SELF-ORGANIZING MAP ALGORITHMS TO IDENTIFY SUSTAINABLE NEIGHBOURHOODS WITH AN EXAMPLE OF SZEGED (HUNGARY)

Tom KAUKO

Department of Geography, Faculty of Social Science and Technology Management, Norwegian University of Science and Technology - NTNU, Universitetssenteret Dragvoll, NO-7091, Trondheim, Norway

Tom.kauko@svt.ntnu.no

Abstract

The housing market involves substantial diversity over space and product categories as it develops across localized city units. Especially amidst city transformations taking place in the Central and Eastern European (CEE) context such processes are dynamic. The paper analyses such localized housing market outcomes and sustainable development in Szeged, Hungary, during the period 2000-2009. The aim is to see to what extent these two patterns overlap. Housing market outcomes comprise data on averaged house prices and sales volumes. Sustainable development is defined in terms of environmental, social, cultural and economic features of the built environment.

The analysis is mainly descriptive and based on interview, field inspection and house price modelling using the self-organizing map (SOM) and the method of fixed time-windows. The analysis replicates prior analysis of Budapest. The results point to a rather heterogeneous urban housing context, which much involves green and mixed settings, often with accessible public transportation. These findings are by definition elements of urban sustainability. The conclusions furthermore suggest that a relatively well-developing regional city such as Szeged might be more successful in generating sustainable urban housing market locations and micro-environment than a capital city such as Budapest.

Keywords: housing market, sustainable development, built environment, the self-organizing map (SOM)

INTRODUCTION

When investigating the changes taking place in urban environments, housing market analysis at the local level is for long time already seen as a relevant research topic. In urban housing economic theory a longstanding issue of interest concerns the various determinants of relative house prices within a city. In the seminal works carried out already in the 1970s large variations in significant factors were found across urban areas: in some cases accessibility to city centre was the key whereas in others the environmental and social factors had a stronger influence on price levels (Ball, 1973; Ball & Kirwan, 1977). Since the 1990s the research gained more momentum due to contributions that accepted a more dynamic and complex mechanism of house price development than the hedonic approach[1] in its simplest form (see Maclennan & Tu, 1996; Adair et al., 1996). Localized price developments and market activities provide signals of how a city is diversifying in a more or less organic sense as some areas are showing growing and others declining tendencies in their economic fortunes when patterns of housing market change is observed. On top of the price development target urban sustainability is more recently being incorporated as a more qualitative target of evaluation (e.g. Bramley et al., 2009).

This study follows such research traditions as the aim is to compare the price development with various sustainability aspects across a variety of locations and typical market segments. This demonstration uses house price and turnover data on Szeged and follows the research design of a prior study on Budapest (the capital of Hungary). The data sets for the years 2000-2009 are run with a quasi-dynamic modelling approach

[1] Hedonic pricing is an increasingly popular method aimed at establishing an equation between the price and the characteristics (descriptors) of a consumer good – in this case the home. Using multiple regression analysis techniques, actual market transaction records and large datasets of the descriptor values, the model can be ‘parametricized’, and, partial prices for each descriptor isolated. The method is well established since the 1970s.

based on the self-organizing map (SOM) and fixed time-windows. The interviewed experts represent real estate business, local government and NGO sectors. Finally, field inspection is carried out. This paper documents some preliminary results of work in progress.

URBAN HOUSING MARKET ANALYSIS

Intra-urban spatial house price differentials and their development in time are interlinked with various locational influences. Among such factors, the physical environment is always important, because it changes slowly. Furthermore, institutions and socio-demographics are growing in importance, when looking at the generic categories of spatial housing market features. In this literature, typically, American, British and Australian cities are covered – relatively little is conducted from mainland Europe, and even less so from the Central and Eastern European (CEE) circumstances. This research is based on traditions of urban economics (pricing, supply, demand) or urban geography (mobility) in their standard forms, that is to say, assuming some form of equilibrium in outcomes. (Kauko, 2006)

Since the 1990s the emphasis on urban housing market analyse has been diverted on behavioural factors, feedback mechanisms and complexity. In this framework an certain area may experience an upward or downward development in value, depending on the time of development and its current image. The investment will either enhance the potential of that location, thereby attracting further investment, which increases the value further, or lead to dilapidation, a loss in potential, absence of investment and a further decrease in price-level. The trend may however be reversed in the area in question. Such processes are temporally and spatially specific, and lead to a mosaic of various price-segments with different degrees of substitutability among them. The questions to answer are how physical, socio-demographic, financial and administrative factors shape the housing choices of individual households and this way the urban form. Again, the CEE context here offers a special interest element due to the quick pace, comprehensive nature and large spatial scale of such changes in the urban mosaic. (Kauko, 2006)

SUSTAINABILITY ASPECTS RELEVANT FOR URBAN DEVELOPMENT

Within the widespread sustainability discourse buildings and land use issues occupy a core position from c. year 2000 onwards. In the physical sense land and buildings are increasingly subject to monetary sustainability evaluations along environmental, social, cultural and economic dimensions. This is evident in the discourse surrounding certification of relevant sustainability features or the lack thereof (see Jackson, 2009; Addae-Dapaah et al, 2009; Runde & Thoyre, 2010; Galuppo & Tu, 2010; Warren-Myers & Reed, 2010). For operational purposes related to such evaluation, I have picked twelve broad issues surrounding the urban land use and real estate sustainability agenda. These are listed as follows, starting from the most localised to the widest scale, based on literature:

1. Energy efficiency in buildings (during their life cycles)
2. Use of renewable energy in buildings (during their life cycles)
3. Pollution control in building (during their life cycles)
4. Real estate quality
5. Real estate affordability
6. Real estate diversity
7. Optimal density for a block/neighbourhood
8. Public transportation availability (functional issue)
9. Traffic pollution (ecologic issue)
10. Social cohesion in the neighbourhood/city/region (including favouring local products and labour)
11. Communicativeness in local/regional planning (governance transparency)

12. Innovativeness of the region (economic sustainability, including financial transparency of corporations).

In the present study the emphasis is on the factors 4, 6, 7, 8 and 9, due to their relatively easy identification during field inspections.

METHODOLOGY

While the sufficiently accurate linear hedonic regression modelling of the housing market has proved successful, alternative approaches allow researchers to capture the complex nature of the housing market relationships. The self-organizing map (SOM, Kohonen 1982) is a type of unsupervised neural network technique with a competitive network architecture.[2] The SOM is best defined as a mapping from a high-dimensional data space onto, a (usually) two-dimensional lattice of points. This way disordered information is profiled into visual patterns, forming a landscape of the phenomenon described by the data set. The SOM produces a feature map of nodes, each of which represents a characteristic combination of attribute levels. In the training procedure of the algorithm the matching is usually determined by the smallest Euclidean distance between observation and response. The results are strongly dependent on the data – all necessary guidance to the analyses is obtained from the sample we feed the network and from the compulsory network parameters. A label is assigned for recognising, for instance a symbol for a particular area, where the particular combination of characteristics is typical (street name in this case).

Thus the output nodes receive ‘hits’ by one or more observations with strong resemblance in terms of the input variables. This technique works in three steps. First, to predefine the surface in terms of the number of potential clusters (nodes) and the parameters for adjustment of this map-like surface (feature map). Second, to train the map using a dataset of m observations (cases) measured as n variables (map layers). Third, to examine the resulting feature map in terms of similarities between nodes and intensities of any nodes with respect to a given map layer. The similarity and intensity of any nodes can also be identified across all map layers when the location of a given node is fixed across these layers by definition. The figures 1 and 3 to 12 illustrate the outcome of this projection, clustering and discrimination process. Here each node (i.e. circle with shading) represents a certain type of observation, its shade denotes the intensity of a given indicator (e.g. price) and its the position notes similarity and difference to other types of observations. AS for the reliability of these clusterings, it depends on the number of ‘hits’ per node, which in turn depends on the data quality and quantity.

The main principle of functioning in the SOM could be described as ‘the winner takes all’. The winner is the node with shortest distance to the observation vector, and its weights are adapted towards the observation. This goes on until all observations are used for training – usually more than once. Neighbouring nodes on the map are being similarly adapted towards the observation, but the extent of this depends on the selected parameters. During the training procedure each node obtains a multidimensional numeric value where the dimensions correspond to input variables. The variation in numeric values across one input variable can be shown visually as a map-layer (see Fig. 1). Here are two important notions: *the similarity* between units within the structure, and *the typical properties* of a given unit with respect to the input dimensions.

[2] The neural network is such a sophisticated statistical method that captures nonlinear, but regular associations (i.e. patterns) within a data set without a pre-defined model.

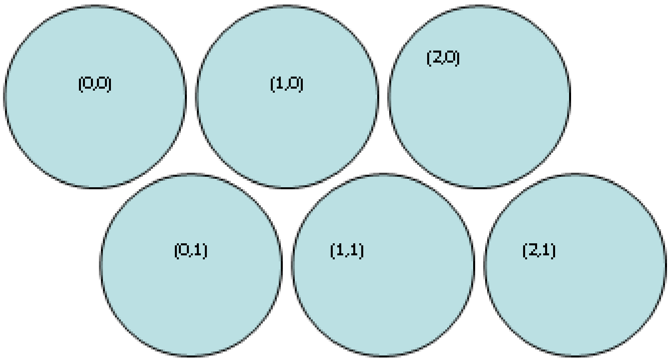

Fig. 1. The situation of the nodes in a three-by-two (3x2) map.

This technique has the following advantages:

- It comprises a holistic, pragmatic and fuzzy method.
- It enables incorporating a qualitative element beyond number crunching.
- It enables detection of combined effects of the input variables together with non-linearity and diversified effects

Unfortunately it also has some caveats that limit its use:

- It requires a *priori* selection of the number of nodes and other parameters.
- It is a 'black box'.
- Any changes that take place between two points in time (i.e. snapshot) are not captured as the time-windows sequence is merely 'quasi-dynamic (see below).

The real challenge for an analysis of price data based on the SOM is to incorporate the dynamic dimension. If time is one variable among others, then the rigid map cannot adapt new observations. However, a solution for this problem is to use fixed time-windows (Carlson, 1998). Using the fixed time-window method by, the procedure is to run cross-sections for each year and from the resulting feature maps identify changes in the trend of a given variable on the one hand, and in the fundamentals of this variable on the other.

To sum up the methodology, the SOM was used as a tool for reduction of dimensions of the input data, and subsequently for clustering of the observations based on these reduced dimensions. Here the time windows technique enabled incorporating temporal dynamics too (albeit in a quasi-dynamic manner). This modelling method was subsequently applied together with more qualitative research strategies (see earlier work on Budapest, Kauko, 2009b).

MORPHOLOGY OF SZEGED

Szeged is a regional city of c. 170,000 inhabitants (thus one tenth of the population of Budapest). While it has a critical mass and universities, it is not being part of a particularly well performing region. Szeged in itself is however a growing city much thanks to having managed to attract EU funding. This also can be seen in the physical structure which in recent years has improved significantly, in particularly in the inner city areas. On the other hand, we are talking about a historical city with vial position in the Austro-Hungarian monarchy. Compared to its size the urban structure is remarkably dense. The city structure of Szeged much is determined by the comprehensive planning and rebuilding that resulted after the great flood in 1879. As seen in Fig. 2, the ring structure is much reminiscent of that of Budapest (but without the elevation of Buda side).

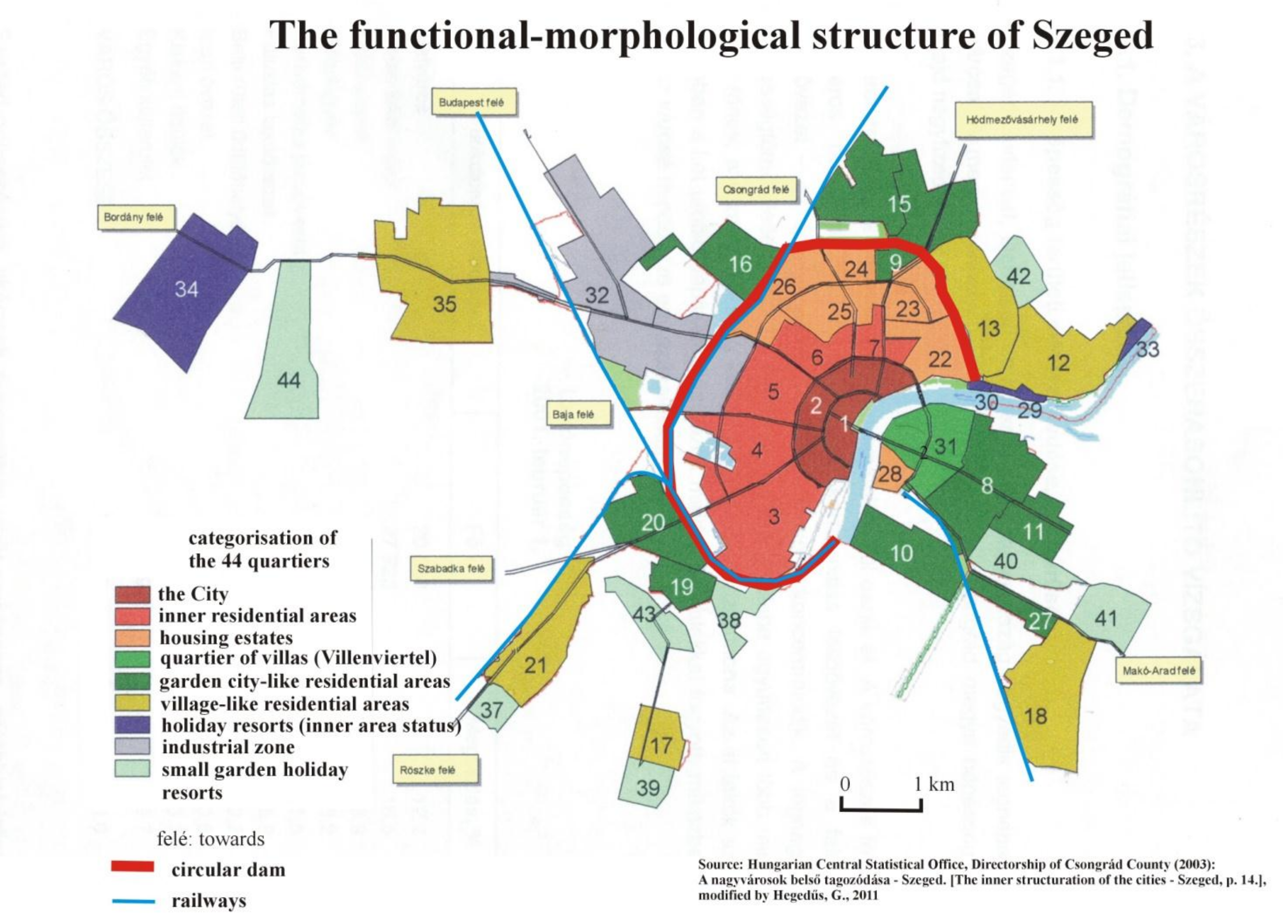

Fig. 2. The city structure of Szeged

In today's circumstances the city development proceeds as follows:

- The inner areas are being filled in rather than the outer expanded.
- Not much row houses since the 80s when it was a popular building type (applies for the whole country).
- The best area (in the sense of having the most expensive property prices) in Szeged:
  - The most expensive plot values: District 2, the Jewish quarters: Belső Városrész (inner parts of the city);
  - The most expensive building values: District 11, Marostő, single family area in Ujszeged.
- The worst area: Dorozsma (district 35): medieval infrastructure, dangerous, used to be a village but now very bad Roma ghetto; the worst in the inner city: Cserepesi sor – Roma neighbourhood with social housing (LKV) in District 4.
- Alsovaros (district 3) and Moravaros (district 4) have undergone gentrification – this has led to dominance of condos instead of peasant housing.
  - Refurbishing.
  - Demolishing and new built.
- Ujszeged (the part west of the river Tisza) is less densely built and thereby does not have so bad areas.
- In Szeged an active mayor Laszlo Botka (socialist party) with a strong lobby got large amount of EU funds. In fact, Szeged received most funds of all Hungarian cities. As a result Szeged is vibrant. (It is

speculated that the voting base of the socialists corresponds to the housing areas 26, 25, 24, 23, 6 and 7 – the 'Red Belt' in the northern parts of Szeged. (These are housing estates, except districts 6 and 7 which are inner city residential areas.)

- Only a few new projects have been launched in recent years:
    - *Arkad* shopping center at the edge of the Jewish Quarter in the inner city (district 2); it is already functional at the time of writing.
    - Due to the expensiveness of plots, residential developments take place outside the inner city: notably some gated communities in districts 22 (*Tiszapalotai*, *Tisza Palace*, a mixed new office and residential block close to and with view over the river Tisza); 25 (*The French Hill*, open condo) and 4 (*Vadaspark*, Zoo); obs: the last of these is amazingly very close to Cserepesi sór.

DATA

Static and dynamic analysis uses data on sales prices and volumes for a ten year period 2000-2009. The data is obtained from the Ingatlanadattar CD-rom, compiled by KSH (Hungarian Statistical Office). Three qualifications are to be observed when evaluating the data:

a) Street names refer to the municipality of Szeged. Some of the outskirts of Szeged belong to autonomous municipalities (such as Domaszek, Deszk and Sándorfalva), but are not included in this analysis.

b) Three standard building types are used: single-family, condominium and panel, as well as a total figure comprising a weighted mean of the price and the sum of transactions. (OBS: gated communities can be both single-family and condo types).

c) The figures comprise the mean price per sqm and volume of sales from the stamp duty calculations. While unreliable as absolute values, these indicators are reliable to compare in relative terms (across space and time).

ANALYSIS WITH THE SOM

The variables comprise a set of indicators which was readily available from KSH (in their annual CD-Rom of house price data). Data is recorded on mean sales prices per sqm and sales volumes aggregated on street and district levels. Both indicators are split onto four variables: single-family, condos, panel, and total figures. This way eight input variables are generated for the analysis with each variable enabling a market related interpretation. As explained above, the aim is to look at particular areas that form clusters on the SOM surface (feature map). In this way a quasi-dynamic approach is afforded through the time windows method.

The feature maps were generated using the following parameters: Software: *SOMPAK*, 12x8 map dimensions, bubble neighbourhood type, hexagonal topology, running length in basic run and fine-tuning 5,000 and 50,000 respectively, alpha (sensitivity parameter) 0.03 and 0.01, radius 10 and 3, and calibration based on street name. Some of the map layers are shown in the figures 3 to 12 (over). These display the variation in total price levels, and are selected intuitively for pedagogic purposes[3]. To interpret the position of the nodes and the grey-shade variations the key is that light colour indicates high price or turnover for a particular group of relatively similar observations, when the similarity between nodes means closeness within the map surface. Here it is to note that the position of a given node remains fixed across all map layers in a one year surface.

[3] All map layers are obtainable from the author upon request. (The total number of comparable map layers is 80 = 10 years x 8 variables; on top of these graphs many others could be shown too, for example a distribution of the number of 'hits' per node for a whole map).

FINDINGS FROM THE SOM ANALYSIS

The interpretation of the map layers is based on intensities of the shading, distances between the nodes and the labels (although special letters such as é and á are replaced by other symbols). Basically two kinds of cases are interesting for this study: (1) assumed niche markets where price is high but turnover low in relative terms; and (2) upper or upper-average markets where both price and turnover is high in relative terms. The locations with exceptionally high price but relatively low turnover are not particularly many for each cross-section. However, they are found everywhere except the northern and south-western neighbourhoods (and obviously the most western part, Doroszma). The more average priced locations with high turnovers in turn are even more spread over the city, which is logical due to their higher number. Thus in Szeged all areas contain cases that are relatively high priced and cases with high turnover, even the northern and south-western parts (and one is even found in Doroszma, the worst neighbourhood). To compare, in Budapest both types of locations were found in very specific places. This is a big difference between the two cities based on these analyses.

It is likely that all the cases that show up with light colour (high prices or turnovers) in the map layers were built at a time when no sustainability or green considerations existed in Hungary. Nevertheless, from each cross-sectional analysis above certain street-addresses were picked for further investigation (in October 2012). In order to follow consistently on the SOM analysis above the further aim was to focus on the two segments namely (1) highest price where turnover is low; and (2) average or above average price where turn-over is high. For reason of convenience, only the most accessible locations were subject to field inspection. The remaining locations were evaluated based on interviews and maps. The next aim was to see why the areas have a higher than average price level, and if there is any kind of logic related to sustainability aspects.

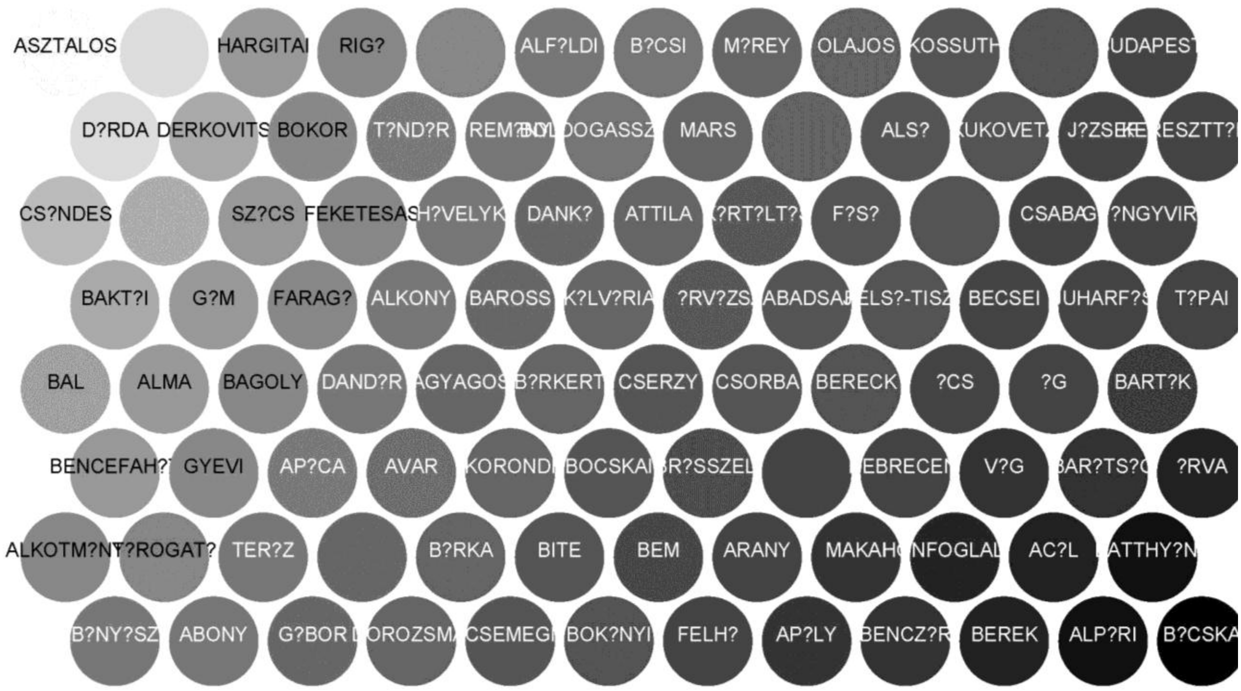

Fig. 3. Map layer: year 2000, total price per sqm.

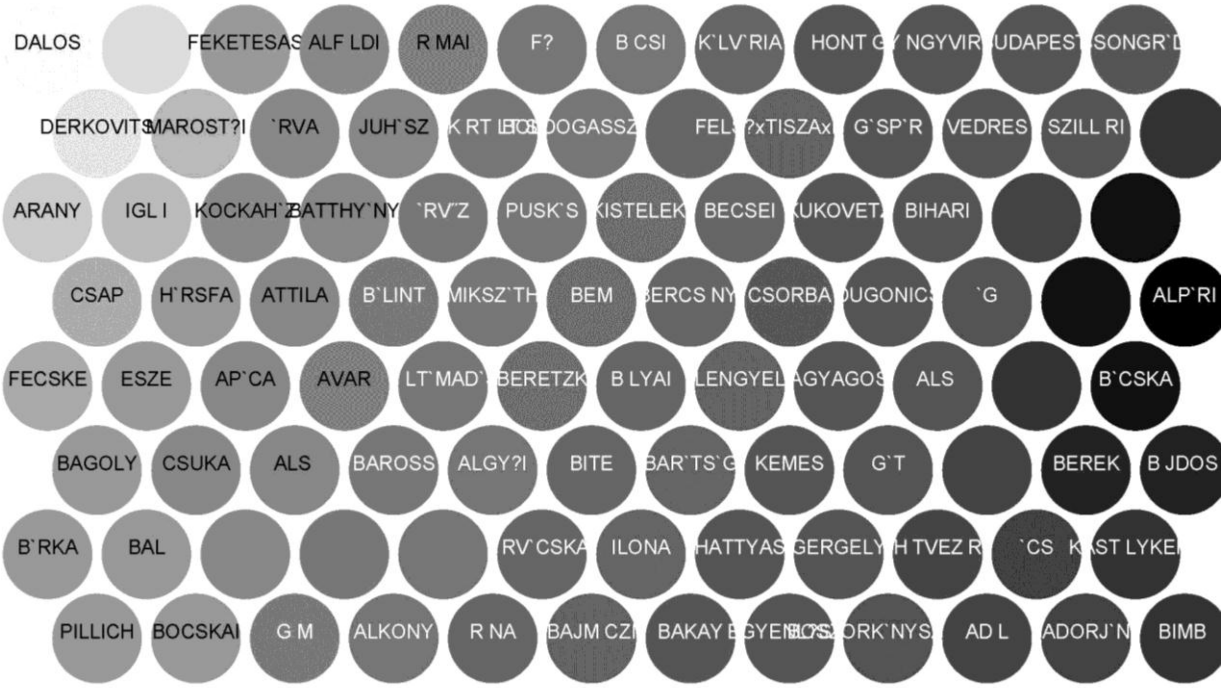

Fig. 4. Map layer: year 2001, total price per sqm.

Fig. 5. Map layer: year 2002, total price per sqm.

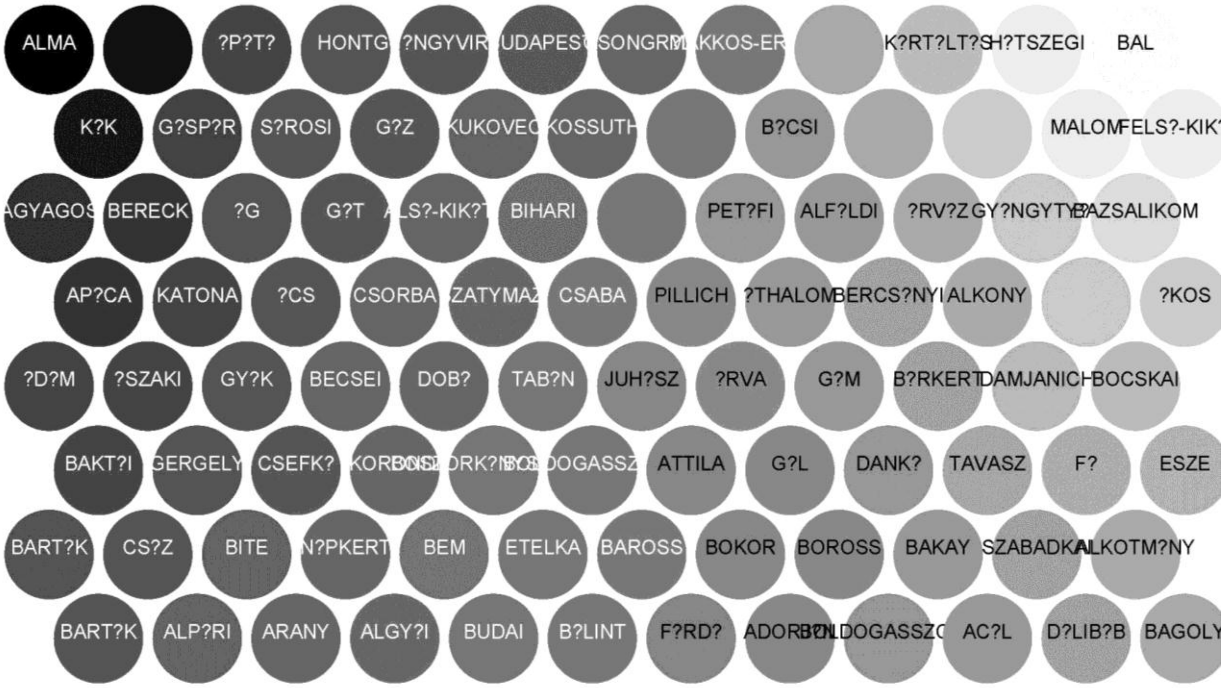

Fig. 6. Map layer: year 2003, total price per sqm.

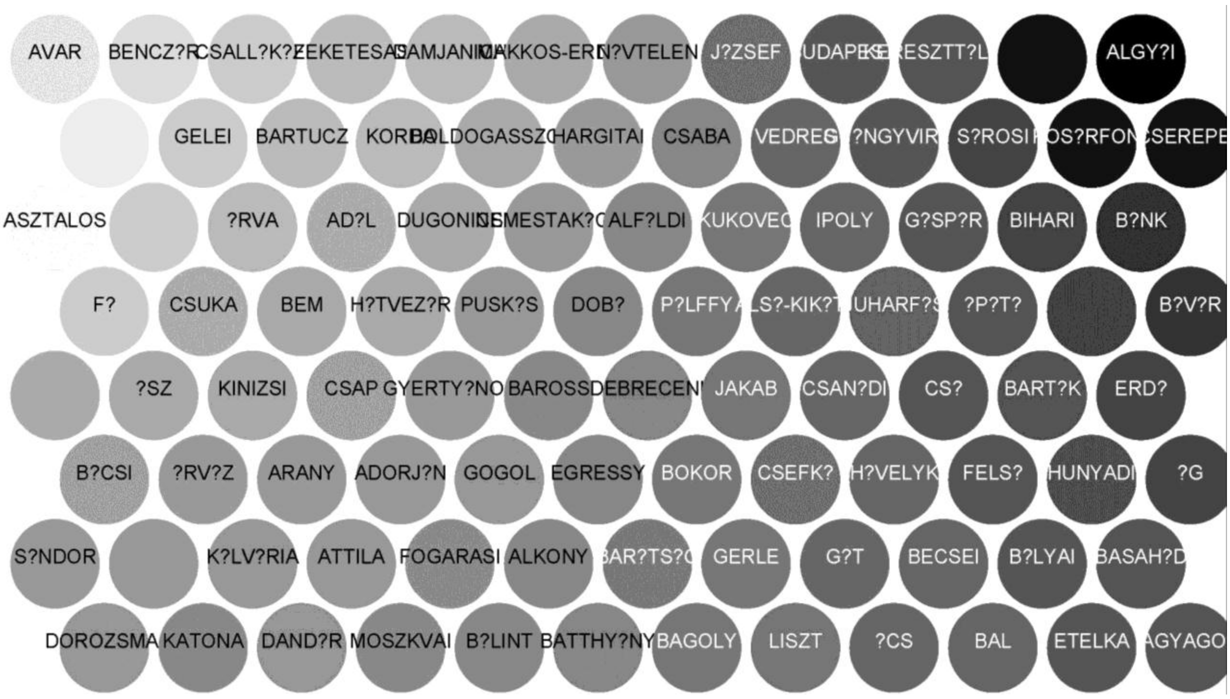

Fig. 7. Map layer: year 2004, total price per sqm.

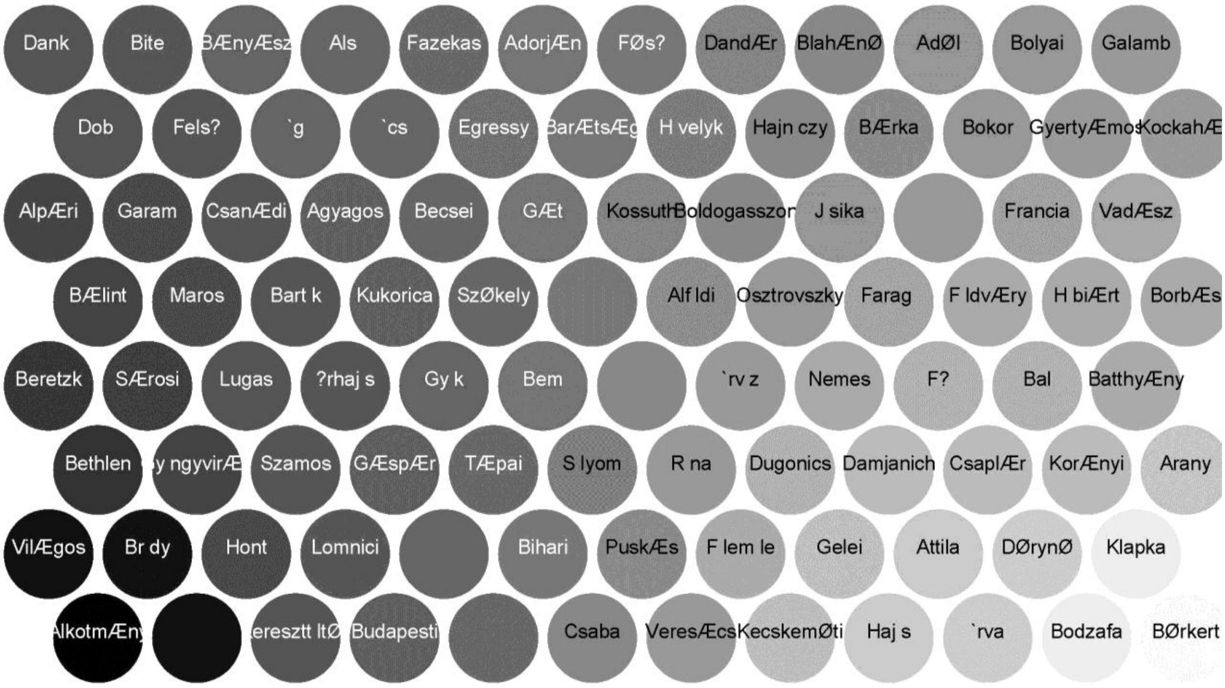

Fig. 8. Map layer: year 2005, total price per sqm.

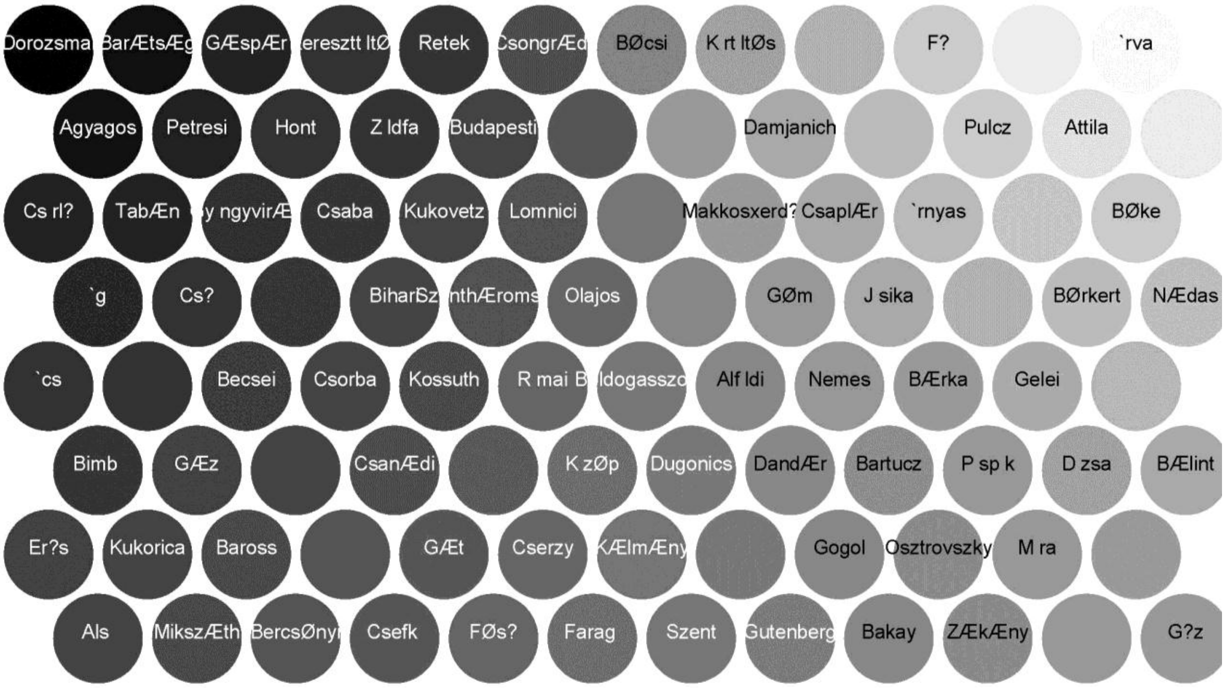

Fig. 9. Map layer: year 2006, total price per sqm.

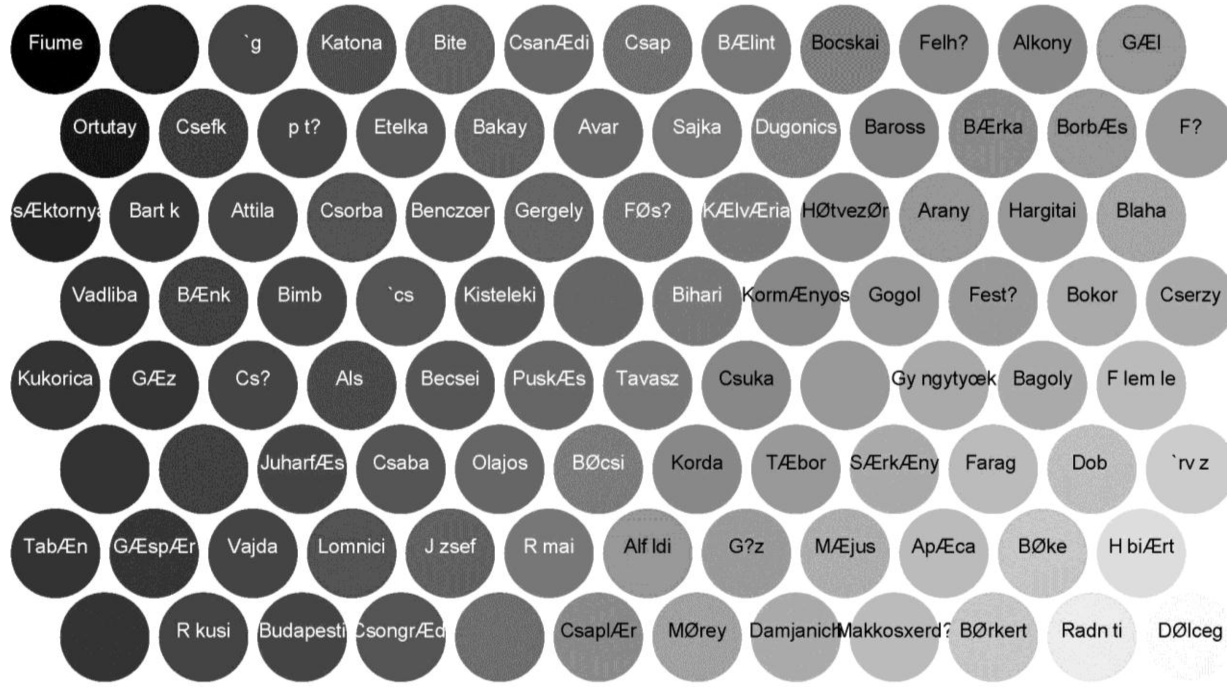

Fig. 10. Map layer: year 2007, total price per sqm.

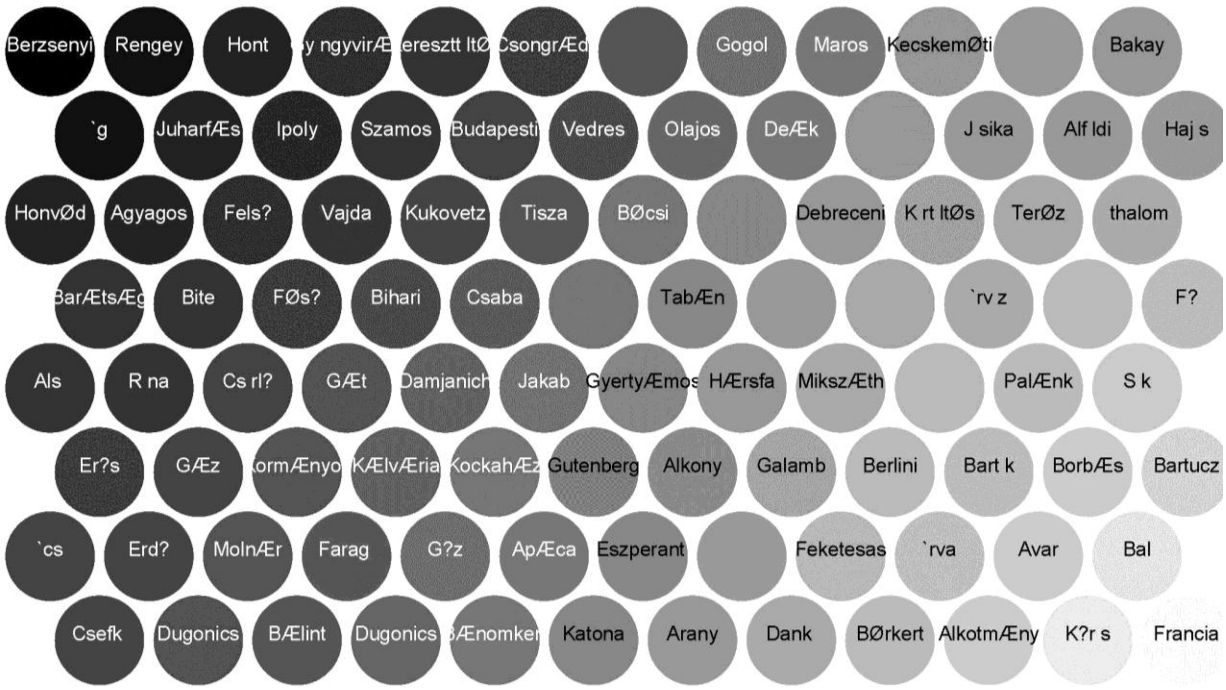

Fig. 11. Map layer: year 2008, total price per sqm.

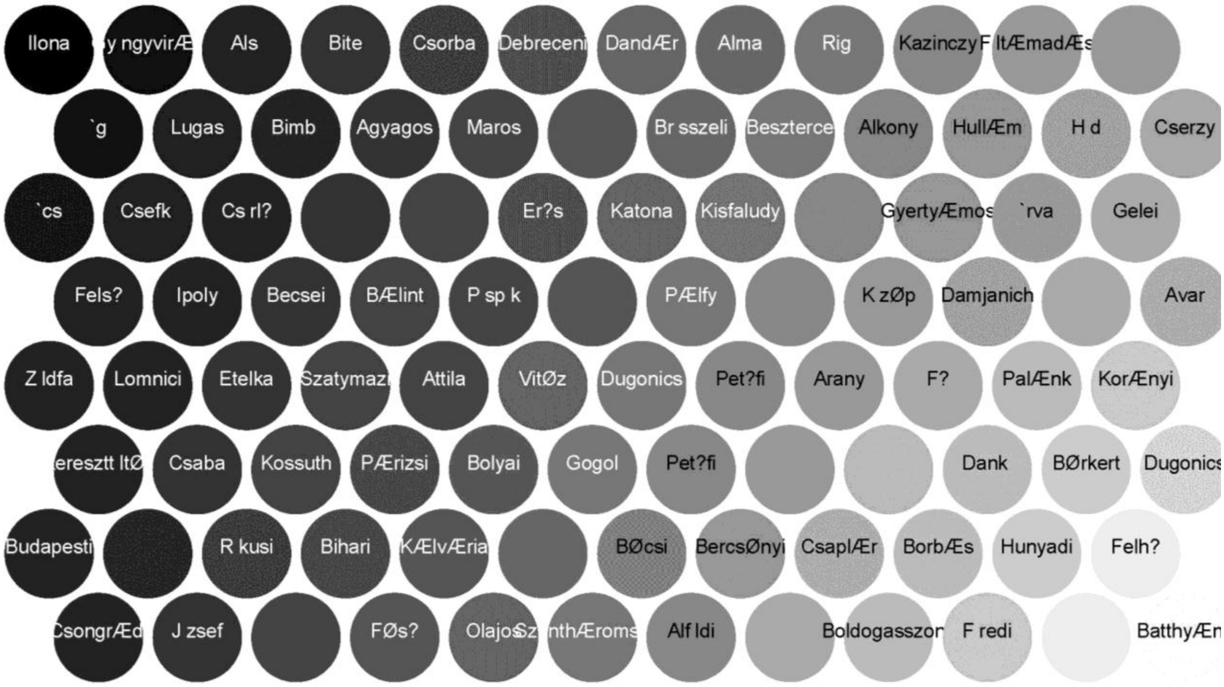

Fig. 12. Map layer: year 2009, total price per sqm.

FINDINGS FROM THE FIELD INSPECTION

Field inspections were carried out for 97 of the total 130 (street and property type specific) cases that showed up in the analyses (75%) with the SOM.

When compared to the corresponding study on Budapest, a number of relationships can be picked up:

1. Mixed areas (1960s housing estate, modest ‘residential parks’ (i.e. gated communities), cultural heritage buildings, converted farmhouses, various types of urban infill, empty plots and so forth) can be found in close proximity, often within one and the same segment of a given street.
2. Residential areas are green, pollution-free and well-maintained.
3. Residential areas have good accessibility and due to the compact city structure the distances are short everywhere. The exception is the area on the eastern side of the river Tisza known as Ujbuda, it is substantially less accessible with public transportation than the rest of the city.
4. Obviously, closer to the centre the structure is more dense (i.e. higher plot efficiency) and the facades better maintained than in the outskirts. The boundaries between the areas are not however as sharp as in Budapest.
5. Only in the ‘red belt’ of the housing estates in the northern segment of the city the areas are internally rather homogeneous and separated from the rest of the urban structure.

The field inspection confirmed furthermore that the following elements command a price premium:

- river proximity (e.g. Tiszapalotai),
- cultural heritage or neotraditional buildings, but in any case ’interesting design’, and
- more middle class neighbourhoods (relative terms).

This both corresponds with findings from Budaepst and text book urban economics theory. Lastly, from a sustainability perspective some positive developments could be verified:

- plenty of row houses (especially in Ujszeged), rather than high rise (which is socially unsustainable) or single-family/villa areas (environmentally unsustainable)

- bicycle friendly size and design of the city (at least its inner parts).

CONCLUDING DISCUSSION

Depending on the selection of data/variables it is possible to illustrate differences in urban structure using the SOM. However, while it is relatively easy to illustrate price premiums it is far more difficult to relate them to any sustainability factors such as cultural heritage, public transport accessibility, walkability, green areas, mixed stock, energy savings or maintenance of communal areas. This paper has documented a SOM-based study on Szeged, a middle-sized city. The field inspection confirms that there is premium for river proximity, heritage buildings and more middle class neighbourhoods (relative terms). This is however still within standard NC urban economics theory rather than any explicit urban sustainability conceptualizations.

From a sustainability perspective some positive developments include plenty of row houses and bicycle friendly city size and design. In a prior study on Budapest, using the same method and data, the analysis suggests that only the cultural dimension is strongly present in the Budapest context of sustainable housing development and housing market (Kauko, 2012). In that respect, the contrast to the analysis of this smaller city is amazing. Namely, that most of the city comprises residential areas that are green, clean and of mixed character, often also with good public transport. These features add up to a picture that is sustainable by definition. The broader significance of this findings might be that in a CEE context big/capital cities have fewer sustainable residential areas than middle-sized/regional cities, using the particular definitions of this study. In fact, the argument of places with fewer inhabitants being more walkable – and thereby more sustainable with respect to this criterion – than larger places was recently backed by Rauterkus and colleagues (2010). In the current paper a caveat is in order however: the analysis has not picked other than upper and upper average housing market areas and segments; furthermore, not even all of these are followed up with field inspection as the study is yet work in progress.

REFERENCES

Adair A.S., Berry J.N. and McGreal W.S. (1996) Hedonic modelling, housing submarkets and residential valuation. Journal of Property Research, 13(2), 67-83.

Addae-Dapaah, K., Hiang, L.K. and Shi, N. Y. (2009) Sustainability of Sustainable Real Property Development. . Journal of Sustainable Real Estate, 1(1), pp. 203-225.

Ball M.J. (1973) Recent Empirical Work on the Determinants of Relative House Prices. Urban Studies, 10, 213-233.

Ball M.J. and Kirwan R.M. (1977) Accessibility and Supply Constraints in the Urban Housing Market. Urban Studies, 14, 11-32.

Bramley, G., Dempsey, N., Power, S., Brown, C. and Watkins, D. (2009) Social sustainability and urban form: evidence from five British cities. Environment and Planning A, 41, pp. 2125-2142.

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

Galuppo, L.A. and Tu, Ch. (2010) Capital Markets and Sustainable Real Estate: What Are the Perceived Risks and Barriers? Journal of Sustainable Real Estate, 2(1), pp. 143-159.

Hegedűs, G. (2009) The Spatial Planning in Relation to Gated Communities in the Central Hungary Region. Unpublished manuscript.

Jackson, J. (2009) How Risky Are Sustainable Real Estate Projects? An Evaluation of LEED and ENERGY STAT Development Options. Journal of Sustainable Real Estate, 1(1), pp. 91-106.

Kauko, T. (2006) Urban housing patterns in a tide of change: Spatial structure and residential property values in Budapest in a comparative perspective. DUP Science Publication, Delft, 2006.

Kauko, T. (2009a) Policy impact and house price development at the neighbourhood-level – a comparison of four urban regeneration areas using the concept of 'artificial' value creation. European Planning Studies, 17(1), pp. 85-107.

Kauko, T. (2009b) The housing market dynamics of two Budapest neighbourhoods, Housing Studies, vol. 24, No. 5, pp. 587-610.

Kauko, T. (2010) An evaluation of the sustainability of inner city residential projects, Housing, Theory and Society, 27:1, pp. 1-20.

Kauko, T. (2012) A demonstration of evolutionary and sustainability arguments using house price data. Unpublished manuscript. Available from the author at tom.kauko@svt.ntnu.no

Maclennan D. and Tu Y. (1996) Economic Perspectives on the Structure of Local Housing Systems. Housing Studies, Vol 11, No. 3, 387-406.

Rauterkus, S.Y., Thrall, G.I. and Hangen, E. (2010) Location Efficiency and Mortgage Default, Journal of Sustainable Real Estate, 2(1), pp. 117-141.

Runde, T. and Thoyre, S. (2010) Intergrating Sustainability and Green Building into the Appraisal Process. Journal of Sustainable Real Estate, 2(1), pp. 221-248.

Warren-Myers, G. and Reed, R. (2010) The Challenges of Identifying and Examining Links between Sustainability and Value: Evidence from Australia and New Zealand. Journal of Sustainable Real Estate, 2(1), pp. 201-220.