Defining housing market areas using commuting and migration algorithms. Catalonia (Spain) as an applied case study

By Vicente Royuela[1], Miguel Vargas[2].

[1]Vicente Royuela
Grup d'Anàlisi Quantitativa Regional-AQR (Universitat de Barcelona)
Facultad CC. Económicas, Av. Diagonal, 690, 08034 Barcelona (Spain)
Tel : +34 934021412, Fax number: +34 934021821.Email: vroyuela@ub.edu

[2]Miguel Vargas
Centre for Spatial and Real Estate Economics (CSPREE)
The University of Reading Business SchoolandFaculty of Economics and Business, Diego Portales University, Chile. Manuel Rodríguez Sur 253, Santiago ChileTel +56(2)6762252Email: m.a.vargas@reading.ac.uk

Abstract
In the literature on housing market areas, different approaches can be found to defining them, for example, using travel-to-work areas and, more recently, making use of migration data. Here we propose a simple exercise to shed light on which approach performs better. Using regional data from Catalonia, Spain, we have computed housing market areas with both commuting data and migration data. In order to decide which procedure shows superior performance, we have looked at uniformity of prices within areas. The main finding is that commuting algorithms present more homogeneous areas in terms of housing prices.

Note: Vicente Royuela acknowledges the support of CICYT SEJ2006-07665 project and the Mobility grant of the Education and Science Ministry Ref PR2005-0253.

1. Introduction

A common formulation of a housing market is the travel-to-work area (TTWA). This concept comes from classical urban economics, in which residential location depends on the distance that the individual faces in his daily commute to work. TTWA is usually measured using commuting data and the parameters of self-containment and minimum size (see Coombes, 1997).

More recently, Jones (2002) has showed that the idea of spatial arbitrage can be used to define housing market areas (HMAs). Thus, migration data can provide an alternative to commuting data for building HMAs when there is, for instance, no commuting data available.

Here we compute HMA using both of these approaches in order to investigate which one performs better. The comparison criterion used, based upon Cournot's definition of a market area, is the extent of price uniformity within an area. Using data from Catalonia, Spain, we calculate hedonic housing prices for 946 municipalities, our basic unit of analysis, in order to combine them into HMAs that exhibit the maximum amount of heterogeneity across them.

Subsequent sections are focused on a literature review of the issue, looking at how housing markets and submarkets are related (section 2), and how a housing market is defined (section 3), as well as considering the two alternatives, TTWA (section 4) and self-contained migration areas (section 5). Further, section 6 describes the database, section 7 sets out the empirical results and section 8 presents comments and conclusions.

2. Housing Submarkets and Housing Markets

In a seminal work, Maclennan (1977) addressed several aspects open to improvement in house price studies. The first of them was to show the difficulties "in the definition of a housing market, the possibilities of market segmentation and the assumption of static equilibrium". These difficulties are still alive. A clear example is that when working with disaggregating the national framework of a housing study, the usual way of looking at data is top-down: since the majority of national studies are defined from a time-series perspective, the sub-national (regional) perspective is usually addressed by working with administrative regions. We assume that "these regions will not correspond with any recognisable housing market area" (Meen, 2001, p. 21), usually because of lack of data availability. This also assumes that labour is relatively immobile between regions, and that arbitrage will eliminate long-term relative changes within metropolitan areas (DiPasquale and Wheaton, 1996).

Of course, the diversity of houses and their multidimensionality, their spatial dispersion and their characteristic of being durable assets lead to spatially and sectorally complex markets with probable short-run disequilibrium. A clear formulation of the complexity of the topic is that, in contrast to the definition of HMAs, a huge amount of literature has been developed to define housing submarkets areas. Most of the studies examine whether a given housing market should be considered as a set of submarkets. Using more or less sophisticated techniques, the answer is usually affirmative. In these studies, an interesting question arises: how was the housing market area defined? The usual way of defining it is simply to assume the administrative boundaries or the greatest unit of physical area in the database used. Several recent examples are Bourassa and Hoesli (2003) for Auckland, Goodman and Thibodeau (2003) for Dallas, Watkins (2001) for Glasgow, and Kauko (2004) for Amsterdam. These and many other studies have emphasised the importance of considering spatial submarkets (see Kauko, 2004, for a recent review), which can arise because the spatial arbitrage process within an HMA is constrained by several problems: transaction costs, search costs, imperfect information and inelastic supply. In any case, houses in different submarkets are still substitutes, although imperfect, within the same broader market, which shows greater stability and, also, a long-term spatial arbitrage pattern.

3. Defining Housing Markets

Classical urban economics sets out clear directions for the physical size or extent of a city. In the literature (Wingo, 1961, Alonso, 1964, Muth, 1968 and Evans, 1973), travel to city centre is the determinant of residential location. Thus, people decide their final location (migration), taking into consideration the time they face every day to go to work (commuting), how much rent they have to pay for their accommodation, and also the fact that they are inside the city, that is, they can have access to several jobs or services that are not available in rural areas (space segmentation). Consequently, we can say that a city or housing market area ends in space when there is no additional household considering their current home as an alternative location for migrating and when no more people are commuting everyday to the city from outside the city. Thus, from our perspective, we assume that a city is the collection of alternative locations that are considered by households as location substitutes, that is, a housing market. According to Jones (2002), the classical Alonso model emphasises travel to the city centre and, therefore, the trade-off between journey-to-work and housing expenditures: "the implicit logic of these models is that the housing market is defined by the TTWA" (Jones, 2002 p. 550).

Maclennan put forward how conventional microeconomic analysis was not fully helpful in defining housing markets: "In Walrasian analysis the market exists at a point in logical space and time where a *uniform* commodity is exchanged. In Marshallian analysis the market exists at a point where a *simple* commodity is exchanged, in conditions where buyers and sellers have had past and frequent market experiences" (Maclennan, 1977, p. 63).

An additional approach is the one that deals with the concepts of spatial arbitrage. Jones (2002) recalls the definition of a market area made by Cournot: *"A market for a good is the area within which the price of a good tends to uniformity, allowance being made for transportation costs"*. Thus, "(…) buyers can do and do consider transactions at any point within the area to be an appropriate substitute and therefore that spatial arbitrage occurs" (Jones, 2002 p. 552). For the majority of goods we buy, it is the commodity that subsumes the transportation costs. But, in housing markets, it is the consumer who moves, not the product. The clearest expression of a housing buyer who moves is migration. Thus, migrants can be considered as households for which spatial arbitrage plays an important role.

Other ways of formulating HMAs have been developed by Forster *et al*. (1995), who suggested a number of alternative approaches to reaching a solution, including the vendor's point of view, search patterns, TTWAs and commuting costs, environmental quality, and a GIS multifactorial model.

These different points of view address exactly the same question, as viewed from alternative perspectives. For instance, Jones (2002) argues that HMAs defined by migration should be nested within TTWAs. Here we directly compare the two alternatives to see if either of them performs better than the other.

The main advantage of the classical urban model is that these perspectives can be clearly explained and related according to the model. Of course, we assume that the main disadvantage is that the real world is infinitely more complex, and that urban models can convey only a small part of that complexity. Partly for that reason, approximations made to determine the boundaries of housing market areas have been always partial. Here such partiality persists, but we try to compare two of the potential approximations.

In order to make the comparison, we consider the total amount of inequality in housing prices that exists between HMAs. If spatial arbitrage is playing the right role, then we should expect price to be more uniform within an HMA. Consequently, the maximum level of price heterogeneity *between* HMAs will be parallel to the maximum level of homogeneity *within* each HMA. As homogeneity is synonymous with equality, computing the amount of price heterogeneity between HMAs requires an inequality index. We have chosen the Theil index, since it has a property of mathematical fractals: it can be decomposed additively between groups, with the Theil index between all municipalities being equal to sum of the Theil index between HMAs and the weighted average of the Theil indices within each HMA. This property greatly simplifies the calculations needed to choose both the best set of parameters and the best approach. Appendix 1 displays the details of the Theil index.

4. Housing Markets as Travel–to-Work Areas (TTWA)

As mentioned before, a market area is an area where the price of goods tends towards uniformity, because spatial arbitrage occurs. As houses are immobile, it is the consumer who moves, not the product. Thus, the concept of TTWA based on commuting patterns has been developed to define a local labour market area. Here the requirement for spatial arbitrage exists, although as Jones (2002) highlights, this principle does not necessarily ensure the internal coherence that Goodman (1970) argued was an essential feature of a local labour market area.

The most frequently used method to create small areas has been the one related to local labour markets, initially proposed by Fox and Kumar (1965). It is based on commuting data. Small areas are grouped hierarchically with bigger areas to which workers travel every day. Coombes et al (1986), Smart (1974) and Ward (see Pacinelli, 1998) have systematised the procedure through the development of algorithms, which have been widely used in many countries and regions, such as in the UK (travel-to-work areas), the Netherlands (daily urban systems, van der Laan, 1998), Italy (*sistemi locali di lavoro*, Pacinelli, 1998), and also in the region of our case study, Catalonia (*àrees funcionals*, Palacio, 1998 and *sistemes urbans*, Royuela and Romaní, 2004).

The algorithm works provided that the areas satisfy several minimum thresholds. In the studies undertaken by Coombes for the UK, the minimum size of a TTWA is a resident workforce of 3,500, and a minimum of 75 percent of journey-to-work trips to or from any TTWA must both start and end in the area in order to qualify. For areas with labour forces of 20,000 or more, the self-containment criterion falls to 70 percent (see Coombes *et al.*, 1985 and 1988). This means that two criteria are at work: minimum size and self-containment.

5. Housing Markets as Self-Contained Migration Areas (SCMAs)

A market search is the first step that households face when they want to migrate. In the end, migration may fail to occur because of financial constraints, for instance, or simply because, after a quick look, a household realises that it is much better off at its current home than anywhere else. As a result, migration is effective or revealed demand. Jones

(2002) lists two main problems with the search approach: problems of practicability, and problems of isolating the search as the only important step of a more general process. Consequently, he proposed using migration patterns—revealed demand—as the key variable for defining an HMA, based on the principle of spatial arbitrage. What is needed is a self-containment level that ensures spatial arbitrage, although “the precise level of self-containment which ensures spatial arbitrage remains an unanswered question (…) that can be only answered empirically” (Jones, 2002, p. 555).

The empirical analysis is very similar to the one employed in defining TTWAs. Jones uses migration data over a 10-year period. He argues that, since migration is not happening every day, the self-containment level should be below 70%. As a consequence, he finally adopts a level equal to 50% of total moves and uses an algorithm similar to the TTWA one, and a threshold of 5% of local sales originating from another settlement. Here the two criteria are the self-containment level and the minimum threshold level.

6. Description of the Database

Our study is undertaken for the region of Catalonia (NUTS II in the European administrative classification), which is one of Spain's most developed regions. It is located in the northeast of the country and borders France. Catalonia had a population of 6,343,110 inhabitants in our base year of 2001 and is, together with Madrid, Spain's most populated and urbanised region. It has 946 municipalities, which are organised in 41 administrative groups, named *comarques,* and four administrative provinces, namely, Barcelona (east: 1,503,884 inhabitants in the city and 4,805,927 inhabitants in the province in 2001, making up 76% of the region), Girona (north: 74,879 and 565,304; 9% of the region), Lleida (west: 112,199 and 362,206; 6% of the region) and Tarragona (south: 113,129 and 609,673; 10% of the region). The 946 municipalities are the basic unit of our study, and we want to combine them into a smaller number of HMAs.

We used commuting data between municipalities corresponding to the population census of 2001. We also used migration data for the 1991-2003 period, from

Variaciones Residenciales published by the Spanish National Statistical Institute (*INE*), that is, for 13 years, the maximum period we had available—as did Jones (2002)—in order to avoid short-run distortions. It should be noted that data was not available for moves within every municipality.

Housing price data of municipalities come from the Spanish Ministry for Housing and always refer to 2001. The database consists of 107,808 dwellings, and takes into account the municipality where the dwelling is located, along with a small number of structural characteristics: age and size.[1] In order to avoid housing prices being influenced by the various characteristics of the dwellings, we computed an overall regression of the database against the squared structural variable. The hedonic prices coming from this regression are the ones we employed in the subsequent sections. We assume that different prices relative to dwelling age and size can happen between markets and submarkets, but this is what we expect to find in the following sections.

7. Empirical Results

The results obtained for each pair of parameters in the two procedures give rise to a series of maps of the various ways of grouping municipalities. Although we reckon the fact that the literature (Coombes et al, 1985) recommends different parameters for urban and rural areas, we did not make this distinction, as we believe that the most important point here is finding the best procedure. Implicitly we are assuming that this differentiation should apply equally to both methodologies. Jones (2002) reports a brief global sensitivity analysis, which clarifies the difficulty of finding the "best" value and does not give any result using a mixed algorithm. In our work we follow the same approach, and for comparison purposes, we will consider both alternatives.[2]

[1] This database draws on appraisals made by officially accredited appraisal firms, because the *real* price of housing transactions is not published anywhere in Spain. As expected, there is a huge diversity in the amount of dwellings per municipality, from 16,511 in the city of Barcelona to other municipalities with just 8 appraisals; the median for the 946 municipalities is 16 appraisals.

[2] In other words, since the purpose of our work is developing a comparison between the two techniques, we are assuming that any differentiation needed between rural and urban municipalities would equally affect both commuting and migration algorithms.

Another difficulty is finding the optimum number of HMAs in a region. Any number might potentially be preferred. Consequently we had to choose a common-sense number of areas. To do so, we briefly return to the number of administrative units within the region: there are 4 provinces and 946 municipalities. Reasonably, we are looking for an intermediate figure. In the 1930s, intermediate units named *comarques* (41 in total) were defined, organised on an agricultural basis. Thus, all the *comarques* should have a main city that could play a market town role for trading farm goods with the rest of the municipalities, so that every farmer could travel to the capital of his/her *comarca* in a single day. Of course, since then transportation costs have experienced a dramatic change, and highways and faster trains allow for daily commuting from more than 100 km (precisely the approximate distance between Barcelona and Tarragona or Girona, two of the other main cities in the region). As a result, we expect an approximate number of HMAs between 4, the number of provinces, and 41, the number of *comarques*.

In order to compare the inequality between HMAs, we have to assume that the measurement of inequality (heterogeneity) between HMAs will be greater in the solutions with a larger number of areas. Hence, we will look at the results carefully and not discard any solution *a priori*.

The structure of the algorithms used for commuting and migration data is almost identical. They are displayed in Appendix 2. In order to be consistent in both procedures, we will use the same criteria: minimum size, self-containment level, and minimum threshold level.

In order to compare the two techniques, both algorithms define the minimum size in terms of population, not workforce (as in the commuting algorithms). In the end, we used a value of 20,000 inhabitants. In addition, for the threshold level used in Jones (2002), we are using migrations instead of local sales originating from another settlement.

7.1. Commuting Algorithms

Firstly, we look at the commuting algorithms. The international literature has shown how the parameters differ depending on the case study, and this also happens in our case. We developed computations with several parameters in order to see the data's sensitivity at different threshold values: self-containment levels between 60% and 80%; a minimum amount of commuting between the small areas in order to create a bigger area from 0% to 20%.

Table 1 displays the results for a set of various parameter values. As can be seen, the highest Theil index of price inequality between HMAs is found at a self-containment level of 65%, using a minimum threshold of 15% of movements between units. Consequently, the highest homogeneity in prices within HMAs is found at these values of self-containment and minimum threshold. Notably, these parameters are slightly different than previous results in Catalonia (75% in Royuela and Romaní, 2004). In addition, we see that our results suggest a number of areas—20—higher than the areas resulting from a higher level of self-containment—16 areas for 75%. Picture 1 displays the 20 different areas.

[TABLE 1]

[PICTURE 1]

7.2. Migration Algorithms

Next, we show the results for the migration algorithms. We started with the self-containment levels used in the literature—Jones uses 50%—and we also explore additional levels of self-containment up to 90%. As we have no data on moves within every municipality, the self-containment level was computed as the proportion of migrations over total population. This procedure will result in significantly higher levels than the ones used by Jones (2002). We also use threshold levels for the minimum amount of migrations from one HMA to another that vary between 0% and 10%. Table 2 displays the results for the grid of parameters.

[TABLE 2]

[PICTURE 2]

As can be seen, the more areas that are found, the higher the Theil index is, as expected. We find the index maximum at 46 HMA in a self-containment level of migration equal to 70%. Nevertheless, if we want to compare both techniques, we have to look at a similar number of areas. Consequently we have to look at self-containment levels between 86% and 88%. What we find are Theil index values much lower than what we found in the commuting algorithm. Picture 2 shows the 19 HMAs that arise from a self-containment level of 87% and a minimum threshold of 10% of migration between units.

What our results suggest is simply that the commuting algorithm leads to more homogeneous HMAs in terms of housing prices, compared to the migration algorithm. It means that the commuting algorithm is preferable when building HMAs.

8. Comments and Conclusions

In this paper we have built housing market areas using commuting and migration data. We have compared the results in terms of housing price homogeneity, and we have found how—with a comparable number of areas within the Catalonia region—commuting HMAs are more homogeneous in terms of prices than migration HMAs.

How can this be explained? Theory tells us that the two procedures should have given similar results, and clearly this is not what is going on after looking at the Theil indices and the maps. So, why is that happening? What is going on with migrations?

First of all we must bear in mind that different kinds of data are involved. As for commuting data, we are using data from a single point in time, namely 2001. By contrast, our migration data refer to a quite wide interval, from 1991 to 2003. That would mean that we could face a trade-off when using migration data: long intervals work to avoid distortions, while short intervals are instead representative of a single point in time. At this stage, one of the requisites of the HMA, relative to submarkets, should be kept in mind. The former needs higher stability and a long-term spatial arbitrage pattern. Consequently, we assume that the comparison being developing here may be an approximation to the ideal one, comparing migration data over a long period of time with yearly commuting data, which is simply not available.

Secondly, we address a major point related to a more complex aspect of migration data. The theoretical urban models show us that if static equilibrium is assumed, then there would be no reason for any migration or it would be negligible. By contrast, assuming permanent disequilibrium would imply that the rationality of individuals does not take spatial adjustment into consideration, and migrations, therefore, could even be random. Evans (1990) puts forward three ideas that could help to reconcile continuing net migration with continuing equilibrium: families migrate according to consistent patterns over their life cycle and differences in growth rates of territories. Further, rising incomes will lead to an increasing demand for an average or superior bundle of amenities. As Evans explains, these points play a major role in an intra-urban model, rather than in an inter-regional one. Consequently, in our framework, it is straightforward to assume that, although some persistent differences in standards of living exist in the territory, they are dynamically corrected through the migration mechanism, with a shorter or longer lag depending on the case as well as on the geographical scope of the analysis. Of course, when a situation of disequilibrium exists—i.e., when a new highway is built—migration will happen until a new static (or long-run) equilibrium is achieved.

This argument clearly expounds one of the major problems with using migration data. We do not know what these flows are saying to us. Are they an expression of disequilibrium? Are they revealed demand playing the role of spatial arbitrage? Probably both explanations are right. In that case, if the former applied, we would expect many migrations not to show spatial arbitrage within an HMA, but between HMAs. Consequently, the algorithm procedure used here would lead to worse results than what could be achieved with the commuting algorithms.

Of course, we do not claim that it must always be so. Looking at one particular case—the region of Catalonia in the nineties—we have found much better results with commuting data than with migration data, and we can say that in that region over that period of time, a migration process to reach equilibrium may have existed, and that would make commuting algorithms preferable. We understand that if that reflects our case, then any other situation making use of migration data should clearly discard the possibility of disequilibrium in order to make the final results acceptable.

9. References

ALONSO, W. (1964) *Location and land use: towards a general theory of land rent,* Harvard University Press, Cambridge, Mass.

BOURASSA, S. C., HOESLI, M. (2003) Do housing submarkets really matter? *Journal of Housing Economics,* 12, pp.12-28.

COOMBES, M. G. (1997) Travel to work areas: frequently asked questions in *Travel-to-Work Areas and the Measurement of Unemployment: Conference Proceedings* in I. Turok (Ed.), Centre for Housing Research and Urban Studies, University of Glasgow, pp. 13–32.

COOMBES, M. G., GREEN, A. E., OPENSHAW, S. (1985) New areas for old: a comparison of the 1978 and 1984 travel-to-work-areas *Area* ,17, pp. 213–219.

COOMBES, M. G., GREEN, A. E., OPENSHAW, S. (1986) An efficient algorythm to generate official statistical reporting areas: the case of the 1984 Travel-to-Work Areas revision in Britain, *Journal of the Operational Research Society,* 37, pp. 943-953.

COOMBES, M. G., GREEN, A. E., OPENSHAW, S. OWEN, D. W. (1988) Substantive issues in the definition of ‘localities’: evidence from sub-group local labour market areas in the West Midlands, *Regional Studies,* 22-4, pp. 303–318.

DIPASQUALE. D., WHEATON, W. (1996) *Urban Economics and Real Estate Markets,* Prentice Hall: Englewood Cliffs, NJ.

Evans, A. (1973) *The economics of residential location*, Macmillan, London.

Evans, A. (1990) The assumption of equilibrium in the analysis of migration and interregional differences: a review of some recent research, *Journal of Regional Science,* 30(4), pp. 515-531.

FORSTER, E., ELLIS, S. P., MCLEERY, A. (1995) Difficulties in defining housing market areas Paper presented to the conference Social Change in the City and the Countryside, Heriot-Watt University.

FOX, K. A., KUMAR, T. K., 1965, The functional economic area: delineation and implications for economic analysis and policy, *Papers and proceedings of the Regional Science Association,* 15, pp. 57-85.

GOODMAN, J. F. B. (1970) The definition and analysis of local labour markets: some empirical problems, *British Journal of Industrial Relations,* 8, pp. 179–196.

GOODMAN, A. C., THIBODEAU, T. T. (2003) Housing market segmentation and hedonic prediction accuracy *Journal of Housing Economics,* 12, pp. 181-201.

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

KAUKO, T. (2004) A Comparative Prespective on Urban Spatial Housing Market Structure: Some More Evidence of Local Sub-markets Based on a Neural Network Classification of Amsterdam, *Urban Studies,* 41(13), pp. 2555-2579.

MACLENNAN, D. (1977) Some Thoughts on the Nature and Purpose of House Price Studies, *Urban Studies,* 14, pp. 59-71.

MEEN, G. (2001) *Modelling Spatial Hosuing Markets. Theory, Analysis and Policy,* Kluwer Academic Publishers, Dordrecht.

MUTH, R. F. (1969) *Cities and housing*, University of Chicago Press, Chicago.

PACINELLI, A. (1998) Sull'autocontenimento del lavoro nei sistemi locali del lavoro dell'Italia peninsulare. Un'analisi esplorativa dei SLL dell'Italia peninsulare, *Economia e Lavoro*, 2.

PALACIO, G. (dir) (1998) *Cohesió, centralitat i cohesió iterativa a Catalunya: 1981-1986-1991-avanç 1996,* Departament d'Obres Públiques, Generalitat de Catalunya, Barcelona.

ROYUELA, V. and ROMANÍ, J. (2004) Defining multicriteria urban areas. An application to Catalonia, Working paper presented in the XXX Reunión de Estudios Regionales, Barcelona.

SMART, M.W. (1974) Labour market areas: uses and definition *Progress in Planning,* 2(4), pp. 238-253.

VAN DER LAAN, L. (1998) Changing urban systems: an empirical analysis at two spatial levels, *Regional Studies,* 32 (5), pp. 235-247.

WATKINS, C. (2001) The definition and identification of housing submarkets, *Environment and Planning A* ,33, pp. 2235-2253.

WINGO, L. (1961) *Transportation and urban land,* Resources for the Future, Washington DC.

APPENDIX 1

The Theil index is an inequality measure based on the idea of entropy

$$Theil\ \ Index = \sum_{i=1}^{n} \frac{y_i}{Y} \log\left(\frac{y_i}{Y} \bigg/ \frac{n_i}{N} \right)$$

where: y_i = Total amount of the variable that belongs to individual *i*.

$Y = \Sigma\, y_i$ = Sum of the whole amount of the variable for all individuals.

n_i = size of individual *i*

N = total amount of individuals

When there is total equality, every individual has the same amount of the variable. Consequently, $\log\left({}^{y_i}\!/_{Y} \big/ {}^{1}\!/_{N} \right)$ would be equal to zero, and the total sum would be equal to zero as well. As the inequality rises, the index gets higher and higher, reaching its maximum value at $-\log(n_i/N)$.

We have computed the index using the computed hedonic housing prices as the key variable (Y), and the individuals are the HMAs. For each HMA, we have used the average price of all appraisals in that HMA. The number of appraisals in each HMA has been used as the size of each individual (N). Consequently, the total number of individuals has been equal to the total amount of appraisals. Finally, we note that the Theil index can be decomposed in several hierarchical levels because of its interpretation as a mathematical fractal.

APPENDIX 2

Algorithms for Constructing Housing Market Areas

Step 1. Identify the *adjacent* settlements that account for at least the given threshold percentage of migration/commuters in the total population of an area.

Step 2. If there is only one greater than the given percentage, then pair the former area with the latter. If there is more than one area over the given percentage, then sum the migration/commuting flows in and out of the area from/to each of the adjacent areas. The settlement is to be paired with the one which gives the highest total migration/commuting flows between areas.

Step 3. If one of the areas is already an HMA, then the other area is incorporated into that HMA. If an area is paired with another area, then they are summed up to create a new area, which is to be included as a whole in future computations until a new HMA is created by adding new municipalities, or until they join an existing HMA.

Step 4. An area can be called an HMA if it has a minimum level of self-containment and a minimum population of 20,000 inhabitants. If an area does not fulfil these criteria, it has to be included in another HMA or put with new municipalities to be tested again.

Step 5. Only after 25 iterations do we relax the requirement of having the given percentage of migration/commuters in the total population of the area, and then the above process (steps one and two) is undertaken for all areas.

Table 1. Commuting. Number of HMAs and the Theil Index of Inequality in Prices between Areas

		Self containment level				
		60%	65%	70%	75%	80%
Threshold level	0%	17 - 3.43%	17 - 3.43%	14 - 3.33%	10 - 3.03%	8 - 3.02%
	5%	20 - 3.33%	18 - 3.12%	16 - 2.86%	14 - 2.82%	10 - 2.51%
	10%	21 - 3.32%	21 - 3.21%	17 - 3.12%	14 - 2.78%	13 - 2.61%
	15%	20 - 3.43%	**20 - 3.44%**	18 - 3.31%	16 - 2.98%	15 - 2.81%
	20%	20 - 3.43%	20 - 3.41%	18 - 3.34%	15 - 2.94%	14 - 2.95%

Picture 1. TTWA-Commuting Data (Self-Containment 65% and 15%). 20 Areas.

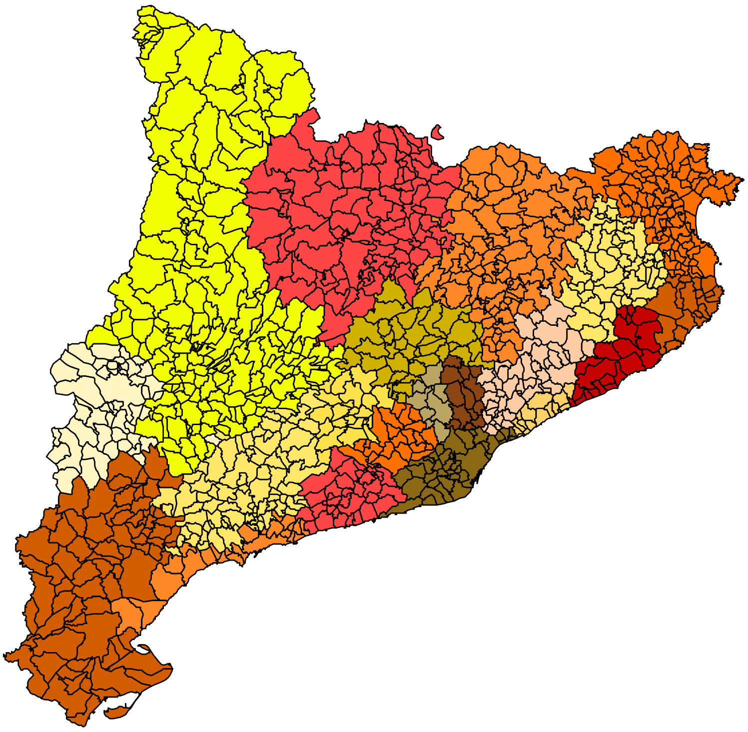

Table 2. Migration. Number of HMAs and the Theil Index of Inequality in Prices between Areas

Threshold level	Self containment level: 50%	55%	60%	65%	70%	75%	80%
0.0%	42 - 3.63%	42 - 3.63%	42 - 3.63%	42 - 3.63%	41 - 3.63%	35 - 3.53%	27 - 3.41%
2.5%	42 - 3.64%	42 - 3.62%	44 - 3.65%	46 - 3.66%	47 - 3.69%	44 - 3.49%	35 - 3.25%
5.0%	41 - 3.64%	41 - 3.63%	45 - 3.65%	47 - 3.67%	43 - 3.67%	44 - 3.48%	37 - 3.39%
7.5%	41 - 3.64%	41 - 3.63%	45 - 3.66%	47 - 3.67%	46 - 3.72%	42 - 3.43%	35 - 3.31%
10.0%	41 - 3.64%	41 - 3.63%	45 - 3.66%	48 - 3.7%	46 - 3.72%	41 - 3.49%	31 - 3.34%
	85 %	86%	87%	88%	89%	90%	
0.0%	14 - 1.65%	10 - 2.4%	7 - 0.63%	4 - 0.19%	3 - 0.15%	3 - 0.15%	
2.5%	27 - 3.28%	21 - 2.92%	17 - 2.75%	10 - 2.13%	10 - 2.27%	8 - 2.21%	
5.0%	29 - 3.04%	24 - 2.75%	21 - 2.57%	15 - 2.39%	14 - 2.39%	9 - 2.38%	
7.5%	25 - 3.11%	23 - 3.09%	19 - 2.67%	17 - 2.69%	11 - 2.32%	9 - 2.39%	
10.0%	24 - 2.91%	20 - 2.9%	**19 - 2.88%**	16 - 2.44%	12 - 2.52%	10 - 2.74%	

Picture 2. Migration HMA (Self-Containment 87% and 10%). 19 Areas.