially for the expensive and high-quality market segment, the demand was great compared with supply (Boelhouwer, 2001). However, the proportion of one-person and two-earner households is increasing as a response to sociocultural changes that started in the 1960s. Notably, the small nonfamily household has become a dominant group in Dutch society. One-person households already are in the majority in the inner cities of the largest cities. Other effects are the ageing ('greying') of the population and the influx of guest workers and immigrants from former Dutch colonies. These changes have led to a bipolar housing market. On the one hand, suburban and inner-city locations are popular, especially among immigrants and young adults but, on the other hand, between these locations less popular neighbourhoods emerge (Bontje, 2001, pages 108–112). Dieleman and Wallet (2003) conclude that the income disparities between affluent suburbs and poorer central cities are particularly pronounced in the largest cities and the Randstad; furthermore, they blame Dutch housing policy for this.

Table 1. Description of variables used in the Dutch analysis, 1999 data (source: VROM, 1999).

Indicators for each district (<i>wijk</i>)	Aggregated or mean values
Addresses per neighbourhood (density proxy)	1–11 856
Extent of urbanisation	1 (highly urban)–5 (least urban)
Population density (inhabitants per km ²)	0–31 001
Percentage of children 10–14 years old	1–65
Percentage of people 15–24 years old	2–96
Percentage of people 25–44 years old	3–76
Percentage of people 45–64 years old	1–71
Percentage of people 65 or over years old	1–98
Percentage of non-Westerners (1st and 2nd generation immigrants)	0–89
Percentage of one-person households	3–99
Number of families	0–141 280
Percentage of families with children	7–93
Average family size	2.1–4.6
Percentage of people on a low income ^a	12–95
Percentage of people on a high income ^a	4–74
Percentage of people 15–65 years old with unemployment benefit as the primary source of income	0–95
Assessed market value of dwelling (total price, in Dutch guilders)	43 000–1 170 000
Percentage of industrial enterprises (including under construction)	0–63
Percentage of commercial enterprises	24–99
Percentage of noncommercial enterprises	0–61

^a As a percentage of all those with an annual income, including the self-employed, students, and certain groups receiving subsidies (such as for rent or childcare).

The dataset is presented in table 1. There are 538 municipalities in the Netherlands, in thirteen provinces. The data are aggregated at the district (*wijk*) level, giving 2382 observations for 1999. The data used are based on an aggregated database of district and subdistrict information “Kerncijfers Wijken en Buurten” (KWB)] maintained by the Dutch Ministry of Housing, Planning and Environment [Het ministerie van Volkshuisvesting, Ruimtelijke Ordening en Milieubeheer (VROM)]. Each entry is either a sum, a percentage, or an average figure of the variable in question with regard to the district concerned. It may be expected that the price formation is differentiated according to such sociodemographic and economic variables. One of the variables is the mean taxation value assessed for municipal property taxation. The averaging procedure inevitably reduces some of the variation, which reduces one’s ability to make conclusions based on the SOM output, the feature map.

3.2 The feature maps

In section 2 I proposed a way of analysing the SOM results when the goals are the identification of different housing submarket structures. Under this method, three types of visual analyses are incorporated: clusters across one map layer, with locational identification based on district (in the Dutch case); associations across the variables (that is, across the map layers); and any relevant discriminating determinants of submarket formation (that is, context boundaries on one or more map layers).

The selection of map dimension is a matter that relates to the desired trade-off between the level of resolution (a larger map has a better resolution) and the level of generalisation (a smaller map is more parsimonious). Here, an experimental approach to the problem was applied, which resulted in the selection of map dimensions of 16×12 neurons [horizontal by vertical (that is, x by y)]. This setup means that each neuron with nearest neighbours is associated with 87 observations. The population density is selected as a particularly interesting variable; this layer is selected and presented in figure 2 (over) for the Dutch case. The lighter the shade, the higher the indicator value of the node, and, conversely, the darker the shade, the lower the value. The label (that is, the text inside the neuron) serves as identification, in this case based on the district location. The population density is strongly associated with indicators of address density, level of urbanisation, share of non-Westerners, and share of one-person households. (For reasons of space, only one map layer is shown; the other layers are obtainable from me on request). This dimension partitions the map into an ‘urban’ left and a ‘rural’ right.

Although not shown here, a clustering also occurs on the basis of the variable ‘average family size’ with areas with a low percentage of families with children and areas with a larger average family size. Areas with a larger average family size also have higher shares of children and people 15–24 years old and a lower density compared with areas with an average family size. Other features that contribute strongly to the structuring of the map are income, unemployment rate, the share of people 15–24 years old, and the share of industrial enterprises.

I have already noted that rural areas are situated on the right-hand side, whereas more urban areas and cities are situated on the left-hand side of the map. The upper left-hand corner may be characterised as ‘truly urban’, whereas the lower left-hand corner is ‘urban periphery’ (either a suburban neighbourhood of a city or a relatively rural area located close to the Randstad). In this type of area the family size is substantially larger than for areas in the upper left-hand corner. The urban areas in the upper left-hand corner also have a high share of people 15–24 years old, non-Westerners, and one-person households (the labels represent districts in municipalities of a very different nature). In contrast, the lower and middle portion of the map is represented by various smaller municipalities in eastern and central parts of the country.

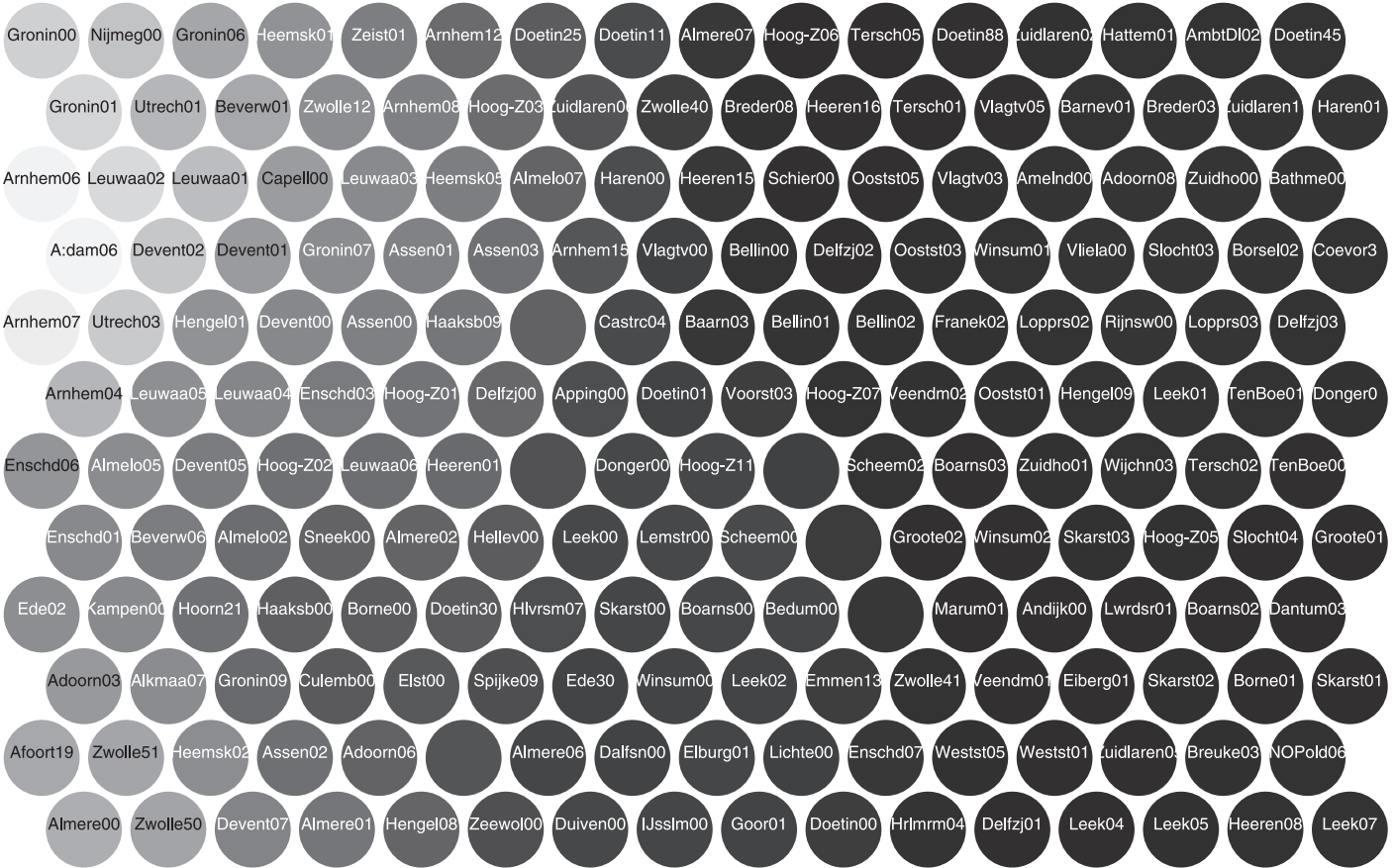


Figure 2. Feature map illustrating population density in Dutch districts. Note: light shading, high density; dark shading, low density. The labels are an abbreviation of the municipality followed by a two-digit code specifying the actual district within that municipality.

The predominantly rural areas are of two types: those positioned in the lower right-hand corner are characterised by large families and a high share of industrial enterprises; those positioned in the upper right-hand corner are characterised by small families, a higher property value, and a high share of noncommercial enterprises. The middle and upper left-hand side of the map is characterised by areas with a high share of low-income and unemployed persons. Most of these areas represent inner-city neighbourhoods in various municipalities. Here, the percentages of people 45–64 years old and 65 years and older, respectively, are high too.

When taking a closer look at certain urban locations, De Baarsjes in Amsterdam and Overvecht in Utrecht stand out from the dataset for variables such as address and population density and share of immigrants and one-person households, but not in terms of price levels. The most expensive areas (as measured as the assessed total value of the houses averaged over the district) represent areas far outside Randstad (with the exceptions of Zeist and Almere). In this respect, the situation is completely the opposite to that of Finland, where the Helsinki city and region clearly stands out as a categorically more expensive location than the rest of the country. Three explanations related to context arise: (1) outside the Randstad the houses are larger than inside the Randstad and, therefore, on average are also more expensive; (2) the ongoing suburbanisation preferences and an increased mobility have generated an out-migration away from urban areas in general and from the Randstad in particular; and (3) the fact that investment demand over the past decade or so has been concentrated in suburban and rural areas is captured in the value variable. (A fourth explanation is that these municipalities have overestimated the value of their property stock. This is a plausible explanation, given the inexact nature of tax assessments; not all municipalities have the resources to undertake accurate assessment on a yearly basis but, rather, adjust the prices by using an index, which may lead to very crude approximations.)

Here the idea is only to show how the data may be structured spatially, based on a few key housing-market features. It can be noted that a number of cities are represented by more than one neuron, and even relatively small municipalities (rural areas or small cities) may have more than one housing-market segment (for example, both Heerenveen and Doetinchem are represented by segments far away from each other on the map; one ‘with children’ and another ‘without children’). In this case, there is considerable variation between spatial segments within municipalities. No municipality is strongly overrepresented (although Apeldoorn has two segments; Leeuwarden has two segments; and Delfzijl has three segments); nor are there many empty nodes on the map.

Only one node out of 192 relates to an Amsterdam district, and that node captures only two of the fifteen observations from Amsterdam (this is what could be expected purely on statistical grounds: the share of observations was 15 out of 2382). The remaining thirteen observations are spread over seven other nodes that relate to similar districts in four large or medium-sized cities (Groningen, Arnhem, Utrecht, and Enschede) and a municipality on the Frisian coast (Dongeradeel). Furthermore, the only Amsterdam-related node has ‘won’ eight observations in total, six of which are from other municipalities and represent neighbourhoods that are similar to Amsterdam. With Utrecht we observe a similar spread: only two out of twelve, and one out of the three observations, that the two Utrecht nodes had won are actually Utrecht observations; the remaining six of the nine Utrecht observations have been won by other nodes. Rotterdam (twenty-five observations) and the Hague (forty observations) are not represented among the labelled nodes at all, as all the observations belonging to these two cities are dispersed all over the map and won by nodes with a variety of labels. This implies a great homogeneity in district types across municipalities.

Nonetheless, there is considerable variation across municipalities in the northern, northeastern, and eastern parts of the country. Moreover, certain areas are represented by two or even three distinct segments in terms of property value levels or other features. The provinces of Friesland, Groningen, Overijssel, Drenthe, Gelderland, and Noord-Holland are strongly overrepresented if one looks at the district labels of the 186 nodes that have won observations.

In order to investigate how certain extreme cases are being positioned on the map let us consider a large city and a small peripheral municipality. In the case of The Hague, the forty districts indeed exhibit broad variation, but each type of area resembles a similar area elsewhere. The Oostduinen district is an affluent coastal area outside the city; hence it is labelled after Hattem, a pseudorural municipality. Duindorp is a socially problematic harbour area comprising old derelict building stock; hence it is labelled after a rather disadvantaged part of Leeuwarden. Schildersbuurt, an inner-city urban renewal area, was ranked the worst neighbourhood in the country; hence it is labelled after an urban, disadvantaged area in Arnhem. Watertuin, an annexed area, is labelled after the urban periphery segment in Almere. In the case of a small municipality in the southeastern part of the country (for example, Venray) a similar logic holds: the urbanised part of the area is labelled after disadvantaged Leeuwarden, whereas the rest of the area is labelled after Boarsterhim—an agricultural municipality. Certain patterns do indeed make sense.

To sum up the overall findings with regard to the situation of the districts on the map, the structure is characterised as five blocks of local housing-market type with fuzzy boundaries: (1) ‘truly urban’; (2) ‘urban periphery’; (3) ‘pseudo-rural’ areas with a concentration of ‘small young families with kids’; (4) ‘agricultural rural’ areas with large families and also industry [for a distinction between categories (3) and (4), see Haartsen et al (2003)]; and (5) areas inbetween, also characterised by ‘concentrations of poorer, middle-aged, and elderly residents’. Finally, it can be noted that the labels are, in some cases, related to absolute location (even though no geographical input was provided), but in other cases they are related to the relative location within the region, so that areas in small suburban municipalities group together (although being far from each other geographically). That two municipalities belong to adjacent nodes may be because they are of similar character and also situated within the same urban system (relative location is then absolute location).

3.3 Evaluation of the findings

As explained above, the evaluation of the Dutch analysis is undertaken not only in relation to a knowledge of the Dutch housing-market context but also in relation to certain hypotheses based on earlier analyses of Finnish data. Below I provide a brief summary of those Finnish results (for full documentation see Kauko, 2002). In the Finnish analyses the clustering illustrated structures in demography, employment, physical environment, and other indicators of a certain locality. The most important factors described the price level, size of the municipality, the relative growth rate, and peripheral situation—in other words, indicators of urban character. The house-specific (physical) factors, in turn, described the internal submarkets of the municipality. In Helsinki, the capital city, the relationship between age of buildings, house type, and price per square metre is especially well observable, and the markets are polarised into A and B classes. In suburban single-family housing areas and old neighbourhoods close to the centre the price levels are high, whereas in multistorey-dominated suburbs price levels are relatively low.

Segmentation can be observed both *between* certain types of municipality on the basis of their composition of municipal variables as well as *within* them on the basis of

their composition of house-specific variables. It may be noted that the variation in household size is potentially a reflection of sociocultural factors, that is to say, of changing norms. Also interesting is to observe how cities that may be separated geographically by hundreds of kilometres are situated close to each other on the map. Two hypotheses concerning market segmentation may now be developed. First, in the feature map, the satellite towns (that is, suburban municipalities) of the cities are not, in some cases, situated near the centre of the region on the map but near each other. Second, in most of the towns there are separate markets for multistorey flats and single-family houses. Also, here, the organisation of the data reflects institutional context.

Based on similar expert knowledge of the Dutch housing-market context, the analysis may now be generalised as the qualitative findings in the two countries are scrutinised and put side-by-side. Compared with the situation in Finland, in the Netherlands the picture is, in fact, more heterogeneous *within* urban areas and regions. In a country such as the Netherlands, certain regional housing markets are very severely constrained. This leads to multiple equilibria in the intraurban housing market—not just in major cities but also in smaller cities and less urban municipalities. From the feature maps this can be observed when the housing market in many cases is differentiated into two to four completely separate submarkets in terms of price level, density, and variety of sociodemographic and local economic variables. Another observation is that, compared with the situation in Finland, in the Netherlands the housing market structure is more homogeneous *across* municipalities, and this applies also to smaller towns and regions with a less urban character. However, segmentation across municipalities is clearer in the northern and eastern parts.

Tendencies to choose suburban and even rural living ahead of urban living, together with recent trends in investment demand away from urban areas, are two factors that do not fit with the emerging model based on similarity. This is, among other things, related to the physical size of the market and the tolerance of residential mobility. In the Randstad region in particular, mobility is very high compared with the situation in Finland, and the formation of preferences may be split into various urban, suburban, and rural (arguably, also counterurban) elements. In particular, one may reside in the (pseudo-)countryside but work in the city. This is not (yet) the case in Finland, as second houses in the countryside are common and rural living in a ‘pseudo’ sense is not traditionally a common preference (on Sweden, see Bontje, 2001). Unfortunately, the method used does not allow for a comprehensive treatment of this aspect.

In both countries, sociodemographic and local economic features are important to the structure. The exact definition of variables was different, yet the results were much the same (for example, in certain regions and parts of cities low income and unemployment go together with low property values). However, whereas these variables in the Finnish case demonstrated the variation across municipalities, in the Dutch case they demonstrated better the variation across districts within the municipalities. Before I move on to the third country case, some interim conclusions may be drawn about what has been achieved so far:

1. The Finnish analyses showed two general patterns: segmentation across the municipalities based on relative location, and intramunicipal segmentation into two or more submarkets determined by density.
2. The Dutch analyses generated a similar pattern; the intermunicipal aspect was weaker as areas of similar size categories were, in some cases, clustered together, regardless of their role in the regional functional hierarchy; intramunicipal spatial

segmentation, in turn, was stronger in the Netherlands than in Finland, and was not confined only to cities but applied also to suburban and even rural municipalities.

3. This allows us to generalise: either both intermunicipal and intramunicipal segmentation, or only intramunicipal segmentation, may be expected from a third context (Hungary), and the result may resemble one of the two cases more than the other.

4. To some extent the empirical findings were possible to justify by recourse to expert knowledge (for example, in relation to the spatial difference between the disadvantaged inner cities and affluent suburbs in the Dutch Randstad).

4 Hungary

4.1 The study area and the data

The Hungarian housing system has recently undergone a rapid transformation in which the rental sector has been wiped away. This trend can be understood as part of a broader qualitative change (1989–90) in which the social order changed from a communist system to a postcommunist and potentially Western democratic system. The immediate consequence has been a shift away from the dominance of political capital towards the dominance of economic capital (see, for example, Treiman and Szelényi, 1990).

According to Kok (1997, page 75), the shift of ownership from state enterprises into private hands has had negative effects in terms of reduced mobility and the abolition of subsidies on new construction. These effects probably outweigh the benefits of the removal of administrative barriers to entry into certain parts of the housing stock. Furthermore, according to Kok (1997, page 20), the transformation has resulted in increased freedom of choice in the labour and housing markets, removal of administrative spatial constraints, and increased regional polarisation, emphasising the push factors in peripheral areas and the pull factors in economically advanced areas with regard to employment and income; it has also resulted in new financial barriers for some. According to Douglas (1997, page 16):

“the determination of success in the current transformation process can be markedly different amongst various socio-economic groups. A clear set of winners and losers is developing in all of [European postsocialist] countries, although for most households, declining living standards have been the norm.”

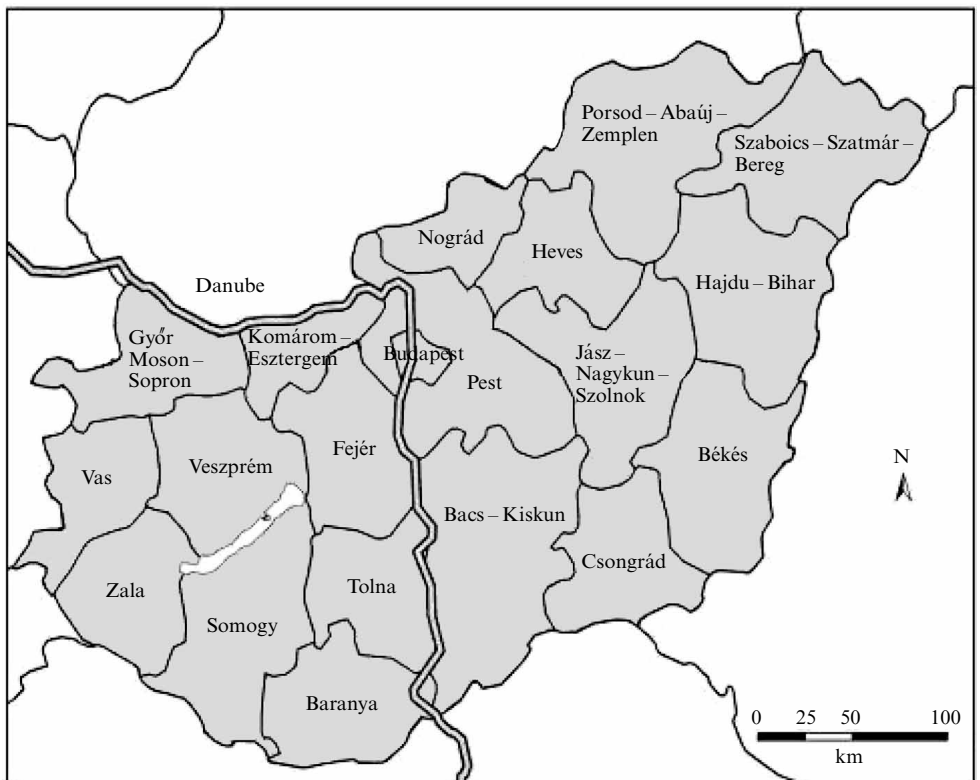
What, then, can be expected from the analysis related to the two identified types of segmentation? Casually observed, the polarisation within urban areas means pleasant new low-density houses and refurbished inner-city flats in contrast to other, more or less derelict, house types. Furthermore, a general lack of purchasing power is a feature that leads to relatively low out-migration and in-migration across housing market areas within the owner-occupied sector in Hungary (for example, Douglas, 1997; Kok, 1997). Therefore, segmentation can be expected to be clearly visible also across geographically defined areas, and not just within them. (It has to be added that the discussions in Douglas’s and Kok’s dissertations refer to the early 1990s. Since then the situation has changed rapidly; in particular, the new housing mortgages introduced in 2000 have had significant impacts on new construction.)

Nine of the variables are aggregated at the county level, there being twenty counties in Hungary,⁽¹⁾ belonging to seven regions (see table 2 and figure 3). The variables are described in table 3 (over). The data were from a register provided by Kolpron Budapest (2002), for May 2001 to January 2002. The variables include both market value and collateral value. Market value corresponds to residential valuations for mortgage lending, with each valuation based on three actual transactions that are

⁽¹⁾ Actually, there are nineteen counties plus Budapest, which does not constitute a county in the strictest sense (Kok, 1997).

Table 2. Geographical areas in Hungary.

Counties	Region	Number
Budapest, Pest	Central Hungary	32
Fejér, Komárom – Esztergom, Veszprém	Central Transdanubia	49
Győr – Moson – Sopron, Vas, Zala	Western Transdanubia	26
Baranya, Somogy, Tolna	Southern Transdanubai	14
Borsod – Abaúj – Zemplén, Heves, Nógrád	Northern Hungary	62
Hajdú – Bihar, Jász – Nagykun – Szolnok,	Northern Great Plain	50
Szabolcs – Szatmár – Bereg	Southern Great Plain	52
Bács – Kiskun, Békés, Csongrád		

**Figure 3.** Hungary: the twenty counties.

comparable for the subject property. Collateral value is the value of a property at a foreclosure of a mortgage loan. In high-risk cases the collateral value may be as small as 50%–60% of the open market value. This indicator incorporates some assessment of risk and is adjusted 5%–25% lower, depending on the location. Furthermore, both indicators were measured per square metre as with the Finnish data.

The county-level variables were collected from the year book of statistics (*KSH* 2000). The income variable was adjusted up by 10%–30% for areas with a high proportion of self-employment. (The reason for this ad hoc correction was to avoid a systematic error, as the total income of self-employed persons are recorded in an unreliable manner, and with a likely downwards bias.) The counties are small enough for employment and services within the county to be within commuting distance;

Table 3. Description of the variables used in the Hungarian analysis, May 2001 to January 2002 (sources: Kolpron Budapest, 2002; *KSH* 2000).

	Unit of measurement ^a
<i>Individual (microlevel) variables</i>	
Market value	(HUF per m ²)/1000
Collateral value	(HUF per m ²)/1000
Age of building	(years ×3)+50
Dwelling format—density	50: single-family (sf) or 400: multistorey/multifamily apartment buildings (mf)
Dwelling format—general prestige	100: Családi ház (sf); 200: Zöldövezeti társasház (garden city, mf); 300: Városi társasház (old urban, mf); 400: Lakótelep (prefab high-rise, mf)
Size	m ²
Inflation effect	(Time of sale ×1.5)+100; where 0 = 11 May 2001
<i>County-level variables (1999 situation)</i>	
Net migration	migration balance/100 inhabitants
Population	as a percentage of the whole country
Population density	number of people per km ²
Average income of the population	adjusted monthly earnings (1000 HUF)
Employment in industry	per 1000 inhabitants
Proportion of children 0–14 years old	as a percentage of all households
Average floor area for new dwellings	m ²
Housing density	inhabitants per 100 dwellings
Housing construction	new dwellings per 1000 inhabitants
^a HUF—Hungarian forints.	

however, a major part of Pest County is the suburban belt of Budapest (Kok, 1997, page 77). The number of observations was 580, distributed across the seven regions (see table 2). For the sake of maximum comparability between two somewhat dissimilar sets of inputs, identical parameters and the same network size, 16×12 neurons as in the Dutch analysis, were selected. Now each of the neurons with its nearest neighbours is affected by twenty-one observations. Because there are fewer data here but the dimensions are the same, the map has a substantially higher resolution than the map of the Netherlands. Thus, on the one hand, more empty neurons are expected on the map surface; on the other hand, greater coverage of the projected input is expected compared with the case for the map of the Netherlands data.

4.2 The feature maps

The aim was to investigate whether similar patterns were to be found in the Hungarian data structure compared with the Finnish and Dutch analyses: namely, segmentation across areal units (here, counties) and segmentation within these areal units. As in the presentation of the Dutch analysis, one of the map layers is shown (figure 4), for a four-value categorisation of 'dwelling format and general prestige' of the predominant housing stock (single-family; garden city; old inner city; and 'prefab' high-rise estates).

The Hungarian data are aggregated on a much coarser level than the data for Finland and the Netherlands, being for only twenty counties. This causes problems for finding comparable evidence; in particular, the proximity and functional hierarchy of two counties cannot be determined with such broad areal definitions. Even so, the feature map of the Hungarian housing market shows clear clusters. A large

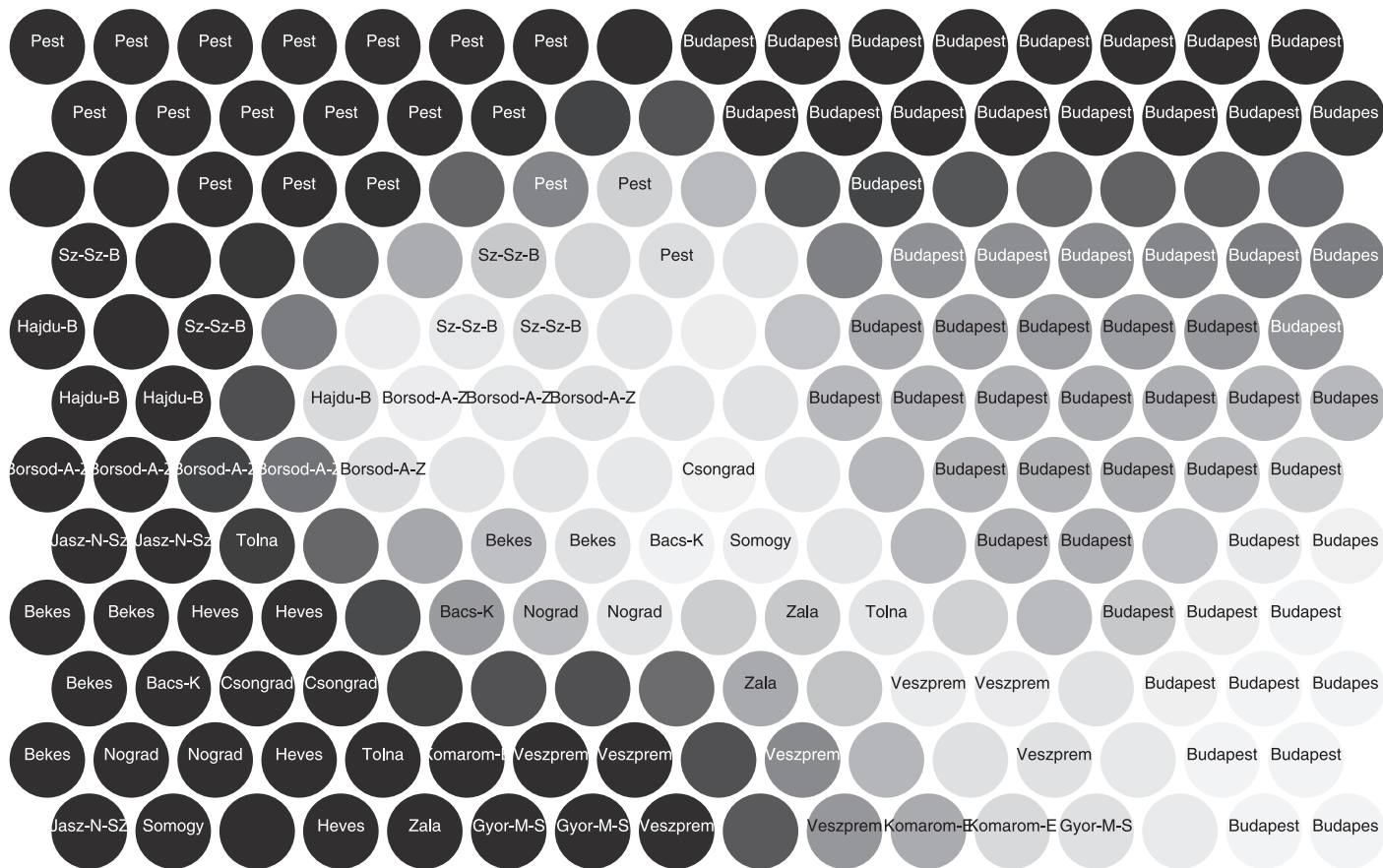


Figure 4. Feature map illustrating dwelling format and prestige of housing stock in Hungarian counties. Note: light shading, old urban and high-rise; dark shading, garden city and single-family housing. The labels are an abbreviation of the municipality followed by a two-digit code specifying the actual district within that municipality.

homogeneous block of Budapest neurons covers a large portion of the map; to a lesser extent, the same applies to Pest County, the large zone around the capital, which includes the large suburban outer ring of metropolitan Budapest. Elsewhere the clustering is relatively clear. Some of the twenty counties do not show up at all, whereas others are represented by more than one neuron or cluster of neurons on the feature map.

The ‘market value per square metre’ indicator partitions two thirds of the Budapest neurons as ‘expensive’; the rest of the map is then ‘on the cheaper side’; the in-between case is the upper left-hand corner neuron labelled after Pest. The collateral value overlaps with that pattern (apart from two Budapest neurons, and the Pest corner neuron, the collateral values of which are low compared with the market value). The age of the building is a relevant criterion in Budapest only; a homogeneous block of old housing stock belongs to the more expensive Budapest segment. Dwelling format, utilising two-valued and four-valued dummies, also shows a clear segmentation within all regions. All regions and most counties have two segments depending sharply on the individual house type (single-family or multifamily); even in the single-family-dominated Pest County a multifamily segment can be found. In Budapest the single-family segment includes both high-priced and low-priced dwellings. For the four-value effect, whereas the modest ‘prefab’ housing type shows up in all four types of basic contexts (Budapest, Pest, Transdanubia, and the rest of the country), the more expensive ‘garden city’ type shows up in Budapest, and as one neuron in Pest and Veszprém. In Budapest, the middle range of the market is captured in the neurons of the old inner city, moderately priced ‘prefab’ types, and some single-family cases. Size does not dominate the whole map; only in Heves is the housing stock divided into larger (typical values of the neurons being 135 m² and 143 m²) and smaller (106 m² and 110 m²) single-family segments.

The time of sale is also not the most dominating feature, although in three cases this feature is the most important partitioning criteria; the single-family (or low-density) segments in Tolna and Szabolcs–Szatmár–Bereg, and the multifamily segment in Bács–Kiskun, are separated into earlier and more recent sales, respectively. Not too much could be expected from this attribute, as the time period covers only a six-month cross-section. However, some minor associations are captured here: the earlier sales are mostly low-density cases (that is, single-family and ‘garden city’ types) and also some more idiosyncratic high-density cases (that is, ‘urban’ or ‘prefab’ types)—this applies to all four macroregional contexts. Based on the individual level variables, it is clear that the spatial housing market structure in all regions (and possibly counties) is partitioned into two or more different submarkets.

The remaining nine (county-level) variables help us to determine the functional and hierarchical position of the counties within the total dataset. The clearest clustering is caused by the indicators for net migration, housing construction, population, population density, and income. Housing construction and net migration shows the character of the Pest County cluster (upper left-hand corner of the map) in relation to the rest of the data structure. Then, in Budapest, there is obviously a much larger population, higher population density, fewer children, and smaller households (indicated by a low ratio of inhabitants per dwelling) than in any of the other nineteen counties. Budapest, most of Pest County, and some of Veszprém have the highest income levels. These counties, together with the northwestern corner of the country (that is, Komárom–Esztergom, Győr–Moson–Sopron; see figure 3) form a clear pattern of high average floor areas for new dwellings; apart from the Pest cluster, this pattern also overlaps with high values along the ‘employment in industry’ dimension. Thus the average size of new dwellings together with the share of employment in industry is a strongly

discriminating feature over the whole map. All dimensions show the functional role of each county. In addition, the level of aggregation shows the hierarchical role of Budapest and its hinterlands, situated partly in Pest County and partly in the Central Transdanubia region (see table 2 and figure 3).

The pattern described above may now be summarised into two general points on segmentation. First, the functional position within the whole country is defined by socioeconomic and other county-specific data, by geographical proximity, and, in the case of the sphere of influence of Budapest, by the hierarchy within the regional system. Second, heterogeneity and segmentation within a county are determined in terms of house type and size, price level, and time of sale. For example, Heves is represented by two typical observations—smaller houses sold earlier and larger houses sold later during the period of data collection—and, in Budapest, the single-family and ‘garden city’ environment is separate from the urban and suburban ‘prefab’ cases. However, the boundaries between the clusters are fuzzy.

4.3 Evaluation of the findings

The next step was to conduct an evaluation of the results in relation to the Finnish and Dutch cases and in relation to the expert knowledge of the Hungarian housing-market context. Together, the patterns show the structure of the dataset, both on a subnational and regional scale and in a more spatially detailed sense in terms of individual dwelling attributes. The structure shares features from both Finland and the Netherlands; features of relative location based on similarities related to the local economy and housing market, population, and geographical proximity can be identified, as can a sharp market segmentation within counties (and regions). The intermunicipal aspect was more explicit in Finland than in the Netherlands, and such findings were expected from the Hungarian analysis too, partly because of the similarity of data used and partly because of the similarity in actual context in terms of centre–periphery relations.

Indeed, the feature maps of Hungary show similarity to the analysis of Finland: segmentation between core and peripheral areas in terms of relative location; and between areas in other respects. Helsinki and its surrounding municipalities clearly were different from the rest of the country, and the same can be said of Budapest (with parts of Pest and Transdanubia) in relation to the rest of the country: a major part of Budapest was shown as being more expensive than the rest of the country. From this it can be noted that Budapest markets are differentiated in a similar sense to Helsinki in the Finnish analysis but, compared with the analysis of Finland, where only Helsinki and Espoo had expensive nodes, the picture shows a more even distribution. As in Helsinki, in Hungary the segments showing the most financial appreciation are the Budapest segments; however, Budapest also has a substantial submarket of relatively low-priced housing.

The Hungarian results also show similarity to the analysis of the Netherlands; with segmentation within all types of areas. The ‘Finnish type’ of segmentation across macrolocations was not, of course, surprising, as the Finnish analyses used similar input variables: individual data together with certain aggregate variables that provide more information on the macrolocation. In the Dutch analysis, in turn, the dataset was aggregated. Because of the greater problem of data incompatibility, any similarity with the Netherlands is therefore more interesting than any similarity with Finland. In Hungary the homogeneity across localities is perhaps greater than is often thought.

As in the cases of Finland and the Netherlands, the results were evaluated against expert knowledge of the context. Because of historical path dependence, a factor not included in this study, the Hungarian housing market is fundamentally different from those of the other two contexts studied. The market is strongly segmented between new

high-quality and old low-quality housing—this goes for inner-city and suburban locations alike. Traditionally, the housing market is constrained (although it has been speculated that accession to the EU may reverse this situation), low-density suburban housing is an absent phenomenon, and residential mobility between regions is poor. Preference formation has not been allowed to lead market processes until very recently, as popular areas have emerged in distant suburbs, on the one hand, and modernised flats in old architectural blocks in the inner city, on the other hand. A related aspect is that rural living is still related to agriculture—not to ‘pseudo-rural’ preferences, as was the case in the Netherlands (see Haartsen et al, 2003).

As the built environment requires at least two or three decades of investment to show any visible changes on an aggregate level, it is not surprising that after only one decade the majority of the Hungarian housing market structure does not show similar features to the two more developed market contexts, where the interplay between planning and market forces has shaped the environment into a mosaic of residential upmarket, middle-market, and downmarket areas. Rather, the opposite has happened: certain ‘hotspots’ developed instantly, whereas the majority of locations and house types are lagging behind in terms of quality. This polarisation is occurring because of an increased exposure to a competitive market in an initially constrained housing-market context.

5 Final remarks

In this cross-national study, an inductive approach to empirical market modelling was applied with the aim of exploring the datasets in question and relating any useful results on housing market structure and submarket formation to what was known about each national context. It was not, however, possible confidently to connect the results generated by the SOM with expert knowledge as well as was hoped at the outset; clearly, one has to be modest about the potential gains to be achieved in such analyses. To combine expertise on different housing market contexts with SOM output requires some creativity. Furthermore, data incompatibility was a potential problem in the sense that the variables themselves and the level of aggregation were, unfortunately, different in the different datasets.

This exercise nevertheless has demonstrated that the SOM-based method reveals patterns of housing-market-related socioeconomic and physical structures. Interpretation of the patterns for each layer allows a relative analysis of aspects of attractiveness as well as identification of more qualitative features across areas with various geographical scales and function, and, at a higher level of abstraction, allows comparisons across completely different institutional market contexts. The classification tells us (1) where the segments are situated (that is, clusters on the map) and (2) their determinants (that is, input variables and/or other features based on interactions between variables). Furthermore, the interesting thing about the method is that a quantitative input may lead to a qualitative output.

The SOM-based, inductive, exploratory method was applied to three nationwide housing-market sets. The dataset for Finland was analysed first and, on the basis of these results, some hypotheses were generated with regard to the identification of subnational submarkets. After that, a dataset for the Netherlands was analysed similarly, and the results compared with regard to a few key findings from the Finnish analysis. Running two datasets and obtaining two sets of results allows a generalisation of the model. In order to test the model, a dataset for Hungary was analysed against the framework set up on the basis of the first two cases.

The advocated SOM-based method of comparing and visualising the attractiveness of localities based on socioeconomic and physical variables is a method with generic

usefulness in spatial analysis. Furthermore, to make the output more understandable for an audience that is not familiar with SOMs it is possible to code the resulting SOM classifications and export the results into a geographical information system (GIS). This research also has a policy aspect: the resulting patterns determine potential policy solutions to fight decay. Prosperous cities and areas should be distinguished from disadvantaged regions, and, the 'ideal places' should be emulated by other, less favourable, types of areas. Interestingly, it may be that the same type of environmental, physical planning, housing, and employment policy will fit two areas that are located hundreds of kilometres apart and that also show considerable differences in many of their input variables.

Acknowledgements. An earlier version of this paper was presented at the ERES conference in Helsinki, 10–13 June 2003. Before that, the method and the Finnish part of the analyses had been presented at the ENHR workshops in Gävle, 26–30 June 2000. I am indebted to helpful comments from many participants at these events and to many other enthusiastic people, especially those who, in informal meetings, have helped me become acquainted with the three country-specific housing markets. Finally, I would like to thank Professor Batty and all the anonymous referees for improving the manuscript to its final form.

References

- Amsden A H, 1995, "Inductive theory in economics development: a tribute to Wassily Leontief on his 90th birthday" *Structural Change and Economic Dynamics* **6** 279–293
- Ball M J, 1973, "Recent empirical work on the determinants of relative house prices" *Urban Studies* **10** 213–233
- Boelhouver P J, 2001, "Capital accumulation via home-ownership: the case of the Netherlands" *European Journal of Housing Policy* **2**(2) 167–181
- Bontje M, 2001 *The Challenge of Planned Urbanisation* doctoral dissertation, Amsterdam Study Center for the Metropolitan Environment, University of Amsterdam, Amsterdam
- Bootsma H G, 1998 *The Myth of Reurbanization: Location Dynamics of Households in the Netherlands* (NETHUR D Publications, Amsterdam)
- Daly J, Gronow S, Jenkins D, Plimmer F, 2003, "Consumer behaviour in the valuation of residential property: a comparative study in the UK, Ireland and Australia" *Property Management* **21**(5) 295–314
- Dieleman F M, Wallet C, 2003, "Income differences between central cities and suburbs in Dutch urban regions" *Tijdschrift voor Economische en Sociale Geografie* **94** 265–275
- Douglas M J, 1997 *A Change of System: Housing System Transformation and Neighbourhood Change in Budapest* doctoral dissertation, Nederlandse Geografische Studies 222, Koninklijk Nederlands Aardrijkskundig Genootschap, Utrecht
- Dutt A K, Costa F J, 1985 *Public Planning in the Netherlands: Perspectives and Change Since the Second World War* (Oxford University Press, New York)
- Haartsen T, Huigen P P P, Groote P, 2003, "Rural areas in the Netherlands" *Tijdschrift voor Economische en Sociale Geografie* **94** 129–136
- Hatzichristos T, 2004, "Delineation of demographic regions with GIS and computational intelligence" *Environment and Planning B: Planning and Design* **31** 39–49
- Izraeli O, 1987, "The effect of environmental attributes on earnings and housing values across SMSAs" *Journal of Urban Economics* **22** 361–376
- James H, Collins A, Lam E, 1994, "The principles and practice of artificial neural networks in property valuation studies", discussion paper, Department of Land and Construction Management, University of Portsmouth, Portsmouth
- 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**(1) 67–90
- Kaski S, 1997 *Data Exploration using Self-Organizing Maps* PhD thesis, published in *Acta Polytechnica Scandinavica: Mathematics, Computing and Management in Engineering Series* 82 Helsinki University of Technology, Espoo
- Kaski S, Kohonen T, 1995, "Structures of welfare and poverty in the world discovered by the self-organizing map", Laboratory of Computer and Information Science, Faculty of Information Technology, Helsinki University of Technology, Helsinki

- Kauko T, 2002 *Modelling Locational Determinants of House Prices: Neural Network and Value Tree Approaches* doctoral dissertation, Faculty of Geographical Sciences, University of Utrecht, <http://www.library.uu.nl/decollectie/proefschriften/11688main.html>
- Kauko T, Hooimeijer P, Hakfoort J, 2002, "Capturing housing market segmentation: an alternative approach based on neural network modelling" *Housing Studies* 17 875–894
- Kohonen T, 1995 *Self-organizing Maps* (Springer, Berlin)
- Kohonen T, Hynninen J, Kangas J, Laaksonen J, 1996, "SOM_PAK: the self-organizing map program package", Report A31, Laboratory of Computer and Information Science, Faculty of Information Technology, Helsinki University of Technology, Helsinki
- Koikkalainen P, 1994, "Progress with the tree-structured self-organizing map", in *Proceedings of the 11th European Conference on Artificial Intelligence, August 8–12* (John Wiley, Chichester, Sussex) pp 209–215
- Kok H J, 1997 *Migration in Hungary and Poland Before and After the Transition* doctoral dissertation, Faculty of Geographical Sciences, Utrecht University, Utrecht
- KSH 2000 *Központi Statisztikai Hivatal* [Statistical Yearbook] <http://portal.ksh.hu>
- Leontief W, 1976, "Mathematics in economics", in *Essays in Economics, Volume 1* (White Plains International Arts and Sciences Press, New York), pp 22–44
- MacLennan D, Tu Y, 1996, "Economic perspectives on the structure of local housing systems" *Housing Studies* 11 387–406
- Meen G, 2001 *Advances in Urban and Regional Economics, Volume 2. Modelling Spatial Housing Markets: Theory, Analysis and Policy* (Kluwer, Boston, MA)
- Openshaw S, 1998, "Neural network, genetic, and fuzzy logic models of spatial interaction" *Environment and Planning A* 30 1857–1872
- Openshaw S, Blake M, Wymer C, 1994, "Using neurocomputing methods to classify Britain's residential areas", WP 94/17, School of Geography, University of Leeds, Leeds
- Potepan M J, 1996, "Explaining intermetropolitan variation in housing prices, rents and land prices" *Real Estate Economics* 24(2) 219–245
- Siikanen A, 1992, "Asuntojen kysyntä, tarjonta ja alueellinen erilaistuneisuus" [The demand, supply and spatial differentiation of dwellings], housing research series number 4, Department of Research and Planning, National Housing Board of Finland, Helsinki
- Treiman D J, Szelényi I, 1990, "Social stratification in Eastern Europe after 1989", Department of Sociology, University of California, Los Angeles, CA
- Tu Y, 2000, "Segmentation of Australian housing markets: 1989–98" *Journal of Property Research* 17(4) 311–327
- VROM, 1999 *Kerncijfers Wijken en Buurten* [Key figures of districts and subdistricts], Het ministerie van Volkshuisvesting, Ruimtelijke Ordening en Milieubeheer, The Hague
- Watkins C A, 2001, "The definition and identification of housing submarkets" *Environment and Planning A* 33 2235–2253
- Wong C, 2001, "The relationship between quality of life and local economic development: an empirical study of local authority areas in England" *Cities* 18(1) 25–32