

# Comparing spatial features of urban housing markets

Recent evidence of submarket formation in metropolitan Helsinki and Amsterdam

Tom Kauko

---

# Comparing spatial features of urban housing markets

Recent evidence of submarket formation in metropolitan Helsinki and Amsterdam

---

## Sustainable Urban Areas 7

---

### Editorial Committee

Prof. dr. ir. R. Dekker (Erasmus Universiteit Rotterdam)

Prof. dr. P. Glasbergen (Utrecht University)

Prof. dr. P. Hooimeijer (Utrecht University)

Prof. ir. W.G. Keeris MRICS (Technische Universiteit Eindhoven)

Prof. dr. T. Kreukels (Utrecht University)

Prof. ir. G.J. Maas (Technische Universiteit Eindhoven)

Prof. dr. N.J.M. Nelissen (University of Nijmegen)

Prof. dr. P. Nijkamp (Vrije Universiteit)

Prof. dr. H.F.L.K. Ottens (Utrecht University)

Prof. dr. ir. J. van der Schaar (Universiteit van Amsterdam)

Prof. dr. J. van Weesep (Utrecht University)

---

---

# Comparing spatial features of urban housing markets

Recent evidence of submarket formation in metropolitan Helsinki and Amsterdam

---

**Tom Kauko**

---

The series **Sustainable Urban Areas**  
are published by DUP Science

DUP Science is an imprint of  
Delft University Press  
Postbus 98  
2600 MG Delft  
The Netherlands  
Phone +31 15 2785678  
Fax +31 15 2785706  
<http://www.library.tudelft.nl/dup/>

---

Sustainable Urban Areas are edited by  
Delft Centre for Sustainable Urban Areas  
c/o OTB Research Institute for Housing, Urban and Mobility Studies  
Delft University of Technology  
Jaffalaan 9  
2628 BX Delft  
The Netherlands  
Phone +31 15 2783005  
Fax +31 15 2784422  
E-mail [mailbox@otb.tudelft.nl](mailto:mailbox@otb.tudelft.nl)  
<http://www.otb.tudelft.nl>

---

The author wishes to acknowledge the financial assistance of the Dutch  
government through the Habiforum Program Innovative Land Use and Delft  
University of Technology through the Delft Centre for Sustainable Urban Areas.

---

Design: Cyril Strijdonk Ontwerpburo, Gaanderen  
Printing: Haveka, Alblasterdam

---

ISSN 1574-6410; 7  
ISBN 90-407-2618-3  
NUGI 755

Subject headings: housing markets, spatial features, segmentation

© Copyright 2005 by Onderzoeksinstituut OTB  
No part of this book may be reproduced in any form by print, photoprint,  
microfilm or any other means, without written permission from the  
copyrightholder.

---

---

# Contents

Preface .....	1
<b>1 The purpose of the study and justification for the approach .....</b>	<b>3</b>
<b>2 Conceptualising housing-market segmentation .....</b>	<b>9</b>
2.1 Differentiation of residential areas, dwelling types and markets .....	9
2.2 Review of the theoretical urban-economics literature .....	11
2.3 Comparing methodologies for analysing the development of submarkets .....	16
<b>3 Empirical analysis of submarkets .....</b>	<b>21</b>
3.1 Overview of the modelling repertoire .....	21
3.2 Neurocomputing and the SOM-LVQ classifier .....	27
3.3 The specific research design .....	34
3.4 Supporting the analysis with expert interviews, using the AHP .....	36
3.5 The comparative perspective and the institutional and behavioural aspects .....	39
<b>4 Results of the submarket classifications in Helsinki .....</b>	<b>41</b>
4.1 Study area and data .....	41
4.2 Analysis with the SOM .....	46
4.3 Analysis with the LVQ .....	48
4.4 Modelling the spatial dynamics of the Helsinki housing market according to a comparison with a later cross-section .....	50
4.5 Expert interviews regarding Helsinki housing markets .....	56
<b>5 Results of the submarket classifications in Amsterdam .....</b>	<b>59</b>
5.1 Study area and data .....	59
5.2 Analysis with the SOM .....	68
5.3 Analysis with the LVQ .....	73
5.4 Modelling the spatial dynamics of the Amsterdam housing market according to a comparison of two cross-sections .....	74
5.5 Expert interviews regarding Amsterdam housing markets .....	79
<b>6 Analysis of Amsterdam, Rotterdam and The Hague with the SOM .....</b>	<b>87</b>

---

---

7	Generalisation of the results: gaining theoretical insight into segmentation . . . . .	91
7.1	The power of the comparative analysis of housing markets . . . . .	91
7.2	Methodological and theoretical considerations . . . . .	96
7.3	Practical application . . . . .	99
	References . . . . .	103
Appendix A	Feature map layers of Helsinki housing markets – 1993 data . . . . .	113
Appendix B	Feature map layers of Helsinki housing markets – 2001 data . . . . .	119
Appendix C	Detailed level aggregated and disaggregated AHP models for Metropolitan Helsinki . . . . .	125
Appendix D	Selected feature map layers of Amsterdam housing markets – KWB/1999 data . . . . .	131
Appendix E	Feature map layers of Amsterdam housing markets – taxation/panel data . . . . .	139
Appendix F	Feature map layers of Amsterdam housing markets – taxation/1992-93 data . . . . .	145
Appendix G	Feature map layers of Amsterdam housing markets – taxation/2000-01 data . . . . .	151
Appendix H	The disaggregated AHP models for the Dutch Randstad . . . . .	157
Appendix I	The variables for the aggregated housing market dataset of Amsterdam, Rotterdam and The Hague . . . . .	161

---

---

# Preface

This study was conducted as a post-doctoral research project, and it continues along the path that was set by my PhD publication of 2002. I am grateful to a number of peers and mentors, and the space here is insufficient to mention all of them. I would first like to thank all of my colleagues at the ‘Sectie-VWM’ of the OTB Research Institute for Housing, Urban and Mobility Studies for listening to my presentations, reading my papers and giving me helpful suggestions. In addition, I appreciate the efforts of Martti Lujanen and the other participants in the ENHR workshop on ‘Housing economics’, which was held in Vienna, Austria in June 2002, for their constructive feedback. This assistance came at a time when I needed it most. Perhaps my greatest debt of gratitude is due to my supervisors, Peter Boelhouwer and Marja Elsinga at OTB, for providing me with valuable advice and for their flexibility when dealing with demands, which were difficult at times. A different kind of gratitude goes to the technical-assistance staff (especially Herman Toneman) at OTB for their practical assistance. I am also thankful to Jacco Hakfoort (Ministry of Economic Affairs), Manuel Aalbers (AME/UvA) and Willem Teune (SWD Amsterdam), for explaining the differences amongst the various Amsterdam neighbourhoods to me. I am particularly indebted to Seppo Laakso (Kaupunkitutkimus) and Marc Francke (OrtaX), as well as Gemeentebelastingen Amsterdam, for making their massive transaction-price datasets available for me. Although it has been a long time, I have also not forgotten the strict but supportive attitude of my doctoral supervisors: Pieter Hooimeijer from Utrecht University and Kauko Viitanen from HUT, Finland. An additional name from my years as a doctoral candidate, and the person to whose memory this book pays tribute is Frans Dieleman, who inspired me in my research projects when I was a doctoral student at Utrecht University. He was particularly influential in convincing me of the importance of building housing-market models, carrying out comparative studies, engaging in inter-disciplinary research across the spatial, economic and social science communities, combining qualitative and quantitative knowledge and conducting – in his own words – ‘research driven by curiosity’.

Tom Kauko, Delft, May 2005

---



---

# 1 The purpose of the study and justification for the approach

Although simple on the surface, the housing market is a complex and multi-faceted topic for scientific inquiry. To date, the socio-spatial dimension has received far less attention in housing-market analyses than has the purely temporal dimension (i.e., market trends and the development of prices and price changes over time). A core American tradition of microeconomics-based land-use and house-price modelling is one notable exception (see Maclennan & Whitehead, 1996, for a brief review of the most important traditions).

Regardless of the current influence of the spatial tradition within housing-economics academia, pragmatic justifications for the object of study – the relative differences between housing-market bundles across an urban area – are not difficult to find. A dwelling in one location is usually not a perfect substitute for a dwelling in another location. The qualitative and discontinuous nature of various location-specific attributes that contribute to the spatial dynamics of housing markets tends to segment the market into submarkets.

Maclennan and Tu (1996) note that progress in housing economics beyond neoclassical reductionism and summary analyses may require explorations outside of the standard framework. According to this contrary view, which is known as commodity variety, consumer choice is but one dimension of a market; space and time are real dimensions as well. In keeping with this argument, theoretical models of housing submarkets should allow for market failures in a way that resembles macroeconomic analyses, and unitary markets fit only within the microeconomic framework. The authors observe that, with regard to other specific factors and circumstances that are related to the formation of submarkets, neighbourhood atmosphere may be impossible to recreate. Furthermore, space is not only an attribute in the preference set; it also acts as a friction and constraint parameter in a spatially dispersed market. For many reasons, new supply in location B does not necessarily remove the price premium in location A, as Maclennan and Tu rightly note. In economic terms, there is no spatial arbitrage in such situations. This is a key concept that underlies much of the debate on submarket formation.

The empirical investigation of segmentation (i.e., the emergence of submarkets) raises a question: if segmentation is observed within a given context, what are the criteria for detecting segmentation? Should segmentation be measured by price level, or should it be measured by other 'objective' socio-economic, demographic or physical features of the location? In other words, can housing submarkets be attributed to certain particular features (i.e., discriminating criteria) of the housing-market area that are empirically observable? To capture the dynamics of housing-market structure, Tu (2003) suggests re-classifying submarkets after a certain time. Has one set of discriminating factors increased or decreased in importance relative to another set, and has the resulting spatial form become more or less regular? If so, what are the most important specific characteristics of this spatial form, according to the analysis? For example, a particular criterion (e.g., CBD distance, synthet-

---

ic and physical environment, neighbourhood status, price, house type, dwelling quality, age of building) may have taken precedence over another criterion. Furthermore, it is not certain that the same spatially defined segments are observable in the same area at two points in time. Is it even feasible to generalise across different urban housing markets, in which contextual exogenous influences – be they the results of governmental intervention or the deeply rooted values and beliefs of housing consumers – contribute to the mosaics of segmentation?

A number of empirical modelling approaches have been proposed that might be able to capture this influence. In this report, the neural-network approach to the classification of market segments is used as an alternative for other, more common methods, which are based on hedonic price, social-area analysis or both. The project follows the pioneering contribution of Kauko (1997; 2000; 2001; 2002), which explains the method. The project compares the results of submarket structure obtained with the neural-network approach from two geographical contexts: the housing markets of Helsinki, Finland and Amsterdam, the Netherlands. Two additional Dutch cities, The Hague and Rotterdam, are subsequently incorporated into the analysis, in order to determine the presence of any national between-cities variations that in any way resemble the cross-national variations. The incorporation of these cities allows the evaluation of similarities between Helsinki and Amsterdam with regard to the effect of the shared country-context for housing-market outcomes. The time-period under study is the 1990s (and early 2000s). Thorough analyses of hedonic house prices in Amsterdam (Needham et al., 1998) and Helsinki (Laakso, 1997) have already been conducted, using the same datasets that are addressed in this report. These analyses will be helpful for the interpretation of the results in each case.

A neural network is a nonlinear and flexible (i.e., model-free, non/semi-parametric) regression technique that requires no pre-specified formal theory. A number of neural network-based applications are in use within the fields of economics and finance (e.g., Yoon et al., 1993), as well as in research on urban issues and planning (e.g., Raju et al., 1998). The proposed specific neural-network classification method is based on the self-organizing map (SOM) and the learning-vector quantification (LVQ). To the best of my knowledge, this method has not been applied to the modelling of housing markets. Nonetheless, a number of recent applications have applied similar logic: in population geography, work by Openshaw and colleagues (1994) on classifying residential areas; in property valuation, Lam (1994), James and colleagues (1994), Jenkins and colleagues (1999) and Kauko and Peltomaa (1998). Because the aspect of housing-market segmentation is arguably linked closely to the aspect of residential valuation (e.g., Adair et al., 1996; Kauko, 1999; Jenkins et al., 1999), it is logical to extend the applicability of SOM-based methods to the modelling of spatial housing-market dynamics in general, and to the classification of hous-

---

ing submarkets in particular.

All of the above-mentioned contributions use the SOM as a tool for reducing various dimensions of the input data and for clustering the observations according to these reduced dimensions, in order to examine the structure of the dataset. In many cases, implicit locational aspect is included within the dimensions. My contribution will follow this path to explore the structural features of housing markets in urban areas. An innovative aspect of this study is that I will conduct the empirical work in two different geographical contexts (and time-periods), in order to link the results to a theoretical framework that captures dynamic and institutional factors that shape local housing markets. Although this research design does not involve the formulation of hypotheses (at least not in the strict positivist sense), the guiding framework arises from expectations regarding a number of key relationships across market areas that are identifiable according to spatial variations in demand, supply and prices.

There are at least three justifications for using these somewhat unconventional methods. First, urban and metropolitan areas in mainland Europe have received considerably less research attention than have their American, British and Australian counterparts. The lack of attention is obviously related to data availability and, perhaps, to issues related to funding. The most important reason, however, is that research interests thus far have been directed towards national housing markets. A large gap remains to be filled. Second, regardless of the context under study, the complex nature of the various housing-market processes arguably requires an approach that is more sophisticated than the combination of market-equilibrium based tools and conventional statistical analysis allows. This point is debatable, however, and the aim of this study is not to develop a rhetorical argument in favour of a more creative approach over one that is more commonplace. At this stage, suffice it to note that the selected approach – or sequence of approaches – is championed on pragmatic grounds, as it is appropriate to the multiple aims of modelling the housing markets in a given set of urban areas, comparing price differentials within each area and comparing the findings across these areas. For other tasks (e.g., property-price determination), other methods are likely to yield better conclusions than would the methods that are applied in this study (see Kauko, 2004, for a discussion). A third justification for the use of these methods is that practical or policy aspects are also frequently involved.

From a more practical point of view, I seek to provide market actors and socially conscious interest groups with a useful tool to aid decision-making with regard to the urban housing-market environment. Such applications can take a variety of forms, ranging from the selection of particular sites according to their value potential to the determination of relative and coarse differences between houses or locations (valuation bands) for tax assessment and other mass-appraisal purposes. I will argue that this study demonstrates the

---

general applicability of the approach for the purposes of classification and assessment. This argument is the last topic of the study.

The extent to which a method based on neural-network modelling provides an alternative for the hedonic regression modelling of housing prices is by no means straightforward (see e.g., Borst, 1995; Worzala et al., 1995; Jenkins et al., 1999). The involvement of certain additional aspects (e.g., segmentation, visualisation and the smoothing of the data set) also makes the technique, at least to some extent, supplemental to hedonic analysis (Kauko, 2002). The neural-network approach, however, differs in two important ways from standard hedonic regression and the more developed space-varying coefficient (SVC, see Pavlov, 2000) techniques: (1) the neural-network approach allows only general and a posteriori theorisation; (2) it requires no strict assumptions regarding the smoothness of the association between price and locational attributes.

Note that the neural network itself is no more than a helpful tool for arranging information; arriving at valid conclusions after the exercise requires theory and local knowledge. For example, why do price structures vary from place to place, even though the various dimensions of price formation are apparently identical? The explanation might be rooted in textbook theory (e.g., Alonso's accessibility-space trade-off in urban housing markets) or connected to the specific context in question (e.g., the location of ethnic minorities).

Openshaw (1998) expresses surprise at the extent to which neural-network models have been neglected. Although the technology is well established, it is surrounded by a 'conservative prejudice', largely fuelled by the 'black box' argument, which must be overcome. Sensitivity analysis and the use of other computational methods to support the modelling procedure are two logical responses to the prejudice. For example, it is possible to compensate for the unsatisfactory rigour of the pure SOM technique by combining it with the LVQ technique, as will be the case in this study. Note also that a recent trend in neural-network applications in various industries focuses on assembling information according to the recognition of patterns, rather than on learning and prediction according to the computation of simple stimulus-response combinations, as was the original idea of neural networks (cf. Nelson & Nelson, 2002).

The sole application of an approach that is based on neural-network modelling and actual housing-market transactions, however, is not sufficient to address all relevant relationships. In particular, such approaches leave the less tangible and more nuanced aspects of the choice process of typical housing consumers with regard to residential location unexamined. This is the basic criticism that is levied against the hedonic-regression approach, and the neural network approach offers little improvement. It is therefore necessary to combine the neural-network approach with an approach that is more sensitive to behaviour and that allows the examination of perceptions, preferenc-

---

es and intentions, in addition to the market choices and prices that the neural-network approach reveals. To support the housing-market classification that is generated by the SOM and the LVQ, a completely different method will be applied to both geographical contexts. In this study, issues of convenience motivated a choice for the analytic hierarchy process (AHP) protocol; the AHP is based on a pair-wise comparison of preferences, and it thus requires hypothetical data on consumer preferences rather than actual market-outcome data, as in the main approach. In addition, the AHP is remarkably pragmatic. The judgements of carefully selected expert respondents can be used to elicit a number of preference profiles and to highlight various dimensions of locational quality. Such information potentially enhances the analysis of typical patterns in the structure of housing markets, as it allows the identification of several relevant buyer segments within a total-market model that has already been generated. We consequently obtain information on the level of typical consumer intentions and preferences, as well as on the aggregate structure level. In this way, each method both compensates for the weaknesses of the other and adds to the accumulated evidence by providing different types of information. Recent house-price analyses have suggested that such triangulation of two different approaches could be valuable, as it enhances both the credibility and the depth of the study (e.g., Strand & Vågnes, 2001; Kauko, 2002).

Finally, this report concerns an interdisciplinary research project on the development of housing markets in various European metropolitan areas. This topic is of considerable interest to the urban-development and housing markets, which are beginning the process of globalisation. How are various locational features related, and what sort of pricing mechanisms are able to explain the property prices in different areas? According to Daly and colleagues (2003), the goals of cross-national studies are ultimately theoretical. Such theoretical ambitions may be achieved through a systematic process that begins with description, thereafter searching for generalisations across the study areas and datasets, together with any idiosyncrasies that may enrich the analysis. This is not to say, however, that the resulting knowledge automatically allows analysts to create theory in this way. In this study, a possible segmentation of the urban housing market, either along purely spatial scale (micro and macro-location), more functional properties (e.g., type, age, size, qualitative characteristics, financing) or transaction price, is merely confirmed with a mode of analysis that is partly inductive and partly descriptive (cf. Kauko, 2002, p. 50). When such analysis is carried out in several geographical/institutional and temporal contexts, considerable generalisation of the findings may be possible. Even if this study is not yet able to arrive at substantial theoretical conclusions, it does provide a thorough exploration of plausible relationships between various features of housing markets, behavioural processes and their broader institutional contexts.

---

This investigation of market segmentation in different institutional and geographical contexts and the methodological evaluation can be summarised in five research questions.

1. How do the spatial housing markets of Metropolitan Helsinki and Amsterdam (together with the rest of the Randstad region) differ in terms of patterns, criteria and dynamics?
2. How are housing prices related to the relevant socio-demographic, physical and institutional features of particular housing-market areas?
3. Do these relationships change over time, when considering each context for approximately the past decade?
4. In comparison to hedonic price analysis and other more traditional methods, how conveniently can we study all of these aspects using the approach based on the SOM and the LVQ, supported by the AHP expert-interviewing technique?
5. How can this tool be used to aid decision-making with regard to the housing market and physical environment?

The text is structured as follows: Chapter 2 provides a brief overview of the broad approaches to studying the identification and classification of housing submarkets. Chapter 3 then follows with a similar discussion of the most common empirical modelling techniques that are currently available. The proposed approach is then applied to the empirical analysis of metropolitan Helsinki and Amsterdam (Chapter 4 and Chapter 5, respectively). These chapters also discuss issues of data quality and comparability. Chapter 6 contains an analysis of the three largest Dutch cities (Amsterdam, The Hague and Rotterdam) using the SOM. Finally, Chapter 7 presents the conclusions (some of which are tentative) that have been drawn from the study, generalising the findings to a moderately theoretical level and providing suggestions for practical applications.

---

## 2 Conceptualising housing-market segmentation

### 2.1 Differentiation of residential areas, dwelling types and markets

Most, if not all, urban dwellers know that different neighbourhoods look different, accommodate different residents and have different levels of price and rent. Some are distant from or poorly connected to employment and service centres, while others are centrally located or well connected by public transport or motorways. Some areas have all kinds of public and private services. Other areas may lack services altogether, but have plenty of green space. Some are clearly delineated by rivers, traffic arteries or other boundaries, while others are merely extensions of the inner city. This is universal, general knowledge, and it is hardly a new phenomenon. Differentiation, however, does not necessarily imply segmentation; markets may be composed of all kinds of areas, such that housing consumers consider the whole city when searching for a new home. Prospective buyers or tenants look for homes in the whole city, and they do not consider it a problem that the city is comprised of various parts, the character of which differs widely. Segmentation is determined by the extent to which the market activities of these various urban component areas overlap. If the moving activity in one area overlaps with moving activity in other areas, these areas are said to be part of a single market. On the other hand, if there is no such movement, the market forms a self-contained submarket or market segment (e.g., Jones, 2002). When buyers and sellers interact within an area that is defined as a submarket, they do not consider other parts of town. Berlin during the time of the wall and Belfast are two classic (albeit naïve) examples of extreme segmentation that illustrate that political and religious considerations are the most important reasons for the strictest possible self-containment.

On the other hand, the definition of segments or submarkets need not be spatial; it can be functional as well. In situations, single locations are divided into two or more submarkets in which the same self-containment principle applies: regardless of physical proximity, markets for completely different dwellings do not overlap. The common structure of the pre-war inner-city condominium blocks in certain central European cities (e.g., German cities and Budapest) offer an illustrative example. These buildings comprise two kinds of dwellings: on the façade side, large and prestigious units face the street; on the back side, small, dark units face the courtyard. It would be impossible to consider these two types of units as parts of a single market; it is often noted how the occupants of the same building actually never even met during their daily movements. More recently, the most usual perception of this kind of segmentation – at least in welfare states – has been the demarcation between private and public housing, and between free-market and regulated prices and rents.

The US-based urban economic literature on residential location, house

---

and land price differentiation and market efficiency is often considered the 'orthodoxy' in the field. Static analysis forms the starting point for this paradigm: does the intra-urban housing-market pattern correspond to a simple equilibrium, which would imply that one part of the city is exclusively populated by a high-income group, while another part of the city is exclusively populated by a low-income group? Alternatively, is the pattern one of multiple equilibria, which would allow for more heterogeneous inner city and suburban areas? What are the exact determinants and boundaries for such segments? Such questions are often (but not always) reformulated and subjected to dynamic analysis: how do the house prices and the choices of moving households, which are actually treated independently in this mode of analysis, respond to the impulses of supply and demand? All of the above-mentioned relationships are quantified using a mechanistic approach that incorporates carefully specified functions and statistical tests. Following a number of standard procedures, particular hypotheses are ultimately either rejected or confirmed, according to the outcome of such models. While this research tradition is much more formal than the one that is applied in this study, the basic goal and the conceptual reasoning are much the same. This line of research is therefore cited frequently throughout the paper, as it provides the only credible starting point for the analysis.

On the other hand, it can be argued that the assumptions for the modeling context in this study also differ greatly from those that pertain to American urban housing-market areas. To what extent can the voluminous US-based findings be transferred to the two urban housing markets under study? This question is complicated by at least three types of compatibility problems. First, the main difference between contexts is that, in the US, the relationship between price and quality is more transparent than it is in Europe. In the less efficient European market context, there is no robust evidence that quality and price are always related. Second, the set of neighbourhood amenities that is relevant to consider in the US is not exactly the same as it is in Finland or in the Netherlands. The need to improve safety and the quality of schools in the district are secondary factors in the European context. In general, the safety of the residential environments in Helsinki or Amsterdam (and other European cities) is adequate, even for young, well-educated populations. Third, in some European contexts, including the ones under study, the plot efficiencies in the inner city blocks may actually be lower than those of some suburban areas that were constructed later. In addition, residents often have a strong preference for the architecture and design of the blocks in inner-city locations; the cityscape tends to increase the attractiveness of the typically urban locations. It is therefore inappropriate to assume that the inner-city areas of these cities are subject to the same repulsion effect that is commonly found in studies of American (and even British) cities. Nonetheless, the European context requires additional consideration, as issues related to path-dependence, local-

---

ised disequilibrium and identity can lead to a remarkable complexity when identifying relevant relationships from the empirical material.

With these general considerations as a corollary framework, it can be noted that differences between segments may or may not be related to one or more factors that pertain to the location within the entire urban or metropolitan area. Examples of these differences include the appearance of the dwelling, the building and its vicinity, characteristics of the people who reside in the block or the neighbourhood, other characteristics of the area, price or rent levels and regulation. Alternatively, the differences may be so insignificant that they are overlooked altogether. Furthermore, segmentation may or may not remain over time. Regardless of the exact definitions that are involved, casual observation alone is sufficient to provide some knowledge of the processes and structures that differentiate the urban space, housing stock and market continuum. If the goal, however, is to develop a more rigorous measure for segmentation and its causes, spatial housing-market theory offers a set of analytical tools for examining the issue of segmentation both conceptually and operationally. This allows a relatively value-free evaluation of the factors behind segmentation, as well as the comparison of actual segments. The following section examines the theoretical treatment of the segmentation topic and related phenomena under study.

## **2.2 Review of the theoretical urban-economics literature**

Theories concerning the ways in which urban spatial housing patterns emerge have appeared within the microeconomics literature since the 1950s. The submarket concept renewed the discipline profoundly, as it pertained to the qualitative aspects of explaining the submarket structure, which was not explicitly recognised in conventional microeconomic models of urban housing structure and residential location. Below, I provide a brief account of the evolution of this research area.

The basic idea of the conventional or neoclassical urban economic theory is derived from the Ricardo's classic rent theory, and it was developed into a 'bid rent' theory of the consumer by Alonso, Muth and Mills during the 1960s and early 1970s. 'Bid rent' implies that there are different land use zones at different distances from the city centre, depending on the willingness of each group to pay (e.g., Mills, 1971; Laakso, 1997; see also Richardson, 1977; Evans, 1985; Bassett & Short, 1980; MacLennan, 1982). Within this approach, location is one argument in the consumption set and utility function of the household (e.g., Laakso, 1997). In the simplest urban model, in which all employment opportunities are situated in the CBD of a mono-centric, uniformly dispersed, round and flat urban area, the land price (or land rent) is assumed to depend

---

on the distance to the CBD and the area of the plot. While housing supply is assumed to be fixed in the short run, the equilibrium locations of households are derived within this static framework as a trade-off between the consumers' demand for space to live and access (i.e., low travel costs and short travel time) to the city centre. For a given level of income, therefore, households make a trade-off between more space and worse access or better access and less space (e.g., Maclennan, 1982; Laakso, 1997). Three standard explanations are given for the household rationale regarding residential location: (1) minimisation of travel costs, (2) minimisation of travel costs and housing costs among the same income group (Wingo, 1961) and (3) income and the availability and conditions of mortgage financing, without any efficiency trade-off, as suggested by the maximum housing expenditure theory of Ellis (1967) and Stegman (1969; cited in Balchin and Kieve, 1977: 31-34). The basic idea of a trade-off between space and access also received a dynamic context in later work (see e.g., Richardson, 1977).

Progressively more elaborate models were developed later, emphasising neighbourhood-level land use and environmental preferences (e.g., Richardson, 1971; Evans, 1985). This represented a shift in emphasis from the basic space-access trade-off to how people perceive the social and physical factors of neighbourhoods. Economic theory examines these factors as either positive or negative externalities that contribute to a certain amenity effect, which is internalised in house prices. The various amenities (and disamenities) of neighbourhoods and municipalities, along with their social and physical composition, are considered relevant factors, given that such land uses are also scarce within a city. This framework allows the incorporation of ecological (e.g., the coverage of green acreage), cultural (e.g., architecture) and social (e.g., status) amenities that together influence the quality of the vicinity, but do not directly depend on distance from the CBD. In most cases, however, locational factors are indicators of both accessibility and the socioeconomic and environmental composition of the surrounding area (see Maclennan, 1977). Furthermore, two significant extensions of the perspective were developed outside urban economics: hedonic price theory, from the price-index research sub-discipline, and capitalisation theory from the local public-economics sub-discipline. Hedonic price theory, developed by Griliches (1971) and Rosen (1974), explains how the implicit market prices of quantitative and qualitative property characteristics are formed by equating the supply and demand for each characteristic within a static framework, and combining them to arrive at the total house price. Capitalisation theory explains changes in the welfare level of an urban area according to costs and benefits that accrue to property owners because of publicly funded changes in a location-specific amenity. The analysis of public goods and neighbourhood quality as determinants of house prices could subsequently be incorporated into the static equilibrium framework (see Richardson, 1977; Evans, 1985; Laakso, 1997).

---

Even in the absence of other substantial influences, an urban area may be segmented within this framework if the preferences and income of the households differ according to the attributes space and accessibility. Similarly, the land use/environmental preferences approach can also be used to explain the occurrence of submarkets, and segmentation may be based on additional factors, including the dominant type of building, area density or even the internal attributes of the dominant type of apartment (see Laakso, 1997; Bourassa et al., 1997; Grigsby et al., 1987). Note that, in urban economics, segregation has been treated as a phenomenon of ambiguous significance for society (e.g., Evans, 1985: 33-34). In a rather complex mode of analysis, differentiation as such is not considered problematic. Although it involves positive aspects of economic efficiency, it simultaneously involves negative social externalities, which consequently present economic problems. This concept is considered when examining the background of the study area with respect to the housing market-indicators in each case.

John Kain was probably the first urban economist to recognise the need for a less elegant but more practically relevant modelling agenda. According to Glaeser and colleagues (2004), Kain moved beyond simple models (e.g., the type described by Alonso & Muth) in order to capture the heterogeneity (including the decentralization of employment) and other features of the urban landscape. Kain's early work (1960s-70s) stressed that economic opportunities are determined by the interrelationship between race and location, which is the core of the spatial mismatch hypothesis.

For the housing-market sub-discipline, the corresponding theory improvements were developed roughly between the late 1970s and the 1990s. The neo-classical microeconomic literature restricts housing-supply factors to physical constraints at the most; no variation is assumed in institutions (beyond a comparative-statics setting for the above mentioned effects of capitalisation). Nonetheless, institutional influences are crucial, as they are seldom market neutral (i.e., they tend to distort the efficiency of the market in one direction or another). For example, depending on the particular regime, a monopoly rent premium may arise. This premium may result from either some sort of external market intervention (i.e., overly strict planning regulations when the planning system has a market constraining function, as in the US and UK) or because of a lack thereof (i.e., insufficiently strict planning regulations when the planning system has a market ensuring function, as in the Netherlands and Finland). This is also the case for demand factors; individual preferences and demand-side institutions are taken for granted.

According to Maclennan (1977), market segmentation is one of four greatly neglected issues in contemporary housing-market research; the other three concern how housing attributes enter the individual's utility function and the non-uniformity of sub-groups, supply decisions and institutions. At the time, Maclennan's paper was probably the first constructive effort to revise

---

the basic hedonic model of the housing market, and could be seen as the beginning of modern urban housing economics. Since its publication, Maclennan's contribution has been cited heavily, and it has proven successful in setting the standards for subsequent house-price research. Several later authors, including Mason and Quigley (1996), have maintained that the existence of submarkets seems to be one reason why the standard hedonic specification does not work. This more refined genre recognises a variety of explanations for why separate housing markets may exist within urban areas (see Bourassa et al., 1997; Grigsby et al., 1987; Rothenberg et al., 1991; Whitehead, 1999; Goodman and Thibodeau, 1998).

Segmentation implies the sale of various types of goods in completely different markets, with variation in both the amounts of money and the preferences of producers and consumers largely diversified (Bourassa et al., 1997). Segmentation can thus be identified according to supply, demand and (quality-adjusted) prices. Although allowing for segmentation does not automatically deny the logic of neoclassical economic theory, the causes of the phenomenon are debated in the literature. The central argument is that submarkets may affect the relationship between location and price. Depending on the specific theoretical perspective, the main criterion is the character of the location, the price level or a combination of both (Tu, 2003).

Heavily influenced by the ongoing debate on economic methodology, a completely new type of approach began to emerge within the fields of housing economics and real estate during the 1990s. These approaches emphasised behavioural factors and complexity. This change in perspective was partially inspired by parallel debates and advances in the financial modelling literature, which questioned the concept of market efficiency (see Shiller, 2003). Although a considerable amount of such state-of-the-art literature exists within the context of residential valuation (see e.g., Daly et al., 2003), it remains to be seen whether there will be any significant diffusion into the housing-market research discipline. Below, I explain the most modest theory adjustments that have emerged from this position.

A certain area may experience upward or downward developments in value, depending on the time of development and the area's current image. This feature is arguably consistent with the evolutionary and Austrian schools of economic thought. In a loosely formulated explanation, the investment (or lack thereof) will either enhance the potential of that location, thereby attracting further investment and increasing the value even further, or lead to dilapidation, a loss in potential, absence of investment and further decreases in the value. In either case, however, the trend may be reversed; inappropriate structures may generate a downward trend in value formation and development activity, and the gentrification of a neighbourhood may lead to an upward trend. The Austrian school allows for a 'feedback framework' between market outcomes and policy formulation (see e.g., Monk et al., 1999). This has impor-

---

tant ramifications for urban housing-market analysis. One inconvenience, however, is that current theory tends to treat these arguments implicitly. For example, when Maclennan and Tu (1996) emphasise adjustment processes, market disturbances and price disequilibrium, it is obvious that their position originated in the Austrian school.

The Austrian school is a versatile line of economic thought that has broad applicability with regard to issues of actors, markets and ownership. The subjective costs that figure into this perspective arise when the role of entrepreneurial discovery is blocked for some reason(s). It then becomes necessary to overcome these impediments. In terms of property rights, the issue concerns non-contractible ex-ante investment. In terms of transaction costs, additional actors, whose agreement must be secured, increase the barrier (i.e., transaction costs) to entrepreneurial initiative. The answer lies in a re-assignment of ownership rights to improve the preconditions for entrepreneurial discovery (Ricketts, 2003).

If the submarket concept is appropriate to this context, it may be assumed that two (or more) potential submarkets exist amongst which price differences are generated by differences in supply constraints, quality or other aspects (e.g., asymmetric information, topography or public sector interventions). In a theoretical sense, the submarket/segment concept implies that, if the current supply in the submarket increases with price levels, the price differences may remain, thus validating the presence of segmentation. If the price difference is levelled due to spatial arbitrage, however, it is inappropriate to speak of separate price submarkets. Maclennan and Tu (1996) point out that spatial arbitrage may or may not exist within a given urban housing-market context.

The premise for this study is that it is possible to derive empirically testable propositions in relation to the most appropriate model of the market and, related to this, the most appropriate model of the relationship between location, land use and preferences. Can the market be characterised as smooth or linear relationship in space? In such a case, price differentials and other indicators do not produce patterns that indicate a segmented market. Alternatively, if the market is idiosyncratic with respect to one or more of its fundamentals (e.g., the house itself, the location in micro or macro terms, land use or other regulation), the differences across locations and housing bundles tend to be qualitative rather than quantitative. In this situation, is it possible to treat urban location in a simple equilibrium, or is it necessary to create another type of tool, based on multiple equilibria? If so, what is the influence of behavioural or institutional circumstances on the formation of submarkets? Indeed, the evidence of the market structure does determine the appropriateness of the model. It is possible, however, that the level of 'examining the market' is also a determining factor, as a more detailed picture of the context inevitably reveals more market disturbances than a more general picture does.

---

## 2.3 Comparing methodologies for analysing the development of submarkets

As already explained, housing-market segmentation refers to the differentiation of housing due to administrative circumstances and the income and preferences of the residents. I will illustrate with a theoretical example. Assume that the households within an urban area represent three income groups, each of which has housing preferences that are distinct from those of the other two groups. Let us then assume that the high and middle-income groups each have two sets of preferences, one of which is common to both income groups. This results in four different demand-side submarkets: three high and middle-income segments and one low-income segment, respectively. Let us now consider additional, institutional features. For example, assume that the local building regulations and the allocation of government subsidies cause further dispersion of the housing market into heavily regulated and other areas, and into subsidised and non-subsidised housing stock. Finally, assume that the regulation criterion differentiates only among the preferences for one of the three high and middle-income segments, whereas the subsidy differentiates among the preferences for the low-income segment. The picture that emerges from this example can be interpreted as a housing market that is partitioned into six submarkets.

Ever since the various behavioural and market mechanisms were outlined by Schnare and Struyk (1976, cited in Leishman 2001) and by Rothenberg and colleagues (1991, ch. 3), they have been a key issue in recent work that has been conducted in several universities and research institutes around the world.<sup>1</sup>

Although the two basic approaches both focus on 'objective' criteria and factors that are measurable on an aggregate level, their views of the segmentation process are diametrically opposed. In the former approach, housing submarkets are assumed to arise due to insufficient competition in the spatial housing market that impedes the equalisation of physical housing attributes. In the latter approach, submarkets represent different price levels of housing that must be adjusted for quality with a hedonic regression model – a standpoint that is more consistent with (neoclassical) economic theory.

Although 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 house-price formation only to the

---

<sup>1</sup> Prominent contributions include works by Adair and colleagues (1996); Maclennan and Tu (1996); Bourassa and colleagues (1997, 1999, 2003); Schwann (1998); Sharkawy and Chotipanich (1998); Morrison and McMurray (1999); Watkins (2001) and Jones and colleagues (1999, 2003).

extent that there is spatial arbitrage. In other words, while the first approach accepts a broader definition of segments, the second recognises market segments only as locations or housing bundles with significantly different quality-adjusted price levels, and the only relevant criterion is therefore the price (or rent) level (e.g., Schwann, 1998). This was also the logic behind the first research question posed in Chapter 1.

The second research question was based on the central argument that submarkets may affect the relationship between location and price. The nature of the relationship between the specific spatially identifiable housing-market characteristics and house prices is unclear. We may distinguish between a single-equilibrium model, in which the relationships between segments according to the price criterion are unambiguously formulated, and a multi-equilibrium model, in which this is not the case, as suburban and city locations may be similarly priced, and the dwellers possessing them may even have the same socio-demographic background. To illustrate with empirical evidence provided by Meen (2001), the London housing market is polarised between wealthy suburbs and a poor inner city. In contrast, evidence from Melbourne shows that wealthy and well-educated households may be accommodated in the city centre as well.<sup>2</sup> Meen (implicitly) applies the price criterion to empirical submarket classification in one urban area, but recognises the possibility of multiple equilibria in another area.

The key issue is therefore which is more relevant: (hedonic) prices or other objective socioeconomic and demographic (henceforth, socio-demographic) or physical partitioning criteria in the two chosen contexts. A third criterion is also plausible, one that is more behavioural and socio-cultural than the ones above and one that requires 'stated' rather than 'revealed' preferences methodology.<sup>3</sup> Such a model would allow the explanation of housing-market structure according to differentiated tastes, lifestyles and similar behavioural factors. In this study, however, such a model is addressed only as a side issue

---

**2** Even in Melbourne, however, the upper-market inner-city segment is relatively small, as Meen later clarified to me.

**3** In the literature, the feasibility and validity of the analysis of actual prices paid and choices made in comparison to analysis of hypothetical prices (property values) and choices (perceptions, preferences and intentions) are frequent subjects of debate. Revealed choice-preference models take one of two forms. In discrete choice models, the dependent variable(s) represent individual choices, and the goal of the estimation (i.e., the betas, part-worth utilities) is to determine the propensities to make these choices. In hedonic models of the housing market, the dependent variable is a proxy for property value, usually transaction price, and the betas constitute shadow prices of each independent characteristic of the regression. The stated choice/preference models, in turn, are comprised of methods in which interviews replace the calculation of market data. For a discussion of the merits and problems of both approaches in a housing-choice context, see Timmermans et al., (1994); and in a valuation context, see Bourassa et al., (2004).

---

in relation to the AHP analyses and their generalisation in the latter parts of the text. Nonetheless, even without such a perspective, qualitative methods may be applied as a support to quantitative methods, in order to add 'flavour' to the analyses that are based on large datasets.

Watkins (2001) concluded that submarkets depend on both structural (house-specific) and spatial (location) criteria. Submarkets may also be driven by demand subgroups, or hedonic quality levels; they may also be manifestations of a non-arbitrage situation. Watkins further argues that the failure of housing economics to account for this relationship is not surprising, because of the complex processes of supply-side and demand-side dynamics that are involved; in other words, how these characteristics influence housing choice and urban form. Jones, Leishman and Watkins (2003) used co-integration analysis on repeat-sales indices to examine a large dataset of Glasgow (over a 14-year period), and found that submarkets are stable through time. This aspect of temporal submarket dynamics necessitates the third research question.

Bourassa, Hoesli and Peng (2003) observe that the classification of submarkets depends on the purpose: the price criterion is suited for mass appraisal, while other criteria are better suited for grouping close substitutes. They added that there is no need for a sophisticated method when determining spatial segments for mass appraisal; established neighbourhood or other urban boundaries are sufficient. They do encourage the use of elaborate statistical methods for identifying submarkets, however, when the aim is to group smaller neighbourhoods into larger ones in order to investigate the structure of cities and the patterns and dynamics of neighbourhoods.

Ley, Tutchener and Cunningham (2002) offered a somewhat vaguer picture of the ways in which house prices may change. They conclude that the analysis of housing-market processes requires moving between different spatial scales. The aim of their own study of the Toronto and Vancouver housing markets was to see how house prices move in response to immigration, polarisation and gentrification. According to their findings, the importance of each of the processes varied in time and across space. Another in-between framework is Weibull's dynamic stock-flow equilibrium, which (as applied by Maclennan & Tu, 1996) advances the theory and methodology by incorporating trade friction and buyer aversion to trade friction into a view of system stability as opposed to instability. The main problem in this framework involves the assumption of uniform preferences among households who, as consumers, belong to the same group with identical income, tastes, knowledge of the market, tolerance for time-consuming search and power positions.

To reiterate, we may postulate that housing markets are segmented according to either price level or other physical or socio-demographic criteria. If submarkets depend solely upon price (or a proxy thereof), the theoretical framework of that context must be based primarily on competition, in which eco-

---

conomic equilibrium and market efficiency are valid assumptions. Otherwise, if the classifications generated by other criteria are superior to those that are generated by price, the assumptions of such a model do not hold. In this report, I compare two European capital cities, Amsterdam and Helsinki, with regard to these aspects.<sup>4</sup>

The fourth research question concerns methodology. The neural-network method is a helpful tool for extracting regularities and developing theory after data analysis. We may seek a posteriori support for a certain theory, given the outcome of the explorations. Information about cross-contextual differences and similarities may allow us to elaborate a theory that disentangles the institutional and behavioural elements. I will return to these aspects in the concluding chapter (7), after presenting the empirical analyses of the two housing markets in Chapters 4 through 6. For the moment, I will depart from this conceptual aspect, turning instead to a more technical discussion in the next chapter. This manoeuvre is necessary to establish a solid platform for carrying out the empirical work.

---

<sup>4</sup> Additional urban housing markets will be included in subsequent analyses. Budapest, Hungary will be included to allow comparison with a presumably entirely different context, in order to emphasise differences from the first two metropolitan areas (see Kauko, 2005).

---



## 3 Empirical analysis of submarkets

### 3.1 Overview of the modelling repertoire

The relevant properties of neural networks (i.e., classification and identification) can be compared to the more common techniques for empirical research on housing-market segmentation that are listed in Table 3.1. The following overview discusses these techniques and presents a few examples of each type of research (see Kauko 2001, for a full review).

The hedonic model of housing markets can be seen as a multidimensional extension of the Alonso-Muth-Mills model. A hedonic regression model cannot actually detect zone boundaries, but the issue can be clarified by using dummy variables. For example in the method proposed by Rothenberg and colleagues (1991, especially 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 according to quality, with regard to characteristics of the house (e.g., number of rooms, age, plumbing facilities, condition and tenure). These classes refer to ranked clusters that are internally substitutable, and can thus be used as a basis for partitioning the total market into submarkets (see also Bourassa et al., 2003).

The partitioning technique is often dictated entirely by practical limitations. According to Maclennan and Tu (1996), however, dwelling units should still be grouped based according to their observable characteristics (including location), rather than in relation to ad hoc aggregation by sector or area. The esti-

**Table 3.1 Summary of empirical research on submarket identification undertaken**

Method	Examples of authors
GIS + descriptive statistics to determine the significance of various factors in the choice of residential environment	Lankinen (1997); several recent studies by AME, URU and other institutions in the Netherlands, for example Deurloo & Musterd (2001)
Hedonic price models	Rothenberg et al., (1991); Leishman (2001); Watkins (2001)
Hedonic models, WTP demand functions for specific socio-demographical groups	Laakso (1997); Bökeman & Feilmayer (1997)
Hierarchy of price groups	Costello (2001)
Repeated-sales in combination with cointegration	Jones et al., (2003)
Projection and clustering, sometimes in combination with hedonic regression analysis	Ball & Kirwan (1977); Bourassa et al., (1997, 1999, 2003); Vaattovaara (1998, 2002); Maury (1997); Ley et al., (2002)
Population surfaces	Martin (1998)
Spatial statistics	Dubin (1992); Dubin et al., (1999); Pavlov (2000)
Non-parametric smoothing and spline functions	Kyllönen & Rätty (2000); Pavlov (2000); Clapp et al., (2002)
Household mobility patterns	Jones et al., (1999,2004); Jones (2002)

mation of hedonic price coefficients is then performed by means of a separate multiple regression analysis for each segment. Next, the mean price of the dwelling is obtained as a function of the input factors for each model. In truly segmented markets, the price equations, and not just the magnitudes of neighbourhood attractiveness, differ across datasets (see e.g., Needham et al., 1998).

Hedonic models build on the principles of 'economic equilibrium' and 'spatial arbitrage', the other basic approach that was outlined in previous chapter, in which submarkets are determined solely by price-related criteria. The 'hierarchy of price groups' approach deployed by Costello (2001) is also in line with such assumptions. The aim of this technique, which is derived from the market-efficiency literature, is to capture price changes for each group, but only for the middle part of the market.

One way of managing the segmentation of data is to 'chain' various statistical methods (e.g., Ball & Kirwan, 1977). First, multidimensional transaction data is summarised into two-dimensional data using factor analysis, which includes such projection methods as principal component analysis (PCA) and multidimensional scaling (MDS). The reduced dimensions are then used as a basis for dividing the data into submarkets using (discriminant, hierarchical or partitional techniques) cluster analysis. Finally, hedonic regression is used to calculate the intrinsic estimation of price for each segment.

Bourassa and colleagues (1997; 1999) used this method with household survey data from Sydney and Melbourne, Australia. They used two datasets: one for local-government areas (43 in Sydney, 56 in Melbourne) and one for individual dwellings. The latter dataset included all of the variables that were contained in the former dataset, in addition to various structural attributes of the dwellings. House values were determined by the owners' estimates of the current value of their residences. As a comparison, five submarkets for each case were determined a priori. Once the hedonic price equations had been estimated for each city as a whole (for both the a priori classifications and the submarkets that had been defined from the data), the weighted mean-square errors were compared to determine the most appropriate classification.

With regard to the results with grouped data, three factors explained more than eighty percent of the variance in the data: (1) distance location (inner/outer city), (2) the socioeconomic factor (indicator of neighbourhood quality) and (3) a residual locational factor (i.e., distance to nearest sub-centre in Sydney; density of persons and dwelling in Melbourne). For the results with individual data, six factors explained more than eighty percent of the variance in the two cities. In addition to variants of the three factors described above, the factors were associated with the age of the dwelling and characteristics of the housing stock. Neither the results that were obtained with the partitional K means nor those obtained with a hierarchical method correspond to the a priori clustering pattern. Not surprisingly, the results of Bourassa and colleagues showed that the performance of all submarket classifications was superior

---

to that of the overall market equation. According to Bourassa and colleagues (1997; 1999), however, the optimal number of submarkets is difficult to determine based on the cluster analysis literature. Laakso (1997), Leishman (2001) and other authors have used principal component analysis as a pre-processing method to overcome problems of multicollinearity in hedonic regression analysis. More recently, Ley and colleagues (2002) employed multivariate analysis using PCA in their study of house prices in Toronto and Vancouver. Their study did not address the prediction of hedonic value.

All of the approaches that are mentioned above fall under the generic classification of the 'partitioning approach'. It is obviously possible to use analysis that is aimed at extracting dimensions from the data and clustering similar observations within the framework of social-area analysis instead of hedonic modelling. Vaattovaara (1998, 2002) and Maury (1997) recently used this technique to analyse the residential areas of Helsinki. Several findings from these studies are used as a reference in Chapter 4.

Spatial regression methods have recently become important in the detection of housing-market segmentation. The idea of kriging utilises the dispersal of residual errors to construct a 'distance-decay' function, which can subsequently 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 and colleagues (1999) emphasise the importance of nearby properties, when the house-price estimate is a function of proximity and degree of spatial dependence (see also Dubin, 1992; Pavlov, 2000, Meen, 2001). This data property is known as spatial autocorrelation. In fact, autocorrelation itself implies market inefficiency (Meen, 2001).

Orford (1999) makes a strong case for multi-level specifications (i.e., property-level, street-level, district-level, community-level) and interaction variables, in order to enhance the efficiency of the value model. He builds several models, first for a more general analysis of the Cardiff housing markets and then for valuation of locational externalities in a part of the city. He focuses on proximity variables that are constructed as interval dummies, based on measured distances to positive locations (e.g., rivers and parks) and negative locations (e.g., heavy industries and railway lines). The model-building process involves the gradual expansion of simple models. The first models operate on aggregate data. From there, micro-level models are formed by adding structural and locational variables and subsequently incorporating structural or spatial drift interactions (i.e., structural or locational variables multiplied with other structural variables to form new independent variables) and multi-level specifications, where each externality effect is measured at an appropriate level.

Although Orford finds the multi-level specification to be more efficient than standard and interaction specifications, he acknowledges a clear problem with the method: the level specification is dependent on administrative

---

boundaries and is therefore not a meaningful measure. For example, multi-level specification does not account for the spillover effect (or similar effects), while interaction specification does. In order to use multi-level specification, levels must be defined according to genuine bottom-up sets (e.g., children within households). This is problematic, as the extent to which this applies to housing markets is not clear, nor is the question of whether locational submarkets are defined within given districts. In other words, it is important for the contextual effect that the partitioning of the dataset is not arbitrary. Because the submarket is an intermediate concept between spatial and non-spatial analysis, this reasoning advances towards an agenda that includes a more explicit spatial dimension.

Flexible (model-free, non-/semi-parametric) regression methods allow for less restricted functional specification in order to enable more adaptable model building that can cope with nonlinearity in the functional form. At the same time, these methods can retain its formality while maintaining their principles of mainstream economic modelling. To give an example of this recently growing trend in research, Kyllönen and Rätty (2000) conducted a hedonic modelling of the housing market(s) of Joensuu, Finland. Their model included a partial spline-function extension which combines both parametric and non-parametric components into an additive model, resulting in a semi-parametric model. Pavlov's (2000) SVC -approach also applies non-parametric smoothing to overcome the specification problem of parametric regression (including spatial regression, as discussed above). The idea underlying this concept is to assign more weight to nearby observations than to observations that are more remote. In their prediction of spatial patterns in house prices, Clapp and colleagues (2002) applied flexibility in the model structure, but within a standard hedonic framework. They argue that capturing the spatial elements is important for hedonic models. They therefore propose a semi-parametric model (local regression) combined with Bayesian inference modelling.

Demand-side segmentation refers to collective preferences based on membership to previously defined ethnic or socio-demographic groups. It is often studied using the specified two-stage procedures of hedonic modelling (e.g., Bökemann & Feilmayr, 1997; Laakso 1997). The hedonic approach provides some help in differentiating the demand side. Maclennan (1982) stressed the role of the hedonic price model as a means of generating demand functions and willingness-to-pay (WTP) estimates for environmental attributes and other housing characteristics. The general form of the demand function is  $W_i = W(NK_i, M_i, A_i)$ , where  $W_i$  is the marginal WTP to pay for the characteristics  $K_i$ ,  $NK_i$  is the amount of  $K$  consumed by the individual,  $M_i$  is income level and  $A_i$  are other demand determinants. In the second stage of the process, the marginal price for each characteristic derived from the price function is equated with the marginal WTP of a household with certain socio-economic and socio-

---

demographic characteristics. This technique is often considered too problematic.

Jones, Leishman and Watkins (1999) explored the relationship between housing submarkets in the Glasgow area, UK, by examining household mobility patterns – an idea originally developed by Grigsby in the 1960s. The analysis was based on data on intra-metropolitan household migration and open-market transactions. Jones (2002) considers the definition of housing-market areas and (related) submarkets according to migration patterns an alternative to the application of statistical tests based on static housing-market outcomes rather than processes (e.g., Bourassa et al., 1997; 1999). Conceptually, this approach is still based on principles of spatial arbitrage and space-access trade-off modelling.

In a follow-up study, Jones and colleagues (2004) make an improvement to the methodology of submarket identification. They use hedonic modelling to examine the Glasgow housing market in such a way that all six a priori submarkets (central, south, south-west, west, north-west, east) are incorporated into migration analysis. The use of only three submarkets (C, S and SW; W and NW; E) would result in the inclusion of links that do not actually exist (in particular, movement between C and S/SW).. In addition, other links that do exist (e.g., movement from W to C) would be missed. Further, a sharp segmentation also exists between new and second-hand homes, flats and rental, as well as between right-to-buy (RTB) and non-RTB. Two conclusions that can be drawn from this information are that (1) there are more than six relevant real submarkets in Glasgow and (2) reducing the number of submarkets to three yields an invalid picture of the market structure with respect to residential location.

The discussion above raises the question of whether an alternative approach to submarket detection is necessary. For example, Watkins (2001) encourages the use of alternative methods, such as space varying regression and submarket tests based on search and migration patterns. Having more options could be expected to improve results, given that submarkets exist within given spatio-temporal contexts. In order for a method or technique to be successful, it must be able to manage the market segmentation on an aggregate level, where the various anomalies that are caused by institutional and physical constraints are discernible. Under these circumstances, it is important to capture outliers that may nevertheless be important in the models. The encouraging findings obtained by Kauko (1997, 2000) indicate that this is possible within an extended model that incorporates the neural network, an emerging 'learning' or 'intelligent' technique. The method chosen for this study is therefore based on neural-network modelling – more specifically, the SOM-LVQ combination – in contrast to the studies that have been reviewed above, which utilise other approaches.

Although econometric methods represent the state of the art, qualita-

---

tive analyses using ethnological research can also prove useful in the search for spatial divisions within the urban housing market (e.g., east/west, north/south). There are also hybrid (qualitative and quantitative) methods that are designed to elicit judgmental information on quality and preferences. For example, an empirical analysis of locational preferences and quality can be performed by using elicitation techniques from multi-criteria decision theory to rank various attributes, using the evaluations of carefully selected respondents as input. The analytic hierarchy process (AHP, Saaty, 1977) is one such technique, which is suitable for many kinds of analyses, including appraisal problems. This method will be used as a supplement to the SOM-based analysis for each context.

The review above reveals one further shortcoming of the current empirical literature: only four of the above-mentioned studies of housing-market segmentation apply a comparative perspective. Of these four, only two represent comparisons between two or more country-contexts.<sup>5</sup> Vaattovaara (2002) shows that the levels of residential differentiation in Helsinki are low, relative to those of six other European cities: Lisbon, Turin, Toulouse, Umeå, London and Dublin. Ley and colleagues (2002) compare the house prices in two Canadian cities. Meen (2001) shows differences between London and Melbourne with regard to housing-market segmentation along socioeconomic dimensions. Bourassa and colleagues (1997, 1999) show that most (but not all) of the main determinants of submarket structures in Sydney and Melbourne are the same.

It is clear that both extensive and intensive methods, as well as combinations of the two, are necessary for conducting this type of research. Comparative research across different urban areas is also crucial. The following is a summary of the main types of methods:

- strict quantitative methods based on parametric hedonic regression (or more generally, economic equilibrium) models – either with or without an explicitly spatial extension;
- flexible (non-/semi-parametric) quantitative methods;
- methods that are based on machine learning; these are ‘intelligent’ methods from the discipline of computer science;
- methods in which interviews and surveys (e.g., multi-criteria decision methods) replace market-data calculation; these methods are usually both quantitative (in the sense that they are mathematical) and qualitative (in the sense that they use judgmental data, including stated-choice and preferences methods).

In this study, the SOM is used to make overall classifications within one local market, and the AHP-elicited expert-preference models<sup>6</sup> are subsequent-

---

<sup>5</sup> Although other comparative (and non-comparative) studies have inevitably been overlooked, the distribution within this relatively broad review is already indicative of the typical genre.

ly used to identify relationships that are more locationally specific. Sections 3.2 through 3.4 explain the methodology used in this study.

### 3.2 Neurocomputing and the SOM-LVQ classifier

The history of applied neural-network research is relatively short, going back to the late 1980s. The basic idea of the artificial neural network can be traced back to the 1940s, when McCulloch and Pitts attempted to simulate human intelligence<sup>7</sup> by studying how the brain functions (cited in Zahedi 1991). It was not until four decades later, however, that computers became capable of handling the requirements of the complex computational processes. This section focuses on the following two aspects of neural networks: how they function and how they are related to the methodology of statistical analysis. The more in-depth discussion is limited to the three neural network algorithms that have been applied within real estate: MLP, SOM and LVQ.

The neural network can be described as a sophisticated statistical method and an estimation method that captures nonlinear, regular associations (i.e., patterns) within a dataset that has no pre-defined model. The basic structural elements of a neural network are called neurons or nodes, the connections between which are determined by weights. Together, neurons process a numerical signal that comes from outside the network in such a way that a connection between input and output information is developed. The connection is referred to as the 'intelligence of the network' when this 'intelligence' is accumulated during a learning process. The iterations in the training process can be based on observed input and expected response values (supervised learning), or on input values alone (unsupervised learning).<sup>8</sup> Figure 3.1 illustrates the principles of the three basic types of network architectures: feed-forward, feedback and competitive networks. The arrows depict connections between the layers of nodes for each type of network. The direction of the calculation process can be input-hidden-output (the feed-forward network), input-output (the competitive network) or unspecified (the feedback process).

The architecture of the feed-forward network consists of an input layer of nodes connected to the observation vector, an output layer of nodes and one or more hidden layers. Like the synapses in the brain, the weights deter-

---

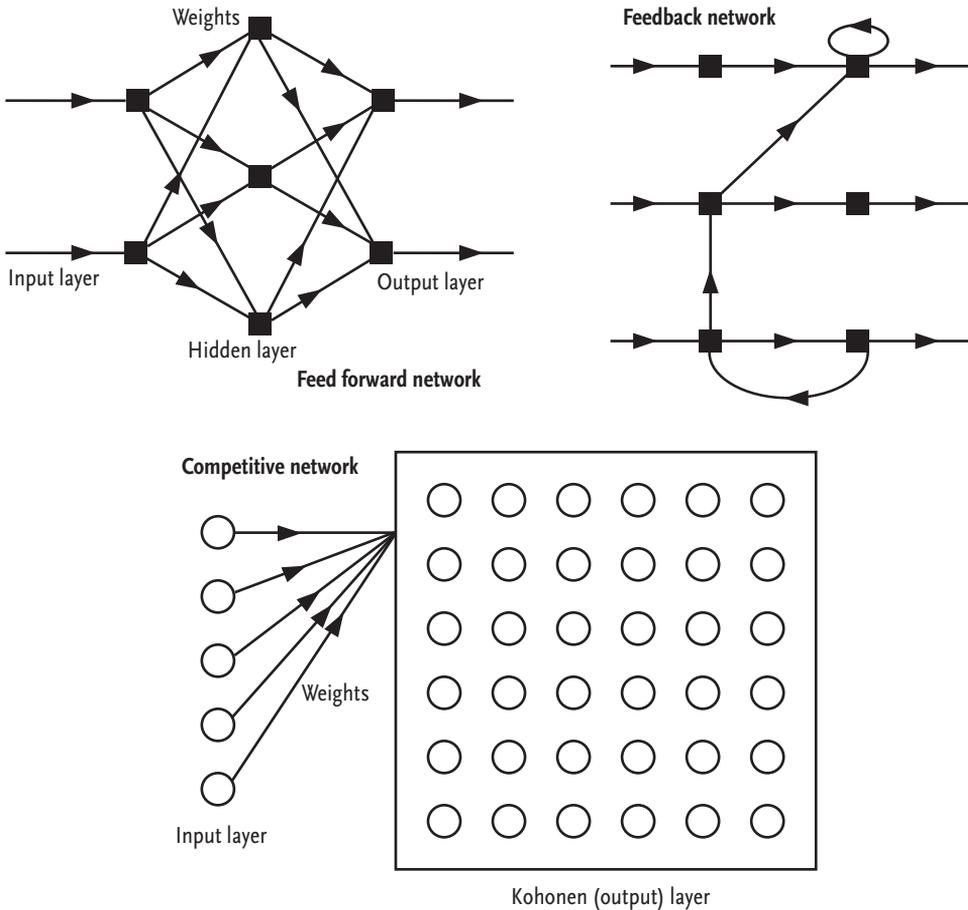
<sup>6</sup> This method is based on the assumption that experts have more detailed and up-to-date knowledge than do households and ordinary buyers. This view broadens the scope of the orthodox model.

<sup>7</sup> The neural network is essentially a stimulus-response technique; it does not reflect real human intelligence (or that of any other species).

<sup>8</sup> There is an analogy to statistics: in particular, supervised learning is equivalent to parameter estimation.

---

Figure 3.1 The principles of the three basic types of network architectures



Source: modified illustration based on Kathmann (1993) and James et al. (1994)

mine the strength of the impulses between the layers. As the training proceeds, the weights are adjusted endogenously until the divergence between the observed output-value and the value that was estimated by the network arrive at the minimum.

The back-propagation-algorithm (developed in 1986 by Rumelhart, Hinton and Williams) is a feed-forward network based on the algorithmic principle of supervised learning. It is by far the most popular neural-network based method. The Multilayer-Perceptron (MLP) feed-forward network is based on the back-propagation of errors, followed by an algorithm to correct the error (i.e., a 'back-propagation algorithm').<sup>9</sup> The error is the actual output, less the output calculated by the network ( $y_k$ ). In a two-layer network, the output can be written as follows:

$$y_k = \sum_j [w_{kj} \sigma(\sum_i w_{ji} x_i) + w_{jH}] \quad (1)$$

$w_{kj}$  = weights between the hidden layer (j) and the output layer (k)

$\sigma(\cdot)$  = nonlinear activation function of the neuron

$w_{ji}$  = weights between the input layer (i) and the hidden layer (j)

$x_i$  = input vector

$w_{jH}$  = a bias-term, where H is a constant

According to experts (e.g., Tay and Ho, 1992), this method is a useful alternative to multiple regression analysis, although it shares many of the same limitations.

The self-organising map (or SOM; Kohonen, 1982) is an unsupervised neural-network technique with a competitive network architecture.

The SOM maps a high-dimensional data space onto a (usually) two-dimensional lattice of points (Kohonen et al., 1996a). This allows disordered information to be profiled into visual patterns, forming a landscape of the phenomenon described by the data set (see Kohonen, 1995).

The SOM produces a feature map of nodes, each of which represents a characteristic combination of attribute levels (see appendices A and B). Some of the patterns may be unanticipated, especially if correlations exist among the most important variables that affect the organisation of the map. Each map layer shows the variation of the dataset with regard to a single input variable, and the position of the nodes is fixed across all layers. From this surface, similar combinations of variables can be considered as a whole and compared with different combinations of variables. It is also possible to interpret a 'typical value' for a given feature of each node. This value can be used as an indicator of goodness.

The first step in using an SOM involves initialising the map by generating random values for each node. The training procedure of the algorithm then proceeds in three stages: (random) selection of a training vector  $x$ ; identification of the best matching neuron (node  $c$ ), which is closest to  $x$ ; and the adjustment of node  $c$  and its neighbours towards the observation  $x$  (e.g., Koikkalainen, 1994). The matching is usually determined by the smallest Euclidean distance (i.e., the distance metrics showing the closeness in an  $n$ -dimensional observation space) between node  $c$  and vector  $x$ , when  $m_i$  represents a parametric reference vector (codebook vector) of every node on map  $i$ . This can be written as follows (e.g., Kohonen et al., 1996a):

$$\|x - m_c\| = \min_i \|x - m_i\| \quad (2)$$

---

**9** Note that, in the US, the MLP (or back-propagation) is used as a synonym for neural network. In Europe, neural networks are defined more broadly.

---

This technique is based on the algorithmic principle of unsupervised competitive learning, which can be described as a 'winner-take-all' situation. The 'winner' is the node that has the shortest distance from the observation vector, and its weights are adapted towards the observation (see Figure 1). The process continues until all observations have been used for training – usually more than once. Neighbouring nodes on the map are similarly adapted towards the observation, but the extent of this depends on the selected parameters.

The SOM has been successfully applied in engineering and many other technical fields in which reduced visual displays of multi-dimensional data are useful (Deboeck & Kohonen, 1998). According to Carlson (1998), the identification of suitable observations and relevant attributes is the most critical point in understanding market behaviour. He argues that the SOM offers a number of general advantages within this problem field:

- self-organisation can be used to create an understanding of typical or less-typical objects in a particular neighbourhood; there is no objective way of specifying the components of property value;
- the SOM enables visualisation and understanding of the market situation; map selection is based on the application;
- the effects of principal components are not destroyed by using the SOM;
- because the SOM can also be applied to portfolio analyses, in which the values are compared, it can also be applied to the analysis of rare components and exceptional cases.

The unsupervised SOM, however, does not provide full verification of the meaning of particular partitioning schemas within the context of study. Determining the success of various clustering solutions generated by the SOM over the entire sample more rigorously than the visual interpretation allows requires two additional techniques. First, the response of the neurons must be compared with the original sample (still unsupervised learning). Second, the correspondence between input and output towards an ideal situation must be improved, minimising the risk of misclassification (supervised learning).

An extension of the SOM, the learning-vector quantisation (LVQ), is based on the algorithmic principle of supervised competitive learning. The LVQ is suitable for testing and improving the classification that is provided by a feature map (Kohonen, 1995), as described above. The LVQ determines the percentage of successful matches between input and corresponding output, when the map (output) is calibrated according to the labelled input. The LVQ requires a small number of meaningful labels to provide a more formal measure of success for the SOM analysis. In this particular application, it can be used to evaluate the most important criteria for segmentation.

The principal idea of this algorithm is to approximate the observations into various classes of the input vector  $x$ , after which  $x$  is assumed to belong

to the same class as the nearest codebook vector  $m_i$ . The classes are determined a priori by assigning a unique label to each observation. The feature map is then calibrated in such a way that each codebook vector receives a corresponding label according to its resemblance (i.e., the closest Euclidean distance) to a certain class of observations. Finally, the accuracy of the classification is determined, preferably using a reserved sample. The performance of the classification is evaluated by recognition accuracy, or the average percent ratio of successful 'hits' over all classes. The following equations define the basic process:

$$\begin{aligned}
 m_c(t+1) &= m_c(t) + \alpha(t) [x(t) - m_c(t)] \\
 \text{if } x \text{ and } m_c \text{ belong to the same class,} \\
 m_c(t+1) &= m_c(t) - \alpha(t) [x(t) - m_c(t)] \\
 \text{if } x \text{ and } m_c \text{ belong to different classes,} \\
 m_i(t+1) &= m_i(t) \text{ for } i \neq c,
 \end{aligned}
 \tag{3}$$

where  $\alpha(t)$  ( $0 < \alpha(t) < 1$ ) may be constant or decrease monotonically with time  $t$ , measured in steps of running (Kohonen et al., 1996b).

The supervised learning properties of the LVQ-algorithm (i.e., the reallocation of the nodes according to an expected match between the labelled observation and response vectors) are used to improve the classification accuracy of the feature maps. The training is conducted with a predefined number of iterations. In the LVQ context, an iteration is formed by making a one-time comparison between the labels of the observation vector of the entire sample and the corresponding response (codebook vectors). The result is then checked for classification accuracy (preferably with an independent sample); if the result is unsatisfactory, the training is continued with a new run of iterations. The appropriate stopping point for the training must be determined by comparing the accuracy results that are obtained with the training to the test samples after each run. Overtraining occurs when the network begins to memorise a training sample instead of learning from it, causing the accuracy of the results to decline. This is the stopping criterion for the training procedure in algorithms that are based on supervised learning (see e.g., Borst, 1995; Worzala et al., 1995).

The quality of the 'organisation' of the feature maps can be determined with either the SOM or the LVQ. The statistic 'Q' denotes the average of the quantisation errors (i.e., the difference between observation vector and codebook vector) over the sample. An alternative measure is the LVQ-classification accuracy (calculated with a set aside sample), given a certain predetermined labelling. In this case, the LVQ is strictly as an unsupervised network.

The following steps are involved in comparing the three algorithms (the SOM, the unsupervised LVQ and the supervised LVQ) with respect to the node-adjustment procedure after confrontation with each new observation:

- SOM: The codebook vectors (of the winner node with neighbourhood) are adapted (i.e., their values are updated) towards the observation vector;
- LVQ/unsupervised: The codebook vectors are not adapted, but simply compared with the observation vector;
- LVQ/supervised: The codebook vectors are adapted towards the observation vector if the classification is correct, and away from it if the classification is incorrect (see eq. 3).

To summarise up to this point, the neural network arrives at results through an iterative process, in which the input is linked to the output, and the linkage is adjusted by weights. The results are strongly dependent on the data, as all necessary guidance to the analyses is obtained from the sample that is fed to the network (and from the network parameters, which are compulsory decisions to make with the standard SOM, as will be explained later). Unfortunately, the lack of a straightforward functional relationship between input and output creates a problem of explicability: the classic 'black-box' argument.

Neuro-computing is sometimes used to answer questions similar questions to those that are addressed by statistics. Neural networks can be seen as both an extension of statistics and as an entirely new paradigm for modelling data. According to White (1989), 'learning methods in neural networks are sophisticated statistical procedures', and 'neural network models provide a novel, elegant, and extremely valuable class of mathematical tools of data analysis' and, eventually, 'statistics and neural network modelling must work together, hand in hand'.

Although there is considerable similarity between neural networks and statistics, and although they tend to generate similar outcomes, there are also a number of differences, which can be considered either drawbacks or benefits. For example, due to the numerical characteristic of the algorithms, a neural network searches only for local optima, causing long running times when trying to obtain an optimal result. On the other hand, these methods do not require certain pre-processing manoeuvres (e.g., the consideration of multicollinearity or the normalisation of data samples). In short, the neural network is characterised by the following properties:

- neural networks require numerical quantitative data (although the outcome may be qualitative);
- neural networks simulate the human mind, using 'black boxes' rather than step-by-step processes;
- neural networks use the terminology of conventional statistics (see section 3.1. above), and they represent non-linear and model-free regression.

Like many econometricians, Schwann (1998b) does not trust neural networks for applications other than forecasting (i.e., not for the classification of sub-

markets), given that they are unrecognisable functional forms. It is common knowledge that, when using neural networks, each new 'neuron model' is influenced to some extent by uncontrollable factors, in addition to the length of the run and the chosen parameters. This creates uncertainty regarding whether variation is caused by the dataset or simply by coincidence. Similarly, in a discussion of data mining, McGreal and colleagues (1998) specify the following drawbacks of neural networks:

- the learning process is slow (not a problem with the capacity of current generation of PC's);
- it is difficult to relate the 'set of numbers' back to the application in a meaningful fashion, due to the 'black-box' nature and difficult interpretation of the learned output;
- performance can be influenced by a range of external factors.

The main methodological findings concern the exploration of the multidimensional complex dataset, the visualisation of patterns and clusters and the classification of potential submarkets according to the findings, with the possibility of improving the models with supervised learning. As mentioned in Chapter 1, the SOM has been applied in the field of property-value modelling (see Lam, 1994; James et al., 1994; Jenkins et al., 1999). In fact, the SOM-based approach can be used as either an alternative or a complement to the partitioning approach, hedonic price modelling and other more mainstream approaches.

The similarities to the combined PCA cluster-analysis hedonic-regression modelling approach (see Table 3.1 on page 21) are notable, as the SOM is more appropriate for detection than it is for estimation. Restricting the application of the SOM is one means of detecting outliers in the data (e.g., James et al., 1994). Differences between the SOM and the k-means classifier become particularly relevant when making analogies to statistical cluster analysis (e.g., Kaski, 1997). The crucial difference is the 'neighbourhood' concept, which is the node that gives the closest response to each observation vector (the 'winner' node) with its adjacent nodes (see e.g., Openshaw et al., 1994).

An additional benefit of the method is its capability to detect submarkets and the idiosyncratic aspect of spatial housing-market structure (Kauko, 1997; 2000; 2002). Furthermore, the inductive approach based on the feature maps that are generated by the SOM may help in the analysis of possible residual aspects of the spatial price-formation structure. The capacity to generate fuzzy, partly qualitative outcomes is an additional advantage over hedonic regression with extensions.

Despite the good qualitative results obtained by Kauko (1997; 2000; 2002), there are a variety of problems regarding the technical presumptions of the analysis that must be considered when using the SOM. The first issue involves how to pre-process the data and, particularly, how to determine the optimal

field-range of a given variable ('scaling', cf. 'assignment of attribute weights', McCluskey & Anand 1999). The second issue concerns the selection of optimal network parameters, which may also have a substantial effect on the outcome (e.g., Kohonen et al., 1996a). Optimal network parameters include the dimensions and the size of the map, the number of steps in training, the initial learning rates and the initial radius of the training. Finally, there are questions concerning the size of the data set and the repeatability of the results. In Kauko's study (1997), the enlargement of the dataset resulted in multidimensional patterns that were more complex, due to the non-parametric nature of the method.<sup>10</sup> On a rough level, however, the clustering prevailed. The remainder of the text focuses on the stationarity of the results over time and space. Although sensitivity, robustness and similar issues are important, this text is not intended to provide a detailed evaluation of the method in this regard.

### 3.3 The specific research design

The brief presentation of the neural network technique above was intended to provide the background necessary to understand the analysis that is reported in this study. It is important to remember that, in the SOM-LVQ classification approach, labels are assigned to each category of observations according to underlying (relative) market characteristics. The label is for recognition purposes (e.g., a label might be a symbol for a particular area in which a particular combination of characteristics is typical). Some segments may be based on criteria other than location. For example, a given area might be divided into building stock from two age categories.

It can be argued that the theory of neural network modelling is not actually a theory at all, except in an open sense, with 'theory' referring only to the process of moving from empirical findings towards generalisation. Assuming that the proposed neural network classification approach is capable of confirming segmentation according to a particular criterion, the identification of appropriate criteria for determining segments becomes a relevant issue. The key question concerns whether submarkets are determined primarily with economic or other criteria. Conclusions about possible similarities or differences between the two contexts are based on the answers to this question. First, transaction-price data from metropolitan Helsinki are analysed with respect to each defined criterion (label) for submarket formation. The criteria for determining the submarkets are independent variables that explain property characteristics, location and other labelling criteria that are chosen in a

---

<sup>10</sup> The results that are generated by non-parametric methods are never exact; the same applies for parametric methods when models have been specified incorrectly.

---

---

more flexible manner. After that the same procedure is applied to the data from Amsterdam.

A number of parameters must be considered when comparing two results that have been obtained with different datasets. These adjustments inevitably induce certain logical expectations for the robustness of the results. First, compared to the use of aggregated data, individual data is expected to generate a 'patchier' feature map, thus providing a more powerful tool for identifying submarkets. Second, enlarging the dataset (possibly because of the previous point: replacing mean values with individual observations) and consequently defining a larger map size generates a more detailed model, creating a better possibility for identifying segments. Third, using panel data from multiple years instead of a single-year cross-section might generate models in which the time trend alone is too important to enable meaningful assessment of submarket classification or other, structural effects of local housing-market dynamics. The validation of all three points (i.e., problems related to the influence of data aggregation, small dataset and map, time trend), however, requires the existence of actual segmentation.

The method for interpreting the outcome of the analysis can now be summarised into a few key stages. First, each observation is assigned a label, and each neuron in the feature map is defined as an n-dimensional codebook vector as a basis for the calibration of the feature map (e.g., Kohonen et al., 1996b). Classification accuracy statistics are then calculated for the successful 'hits' between observations and the feature map, according to this class. Repeating this procedure for each labelling solution can determine which factors contribute to high classification accuracy and are therefore relevant for the observed segmentation. Finally, a loosely formulated theory combined with additional knowledge of the local market context is necessary to guide the analyses. It may therefore be worthwhile to perform the analysis with similar types of input in two or more different contexts, with the goal of extracting a new, institutionally sensitive theory. As noted above, however, this would require controlling for the effects of aggregation, enlargement and the temporality of the dataset.

The approach to modelling spatial housing-market structure (submarket classification in particular) that has been elaborated in this chapter contains two general points:

1. All available data sources and previous studies may be used to obtain information on the diversified (multiple equilibria), fuzzy (based on images and bundles of intangible concepts) and truly nonlinear (e.g., not just log-linear) phenomenon of spatial housing-market dynamics (i.e., how different locations and housing bundles within one urban area differ from each other, both quantitatively and qualitatively).
  2. Because all locations are different by definition, location always has a residual influence on the structure and dynamics of the housing market (regard-
-

less of how well supply and demand factors can be approximated). This idiosyncratic element has not been fully utilised in the conventional intra-urban location-modelling literature. A bottom-up approach is consequently more efficient than a top-down approach, and induction is a more valid research strategy than deduction is. It is therefore useful to build up the theoretical framework by generalising from particular cases (instead of specifying an a priori theoretical model) and focusing on average market behaviour, as in the more commonplace approach.

### 3.4 Supporting the analysis with expert interviews, using the AHP

To supplement the quantitative analysis, several interviews were conducted with experts from each context. The results of these interviews are presented at the end of Chapters 4 and 5. Before proceeding, a brief introduction to this approach is necessary.

The analytic hierarchy process (AHP, Saaty 1977) technique is based on the pair-wise preference comparison of elements (attributes or alternatives), and results in a comparison matrix in which the relative importance of each element is determined as a ratio between 0 and 1. This technique is suitable for many kinds of analyses, including appraisal problems. The basic principle of the method is given below.

In sharp contrast to the classical multi-attribute value-tree modelling approach, which is based on the assumption that utility functions can be explained, the AHP does not assume that the evaluator is able to express the overall elicitation of the problem as a single function. Instead, the AHP is based on the assumption that the relevant dominance of one attribute over another can be measured with a systematic, pair-wise comparison of preferences at each level of a hierarchy of factors, presented as a value tree (e.g., Ball & Srinivasan, 1994). The overall objective of the decision stands at the top of the hierarchy, with lower-level objectives or attributes at the lower levels (e.g., Zahedi, 1986).

The comparison begins at the lowest level of the tree, where the elements (attributes or alternatives) are usually elicited with an ordinal scale from 1 to 9, with the values corresponding to verbal expressions. A value of 1 means that 'both are of equal importance', and a value of 9 means that 'A has an extreme importance over B'. The comparisons are then converted into cardinal rankings (e.g., Erkut and Moran, 1991). Balancing the pair-wise ranks in this way involves the use of measurement theory, as pair-wise judgments cannot be assumed consistent across the entire set of comparisons (e.g., Ball and Srinivasan, 1994).

Following Saaty (e.g., 1990), the functioning of the AHP technique is

explained in terms of a matrix equation. Consider the elements:  $A_1, A_2, \dots, A_n$  within one level of the tree hierarchy. In practice, the maximum number of elements to compare within a single comparison matrix is nine (the 'Expert Choice' software actually has a maximum of seven elements), although there is theoretically no upper limit to the number of elements to compare. The comparisons among all of the elements ( $A_1:A_2, \dots, A_{n-1}:A_n$ ) then generate the following matrix:

$$A = \begin{matrix} A_1 \\ A_2 \\ A_3 \\ \cdot \\ \cdot \\ \cdot \\ A_n \end{matrix} \begin{bmatrix} w_1/w_1 & w_1/w_2 & w_1/w_3 & \dots & w_1/w_n \\ w_2/w_1 & w_2/w_2 & w_2/w_3 & \dots & w_2/w_n \\ w_3/w_1 & w_3/w_2 & w_3/w_3 & \dots & w_3/w_n \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ w_n/w_1 & w_n/w_2 & w_n/w_3 & \dots & w_n/w_n \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \cdot \\ \cdot \\ \cdot \\ w_n \end{bmatrix} = n \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \cdot \\ \cdot \\ \cdot \\ w_n \end{bmatrix} \quad (4)$$

The total number of comparisons is  $(A_{n-1} \times A_n)/2$ . For example, a matrix of four elements generates six comparisons. Each comparison generates a pair-wise ratio, (e.g.,  $w_1/w_2, w_2/w_1$ ). All of the ratios along the diagonal are obviously equal to 1, as it is not necessary to compare elements with themselves. The overall weight is indicated by the priority vector.

The most common way to estimate the relative weights from the matrix of pair-wise comparisons is the 'eigenvalue' method (see e.g., Zahedi, 1986, for a full discussion).

The matrix formula  $A_w = n_w$  applies only for the theoretical ideal situation in which the comparison is fully consistent. This is usually not the case in observed pair-wise comparisons (unless the comparison is unambiguous and the matrix is very small, e.g., 3 elements that compare 2:1, 2:1 and 4:1), and the estimate  $\lambda_{\max}$  is therefore used instead of the exact  $n$ . To enable approximation of a less than fully consistent comparison matrix, there must be more observations than weights. In fact, as Saaty (1990) demonstrated,  $\lambda_{\max}$  is always greater than or equal to  $n$  and, as it approaches  $n$ , the values of  $A$  become more consistent. In the terminology of AHP, this property has led to construction of the consistency index (CI) as follows:

$$CI = (\lambda_{\max} - n) / (n - 1), \quad (5)$$

The consistency of the comparisons is measured with the consistency ratio (CR), which is calculated according to the expected results of consistent pair-wise comparisons across the matrix, as follows:

$$CR = (CI/ACI) \times 100, \quad (6)$$

The ACI is the average index of randomly generated weights (Cited in Zahedi, 1986). Using analogous terminology from statistics, substituting  $\lambda_{\max}$  for  $n$  generates a number of equations that exceeds the number of unknown parameters to be estimated. The CR should be very small. There are several opinions about the relevance of the CR; for example, it may be used as a filter. This measure is disregarded in the exercise that is reported below.

Finally, local weights are transformed into global weights. The most attractive choice is determined by aggregating the local priorities into global priorities (i.e., 'quality-ranks' or Q-values). This process quantifies the relative contribution of each element in the value tree to the overall goal.

Using the analytic hierarchy process (AHP), quality ranks were generated for various bundles of locational attributes, using interactive data. In this exercise, the respondents were required to meet two criteria: (1) a pursuit as stakeholder, based on professional responsibility in business or administration and (2) a deep local knowledge of the spatial housing-market structure, gained through professional experience. The experts represented transaction-related services (e.g., estate agents and assessors), land and property ownership (e.g., builders, municipalities as landowners and other investors) and user-oriented interest groups (e.g., planners, rental agents and other administrators). There is no fundamental reason either for or against adapting the method by including the owners and renters of housing as experts. In contrast to the better-informed professional expert groups, these informants are likely to have somewhat less variation in the attributes determining their location choices or property-appraisal decisions, as households tend to have much less information at their disposal than do professionals in the field.

The 'behavioural paradigm' in residential valuation, which is propagated by Daly and colleagues (2003), places more emphasis on demand or consumer-driven factors that relate to preferences and intangible components of quality. Further, it evaluates the performance of a given method with regard to these aspects. This approach offers both a contrasting alternative and a supplement to the main approach (as in the case of this study). Because of problems that are associated with listed-price data (e.g., scarcity, unreliability and low quality), conventional methods do not apply. Arguably, the inclusion of consumer behaviour and quality in the method can improve its conceptual soundness. Multi-criteria decision-making analysis is therefore the most conceptually sound approach to valuation, as it explicitly deals with such elements. Particular goods, including housing, may have fashionable symbolic meanings (sign value), and demographically similar groups may have fundamentally different ways of life (Bourdieu). Scarcity value is one obvious aspect of this. Adding the conceptualisation of the ways in which particular products (in this case, residential areas and housing packages) become fashionable (Beck) and attractive targets for trendsetters the argument moves toward the discussion on immaterial sign values in consumer sociology (Cited in Uuskallio, 2001). As recent-

---

ly noted, ‘new life-styles have developed that emphasise “hedonistic individualism” and that are characterised by patterns of consumption that emphasise symbolic values in conjunction with articulated life-styles’ (Mingione & Scott, cited in Kloosterman & Lambregts, 2001). The same consumer-oriented processes of attributing symbolic and sentimental value to the home also occur in less-developed societies such as Turkey (Tekkaya, 2002).

### 3.5 The comparative perspective and the institutional and behavioural aspects

In social science, the principle of comparative research advises going beyond merely describing sets of countries, regions or areas, towards the generalisation of contextual effects that are related to given socio-economic structures and processes. Within the field of property-valuation methodology, Daly and colleagues (2003) conducted an excellent piece of comparative research using interviews of housing consumers and residential valuers. They noted that problems (e.g., scarcity, unreliability and low quality) are often associated with listed-price data. Because of the limitations of approaches that are based on market-outcome data, they propose a ‘behavioural paradigm’ in residential valuation, which puts more emphasis on demand or consumer-driven factors that are related to the buyer’s perception of the market situation. Furthermore, comparative research is the most appropriate tool for gathering knowledge of differences in institutional structures. The approach that is used by Daly and colleagues is interdisciplinary, drawing on economics, sociology and psychology. In addition, they apply a qualitative perspective, and design their research to allow the identification of the attitudes and behaviour of homeowners with regard to the key attributes that influence value in cross-cultural settings.

Using this approach, however, makes it necessary to overcome two difficulties of comparative methodology, as Daly and colleagues implicitly show. The first problem involves how to link the influences of ‘globally-oriented economic functions’ to micro-based behavioural theoretical perspectives. The second concerns what must be done to ensure the comparability of contexts. According to Daly and colleagues, the ‘categories targeted’ should be ‘broadly similar to ensure that the samples targeted are likely to demonstrate the same characteristics and are representative of comparable populations’. It is difficult (if not impossible), however, to make two or more sets of data exactly similar, due to fundamental differences between contexts (e.g., attitudes, procedures, products). As the study by Daly and colleagues demonstrates, the addition of a comparative dimension to a study that is concerned with behavioural processes is nonetheless beneficial.

Thus a behavioural valuation study with a comparative dimension sketched

---

above enables qualitative generalisations to be made. The qualitative variation in value across locations and localities is assumed attributable to two factors: institutions or the role of individual experience in preference formation. An institution can either mediate the formation of preferences or directly mediate the market outcome. Beck (1993, pp. 130-313) refers to institutionally (e.g., market, law, welfare, education, fashions) dependent individual situations'.<sup>11</sup> From the experience-based perspective, Gram-Hanssen and Bech Danielsen (2002) conclude that life-styles, which are either symbolic or functional, are important determinants of housing consumption. The key issue is the relationship between home and everyday life. Preferences can include features of physical location, status, or the external or internal design of the house. Identity is obviously a difficult concept to validate. The reason for this preoccupation with institutions in a seemingly market-based context is that formal and informal institutions have explanatory power when neither pure rationality (economic model) nor individual experience (psychological model) is sufficient to explain how the results represent more differentiated patterns of market behaviour (see e.g., Ball, 1998 for a review).

---

<sup>11</sup> According to Guy and Henneberry (2000), institutions do more than merely reshape the markets, as investment actors actively engage in a constitutive role. The notion of an 'active' institution actually underpins the classical (or 'old') institutionalist idea of 'value creation' (i.e., the Kaleckian model of pricing behaviour, 1954).

---

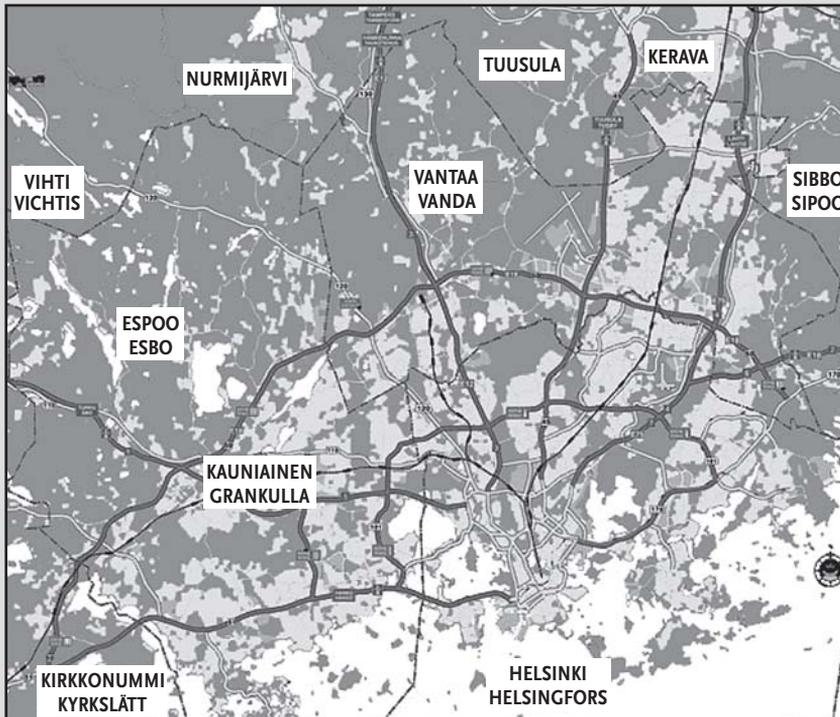
# 4 Results of the submarket classifications in Helsinki

## 4.1 Study area and data

The Helsinki metropolitan area consists of four municipalities: Helsinki, Espoo, Vantaa and Kauniainen. The population of the entire area is approximately 950,000, sixty percent of which reside in the city of Helsinki. The Helsinki metropolitan area is the central part of the greater Helsinki region, which is by far the largest agglomeration in the country, accounting for approximately one fifth of the Finnish population (see Figure 4.1).

There are 400,000 dwellings in Helsinki, approximately sixty percent of which are owner-occupied. In 1993, the Finnish housing market was still in a recession. At the time of this writing, a decade later, price levels were substantially higher in most places. Nevertheless, price levels reflect structural differences that depend on the differences in attractiveness between areas,

Figure 4.1 Helsinki metropolitan area



OBS: The data and study area only cover Helsinki, Vantaa, Espoo and Kauniainen.

Source: modified illustration based on <http://www.nls.fi/ptk/wwwurl/localmap.html>

most of which prevail independently from cyclical price fluctuations. The literature sometimes makes a distinction between local relative price components and those that are based on market trends (see e.g., Bramley, 1999). This implies that it is possible to treat market price as an outcome of fundamentals: relative price differentials across an area that remain constant in time, and cycles and bubbles that occur through time due to market disequilibria and that comprise an extra element in addition to the fundamental property value.

According to Doling (1990), there was formerly 'relatively little price variation within any one urban area' in Finland; today the variation is larger. Prices are high in the city centre of Helsinki (usually understood as the neighbourhoods of Kamppi and Kluuvi), the suburbs on the city's western coastline (e.g., Lauttasaari) and southern Espoo (areas west of Helsinki, e.g., Westend), as well as several suburbs on the city's eastern coast (e.g., Kulosaari), the tiny municipality of Kauniainen and some of the low-density areas in northern Helsinki (e.g., Pakila). Some other areas are comprised of inexpensive housing that was built in large housing estates in the late 1960s and 1970s (e.g., Jakomäki in Helsinki and Myyrmäki in Vantaa).

In many cases, the population base of the less-expensive areas tends to be underprivileged. For example, Kortteinen and Vaattovaara (1999) conclude that there is a clear trend towards the spatial concentration of pockets of poverty in the northern and eastern parts of Helsinki. Such trends are deemed undesirable in the segregation literature. On the other hand, the majority of the building land within the city of Helsinki is municipally owned, enabling the city to use pro-active measures to prevent any form of segregation.

The price premium at a given location can be defined as the positive difference in actual price per square metre, minus the average price per square metre. Consideration of the price per square metre instead of the total transaction price generates relative price premiums for central locations (i.e., the areas close to the CBD of Helsinki). Many of the outer neighbourhoods that are comprised of spacious and expensive dwellings show relatively modest price premiums. The land-price gradient thus plays a substantial role in explaining house prices (per square metre) in Helsinki.<sup>12</sup> In addition, some locational differences are qualitative and discontinuous, even more so than gradational changes in consumer purchasing power and distance decay. The most important findings from recent urban housing-modelling studies on Helsinki are compared below.

Laakso (1997) initially partitioned a dataset from Helsinki into two segments, according to whether travel time to the CBD was more or less than twen-

---

<sup>12</sup> In this study, however, the input variables are direct measures of the physical and social compositions of different locations, and only possible indirect measures of the accessibility factor.

---

A 1990s suburban housing estate in Helsinki.

ty minutes. These segments were subsequently partitioned into separate segments of multi-storey and single-family housing. The hedonic price models that were estimated for each of the partial datasets differed significantly. He also discovered that, in addition to the house type, dwelling size is also an important criterion for segmentation.

Maury's (1997) classic factor-ecology approach (developed by Frank Sweetser in 1960), which is based on area-level information, age and house type and supplemented by demand-side data (26 socio-demographic variables), revealed four basic clusters: (1) old Helsinki (i.e., the inner city); (2) early multi-storey suburbs; mostly affluent neighbourhoods, partly in the inner city (e.g., Lauttasaari and Munkkiniemi), as well as several deprived areas; (3) multi-storey suburbs that were built in the 1960s and 1970s; and (4) single-family housing areas. These clusters were considered natural and meaningful segments, considering the data that were used.

Lankinen (1997) emphasised proximity when moving to certain areas, some of which are inexpensive and others expensive. Although differences between areas decreased during the market upswing of the 1980s, a dispersal occurred in the early 1990s. According to Lankinen's findings, distance to CBD combined with income and status to explain between sixty and seventy percent of the price dispersal in the Helsinki housing market. This comprises a mosaic pattern: the CBD has a regular distance decay association with price; some income and status differences between sectors and blocks have irregular associations as well.

Vaattovaara (1998) segmented the residential areas of Helsinki into four groups according to type of neighbourhood. Three of these groups represent owner-occupied housing using a factor-ecological approach together with GIS: (1) families with children (1/3 of the households) in suburban multi-stor-



ey apartments; (2) urban residents in inner-city locations; some areas (e.g., Kallio) have an identity of their own; (3) wealthy residents and spacious living in suburban single-family housing. In a subsequent study, Vaattovaara (2002) demonstrated the presence of segmentation – and even segregation – on some level in Helsinki. Spatially identifiable socio-demographic data from 1980 to 1994, however, show that, while spatial polarisation has occurred, social polarisation has not. Although the information sector is at the forefront of the economic upswing, there is no simultaneous growth in poverty. On the contrary, the widening spatial differences (notably, the polarisation between the educated prosperous western part and the less-advantaged eastern and northeastern part) have occurred because different areas have developed at different paces. The ‘east’ had already been in a poor position in the 1980s; it was most strongly affected by the recession, and it experienced the slowest recovery during the upswing of the late 1990s. Vaattovaara notes, however, that the differences are small in an international comparison, due to the strong combined effect of the Nordic welfare state and the above-mentioned strong, anti-market city-planning apparatus that also managed to cope well with the recession of the early 1990s recession. Further, image plays a crucial role in this context: areas that people regard as unattractive tend to generate low prices and rents, attract impoverished migrants and become dead-ends for those who remain in the area, making it even less attractive.

These studies reach a consensus about at least three distinctive housing submarkets: (1) the inner part of Helsinki (if necessary, this may be split into two further submarkets according to micro-location status), (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 observation confirms this segmentation. The use of a dynamic perspective introduces a fourth segmentation tendency: the ongoing divide between the eastern and western suburbs. Note that there are several relevant determinants of submarkets: determinants that are primarily physical, socio-demographic and related to price; and determinants that relate to behaviour and institutions (cf. Kortteinen & Vaattovaara, 1999). I attempt to consider all of these aspects in the empirical analyses that are documented in the remainder of this chapter, using the SOM, the LVQ and the AHP.

The exploration of house prices in Helsinki is based on a full cross-section of 18592 condominium (i.e., securitised dwelling) transactions<sup>13</sup> during one

---

**13** The data do not contain transactions for real property (i.e., landed property, including the majority of detached houses), however, which are fewer in number and maintained in a separate system (by the National Land Survey). This dualism between securitised housing and residential property is a curiosity of the Finnish cadastral system. Because these data are based on the former source, detached houses are relatively rare in comparison to semi-detached houses.

---

**Table 4.1 Description of the variables in the Helsinki dataset**

Micro-level variables	
1	Price of the dwelling per square metre (1000 FIM)
2	Age of the building (10 years)
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 subarea / 10
6	Number of public services in the subarea
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 somewhat stigmatised 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

year (1993) in the Helsinki metropolitan area, with locational attributes aggregated at the statistical sub-area level for each transaction. The raw data were obtained from Statistics Finland. Four house-specific and six locational variables are included in the dataset (see Table 4.1).

Three of the locational variables reflect the availability of commercial and public services, and the share of open space in the surrounding area. The remaining three locational variables were constructed by principle component analysis, a common way of reducing the dimensions of the original data according to the loadings of each principal component.<sup>14</sup> Two of the variables can be classified as institutional: 'public services' captures the effects of governmental investments, and 'open space' captures the effects of land-use planning and land ownership. Note that two of the locational variables (8, 9) are predominantly socio-demographic and are constructed according to indicators for education, income, unemployment and share of foreigners, whereas the remaining variables concern the physical environment (7, 10) and the availability of services (5, 6). Further, some of the variables (notably, the urbanisation indicator) are proxies for CBD distance. This is justified by the fact that Helsinki is a strongly mono-centric urban area, as confirmed by Laakso (1997). This assumption, however, does not hold for all urban areas. For example, Amsterdam is not mono-centric as the main employment centres are elsewhere (see next chapter<sup>15</sup>).

For the visual interpretation and classification using the LVQ, location and other labelling options were added to the observations in both datasets. The labels did not affect the calculations of the SOM algorithm. Finally, the total

<sup>14</sup> The variables were prepared by Seppo Laakso.

<sup>15</sup> There are hints that a similar situation is emerging in Helsinki as well. Due to the ambiguous nature of the evidence, however, this point will not be elaborated further. New results on spatial house-price formation and land gradient patterns in contemporary Helsinki are needed.

dataset was divided into a training set and two separate validation sets for the LVQ testing and for evaluating the reliability of the feature maps.

## 4.2 Analysis with the SOM

Based on the dataset, the SOM produces<sup>16</sup> a feature map of several characteristic combinations of attributes. It is possible to interpret a 'typical value' for each node according to a given feature, as an indicator for the given combination of attributes that may indicate a particular neighbourhood. To enable visual examination of the feature maps, differences in this value estimate across the map are depicted with grey shading; lighter shades represent the higher values. For example, when examining the map layers for price and age, light colours indicate areas with high price per square metre and old buildings, whereas dark colours indicate areas with low price per square metre and new buildings. This makes it possible to see at a glance which areas have a high per-unit price level, which areas have an old building stock and the extent to which these two layers overlap. In other words, it shows whether there are any associations between these two factors. For example, the feature-map shows that old buildings contribute to the segmentation of the data set, and that these areas belong to the more expensive cases. A similar visual analysis can be made for all of the input variables. For a more quantitative analysis, the 'typical values' of the nodes can be post-processed, using another computational technique.

The resulting feature map is shown by layers in Appendix A. The labels are based on sub-districts, and they indicate location within metropolitan Helsinki; they are used to calibration of the feature map. In this way, a particular locational label becomes a symbol for a particular combination of variables and submarket structure. The following results were interpreted from the map:

- single-family housing forms two separate homogeneous clusters: (1) a larger group comprised of areas of mixed nature from all three main municipalities (the darker neurons on the lower right of the map), suggesting a physically homogeneous space across municipal boundaries and (2) another, much smaller group in southern Espoo (the darker neurons on the lower left side of the map);
- the most expensive (per square metre) areas are the most urbanised, with the least open space and the best commercial services (inner Helsinki neurons in the upper and lower left corners of the map);

---

<sup>16</sup> The software packages used in the study are SOM\_PAK and LVQ\_PAK, in MS/DOS, produced by the Laboratory of Computer and Information Science at the Helsinki University of Technology.

---

- the least expensive areas are positioned along the upper middle and right corner, most notably in the outlying neighbourhood of Jakomäki (914142), which is a symbol of poverty and social externalities;
- the newest building stock is positioned on the right and the oldest on the left, forming two clearly distinct submarkets;
- four or five clusters indicate submarkets with larger dwellings (i.e., three or more rooms);
- low-status areas overlap with areas that have many social externalities and vice versa; high status areas overlap with areas that have few externalities;
- shoreline proximity brings a clear price premium for the neurons on the left of the map; it also indicates the Espoo high-status areas in the lower corners of the map (the feature can be seen from the 'open space' indicator, because being surrounded by water automatically means the absence of undeveloped land)<sup>17</sup>;
- the average, less interesting cases are situated in the centre of the map (but they may still indicate market segmentation).

The main idea is that the visual SOM analysis generates the same three or four submarkets that could be expected according to a priori knowledge of the Helsinki housing market<sup>18</sup>:

- (1) locations in the inner city and the nearest old suburbs
  - (1a) absolute top location; high-price areas
  - (1b) older, low-status working-class areas; still relatively high prices
- (2) other locations (e.g., multi-storey housing, low status, low price)
- (3) detached and terraced housing; also low-density, multi-storey housing.

---

**17** Note that the 'urbanisation' indicator and the 'open space' indicator overlap with regard to central Helsinki neurons on the left side of the map, but not with regard to the suburbs that are positioned in the lower right corner of the map. In the latter case, proximity to the seashore is reflected only in the open-space indicator. In the former case, it is reflected in both indicators, as the city centre of Helsinki is surrounded by water and urban by definition. To give an impression of the magnitude in price premiums: shoreline proximity brings a clear price premium for two segments defined by location and house type. It was thus possible to select two combinations of housing-attribute levels, approximated as two locations that are of same magnitude in all other characteristics that are used as input (i.e., house type, age, rooms, price level, status, social externalities, services and distance to CBD), but that differ with regard to water. One location on Tammisalo (an island) and another in Hakuninmaa-Maununneva (in northwestern Helsinki) are as far away as possible from any water. It was then possible to calculate the difference between these two areas in terms of price per square metre. The premium in favour of the island location was FIM 731/ FIM 6026 = 12%. (This is substantially lower than the result obtained by Laakso, which were based direct or immediate vicinity of the coast.

**18** For a more detailed elaboration including a comparison with a k-means analysis on the same data, see Kauko (2002).

---

We cannot conclude, however, that these segments overlap with any definite spatial boundaries. For example, the western part of Helsinki alone comprises areas from all four segments. The neighbourhoods of Lauttasaari and Munkkiniemi belong to segment (1a), even though they are not situated on the peninsula. Meilahti and Ruskeasuo are represented in both segments (1a) and (1b). Konala clearly belongs to segment (2), while Munkkivuori may be classified under segment (3). Nevertheless, the identification labels based on sub-district can serve as proxies for a dominant combination of observed-attribute levels, and can therefore help to structure the housing-market data, first visually and then more formally with the LVQ.

### 4.3 Analysis with the LVQ

Following the method that was outlined in Section 3.2, the meaning of the clustering pattern was tested using two independent samples. Efforts were then made to improve the classification accuracy, using the LVQ-network. Each observation was identified with a label according to its attributes. Following a calibration procedure, labels were also assigned to the output (i.e., feature maps), in order to allow comparison between the labelled data and the corresponding neurons on the feature map, each of which represented a particular category of observations. Classification accuracy refers to the percentage of successful matches, when the observations are classified according to their codebook vectors. It is important to be aware of poor classifying accuracy, due to ambiguities associated with the labelling of the samples. For example, if all of the expensive observations are old (and thus labelled as 'old'), it may be difficult to assign the correct label to an independent observation that is both expensive and new.

Table 4.2 shows a comparison of a variety of classification criteria according to recognition accuracy, while holding the sample, map size and network parameters constant across the runs. In theory, having fewer labels eases the task for the algorithm and improves the expected classification result. When only a few labels are used, the captured market segments tend to be logical and coherent.

The high levels of classification accuracy imply segmentation within the dataset. Somewhat surprisingly, the best classification result is obtained with the dichotomous 'open-space' indicator. As expected, the urbanisation indicator, which proxies CBD accessibility, generates good results. Also as expected, house type does matter. 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/design), and it is very important. Both types of services are important, as is macro-location; either municipality (Helsinki, Vantaa or Espoo) or a more specific grouping based on the combined effect of age, price

**Table 4.2 LVQ-classification of Helsinki market segments 1993: verification of the segmentation based on classification accuracy results for various labelling criteria<sup>1)</sup>**

Number 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 2 km range: 0-4.99 km <sup>2</sup> /5.00+ km <sup>2</sup>	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 years / 50+ years	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	Number of public services in the area: 0-49/50+	94.95% (95.35%)	-
2 commercial services	Number 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: two 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 years/25-74 years/75+ years	81.27% (83.61%)	Improvement => 91.67%, but 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/+ 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

1) Some comments related to Table 4.2: (1) The map size is 12x8. A bigger map (e.g., 24x16), would give a better classification accuracy. (2) In general: defining more classes gives a lower classification accuracy, if the criterion is the same. (3) The classification is based either on an a priori chosen criterion or an a posteriori chosen clustering. In general, labels chosen based on the a posteriori clustering gives a better accuracy than the a priori chosen labels. (4) A dichotomous classification, where the number of observations per class in one class is large (tenfold) compared to the other was excluded, even if the result was superior.

and number of rooms, tends to segment certain suburbs that are located far away from the city centre. These clusters represent two groups of dwellings in northern, eastern and northwestern Helsinki, Espoo and Vantaa that have poor services, low density and that are located relatively far from the centre of Helsinki. These neighbourhoods are based on a posteriori clustering. The 'typical dwellings' in these areas have an average or low price per square metre, are situated in average or new building stocks and have three or more rooms.

The conclusion is that a segmentation based on house type, location and other factors (age?) seems more appropriate for 'objective' submarket identification in Helsinki than is a segmentation based on price levels (i.e., either price per square metre or total price). This conclusion is strengthened by a comparison of the partitioning based on the SOM to chi-squared automatic interaction detection (CHAID) – a technique of variance analysis – using the same dataset. According to this run, the most important factor affecting price segmentation is the number of rooms, which is a proxy for size, as categorised into four classes (see lower in the table). Of the ten variables used, 'number of rooms' may be the best proxy for segmentation by price level, although the use of four size categories does not segment the dataset with even fifty-percent classification accuracy, which is substantially lower than the levels that are reached by other criteria.

In addition, supervised training was quite successful for most labelling solutions. The accuracy percentages were improved in most cases, but clear over-training did not occur. This may be due to the fact that accuracy levels were already high. (Over-training occurs when the ANN begins to memorise the sample patterns instead of learning them. Over-training can be detected by using an individual sample. Unfortunately, this procedure is not a feasible with small datasets.) The classification accuracy of the two best classification results for the unsupervised map was already above ninety-five percent. It is unlikely, therefore, that supervised training would have had any added value.

#### **4.4 Modelling the spatial dynamics of the Helsinki housing market according to a comparison with a later cross-section**

The analyses continue with the application of the SOM and LVQ to a new dataset. When evaluating the importance of physical supply-side variables in relation to social demand-side variables, it is interesting to conduct new analyses with a cross-section from 2001, comparing the results with those of 1993. In Helsinki, considerable large-scale development took place along the coast (notably, in Vuosaari, Herttoniemi and Ruoholahti) during the late 1990s, which may have generated residential patterns that differ from those that are

---

reported here. One problem with ex-post research designs is that they can address only changes in the demand factors over a decade. The prices of similar properties and locations are therefore expected to be different at another time point, primarily due to the instability of demand factors.

A few causal statements can be made based on the results of the SOM and LVQ explorations that are reported above. As the analysis showed, the Helsinki housing market is (or, at least in 1993, was) segmented primarily by a locational criterion (i.e., proximity to the CBD). Particularly in Helsinki, the relationship between building age, house type and price per square metre is readily observable; markets are polarised into A and B classes. In single-family housing areas and in old neighbourhoods that are close to the centre, prices are high, whereas they are low in suburbs, which are dominated by multi-storey construction.

Indirectly, location in the city core reflects the premium for locations in old, monumental buildings, possibly with a seaside view. In Chapter 1, I argued that neural-network modelling should be supported by local expertise if it is to have any meaningful relevance beyond 'number crunching'. Only then can speculations be regarding any causality that might be attributable to the results. It is known that the roots of prestige for the older, central areas lie in the 19th and early 20th century. These areas on the Helsinki peninsula were occupied by the aristocracy and the bourgeoisie, and they are sharply segregated from the working-class areas on the 'wrong side of the long bridge' (*pitkä silta*, referring to the approximately fifty-metre long bridge that separates the peninsula from the northern and eastern inner-city neighbourhoods. The bridge connects the Kluuvi and Kallio neighbourhoods.)

Repeating the analyses in a different market situation will generate different results. A specific question in this setting concerns the impact of the Finnish recession in the 1990s on segmentation in the Helsinki housing market. It is conceivable that different submarkets based on geographic location, form of occupancy, format of dwelling and other factors emerged as a result of this recession (e.g., Kortteinen & Vaattovaara, 1999). To investigate the matter further, an ex-post analysis of two cross-sections was undertaken in a follow-up study, in order to detect any of these expected changes. Physical factors, however, are unlikely to change within ten years, meaning that only demand side factors will contribute to changes in submarket structure.

The same SOM and LVQ-based analyses were then performed using a later cross-section of Helsinki. The map was run using a dataset from 2001 (N=6,600; once again, 1/3 of total sample) with eight variables. Some of the variables are the same as those that were used in the 1993 analysis, but a number of new variables were added, and there were no aggregated location variables. The eight variables were as follows: surface, format, rooms, public subsidy (either ARAVA or HITAS), age of building, 'starter' status, price per square metre and share of mortgage. The variables for public subsidy, starter and mortgage are

---

not expected to be fully associated with price, but may generate interesting patterns, possibly functioning as relevant criteria for segmentation.

For the SOM, the dimensions of the map were 24 by 16, as with the map that was generated by the data from 1993. A four-class spatial labelling was used to calibrate the maps, in accordance with the domain knowledge of the Helsinki housing market (partly derived from the 1993 analysis): central and western Helsinki, northeast Helsinki, Espoo and Kauniainen, and Vantaa (see App. B: the four labels refer to these parts of metropolitan Helsinki). First, the visual analysis of the map layers revealed the following patterns:

- large dwellings exist in all four spatial segments, as measured by number of rooms and surface area (a relatively small cluster, if measured in surface area);
- typical detached and terraced housing exists in all four spatial segments (Central-Western Helsinki, North-East Helsinki, Espoo and Vantaa);
- a very small cluster of three nodes contains subsidised apartments in Vantaa and Espoo. They thus represent 0.78% of the map surface, whereas their share of the data is only 0.6%;
- there are three very different clusters of old buildings (in the city of Helsinki only): one for starters, one for non-starters and a third cluster;
- the starter dimension sharply discriminates the map between two blocks with approximately 1/3 and 2/3 (starters and non-starters, respectively) of the neurons;
- four expensive clusters, which are very different from each other, emerge. One consists partly of old buildings and starters, and one consists partly of substantially mortgaged dwellings; another cluster consists partly of large, single-family dwellings with substantial mortgages, and the fourth consists partly of substantially mortgaged dwellings that differ from the other clusters with respect to some feature that was not captured by the input variables;
- there are three different clusters with substantial mortgage.

According to the analysis of the 1993 data, house type, size and the age of the building are all important, as are the socio-demographic characteristics of the dwellers. A new aspect that emerges from the analysis is the use of two types of institutional arrangement. These arrangements (i.e., mortgage uptake and subsidies) were expected to have an impact on segmentation, and location, as indicated with a four-class label, was expected to be more important than the 1993 analysis showed. The dataset contains only individual data, however, whereas aggregate-level location variables dominated the 1993 analysis. Bear in mind that the two institutional variables (e.g., public services and open space) were among the most important factors.

Although it is possible to speculate about whether the ARAVA and HITAS subsidy programmes actually pertain to different submarkets, these factors

are not associated with price. The share of mortgage, on the other hand, is strongly associated with high house price, and not only a different submarket.

To investigate the matter further, a classification was conducted using the LVQ algorithms. For the LVQ, a 12 by 8 map was chosen, as with the analysis that was conducted using the 1993 data. (In both analyses, the 24 by 16 maps were classified using the LVQ. Only the results for the 12 by 8 map, which were more informative, are reported in Table 4.3). The a priori labelling was as follows:

- format (i.e., house type): This is expected to yield high classification accuracy, given the 1993 analysis;
- location based on districts: This is expected to be significant, but may not show up, due to the absence of location within the direct input;
- age of the building: Because this is a partial proxy for CBD location, it is expected to yield high classification accuracy (as with the 1993 data);
- price per square metre: This is always an interesting variable to examine from the perspective of equilibrium economic theory (note: an almost double price increase from 1993 to 2001 required strongly redefined labelling categories);
- mortgage: The amount of loan is the first explicitly institutional variable to be included in the analysis thus far;
- number of rooms: 1 to 5;
- subsidy: This is a dummy variable (although the distribution of the data is not sensitive to classification according to this variable, as only 0.6% of the observations are subsidised);
- starter: A dummy variable.

Table 4.3 shows the results of the classification of the 2001 Helsinki data. The dimensions of the map are 12 by 8, and classification is based on two and three labels (with four labels for location):

- The a posteriori classification is based on clusters on the map surface, and is a combination of several factors. The three-label solution (two inner-city segments and a suburban segment) yields the highest accuracy (even higher than the two-label solution);
- Location is approximated as municipality (3 labels: Helsinki/Vantaa/Espoo-Kauniainen; 2 labels: Helsinki/other) or as municipality combined with direction (4 labels: central-west Helsinki; northeast Helsinki; Espoo-Kauniainen;

**Table 4.3 LVQ-classification of Helsinki market segments 2001: verification of the segmentation based on classification accuracy results for various labelling criteria**

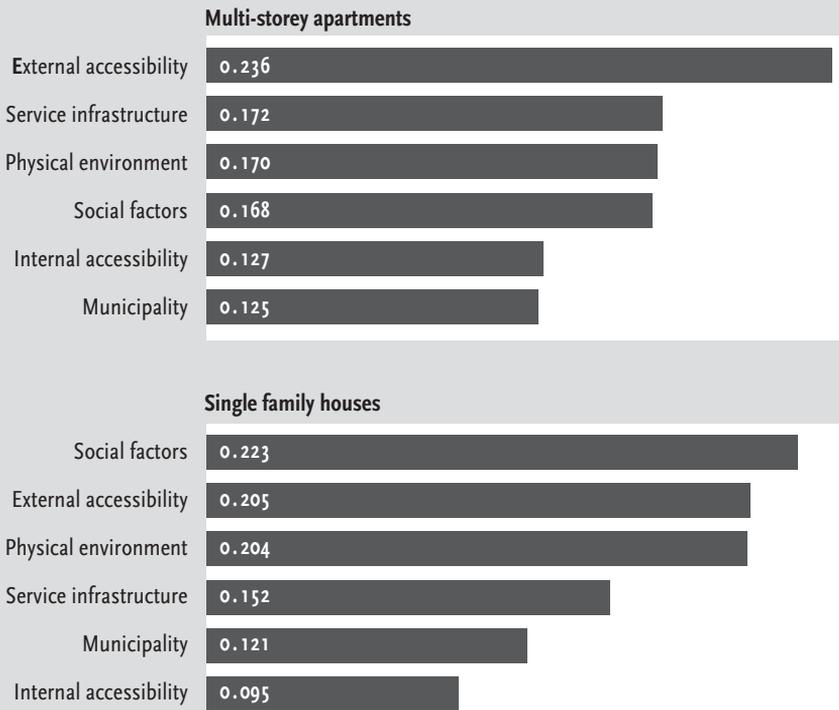
A posteriori, 3 labels	98.40% (98.43%)
Format, 2 labels	97.62% (99.48%)
Rooms, 2 labels	97.23% (99.17%)
Rooms, 3 labels	95.45% (95.60%)
Price, 2 labels	94.93% (94.39%)
Surface area, 2 labels	92.30% (92.77%)
Mortgage, 2 labels	86.93% (92.87%)
Age, 2 labels	83.60% (86.30%)
Price, 3 labels	76.93% (77.58%)
Surface area, 3 labels	75.00% (76.20%)
Location/municipality, 2 labels	70.65% (71.27%)
A posteriori, 2 labels	67.98% (69.95%)
Location/municipality, 3 labels	63.58% (64.51%)
Age, 3 labels	55.82% (61.08%)
Location/municipality, 4 labels	47.22% (49.53%)
Starter, 2 labels	40.88% (43.10%)

Vantaa). At best, this variable is only a moderately important criterion;

- Age is not as important a criterion as it was in 1993. Using two labels (the same labelling as 1993: 1948-2001; 1869-1947), however, it yields relatively high accuracy;
- Format is not sensitive to labelling in three classes, as the categories of the third class become too small. The only valid category is the multi-storey/other type variable, which yields high accuracy (highest for two labels);
- Number of rooms is defined in the same way as in 1993, with two labels (1-2; 3+). The result now is much better than it was in 1993 (the second best classification for both the two-label and the three-label solutions);
- Because of the previous point, measuring size by surface was tried. This proxy for size, however, yielded substantially lower accuracy;
- Mortgage is not sensitive to classification into more categories than a 'mortgage/no mortgage' dichotomy. Nonetheless, the classification accuracy of the 'no mortgage' observations was one hundred percent, whereas the accuracy of the remaining 'mortgage' category was zero percent. Even though this variable had the highest accuracy of all of the variables (99.67%), it is not reported in the table, because of the uneven distribution of classes in the dataset (more than 99% have no subsidy).
- Price per square metre: This variable consists of three labels (€1,500 or less; €1,500-3,500; above €3,500; cf. 1993, ca. double price increase). The two-label classification (< €2,000, and €2,000 or more) is highly accurate. The three-label classification is reasonably accurate, although the most expensive group is too heterogeneous and obtains an accuracy of zero percent.
- Starters: One third of the observations are starters and two thirds are not. Although visual analysis of the feature maps had raised high expectations, the classification accuracy turned out to be low. (Note that all starters were classified incorrectly, suggesting that the importance of this criterion is low.)

Using a dichotomous label, format yields the best classification accuracy, although it did not obtain the highest overall accuracy. Using a three-valued label based on a posteriori clusters generated even better results. In the 1993 results, the a posteriori clustering obtained the best results for three labels, and the two-label solution yielded a better classification than did house type, as shown in Table 4.2. It therefore appears that another locational feature, which cannot be precisely defined by this study, dominates the house-type variable when searching for the most important criterion for segmentation; this order remains constant over time.

In light of the LVQ analysis, we can conclude that age has declined in importance, whereas the number of rooms and (most likely) price increased in importance, relative to the situation eight years earlier. Furthermore, when comparing the results for two locational proxies (age, which is a partial proxy for CBD distance, and municipality), both criteria for segmentation

**Figure 4.2 Synthesised elicitations for both suburban Helsinki models**

have remarkably declined in importance over the eight-year period. In particular, the classification accuracy of age was much higher in the 1993 analysis (97% with two labels, 81% with three labels) than it was in the 2001 analysis (84%, 56%). This means that the owner-occupied housing market in metropolitan Helsinki is still segmented, although the most relevant discriminating criteria appear to have shifted away from age (and, indirectly, CBD distance) and towards the size and price of the dwelling and location in terms of neighbourhood (inner city or suburb; in some cases also a more sectoral or patchy picture). It should be stressed, however, that the locational and institutional variables 'open space indicator' and 'public services' were not explicit in the 2001 analysis.

The main conclusion that can be drawn from this analysis is that location and its many different dimensions are still of critical importance to the formation of submarkets. Furthermore, the housing-market structure is more reminiscent of a mosaic than it was in 1993, due to the decreased importance of CBD distance and the increased importance of neighbourhood character. The location model indicates a fragmentation of the space-distance relation towards multiple equilibria, in which a combination of location, house type and price are important. While macro-locations of homogeneous blocks remain important price determinants within the metropolitan area, the effect is distorted by interactions that cause non-linearity (e.g., with respect to building age and dwelling format).

## 4.5 Expert interviews regarding Helsinki housing markets

In Chapter 2, it was proposed that a more qualitative analysis based on hypothetical data might provide additional insight for determining the dominant housing-market structure. Below is a brief documentation of the supplemental AHP-analysis. The discussion begins with a brief repetition of the method itself and then of the aggregate results. The AHP is based on a pair-wise preference comparison of elements (attributes or alternatives). The pair-wise comparison is usually performed using a standard transformation, in which a scale with values ranging from 1 to 9 is analogous to nine verbal statements regarding the importance of element  $A_1$  in relation to element  $A_2$ . A value of '9' means that 'attribute/alternative  $A_1$  is much more important than attribute/alternative  $A_2$ ', whereas a value of '1' indicates equal importance between  $A_1$  and  $A_2$ .<sup>19</sup> In this way, an ordinal ratio  $A_1:A_2$  is generated. Conducting the reciprocal comparisons over the entire set of elements ( $A_1, A_2, \dots, A_N$ ) results in a comparison matrix in which the relative importance of each element is determined as a cardinal ratio between 0 and 1, in such a way that the sum of the ratios equals one. This applies on a global level across the entire model, or on a local level for any hierarchical level of the model. The 'eigenvalue' technique is the most common way of estimating weights from pair-wise comparisons, as it allows the 'goodness' of the resulting model to be evaluated with a measure of inconsistency (for a full presentation, see e.g., Saaty, 1990; Zahedi, 1986).

The interviews were conducted in 1998, halfway through the eight-year period between the data cross-sections. To reduce dimensionality, only the suburban context was chosen (see Fig. 4.2). The ratios indicate the relative importance of the elements. They can be interpreted according to order or as relative weights according to their magnitude. The sum of all ratios in a single model is 1. The elicitations of twenty-two expert stakeholders showed that external accessibility (distances to CBD and public transport) is the most important attribute in the Helsinki multi-storey segment. In the single-family housing segment, this attribute is the second most important, and social factors (e.g., status, social externalities) are even more important. The internal distances (e.g., to parks, seashore, shopping centre, schools) are less important, and the availability of services is relevant only for the multi-storey segment. Comparison of the two graphs shows how the weights vary between the two segments. For example, the physical environment (including area density and such intangible factors as proximity to nature, aesthetics and

---

<sup>19</sup> Other types of scales (e.g., exponential transformation, direct percentages) are occasionally used instead of the standard linear 1 to 9.

satisfaction) received twenty percent; social factors were even stronger, with single-family houses thirty-three percent more important than multi-storey apartments.

The disaggregated elicitations are presented in Appendix C. The responses were also classified into 3-5 groups for both segments, using an inductive and rough grouping based on similar patterns of preferences. For flats (multi-storey apartments), the following profiles were obtained:

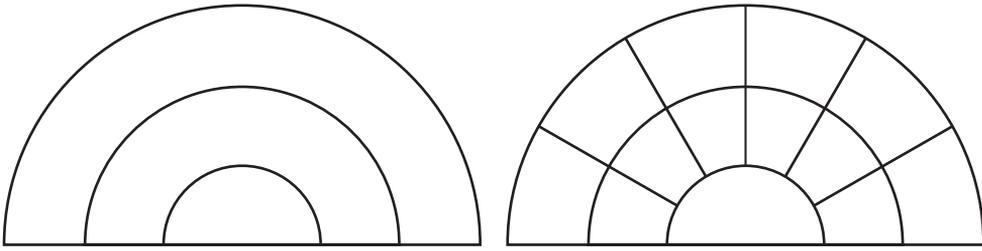
- Group Ia: External accessibility is most important and internal accessibility is also important; status is less important, and density least important. The seven respondents included three estate agents, a planning officer, a researcher, an owner (builder) and the housing manager.
- Group Ib: Accessibility is most important, and status has relatively little importance (close to Ia). The five respondents included three planning officers and two estate agents.
- Group II: Social factors with good external accessibility are most important, and density is among the least important attributes. The three respondents included two owners (both representing the municipality) and an estate agent.
- Group III: In the 'residual' group, some factor(s) other than accessibility (i.e., status, commercial services or municipality) matter most, but density and closeness to nature do not matter at all. The seven respondents included four owners (builders), an estate agent, the rental agent and a researcher.

The same differentiation of preferences was then performed for the single-family segment. Using the same notions as above, the three resulting models were as follows:

- Group I: Status is by far the most important attribute (among the three highest magnitudes in every elicitation); scenery and density are the least important. The ten respondents included five owners (three builders and two representing the municipality), four estate agents and the rental agent.
- Group II: Commercial services are always more important than the status factor (among the three highest magnitudes in every elicitation), and density is the least important attribute. The four respondents included two owners (both builders), an estate agent and a researcher.
- Group III: In the 'residual' group, accessibility, satisfaction with living or various other factors (except taxation) are dominant. The eight respondents included four planners, two estate agents, a researcher and the housing manager.

Given the mono-centric structure of the Helsinki housing market, it is not surprising that external accessibility is the single most important attribute. Social factors and services are also relatively important, however, and the municipality is important as well, albeit to a lesser extent. This result sug-

Figure 4.3 Theoretical spatial urban location models of Helsinki 1993 (left) and 2001 (right)



gests similarities to a Tiebout-type trade-off between various local governmental packages. In that case, prospective investors and residents make locational choices based on the net benefit of municipality characteristics, including taxation. The models that were generated by the expert interviews confirmed the results that were obtained from the analyses conducted with the SOM and with the LVQ: house type matters, as does location, measured as a composite variable of three to five features, which are defined as social factors, physical environment, service level, accessibility and municipality.

To summarise the Helsinki analysis, the spatio-temporal context in question pertains to a relatively well-behaved and compact housing market. When using hedonic modelling or social area analysis data is plentiful and of reasonably good quality, despite the fact that only a few similar studies on Helsinki exist. The results obtained from running the neural networks were informative and partly explicable. In 1993, CBD distance (proxied by the age of the building) and dwelling format (i.e., type) were the two main determinants of submarket structure. Eight years later, they were still important, although price, size and neighbourhood location were gaining importance. Neighbourhood location is determined according to a variety of factors, and it reflects not only the differences between different circular CBD distance zones, but also sharp differences between spatial sectors, including macro-location (e.g., east-west differences). When the purely spatial dimensions are isolated and the complexity is reduced into a 'Burgessian' urban-location model, the dynamics can be illustrated with the simple graphs that are shown in Figure 4.3. When comparing spatial features for two points in time, the basic pattern remains, and a completely new pattern emerges.

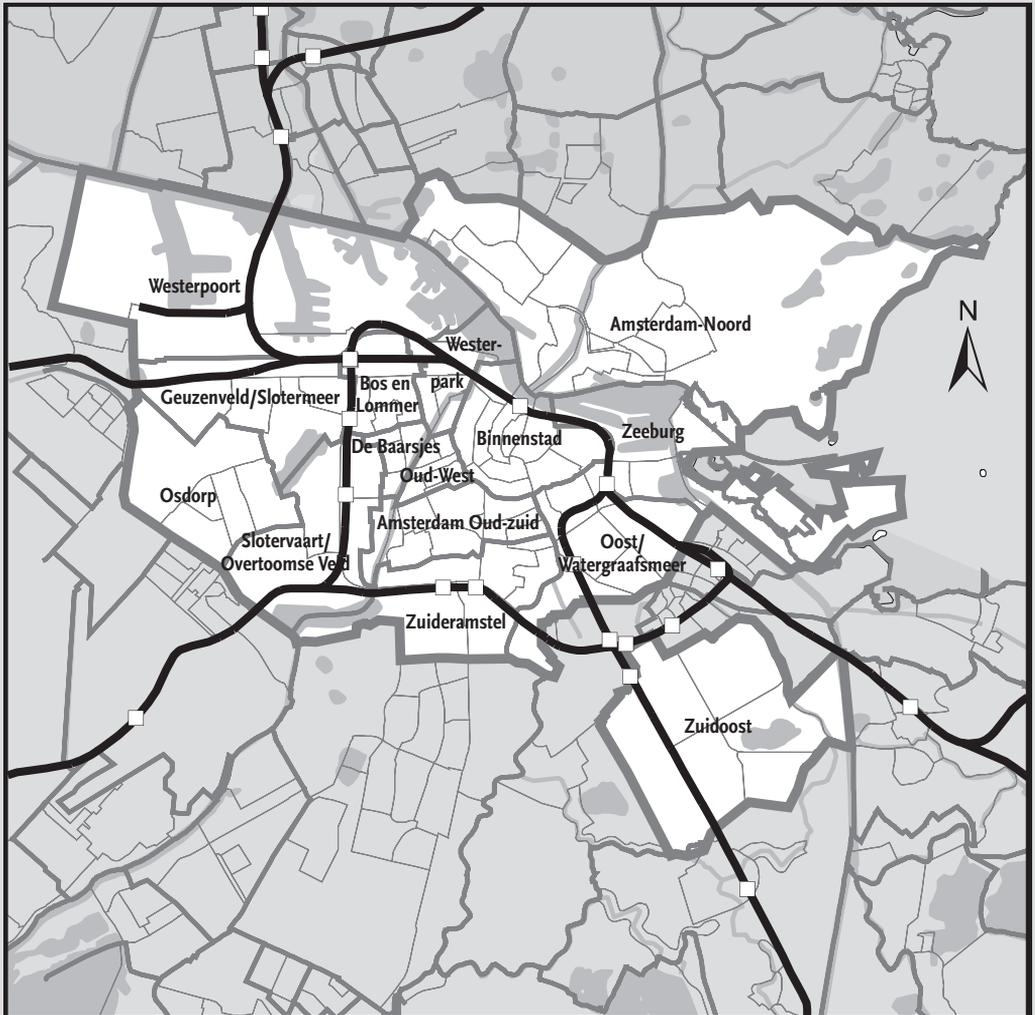
Finally, the AHP analysis suggests that, in 1998, accessibility was the strongest preference of housing consumers in both suburban segments, and that social factors were most important in the single-family housing segment. The differentiated preference profiles also recognised services and the municipality as important factors for both segments, and certain additional physical factors for the single-family segment. While the AHP findings do not say anything about the market mechanism or actual price-formation structures, they do support the notion of that the pattern related to distance-accessibility is relevant, as is the patchier pattern.

# 5 Results of the submarket classifications in Amsterdam

## 5.1 Study area and data

The results from Helsinki suggested a segmentation based on house type (multi-storey apartment or single-family house) and location (initially representing accessibility to the local CBD and later showing a more complex pattern). This result raises the question of whether Amsterdam's housing-market

Figure 5.1 Amsterdam metropolitan area



Source: CBS Netherlands and Geodan IT/Andes

structure shows any similar replications. As used in the previous chapter, the SOM-based method generates only a snapshot of a particular housing-market structure. Unless the results are generalised and compared to those from another urban housing-market setting, no even remotely theoretical analysis of potential spatial housing market patterns is possible, as outlined in Chapter 1. Although Helsinki already was split into two temporal contexts, the division becomes even clearer in comparison to another geographic context. I have therefore chosen a number of new urban areas within which to conduct an exercise similar to the one that was conducted with Helsinki. The first of the new contexts is Amsterdam (see Fig. 5.1), which is slightly larger than Helsinki (ca. 720,000 inhabitants).

Switching to another study area creates two problems: (1) a dearth of experience with the local housing-market contexts and (2) data compatibility. The latter problem, however, was already present when switching to a later data set within the same study area. The first task, therefore, is to look for recent studies that are similar. In other words, it is necessary to find studies on spatial variations in house-price formation, physical features and certain relevant socio-demographic groups. The second task is to find the same variables that were used in the Helsinki exercise (see Table 4.1 in the previous chapter), or at least reasonably similar variables.

Even among Dutch cities, there is reason to believe that the price levels of similar houses differ across neighbourhoods (Spit & Needham, 1987). In a tax-assessment application for dwellings and several types of property in Amsterdam, Needham and colleagues (1998) captured the ‘neighbourhood effect’, which is based on average house-price levels. Similarly, Luttik (2000) estimates shadow prices for environmental factors in three different Dutch localities (Emmen, Apeldoorn, Leiden).<sup>20</sup> In particular, the study confirmed the expectation that the presence of and proximity to water is an important factor in the housing-price bundle. The analysis regarding parks was not convincing, however, although that may have been because the variable was too heterogeneous to operationalise well.

In the Helsinki case, the aspects of ethnicity and segregation were not (yet)<sup>21</sup> relevant. Although data for studying these aspects is available, the topic has not been studied in Helsinki as frequently as it is has been studied in

---

**20** These were ‘garden facing water’ (as high as 28%); ‘pleasant view overlooking water’ (8-10%); ‘pleasant view overlooking open space’ (6-12%) and ‘landscape type’ (5-12%).

**21** Kortteinen and Vaattovaara (1999), however, concluded that social exclusion features are rising.

**22** When defining local housing and labour markets, it is important to consider that, while metropolitan Helsinki is strongly mono-centric, Amsterdam is part of the Randstad region, in which there are several employment centres. Furthermore, the Randstad is considered a polycentric urban field, in which residents have a wide range of residential and commuting options.

---

the larger Dutch cities. Several studies of the residential patterns in the *Randstad*<sup>22</sup> have been made in recent years. Musterd and van Zelm (2001) concluded that the current demand for residential environments is differentiated due to changing household compositions and related changes in activity patterns and residential preferences. This is largely related to rising levels of ethnic segregation in all of the major Dutch cities. Deurloo and Musterd (2001) studied the segregation of Surinamese and Moroccan residents, and Musterd and Smakman (2000) investigated the segregation of Caribbeans (i.e., Surinamese and Antilleans).<sup>23</sup> Bolt and colleagues (2002) noted increasing levels of concentration of ethnic minorities (particularly Turks and Moroccans) in the housing areas that were built between 1945 and 1975, notably in Amsterdam West.<sup>24</sup> There is also growing concern in Amsterdam about spatial segregation and polarisation, a result of the neo-liberal restructuring of the welfare state. In the above-mentioned study, Bolt and colleagues conclude that the trend towards spatial segregation should be alleviated by using correct spatial and housing policy tools.<sup>25</sup>

Caribbean immigrants, who seem to be over-represented and concentrated only in the southeast of Amsterdam (*Zuidoost*), are a particularly vulnerable group. The *Bijlmermeer* area is often labelled an 'ethnic neighbourhood'. The inner suburbs and the inner city are more heterogeneous in this respect.

The open-ended expert interviews that were conducted in 2003 suggested that, similar to Helsinki, the Amsterdam inner city is popular among a certain proportion of the upper middle class that are known as 'gentrifiers' (see Section 5.5). The fact that land use is varied, there are plenty of monuments, water (in the form of canals) is abundant, the infrastructure functions well and – obviously – the famous 'liberal' attitudes are all factors that contribute to the relative attractiveness of the inner-city segment. It is also crucial to note that the urban structure of Amsterdam has developed throughout a long history; during the 'golden era' in the 17th century, Amsterdam played a leading role in global trade (Theebe 2002, 38). The small-scale road network and physical housing structure in the historic centre have also remained remark-

---

**23** Aalbers and Deurloo (2003) confirm that, while there are no ghettos in Amsterdam, 'ex-pats' and business-people from industrial countries, who usually return home to their countries of origin, are more spatially concentrated than are other ethnic immigrant groups.

**24** Bolt and colleagues (2002) demonstrated how ethnic segregation has increased in Amsterdam due to the concentration of minority groups in post-war housing areas, particularly in Amsterdam West. Segregation has actually declined in Rotterdam and The Hague, due to a contrasting pattern of ethnic-minority concentration. Although these groups have traditionally been concentrated in the pre-war areas of the inner city, current trends are shifting the concentrations towards post-war areas in both of these cities as well.

**25** On the other hand, Deurloo and Musterd (2001) predict that the Dutch government's current ethnic desegregation policy may prove inefficient or even perverse.

---

**Table 5.1 General submarket classifications in Amsterdam**

Segment	Price	Dominant physical features	Ethnic concentration aspects
1	+/=	Canals, old architecture	Moderate
2	+/=/-	Parks; mixed architecture (pre-/post-war areas), some urban renewal areas, some 19th century architecture	Stable levels (Turks have lesser degree of concentration than Moroccans)
3a	-	Office blocks, prefab housing, annexed area outside the ring-road	High but stable levels (Surinamese)
3b	=/-	Office blocks, prefab housing, garden city; outside the ring-road, urban renewal areas	Increasing levels (Turks and Moroccans in western Amsterdam) <sup>1)</sup>

1) *Musterd & Deurloo (2002)*

ably intact; in many micro-locations, narrow pedestrian alleys still exist, although they represent the scale of the traffic from that era. The reason is that the inner city did not experience the type of reconstruction that took place in Brussels, Paris, Rome, Budapest and other cities in the 19th century (Wagenaar, cited in Deurloo & Musterd, 2001). Such reconstruction is thought to have had important effects on the development of social patterns in these cities. The fact that Amsterdam did not explicitly divide its territory into rich and poor areas in the 19th century might have had a cushioning effect on today's segregation patterns.

The changed attractiveness of the inner city of Amsterdam is described by Van Duren (1992). He argues that the city core of Amsterdam has experienced an essentially qualitative change, in which its character has shifted towards cheaper and different retail, in order to cater for the preferences of younger customers.

Table 5.1 presents a number of observations regarding the segmentation of the Amsterdam housing market according to casual observation, along with a synthesis of the practical literature and expert interviews. This general segmentation serves as a background for the analysis, and it suggests certain expectations. The general picture is that, as in Helsinki, the urban housing-market structure may be partitioned into three or four main spatial submarkets: (1) the inner city (*binnenstad*), (2) the inner suburbs and (3) the outer suburbs. The outer suburbs may be further divided into (3a) the Southeast (*Zuidoost*), and the (3b) Western Garden cities (*Westelijke tuinsteden*) of *Osdorp*, *Geuzenveld* and other areas in the outer western part; *Amsterdam-Noord* (which also is a rather differentiated part of the city) and other similar areas. This conclusion is based on three types of factors: (1) price variation; (2) physical features, including canals (*grachten*), bridges, parks, architecture and squares (*pleinen*) and (3) social, economic and cultural segregation aspects that may generate additional externalities.

Although the above-mentioned pattern is likely to be roughly applicable, there is considerable diversity within the districts (*stadsdeelraden*) of which Amsterdam is comprised. For example, although they both qualify as 'inner suburbs' according to the classification in Table 5.1, *Amsterdam Zuid* (a luxurious neighbourhood) is quite different from *De Pijp* (a rather cosy, yet densely built neighbourhood not far from the inner city) in terms of both the quality and house price. Both neighbourhoods are also quite different from the

*Indische Buurt*, which is a lower-class area in the eastern part of the city. *Zuidoost*, *Noord* and the *Westelijke tuinsteden* are quite different from the areas inside the ring road around Amsterdam.

The inner-suburbs segment (2) is particularly heterogeneous with regard to price levels<sup>26</sup>: the most expensive and most popular neighbourhoods (e.g., *Willemspark*, *Apollolaan*) are found within this segment, as are some of the areas with lowest property values and worst reputations (e.g., *Staatsliedenbuurt*, *Indische Buurt*). As already noted, *Deurloo* and *Musterd* (2001) show how Surinamese residents are overrepresented in segment (3a), while Moroccans are somewhat concentrated in certain areas of segments (1) and (2) (see Bolt et al., 2002). Turks tend to be less concentrated than Moroccans are (*Musterd & Deurloo*, 2002), although it is not clear at this stage whether this feature is relevant for the SOM analysis that follows. At any rate, the studies cited above do suggest that a connection might exist between the share of minorities and market segmentation.

The physical structure of the city evolved as follows: the inner city developed until around 1800; Segment 2 then developed roughly between 1800 and 1940; (this segment also includes the urban renewal of the housing stock); finally, Segment 3 developed during the post-war period (since 1945). There is no border between the nineteenth-century and pre-war/inter-war neighbourhoods (Segment 2), although there is a strict boundary between pre-war (2) and post-war (3) areas. In addition, the western and southern parts are defined by the highway, which was completed in the 1950s and 1960s. (*Aalbers*, 2002)

Amsterdam markets are differentiated according to the following logic (*Aalbers*, 2002; *Teune*, 2002):

- The cheapest apartments are found in *Noord*, *Zuidoost* and *Westelijke tuinsteden*.
- The most expensive addresses (per square metre) are the apartment buildings in the *Grachtengordel*, *Museumplein* and south of the *Vondelpark* (e.g., *Apollolaan*). The rate of homeownership is high in the most expensive segments.
- The lowest share of housing corporations (30-40%) is in the *Binnenstad* and *Zuid* (see the map of the AFWC website [www.afwc.nl](http://www.afwc.nl)).

---

<sup>26</sup> Because of this heterogeneity, a valid alternative or supplement to this partitioning based on three concentric 'circles' would be to use sector 'slices' based on five geographical directions; as shown here, the inner city remains one segment. Thus defined, the submarkets would be (counter-clockwise) as follows: (1) the inner city, (2) West, (3) South, (4) South-East, (5) East and (6) North. A number of patterns are visible from these segments. Segment (3) is characterized by upper-middle and upper-income groups, and house price (as measured by the total market price of dwellings) is consequently the highest. This segment also contains the largest dwellings. Segments (2), (4), (5) and (6) are characterised by lower-middle and lower-income groups. (This point emerged in a discussion with Manuel Aalbers, AME, on July 2002).

---

- Segment 2 is comprised of areas with very different price levels (cf. Bos & Lommer, *Oud-Zuid*)
- A locational market also exists within the social housing sector. Because different parts of the city differ according to real attractiveness potential, this is not merely an administrative issue.
- Casually observed, Amsterdam is somewhat segregated, as are most European cities. High-income groups prefer to remain in certain parts of the city. In contrast to the *Zuid* and *Oud-West* regions of the city, nearly half (46%) of the *Indische Buurt* and the *Tuinstiteden* are foreigners. The fact that ‘scientific’ methods may be incapable of detecting segregation, however, raises the question of how to define segregation. Should it be defined according to everyday discourse, or are objective measures necessary?
- All of the housing land (*erfpacht*) in Amsterdam is owned by the city (Helsinki has the same policy); the municipality can therefore build owner-occupied dwellings in less-desirable areas as well, if the aim is to equalise value potentials across the city.

A number of urban redevelopment projects have been undertaken recently. For example, Majoor (2002) documented the South Axis (*Zuidas*) project, in which the city government changed its strategy to reflect a pro-market orientation. This area (which includes the Amsterdam World Trade Centre) provides a physical barrier between downtown (Segment 2 in my classification), and *Buitenveldert* (Segment 3b in my classification). The functional urban area *Zuid* continues into *Amstelveen*, even though this area is in another municipality (and is therefore not included in the dataset). The project also includes housing development that was intended to ease the housing shortage in Amsterdam (see also Sluis & Kauko, 2003, on another area known as *Buurt Negen* in the western part of the town.)

Two different datasets were used to run the SOM (i.e., to generate the feature maps of the spatial housing-market structure of Amsterdam; see Table 5.2). The first was based on the KWB (*Kerncijfers Wijken en Buurten* or ‘basic statistics on neighbourhoods and sub-districts’) area database, which contains aggregate demographic and socio-economic data. The KWB database contains 94 observations, with each observation representing the aggregate values (sums or averages) of each variable in one sub-district (*buurt*) of Amsterdam. The database is maintained by the Dutch Ministry for Housing, Regional Development and Environment (VROM). It is a general data source, containing area information aggregated at the sub-district (*buurt*) level.<sup>27</sup> As shown in the

---

<sup>27</sup> I used this dataset for preliminary analysis of Amsterdam housing locations and as a ‘safety net’, should any individual-level data be unavailable in time. The fact that more than one data source has been used should be considered a methodological strength.

**Table 5.2 Description of the variables in the Amsterdam datasets**

K W B , a g g r e g a t e d ( s u b d i s t r i c t ) d a t a o n a r e a s 1 9 9 9	
1	Addresses per neighbourhood (straightforward density-proxy, n-hood defined as a 500 x 500 sq. m.)
2	Extent of urbanisation (evaluated based on a 5-point scale, where 1 indicates more than 2500 addresses/sq. km. and 'very strongly urban area', and 5 less than 500 addresses/sq. km. and 'not urban area')
3	Sq. m. of area including water
4	Sq. m. of land
5	Population density (inhabitants per sq. km.)
6	Total population
7	Population of males
8	Population of females
9	Percentage of 0-14 years old children
10	Percentage of 15-24 years old
11	Percentage of 25-44 years old
12	Percentage of 45-64 years old
13	Percentage of 65+ years old
14	Percentage of non-westerners (first and second generation immigrants)
15	Percentage of 1-person households
16	Number of families
17	Percentage of families with children
18	Average family size
19	Average net income including subsidies per resident
20	Average net income including subsidies per income taker
21	Percentage of low income takers
22	Percentage of high income takers
23	Percentage of 15-65 years old with unemployment benefit as the primary source of income
24	Number of dwellings
25	Mean WOZ-value: assessed total market value of dwelling (price in 1.000 NLG)
26	Number of (urban) firms in the neighbourhood (9 categories)
27	Percentage of industrial enterprises (including construction)
28	Percentage of commercial enterprises
29	Percentage of non-commercial enterprises
I n c o m e t a x e s , i n d i v i d u a l d a t a o n t a x a b l e p r o p e r t i e s , 1 9 8 6 - 2 0 0 2	
1	Transaction price and transaction price per sq. m.
2	Building year
3	Type of house
4	House size (sq. m.)
5-7	Marks 1 (very bad)..10 (perfect) for other dwelling-specific variables: quality, situation, and maintenance (0 is empty or unknown).
8	Lot size (sq. m., incl. possible garden; in case of multi-storey apartment buildings, this indicates the size of the garden only)
9	If the house is situated by a canal
10	Date and year of transaction
11	Municipality land lease (Erfpacht)

table, several of the variables are proxies for a single effect.

The second, and in many ways preferable, type of data, which concerns the

sales of approximately 46,000 dwellings, was prepared by the municipal tax authorities of Amsterdam (*Gemeentebelastingen Amsterdam*) for the calculation of property taxes. This dataset resembles the one that was used in the Helsinki case, in the sense that each observation indicates a specific dwelling transaction, at a particular address and with recorded information concerning the selected attributes. In this case as well, the variables include both neighbourhood-specific and house-specific variables.

The basic data that were obtained from the Amsterdam tax authorities (henceforth taxation data) consist of a seventeen-year (1/1/1986-28/2/2002) panel dataset. Instead of selecting a particular year as a cross-section, the initial plan was to pool several years into the same dataset. (This strategy yielded more data – a relevant point to consider, as the body of owner-occupied market transactions (i.e., the annual turnover) in Amsterdam is relatively light in comparison to Helsinki.<sup>28</sup> The owner-occupied dwelling stock in Helsinki consists of 240,000 units and in Amsterdam, 51,000 (almost a fivefold difference).<sup>29</sup> The dwelling stocks in metropolitan Helsinki and Amsterdam, however, are approximately the same size (ca. 400,000).

The solution of incorporating all observations into a single database requires the addition of the year of transaction as a variable, because of the Dutch housing-market boom in the 1990s. The role of the market situation (i.e., the effect of the market fluctuations) in partitioning the data, however, may be too strong to allow the analysis of the other relevant variables. A nearly three-fold increase in price levels across transactions took place between the beginning and the end of the period. Instead of spatial submarkets, therefore, the analysis would yield information primarily about the temporal factor of market development. Such an outcome would suggest that cross-sectional data is a better solution after all, as it allows the comparison of two cross-sections (e.g., 1993 and 2001) for changes in the spatial patterns of demand-sided attributes, according to the method that was proposed in Section 4.4. (There would nevertheless be sufficient observations for any one of the seventeen years; pooling two years together would also generate a broader base.)

The study area has been a frequent subject of research on all of the aspects that are relevant to this topic; furthermore, plenty of housing-market and

---

<sup>28</sup> The Amsterdam housing market is very different from the rest of the Netherlands, however, even from the other major cities. The homeownership rate for the entire country is fifty percent, while the rate in Amsterdam is only fourteen percent (*Volkshuisvesting in cijfers*, cited in Musterd & Smakman 1998.) In Amsterdam, the owner-occupied market does not function efficiently, due to serious impediments on the investment side. It can be argued that, in this respect, Amsterdam differs more strongly from the rest of the Netherlands than Helsinki differs from the rest of Finland.

<sup>29</sup> Price data were collected from owner-occupied dwelling transactions, although the tax authority (*Gemeentebelastingen*) actually extrapolates from this approximate twenty percent to all Amsterdam housing.

---



Suburban housing in Buitenveldert, southern Amsterdam.

house-price data exist, and they are apparently suitable for the SOM-LVQ method. Several additional points concerning the intricacies of the data are worth noting:

- As in the Helsinki case, the dataset (45,899 observations remained after removing clear errors) was split into three parts for testing and validation purposes.
- As in the Helsinki case, a locational identification code was assigned to each observation. In the KWB dataset, the codes reflect the sub-district (*buurt*) level; codes in the taxation dataset are based on the name of the street in question, the district (*wijk*) and sub-district (*buurt*) levels. In the latter dataset, the identification was also based on the specific format (archetype) of the house.
- As in the Helsinki case, the taxation dataset allows the use of price per square metre as a price variable, although the use of total price is a more common procedure in the Netherlands (in both academia and practice). In the categorisations, the size of the house and plot are split into four areas: A, B, C and D. A represents the main part of the area that is in use, while B, C and D are negligible areas such as cellars, sheds or lofts. In the selected definition of the variables 'transaction price per sq.m.', 'house size' and 'lot size', only the area A is included for each transaction.
- The taxation dataset contains subjective quality variables for the quality and state of maintenance of the house and its situation within the site, which may be an advantage over the Helsinki dataset, as these variables are considered useful for valuation in contexts within which many relevant factors may influence value. On the other hand, such variables are not ideal for neural-network processing, which requires numerical variables. The variables are useful for appraisers, who assign scores from 0 to 10 (in practice from 6 to 9) for each of the three quality characteristics.
- The *Erfpacht* variable indicates a favourable land-lease contract and often a reduced price. These contracts ceased to be renewed after 1 January 2000, which is expected to have brought strong price increases (up to 50%) in certain cases.

## 5.2 Analysis with the SOM

As in the Helsinki case, the results of the SOM analysis (i.e., the Amsterdam feature maps), were examined according to the surface of ‘typical values’, as well as patterns and clusters. Regarding the smaller sample, the resulting 6 by 4 maps indicated levels of physical and demographic density, property values and other features, across all Amsterdam neighbourhoods.

An old inner city (mainly the Old City Centre, *Jordaan*, *Grachtengordel*, the Old Jewish Quarters and *Plantagebuurt*), in which a large proportion of the housing consists of dwellings that are situated along a canal (*grachtenpandjes*) obviously differs substantially in appearance from the suburbs. Even among the suburbs, however, inner suburbs, which have a variety of parks, mixed-density housing and building stock, can be distinguished from outer suburbs, with their office blocks and prefab post-war housing estates. Although the latter type of location is found in the satellite city of *Zuidoost* (the area with a high concentration of Surinamese residents and low house prices), they are also present to some extent in other outer locations, including *Amsterdam-Noord*.

Some of the resulting feature-map layers are shown in Appendix C. The use of variables in each dataset allows certain general comparisons with the results from the Helsinki analysis that were reported in Section 4.2. Although the variables are not the same as in the Helsinki case, they are similar enough to allow certain conclusions about the Amsterdam context, according to a few key dimensions of spatial housing-market features.

The KWB variables of interest here involve density, urbanisation, share of immigrants (i.e., the percentage of first or second-generation non-western residents in the total neighbourhood population), income, share of unemployed residents, number of commercial firms and property-value levels. In addition, the use of labels for street address and the presence of water (the water-coverage indicator is defined as the ratio of two variables: ‘Sq.m. of area including water’ and ‘Sq.m. of land area’, and takes values between 0 and 1.9) in the sub-district allows some conclusions about the aspects that were identified in the Helsinki analysis. The models based on the KWB dataset may now be compared (qualitatively) with the situation in Helsinki, as described in the previous chapter.

By comparing the patterns across the layer that depicts the distribution of price levels with other layers, we can observe a variety of associations with other variables (see App. D). First, the most expensive areas (sub-districts with assessed value levels) are situated in the lower part of the map. The number of commercial firms as a proportion of all firms in the area is relatively high, and the number of non-commercial firms is even higher, as compared to industrial firms. There is no direct association, however, between price and the ‘straightforward’ density variable, which measures the density of homes in an area. The ‘urbanisation’ variable was less important to the organisation

**Table 5.3 Overlap between presence of water and expensive areas in Amsterdam**

No water/cheap area	Plenty of water/Quite expensive area
Indische Buurt	Oostelijk Havengebied

**Table 5.4 Association between water and density in Amsterdam**

No water/high density	Moderately water/Medium (high) density	Plenty of water/Medium (low) density	Plenty of water/low density
Cheap area (Indische Buurt)	Very expensive (Willemspark, although other factors than water bring the premium)	Quite expensive (Oostelijk Havengebied)	Quite cheap (Houthavens)

of the whole map (and not just for showing the cluster of three ‘non-urban’ nodes in the upper left corner of the map, see App. D), than it was in the Helsinki case. This is because most Amsterdam areas may be described as urban anyway, as measured by the number of addresses in each neighbourhood.

The cheapest areas are situated in the upper-right corner of the map, most notably the *Indische Buurt* neighbourhood (as well as in *Landlust* in *Bos & Lommer* and the *Staatsliedenbuurt* in *Westerpark*); these areas have relatively high shares of non-westerners, low-income wage earners and unemployed people, and relatively low average net-income levels.<sup>30</sup> Low average net income overlaps with high proportions of low-income wage earners, unemployed people and non-westerners; conversely, high average net-income patterns overlap with low proportions of low-income wage earners, unemployed people and non-westerners (cf. the situation with the indicators of status and social externalities in Helsinki).

As in the Helsinki case, the presence of water (as measured and visualised with the water-indicator label, as explained above) is important as a determinant of price and other segments in Amsterdam, but not in a linear sense. Using crude and informal notions, some overlap between the presence of water and expensive areas was observable as follows (see Table 5.3 and App. D).

Unlike the shoreline in Helsinki, however, the aspect of water coverage is more strongly related to the average density of the given group of observations than it is to house-price formation or the socio-demographic characteristics of the area. The clearest association is between more water and lower density; there is less association between either of these two factors and appreciation levels (see Table 5.4).

Compared to the layer-by-layer analysis of the Helsinki feature map, the Amsterdam feature map shows more variation and fewer distinguishable clusters. It is also much smaller, and the input data are aggregated by sub-

<sup>30</sup> This result is also robust. The result was the same in a repeated run, and the same two low-status neighbourhoods (*Indische Buurt* and *Staatsliedenbuurt*) were again situated in the adjacent nodes. Only the corner of the map was different.

district, which gives us reason to consider the robustness of the data (as discussed in Section 3.3). According to visual examination, density (as proxied by addresses/neighbourhood), population density and the share of one-person households seem to be particularly strong determinants of the organisation of the map as a whole.<sup>31</sup> Because these factors contribute to the organisation of the map along the longer, horizontal direction (in a later run, these factors contributed to the organisation of the map diagonally), they can be understood as the most important determinants of the organisation. The more composite 'appreciation factor', which was operationalised by income and certain other socio-demographic proxies (i.e., the variables for the [%] proportions of the population aged 15-24 and 45-64 years of age, non-westerners, low-income wage earners, high-income wage earners and unemployed people, as well as the total number of families, seems the second relevant determinant of the dataset (see App. D).

Let us now briefly compare these results to those from Helsinki. It is interesting to note that Helsinki emerges as the more homogeneous submarket context, even though the observations in Helsinki consist of individual dwellings, while the average values in Amsterdam are aggregated by sub-districts. On the other hand, it is more difficult to draw conclusions about segmentation with a parsimonious model than it is with a detailed model. The smaller dataset and map size generated a smoother surface for Amsterdam than it did for Helsinki, where the data and submarket structures were patchier. The analysis continues, however, and a profound analysis of the taxation dataset of Amsterdam may provide a clarifying illustration of the situation, especially when isolated from the time-trend component. This dataset is relatively comparable to the dataset that was used in the Helsinki analysis.

In the taxation dataset, the variables of interest are house type, price, age, dwelling size, quality of location (district, sub-district or even the site itself), and location by a canal (see App. E). Based on a preliminary run, the models generated with this dataset indicate the following:

- Price contributes to the organisation of the map as follows: three clusters with very high total-price levels, two of them for recent transactions; three clusters with very low total-price levels, one representing transactions in the late 1980s and early 1990s transactions. Clustering based on price per square metre differs from clustering based on total price.
- The age of the building stock partitions the data into a number of segments,<sup>32</sup> visible partial associations between old buildings, high total price

---

<sup>31</sup> Of these three indicators, only the density (although differently defined) was used in the Helsinki analysis. In an exploration using data from the entire country of Finland (see Kauko, 2002), the municipal population and the share of one-person households in the municipality proved important determinants of the organisation of the map.

---

and, in some cases, high price per square metre. Partial associations were also observed between new buildings and low total price.

- Low area density, as proxied by single-family housing and large plot size (which are layers that overlap considerably), is partly associated with expensive areas (both total and price levels per square metre); due to a shortage of space, there are fewer green areas than in Helsinki.
- High area density also carries a price premium. This suggests a diversified and non-linear association between high price levels (total or per sq.m.) and low-density development (large plot size and/or single-family homes).
- One submarket with larger dwellings, including the most expensive ones (total price), can be identified; large house size is partially associated with expensive areas (both total and per sq.m. price levels) in one cluster, which contains the *Plantagebuurt* (in the *Binnenstad*), *Chopinstraat* and *Schubertstraat* (both in *Oud-Zuid*). These cases also demonstrate a logical association between large dwellings and large plot size.
- Subjective quality ranks vary widely. Considerable variation can also be observed within a given house size, house type, year of construction and price category. In other words, quality contributes to heterogeneity. It is not associated with price levels, however, with the exception of the lowest quality cluster, which also shows low total price and relatively old buildings (e.g., *Vrolijkstraat* in *Oost/Watergraafsmeer*). Other cases show an association between high price levels and relatively high quality (e.g., the *Schubertstraat*, *Chopinstraat* cluster above). Maintenance rankings logically overlap with quality rankings. However, a cluster of poorly maintained but not essentially poor-quality cases is found in the upper-right corner of the map. Examples include the *Dusartstraat* (in *de Pijp* part of *Oud-Zuid*) and *Warmondstraat* (in the northwestern corner of *Oud-Zuid*, close to *de Baarsjes* and the main street *Sloterkade*).
- The subjective ranking for situation also reflects considerable heterogeneity. The least favourable situations are associated with low total price and old building stock (e.g., *Raamstraat*, in the southwestern part of the city core, close to the *Leidseplein*), and some of the most favourable situations are associated with high total price (e.g., the *Schubertstraat* cluster, as described above).
- In some cases, location by a canal also emerges as a clear determinant, generating a price premium. Location by a canal is sometimes associated with expensive or spacious areas. For comparable houses (by my own estimation), canal frontage (*Lijnbaansgracht* in the *Binnenstad* is on a canal; *Dusart-*

---

**32** Note that the colour scheme of the feature-map layer depicting the influence of age in Amsterdam is different from the colour scheme in the feature-map layer of Helsinki. In this map, new buildings are shown in light shading, and old buildings in dark shading.

---

straat in *Oud-Zuid* is not) has a positive impact on the price per square metre (missing data makes estimates unreasonable), but it has a negative impact (17.5%) on total sales price. Water is therefore apparently a proxy for another variable. Dwellings that are situated along a canal are typically better situated in general; curiously, their plots might also be larger. This fits the model that emerged from the KWB data.

- A strong time trend is evident: four clusters with recent sales and three clusters with sales from the late 1980s and early 1990s.

A notable result was that low density and large plot size possibly generate a price premium in top locations (e.g., *Apollolaan*), somewhat in contrast to the findings from Helsinki. This might be due to the more crowded context, and that space is more highly valued. As in the rest of the *Randstad* region, the physical constraints in Amsterdam are obviously severe. Nonetheless, the SOM models suggest that a few very dense areas (e.g., *Weesperbuurt Plantage*) are clearly appreciated. Furthermore, some of the older, well-kept houses show an association with high price per square metre.

The analysis also shows considerable heterogeneity with regard to the neighbourhoods. The variation in subjective quality, situation and maintenance is remarkable, and some (but by no means all) of the cases that have good rankings for all three attributes are associated with high price levels. Particularly in the inner suburbs (Segment 2), the housing stock varies widely in terms of physical features related to either design or quality. According to the SOM analysis, however, the absolute top locations are not ambiguous in relation to their actual situation; they are located along the river *Amstel*, on the *Churchillaan* in *Nieuw-Zuid* and on *Apollolaan* in *Amsterdam Zuid*. Recall also that a later sales year is associated with higher price.<sup>33</sup> This factor thus contributes strongly to the organisation of the map; in this sense, it is possibly the most important feature. It is therefore necessary to run the SOM on a single cross-section at a time (for the same years that were addressed in the Helsinki study).

---

**33** This price development was not corrected for in the data sets, because comparing price levels in absolute terms between two points in time was not the topic of the analysis. For the whole country, Boelhouwer (2004) gives an increase of 59% in real terms for the 1995-2003 period. In nominal terms the increase in Amsterdam is of course even higher. Simple calculations with the data shows that the mean price level was only €494 per sq.m. in 1993, whereas it was as high as €1,332 per sq.m. in 2001 – an increase of 170%! (For Helsinki the same figures were €1,056 in 1993 and €1,885 in 2001, respectively – thus only an increase of 79%). In a longitudinal or time-series study, or even in a cross-sectional study with direct practical applicability, of course the issue of deflating the price variable is a crucial one.

---

### 5.3 Analysis with the LVQ

As with the Helsinki analysis, the intent of this analysis was to determine the discriminatory power of various formal partitioning criteria. In other words, what is the most important factor that is associated with segmentation? Is it an expression of the physical characteristics of the house itself and its vicinity, or does it express particular aspects of the location (e.g., density, CBD accessibility, socio-demographic indicators or the price level)? The analysis of Amsterdam considered price (i.e., assessed value, variable 25 in the upper portion of Table 5.2) in comparison to the following six 'objective criteria': indicators for density (Variable 1), urbanisation (2), population density (5), proportion of non-western immigrants (14), proportion of the population aged 25-44 years (11) and the water-coverage indicator label (see Section 5.2).

The following classification emerged from the analysis. Using three values<sup>34</sup>, the results indicated that density (measured either as the straightforward address/neighbourhood indicator of an area, or as the extent of urbanisation in an area) is the most important criterion, but that price also plays a substantial role (see Table 5.5). The weakest results were generated by the proportion of the population aged 25-44 years, a socio-demographic indicator.

The fact that the two measures of physical density (urbanisation and straightforward density) yielded results that exceeded property value by 4-5 percentage points as criteria for discrimination is not surprising. According to these criteria, Amsterdam can be partitioned into three spatially distinct segments, thereby supporting the results of the visual analysis. The physical element of house location is a relevant partitioning criterion. The above-mentioned margin, however, is too narrow to warrant a definite conclusion.

The primary results did not change when the classification was continued using four labels: the most important criteria were density and urbanisation. The extended analysis actually strengthened this conclusion. The

**Table 5.5 Results of the LVQ-classification of Amsterdam submarkets with KWB<sup>1)</sup> data**

Number of labels and definition of labelling criterion	Classification accuracy (training sample) in %
3 labels urbanisation	93.62
3 straightforward density	92.55
4 straightforward density	92.47
4 urbanisation	90.43
3 population density	89.36
3 property value level	88.51
3 proportion of non-western immigrants	88.51
4 population density	85.11
4 proportion of 25-44 year old	83.87
3 amount of water coverage	83.87
3 proportion of 25-44 year old	79.35
4 property value level	78.16
4 amount of water coverage	72.04
4 proportion of non-western immigrants	71.91

1) KWB = *Kerncijfers wijken en buurten* (Core figures of districts and neighbourhoods)

<sup>34</sup> The two-value label generated no useful results; all indicators reflected very high recognition accuracy. Results from this label are therefore not reported.

four-label classification for these two criteria yielded a better result than the three-label classification for the other five criteria did. Additionally, the analysis revealed a number of more specific points:

- The straightforward density criterion is indeed a very strong discriminating feature, as the classification result is almost as high with four as it is with three labels.
- The classification result for the proportion of the population aged 25-44 years was even higher with four than it was with three labels, suggesting that a four-label classification is more meaningful than a three-label classification is.
- For the property-value level, the four-label classification result was very high (up to 87%) for both the lowest and the highest class. This result indicates that the ends of the price continuum are easier to identify as separate segments than the middle part of the price continuum are.

This analysis also reveals several differences with the Helsinki context. In Amsterdam, the market structure is comprised of fewer monotonous housing areas, and the spatial association between density and price is greater (albeit nonlinear and differentiated, as in Helsinki). On balance, the results obtained with the KWB dataset show that the price structure in Amsterdam is heterogeneous in all segments. In particular, the middle part of the market is remarkably mixed in urban space, as compared to its Helsinki counterpart (with reference to the situations in 1993 and in 2001). The small size of the dataset, however, did not allow a separate test sample for validation; only a training sample was deployed. For this reason, the same LVQ analysis was also performed with the taxation dataset, which is substantially larger than the KWB dataset. This provides more rigour for the analysis of the two cross-sections that follow. Note that the modelling process is partly a matter of data.

After running the entire massive dataset of Amsterdam house prices, I split the dataset into cross-sections for the sake of manageability and, obviously, to allow comparison with the method that was used in the Helsinki case, particularly with regard to the LVQ classification.

## **5.4 Modelling the spatial dynamics of the Amsterdam housing market based on a comparison of two cross-sections**

As in Section 4.4, this exercise was conducted with a follow-up comparison of two cross-sections. The ex-post analysis of the dynamics and structure of segmentation was conducted with an earlier and a later cross-section (1993 and 2001, respectively, for Helsinki). With Amsterdam, the periods 1992-93 and

2000-2001 were selected, as they correspond to the years that were addressed in the Helsinki analysis. Note that, for Amsterdam, two years are combined into a single cross-section to compensate for the fact that the annual turnover among owner-occupied dwellings in Amsterdam is lower than it is in Helsinki.

The results of the new runs are as follows:

- 12 by 8 maps
- length of the runs: 4,800/48,000
- alpha: 0.05/0.02
- radius: 10/3.

Feature-map layers for the 1992-93 runs are shown in Appendix F. The feature maps that were generated with the 1992-93 cross-section demonstrate the following:

- House price: Price levels correspond to lot/garden size and to house type, at least to the extent that the most expensive houses belong to two typical suburban cases: (1) single-family dwellings with two or more storeys, built 1960 or later (*Zuidoost*) and (2) dwellings in modern free-standing buildings of six or more storeys (e.g., *Zuideramstel*). The price association with diversified area density is also applicable.
- House type: The map shows clear clustering with regard to house type: 1/3 of the map consists of single-family dwellings, including the very old and inexpensive as well as the most expensive; 2/3 of the map is comprised of multi-storey housing from all price levels.
- Year of construction: The oldest houses are inexpensive, but well situated (archetypes 1, 2, 5 and 18). The newest are very expensive, but most are poorly situated (archetypes 7, 10, 12, 13 and 15). Dwellings that were built between 1994 and 1996 were sold before the building project was completed, enabling a six-percent tax deduction (taxation is an institutional influence). These cases typically belong to one of the two most expensive multi-storey segments. In some cases, classic old-style buildings (archetype 2) have been (re-)built very recently.
- House size: The largest houses are in *Oud-West* and *Geuzenveld/Slotermeer*. These houses, however, are not the most expensive.
- Plot/garden size: Plot or garden size corresponds to higher prices (see above). The single-family segment has larger plots, and the multi-family segment has larger gardens. Pure space generates a price premium.
- Quality: One clear cluster appears in the *Binnenstad*, where quality is high. There is no association between quality and price.
- Situation: The map reveals two clusters in which the micro-situation is good: (1) mixed housing in the *Binnenstad* and *Zeeburg* and (2) partly high-priced single-family housing.
- Maintenance: A clear cluster partly overlaps with quality, but not with situation. Maintenance is not associated with price.

- Canal situation: A cluster consisting of 1/3 of the observations represents dwellings that are situated by a canal, typically in the *Binnenstad*, *Zuideramstel* and *Zeeburg*. These areas are partly comprised of single-family housing, typically small, inexpensive, with small gardens or lots, but well situated. To some extent, houses with canal frontage do represent a unique part of the market, but no clear price premium (or discount) could be detected for them. Canal situation is thus an extremely important determinant of the overall market structure, and it has a partial association with low density (i.e., in many cases, more water is associated with lower density), although it has no clear association with price levels. Although this finding seems counterintuitive, the analysis with the total dataset had produced somewhat similar results, as discussed above. An association between price and canal, however, emerged from the panel dataset (see Section 3.2). This unanticipated result might have been due to multicollinearity.
- *Erfpacht*: Municipal land lease contributed to a division into two types. (1) Counter-intuitively, the majority of the most expensive houses had *Erfpacht*; as lower land leases had been expected to indicate lower house prices. There is reason to suspect that, in this context, this dummy (like the canal dummy) is a proxy for some other (missing) variable. (2) A smaller segment comprised of houses with *Erfpacht* but with lower prices (also canal situation) is also distinguishable.

The urban-suburban division (2/3 and 1/3, respectively) is indicated by the following discriminant factors. (1) With regard to house type, most multi-storey buildings are urban, and most single-family dwellings are suburban. (2) A minority of the cases within each house type represent the opposite segment: urban, single-family homes (archetype 18), typically in the *Binnenstad*, and suburban multi-storey buildings of two types (typically archetype 12). (A) Well-maintained multi-storey buildings are typical of *Geuzeveld-Slotermeer*, and (B) very expensive multi-storey buildings with large gardens and other features that are not visible from the map layers are typical of *Zuideramstel*.

The most relevant findings from the feature maps for the 2000-2001 cross-section are presented below (see App. G; cf. App. F.):

- Price discriminates the map very clearly; the right side is 'expensive'. This pattern also emerged from the 1992-93 data.
- New (and inexpensive) cases discriminate the upper-left corner of the map into one clear cluster. This pattern differs from the pattern that emerged from the 1992-93 data, in which no such association was identifiable.
- House type shows up clearly, as the majority of the map consists of multi-storey buildings. The lower-left corner indicates inexpensive single-family homes, and the lower-right corner indicates expensive single-family homes. This pattern is quite different from that of the 1992-93 data, in which the single-family clustering was more homogenous.

- House size generates a pattern that overlaps completely with both total price and price per square metre. This pattern differs from the 1992-93 pattern.
- The plot/garden-size variable generates a pattern that overlaps with size and price. The upper portion of the map indicates multi-storey buildings that have no gardens, and the lower portion of the map indicates single-family dwellings on large plots. This pattern is the same as the pattern that emerged from the 1992-93 data.
- The quality of the dwelling generates a pattern that sharply separates (1) high-quality, inexpensive, small multi-storey cases in the middle and lower middle from (2) low-quality cases of at least three types. The low-quality cases include (a) new, inexpensive, small multi-storey buildings; (b) large, expensive multi-storey buildings of low quality and (c) large, expensive, single-family dwellings of low quality. This pattern diverges somewhat from the pattern that emerged from the 1992-93 data.
- The quality of the situation generates a pattern that highlights a minority of cases that are well situated but nevertheless inexpensive (multi-storey buildings with small or no gardens). Because this cluster is small, the pattern differs somewhat from the 1992-93 pattern.
- Maintenance generates a pattern that highlights a minority of cases that are well maintained but nevertheless small, inexpensive, older multi-storey apartments (with small or no gardens). In general, this segment is very heterogeneous, a pattern that is relatively consistent with the pattern that was generated by the 1992-93 data.
- Canal situation generates patterns that resemble the 1992-93 pattern. The category of dwellings with canal frontage includes all types of housing: inexpensive as well as expensive, high-density as well as low-density and multi-storey as well as single-family dwellings.

The conclusion from the SOM analysis is that four features are relatively stable between the cross-sections: transaction price, plot/garden size, maintenance and canal situation. Price currently has the same relative weight as it had eight years earlier, but it is now more clearly associated with house size, which is similar to the Helsinki findings. The rest of the features, however, differ from the 1992-93 situation. Note also that both of the institutional aspects – taxation and *Erfpacht* – are completely missing from the later cross-section.

To summarise the visual analysis, patterns for building age and district (*wijk*) location are the strongest determinants of the data structure in the analysis of the earlier cross-section, while patterns for dwelling quality and micro-location (i.e., immediate vicinity of the dwelling) are the strongest determinants in the later cross-section. Another change in the patterns between the two points over time is a fragmentation of the spatial patterns of the inner suburban segment in Amsterdam (i.e., Segment 2 in Table 5.1).

**Table 5.6 Results of the LVQ-classification of Amsterdam submarkets with 1992-93 and 2000-01 taxation data**

Criterion	1992-'93 in %	2000-'01 in %	Change in position 1992/'93 – 2000/'01
Canal, 2 labels	99.75 (100.00)	99.32 (100.00)	1. – 2.
Format, 2 labels	99.75 (100.00)	98.06 (100.00)	2. – 3.
Land lease ( <i>Erfpacht</i> ), 2 labels	98.46 (98.77)	-	3. – (-)
A posteriori: urban/suburban, 2 labels	90.33 (90.56)	99.96 (99.95)	4. – 1.
Building year, 2 labels	77.09 (77.55)	64.27 (62.20)	5. – 8.
Transaction price, 2 labels	69.53 (70.23)	85.96 (96.70)	6. – 7.
Situation of the house, 2 labels	65.76 (68.27)	91.13 (94.30)	7. – 4.
Quality of the house, 2 labels	64.96 (67.40)	90.38 (93.50)	8. – 5.
Maintenance of the house, 2 labels	64.47 (66.85)	89.62 (93.15)	9. – 6.
Location (neighbourhoods, 20 labels)	34.85 (37.39)	21.70 (23.70)	
House type (archetype, 20-25 labels)	32.57 (36.35)	24.74 (30.65)	

The results of the LVQ analysis for both 1992-93 and 2000-2001 are presented in Table 5.6. The map dimensions were 12 by 8; the size of the test sample included 1,624 observations, and the training sample (in brackets) included 1,626 observations. A two-label solution was tested in this analysis, together with a separate comparison between location and house type, which involved several labels and obviously resulted in much lower classification accuracies.

The results regarding canal situation and *Erfpacht* were frustrating, as the anticipated association with price (canal brings premium; *Erfpacht* brings discount) was not evident in the feature maps. Both of these criteria, however, proved very important as determinants of the overall data structure, as the LVQ classification shows. The interpretation must be that these features do have some meaning in this context, but not in relation to a price premium.

The LVQ classification largely confirms the visual analysis of the feature maps, revealing several new findings as well. When we compare the analysis with the situation 8-10 years later, we can note the following:

- The three most important criteria remain the same (four, if *Erfpacht* is counted): canal, format and urban/suburban location (and *Erfpacht*).
- The classification accuracy of this group of criteria differs substantially from that of the other groups (e.g., year of construction, transaction price).
- The order of classification accuracy remains largely between the criteria (canal and format > situation > quality > maintenance), suggesting that the immediate surroundings (i.e., micro-location), which are almost impossible to change, are more important than the level of facilities, which is difficult but not impossible to change. The level of facilities, in turn, is more important than the condition of the house, which is easier to change. Nevertheless, situation, quality and maintenance have become more important as criteria for segmentation over the eight-year period.
- In the earlier analysis, location (approximated as district or *wijk*) was more important than house type. In the later analysis, however, house type is the more important of the two (cf. the AHP analysis: both views were expressed in 2003, at least in the VINEX segment).

- The urban-suburban distinction, a new feature that was identified in the SOM output but not included in the input, is a crucial criterion for segmentation, which cannot be concealed. This result demonstrates the capacity of the method to identify residual factors. Furthermore, the importance of the urban-suburban distinction increased over the ten-year period.
- Year of construction and location (at the district or *wijk* level) have decreased in relative importance. The importance of all other factors has either increased or remained the same (cf. the SOM analysis above: canal, price and maintenance appeared as static criteria).

## 5.5 Expert interviews regarding Amsterdam housing markets

The expert-interview component of the Amsterdam study had the same intent as in the Helsinki study: to obtain additional information concerning preference profiles as a demand-side determinant of price, and to incorporate the intangible quality component (soft factors) of housing choice into the housing-market analysis. Seventeen expert interviews were conducted during 2003, with questions and target groups that were similar to those that were used in the Helsinki analysis. Several alterations in the research design were necessary, however, partly because of the difference in urban context, and partly because of lessons that had been learned during the interview phase of the earlier Helsinki study.

First, the AHP analysis concerns the urban and suburban (but not peripheral) areas in the entire *Randstad* region of the Netherlands. The entire region was chosen because, unlike the mono-centric Helsinki, Amsterdam belongs to the polycentric *Randstad*, which is usually considered a single, tightly integrated metropolitan region – at least by its more mobile residents and professional employees.<sup>35</sup> Furthermore, the urban areas outside the city of Amsterdam were retained as a separate segment for the AHP elicitation. The initial partitioning was between Amsterdam and the (other large urban areas in) *Randstad*. On the other hand, there is a substantial difference between what housing consumers seek in the ‘old town’ and ‘new town’ segments. After the first interviews, it became evident that developers (in contrast to other actors) tend to consider housing consumption from the perspective of the new-build market; the comparison between VINEX (Greenfield) and the older cities (Brownfield) is therefore a key issue. The other expert groups, in turn, were more likely to view housing consumption through the lens of the second-hand market. As it was not clear which perspective represented the best possible segmentation, it was necessary to incorporate all three segments. In some interviews, the conclusion was that Amsterdam is not significantly different from the other cities with regard to any of the comparisons.

The second difference concerns the hierarchical structure of the model. Although the Helsinki model applied a two-level structure in attributes (see App. C), this model places all attributes on a single level, splitting the physical environment further only into soft and hard factors.<sup>36</sup> Accessibility is not split further into internal and external because, as already noted, the *Randstad* region is polycentric. In addition to the attributes that were included in the design of the Helsinki analysis, supply-side friction is included, because of the serious shortage of space in Amsterdam. This element was indeed discernable from the features maps that are illustrated in Appendices D through G. The resulting graphs (AHP calculations for the aggregate models, similarly as for Helsinki) are presented in Figure 5.2, in a manner that is similar to its counterpart figure in the Helsinki analysis. A brief explanation follows.

First, all three *a priori* segments revealed a similar result: accessibility and municipality (Amsterdam is a single municipality) have low priority in all of the aggregated models. On the other hand, substantial differences were also observed:

- ‘Brownfield’ or urban Amsterdam (no substantial difference to the other old cities within the *Randstad*) is characterised by two dominant features: (1) shortage of space and (2) social factors.
- ‘Greenfield’ or VINEX: this segment is completely different from above, and is characterised by two dominant features: (1) physical factors (these were split further into ‘hard’ and ‘soft’ factors; see below) and (2) services. This is consistent with the SOM analysis. Density, urbanisation, house quality and the quality of the location matter, thus producing segmentation along two dimensions: Brownfield/Greenfield and price level.

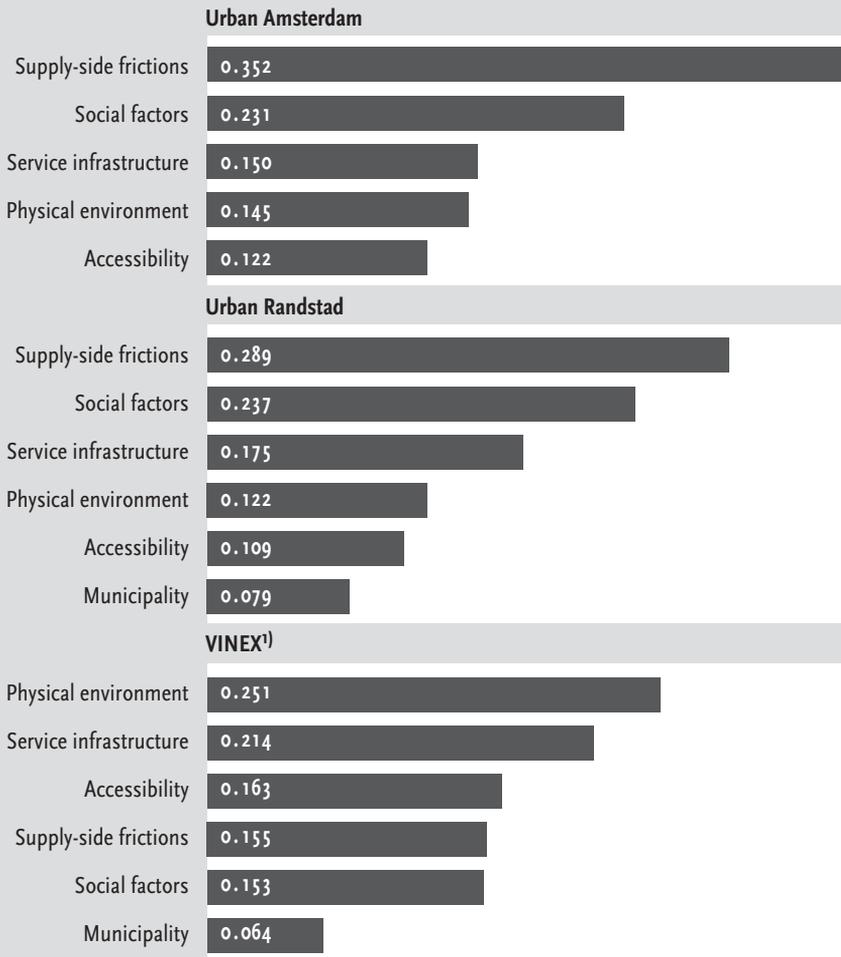
---

**35** All of the contextual definitions are obviously open to debate. First, why was the *Randstad* chosen, and not a larger or smaller region? One interviewee noted that the study area should include the ‘*Zandstad*’ (the ‘Sandy city’; i.e., the region of Breda, Tilburg and Eindhoven), southeast of the *Randstad* region. This region has a considerable IT industry, and it forms a corridor towards Antwerp in Belgium and the *Ruhrgebiet* in Germany. According to this view, Amsterdam is not significantly different from the other cities with regard to any of the attributes that are included in the AHP exercise. Another interviewee, in turn, offered the opposite argument: Amsterdam is not polycentric for lower-middle and working-class residents. For these groups, Amsterdam is unique, particularly because of its proximity to Schiphol Airport. At the most, the northern part of the *Randstad* (which includes Almere, Zaanstad and other municipalities) should be defined as the context in which locations are substitutes.

**36** The comparisons were not split into several hierarchical sub-criteria and attributes, as in the Helsinki study, which was a pioneering exercise that had been conducted five years earlier. The decision not to split the comparisons was intended to prevent mistakes that had been made during that study. Specifically, the hierarchical partitioning of the comparison into more detailed level of questions in the Helsinki analysis resulted in time-consuming interview sessions and more than reasonable overlap and ambiguity in the comparisons amongst the attributes, when these were clarified to the respondents.

---

**Figure 5.2 Synthesised elicitations for the Randstad models**



1) New locations constructed under the Fourth Report on Spatial Planning Extra.

The disaggregated elicitations are presented in Appendix H. The interviews involved the following logic.

- **Accessibility and proximity** - As shown in Figure 5.2, although this factor is not as important for the old cities, it is of moderate importance for the VINEX-locations. As shown in Appendix H, this factor is the second most important factor in one of the three profiles for the urban *Randstad* (i.e., the profile in which services were of highest importance); it was otherwise of little importance in the profiles that were elicited for the three segments. The explanation was that, in general, accessibility is good throughout the *Randstad*; in particular, the inner city of Amsterdam is accessible for everyone. This is clearly different from Helsinki, where accessibility was the most important factor (external accessibility to the CBD was most important and internal accessibility, including distance to the closest services, was second). Although these factors had posed a major problem for the *Randstad* in

the past, the situation has been completely resolved. Two exceptional types of location in Amsterdam, however, do currently resemble Helsinki:

- VINEX locations, which are not yet optimally accessible (e.g., *IJburg* near Amsterdam has accessibility problems). In several years, however, the infrastructure in these locations will be good, as road construction tends to follow housing construction.
- The gentrified inner city of Rotterdam (*Kop van Zuid* and the northern strand of the Maas) and Amsterdam (*de Pijp*): People want to live within walking distance of the city centre.
- *Social factors of the neighbourhood* – Figure 5.2 shows that these factors are important for the urban segments, but not for VINEX dwellers. As shown in Appendix H, they are relatively important in seven of the eight profiles, including one of the two VINEX profiles. While this set of factors is probably of general importance across all of the contexts, they are less important in VINEX locations than they are in inner-city segment(s), according to most of the interviews. Although this group of factors is the most important for existing neighbourhoods, it is the least important determinant of attractiveness for VINEX locations.

The negative externality side of this attribute may be the only reason that customers have for rejecting or accepting certain neighbourhoods. There is a growing fear about neighbourhoods in which the share of immigrants is high. These residents have not returned to their countries of origin and they have not integrated into Dutch society. (This point is related to the discussion of the soft physical factors, which appears below.) Furthermore, the social factors are completely insignificant for the VINEX segment.

- *Service infrastructure in the neighbourhood* - As shown in Figure 5.2, these factors are of moderate importance for all aggregated profiles. Appendix H shows further that this group constitutes the most important attribute in one of the three *Randstad* profiles. These factors are of relative importance in all of the other profiles. In Amsterdam, services are better than they are elsewhere in the *Randstad*. Services in urban Amsterdam therefore ranked lower in the aggregated models than they did in the other two segments. Neighbourhood service infrastructure is a strongly differentiated aspect. Young people want bars; elderly people want churches and hospitals (for them, this factor may become more important in the future). Schools, hospitals and playgrounds are the most important services in the VINEX locations. In the inner city, ‘cultural infrastructure’ is the most relevant factor is. Throughout the *Randstad*, most areas already have schools, shopping centres and similar facilities.
- *Physical environment* - There is a fuzzy line between the soft and hard characteristics of the physical environment. For example, noise and pollution are defined here as *soft*, but they could also be considered *hard*, because they are measurable. As shown in Figure 5.2 and Appendix H, however, this

composite factor encompasses strongly divided views. For the aggregated profiles, physical environment was the most important factor in the VINEX segment, but it was a relatively minor factor in the two urban segments. For the disaggregated profiles, it was the most important factor in one profile each in the urban Amsterdam and VINEX segments, but it was not more than moderately important in any of the other six profiles.

A number of extreme cases regarding these features were mentioned. Although the density of Osdorp in Amsterdam is low, is the area is otherwise considered unpleasant. *De Pijp* is a high-density area, but it also a pleasant living area. This neighbourhood is currently regarded as a major success story, as it contains a mix of all income groups. A number of years earlier, the area had been slated for demolition (on strictly rational grounds, because of its location in the middle of 'the golden axis' of four highly popular areas: the *Binnenstad*, *Oud-Zuid*, the *Rivierenbuurt* and *Buitenveldert*). According to one respondent, areas in the inner-suburb segment in Amsterdam (e.g., the *Transvaalbuurt* and even the *Indische Buurt*), whose relative location is perfect, but whose social and physical factors are unfavourable, are currently improving and could emulate the trajectory of *de Pijp*, given continued investment. In general, *De Kolenkit* is the worst area in Amsterdam.

Hard physical factors matter more for the VINEX locations than they do for the 'old towns' (multiple weights were given in some of the elicitations). People want space and low density. In old towns, the aesthetics of the urban milieu and other soft factors matter more, possibly enticing customers to accept smaller houses.

- **Municipality** - This factor was not of even moderate importance in any of the seven profiles of the VINEX and urban *Randstad* areas (see Fig. 5.2 and App. H). The municipal image and local government policy do not matter much for the decision to reside in one city or another. Amsterdam, however, is considered more attractive than are other urban municipalities in the *Randstad* (particularly Rotterdam and The Hague).<sup>37</sup> In general, Amsterdam is attractive and has a favourable policy (at least according to the respondents from Amsterdam).

Traditionally, Amsterdam's policy (e.g., *Erfpacht*) has been considerably more important than its image. Current rules and regulations are not as strict, however, and they have declined in importance as factors that influ-

---

**37** According to one of the respondents who were affiliated with housing consumer stakeholders, different income and age groups appreciate different aspects of the location. In principle, however, Amsterdam's attraction lies in the presence of firms, services, Schiphol airport and social factors (e.g., status, nuisance, belonging). Although the hypothetical consumer in this study is assumed to be upper or middle-class, Amsterdam tends to focus more on problem groups than do the other cities in the *Randstad*.

---

ence residential location. VINEX areas have even fewer regulations. In general, the central tax system and the relatively small differences in character amongst the communities makes this factor less important in the Netherlands than it is in Finland. Image does matter, but it matters less than any of the neighbourhood level attributes do. The VINEX locations also have an image of their own, which is far more important than any particular municipality.

- *Supply-side friction* - In the aggregated profiles, this is the most important factor for urban cities, but not for the VINEX locations (see Fig. 5.2). In the disaggregated profiles, supply-side friction is the most important factor for one profile in each segment (App. H). The in-depth interviews confirmed that this attribute also split the judgements into two main profiles. Because of the friction factor, people tend to be satisfied with a poorer-quality environment. This friction is thus one of the main overall factors, according to many comments.

The presence of friction within a housing-market area depends largely on the rent policy. In particular, many movers do not consider Amsterdam, as it has the most severe scarcity problems.<sup>38</sup> It was pointed out, however, that even in serious shortage situations, people do not choose these areas if the social or physical attributes are inappropriate, even if they are affordable. Taken together, the demand factors are ultimately more important than the supply factors are.

In the Netherlands during the 1990s, the supply-side market shifted towards the demand side (but not everywhere). Recently, however, the tide appears to have turned, such that a supply-side market once again prevails in the Netherlands. In addition, the production of houses is relatively low throughout the *Randstad*, due to regulations and appeals against the plans (according to several private-developer respondents).<sup>39</sup> According to Boelhouwer (2004), decreasing output of new homes poses a serious problem for situations in which demand has been boosted. This situation is currently present in the Dutch housing market, largely due to governmental planning policies and building regulations. The VINEX locations are not faring as badly in this respect; there is less 'movement in the market' in these areas than there is in older cities. In other words, supply-side friction is not as substantial a factor in VINEX locations as it is in the urban context (with

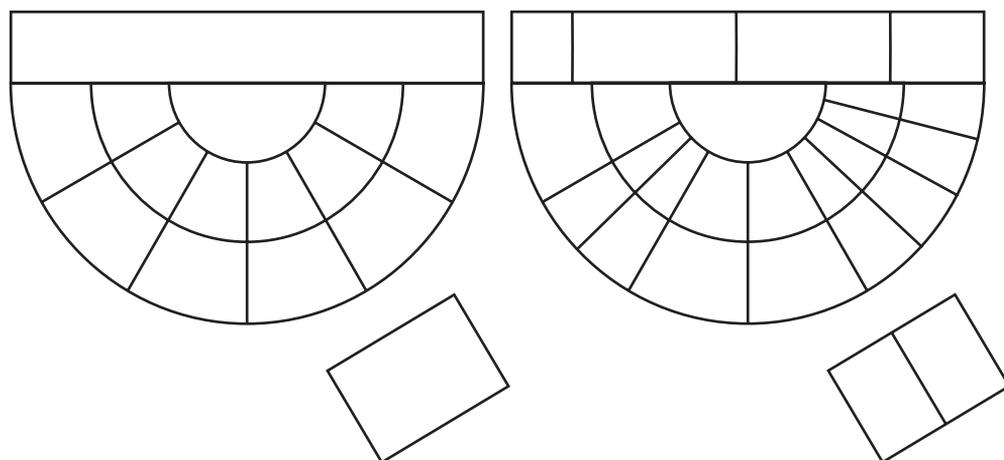
---

<sup>38</sup> The dwelling shortage is 'not so bad' in de *Westelijke tuinsteden* (the garden city suburbs of Western Amsterdam), but it is 'very bad' in the south and central parts of Amsterdam. Furthermore, until the 1990s, The Hague also had considerable housing shortage; the current situation is more comparable to that of the less constrained Rotterdam than it is to Amsterdam.

<sup>39</sup> De Vries and colleagues (2003) note that municipalities perceive the problems with the various 'bottlenecks' that impede the building market less than private developers does, as they operate on a longer-term perspective.

---

Figure 5.3 Theoretical spatial urban location models of Amsterdam 1992-93 (left) and 2000-01 (right)



the Amsterdam housing-market area as the worst example in this respect). This situation imposes fewer constraints on the choice of house and location. Supply-side friction is thus less important for Amsterdam as a whole than it is for the old city.

To summarise the Amsterdam analysis, the context in question represents a fragmented housing market, in which regulation plays a substantial role. More studies have been conducted in this context than in Helsinki, using hedonic modelling, social area analysis and large-scale questionnaire surveys. The data are of excellent quality. Although the sales data are less frequent than are those from metropolitan Helsinki, the total stock is the same. Multiple datasets were run with the SOM and the LVQ. In most cases, the results that were obtained were informative, if not always explainable. The small aggregate dataset from 1999 and the total massive 1986-2002 panel dataset revealed considerable heterogeneity in spatial structure, and it highlighted the importance of the density factor in the formation of submarkets. Splitting the latter dataset into two cross-sections (1992-93 and 2000-01) revealed that the spatial division of submarkets along 'sectoral slices' has become more detailed, and that building age and district location has been overtaken by other factors (e.g., dwelling quality and micro-location) as the most important criteria for submarket formation. In Figure 5.3, the spatial submarket pattern is illustrated in its most simple terms (as in the Helsinki analysis) using concentric (semi-)circles, which have remained the same, as well as sectors, which have become denser and have not necessarily remained the same between the two points in time.

The AHP analysis (of the entire *Randstad* region) suggests that the preferences of housing consumers in 2003 were strongest for social factors in the urban segments and physical environment in the VINEX segment. Supply-side shortages were also of extreme importance in most choice profiles. The differentiated preference profiles also recognised services (in all segments) and

accessibility (in the VINEX segment) as important factors. It is likely that all of these factors, in different and intricate ways, have contributed to the mosaic-like submarket pattern that was revealed by the analysis of price data. The Amsterdam market is more differentiated than is that of Helsinki, but it nonetheless has its own logic. The dynamics depend greatly on the spatial containment caused by overall land shortage and supply-side friction, given the presence of more physical and institutional barriers than there are in Finland.

## 6 Analysis of Amsterdam, Rotterdam and The Hague with the SOM

A full comparison of the two cases under study is presented in the last Chapter 7. Before that, a further empirical issue was addressed: the SOM analysis of all three major Dutch cities was conducted using a common set of housing-market variables as input. The aim of this procedure was to determine whether any further information could be extracted that would support or distort any of the conclusions that have been made so far. While some similarity between markets within the same national boundaries could be expected, the nature of the housing-market structures of The Hague and Rotterdam (the other two major Dutch cities) is essentially different from that of Amsterdam.

Only a few key dimensions will be addressed as necessary background information for the SOM modelling, as all possible differences and similarities would require a separate article. How similar are Amsterdam, Rotterdam and The Hague? Recent literature suggests a number of substantial differences: The Hague is the political and administrative capital of the Netherlands, and Amsterdam is the financial capital. Rotterdam is the 'working class' city, as reflected in house price, which is lower in Rotterdam than it is in the other two cities. With regard to the revitalisation of densely built residential areas, Wassenberg and van Kempen (2004, pp. 141-142) argue that Dutch cities have clearly become subject to a variety of social, economic and physical problems, according to the logic of the Big Cities Policy that was developed in the late 1990s. The types and extent of problems, however, are expected to differ somewhat across the three cities. For example, the share of multi-storey buildings is much larger in the housing stock of Amsterdam than it is in that of either Rotterdam or The Hague. Furthermore, the proportion of one-person households is higher in Amsterdam than it is in either The Hague or Rotterdam (Boelhouwer, 2002).

Aalbers (2005) compares the redlining practices of financial institutions – a form of place-based social exclusion – in Rotterdam and Amsterdam. He discovers that claims about the existence of redlining tendencies are valid only in Rotterdam and that this is due to the combined effect of at least three types of substantial differences in terms of the socio-demographic, institutional and housing-market environment:<sup>40</sup>

- 1 socio-economic differences: Amsterdam has a much larger share of a middle-class population within the city boundaries than Rotterdam does;
- 2 differences in home-mortgage finance: The National Mortgage Guarantee Fund was applicable in Amsterdam five years earlier than it was in Rotterdam;
- 3 differences in housing-market tightness: Price levels in Amsterdam are

---

<sup>40</sup> According to Aalbers' study, the fourth determinant of such tendencies is the difference in the tightness of the mortgage market; redlining is more likely to occur in a tight mortgage-market situation. Because this issue concerns the national financial market, however, no spatial component is involved.

---

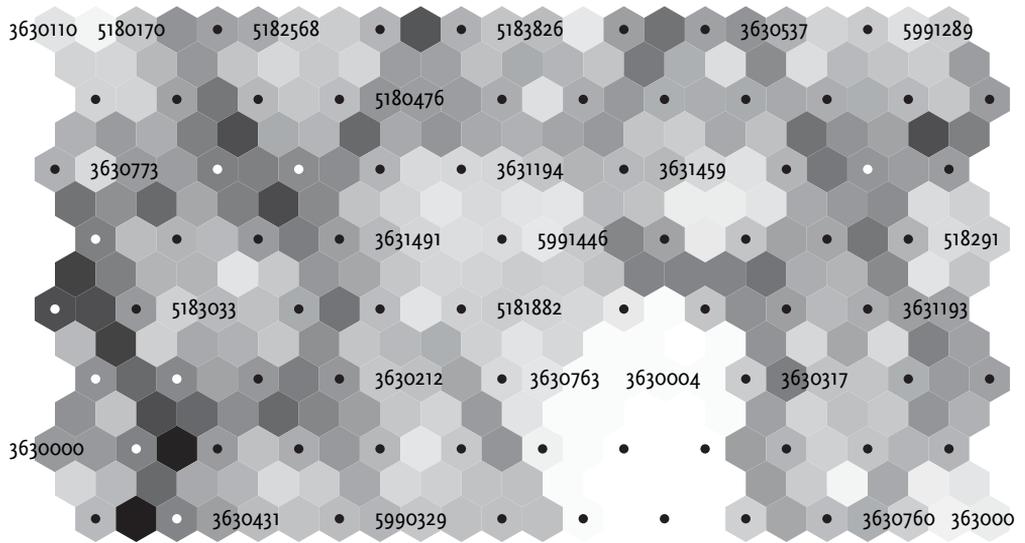
2.5 times higher than they are in Rotterdam, which has also produced a remarkable difference in the levels of investment in existing housing stock.

On the other hand, these three cities are likely to share a number of problems that have been documented for all four 'large Dutch cities' (the fourth being Utrecht). For example, as Dieleman and Wallet (2003) observed, in spite of steady government attempts to ameliorate income differences in Dutch society, such disparities continue to exist at the spatial level, at least to some extent. In fact, the difference between the income levels in the poorer central cities and in the affluent suburbs has been increasing since the 1970s. According to Dieleman and Wallet and to Boelhouwer (2002), relatively low incomes are to be found in all of the largest cities.

The analysis of Amsterdam, Rotterdam and The Hague was conducted partly using the SOM on the KWB-WBO dataset, and partly using expert interviews. The KWB data served as a basis (1999), combined with the WBO housing-demand survey that is maintained by Statistics Netherlands (1998-2000). This allowed the combination of socio-demographic data with data on people's opinions about the vicinity of their homes. The meaning and content of the KWB data has already been explained (see Section 5.1 and Table 5.2). The other dataset, the Housing Demand Survey (*Woningbehoefte-onderzoek*, or WBO) is administered by Statistics Netherlands to a representative sample of 120,000 Dutch residents. The variables are listed in Appendix I. The two datasets are linked with the help of postcodes. The first 25 variables are subjective judgements of the residents, expressed as two, three or five-point scales, with the exception of the weighting variable. Due to averaging over the whole sub-district area, the values do not cover the entire range from one to five, or even from one to three. For example, the variable 'satisfaction with the quality of the vicinity' ranges in value from 1.28 to 3.05; the 'attractiveness of built environment' ranges 1.41 and 3.26, and the 'annoyance of the neighbourhood' ranges from 3.44 and 4.69. The data are then linked to the KWB statistics to determine the extent to which the cases (in terms of opinions) correspond with the cases as measured with the indicators that were used for the KWB dataset.

The combined KWB-WBO and SOM datasets generated the output that is shown in Figure 6.1. This graph differs from those that are included in the appendices. It is suitable only for the identification of various elements within the data structure; it is not appropriate for the assessment of particular cases. Instead of the layer-by-layer analysis that has been presented thus far, this figure represents the map as a whole in terms of distances between the reference vectors of neighbouring map units on the surface (Kohonen et al., 1996a). The valleys and peaks are thus unrelated to any measurements of variables, referring only to similarities between cases, with similarity is defined as nearness between two nodes, measured in 54-dimensional space (54 var-

**Figure 6.1** Feature map of the three largest Dutch cities, showing the distances between reference vectors of neighbouring map units (lighter shading indicates greater distances)



ables were used: 25 from WBO and 29 from KWB, see App. I). The figure is intended to demonstrate similarities and differences amongst the districts in each city, and how they compare to districts in the other two cities.

It is immediately obvious that similarities or differences between neighbourhoods are not due to the specific attributes of a particular city (sub-districts in Amsterdam begin with 363; sub-districts in The Hague begin with 518, and Rotterdam sub-districts begin with 599). Fundamentally, the neighbourhoods in a particular city are as different from each other as they are from the neighbourhoods in another city. The western harbour area (*Westelijk havengebied*) in Amsterdam (the upper left corner neuron, labelled 3630110) and *Oostduinen* in The Hague (5180170 in the same corner) are situated much more closely to each other on the map than they are to other sub-districts in their respective cities. (Both of the areas are located along the shoreline.) Distances on the map (i.e., differences) are also relatively small between *Landlust* in Amsterdam (3630537) and *Groot-IJsselmonde* in Rotterdam (5991289), and between *Schildersbuurt-Oost* in The Hague (5182917) and *Bijlmer-centrum* in Amsterdam (3631193), especially considering the fact that distances between many neighbourhoods within one city are much larger on the map. On this map, *Landlust* and *Groot-IJsselmonde* represent two relatively poorly connected, unpopular areas, while *Schildersbuurt-Oost* and *Bijlmer-centrum* represent better-connected areas that are nonetheless unpopular. All four corners of the map and the middle part are labelled after Amsterdam. With the exception of the lower-right corner, the same applies for The Hague and, with the exception of the lower right and upper left corners, for Rotterdam.

In some cases, however, areas within the same city form a visual block on the map; the labels of the neurons in the upper-middle part of the map (*Mar-*



**Suburban post-war housing estate with considerable ethnic population in southwestern Amsterdam**

lot, Noordpolderbuurt and Duin-zigt) are all from The Hague, and those in the lower-middle part of the map (*Houthavens*, *Tuindorp Buiksloot* and *Nieuwmarkt Lastage*) are all from Amsterdam. The lower left and lower right corners of the map belong to Amsterdam,

whereas Rotterdam is represented by an isolated node on the lower side of the map (5990329, *Schiemond*). Such homogeneity is minor compared to the heterogeneity across the whole data structure. In terms of the characteristics of their sub-districts, therefore, the three largest cities are not so different from each other after all, considering the differences between the sub-districts within one city. They are considerably well mixed, with a few exceptions, most of which are in Amsterdam. Neighbourhoods (*buurten*) in each city vary widely, particularly in Amsterdam. Areas that are close to each other geographically are likely to be located at considerably distant conceptually.

The aim of this module was to provide a comparison of Amsterdam, Rotterdam and The Hague, with regard to spatial housing-market structures (economic, socio-demographic, natural and institutional features), house prices (i.e., property-tax assessment values) and locational preferences, aggregated at the district level. The application of the visual SOM approach in a manner similar to that applied in Chapters 4 and 5 revealed similarities and differences between these three city contexts. The SOM analysis of Amsterdam, Rotterdam and The Hague was not consistent with the findings from the expert interviews (or with the assumptions of the *a priori* segmentation) that were presented in Section 5.5. In particular, information from the expert interviews suggested that Amsterdam differs from the other two large cities in the sense that it poses more problems for mass-appraisal (e.g., for taxation purposes). According to Ruud Kathmann of the Dutch Valuation Board, property valuation in Amsterdam involves more non-market information (50% of the explanatory power is derived from such 'soft', intangible factors as image) than it does in the other cities, which fit into a simpler model (90-95% of the variation is explained by such 'hard', tangible factors as building efficiency). Ostensibly, many of the differences are caused by policies that are implemented by the local government. The KWB data can be used to investigate differences that stem from the potential of the areas, while the WBO survey data concern preferences for the current residential milieu, as stated by the residents themselves. These differences, however, were not captured by the SOM analysis.

---

# 7 Generalisation of the results

## Gaining theoretical insight into segmentation

### 7.1 The power of the comparative analysis of housing markets

In general, comparative analysis can be used to elaborate theory that untangles the institutional and behavioural elements of housing markets. The success of such analyses obviously depends on the ability to overcome problems caused by data incompatibility (cf. Daly et al., 2003). The objective of this comparative study is to develop an image of the housing markets in different European metropolitan areas. This objective is highly relevant for the urban development and the real estate markets, which are beginning the process of globalisation. What pricing mechanisms are able to explain the real estate prices in these different areas? Place A might generate a market outcome that is different from the outcome in place B. The general goal of testing and comparing the outcomes of different approaches is to develop a conceptual model of spatial housing-market structure in relation to institutional and behavioural elements.

I will now return to the five research questions that were stated in the introduction to this report. (1) What are the relevant differences between the two contexts? (2) Which features have a notable association with prices in each context? (3) How does the submarket structure and its determinants change over time in each context? (4) What is the value of the SOM-LVQ classifier in relation to hedonic price analysis? (5) How can this tool be used for decision-making? The first three questions are interlinked and pertain to the empirical findings; they examine the relevant similarities, differences and changes in empirical terms. This section, therefore, address these three questions in the same account. The last two questions, which pertain to the advantages of this approach, are addressed in Sections 7.2 and 7.3, respectively.

The first research question is the most difficult. To what extent does the balance between economic, socio-demographic, institutional and physical factors change, when moving between institutional and geographical contexts? Do fundamental differences prevail, as could be assumed from the outset? The answer requires elaboration on the findings. While housing-market structures of both Helsinki and Amsterdam are characterised by multiple equilibria with regard to the balance between price, socio-demographic indicators, physical characteristics and government regulation, the resulting spatial patterns are completely different. In Helsinki, the residential location is modelled as a set of large homogeneous blocks and belts. In Amsterdam, it is modelled as a more heterogeneous mosaic of smaller patches. Had the temporal changes not been incorporated in Sections 4.4 and 5.4, the submarket patterns would have taken on a completely different shape: the Helsinki market would have been circular, and the Amsterdam market an overlap of both circles and sector-slices. The temporal changes, however, reveal that sectoral slices have emerged in Helsinki as well. There are thus no differences in the

---

basic shape of the segmentation, even though the number of submarkets is larger in Amsterdam. Because the two urban areas are of comparable size, the Amsterdam pattern is far more detailed and shows smaller submarkets than does the Helsinki pattern.

According to the analysis, the submarket structures of the two urban housing markets are quite similar. In both cases, the most important factors involve area densities and other physical and relational aspects of the location and house-specific characteristics. Additional factors that contribute to the formation of submarkets are house price (or value) and government regulation; these factors contribute significantly less to the process than physical factors do. Examination of the measured variables in general, and the observed property value levels in particular, however, reveals that the variation within any one submarket is much wider in Amsterdam than it is in Helsinki. In Amsterdam, there are more relevant factors behind submarket formation than there are in Helsinki. In Helsinki, the spatial differentiation of the market (and preferences) occurs *across* segments, which differ primarily in terms of CBD distance; in Amsterdam, differentiation occurs within the suburban inner-city segment (but across radial sectors within this submarket), if segmentation is based on sectors and half-circles, as it is in Helsinki (cf. Figures 4.3 and 5.3). This example demonstrates that boundary definition plays a decisive role.

The second question concerned possible price determinants. This question was investigated primarily through the associations that were identified across the layers of feature maps, and by analysing expert-elicited consumer housing choices (i.e., attractiveness, as a proxy for price), using the AHP. It is necessary to note, however, that the AHP-elicited choice-modelling exercise as such is, at best, a weak and indirect method of price determination. Using the visual SOM, the most important price associations in both cities were with house size (square metres or rooms), type (single-family is most attractive), age (nonlinear association and proxy for CBD distance), density indicators (nonlinear and differentiated association) and location (district, neighbourhood). In Helsinki, CBD distance and neighbourhood effects showed clearer and more linear price associations than they did in Amsterdam. According to the AHP, social factors are important determinants of attractiveness in both 'single-family Helsinki' and 'urban Amsterdam'. In all other respects, however, the determinants differ greatly across the two cities: in Helsinki, accessibility is important, while the availability of houses matters the most in Amsterdam. Other important factors in Amsterdam include services and the physical environment in VINEX locations. In general, the two areas appear to be more similar than they are different.

What types of features must be considered on a theoretical level? The third research question was addressed in Sections 4.4 and 5.4, which concerned how various patterns change over time. According to the findings, although

price and demand factors (e.g., income) increased in importance during the late 1990s in both Helsinki and Amsterdam, supply factors have remained as key criteria. These supply-related constraints are partly topographical and partly the effects of institutions. In other words, the results also suggest that, in each of the two urban areas under study, the natural and fabricated supply factors were and still are more important determinants of submarket formation than are socio-demographic demand factors (which are obviously more difficult to measure). Despite the results of the AHP analysis regarding preferences, indicators for 'status' and 'social externalities' did not turn out to be as important in Helsinki as indicators of physical environment were. The evidence from Amsterdam also supports the view that area density is more important than socio-demographic indicators are.<sup>41</sup> This may be related to the finding that the relevant criteria for segmentation changed only marginally during the eight-to-nine-year period in each case. Most of the criteria that were important determinants of submarkets in the early 1990s thus remained important throughout the decade.

The third research questions concerned the development of spatial housing patterns in time (here for a period of less than ten years). The convenient metaphors of circles and sectoral 'slices' help to illustrate how the segmentation in Helsinki implies a fragmentation of the neat (semi-)circular model into a model in which the circular and sectoral submarkets overlap. A similar pattern of overlapping circles and sectors already existed in Amsterdam in the earlier 1990s, and it has become denser since then, implying an increase in the variation within a single *a priori* submarket.

The analysis also revealed a number of crucial differences in relation to temporal dynamics: in Helsinki, building age (and therefore CBD distance) and house type were the main criteria for segmentation in 1993, while house type, price and size, together with neighbourhood and macro-location – and possibly institutional arrangements regarding subsidy and mortgage – were the main criteria eight years later. The information from Amsterdam revealed a similar shift in the main criteria from building age and district location to dwelling quality and micro-location (i.e., the immediate surroundings of the house). Furthermore, the segments in Amsterdam may not have remained the same, as the spatial division of submarkets fluctuates. This is particularly true for the various submarket sectors within the inner suburban belt.

In relation to the link between segmentation and segregation (see Section 2.2), there are growing concerns in both Finnish and Dutch society about the increased polarisation of urban areas. In this discussion, residential differ-

---

<sup>41</sup> In an interview, W. Teune (*Stedelijke Woningdienst Amsterdam*) disagrees and suggests that, in Amsterdam, status is the most important determinant of locational choice. Nonetheless, objectively physical factors carry more weight than do price and social factors, although they had not changed substantially in ten years.

---

entiation is often understood as a phenomenon that is somehow associated with increasing income differentiation (e.g., Siikanen, 1992). The same applies for segmentation in the housing market: it is either a direct submarket division based on income level, or a division based on a combination of income, race and ethnicity (see e.g., Van Kempen & Priemus, 1999). These results indicate that Amsterdam is particularly likely to follow the latter path.

Linking the results to what we already knew about both market contexts revealed notable similarities between the housing markets of Helsinki (1993 to 2001) and Amsterdam (1985 to 2002). The similarity was evident in the formal criteria of classification. Price levels do not matter; various locational factors do. A number of contextual similarities are evident even without neural-network analysis. In particular, both inner-city areas are small and surrounded by water, their socio-demographic characteristics are similar and, in both cases, policymakers are concerned about increasing social inequality throughout the entire society, which has spatial implications for the housing market.

On a more particular level, however, the differences are immense (and it was never the intention of the study to emphasise all differences related to physical environment, planning and housing system, and cultural aspects): results of the analysis indicate that water, accessibility and municipality are important determinants of house price in metropolitan Helsinki but not in Amsterdam. As a traditional administrative city, the urban structure of Helsinki obviously differs from that of Amsterdam, which is an old commercial city and seaport that transformed into a service and consumer-oriented centre. Everyone who has been to a Northern European and a Central European urban area know that the structural differences between any two cities from each category are substantial. The former type of area is predominantly modern, while the latter type tends to be a mixture of old historic sites and modern buildings. Land use is not the only difference; considerable variety may be also observed in the attitudes and values of the residents. The aspects of similarity and differences are not ambiguous, as the comparisons differ with the generality of the analysis. We can therefore derive the following conclusions:

A. General similarities between Helsinki and Amsterdam:

- in both cases, the segments are (semi-) circles and sector-slices that become increasingly detailed;
- while price and demand factors (income and other socio-demographics) have increased in importance, supply factors (physical and institutional) remain as key criteria in both contexts.

B. Particular differences between Helsinki and Amsterdam:

- the housing-market structure of Amsterdam is more fragmented and detailed than is that of Helsinki;
- the main discriminating housing-market features and the ways in which they have changed over time are somewhat different for each case.

Each city has its own sets of particular explanations for a particular outcome,

---

and these explanations need not be counter-intuitive. The Helsinki findings were all to be expected, according to the literature. In Helsinki, the growing importance of neighbourhood location can be explained by the demand of the high-tech professionals for suburbs that are situated in the western part of the town and Espoo. The eastern portions of the city have become stigmatized by growing concerns over possible concentrations of unemployed people in many housing-estate areas. Further, income segmentation is increasing. It may be that remnants of the Tiebout-type effect that was found by Laakso (1997) remain, thus differentiating between the four metropolitan municipalities. Traditionally, Finnish dwellings have been small, and they have only recently begun to catch up with other western European standards. The share of high-tech, graphic and other professionals who demand luxurious dwellings is probably part of the explanation.

In Amsterdam, micro-location was more important than district level was. New islands of regeneration and gentrification have emerged (and continue to emerge) in the inner city. These often develop as small, piecemeal 'hotspots', parallel with the emergence of disadvantaged inner suburbs. As in Helsinki, several locations in Amsterdam have become stigmatized, as the ethnic diversity of the population and its associated segmentation have increased in recent years, leading to social problems in many cases. The state of neglect in the inner-city pre-war housing stock clearly illustrates the importance of dwelling quality in Amsterdam.

Some of the most obvious explanations for differences between the two contexts concern the way in which the data on dwelling format are coded: Statistics Finland distinguishes among four categories of house types: detached, semi-detached, terraced and multi-storey dwellings. In contrast, the tax office in Amsterdam works with a system of twenty-two different codes for house type. The codes reflect the administrative unit of a (partial or whole) house, the age of the building and various architectural characteristics (e.g., type of roof). Recently developed innovations (e.g., the Dutch 'floating homes', which are dwellings that are situated by the shore, but which are unaffected by floods) are a related issue. Even on this casual level, it is obvious that the housing supply in Amsterdam is more diverse than it is in Helsinki.

The differences and similarities between the two city, country and time-specific contexts are difficult to investigate responsibly according to the dimensions that are considered in the analysis. To what extent are the similarities between the cases based on general factors that shape urban housing markets? Answering this question requires some sort of benchmark with which to evaluate what is different and what is similar. The basic assumption is that different places generate different outcomes in terms of spatial housing-market dynamics. Findings that support similarity may therefore suggest that both contexts share the same assumptions regarding market structure. In both cases, particular indicators of physical density and design, urbanisation and rel-

---

ative location were more important criteria for segmentation than were price levels alone. Because the study addressed only two contexts, however, any conclusions regarding the assumptions of our theoretical model would be premature. Even though some level of similarity may be observed, the relative significance of that similarity is unclear, given all of the possible idiosyncrasies that could prevail. Excessive similarity between the two cases in terms of spatial market structure and price formation would require broadening the variation and extending the comparison to a clearly different context.

On the other hand, the comparison was also extended into other market contexts within the same country to confirm how much the national context serves as a common denominator to influence the structure of the urban housing market. This module may therefore help to consider the issue of similarity and difference between Helsinki and Amsterdam. Extensive differences across the three Dutch cities would suggest that no such national-level influence exists, thus bringing the results from Helsinki and Amsterdam closer to each other. The results of The SOM-LVQ analysis, however, indicated great similarity across the three *Randstad* cities, with regard to the character of their sub-districts. The sub-districts belonged to a variety of neighbourhood types, and only some were associated with geographical location (i.e., the fact that they are part of the same city). Intra-urban segmentation is thus stronger than inter-urban segmentation. The analyses of Rotterdam and The Hague indicate that there are more similarities between Amsterdam, The Hague and Rotterdam than there are between Amsterdam and Helsinki. This conclusion is not surprising. When comparing cities from two different countries, similarities tend to be on a general level, while differences tend to be more detailed. Cities within a single country, however, tend to be more similar, at least for these particular cities and according to these indicators.

## 7.2 Methodological and theoretical considerations

The evidence that was obtained with the SOM-LVQ classifier (and enhanced by the AHP judgement) in the previous three chapters has generated ideas for further elaboration. A submarket-classification approach, such as the one documented above, could help to break down the price development trend into structural components. This may shed extra light on the extent to which socio-demographic factors (income, unemployment and immigrants), physical environment (density and urbanisation), provision of services and distribution of employment, and property-value levels overlap to produce patterns and combined effects. How these effects dominate each other may provide useful information about the qualitative and quantitative key dimensions along which intra-urban locations are to be assessed.

---

Two features are of particular importance: (1) the main determinants of price and (2) the main determinants of segments. One interesting finding is that the determinants of segments become relevant when the determinants of price are dependent upon the context. Conversely, if the context is uniform everywhere, segmentation is obviously not necessary for modelling price; the hedonic-price model would therefore provide an adequate framework. A comparison of our empirical results with those that appear in other city comparisons of residential differentiation and housing market structure (i.e., Vaatovaara, 2002; Ley et al., 2002; Meen, 2001; and Bourassa et al., 1997, 1999) demonstrates the multi-functionality of the SOM. This method offers dimension reduction, clustering and estimation in a single package, and its visual properties allow the consideration of boundaries between the segments from a fuzzy perspective.

Despite the empirical advances that it allows, the key aspect of the SOM-LVQ approach is that it allows the examination of findings concerning the first three research questions in light of broader theory on the topic (see Section 2.2). The results that are addressed in the following discussion are therefore related to the current theoretical approaches of segmentation: the strict economic-equilibrium model, the multi-equilibrium model (i.e. allowing for multiple equilibria and localised disequilibrium) and the explicitly behavioural and cultural model. It must be noted that the level of analysis (ranging from purely conceptual to fully operational) must remain constant when comparing the results. Hedonic modelling would have allowed only an operational analysis to confirm or reject hypotheses that had been derived from a strict equilibrium-modelling framework.

As explained in Section 2.2, the trade-off theory of residential location (Alonso, Muth, Mills) is also applicable to segmentation, even though it is merely an equilibrium model within an urban land-use setting. Even in the absence of other substantial value factors, an urban area might be segmented if the preferences and income of the households differ according to space and accessibility. The more recent land-use/environmental-preferences approach (Evans, Richardson, Wheaton) can be used to explain the occurrence of sub-markets. In addition, segmentation within an urban area may be based on a variety of other factors, including the dominant type of building, plot efficiency or the internal attributes of the dominant type of apartment (see Laakso, 1997; Bourassa et al., 1997; Grigsby et al., 1987). In general, such models would predict that differences across dwellings and locations would be so small that they would disappear over time due to the arbitrage of prudent consumers.

Both Helsinki and Amsterdam fit partly within these rather one-dimensional models of urban housing market structure, although Helsinki fit better 1993 than it did in 2001. In reality, however, individual buyers have differing housing preferences and face a variety of housing alternatives, which might not comprise a single market. MacLennan and Tu (1996) therefore suggested

---

a non-coordinated or partly coordinated view as an alternative to the dominant 'unitary-equilibrium' approach, arguing that attempts to model housing markets within an instantaneous-equilibrium model are futile. Because such frameworks focus on processes of adjustment rather than on 'standard outcome' data, they produce a 'persistent localised disequilibrium' caused by both spatial and sectoral factors combined with diversification of either supply or demand. The segments are defined according to the total vector of market characteristics of the location and housing bundle, rather than by price levels alone. The information from Helsinki fits particularly well in this regard, as suburban locations clearly may be associated with both price premiums and discounts. Temporal dynamics, however, have also affected the picture. In 2001, neighbourhood and size (and probably price) mattered more than they had in 1993.

Amsterdam is perhaps too heterogeneous to be able to arrive at this general conclusion, although both ends of the price range in this area are situated within multiple clearly defined types of location. For example, two areas (*Middenmeer* and *Willemspark*) apparently belong to completely different market segments, even though they both have high price and income levels and few children. *Middenmeer* is populated primarily by elderly couples within a low (area) density setting (in *Watergraafsmeer*, a child-friendly area south-east of the city centre). In contrast, the population of *Willemspark* consists of singles between the ages of 25 and 44 years old within a high-density setting (*Oud-Zuid*). The multi-equilibrium model, therefore, is probably also valid in Amsterdam (cf. affluent, suburban *Westend* and affluent, urban *Eira* in Helsinki). The fact that such factors as density proxies and canal situation have nonlinear and diversified price associations provides further support for this conclusion. Temporal dynamics matter here as well. The areas have differentiated according to price level, urbanisation and density of location, as well as the quality of the houses (also situation and maintenance), but some of the criteria (e.g., building age and district location) have decreased in importance. Furthermore, the urban/suburban dimension and the single-family/multi-storey dimension remain strong submarket criteria in themselves, even without the link to price formation.

Recent theoretical improvements have emerged from socio-cultural and actor-led institutionalist positions. One distinct and relevant characteristic of the equilibrium-based approaches that are outlined above is that, according to these approaches, housing-market behaviour cannot be explained solely by a set of 'objectively' measured variables, as the factors that really matter are dependent on differentiated tastes and lifestyles. Unfortunately, the predominantly quantitative, aggregated and explorative approach that has been applied in this report can allow only brief discussion of this demand-side aspect. It is reasonable to assume, however, that additional data collected from questionnaire surveys (e.g., the Housing Demand Survey [WBO] in the

---

Netherlands) could allow the use of this theoretical cluster as a background in the search for plausible explanations for particular housing-market structures. The application of the AHP in this study yielded somewhat unexpected results.

Finally, it is important to note that this is the view and analysis of one researcher, who has applied a ‘fundamentally modified neoclassical perspective’ and multidimensional approach to the analysis of local housing markets, house prices and the locational preferences of housing consumers. The study began with a justification for the approach, and it is convenient to conclude in the same way. The goal of the study was to create a measure for segmentation. As no single approach can cover all of the aspects that are involved, the need for a variety of data and approaches is clear. In area assessments, the SOM-LVQ classification approach represents the ‘middle level’ with regard to trade-offs between feasibility, model performance and conceptual soundness. In other words, it lies between standard simple equilibrium frameworks and complex behavioural-institutional frameworks.

The approach that is advocated in this study is largely inter-disciplinary, borrowing concepts from real estate, housing economics, spatial planning, computer simulation and behavioural science. While no exactly comparable previous research on the topic exists to provide an anchor for the study, three different literatures have been consulted. The literature on SOM applications in property valuation applies a point-specific approach. Although these methods are strongly inductive, they are too localised to be able to serve as models for this study on their own. Spatial housing-market theory was therefore examined as well, in both its urban and regional dimensions. Nonetheless, these methods are predominantly deductive and too formal for this purpose. Because of these restrictions, it was necessary to incorporate a third type of literature: cross-contextual/cultural housing research tradition. Although inductive and theoretic, this tradition does not usually address economic value. Because it operates at the interface of all three paradigms, this study is obviously unconventional. It nevertheless covers new ground, as suggested in the discussion on increased collaboration across scientific disciplines when studying housing-choice patterns, at least in the Netherlands (see e.g., Mulder & Dieleman, 2002).

### 7.3 Practical application

In today’s increasingly competitive context, market actors need useful tools to aid their decision-making with regard to the spatial and functional allocation of resources within urban housing markets. One of the goals of this research was therefore to transcend academic discourse, demonstrating relevance from the perspective of practice as well. We must therefore consider

---

**Table 7.1 The generic applicability categories of the SOM-LVQ related to housing market analysis**

Generic application	How to use the SOM-LVQ output map/matrix
Mass-appraisal	All resulting clusters/classes are identified and in the end assigned a value.
Evaluation of attractiveness	All resulting cluster/classes are identified and compared with each other, but only one particular type of them is assigned a value or action-plan as a basis for decisions about allocation.
Outlier analysis	Only one particular cluster/class (at the time) is identified and assigned an action plan for further processing; no comparison with the rest of the clusters/classes.

the applicability of the ideas that have been presented throughout this study to policy, planning or business situations.

The analyses have indeed opened up possibilities for policy application (an area that should be strengthened, according to Mulder and Dieleman). Even though no real spatial segregation could be observed in either Helsinki or Amsterdam, the issue of monitoring the processes of residential dynamics is highly relevant, for a variety of reasons.<sup>42</sup> Observing changes in the quality and the population of areas requires the analysis of multiple (one-year) cross-sections, allowing the identification of socio-demographic and price-related demand patterns. Given sufficient time between cross-sections, we may even be able to monitor changes in the physical environment.

In principle, such monitoring is possible, as feature maps enable convenient comparisons across locations and housing packages. These applications are aimed either at decision-making in investment analysis or aimed at public policy. Currently, however, (local) governments are able to act as decision-makers with profit-motivated strategies that are similar to those of private companies. It is thus possible to show three types of general purposes for the SOM-LVQ classifier: mass-appraisal, evaluation of attractiveness and outlier analysis (see Table 7.1).

The most obvious application is the mass appraisal of residential property. Mass appraisal is intended to allow an 'assembly-line' form of assessment for a large number of properties for the same purpose at a certain time. The issue of mass appraisal becomes relevant for municipalities or corporations when the single-property valuation approach is either too expensive or too time consuming (e.g., Spit & Needham, 1987). A number plots or buildings must be valued quickly, and there is a trade-off between saving resources and valuation accuracy. In some cases, the ground must be valued separately from the built structures. Because the emphasis is on larger residential areas rather than single houses or plots, the results of the study can be applied in urban planning as well. Mass-appraisal may be used for property and land tax assessments (e.g., Spit & Needham, 1987; Needham et al., 1998) or for determining changes in the existing portfolio, in which the value of a heterogeneous selection of houses is referred to as a 'portfolio'. This method could also be applied to municipal plot disposal (see Needham et al., 1993),

<sup>42</sup> On the other hand, 'segregation' is a more loaded concept than 'segmentation'; in the Netherlands, these two concepts belong to two different discussions.

---

which refers to the elicitation of weights for various locations and land uses within the municipality, such that the plots to be sold or leased are differentially priced according to size, internal quality and location.

The second type of application is comprised of several specific purposes. Investment appraisal (e.g., Barlowe, 1986; Balchin & Kieve, 1977) involves making decisions concerning whether an investment (in dwellings or land) is worth undertaking, or to choose the most profitable of several possible investment alternatives. For developers or building companies, the questions concern site selection for buildings. Property investors are interested in determining the various discount rates and periods of investment; investment periods increase with the attractiveness of an area. This application can be modified for the evaluation of the social and economic costs and benefits of a planning project (e.g., Larsson, 1995). A profit motive may also apply when a comparative evaluation of attractiveness is required for policy and planning purposes. Alternatively, if a local government adopts an 'anti-market policy', the goal may be the opposite: allocation of resources into unfavourable locations and housing packages. A related application is the evaluation of a built (up) area for urban management and territorial competition (e.g., van der Krabben & Lambooy, 1993; D'Arcy & Keogh, 1998). Market research is conducted by estate agents and other entities in order to map consumer preferences. Somewhat pragmatically, prospective purchasers or tenants state their explicit preferences as input for the system, after which operators provide the closest corresponding locations or combinations of characteristics as output.

The third type of application involves the use of SOM output to screen out particular rare or extreme cases (outliers) from the data set for further, possibly more qualitative analysis. Economic impact analysis for urban planning purposes involves the evaluation of the effect of such externalities as parks, metro, highway and shopping centres for the community in monetary terms, or the determination of the grounds and magnitude of compensation for property owners due to a source of nuisance. This approach is also useful for the identification of contexts for urban rehabilitation neighbourhoods. Further, this method could be incorporated into the tax-assessment application (see above) in order to identify non-market cases (i.e., properties that are hardly ever sold) that require particular treatment.

The method can thus be used as a tool to aid decision-making in (at least) three different contexts: relative appraisal of location and housing packages, policy formulation (either pro-market or anti-market) and the outlier analysis of certain cases that require separate treatment. The overall conclusion regarding the applicability of the SOM-LVQ method is that it can be used for the assessment of particular locations or locationally determined bundles of housing-related attributes. An empirical modelling approach to locational value formation would obviously be essential for improving policy decisions.

---



# References

- Aalbers, M., 2002, Researcher, University of Amsterdam, AME. Discussion 25-7.
- Aalbers, M.B., 2005, Place-based social exclusion: redlining in the Netherlands, in: *Area*, (37) 1, pp. 100-109.
- Aalbers, M.B. & R. Deurloo, 2003, Concentrated and Condemned? Residential Patterns of Immigrants from Industrial and Non-industrial countries in Amsterdam, in: *Housing, Theory and Society*, 20, pp. 197-208.
- Adair, A.S., J.N. Berry, & W.S. McGreal, 1996, Hedonic modelling, housing submarkets and residential valuation, in: *Journal of Property Research*, 13 (2) pp. 67-83.
- Ball, M.J. & R.M. Kirwan, 1977, Accessibility and Supply Constraints in the Urban Housing Market, in: *Urban Studies* 14, pp. 11-32.
- Ball, M., 1998, Institutions in British Property Research: A Review, in: *Urban Studies*, 35 (9), pp. 1501-1517.
- Beck, U., 1993, *Risk Society – Towards a New Modernity*, London, (Sage Publications).
- Boelhouwer, P.J., 2002, Segregation and problem accumulation in the Netherlands: the case of The Hague, in: *Urban Geography* 23 (6) pp. 560-580.
- Boelhouwer, P., 2004, **The disfunctionality of the Dutch housing market: exploding house prices versus falling housing production.** Paper presented at the ENHR conference Housing: Growth and Regeneration, Cambridge, UK, July 2-6.
- Bolt, G., P. Hooimeijer, & R. van Kempen, 2002, Ethnic segregation in the Netherlands: new patterns, new policies?, in: *Tijdschrift voor Economische en Sociale Geografie*, 93 (2), pp. 214-220.
- Borst, R.A., 1995, Artificial Neural Networks in Mass Appraisal, in: *Journal of Property Tax Assessment and Administration*, 1, pp. 5-15.
- Bourassa, S.C., F. Hamelink, M. Hoesli, & B.D. MacGregor, 1997, **Defining Residential Submarkets: Evidence from Sydney and Melbourne**, Real Estate Research Unit Working Paper series No. 3, University of Auckland, Department of Property.

Bourassa, S.C., F. Hamelink, M. Hoesli, & B.D. MacGregor, 1999, Defining Housing Submarkets, in: **Journal of Housing Economics**, 8, pp. 160-183.

Bourassa, S.C., M. Hoesli, & V. S. Peng, 2003, Do housing submarkets really matter?, in: **Journal of Housing Economics**, 12, pp. 12-28.

Bourassa, S.C., M. Hoesli, & J. Sun, 2004, What's in a view?, in: **Environment and Planning A** 36, pp. 1427-1450.

Bramley, G., 1999, Housing market adjustment and land-supply constraints, in: **Environment and Planning A**, vol. 31, pp. 1169-1188.

Bökemann, D. & W. Feilmayr, 1997, **Differences Among Social Groups in their Willingness to Pay for Distinct Real Estate Properties**. Conference paper. American Real Estate and Urban Economics Association 6th International Conference, Berkeley May 31-June 2.

Carlson, E., 1998, Real Estate Investment Appraisal of Land Properties using SOM, in: Deboeck G. and T. Kohonen (eds.), **Visual Explorations in Finance with Self-Organizing Maps**, pp. 117-127, New York (Springer).

Clapp, J.M., H-J. Kim, & A.E. Gelfand, 2002, Predicting Spatial Patterns of House Prices Using LPR and Bayesian Smoothing, in: **Real Estate Economics**, 30 (4), winter, pp. 505-532.

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.

Daly, J., S. Gronow, D. Jenkins & F. Plimmer, 2003, Consumer behaviour in the valuation of residential property. A Comparative study in the UK, Ireland and Australia, in: **Property Management**, 21 (5), pp. 295-314.

Deboeck, G. and T. Kohonen, (eds.), 1998, **Visual Explorations in Finance with Self-Organizing Maps**, New York (Springer).

Deurloo, R. & S. Musterd, 2001, Residential Profile of Surinamese and Moroccans in Amsterdam, in: **Urban Studies**, 38 (3), pp. 467-485.

De Vries, P., C.P. Dol & H.M.H. van der Heijden, 2003, Knelpuntenmonitor Woningbouw 2003 (in Dutch), **Bouwmarkt** 11, november, pp. 5-7.

---

Dieleman, F. M. & C. Wallet, 2003, Income differences between central cities and suburbs in Dutch urban regions, in: *Tijdschrift voor Economische en Sociale Geografie*, 94 (2), pp. 265-275.

Doling, J., 1990, Housing Finance in Finland, in: *Urban Studies*, 27 (6), pp. 951-969.

Dubin, R.A., 1992, Spatial autocorrelation and neighborhood quality, in: *Regional Science and Urban Economics*, 22, pp. 433-452.

Dubin, R., K.R. Pace & T.G. Thibodeau, 1999, Spatial Autoregression Techniques for Real Estate Data, in: *Journal of Real Estate Literature*, 7 (1), January, pp. 79-95.

Evans, A.W., 1985, *Urban Economics – an Introduction*, Oxford (Blackwell).

Glaeser, E.L., E.A. Hanushek & J.M. Quigley, 2004, Opportunities, race, and urban location: the influence of John Kain, in: *Journal of Urban Economics*, 56 (1), July, pp. 70-79.

Gram-Hanssen, K. & C. Bech-Danielsen, 2002, **Home-building and identity in detached houses – new ways to interpret resource consumption in households**, paper presented at the European Network of Housing Research conference workshops, Vienna, 1-5 July.

Grigsby, W., M. Baratz, G. Galster & D. Maclennan, 1987, The Dynamics of Neighbourhood Change and Decline, in: *Progress in Planning*, 28 (1), pp. 1-76.

Guy, S. & J. Henneberry, 2000, Understanding Urban Development Processes: Integrating the Economic and the Social in Property Research, in: *Urban Studies* 37 (13), pp. 2399-2416.

James, H., A. Collins & E. Lam, 1994, **The Principles and Practice of Artificial Neural Networks in Property Valuation Studies**, Discussion Paper, University of Portsmouth, Department of Land and Construction Management.

Jenkins, D.H., O.M. Lewis, N. Almond, S.A. Gronow & J.A. Ware, 1999, Towards an intelligent residential appraisal model, in: *Journal of Property Research*, 16 (1), pp. 67-90.

Jones, C., 2002, The Definition of Housing Market Areas and Strategic Planning, in: *Urban Studies*, 39 (3), pp. 549-564.

---

Jones, C., C. Leishman & C. Watkins, 1999, **Exploring linkages between housing submarkets: theory and evidence**. Conference paper, Cutting Edge, Cambridge, September 5-7.

Jones, C., C. Leishman & C. Watkins, 2003, Structural change in a local urban housing market, in: **Environment and Planning A**, 35, pp. 1315-1326.

Jones, C., C. Leishman & C. Watkins, 2004, Intra-Urban Migration and Housing Submarkets: Theory and Evidence, in: **Housing Studies**, 19 (2), pp. 269-283.

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** (Visualising differences in attractiveness between localities with a neural network; in Finnish), *Maanmittaus*, 72 (1-2), pp. 37-70.

Kauko, T., 1999, The Spatial Dimension of Residential Valuation, in: Artimo K. and Haggren H. (Eds.), **Spatial statistics in GIS applications**, Helsinki University of Technology, Carthography and Geoinformatics, pp. 85-110, Espoo.

Kauko, T., 2000, **Can housing market segmentation be captured with a neural network modelling approach?**, the ENHR Conference in Gävle 26-30 June (with YHR pre-conference seminar), published as a CD ROM.

Kauko T., 2001, **Modelling location in house prices - a literature review of the empirical research methodology**, Discussion paper. Utrecht: Urban Research centre Utrecht (URU).

Kauko, T., 2002, **Modelling the locational determinants of house prices: neural network and value tree approaches**, PhD Thesis. Utrecht (Labor Grafimedia), <http://www.library.uu.nl/decollectie/proefschriften/11688main.html>

Kauko, T., 2004, Towards the 4th generation – An essay on innovations in residential property value modelling expertise, in: **Journal of Property Research**, 21 (1), March, pp. 75-97.

Kauko, T., 2005, **The Budapest housing market structure from a heterodox economics perspective and with a neural network classification**, The 11th Annual Conference of the Pacific Rim Real Estate Society, January 23-26, 2005, Melbourne, Australia, <http://business2.unisa.edu.au/prres/>

Kauko, T. & H. Peltomaa, 1998, A Comparison Between Neural Network and Multiple Regression Approaches: Effects of Transmission Line on Residential Property Values, in: **Surveying Science in Finland**, 16 (1-2), pp. 52-87.

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

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

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

Kortteinen, M. & M. Vaattovaara, 1999, Huono-osaisuus pääkaupunkiseudulla 1980- ja 1990-luvuilla -käännepöytäkaupunkiseudun kehityksessä? (Distribution of the underprivileged in the metropolitan areas of Helsinki in the 1980's and 1990's – turn point in the development?; in Finnish), in: **Terra**, 111 (3), pp. 133-145.

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

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

Lam, E., 1994, **An Artificial Neural Network Approach to Residential Valuation: the Self-Organising Map**, Working Paper Nr. 1 in Adaptive Algorithms: University of Portsmouth, Department of Land and Construction Management.

Lankinen, M., 1997, Asuinpaikan valinta: status, etäisyys vai totunnaisuus? (in Finnish), in: Ilmonen, M., 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), in: **Pääkaupunkiseudun julkaisusarja B**, 2, pp. 28-51.

Leishman, C., 2001, House building and product differentiation: An hedonic price approach, in: **Journal of Housing and the Built Environment**, 16 (2), pp. 131-152.

Ley, D., J. Tutchener & G. Cunningham, 2002, Immigration, Polarization, or gentrification? Accounting for changing house prices and dwelling values in gateway cities, in: **Urban Geography**, 23 (8), pp. 703-727.

Luttik, J., 2000, The value of trees, water and open space as reflected by house prices in the Netherlands, in: **Landscape and Urban Planning**, 48, pp. 161-167.

Maclennan, D., 1977, Some Thoughts on the Nature and Purpose of House Price Studies, in: **Urban Studies**, 14, pp. 59-71.

Maclennan, D., 1982, **Housing economics**, New York (Longman).

Maclennan, D. & Y. Tu, 1996, Economic Perspectives on the Structure of Local Housing Systems, in: **Housing Studies**, 11 (3), pp. 387-406.

Maclennan, D. & C. Whitehead, 1996, **Editorial: Housing Economics – An Evolving Agenda**, 11 (3), pp. 341-344.

Majoor, S., 2002, **Multiple Intensive Land Use as an Urban Strategy: Lessons from Amsterdam**, NETHUR school Land use planning and policy, 11 October, Utrecht.

Martin, D., 1998, Automatic Neighbourhood Identification from Population Surfaces, in: **Computers, Environment and Urban Systems**, 22, pp. 107-120.

Maury, J., 1997, Ilmakuva pääkaupunkiseudun asuinalueiden sosiaalisista eroista 1990-luvun alussa. (An aerial view of the social differences in the Helsinki region; in Finnish), Ilmonen, M., 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), in: **Pääkaupunkiseudun julkaisusarja B**, 2, pp. 55-87.

McCluskey, W.J. & S. Anand, 1999, The application of intelligent hybrid techniques for the mass appraisal of residential properties, in: **Journal of Property Investment and Finance**, 17 (3), pp. 218-238.

Meen, G., 2001, **Modelling Spatial Housing Markets: Theory, Analysis and Policy (Advances in Urban and Regional Economics, V.2)**, USA (Kluwer).

Morrison, P.S. & S. McMurray, 1999, The Inner-city Apartments versus the Suburb: Housing Sub-markets in a New Zealand City, in: **Urban Studies**, 36 (2), pp. 377-397.

---

Mulder, C.H. & F.H. Dieleman, 2002, Living arrangements and housing arrangements: Introduction to the special issue, in: **Journal of Housing and the Built Environment**, 17 (3), pp. 209-213.

Musterd, S. & R. Deurloo, 2002, Unstable Immigrant Concentrations in Amsterdam: Spatial Segregation and Integration of Newcomers, in: **Housing Studies**, 17 (3), May, pp. 487-503.

Musterd, S. & N. Smakman, 2000, Integration of Caribbean Immigrants in a Welfare State City: Surinamese and Antilleans in Amsterdam, in: **International Journal of Population Geography**, 6, pp. 303-320.

Musterd, S. & I. van Zelm, 2001, Polycentricity, Households and the Identity of Places, in: **Urban Studies**, 38 (4), pp. 679-696.

Needham, B., M. Franke & P. Bosma, 1998, How the city of Amsterdam is using econometric modelling to value real estate, in: **Journal of Property Tax Assessment and Administration**, 3 (2), pp. 25-46.

Nelson, K. & R.R. Nelson, 2002, On the nature and evolution of human know-how, in: **Research Policy**, 31, pp. 719-733.

**NUL20**, 2003, Tijdschrift voor Amsterdams woonbeleid (in Dutch), 10, September.

Openshaw, S., M. Blake & C. Wymer, 1994, **Using neurocomputing methods to classify Britain's residential areas**, Working paper 94/17, School of Geography, University of Leeds.

Openshaw, S., 1998, Neural network, genetic, and fuzzy logic models of spatial interaction, in: **Environment and Planning A**, 30, pp. 1857-1872.

Pavlov, A.D., 2000, Space-Varying Regression Coefficients: A Semi-parametric Approach Applied to Real Estate Markets, in: **Real Estate Economics**, 28 (2), pp. 249-283.

Raju, K.A., P.K. Sikdar & S.L. Dhingra, 1998, Micro-simulation of residential location choice and its variation, in: **Computers, Environment and Urban Systems** 22, pp. 203-18.

Richardson, H.W., 1977, **The New Urban Economics**, London (Pion).

---

Ricketts, M., 2003, Alternative Explanations for Changes in Ownership Structure, in: **Journal of Institutional and Theoretical Economics**, 159 (4), pp. 688-697.

Rothenberg, J., G. Galster, R. Butler & J. Pitkin, 1991, **The maze of urban housing markets: theory, evidence and policy**, Chicago (University of Chicago Press).

Saaty, T. L., 1990, How to make a decision: The Analytic Hierarchy Process, in: **European Journal of Operational Research**, 48, pp. 9-26.

Schwann, G.M., 1998, **Distinguishing Residential Submarkets**, Conference paper, ERES & AREUEA Conference, Maastricht, June 10-13.

Sharkawy, M.A. & S. Chotipanich, 1998, Housing Segmentation in Developing Countries in Transition: A Recent Case Study of Residential Development in Bangkok, in: **Journal of Real Estate Literature**, 6, pp. 29-42.

Shiller, R. J., 2003, From Efficient Markets Theory to Behavioral Finance, in: **Journal of Economic Perspectives**, 17 (1), pp. 83-104.

Siikanen, A., 1992, Asuntojen kysyntä, tarjonta ja alueellinen erilaistuneisuus (in Finnish), **Asuntohallitus, tutkimus- ja suunnitteluosasto, asuntotutkimuksia** 4, Helsinki.

Sluis, R.-J. & T. Kauko, 2003, **Local housing market activity and planning practice**, ENHR-workshops, Tirana, Albania, May 26-30.

Spit, T. & B. Needham, 1987, A model of house prices in a Dutch city, in: **Netherlands Journal of Housing and Environmental Research**, 2 (1), pp. 53-60.

Strand, J. & M. Vågnes, 2001, The relationship between property values and railroad proximity: a study based on hedonic prices and real estate brokers appraisals, in: **Transportation**, 28, pp. 137-156.

Teune, W., 2002, Research Officer, SWD Amsterdam. Discussion 25-7.

Theebe, M., 2002, **Housing Market Risks**, UvA Research Series (Tinbergen Institute).

Timmermans, H., E. Molin & L. van Noortwijk, 1994, Housing choice processes: stated versus revealed modelling approaches, in: **Netherlands Journal of Housing and the Built Environment**, 9 (3) pp. 215-227.

Tu, Y., 2003, Segmentation, Adjustment and Disequilibrium, in: O'Sullivan, T. & K. Gibb, (Eds.): **Housing Economics & Public Policy**, (Real Estate Issues), pp. 38-55, Blackwell & RICS Foundation.

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, 7.

Vaattovaara, M., 2002, Future Development of Residential Differentiation in the Helsinki Metropolitan Area: Are We Following the European Model?, in: **Yearbook of Population Research in Finland**, 38, pp. 107-123, Helsinki, (The population research institute).

Van Duren, A.J., 1992, Changes in the Attraction of Amsterdam City, in: **Built Environment**, 18 (2), pp. 123-137.

Van Kempen, R. & H. Priemus, 1999, Undivided Cities in the Netherlands: Present Situation and Political Rhetoric, in: **Housing Studies** 14, pp. 641-657.

Wassenberg, F. & R. van Kempen, 2004, The Netherlands. Modernist housing in a developed welfare state, in: Turkington, R., R. van Kempen, & F. Wassenberg (eds.), **High-rise housing in Europe. Current trends and future prospects**, pp. 129-146, Housing and Urban Policy Studies, 28, Delft (Delft University Press).

Watkins, C. A., 2001, The definition and identification of housing submarkets, in: **Environment and Planning A**, 33, pp. 2235-2253.

Worzala, E., M. Lenk & A. Silva, 1995, An Exploration of Neural Networks and Its Application to Real Estate Valuation, in: **The Journal of Real Estate Research**, 10 (2), pp. 185-201.

Yoon Y., G. Swales Jr. & T.M. Margavio, 1993, A Comparison of Discriminant Analysis versus Artificial Neural Networks (Stock Price Application), in: **Journal of the Operational Research Society**, 44, pp. 51-60.

Zahedi, F., 1986, The Analytic Hierarchy Process – A Survey of the Method and its Applications, in: **Interfaces**, 16 (4) July-August, pp. 96-108.

**Internet:**

[www.stat.fi](http://www.stat.fi) - homepage of Statistics Finland (Tilastokeskus).



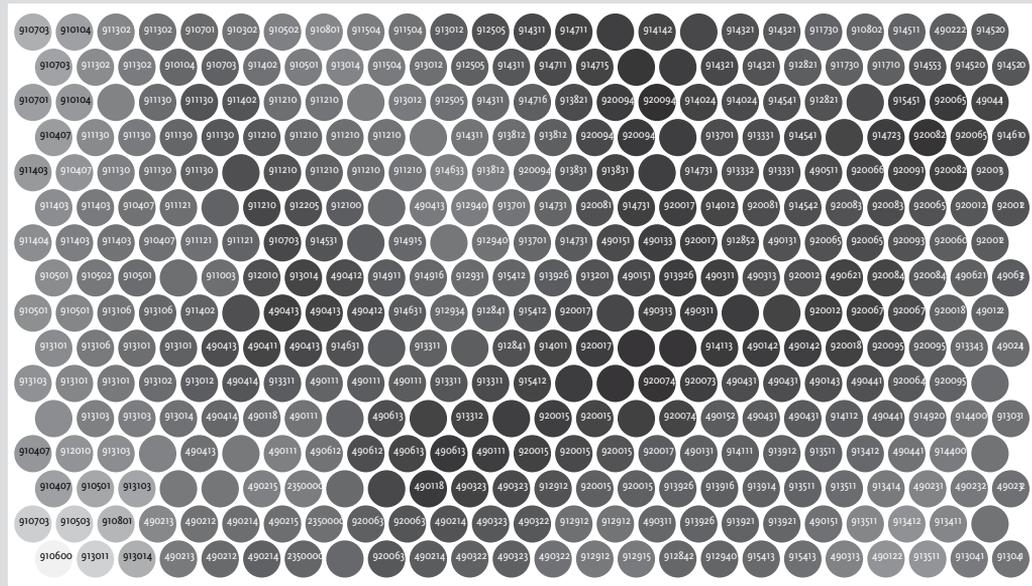
---

## Appendix A **Feature map layers of Helsinki housing markets, 1993 data**

How to interpret the maps in Appendix A, B, D, E, F and G?

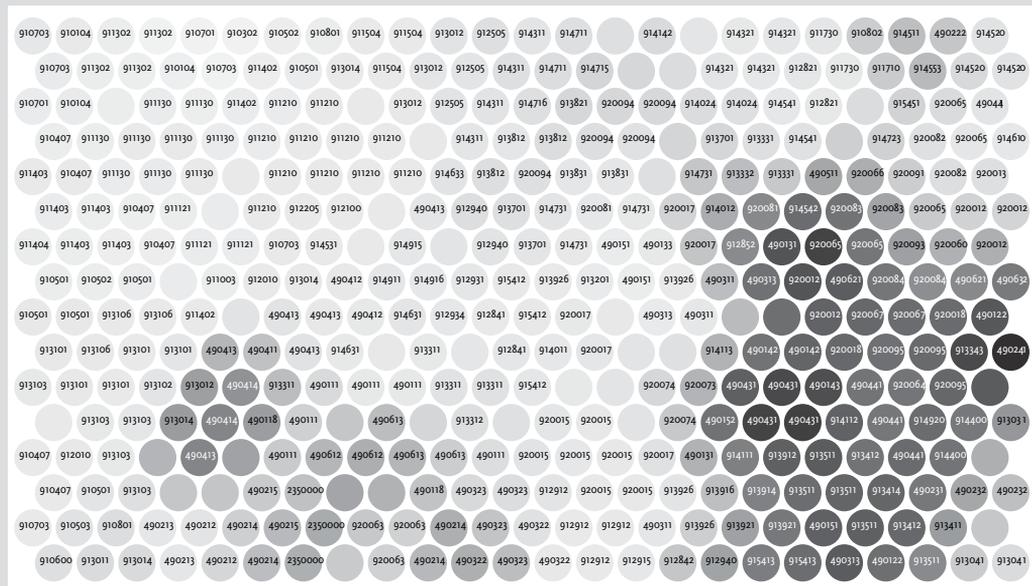
- Each neuron in the map represents a combination of attribute levels for all input variables (i.e., map layers).
  - For one and the same map, the map layers can be compared.
  - The position of each neuron is fixed across the map layers.
  - The greyscale indicates the intensity of the given variable.
  - The label is only for identification, and refers to a location.
-

## Feature map layer A1 Price levels



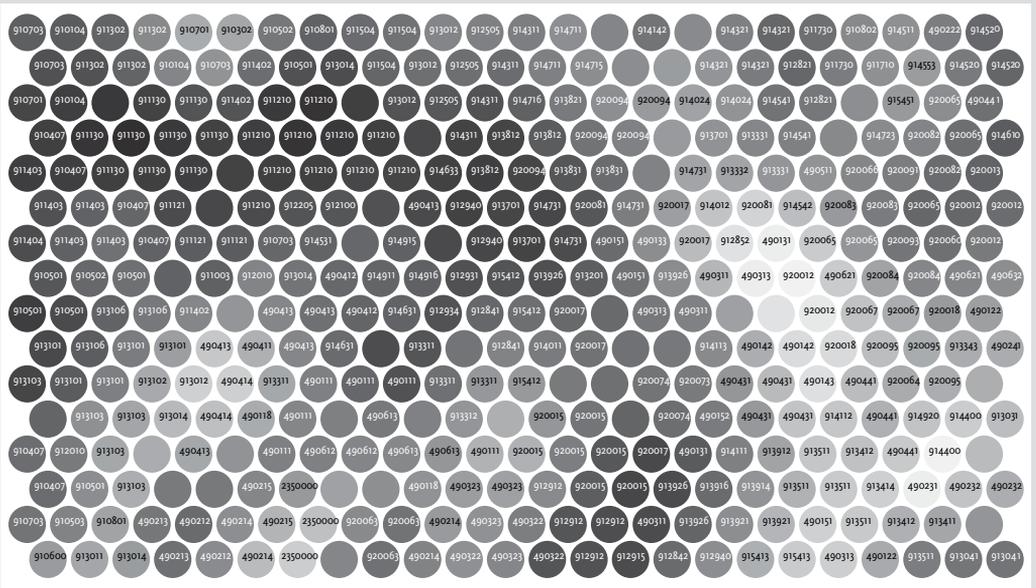
dark = cheap area; light = expensive area

## Feature map layer A2 Dwelling format



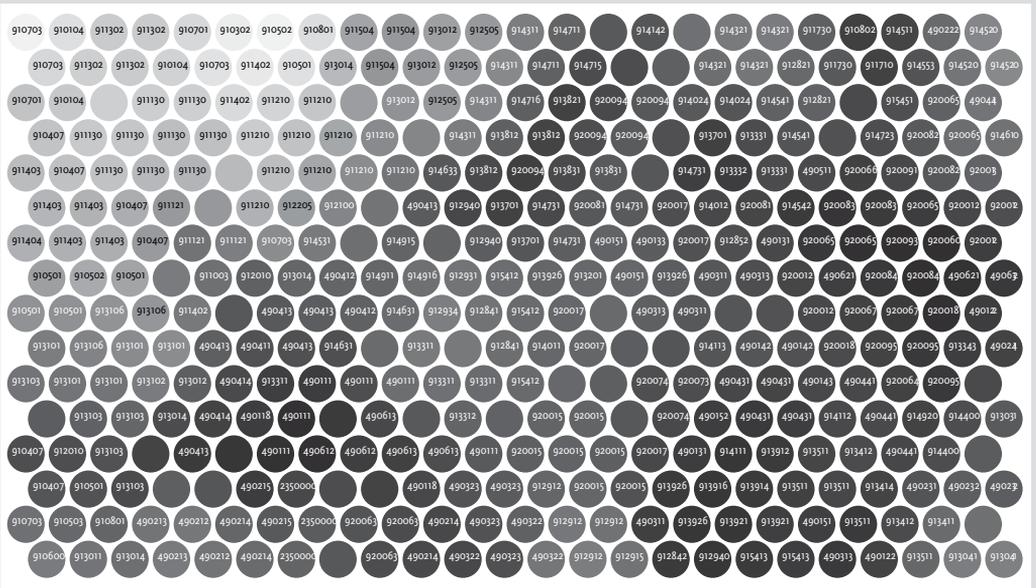
dark = single-family houses; light = multi-storey houses

### Feature map layer A3 Number of rooms



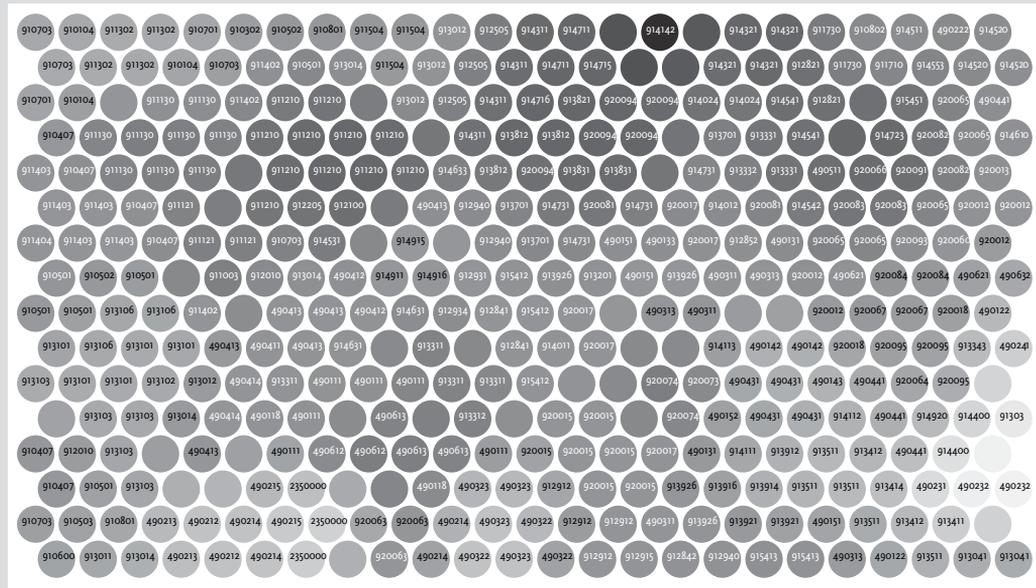
dark = 1-2 rooms; light = 3+ rooms

### Feature map layer A4 Building age



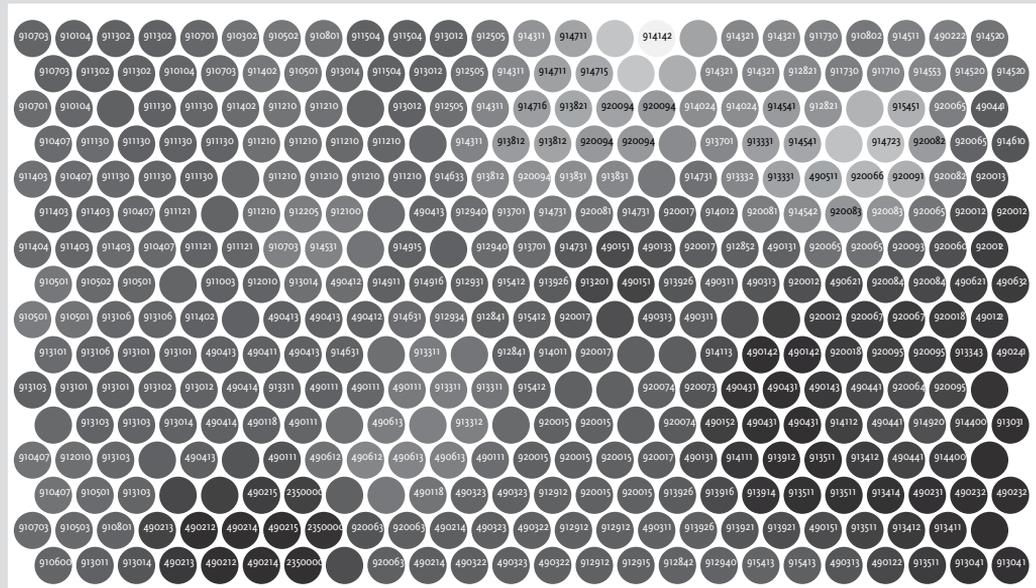
dark = new buildings; light = old buildings

### Feature map layer A5 Status



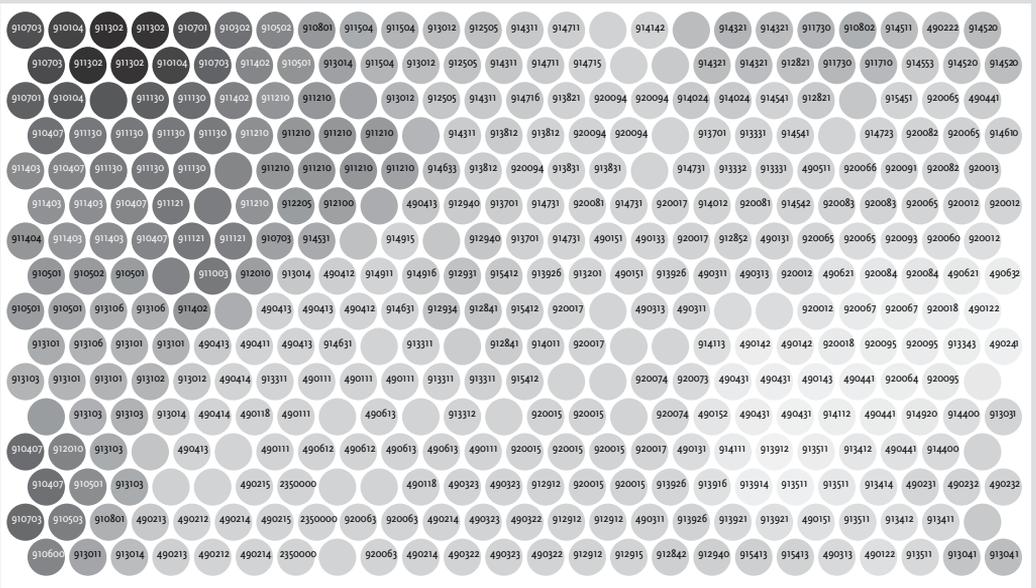
dark = low status; light = high status

### Feature map layer A6 Negative social externalities



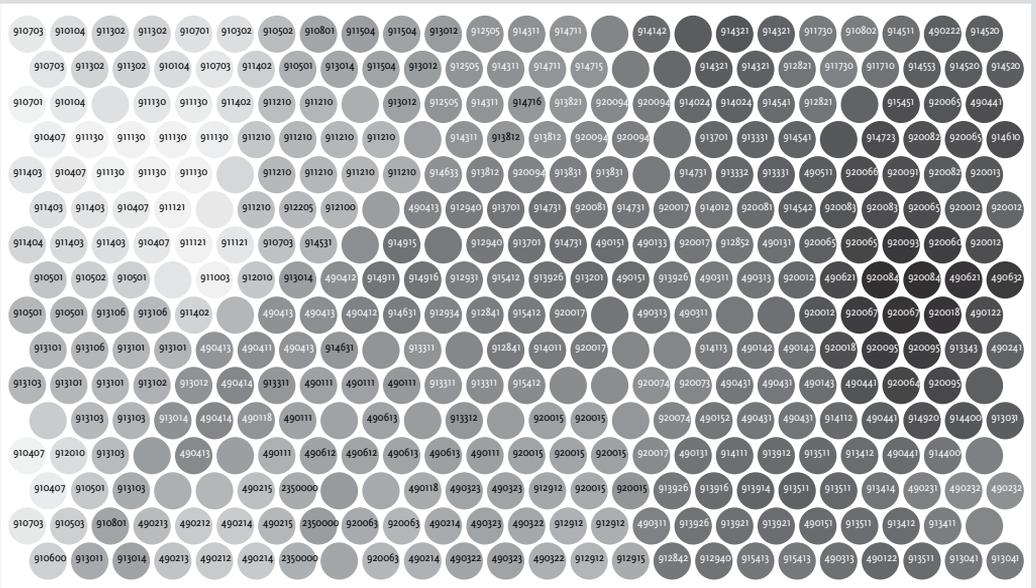
dark = low levels; light = high levels

### Feature map layer A7 'Urban' indicator



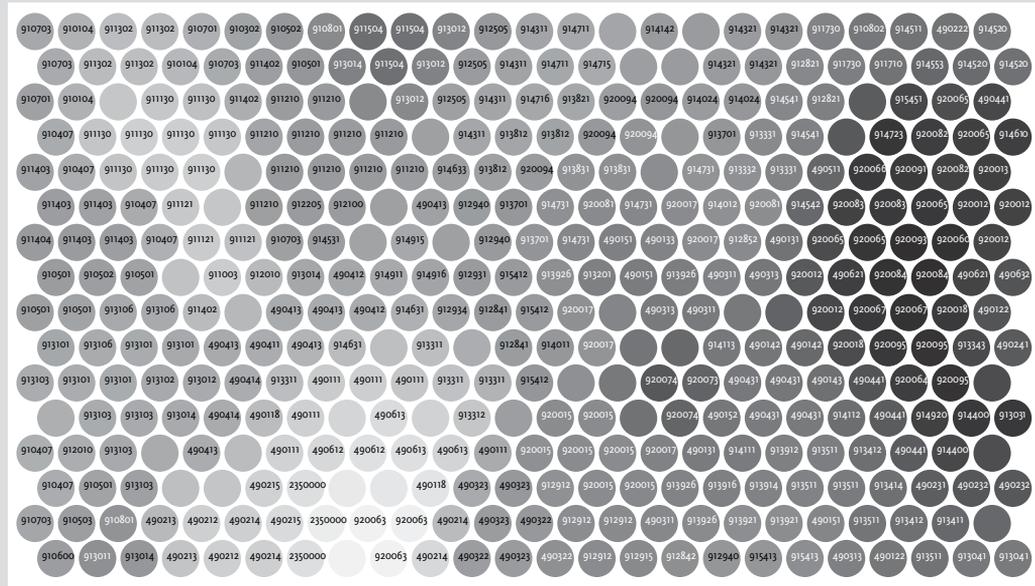
dark = most urban areas; light = least urban areas

### Feature map layer A8 Commercial services



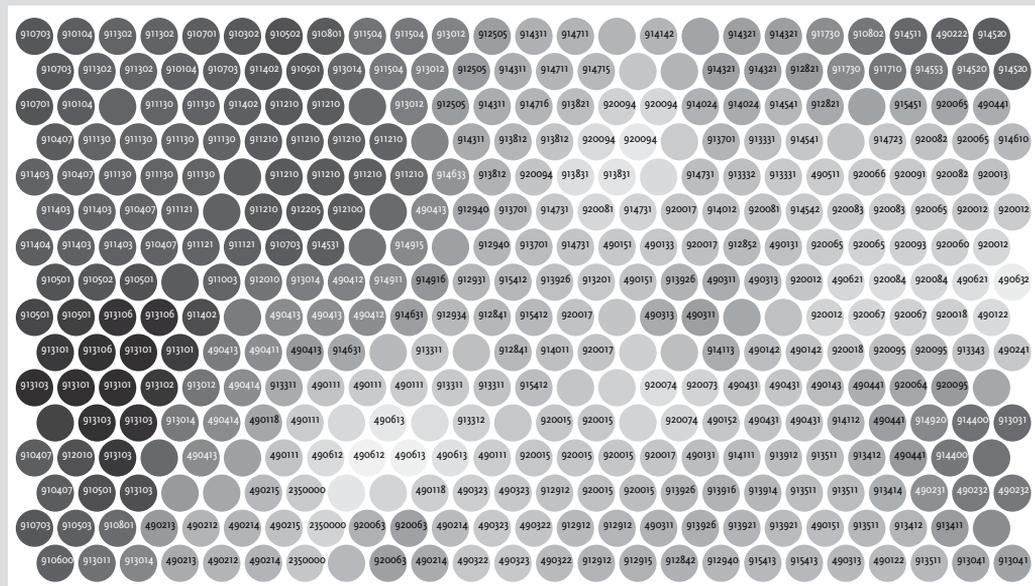
dark = low levels of service; light = high levels of service

## Feature map layer A9 Public services



dark = low levels of service; light = high levels of service

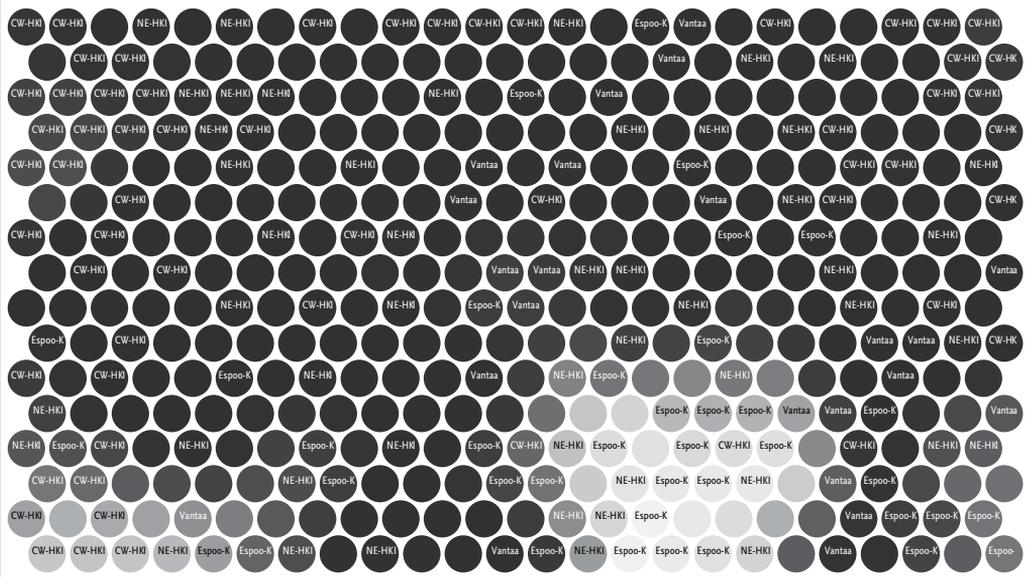
## Feature map layer A10 'Open space' indicator



dark = least undeveloped land in the area; light = most undeveloped land in the area

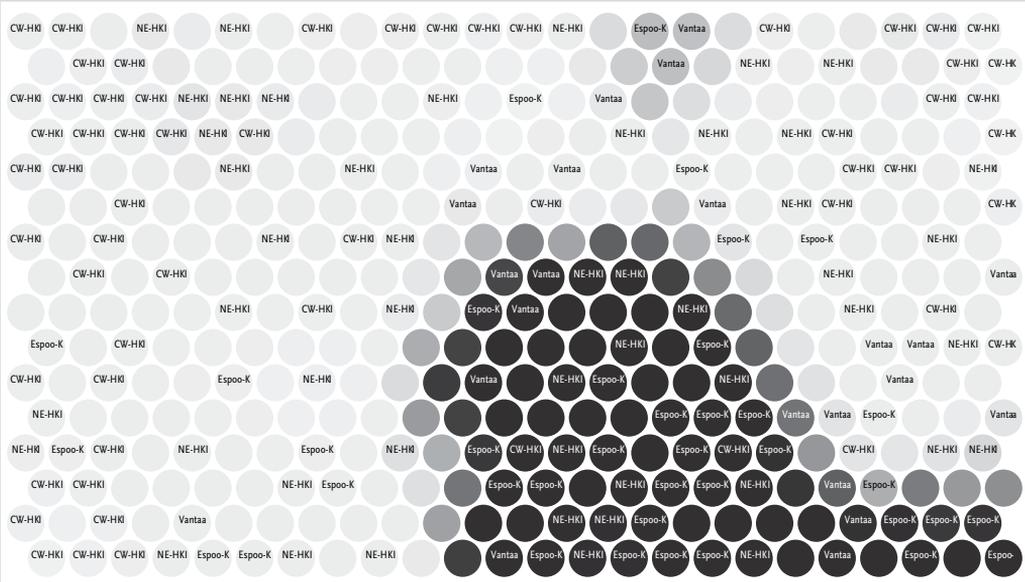
Appendix B **Feature map layers of  
Helsinki housing  
markets, 2001 data**

### Feature map layer B1 Dwelling size (sq. m.)



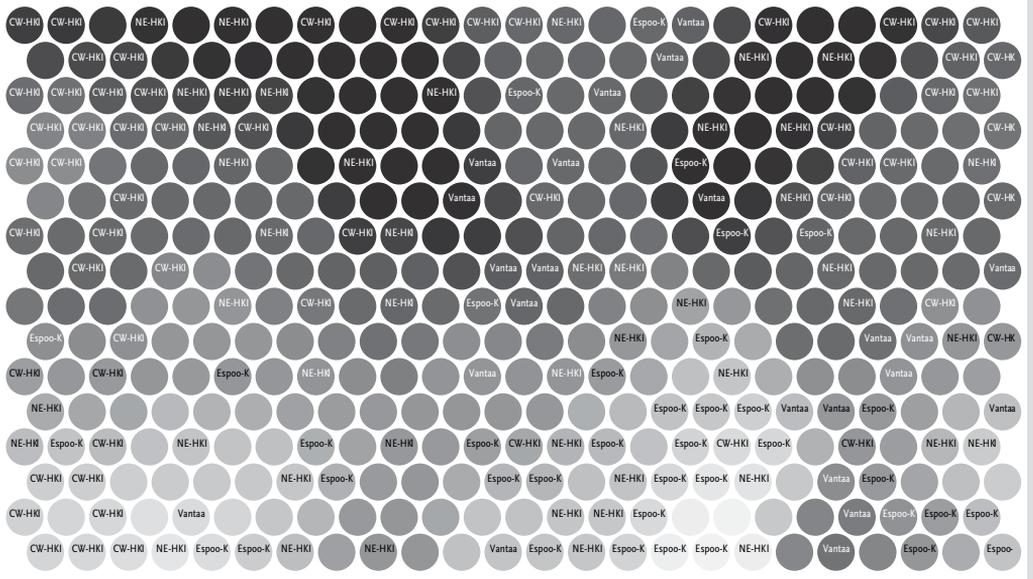
dark = small dwellings; light = large dwellings

### Feature map layer B2 Dwelling format



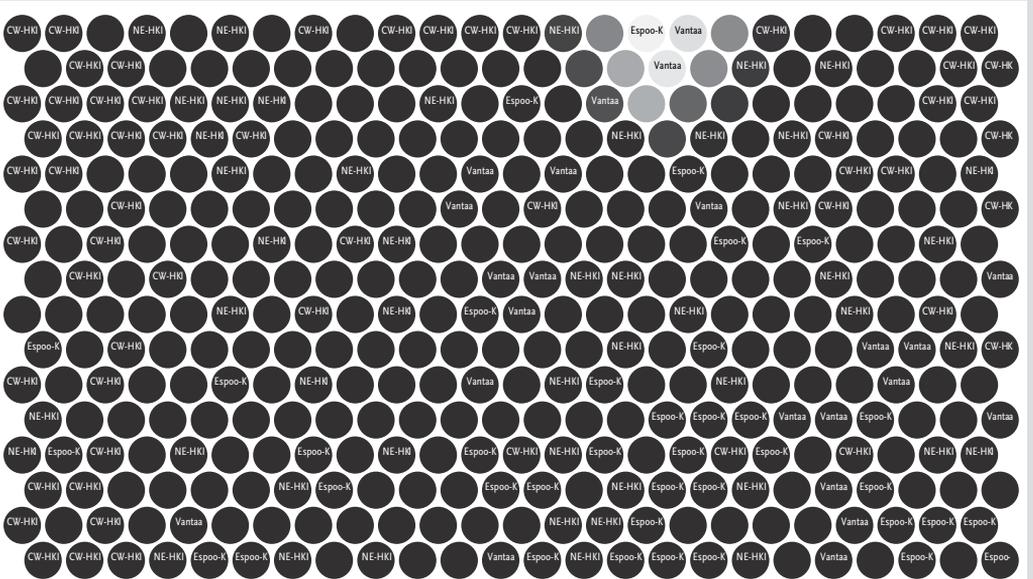
dark = Single-family houses; light = Multi-storey apartments

### Feature map layer B3 Number of rooms



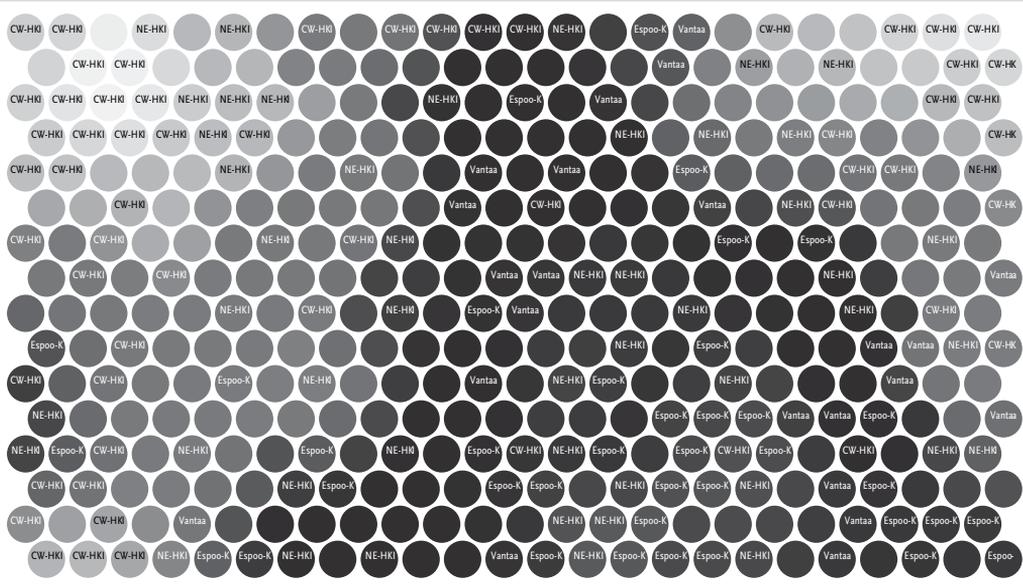
dark = 1-2 rooms; light = 3+ rooms

### Feature map layer B4 Subsidy



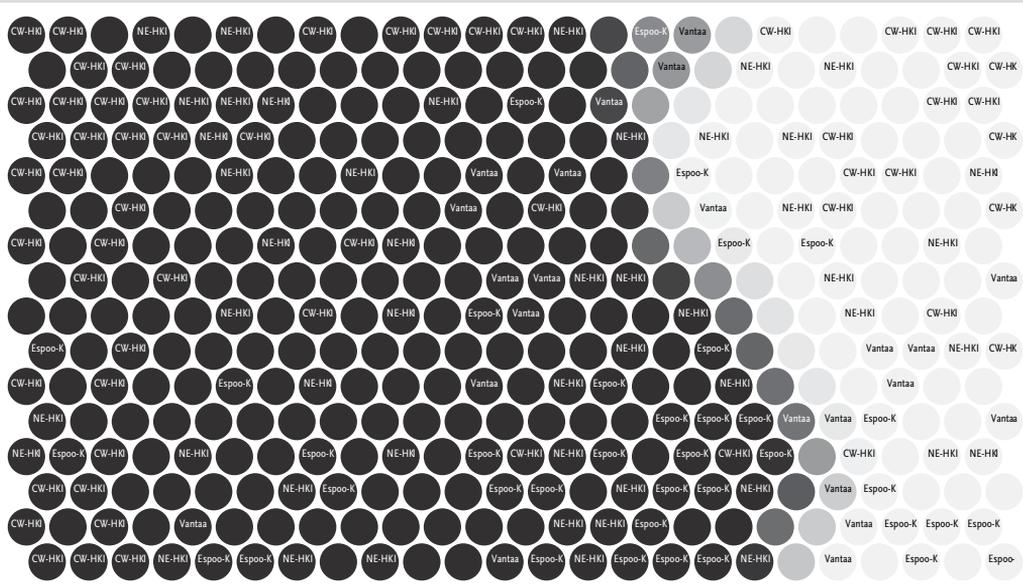
dark = no; light = yes

### Feature map layer B5 Building age



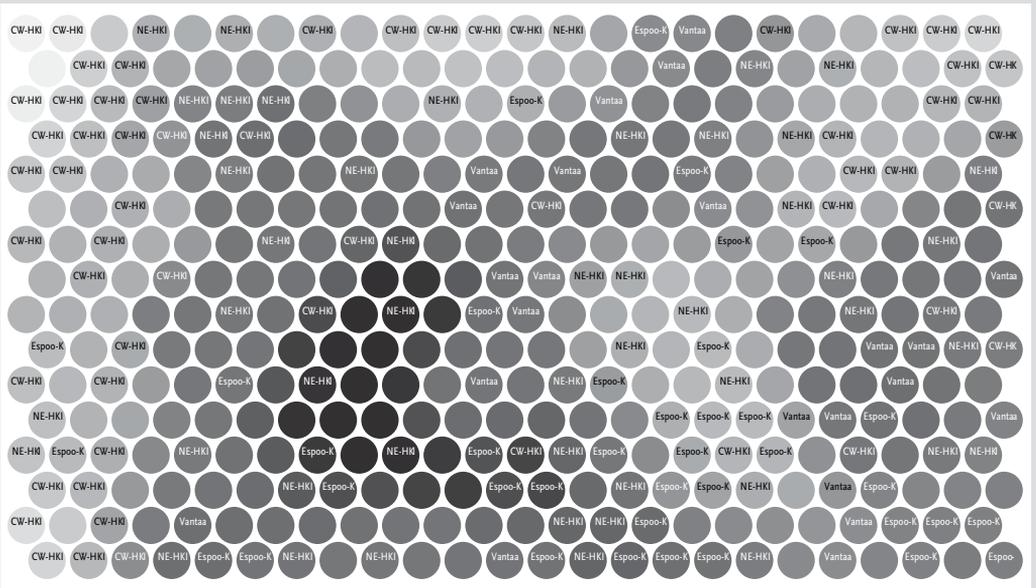
dark = new buildings; light = old buildings

### Feature map layer B6 Starter



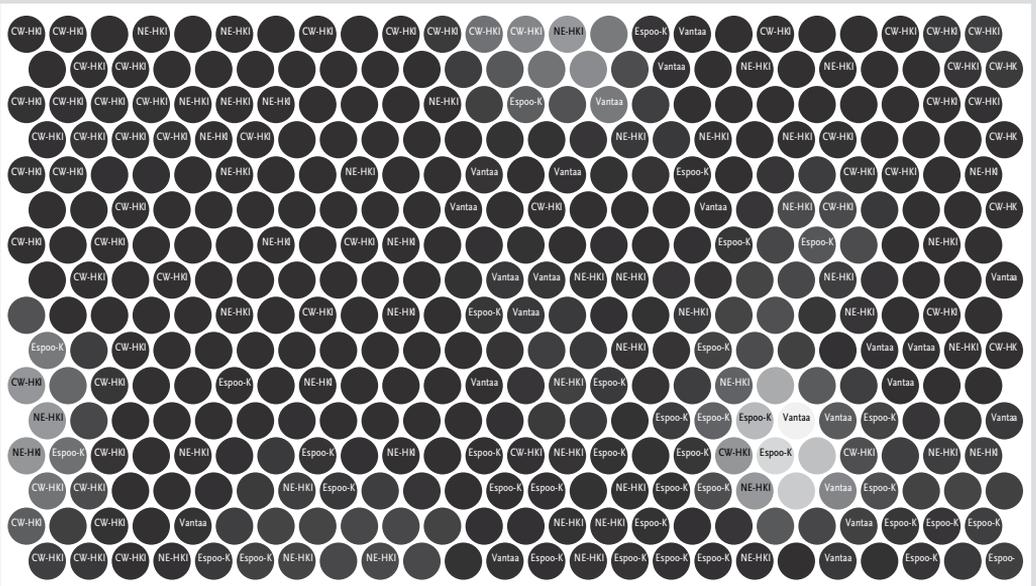
dark = no; light = yes

### Feature map layer B7 Price per sq. m.



dark = cheap; light = expensive

### Feature map layer B8 Mortgage



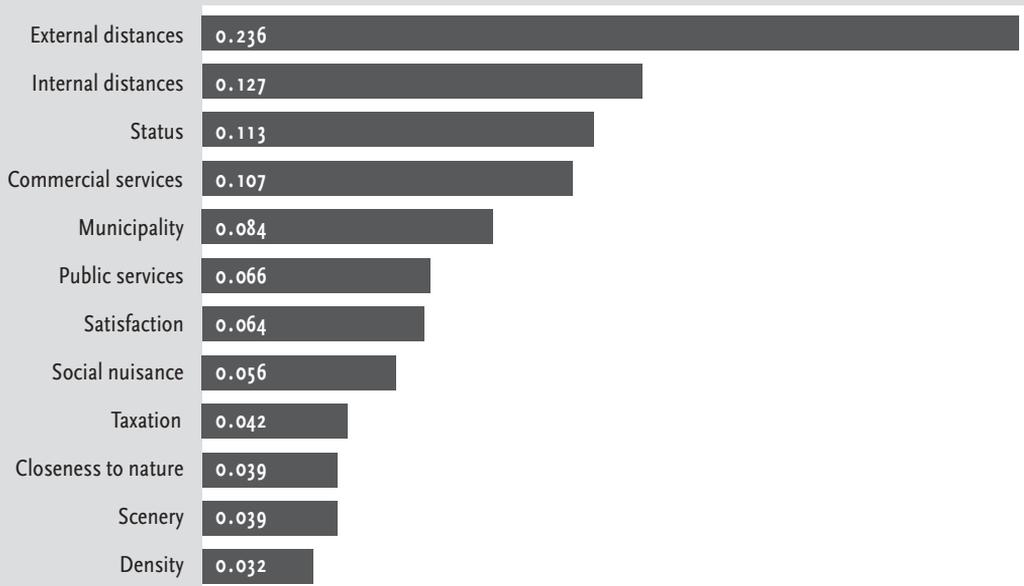
dark = no; light = substantial amount



## Appendix C Detailed level of aggregated and disaggregated AHP models for Metropolitan Helsinki

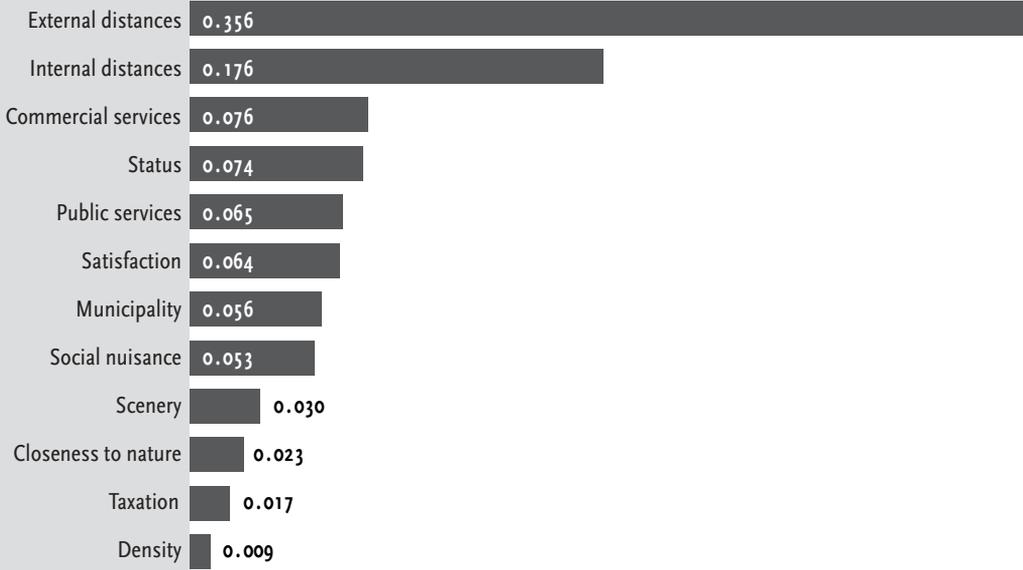
### C.1 The relative importance of the detailed locational attributes for multi-storey apartments

Figure C1.1 The aggregated model for all 22 respondents

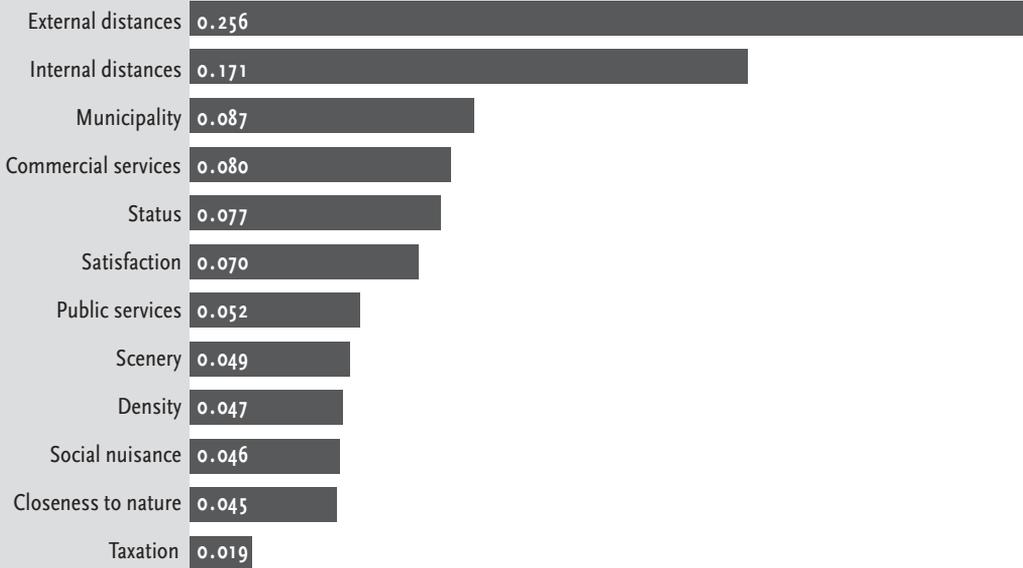


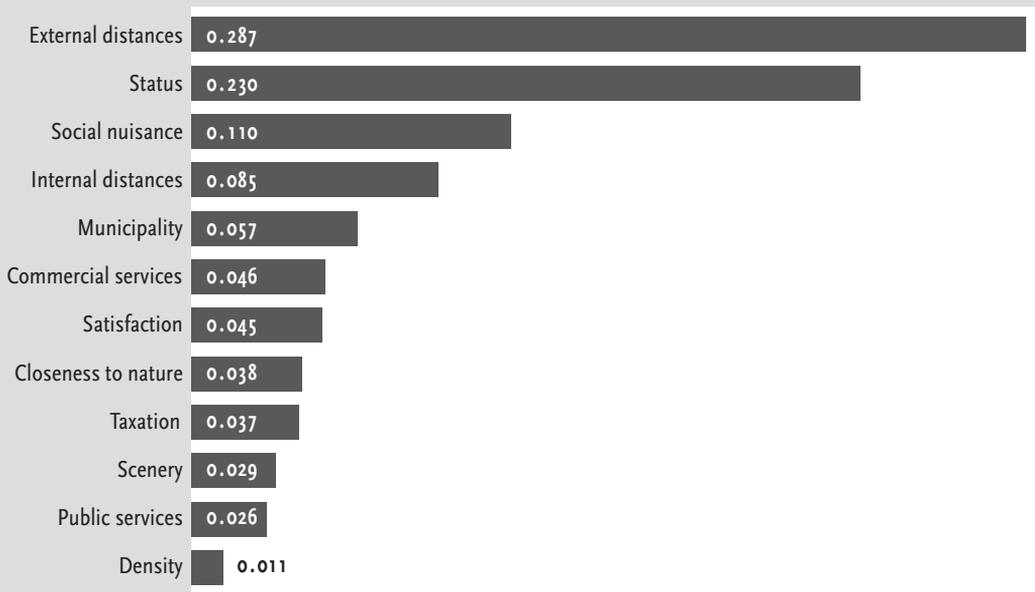
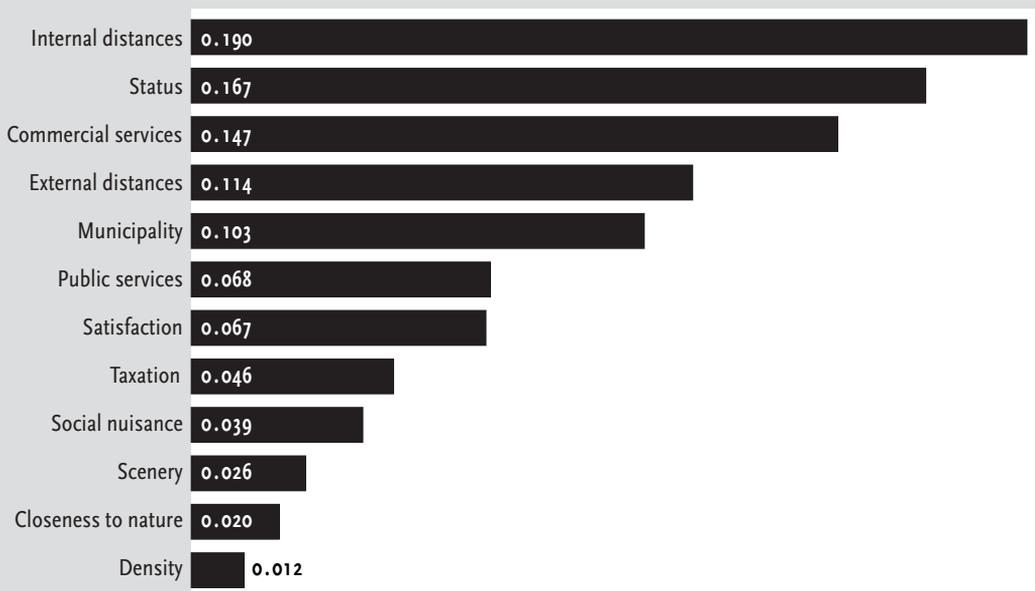
Closeness to nature	closeness of the area to nature
Commercial services	postoffice, bank, grocery store, bar etc.
Density	building efficiency in the area: scarce density is preferred
External distances	external distances/accessibility out to the area: distance to CBD, public transportation, journey to work
Internal distances	internal distances/accessibility within the area: to comprehensive school, services, parks, seaside etc.
Scenery	the aesthetic values of the area
Municipality	other municipality level factors: attractiveness, price-level, employment etc.
Public services	social, culture, sport/recreation, health, school, maintenance services
Satisfaction	other aspects of satisfaction with living in the area: homogeneity, pollution, safety, own character etc.
Social nuisance	the social nuisance/disturbance factors of the area: unemployment, social housing estates, crimes etc.
Status	the status of the area: level of income, education, share of owner-occupied housing
Taxation	the level of taxation: municipal income tax rate, property tax rate

**Figure C1.2 Dis-aggregated model for group Ia (7 respondents)**

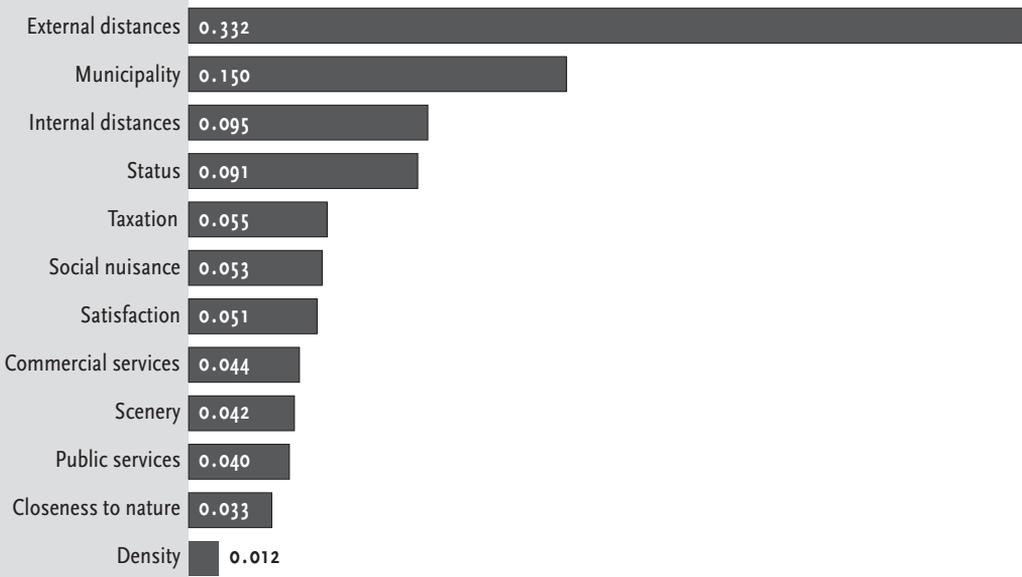


**Figure C1.3 Dis-aggregated model for group Ib (5 respondents)**



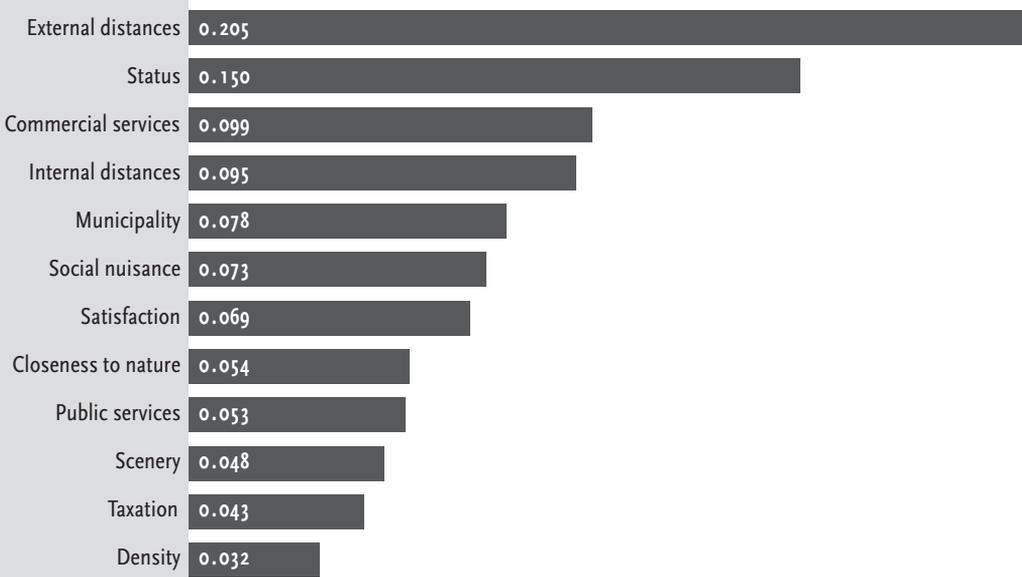
**Figure C1.4 Dis-aggregated model for group II (3 respondents)****Figure C1.5 Dis-aggregated model for group III (7 respondents)**

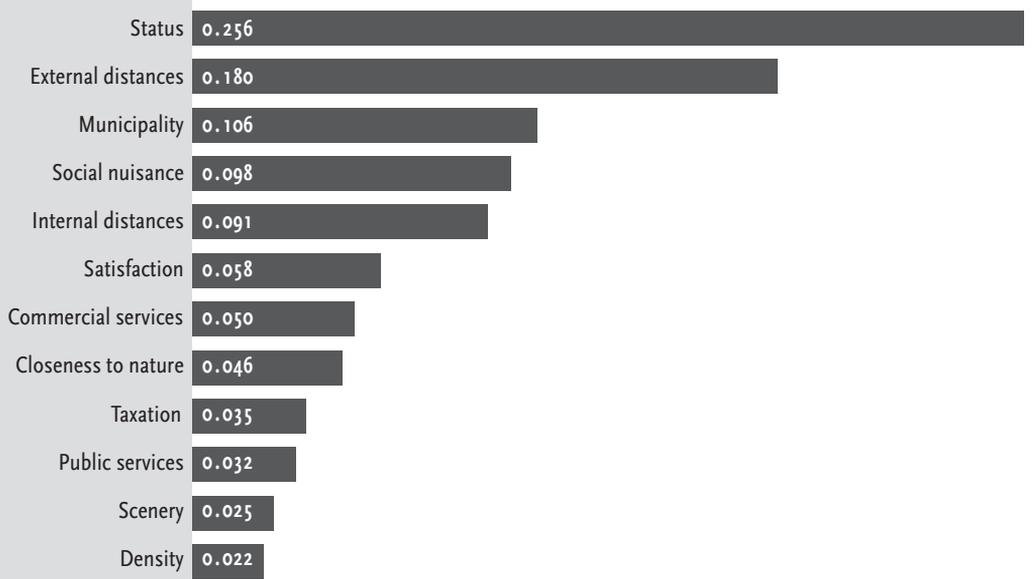
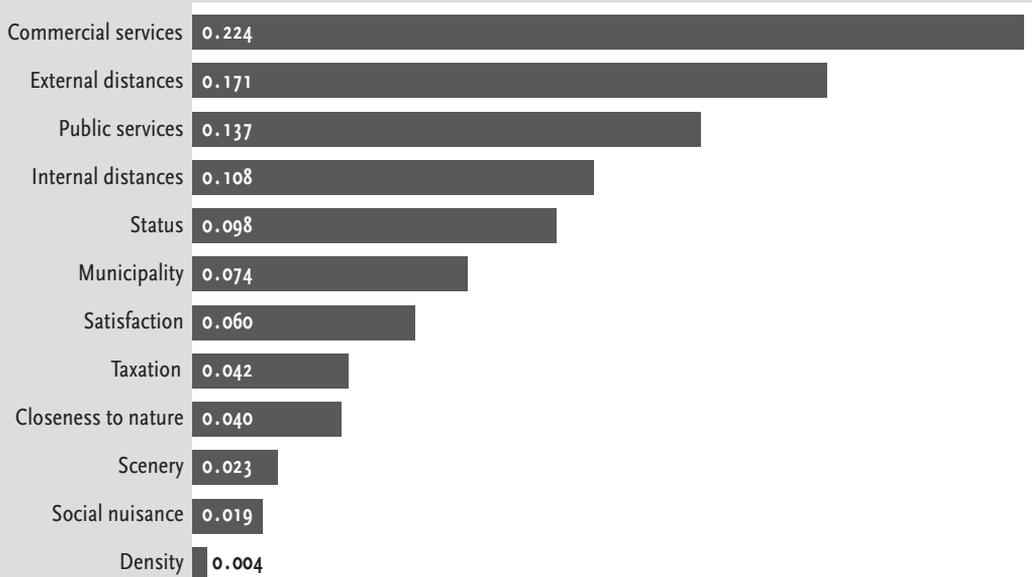
**Figure C1.6 Aggregated model with a 10% cut-off rule in inconsistency (3 respondents)**



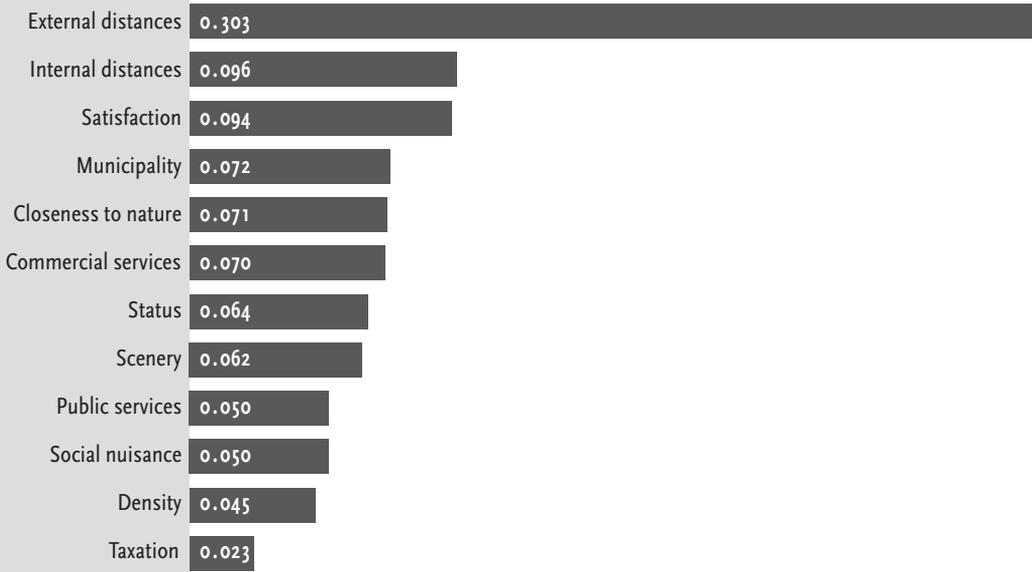
## C.2 The relative importance of the detailed locational attributes for single-family houses

**Figure C2.1 The aggregate model for all 22 respondents**



**Figure C2.2 Dis-aggregated model for group I (10 respondents)****Figure C2.3 Dis-aggregated model for group II (4 respondents)**

**Figure C2.4 Dis-aggregated model for group III (8 respondents)**



**Figure C2.5 Aggregated model with a 10% cut-off rule in inconsistency (4 respondents)**

