Universidade de Aveiro Departamento de Ciências Sociais, Políticas e do Território
Ano 2012

JOÃO JOSÉ LOURENÇO MARQUES

A NOÇÃO DE ESPAÇO NOS MERCADOS HABITACIONAIS URBANOS

THE NOTION OF SPACE IN URBAN HOUSING MARKETS

Universidade de Aveiro
Ano 2012

Departamento de Ciências Sociais,
Políticas e do Território

JOÃO JOSÉ LOURENÇO MARQUES

A NOÇÃO DE ESPAÇO NOS MERCADOS HABITACIONAIS URBANOS

THE NOTION OF SPACE IN URBAN HOUSING MARKETS

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Doutor em Ciências Sociais, realizada sob a orientação científica do Doutor Eduardo Anselmo Moreira Fernandes de Castro, (Professor Associado) do Departamento de Ciências Sociais, Políticas e do Território da Universidade de Aveiro Doutor Arnab Bhattacharjee, Professor (Reader) do Economic Studies Department da University of Dundee.

o júri

presidente

Prof. Doutor Manuel João Senos Matias
Professor Catedrático do Departamento de Geociências da Universidade de Aveiro

Prof. Doutor Eduardo Anselmo Moreira Fernandes de Castro
Professor Associado do Departamento de Ciências Sociais, Políticas e do Território da Universidade de Aveiro

Prof. Doutor Arnab Bhattacharjee
Reader no Department of Economic Studies da Universidade de Dundee (Reino Unido)

Prof. Doutor Duncan Maclennan
Professor na School of Geography & Geosciences da Universidade de St Andrews (Reino Unido)

Prof. Doutor José da Silva Costa
Professor Catedrático da Faculdade de Economia da Universidade do Porto

Prof. Doutor Monojit Chatterji
Professor no Department of Economic Studies da Universidade de Dundee (Reino Unido)

Prof. Doutor Joaquim Carlos da Costa Pinho
Professor Auxiliar do Departamento de Economia, Gestão e Engenharia Industrial da Universidade de Aveiro

Prof. Doutor Carlos José de Oliveira e Silva Rodrigues
Professor Auxiliar do Departamento de Ciências Sociais, Políticas e do Território da Universidade de Aveiro

Acknowledgments

In the following lines I should like to mention how grateful I am to those who directly or indirectly have contributed to the completion of this stage of my academic career. Apart from being considered a formal procedure, it is only fair to recognize the importance of some people without whom this work would not have been possible. It is a great pleasure to thank all of those whom made this thesis achievable.

First and foremost I would like to thank my supervisors, Professors Eduardo Anselmo Castro and Arnab Bhattacharjee for their sapience, unwavering support and enthusiasm for the research process. They were responsible for putting me in this “boat” and encouraging the pursuit of increasingly higher levels of demand. Their support was beyond the strict relationship of academic cooperation.

My sincere gratitude also extends to Nandini Sen, who since the beginning received me as a member of her family.

I am grateful to the School of Economics & Finance of the University of the St Andrews and the Economic Studies Department of the University of Dundee for providing an exceptional reception during my visits. I had the pleasure to meet some fine people which excel at combining work duties with one of the beautiful pleasures of life, that of friendship. My stay in these universities would truly be less remarkable without the presence of: Chaiyasith Boonyanate, Chang Yee Kwan, Dilshad Jahan, Homagni Choudhury, Long Wang, Rosen Chowdhury, Sayantan Dastidar, Shaolin Wang and Xuan Wang. Thanks also go to the staff of both Universities, namely, Professors Catia Montagna, Hassan Molana, Monojit Chatterji and Rod McCrorie.

Special thanks also go to Professors George Galster and Geoffrey Hewings who gave me extremely valuable insights and incentives into the progress of the work.

To professor Christopher Jensen-Butler the acknowledgment of its crucial importance in this ongoing and enthusiastic research.

I also want to express my special thanks to my colleagues and friends of the University of Aveiro, particularly, the Department of Social, Political and Territorial Sciences for their interest, sharing of knowledge, comradeship and friendship.

I would also like to thank all the members of the Research Group in Planning and Innovation (GETIN) for the good environment and friendship. Among them, a special thanks to Professor José Manuel, Paulo Batista and Marta Marques for their unquestionable availability.

The author is grateful to the Portuguese Science and Technology Foundation (FCT) for the financial support of the Project DONUTS (Drivers of housing demand in Portuguese urban system - PTDC/AUR-URB/100592/2008) which allowed the work in progress to be presented at several national and international scientific conferences and support three training periods in Scotland.

I also wish to thank the two real estate agencies, ERA Aveiro and Janela Digital, for providing housing data, a critical part of the empirical research.

Finally, my greatest thanks go to my family: Joana, my sister and my parents; without their unlimited support and comprehension, this research would have been harder. Eva you appeared in the end of this journey, however, at a crucial time to put each thing in its right place... it’s just a PhD thesis.

Obrigado, não teria chegado aqui sem o apoio de todos vós.

palavras-chave

Estrutura Espacial, Mercado Habitacional, Dependência Espacial, Heterogeneidade Espacial, Escala Espacial.

resumo

O mercado imobiliário tem um papel importante nas economias modernas, tanto a nível macro como a nível micro. Ao nível macro, a construção de habitação representa um sector importante e influente na economia, com efeitos multiplicadores significativos sobre a produção e o emprego. Ao nível micro, uma residência representa o activo mais valioso da maioria dos indivíduos e uma parcela muito relevante da riqueza das famílias. Para estas, o custo e a qualidade das suas habitações influencia directa e indirectamente a sua qualidade de vida.
A habitação é por isso mesmo um tema, que avaliado nas suas múltiplas dimensões, se caracteriza por ser bastante complexo, mas também ao mesmo tempo desafiante. De modo a delimitar o objecto de análise do trabalho de investigação, esta tese realça os aspectos de localização e distribuição espacial das habitações urbanas. Será desenvolvido um quadro conceptual e respectiva metodologia para a compreender a estrutura espacial da habitação urbana realçando os três aspectos fundamentais da análise espacial: heterogenidade espacial, dependência espacial e escala espacial. A metodologia, aplicada à área urbana de Aveiro e Ílhavo é baseada numa análise hedónica factorial de preços e na noção não geométrica do espaço. Primeiro, é fixada uma escala territorial e são definidos submercados habitacional. Posteriormente, quer a heterogeneidade quer a dependência espaciais são estudados utilizando métodos econométricos, sem considerar qualquer padrão fixo e conhecido de interações espaciais. Em vez disso, são desenvolvidos novos métodos,tendo como base o modelo hedónico factorial, para inferir sobre os potenciais drivers de difusão espacial no valor de uma habitação. Este modelo, foi aplicado a duas diferentes escalas espaciais, para compreender as preferências dos indivíduos em Aveiro ao escolher os seus locais de residencia, e como estas afectam os preços da habitação. O trabalho empírico, utilizando duas bases de dados de habitação distintas, aplicadas ao mercado de habitação de Aveiro mostram: i) em linha com a literatura, a dificuldade de definir submercados e compreender as inter-relações entre esses mercados; ii) a utilidade de uma abordagem híbrida, combinando análise factorial com regressão; iii) a importância fundamental que o efeito escala espacial desempenha no estudo da heterogeneidade e dos spillovers e, finalmente, iv) uma metodologia inovadora para analisar spillovers sem assumir aprioristicamente uma estrutura espacial específica de difusão espacial. Esta metodologia considera a matriz de pesos espaciais (W) desconhecida e estimatima as interações espaciais dentro e entre submercados habitação.

keywords

Spatial structure, Housing Market, Spatial Spillovers, Spatial Heterogeneity, Spatial Scale.

abstract

The housing market plays an important role in modern economies, both at the macro and micro levels. At the macro level, housing construction represents an important and influential sector in the economy with large multiplier effects on production and employment. At the micro level, a residence represents the most valuable single asset owned by most individuals as well as a very important consumption good, and a very large share of household wealth. For households, the cost and quality of their houses greatly influence their quality of life.

Housing is a very complex issue, when analysed in its multiple dimensions, but at the same time it is also challenging. In order to delimit the analysis of this research, the thesis highlights the aspects of location and spatial distribution of urban housing.

A conceptual framework is developed along with the corresponding methodology for understanding urban spatial housing structure, emphasizing three relevant aspects of space: spatial heterogeneity, spatial spillovers and spatial scale. The methodology applied to the urban housing market of Aveiro and Ílhavo is based on a factor hedonic analysis and on the notion of multi-dimensional non-Euclidean space. First, spatial scale is fixed and housing submarkets are defined at this scale. Next, heterogeneity and spillovers are studied using spatial econometric methods. Importantly, no fixed and known pattern of spatial interactions is assumed. Rather, new methods are developed based on the factor model to help infer the potential drivers of spatial diffusion in residential value. This model is used, on two different spatial scales, to understand how households in Aveiro choose their residential locations, and how this in turn affects their house prices. The empirical work, using two distinct housing databases, applied to the urban housing market of Aveiro reveals: i) the difficulty of defining submarkets and the understanding the inter-relations between these markets in line with the literature; ii) the usefulness of employing a hybrid approach, combining factor analysis with regression; iii) the crucial rule that the spatial scale plays in the study of heterogeneity and spillovers; and finally iv) an innovative methodology to analyse spatial spillovers without assuming specific drivers of diffusion. This methodology considers the spatial weight matrix (W) as unknown and estimates the spatial interactions within and across housing submarkets.

TABLE OF CONTENTS

Index of Tables

Index of Figures

PART 1

Introduction

I.Introduction

Why is housing singled out for this work when it is just one of several goods in the economy? Many justifications can be given to answer to this question, but for this particular purpose one single reason deserves special attention: the straight relation between housing characteristics (and functions) and space. Since the housing market shapes space, and space defines how the housing markets operate, interesting conceptual and empirical modelling challenges present themselves.

This research focuses on examining the role of space in the housing markets. Related to this main goal, the thesis has three main contributions: i) it develops a new framework to analyse urban spatial structures considering three key aspects of space, spatial heterogeneity (structural differences between housing markets or housings), spatial dependence (spatial interactions across submarkets or housings) and spatial scale (the territorial level where these phenomena occur); ii) it develops a new methodology based on factor analysis and a multi-dimensional non-Euclidean notion of space to understand spatial hedonic models (this factor analysis approach is innovative to the determination of submarkets, and therefore, heterogeneity across submarkets, and unrestricted spillovers between submarkets); and, finally, the third contribution iii) the application of this framework and methodology to the urban housing market of Aveiro. For this purpose new databases for developing empirical analyses are used. Despite the well known lack of available housing information, for some levels of disaggregation the definition of the most appropriate scale to capture the relevant aspects necessary to analyse housing spatial structures in urban spaces, is a key issue.

I.1. The relevance of the subject

The housing markets and sectors linked to housing play important roles in regional and national economic growth. In modern economies; when major changes take place in these markets they often have impacts on other sectors of the economy and consequently on the social development of their populations.

The impacts of the housing market are visible both at the macro and micro levels in three distinct areas of the economy: i) in the production system, ii) in household expenditure, and iii) in urban and regional planning.

In the production system, the housing sector has significant multiplier effects on employment, output and investment, resulting from its relationship to a vast commodity chain. The housing market is considered a major element in the composition of national income in modern economies. For example, in the United States housing construction and related services represent about 14% of GDP (Bureau of Labour Statistics, 2001), in Portugal housing investment accounted for 18% of Gross Fixed Capital Formation, in 2003, while equivalent figures for Spain and the European Union were 29% and 25%, respectively (Eurostat - Statistical Office of the European Union). However, in recent years construction of new houses in Portugal has been subject to considerable decline, as reported by the Portuguese National Statistical Institute.

As a part of household expenditure, the residence represents the most valuable single asset owned by most individuals, and a very large proportion of household wealth. The share of income spent on housing represents a very large percentage of total expenditure and a permanent source of direct expenses (rent, interest rate and amortisation, repair and renovation, etc.) and indirect costs (energy, water, telecommunications, furniture and other domestic goods, etc.). For this reason, for households, the cost and quality of their houses greatly influence their quality of life. The Portuguese National Bank estimates that average housing expenditure in Portugal currently represents 30% to 40% percent of monthly income, while the Bureau of Labour Statistics (U.S.), in its questionnaire on the family budgeting, reports that the proportion of total expenditure allocated to housing was 34% in 2008. Beyond its impact at the household level, major and persistent expenditures on housing lead, at the macroeconomic level, to high levels of external indebtedness affecting countries as diverse as Portugal and the United States.

Finally, the housing market has a significant impact on the growth of cities, which in turn leads to changes in the concentration or sprawl patterns of urban populations. Consequently, this leads to pressures on the use and preservation of rural areas and natural heritage, but especially on the provision of adequate infrastructure and facilities for economic growth.

The rise in the importance of the housing sector is due to the abovementioned multiplier effect and corresponding policies essential for maintaining high levels of economic growth and employment supply. The flip side of the coin corresponds to the negative effects from oversupply of housing that is recorded for most developed countries (ESDP, 1999; ECB, 2002), and which in turn, is partially explained by the expansion of the number of second dwellings. In the case of Portugal, the number of houses increased from 2.6 million, in 1970, to more than 4.8 million in 2001, representing a growth rate significantly higher than the number of households. This also represents the second highest ratio of housing per household within the EU (1.38), exceeded only by the value of Spain (1.44) (Eurostat - Statistical Office of the European Union). Therefore, control of oversupply and resulting pressures on land use along with avoiding negative impacts from the decline of the housing sector, are fundamental tasks that must be supported by a thorough knowledge of the housing market. The ultimate objective of urban and regional planning is to make qualitative changes in the housing sector, involving investment in two dimensions: promotion of quality and comfort of accommodation and its associated energy efficiency and, the repair and restoration of buildings as an alternative to continued expansion in housing supply.

Despite the importance of housing markets, however, issues relating to the measurement of housing characteristics (both tangible and intangible) and lack of knowledge regarding the mechanisms of this market make the analysis of this theme a complex task. Evidence of this complexity can be seen in the persistent variability and uncertainty, spatial and temporal, which requires a profound and prospective analysis on the nature and specificities of these transformations.

At the temporal level, the successive changes in population structure contribute significantly to the variation in housing markets over time. Understanding the nature of such temporal variation is important when using deterministic and stochastic methods to forecast future trends, which in turn are extremely important for policy. At a theoretical level this issue is presented in the Section IV.2, where the relevance and limitations of methods used for analysing the temporal aspects of the housing market are highlighted. Despite the relevance of temporal perspectives in the analysis of housing markets, it is an aspect not developed in the empirical work; only a short analysis of the evolution of time coefficients is presented (the spatial perspective is the focus of this research).

At the spatial level, the configuration of urban settlements is sometimes ambiguous. At the urban scale, the prices of housing and location patterns vary significantly across different urban centres and within each part of the city, rendering research on spatial variation a complex task, while at the regional and national scale, urban concentration and dispersion of the urban population is a widely studied but controversial research topic (see CBOD, 2011; Castro, Marques *et al.*, Forthcoming-a; Castro, Marques *et al.*, Forthcoming-b). Spatial patterns in the housing market arise, according to the spatial econometric literature (Anselin, 1988; LeSage and Pace, 2009; Anselin, 2010), from a combination of two distinctive aspects: spatial heterogeneity and spatial dependence. The former is related to the characteristics of demand or supply that cause differentials across space (neighbourhood, or other spatial disaggregation) in the way that housing is valued. The challenge is to identify housing submarkets, to capture the spatial variability of housing value across space. On the other hand, the latter means that housing prices or other characteristics for a particular locality show a high degree of similarity when compared to more distant locations, resulting from some form of spillover. As mentioned by Le Sage and Pace (2009) the value of a house in a particular place may be based on recent housing price appreciation of the neighbourhood.

I.2. Housing market challenges

One of the essential conditions for efficient performance of any market is that all agents operating there possess full and correct information to support their decisions. The size and complexity of housing markets, the associated externalities, the susceptibility to speculative activities and the impact that the housing market has on macro-economic dynamics, reinforce the importance of providing quality information to support decision making.

In Portugal there is a considerable amount of information about the housing market, collected by the National Statistics Institute (INE) and by a variety of other institutions or firms, public and private, involved in real estate and housing related activities. However, this information is fragmented, poorly systematized and dispersed over a large number of agencies, who tend not to cooperate with each other. Due to the lack of quality information, and perhaps more importantly, inadequate tools to support decision making (both for real estate agents and public institutions), the housing market is not as transparent as desired. In particular, there is lack of clear understanding concerning the supply side of the market, the preferences of buyers and consequently the price formation mechanisms.

The problem starts with the price of land, which has been subject to intensive growth, an issue not easily explained by economic fundamentals. Price is affected by complex dynamics, driven by divergent interests of owners, builders, real estate agents, local and fiscal authorities etc., and involving externalities associated with both the patterns of centrality in permanent reconfiguration, and the processes of planning and urban management. Adding to the complexity of the land market, there is the heterogeneity inherent in housing markets, rendering systematic research of spatial-temporal variation of the housing market limited, standard and largely unsatisfactory. The challenges of identifying and analysing the structure of housing markets, its relationships to other areas of economic activity, and individual preferences for housing as a consumption and investment good, are truly complex and enormous.

Five selected topics have been chosen here for discussion, encompassing supply and demand sides as well as price formation and policy, that can be considered fundamental for understanding the housing market. See also Smith, Rosen *et al.* (1988), and McMaster and Watkins (1999) for a comprehensive, detailed, but somewhat dated, review.

i) Land use patterns and urban land development

On the supply side the analysis of land value, which depends crucially on fertility (Ricardo, 1821) and built heritage, has occupied an important place in the economic sciences. Unfortunately, this does not imply that the analysis of the land market has consolidated into a mature and uncontroversial area, since the accumulation of empirical knowledge and theory has been accompanied, and often overtaken, by empirical observations on the evolution of spatial and temporal patterns. The traditional dichotomy between city and countryside was accompanied by simple patterns of inter-urban hierarchies, based on centre-periphery transitions; see, for example, the Christäller (1933 [1966]) model emphasizing hierarchy of central places. This in turn gave rise to complex territorial configurations which have manifested themselves in concepts such as the diffused city, emergent city, metropolis and urbanized countryside; see Hall (1966) for pioneering research in the area and Lacour and Puissant (2007) for an overview of the current literature. Consequently, patterns of land value, depending on their location at macro level or on their local surroundings, significantly increased in spatial complexity and temporal variability. This variability, in conflict with the role of scarcity rent traditionally assigned to land, opens the way for speculative dynamics, and generates reduced transparency in land market behaviour. Such behaviour is increasingly dependent on divergent incentives of real estate agents and financial intermediaries, and the reaction of these agents to fiscal policies and planning instruments.

ii) Construction techniques and housing types

The evolution of construction techniques and building materials, the growing sensitivity to energy efficiency issues and the multiplicity of infrastructures, equipments and services associated with the electronic revolution, computing and telecommunications, have all had a profound impact on housing supply. Building techniques, quality standards and the level of urban facilities considered essential have also evolved rapidly, which in turn has affected the cost of construction and technical requirements imposed on various agents in the supply side of the market (Rosenthal, 1999; Malpezzi and Maclennan, 2001). The evolution of information and communication technology also had a decisive role in the emergence of new methods to advertise and sell real estate products, contributing, in a complex and sometimes counterintuitive way, to transparency in the real estate market. Zillow[1], Zoopla[2], Eppraisal[3] and Real Estate ABC[4] are some of the most current real estate portals in U.S. and U.K. In these portals, beyond the usual statistics of housing markets, it is possible to find useful tools for those who want to buy or sell a home (Zestimate, Walkability-Score, descriptive reports for a specific place based on predefined criteria, and the Zed-Index, for instance).

iii) Demand side dynamics

On the demand side, demographic dynamics have a crucial influence on the evolution of housing demand, both in terms of quantity of houses and the desirable features of such housing. The contrasts between areas that attract and expel population, and the increasing instability of the standards of attractiveness, together with investment value and durability of a dwelling, are important factors of mismatch between supply and demand. This is highlighted by patterns in socio-cultural changes which, in turn, induce qualitative changes in housing demand. The aging population, leading to growth in the number of elderly couples or isolated widows, together with a reduction in birth rates and an increasing number of divorces, has resulted in a dramatic decrease in household size, which underpins the increased number of houses and a decrease in their average size. The growing instability in the labour market, as well as the need to conciliate professional obligations with affordability, children's education and desirable neighbourhoods, also contributes to the increasing complexity of

[1] Zillow.com. url: http://www.zillow.com.
[2] Zoopla.co.uk. url: http://www.zoopla.co.uk.
[3] Eppraisal. url: http://www.eppraisal.com.
[4] Real Estate ABC. url: http://www.realestateabc.com.

housing demand; for further discussion, see among others, Mayo (1981) and Zabel (2004).

iv) Spatio-temporal dynamics and price formation

Substantial literature on the determination of house prices has accumulated over the past four decades. Research highlights mismatch between demand and supply at least in a localised context (in terms of region and type of housing, for example), a low and declining price-elasticity of supply, and a reduced response from demand to price signals when compared with changes in income. Significant and persistent spatial variation and dynamics in prices and volatility, as well as elasticities, have been attributed to differences in features of the local economies as well as to local supply constraints that limit the response of prices to changes in the economic environment (DiPasquale, 1999). The implications of differences in housing markets in terms of reduced mobility and a growing spatial inequality have also been discussed.

There are two distinct approaches to modelling housing markets in the economics literature. First, there is a well-developed empirical literature analysing supply, demand and prices across regions and over time based upon economic models (see Smith, Rosen *et al.*, 1988). The general approach in this literature is to construct three behavioural equations (for endogenous demand, supply and prices) which link exogenous independent variables to the property market. Unlike many other markets, the link between demand and supply in housing markets is not direct, and operates indirectly through vacancy rates. However, data on vacancy rates are not always readily available, which places constraints on the empirical implementation of such models.

Following Wheaton (1999), the second approach in the literature examines search and bargaining and its effect on price formation in local housing markets. These repeated searches, matching and bargaining models highlight the importance of time-on-the-market and degree of overpricing in the price-setting process. Importantly, this offers an alternative micro-founded approach where, in the absence of quality data on vacancy rates, time-on-the-market (and sometimes also overpricing) can be used to identify the wedge between demand and supply; for a recent application, see Bhattacharjee and Jensen-Butler (2011).

Further, hedonic and repeated sales models of regional prices reflect not only geographically varying price effects, but also substantial spatial dependence; see, for example, Rosenthal (1999) and Malpezzi (2003). Attempts have been made to explain spatial diffusion, particularly in terms of neighbourhood characteristics such as crime rates, schooling, transport infrastructure and quality of public services, and social interaction and segregation (Rothenberg, Galster *et al.*, 1991).

The above literature abounds in its implicit acknowledgement of the strong spatio-temporal dependence by features of regional or local housing markets. However, what is distinctly missing in the literature is adequate understanding of the reasons behind spatial or spatio-temporal interactions (Bhattacharjee and Holly, 2011). Whereas traditional empirical spatial models hold the nature and strength of spatial spillovers as given, the choice of an appropriate economic distance measure is by no means obvious: these may be based on one of geographic distances, transport costs, transport time, or socio-cultural interactions. These different drivers have widely varying implications for policies relating to neighbourhood improvement and revitalisation, quality of public services and employment opportunities.

v) Financial markets, public policy and decision-support information

Housing investment as a share of household expenditure, the traditional role of real estate investments as a destination for savings, the multiplicity of agents intervening on the supply side of the market, and the large time lags between real estate planning, construction, sale and financial transactions in housing all necessitate that the financial system plays a key role in the residential property market. Furthermore, it is a common observation that cycles in the real estate sector have a strong impact on financial markets and on the sustainability of national accounts, at the same time as fluctuations in interest rates and credit availability affect housing market investment (Wheaton, 1999; Mayer and Somerville, 2000). Simultaneously, fiscal policy, planning and urban management, and government intervention (in terms of supply of land to support urban renewal actions or development of social housing) are factors which determine the dynamics of the market and render its systematic and scientific study challenging.

The complexity of these dynamics, the scarcity of information, and the incomplete understanding of market structure and its response to exogenous impacts (such as planning and regulation) and endogenous shocks (for example, housing and business cycles) mitigate against the growing need to formulate appropriate models for decision support and the need for necessary information on housing markets. Without these models, the analysis of quality-adjusted housing supply and its relation to demand, as well as understanding the integration of housing markets into the urban structure, is rendered difficult, perhaps even impossible. It is therefore essential that policy-makers and planners, economists and managers, architects and geographers understand the structure and spatio-temporal dynamics of demand, supply and prices in the housing market.

The work developed in this research addresses, albeit with the main focus on spatial issues, some important insights into such challenges.

I.3. Research aims and questions

From the previous two sections emerged the idea: i) of the importance and the impact of housing, at the macro and micro level; ii) the large set of dimensions involved in the issue of housing, both from demand and supply side perspectives; and iii) its permanent variability, at a temporal and spatial level. Despite this complexity, the lack of mechanisms to quantitatively analyse the housing market, mainly at a more urban scale, requires additional efforts to better understand this phenomenon.

Hence, the main objective of this thesis is to improve the understanding of the housing market, emphasising the importance of spatial dimension and the uncertainty of distance measures used to capture structure and interdependences between objects (housings or housing submarkets) distributed over space.

More specific objectives, both in terms of theoretical and empirical perspectives, can be described as follows:

Theoretically;

- To understand historical evidence of urban growth and the residential housing market that contributed to different perceptions and functions of space.
- To analyse how different urban study disciplines, namely in urban economics, urban geography and urban planning, dealt with the notion of space over time.
- To collate available methodologies and techniques in the literature, that contribute to an understanding of the structure of space within housing markets (heterogeneity and dependence) in an urban context.

Empirically;

- To analyse the willingness-to-pay for increments in the corresponding structural and location property characteristics in the housing urban area of Aveiro and Ílhavo (two municipalities located in the coastal region of Portugal), giving some insights into how households choose their dwelling units.
- To define and to describe housing segmentation in the urban centre of Aveiro-Ílhavo.
- To develop, apply and evaluate a methodology to examine the role of space in urban areas in terms of spatial interaction based on factor analysis and an abstract notion of geographical space.

Based on the empirical research this thesis addresses two main research questions:

1. To what extent is space important in a specific example of the urban housing context (Aveiro and Ílhavo)?
2. What is the added value of considering the multi-dimensional non-Euclidean space to analyse the spatial spillovers across submarkets?

To achieve the main objective, answering at same time the previous research questions, a framework and corresponding methodology are developed, based on a statistical factor analysis and on a multi-dimensional non-Euclidean notion of space, to analyse and understand the importance of space in the urban housing context. Three characteristics of space are highlighted: i) spatial heterogeneity (existence of different spatial housing submarkets in a housing market context); ii) spatial spillover or spatial dependence (existence of influences from housing located in the neighbourhood); and iii) spatial scale (existence of different vertical territorial levels to analyse spatial structures). In short, the proposed methodology analyses the spatial heterogeneity across submarkets and unrestricted spatial spillovers between submarkets, considering two different spatial scales.

Most applications in spatial econometrics studies typically use Euclidean notions of distance to analyse geographic interaction between spatial objects (houses or housing markets). However, there is no reasonable explanation for the fact that spatial interactions should need to be limited to geographic or bi-dimensional Euclidean distance. Thus, conceptual aspects of space are quantified and empirically studied using spatial econometric methods, without previously assuming any fixed and known pattern of spatial interactions. Rather, new methods based on the factor model are developed, to shed light on the potential drivers of spatial diffusion in residential value.

In short, the basic argument of this dissertation is that traditional methods of analysing spatial interactions based on a pure geometric notion of distances may give erroneous insights into the features of the housing market.

I.4. Outline and methodology of the dissertation

In order to achieve the objectives mentioned above, the thesis is composed of five parts, including this introductory part and is laid out as follows.

Part 2 defines the theoretical background and is organized into three main chapters. Chapter II seeks to understand historical evidence of urban growth and the residential housing market that contributed to different perceptions and functions of space. Chapter III lays out the theoretical dimensions of the research, providing a framework to understand the importance of space in the urban studies literature, more specifically, in the scientific domain of urban economics, urban geography and urban planning. And finally, in Chapter IV, methods and techniques to analyse the housing urban market, both in terms of spatial and temporal perspective, are addressed. This last chapter highlights the importance of data analysis techniques, both to support the theories explaining the mechanisms underlying the housing market dynamics, to develop decision support tools and to sustain the design and evaluation of policies.

Part 3, composed of two chapters, presents the development of a new framework (Chapter V) and the corresponding methodology (Chapter VI) to capture the commonly discussed features of spatial housing data, spatial heterogeneity, spatial dependence (considering unrestricted spatial spillovers between submarkets), and spatial scale. This methodology, based on factor analysis and a multi-dimensional non-Euclidean notion of space, is placed within the context of the emerging literature providing insight into unknown spatial interactions.

In Part 4 a multi-dimensional non-Euclidean notion of space is empirically developed to understand spatial interaction in the housing market within the context of Aveiro and Ílhavo (Portugal). Two different datasets are used to capture the scale effect in the analysis of spatial heterogeneity and spatial dependence, that is, to assess the importance of spatial scale to understand the urban housing spatial patterns: firstly, considering the urban housing market of the city of Aveiro, and subsequently encompassing a more peri-urban territory, including two municipalities of Aveiro and Ílhavo. Some insight into how to produce an appropriate database to analyse urban housing markets is also addressed in this fourth part.

The last part draws upon the entire thesis, tying up the various theoretical and empirical strands, discussing the implications of the main findings, limitations, and important insights for further research on the topic of housing markets.

The main structure of this thesis and how the role of space in the urban housing market has been analysed in the research is shown in *Figure 1*.

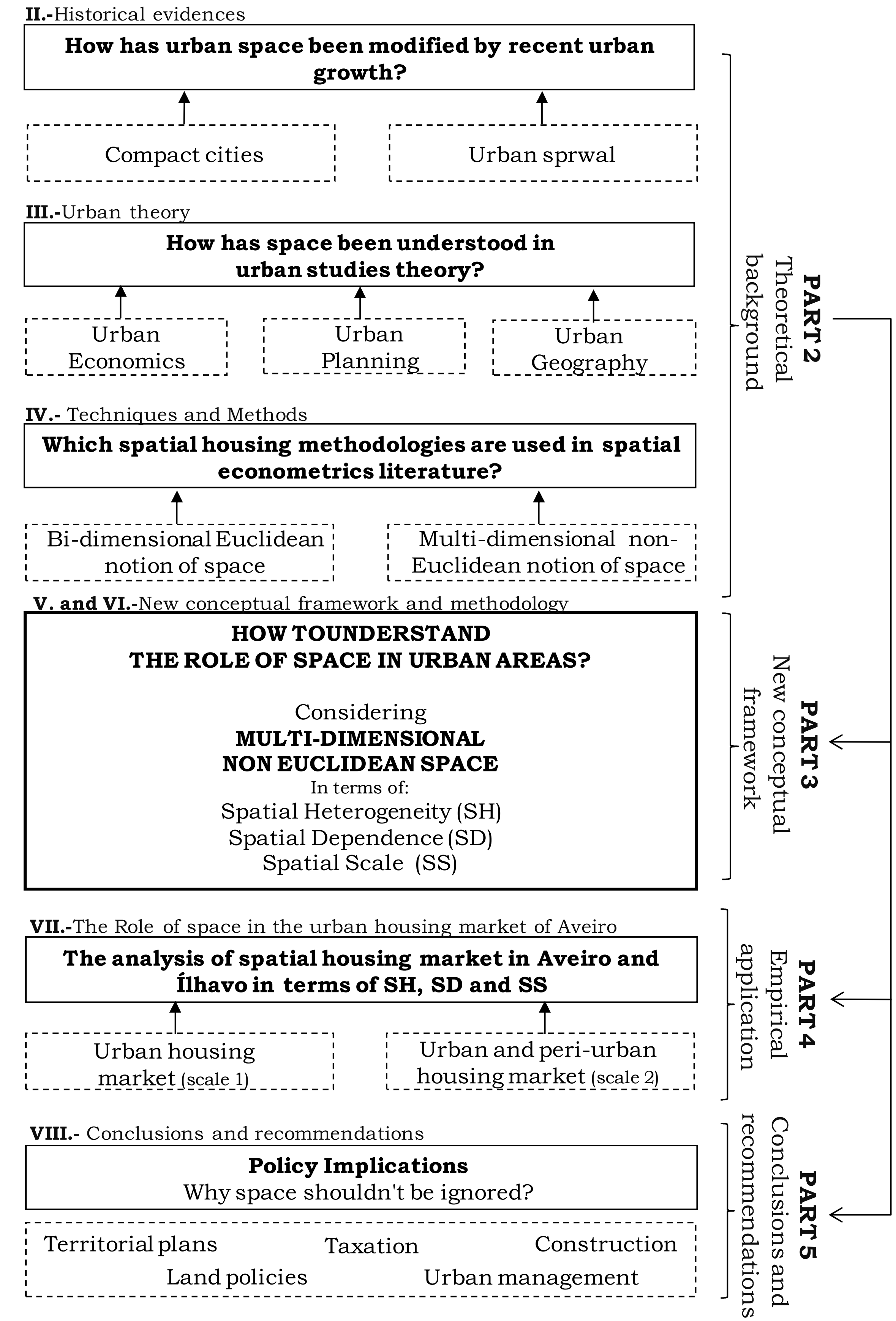


Figure 1 - Main structure of the research

PART 2

Theoretical Background

II. Historical perspective of urban growth

Cities are permanently changing, being the object of an ongoing process of urbanization, simply defined as the change from a predominantly rural to an urban society, and therefore urban landscape. It necessarily involves an increase in the number of people living in urban areas, becoming over the last centuries the most important space to live. This argument is widely sustained by the fact that cities account for half of the population of the world. According to the projections of United Nations, in 2008, for the first time in history, more than half of the world's population lived in cities; this is expected to increase to 60% in 2030 and 70% in 2050 while in the case of Europe, more than 80% of its inhabitants live in cities. These transformations have renewed interest, not only when thinking about the role of cities in the international system, but also on how spatial location and relationships of various physical objects within and between urban spaces are organised and have been changing over time.

In the 19^{th} century the urban areas were concentrated around urban centres, places of economic opportunities and where the costs of transportation and trade were minimised. Later on, in the latter 19th century, with the industrial era, and because of technological developments in transport, mainly train and trams, cities have been transformed into more disparate structures. The most notable urban transformations in cities were the enormous spatial expansion of the suburbs. Several sub-centres started to emerge in a process of territorial decentralization of labour, residences and other activities. The technological revolution of the automobile lead to a new stage of urban development (see Newman and Kenworthy, 1989; Hall, 2002; Bruegmann, 2005 for a comprehensive analysis of the urbanization process).

In the emerging cities, patterns of mobility caused by the development of transportation and communication systems, among other factors, have transformed social and spatial relations; old cities characterised for being compact and continuous have been transformed into more fragmented and disperse urban areas, where urban settlements interpenetrate with green areas, transforming rural land into building space. Of course that the development of such urban forms did not result in a common and standard urban spatial pattern, for example: in Europe, the expansion processes and urban growth dispersion appeared in contrast to long-established dense urban centres and links with pre-existing rural nuclei. In the United States of America, the unlimited outward extension into undeveloped areas (commonly in areas with limited infrastructure or public services) and of low density[5] is often referred to as traditional, the so-called urban sprawl. The dream of suburban home ownership, leading to the urban sprawl, is particularly near and dear in the United States, where the new developments are built traditionally beyond the metropolitan core (Burchell, Downs *et al.*, 2005). In short, the pattern of urbanization in Europe is more dense, more centrally orientated (essentially in Western Europe), and evenly distributed (essentially in Central Europe), when compared with the United States of America, where the notion of sprawl is much more visible (see Cheshire, Summers *et al.*, 1999 for a historical understanding of the process of urbanization in Europe and United States of America).

Urban sprawl, being a term widely used in the literature, does not have a clear definition and is the subject of many contradictions.

Regarding the definition, and following Brueckner (2000, p.163), sprawl means different things to different people, but in its simplest form can be characterised as "(…) *excessive spatial growth of cities*". The author considers that cities must grow spatially to accommodate a growing population, however the growth is much more than what would be considered adequate. Despite the fact that there is an inherent difficulty in finding an accurate definition of urban sprawl, Nelson, Duncan *et al.* (1995) have summarised the various concepts of sprawl into the following definition: "(…) *unplanned, uncontrolled, and*

[5] Note that density, as mentioned by Burchell, Downs *et al.* (2005), has to be analysed in its context because what is dense in one place could be not considered dense somewhere else, for example, in the United States the densities are approximately one-tenth what they are in Western Europe. Urban sprawl can, for this reason, naturally occur in different ways, and it varies between different countries and regions.

uncoordinated single use development that does not provide for a functional mix of uses and/or is not functionally related to surrounding land uses and which variously appears as low-density, ribbon or strip, scattered, leapfrog, or isolated development." Other attempts have been made by many other authors, for example, Galster, Hanson *et al.*, (2001), which consider that there are different types of sprawls, defined by a low level of some combination of eight distinct characteristics or dimensions across many urban areas. Such dimensions are: density, continuity, concentration, clustering, centrality, nuclearity, mixed uses and proximity The latter, also suggest that is more fruitful to understand the phenomena of urban sprawl as a process of development, that occurs over some period of time as an urban area expands, rather than a static condition of a specific distribution of land use. Following this argument, the authors pointed out that, because of social and economic reasons, the demand for space in the city can decrease, while, it may increase on the outskirts. In this way, and in line with all spatial economic theories, whose fundamental assumption is that having good access is more attractive and has a higher market value than peripheral location, house pricing and the level of income are, therefore, key elements in this urban spatial development. These effects of non-contiguity of spatial dependences give important insights for the argument developed in this research, that is, the notion that spatial interactions are much more complex than the simple analysis of a multi-dimensional Euclidean notion of space.

Contradictions arise directly from the elements contained in the definition of urban sprawl by Nelson and Duncan *et al.*, (1995). It is implied that this type of urban growth is characterised by an insufficient political control ("unplanned", "uncontrolled" and "uncoordinated" are the terms that support this argument), however, it is not absolutely correct, since municipalities are responsible for the promotion of policies of urban sprawl based on a goal of attracting new residents by promoting land at very low cost when compared with consolidated areas. Hence, for an individual household, the cost of locating in the periphery is much lower than in a high density central area, but in general, the whole system of suburban sprawl is more expensive to operate and more costly to maintain. For this reason, dispersed urban growth or urban sprawl of development has been viewed negatively in the planning literature. Factors related to: i) the excessive land consumption, imposing a higher and a never-ending spiral of costs to the municipalities (or to other government administrative level) due to the provision of services and infrastructures to new areas; ii) the traffic congestion due to increased commuting and consequent increase of the air pollution and other risks; and finally, iii) the socio economic segregation due to the concentration of poor households far from the city centre, where the price of the land is high; are all considered by Duany, Plater-Zyberk *et al.*, (2000), Carruthers and Ulfarsson (2002), Burchell (2000) and Burchell, Downs *et al.* (2005) as important characteristics of urban sprawl that do not sit well with a sustainable land use management policy.

Despite its contradictions, there is a general agreement that the car is placed at the centre of the problem; that is, urban sprawl is reinforced by an increased use of the individual car, without which the idea of living towards the outskirts of the urban area would not be so attractive. Thus, this dependence generates urban morphologies appropriate to the car, resulting in a vicious circle: the car has shaped the dispersed urban form and has caused the cities to depend on the car (Newman and Kenworthy, 1989). Following this principle, it is easy to understand that this chaotically urban expansion occurred mainly in the suburban parts of cities, and is linked to increasing spatial segregated land use, considered an important characteristic of urban sprawl.

The main proposed alternative to the unsustainable urban sprawl has been the compact city model for urban growth (known as New Urbanism[6]). Support for compact cities has arisen in reaction to the effects of dispersed and continual expansion of cities outwards, the segregation of its residents and the role of the private car encouraging decentralization and contributing, in this way, to global warming through increasing CO_2 emissions. This new form of thinking about the city differs significantly from dispersed patterns of urban development, in the sense that is focused on: i) the urban intensification and stressing density, ii) the definition of the limits to urban growth, and at the same time iii) the encouragement of multiple and mixed use combining social and cultural diversity (CEC, 1990). Linked to the goal of sustainable development these compact city policies were implemented by many planning authorities, throughout Europe, following the guidelines of the "*Green Paper on the Urban Environment*", published in 1990, by the Commission of European Communities (CEC). Their

[6] Considered by Bruegmann (2005), the strongest attack on the alleged aesthetic deficiencies of sprawl.

vision for the future was based on the idea that the compact city would be modelled "(...) *on old traditional life of the European City*" (CEC, 1990, p.42). Nevertheless, there are serious doubts as to whether the compact city could fulfil its fundamentals of delivering sustainability. William (1999, p.175) sustains this argument concluding that "(...) *intensification policies are fraught with contradictions and difficulties*", not only because it is so difficult to implement, in the sense that it brings such radical change to the existent urban environment, but also for the reason that, urban sprawl is unsurprisingly attractive for individual homeowners and property developers[7].

A whole range of advantages and disadvantages of different types of urban growth are brought forward in the literature[8]. The most prominent among these is, as has been mentioned, the dichotomy between the compact city (related with the notion of high density urbanism) versus the sprawl city (related to the notion of low density urbanism); the latter has come to dominate the urban environment in the past fifty years (Jenks, Burton *et al.*, 1996). A full discussion of the theory of urban development lies beyond the scope of this study. Nevertheless, apart from the lack of consensus about the specific role of academics, planners and politicians to regulate spatial development of cities, what should be noted is the idea that the space and the way that the urban structure is organized are permanently changing: generally, the urban transformation is characterised from a compact and continuous to a fragmented and dispersed type of occupation, which reinforces the need for analysing space with new methodologies, able to capture the logic of these more complex spatial structures and spatial interactions. Cities have expanded beyond their territorial boundaries and their CBDs are no more the unique centre of a city. The existence of sub-centres (leading to the existence of different submarkets) with different levels of importance among the space (different levels of economies of agglomeration and externalities) is a phenomena of the recent phase of urban transformation and of which the spatial econometric methods should be aware.

The main urban transformations and the debate about the major challenges of the urban spatial development have been presented above. From

---

[7] In 2000, a survey of Americans carried out by the National Association of Home Builders showed that most people want to have their own homes in their own lots.

[8] As referred by Malpezzi and Guo (2001), the literature on urban sprawl is huge, but for a broader review see the seminal work of Ewing (1997) and Gordon and Harry (1997).

the debate, strong arguments emerged referring to the compact city as the most sustainable urban form, in contrast to the development of urban sprawl, where the inefficient use of the land, the increase in energy consumption and air pollution, and increased social and geographic segregation, are all considered challenging factors for urban development. Thus, the above complexity expressed either by urban sprawl or the compact city (linked with the concept of the New Urbanism) emphasizes the particular importance of not reducing the notion of space to the essence of geometry.

III. Spatial theories and theoretical models in urban studies

This chapter develops a framework for understanding the importance of space in urban studies. The aim is to find in the literature different philosophical perspectives and conceptual models to analyse space. Since no research theory takes place in a philosophical vacuum is important to understand how the most important urban studies framework provides a distinctive view of the nature of space.

Urban studies has long been an interdisciplinary field, in which several disciplines contribute to an understanding of how socio-economic processes produced and transformed urban space: geography, economics, planning, philosophy, history, anthropology, sociology, among others. Given the impossibility of studying them all, this chapter focuses on three main urban study disciplines: urban economics; urban geography; and urban planning; while it is recognized that the delimitation of these specific domains is somewhat ambiguous and controversial. Thus, interdependences between such areas are clearly evident, for example: geography issues are a key aspect in economics; philosophical and social perspectives are predominant elements in geography, etc.

Before embarking on the central issue of this chapter, a brief clarification of some concepts related to space it is presented below.

III.1. The notion of space

The debate about space (considered a more abstract concept than place) has a far longer history and became an important topic in a wide range of scientific domains, including urban studies disciplines.

Space is a concept that refers to various things with a variety of uses and meanings to different people (Massey, 1992). Starting from the definition of The Pocket Oxford Dictionary (*cited in* Couclelis, 1992), space is defined as "(...) *interval between two things, this regarded as empty space of matter, sum of these as opposed to matter, this together with the room taken up by matter regarded as containing all things, any part of such space, regions beyond ken, a distance, an area, room available or required a period of or interval of time*". From this definition, space means one thing and its opposite. As mentioned by Couclelis (1992) space is viewed as the gaps between things but also as a larger container into which thing are inserted; space is at same time expansive and confined: it is empty of matter and defined by the matter. This concept of space gives the idea of a neutral, homogeneous and insignificant, meaningless space, where only things which occupy space are of significance to define space itself (West-Pavlov, 2009). This concept is in line with the notion of a bi-dimensional Euclidean understanding of space, conceived as abstract geometries (distance, direction, size, shape, volume etc.) and disconnected from meaning and values. However, Lefebvre (1974 [1991]; 1979), among several other philosophers, sociologists and geographers, has been developing seminal works underlying the importance of the symbolic meanings and lived space, which emerged in opposition to the pure notion of the geometry of space. Since space is defined through the social relations, emerging from different human interventions and several contexts, the focus should be, as according to Aase (1994), on the understanding of those contexts rather than searching for a real, once-and-for-all definition of space. The same argument is emphasised by Harvey (2006) where the author pointed out that is the diversity of contexts (metaphorical, liminal, personal, social or psychic space) and the range of applications (such as, spaces of fear, of play, of cosmology, of dreams, of anger, of particle physics, of capital, of geopolitical tension, of hope, of memory, or of ecological interaction) that define the meaning of space.

The complexity and extent of the concept of space is also justified by the use of some words that can be confounded with space itself, such as, location, place, territory, locality, local, spatiality, region etc.. In short, location is a specific point or area in space, in which an object is precisely referenced by a system of coordinates, for instance, or some other distance measures. Place is a wider concept referring to a description of the human and physical characteristics of a location. According to Gieryn (2000), place is space filled up by people, practices, objects and representations. Place is a special site in space with a particular location and having finitude but should not be confused with the use of geographic metaphors, such as, boundaries and territories that define conceptual or analytical spaces. In places, boundaries are elastic. Unlike places, territory does not signify the possibility of differentiation. It is a term, depending on the spatial scale, used to reference the area of bounded space occupied by an individual or a collective (Storey, 2001). The idea of precisely defined boundaries is also pointed out by Sack (1986, p.19) which defines territory as the geographic area in which an individual or group attempt "(...) *to affect, influence, or control people, phenomena, and relationships by delimiting and asserting control over a geographic area"*. The author also emphasised that a territory is a kind of place, however, unlike many ordinary places, territories require constant effort to establish and maintain, and are not created by circumscribing or delimiting an area on a map. This space becomes a territory *"(...) only when its boundaries are used to affect behaviour by controlling access"* (Sack, 1986, p.19). A further discussion about various dimensions of space can be seen in Sack (1986), Couclelis (1992), Storey (2001), and Massey (2005), among others.

In the opinion of Harvey (2006) space can be listed as one of the most complicated words in our language, not only because of any inherent complexity of the notion of space itself, but the permanent modifications that the concept suffers. In short, space can be more than bi-dimensional Euclidean geometry and can be more than a medium in which social, economic and political processes operate. These nested hierarchies of different concepts of space are illustrated in

Couclelis (1992) and is shown in the *Figure 2*. Several terminologies are associated with different types of spaces, as well.

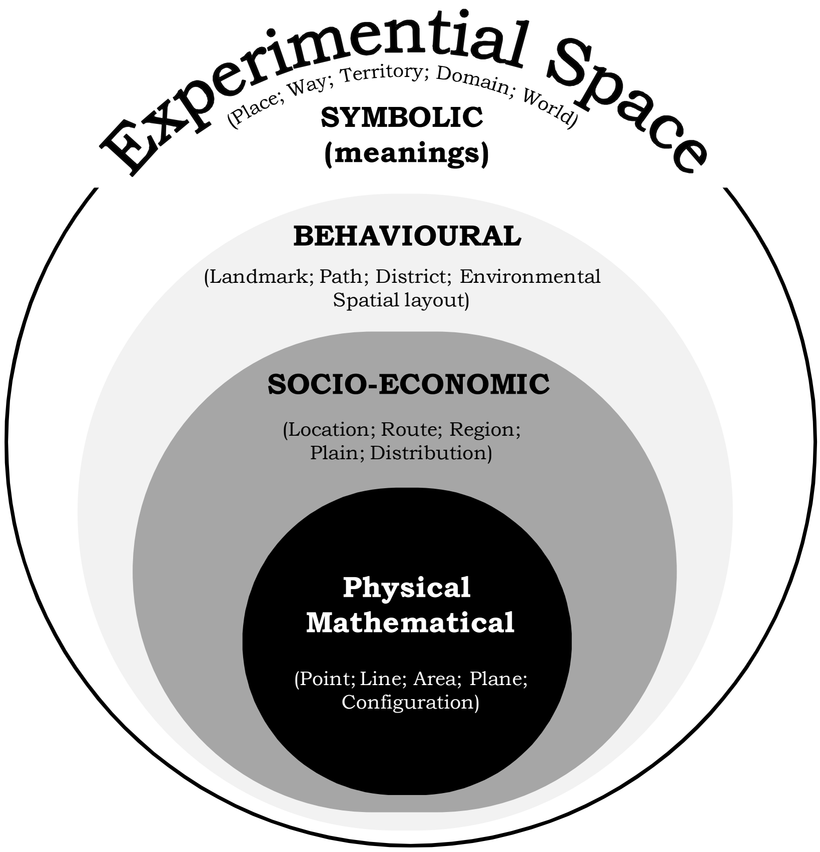


Figure 2 - Mathematical, Physical, Socioeconomic, Behavioural and experiential Space as a Nested Hierarchy.

Source: based on Couclelis (1992)

III.2. The role of space in urban economics

A literature review to address the importance of space in the urban economic field is presented in this chapter. This analysis focuses on the main causes of the formation of the various types of spatial organization.

As pointed out by Ekelund and Hébert (1993)[9] the consideration of the influence of spatial (geography and distances) questions in economics dates back to the 18th century and James Stuart (1767), Adam Smith (1776) and Abbot Condillac (1976). Adam Smith in his book "*The Wealth of Nations*" (1776) assumed the importance and the effect of space in economics, when he considers: i) the division of labour determined by the size of the markets and the importance of concentration for protectionism (spatial scale as a relevant dimension); ii) the prices of the good as determined by spatial variation of the production costs (assuming spatial heterogeneity); and iii) the importance of the emergence of cities and its relationship to our understanding of domestic trade (spatial spillover effects being a preponderant element in the analysis). However, despite its recognized importance, the spatial dimension has sometimes been overlooked in urban economic models.

III.2.1. Schools, theories and scale approaches of urban economics

Urban economics is the study of economies that are organized as urban areas. Depending on the ethical objectives and type of influences, four different schools of thought can be defined, as according McDonald and McMillan (2010): i)

[9] *Cited in* Backhouse (2004)

mainstream (neoclassical) economics; ii) behavioural economics; iii) conservative economics; and iv) Marxian economics.

The first category is where most urban economists belong; having as the main ethical objective the maximization of the utility of members of a society which act rationally in pursuit of their own purposes. The idea that markets operate in a self-regulated system and the assumption of oversimplified notions of space were the subject of criticism by the other three schools of thought: behavioural economics considers that mainstream economics uses unrealistic assumptions about human behaviour; the conservative economists argue that they do not pay attention to human freedom as an ultimate goal; and finally, Marxist economics considers that class struggles are ignored.

Economists within the school of behavioural economics are a combination of psychologists and economists which investigate what happens in the market where agents with human limitations are included. Thus, it is a field of research that is focused on investigating reactions of agents in different situations, stimulus and interactions. Is not assumed that these agents have perfect information and able to be fully rational. However, this position does not imply a rebuttal of the neoclassical approach to economics, based on utility maximization, equilibrium, and efficiency (Camerer and Loewenstein, 2004). It only considers that people have limited power of problem solving; are influenced by conventional wisdom and by how problems are framed, have limited willpower, and tend sometimes to sacrifice their own interest for the sake of someone else (McDonald and McMillan, 2010). According to Camerer and Loewenstein (2004) the methods used in behavioural economics are not different from those in other areas of economics. Most of the papers in behavioural economics adjust one or two assumptions in standard theory in the direction of greater psychological realism[10]. Behavioural economics has proved to be important in urban economics and particularly in real estate, in the sense that markets, in general, and real estate markets, in particular, are not fully efficient, essentially because market prices do not completely reflect all available information and bubble episodes or social contagion can easily occur.

The third main school of urban economics thought, mentioned by McDonald and McMillan (2010), is that of conservative economics, which is

[10] See Camerer and Loewenstein *et al.* (2004) for a representative collection of papers on behavioural economics.

based on a rational examination of the ends before discussing the means to achieve those ends. Quoting Campbell (1982, p.38) conservative economics has "(…) *a great respect for one's own, family, blood, private property, territory and nation, but it refuses to spin theoretical systems on them as absolute principle"*. This respect is considered by McDonald and McMillan (2010) as closely associated with the development of the *laissez-faire* market economy.

Lastly, the main theme of Marxian economics is the role of labour, associated with the two primary factors of production, capital and labour, in the development of an economy. This school of thought, based on of the writings of Karl Marx (1818-1883), provides a critical analysis of capitalism and identifies its fundamentally conflictual and exploitative character, where the dominant issue is the class struggle. According to McDonald and McMillan (2010) the term Marxian economics cannot be considered a homogeneous body of work, because considerable diversity of debate over the interpretation and the validity of Marx's work has been developed. In the specific context of urban economics, Marxist theory considers that capital accumulation is taking place in the suburbs, at the least expensive location, and the consequences of this situation is the loss of economic opportunities for central city residents and financial decline of various central cities and the appearance of social problems associated with the urban underclass (see Edel, 1992, to introduce readers to the main concepts of Marxism in the context of urban and regional economics).

A different taxonomy of how urban economic theory is organized can be found in Fujita's (1989) seminal work. The author describes two types of theories where the regularities in the spatial structure of different urban areas can be analysed: i) the *positivist approach*, which provides explanations for the existence of regularities in the spatial structure of different urban areas (it is a mere confirmation of regularities), suggesting testable hypothesis for further investigation; and ii) the *normative approach*, which identifies the efficient spatial structure and size of cities, suggesting how to achieve them.

The spatial scale is another important criterion to categorise the analysis of the urban space organization. MacDonald (2010) distinguishes two different kinds of approach. The first is the *macro scale level,* which focuses essentially on the growth or decline of the economy of the urban area, and how urban areas are organized in a larger economy. In this perspective, the urban economy is analysed as an integrated system. A second approach is at *micro scale level* which seeks to understand the spatial patterns of economic activities within an urban area. For instance, studying decisions of where to locate by households, firms, and industries is included in this perspective of analysing urban space[11].

Independent of the school of thought (mainstream economics; behavioural economics, conservative economics and Marxian economics), the type of theory followed (positivist or normative) and the level of the approach (micro or macro) urban land use models do not explain all regularities that occur in space.

In the remainder of this chapter an overview of a selective set of models is presented, developed over urban economics history, to explain the location of agents (individuals, firms, organizations) and their behaviour in a competitive market. From von Thünen to New Economic Geography, the main drivers of the more influential location theories, highlighting the three distinct but related aspects of space (spatial heterogeneity, spatial spillovers and spatial scale) are described, as well as the complexity of modelling these particularities in urban economics.

III.2.2. Classical location theories

The pioneering theories of location have always had the intention of explaining location and the spatial organization of several economic sectors across space: it was the case of von Thünen, in structuring of land in agriculture; Weber in the industries and Christäller in the services, just to mention a few. Paradoxically, in spite of the importance of these contributions, space has remained as a secondary figure in the context of economic analysis. In this respect, the contribution of Walter Isard (1956) was decisive. In his work, *"Location and Space Economy"*, the principal founder of the discipline of Regional Science, developed principles for a general theory of location, attempting to unify location theory and neoclassical economics, pulling together classical location theories in an intelligible whole (Correia-da-Silva, 2004).

[11] Note that, the evolution and changes of urban patterns are also an important object of analysis in urban economics.

The history of spatial economics is very rich, but, at some point considered perplexing by Fujita (2010). The richness of this evolution is illustrated by the variety of pioneering ideas that have been developed periodically by great location theorists, geographers and economists. Since the work of Johann von Thünen (1826 [1966]), urban economic theory advances rapidly. He is considered the "*founding god*" of economic geography and location theory (Samuelson, 1983 p.1469 *cited in* Fujita and Krugman, 2003). Although following von Thünen´s work, other important authors must be mentioned, such us: Launhardt (1885 [1993]), Marshall (1890), Weber (1909 [1957]), Hotelling (1929), Ohlin (1933 [1968]), Christaller (1933 [1966]), Palander (1935), Kaldor (1935), Lösch (1940 [1954]), and Isard (1949), only to mention the most important contributions that took place in the first half of the last century. The perplexity emphasized by Fujita (2010) is based on the apparent contradiction between the long and deep intellectual tradition of spatial economics and the peripheral situation that spatial economics occupied in economic science. Quoting Paul Samuelson, Fujita and Krugman (2003) remember that, almost two centuries ago, the seminal work of von Thünen appeared in the opposite direction to the mainstream trade theory, where spatial issues were frequently neglected. Paul Samuelson, to characterise von Thünen's model, states in "*Thünen at Two hundred*" (Samuelson, 1983, p.1482) the following: "*Ricardian trade theory traditionally assumes zero factor mobility and 100% commodity mobility between countries or regions. Thünen's model works out the opposite case. Within a region, labour moves freely (on immobile land); goods move only at a cost. Where labour will locate was not a question that trade theory considered, but Thünen did.*"

Throughout history there have been a variety of contributions to the body of literature referred to as urban economic theory. Back to its origins, nearly 200 years ago, 19th Century economists David Ricardo and Johan von Thünen[12] developed models of agricultural location that were later applied by William Alonso to an urban land context. The general idea of Ricardo and von Thünen's model is to explain why there are differences in the cost of producing agricultural products, resulting from utilization of land of different quality and location. While Ricardo focused on differences in soil fertility, von Thünen concentrated his analysis on differences in land location. In the next paragraphs, the necessary

[12] They started out from Adam Smith's idea of "*economic man*": that the farmer is expected to maximize his profit ("*economic rent*") from his farmland (see Ricardo, 1821: Chapter 24 - Doctrine of Adam Smith concerning the Rent of Land).

discussion is laid out for understanding the variation in land rents in a more comprehensive way.

David Ricardo, English political economist (1772-1823), in his book "*On the Principles of Political Economy and Taxation*", published in 1821, developed a model that explains the variation of rents according to the fertility of the soil. The basic idea of his model is that farmers will be willing to pay more for high-fertility land, being available to pay more for it. As a result a variation of land rents over the space will appear. Ricardo introduced crucial concepts in the theory of land rent, namely: land varies in its natural endowment or advantage for the user; land of a given level of natural endowment or advantage is fixed in supply; land market is governed by perfect competition; and finally, land rent is determined by the natural endowment or advantage of land (see, Fujita, 1989; Arnott and McMillen, 2006; McDonald and McMillan, 2010).

In the same decade, a North German landowner from the Mecklenberg area, Johan von Thünen (1783-1850), in his book "*Der Isolierte Staat"* (*The isolate state),* published in 1826, incorporated another element into Ricardo's land rent model, that of transportation costs. Von Thünen's model also explains the use and structuring of land in agriculture, however it assumes that the fertility of the soil is the same everywhere. The quality of land varies not with fertility but with respect to location or distance to the marketplace. Farmers cultivate different crops in a uniformly fertile terrain and must ship their product to a central market place to sell it. As such, the land rents vary according to access to the central marketplace. Since shipping is costly, farmers will bid more for land closer to the marketplace.

Von Thünen's *Isolated State Model* is a clear example of a normative model and is partly based upon empirical evidence relating to the economic conditions in the early 19th century. It gives some insights on how the land is distributed in the case of free competition among farmers and landowners. Thünen showed that competition will lead to a gradient of land rents from his maximum values in the marketplace (in the town) to insignificant values at furthest limit of the cultivation. Thus, each farmer is confronted with a trade-off between land rent and transportation costs, which are proportional to the distance from the centre. Assuming these principles, von Thünen was the first author to develop a basic analytical model considering the interactions between markets, production, and distance (Fujita, 2010).

Figure 3 schematically illustrates the basic principles of the land rent theory in the case of four agriculture crops. It assumes a centre which represents a desirable location with a high level of accessibility, and several land uses located at 1, 5, 25 and 50 kilometres from the city centre. The closest area, within a radius of 1 kilometre from the city (land use 1), becomes profitable to produce product 1. The descending straight line represents the profit for a specific type of crop at a certain distance from the city, thus for land closer to the city, farmers are willing to pay a land rent which is at most the profit they make at that location. Equilibrium bid-rent curves are shown in the upper part of *Figure 3*, that represent the rent that farmers would be willing to pay at any given distance from the city centre for four different types of crops. More specifically, the heavy line, the envelope of bid-rent curves, defines the rent gradient. Along each of the four segments of that line, farmers of one of the crops are willing to pay more for land than the others. The result of this allocation is a concentric distribution of cultivation, represented as quarter sections of the layout in the bottom half of *Figure 3* (see, among others, the following references: Fujita, 1989; Dicken and Lloyd, 1990; Fujita, Krugman *et al.*, 1999).

This representation contributes to defining concentric rings of cultivation around the market that emerge from the location occupied by the crop that offers the highest bid rent.

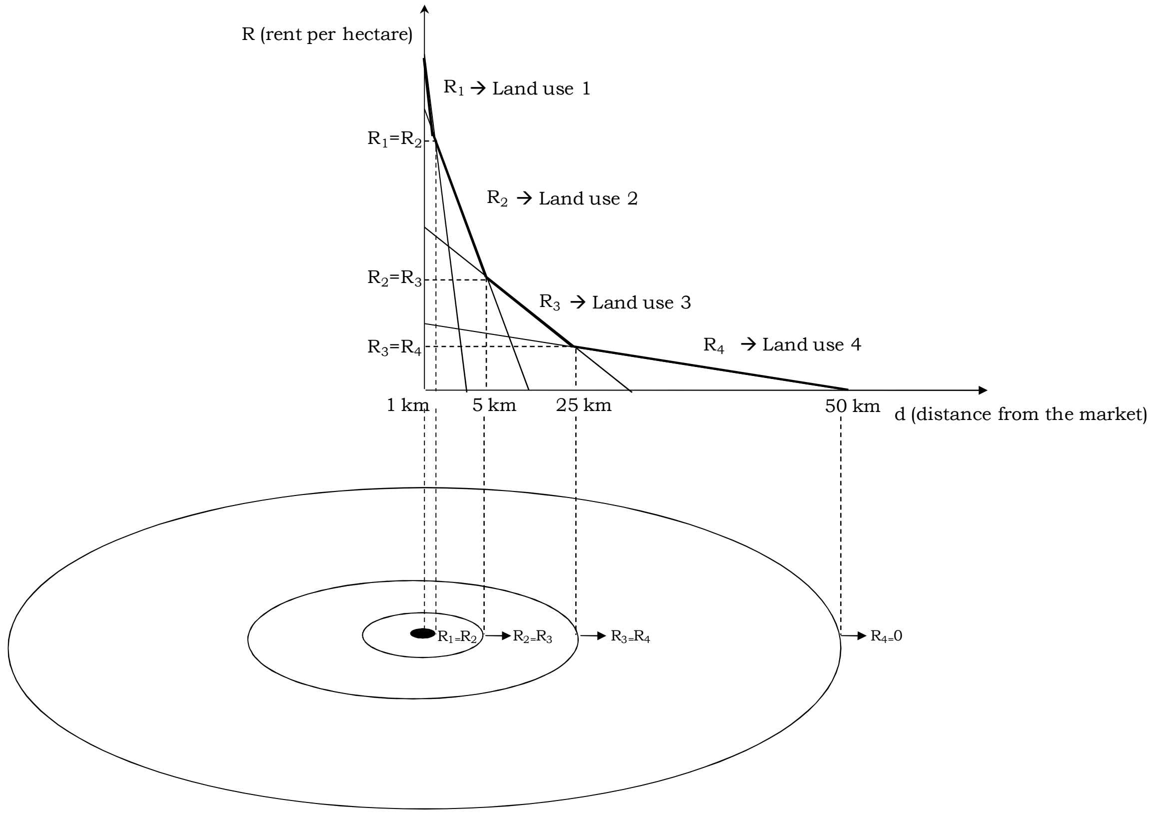


Figure 3 - Von Thünen land rent profile with four crops

Source: based on Fujita (1989)

The model, that compares the relationships between production cost, the market price and the transport cost of a crop is expressed as follows:

$$R = Y(p - C)YFm \qquad \text{Eq. 1}$$

Where:

R=Rent (e.g.: per hectare);
Y=Yield (hectare);
p=price for a unit of product at market;
C=production cost per unit of product;
YFm=total transport cost (per hectare);
F=Freight rate (per unit of product per kilometre)
m=Distances in kilometres from the central market.

Independently of the von Thünen tradition, which depends on an exogenous focal point for trade (the isolated sate), another tradition was developed following Weber's work (1909 [1957]).

Alfred Weber (1868 –1958)[13], with his book "*Über den Standort der Industrie*" (Theory of the Location of Industries) published in 1909, developed the first general theory applied to industrial location and has established the foundations of modern location theories. Weber's great contribution is in the domains of pure deductive theory (Correia-da-Silva, 2004). Weber's theory is a generalization of analysis developed by Launhardt[14] and seeks to define an optimal industrial location, the one with minimal production cost, based on three main factors: transport costs, labour costs and agglomeration economies.

Formally, the model can be described assuming a set of simplifications. Like von Thunen's agricultural land use model, Weber's model assumes that economic activities are placed in an isotropic or uniform space where several natural resources are ubiquitous. It means that there are no variations in transport costs except a simple function of distance. The economic landscape is considered as a given, composed by one market within an isolate region with no external influences; only the location of a new firm (plant) is determined, given the known location of the existing firms. The model also assumes a perfect competition, where both buyers and suppliers have perfect knowledge of market conditions; a firm produces a specific quantity of output and uses two inputs, with fixed technology coefficients and are either market oriented or resource oriented, while many production inputs such as labour, fuel and minerals are available at specific locations (Weber, 1909 [1957]; McCann, 2002; Ottaviano and Thisse, 2004). Theoretically, there must be a point in space at which these transport costs will be minimized.

The model has the main objective of minimizing the weighted sum of Euclidean distances from that plant to a finite number of sites corresponding to the markets where the plant purchases its inputs and sells its outputs. The weights represent the quantities of inputs and outputs bought and sold by the plant, multiplied by the appropriate freight rates. Therefore, the problem of

[13] He started his academic career in Germany as an economist, and then became a sociologist.
[14] The great contribution of Weber to the designated Launhardt-Weber model was the introduction of differential labour costs and agglomeration economies. Wilhelm Launhardt in 1882 showed how to determine the optimal location in a system with transport cost.

location comes down to the firm's willingness to be near the markets and the factors of production.

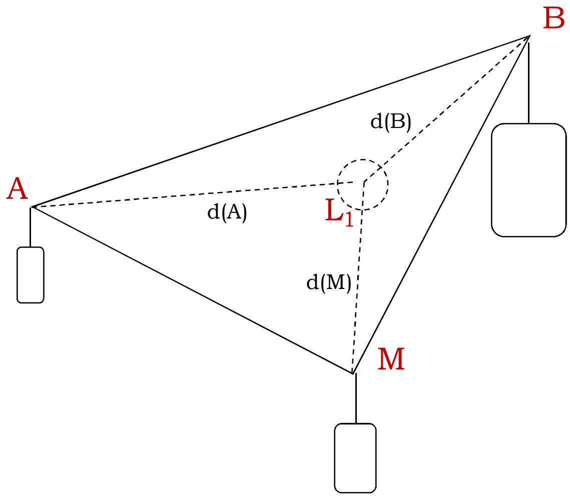


Figure 4 - Weber's Location Triangle

Source: based on McCann (2002)

Figure 4 illustrates the issue of minimizing transport costs, where, A and B are the source of raw materials and M is the market. Transportation is the most important element of the model, although adjustment effects caused by spatial variation in labour costs and other agglomeration economies are also considered. When production costs are the same everywhere, transport costs will be the determinating factor in the choice of location, and two situations can occur: on one hand, the optimal location will be close to the source of raw material if the transportation of raw material implies loss of weight (as a percentage of product loss); on the other hand, the firms will choose their location close to the markets if the weights of products are higher than the weights of the materials used in its production. The problem resides, as shown in *Figure 4*, in finding an optimal point (L_1) located at the respective distances of d(M), d(A) and d(B), that corresponds to the point that balances the triangle (gravity centre of the triangle). In Weber's model, in contrast with what is considered by Ricardo and von Thünen, the different uses of land are not important (both in terms of fertility or rent), only the optimal location.

Harold Hotelling (1895-1973) was also an important reference in spatial economics research. Hotelling (1929) also developed a location model, but went in a different direction. His model considers a duopoly where two firms compete for a location and for prices, suggesting that the firms compete in two stages: first, firms choose their location and then compete on price. Thus, the fundamental decision is a trade-off between central location, in which a large share of the market is captured, and more peripheral locations, which allow them to attenuate the intensity of competition. Firms produce a single product, the market demand is homogenously distributed over a finite segment and transportation costs are supported by the consumers that are evenly distributed over the linear market. In the case of two firms and considering these conditions, Hotelling's principle of agglomeration advocates that both firms will decide to locate in the centre in order to capture half market each.

Walter Christäller (1893-1969) and August Lösch (1906-1945) developed the theory of central place defined by a hexagonal pattern. Christäller's Hierarchy Principle shows how the range and type of goods in the city differs systematically according to the size of the city (Cheshire, 1999). At a conceptual level, market area theory as considered in this model is very similar to von Thünen's isolated town, however Christäller's took into account points of supply surrounded by spatially dispersed demand. Lösch (1940 [1954]) formalises a model of general spatial equilibrium based on original Christäller principals but in a somewhat more formal way. The main difference between the models of Christäller (1933 [1966]) and Lösch (1940 [1954]) is that the former assumes the spatially distributed units of demand are continuously distributed throughout space, whereas Lösch assumes that they are discretely distributed. Thus, Christäller's market areas comprise an infinite set of possible orientations, whereas, those of Lösch have only a finite set (Cheshire, 1999). The Christäller-Lösch problem consists, in a homogeneous economic space, of finding the lattice which minimizes the production and transportation cost, per unit area, for a given level of demand. The optimal configuration is obtained when each producer is equidistant from exactly six other producers, located at the vertices of a regular hexagon (Beckmann and Thisse, 1987).

In short, the relevance of classical location tradition relies on its simplicity in explaining the spatial distribution of economic agents (concentrating on a limited number of factors to capture the main structure of reality) and its provision of foundations for developing more complex theories and models.

III.2.3. Urban location theories and residential choice

The general monocentric model developed early on by von Thünen is still applicable to rural areas, but its relevance lies in the study of urban rents and the location of households and firms in cities. Von Thünen's model is now in its third century, and its importance to questions of economic geography is incontestable.

Several models have been developed to conceptualise urban spatial structure and consequently, to explain the complexity of both contemporary and primordial urban systems. According to Anas *et al.* (1998) and Fujita (1989) there are two basic standard models to understand residential land use: *monocentric models* and *polycentric models*.

A monocentric model is a model of demand for residential location, originated by Alonso (1964) that is formally equivalent to the land use model of a monocentric economy developed by Thünen (1826 [1966]). Following the pioneering work of Isard (1956), Beckmann (1957), and Wingo (1961), William Alonso[15] (1964) succeeded in generalizing von Thünen's and Ricardo's central concept of *bid rent curves* to understand land use patterns in urban areas (Fujita, 2010).

William Alonso is considered the founder of urban economic theory. In his book "*Location and Land Use*" (1964) he reinterpreted monocentric models and developed a modern version of Von Thünen's ideas (William Alonso's bid-rent function theory): the isolated state is replaced by the central business district (CBD); the agricultural land surrounding the city is replaced by residential areas, and farmers by commuters. This theoretical model is focused on location patterns within urban areas and it remains to this days the basis for an extensive theoretical and empirical literature (Fujita, 1989; Fujita, Krugman *et al.*, 1999; Fujita, 2010; McDonald and McMillan, 2010).

---

[15] He was an Argentinian-born American planner and economist.

The land use model of the monocentric city in its simplest form described the city as a circular residential area surrounding a CBD in which all activities of the city are supposed to take place and all workers living in the surroundings are supposed to commute to the centre (that is, households work in a single location in the city but choose between living in the city centre and out-of-town). The model also considers some assumptions about the spatial character of the urban area, namely: the city has a single centre of a fixed size (a CBD is *a priori* assumed)[16] where all jobs are located; the transport system is organized in a radial fashion, where only travel between places of work and residence is considered (accessibility is partially considered, travel within CBD is ignored); and the land is isotropic or homogeneous[17] (in which, as regards environmental amenities, no neighbourhood externalities are considered).

In this scenario, households and firms are willing to make bids for land at various places and the location decisions can be explained by comparing the bids made by different types of households and firms. It has been assumed that an individual, living in a house (out of town) that commutes to a job in the CBD, maximises their utility, paying a premium for sites that lead to lower commuting costs. Thus, equilibrium requires the household to locate where the marginal increase in commuting costs is exactly equal to the marginal reduction in land price (Meen, 2001; Arnott and McMillen, 2006).

The concept of bid rent function has an important role in this model. Hence, bid rent function b(x,u) at location x is defined as the maximum rent per unit land area that someone (households, firms or government) is willing to pay for a unit of land (as a function of distance from the central business area) in order to achieve a certain level of profits, satisfaction or utility (u) (Alonso, 1964; Fujita, 1989). Residential bid price would be expressed by the following equation (Anas, Arnott *et al.*, 1998, p. 1434; McDonald and McMillan, 2010, p. 86):

$$b(x,\bar{u}) = \max_{\mathrm{z,L}} = \frac{\mathrm{y - T(x) - z}}{\mathrm{L}} \quad s.t.\, u(z,L) \geq \bar{u} \qquad \text{Eq. 2}$$

[16] Considering the existence of only one centre is a major drawback of this kind of model because as cities grow they can generate secondary centres and as such the process of sub-centre formation within existing cities competes with the emergence of new cities.
[17] A homogeneous space is when the production set of a firm is the same in all locations and consumers' preferences are the same at all locations (Ottaviano and Thisse, 2004).

Where, $u(z,L)$ is the utility from a numeraire good z and a residential lot of size L; x is the location (distance from the CBD) of a specific household; T(x) is the commuting cost; and y is the household exogenous income (Anas, Arnott *et al.*, 1998; McDonald and McMillan, 2010).

The following equation represents the slope of the bid rent function. It is the most basic result of the monocentric model and expresses the additional transport cost for a household located a small additional distance dx from the CBD:

$$\frac{db(x,\bar{u})}{dx} = -\frac{\mathrm{T}'(\mathrm{x})}{\mathrm{L}[\mathrm{y}-\mathrm{T}(\mathrm{x}),\bar{\mathrm{u}}]} \qquad \text{Eq. 3}$$

To keep the household location indifferent, the lot rent must be lower at the more distant location, in order to comply with the following relation (Anas, Arnott *et al.*, 1998): additional transport cost

$$Ldb = -T'(x)dx \qquad \text{Eq. 4}$$

As verified by the previous formulation there is a family of residential bid-rent functions. According to Alonso (1964, p.59) a *Bid-Price Curve* is a "(...) *set of combinations of land prices and distances among which the individual is indifferent*". Three factors are important to characterise the bid price curve: i) each individual or household has her own bid price curve; ii) the bid price curve represents a given utility level, thus different utility levels or bid price curves can be represented; iii) prices represented by the bid price curve have no necessary relationship to actual prices, in other words, a bid price is hypothetical, meaning that, if the price of land were such, the individual would be satisfied to a given degree (Alonso, 1964).

Figure 5 represents the Alonso-Mills-Muth (AMM) model with different variation in land rents considering different users from the city centre. Rents generally tend to fall according to the cost of transportation, generating different bid-rent curves and corresponding to different forms of land use (retail, service, industrial, apartments, and single houses). Additionally, the figure represents a simulation of negative and positive externalities (caused by any urban amenity or natural resource) which influences the nature of the bid-rent curves.

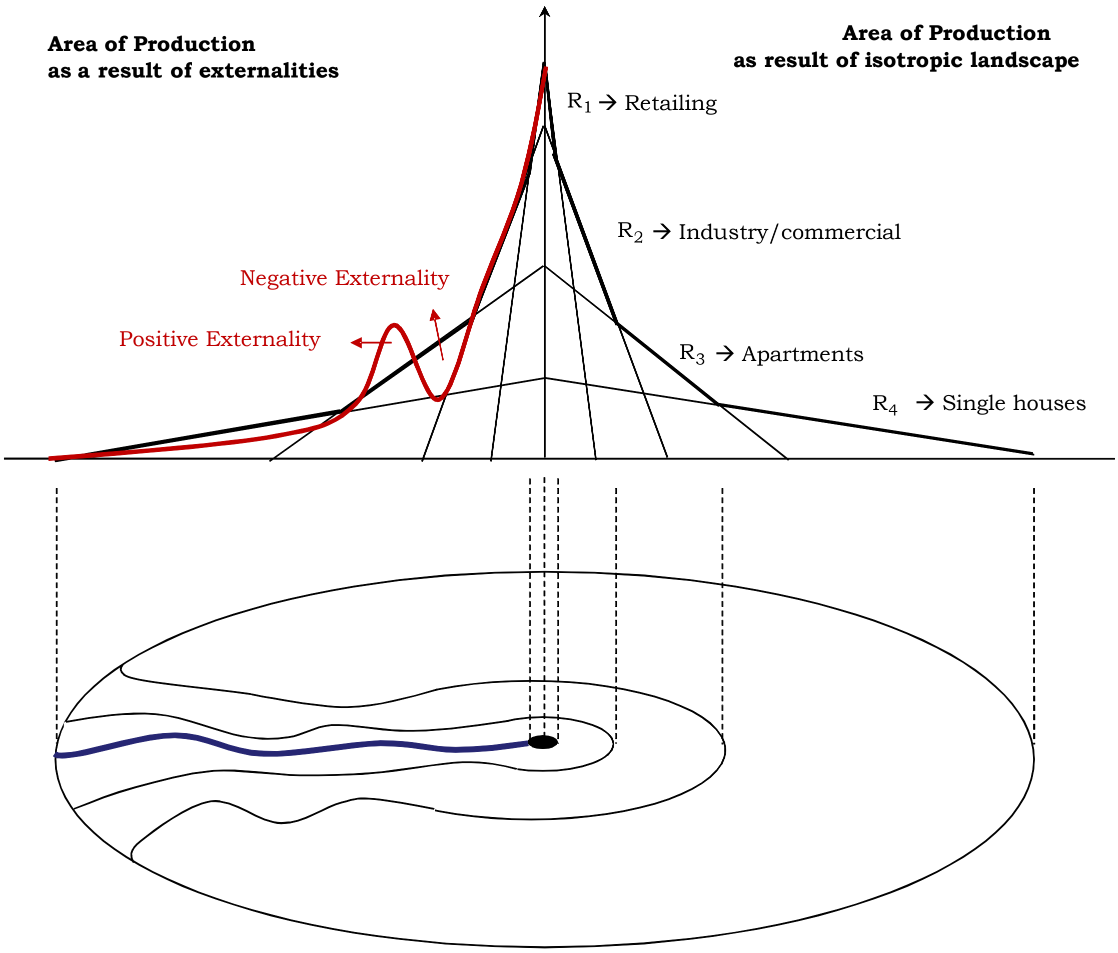


Figure 5 - Overlay of bid rent curves

(adapted from Hoover and Giarratani, 1999)

The complexity of urban land use patterns imposed some simplifications on the models of demand for residential location, in other words, many special characteristics of housing were ignored, or simplifying assumptions about space were imposed. According to Fujita (1989) two main reasons summarise the complexity of land use theory (easily extended to residential choice): the characteristics of land and the complexity of the set of judgements involved in a decision process by firms, households and governments. Land (and therefore housing) has some particular defining characteristics as an object. On the one hand, it is a commodity in the usual sense, and on the other hand, is completely immobile, meaning that each price of land (or housing) is associated with a unique location in geographical space (Fujita, 1989, p.34). The set of judgements that different territorial agents make can be summarized and represented by three basic factors (*Figure 6*): i) *accessibility*, which considers the pecuniary and time costs associated with getting to and from work, leisure and other such activities; ii) *space,* which regards size and quality of the house itself; and iii) *environmental amenities,* which includes natural features (scenic view) and neighbourhood characteristics (such as, quality of schools, racial composition etc.) (Fujita, 1989). The author includes budget and time constrains in the set mentioned above to make an appropriate residential choice. Some of the decisions are made sacrificing one of these factors. The environmental amenities dimension can be extended to the concept of an externality (positive or negative) and are crucial in determining the household's choice.

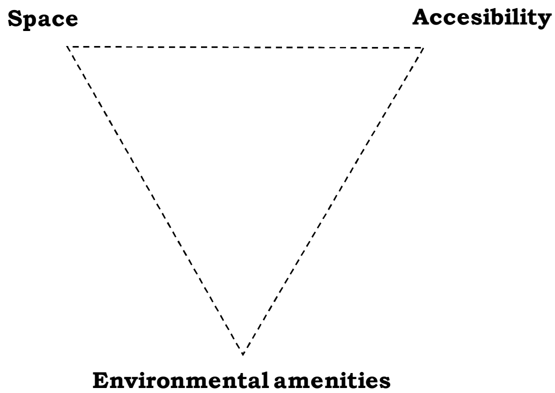


Figure 6 - Trade-off problem of a location decision

As described above in monocentric models the city is *a priori* assumed, the space is isotropic and homogeneous, there are no market externalities in the production and in the consumption, and finally the market is perfectly competitive[18]. Land use decision is based on the trade-off between accessibility and space, in which, the only spatial characteristic that is considered is the distance from the CBD in a unidirectional way. All other aspects of location are ignored, thus no agglomerative effects are presented. However, the monocentric city model developed by Alonso (1964) was extended to consider production, transport and housing by Mills[19] (1967; 1972b; a) and Muth (1961).

Alonso-Mills-Muth spearheaded a theory of the internal structure of a city emphasizing land markets and land use; which in combination with the concept

[18] In order to consider economic agglomeration, useful insights are provided by Fujita (1986), based on the spatial impossibility theorem of Starrett (1978) (see Mori, 2006).

[19] The author considers two important factors to help understand the evolution of the city: one encouraging agglomeration (economies of scale) and other encouraging diseconomies (commuting costs).

of Marshallian externalities, considered crucial in the formation of economic agglomerations, comprise the basis to the beginnings of the systems-of-cities literature within modern urban economics (Abdel-Rahman and Anas, 2004). Henderson (1974) and Fujita (1989) have the merit of integrating into a unified framework these traditional models to help explain the concentration of economic activities in cities.

Besides its simplicity, the monocentric model remained the most influential representation of the urban structure until the 1970s. According to Anas, Arnott *et al.* (1998) two of the most important facts about urban structure are provided by these models: the notion that density declines with the distance from the centre; and, the idea that most cities have been progressively decentralizing, for a century or more.

Nevertheless, the monocentric model became progressively less important in helping understand and explain internal urban structures; in part because of the geographical 'spreading out' of urban growth, induced by: the increasing decentralisation of economic activities; the evolution of new transport technologies and the mobility that comes from it; the variety and multiplicity of travel patterns and complex cross-commuting, and finally; changes in household structure and lifestyle. Other factors can be mentioned that support a polycentric form as a more appropriate approach to describe urban spatial structure: the significance of economic agglomeration for the distribution of firms and population, and the propensity of firms to cluster when spatial transactional costs are high (Davoudi, 2002).

The tendency for the cities to be larger, characterised by the dispersion of population and employment, justifies the requirement of other types of models. Because space is not homogeneous, economic activities (and individuals) tend to be clustered in several interacting centres of activity. Modern metropolitan areas are not monocentric and even those that still have a traditional downtown typically have a number of sub-centres that compete directly with the traditional city centre, in terms of employment and appraisable places.

In general, a polycentric model considers a large centre and a number of concentrated sub-centres, with high population and employment density. It results from a process which involves two opposing types of forces: agglomerative or centripetal and dispersive or centrifugal. Thus, activities tend to be located close to each other, with a high level of concentration, in the presence of positive externalities; and tend to be disperse in the presence of negative externalities (Anas, Arnott *et al.*, 1998; Fujita and Thisse, 2002). According to Mori (2006), the first polycentric model was developed by Fujita and Ogawa (1982), considering that the formation of multiple equilibria (centres) appear when commuting costs between firms become relatively high (the negative exponential function of the distance between them being a benefit from this interaction).

According to Davoudi (2002) sub-centres can emerge because of the following two reasons: old towns that have gradually become incorporated into an expanded urban area, and from newly spawned centres which appear in the nodes of transportation networks. Apart from the process of sub-centre formation, advances in transport and communications are considered the important factors in shaping the spatial structure of modern cities. In response to ongoing transformation in urban form, the planning community has tended to advocate policies aimed at reversing decentralization: reducing automobile use and revitalizing the downtown core (see Chapter II).

Based on a review of empirical descriptions of polycentric forms some evidence regarding the nature and the rules of sub-centres is presented by Anas, Arnott *et al.*, (1998): i) sub-centres exist in both new and old cities; ii) the definition of the number of sub-centres and their boundaries is a quite sensitive issue[20]; iii) sub-centres are sometimes arrayed in corridors[21]; iv) employment centres help to explain surrounding employment and population, because sub-centres are viewed as perfect substitutes; v) sub-centres have not eliminated the importance of the main centre, which continues to have larger densities and land price than any sub-centre; vi) most jobs are outside of the centres, mainly placed in the edge cities, which are well known as important sites of office location, demonstrating that they serve as nodes of information exchange; vii) commuting is not well explained by standard urban models, either monocentric or polycentric, where heterogeneity of idiosyncratic preferences for particular residences assumes a more important aspect in the explanation of urban residential location decision. This evidence emerged from large metropolitan

[20] Such sensitivity is not surprising since the spatial pattern is strictly related to spatial scales (more fine or more general) in which the different types of spatial agglomeration and its interaction are analysed).
[21] For example: the megalopolies corridor between Boston and Washington.

areas in the United Stated of America but can be generalised to other developed economies.

Although the definition of polycentrism seems clear, Davoudi (2002) argues that it can have different meanings to different people: what some people consider as a polycentric city or as an organised system of sub-centres, others conceptualise it as a disperse city or as an unorganised urban sprawl. Even for professionals that should deal with this concept in a more rigorous way, some uncertainty in its definition exists: for example, and quoting Davoudi (2002, p.2) *"(...) urban planners use the concept as a strategic spatial planning tool; economic and human geographers use it to explain a specific form of urban structure and growth; EU Commissioners and their counterparts in member states promote the concept as a socio-economic policy goal; and civic leaders use the term for place-marketing, presenting it as synonymous with dynamism, pluralism, multi-culturalism as well as a symbol of a modern lifestyle."* According to several authors (Anas, Arnott *et al.*, 1998; Davoudi, 2002; Fujita and Thisse, 2002) such distinctions depend on which criteria, thresholds and scale of observation are being used as the basis of the analysis. A spatial configuration at a certain scale is not necessarily the same as at another, leading to the so called "*ecological fallacy*" (Fujita and Thisse, 2002).

Thus, geographical scale matters and reveals itself as a crucial aspect in the analysis of urban spatial patterns. For example, as Fujita and Thisse, (2002, p.2) mentioned, whether "(...) *Los Angeles or Chicago may be considered as a megacentre or as a collection of several large sub-centres depends very much on the scale of observation.*" One of the reason for this uncertainty, and according to Anas, Arnott *et al.* (1998), is the different effects of agglomeration economies, that emerge at a specific scales.

The analysis of the exact nature of the urban structure is not obvious. Because economic agglomerations appear at different geographical scales and involve several levels of spatial disaggregation, Papageorgiou (1983), *cited in* Fujita and Thisse (2002), supports that it would be useless to look for the model explaining all the different types of economic agglomerations. Since different externalities operate at different scales it is quite possible for a spatial pattern of economic activity to be centralized at one scale (e.g. large city) and dispersed at another (e.g. sub-centres that are too small).

These principles were critical for the emergence of New Economic Geography, analysed in more detail in the next section.

III.2.4. New Economic Geography viewed in a spatial perspective

Three authors are considered the pioneers of New Economic Geography (henceforth NEG), Fujita (1988), Krugman (1991) and Venables (1996). This branch of economic and geography literature has grown exponentially since Paul Krugman's 2008 Nobel Prize work[22], which culminated with the publication of Fujita, Krugman, and Venables's (1999) book: "*The Spatial Economy: Cities, Regions and International Trade*". This book integrates the prime insights from the NEG literature in a consistent general equilibrium framework[23].

The main contribution of NEG lies not in presenting specific forms of models but in developing a unified framework, combining old ingredients through a new recipe (Ottaviano and Thisse, 2004). By now, it is generally accepted that many ideas used in NEG had been already developed in seminal works[24] of economists, geographers, planners, regional scientists and location theorists that lay for a long time in the periphery of mainstream economic theory.

Like prior frameworks in regional science, NEG deals with the basic questions of analysing and understanding which factors influence the geographical distribution of economic activities, trying to give some insights to where and how economic activities are sited and related. Particularly, NEG explains why many economic agglomerations occur in geographical space. To analyse these important questions related to space, three analytical ingredients

[22] The Royal Swedish Academy of Sciences, announced in its scientific background report, that: "(...) *traditionally, trade theory and economic geography evolved as separate subfields of economics. More recently, however, they have converged [to] become more and more united through new theoretical insights, which emphasize that the same basic forces simultaneously determine specialization across countries for a given international distribution of factors of production (trade theory) and the long-run location of those factors across countries (economic geography)."(Committee, 2008, p.1)*

[23] This theoretical framework has the acknowledged merit of generating a wave of empirical research. Nevertheless, despite the extensive theoretical bases of this domain, empirical research remains comparatively less developed. For a extensive review concerning the existing empirical literature on New Economic Geography, see Redding (2010).

[24] For instance, Ottaviano and Thisse (2004) point out in their chapter of this Handbook that many of the ingredients of New Economic Geography were developed many decades before Krugman's (1991) paper.

are considered crucial in the NEG approach (Fujita, Krugman *et al.*, 1999; Fujita and Krugman, 2003; Ottaviano and Thisse, 2005; Venables, 2005; Fujita, 2010).

i) *Increasing returns to scale or indivisibilities of a fixed factor*: responsible for making location choice advantageous, and essential for the economy not to degenerate into "*backyard capitalism*", in which small groups of firms or households produce most items themselves. Spatially concentrated increasing returns to scale encourage activities to concentrate in space.

ii) *Transport costs*: the recognition that spatial interactions are costly and are directly influenced by distances, communication infrastructures, geography and by nature of interactions, leading to a situation of less effective interaction between workers if the proximity is not sufficient. If one wants to have an integrated picture of the economy, transportation costs strictly related with the used resources and generated incomes should be considered. If transport costs are not included in the analysis, space is becoming immaterial.

iii) *The movement of productive factors and consumers*: facilitates the location choice (this assumption distinguishes NEG from trade theory) and is a prerequisite for agglomeration. Assumptions about mobility, both between and within counties, regions or urban areas, covering the geographical mobility of goods, services, ideas, technologies and primary factors (land as being immobile, capital taken as often to be highly mobile)[25] are crucial to determine the spatial structure in an economy. Of course transportation costs are not independent of these aspects.

Surprisingly, these key elements of the NEG, crucial in the explanation of spatial distribution, and for this reason essential to urban and regional science, have always posed difficulties for economic theorists and have been paid little attention by the mainstream economists. The reason for such exclusion, as noted by Ottaviano and Thisse (2005), was due to the difficulty, until very recently (early 1990s) for the competitive paradigm to explain the formation of economic agglomeration. Whereas traditional neoclassical explanations for the uneven distribution of economic activity (population, employment and wealth) across space (countries, regions, cities and neighbours) emphasize first nature geography (e.g., natural landscape, for example, specific types of climates, differences in the fertility of land, raw materials or accessibility to natural ways of

[25] The imperfect mobility of labour is a fundamental issue that justifies a special treatment.

communications) instead, this new body of literature (NEG) gives special importance to the role of second nature geography, related to human actions (Krugman, 1993). To be more precise, the first nature features are those that are intrinsic to the physical site itself, and refer to exogenously given characteristics of different sites, independent of any development that may previously have occurred there. These are spatial endowments that cannot be easily changed. On the other hand, the second nature features of a location are those that are dependent on the spatial interactions between economic agents. In a practical perspective, homogeneous space in some classical models (as shown in the previous section) is assumed to control the impact of first nature, being possible then to find economic mechanisms "(...) *which emerges as the outcome of human beings' actions to improve upon the first one*" (Ottaviano and Thisse, 2004, p.2565). Thus, roughly speaking, first nature geography can be associated with the notion of bi-dimensional Euclidean space and second nature geography associated with multi-dimensional non-Euclidean space, where things affect and deform space (roads, bridges etc. make that space can only be described in n dimensions) (see Chapter V).

Aspects like constant returns to scale in production and the lack of consideration of transport costs exclude any possibility of geographic agglomeration of economic activity, leading to a uniform distribution of economic activities. The real world where people live is characterised by a cumulative and self-reinforcing economy of agglomeration processes, in which forward and backward linkages create a circular logic, where, other things being equal, producers would want to locate near to both their suppliers and customers. This circular and cumulative causation mechanism, first emphasised by Myrdal (1957), is responsible for agglomeration of economic activities, caused by firms in the process of finding larger markets, and by consumers in the process of finding cheaper and more diverse supplies.

Hence, assuming the existence of increasing returns to scale (which precludes perfect competition) opens the door to a much more complex reality, with multi equilibria and multiple levels of spatial aggregation, where many regions or agents can proliferate[26] and economic agglomerations (or concentration) can occur at many geographical levels[27].

In short, NEG literature, based on rigorous micro-economic foundations of geographical economics and on modern tools of economic theory, may be viewed as an attempt to combine a large number of theoretical models, designed to describe various aspects of geographic forces[28].

The existence of two different and opposite types of forces, centripetal and centrifugal, which are acting in a situation of increasing returns, is considered by Krugman (1998) the obvious reason for the evolution of the shape of the economy's spatial structure. The geographical distribution of the economic activities and gentrification over space result from a balance of centripetal forces, working toward spatial concentration of economic activity, and centrifugal forces that oppose such concentration. A list of such forces is summarised in the following *Table 1*:

[26] A similar framework has been applied in the context of industrial organization by of Dixit and Stiglitz in 1977.

[27] The example described by Fujita and Krugman (2003) highlights this situation: on a small scale one type of agglomeration may arises when small shops and restaurants are clustered in a neighbourhood; on a bigger scale, many other types of agglomeration can occur in the formation of cities. The emergence of this variety of agglomerations is responsible, in their point of view, for the existence of strong spatial regional disparities.

[28] A general equilibrium approach has been used based on Dixit and Stiglitz (1977) model of product differentiation and a proposed formalization of Myrdal's (1957) circular and cumulative causation (Fujita, Krugman *et al.*, 1999; Ottaviano and Thisse, 2004).

Table 1 - Forces affecting geographical concentration and dispersion

Centripetal forces[29]	Centrifugal forces
▪ *Linkages derived from market-size:* The idea expressed here is that a large local market creates both backward and forward linkages, reflecting preferences for sites with good access to large markets: for the production of goods, subject to economies of scale, and to support the local production of intermediate goods.	▪ *Immobile factors:* Immobility of some resources (land and natural resources) are centrifugal forces because they contribute to a dispersed market. The choice is between being close to workers (supply side) or locating close to the consumers (demand side).
▪ *Thick local labour market:* Supported by a concentration of the labour force, where specialized skills are located: employees find it easier to find employers and vice versa.	▪ *Land rent/ commuting:* High house prices may be caused by increased demand, providing a disincentive for further concentration.
▪ *Pure external economies and knowledge spillovers:* External economies via information spillovers may be induced in a more concentrated environment.	▪ *Congestions and other pure external diseconomies:* The excessive concentration may generate more or less pure external diseconomies (ex.: pollution, congestion).

Source: based on Krugman (1998), Fujita, Krugman et al. (1999) and Fujita and Krugman (2003)

[29] The three classical Marshallian sources of external economies, also called *Marshallian trinity*. The concept of external economies, according to Fujita, Krugman *et al.* (1999), was introduced by Alfred Marshall in 1890. Three main reasons have been identified by Marshall to explain why a geographical proximity or concentration of firms is advantageous: i) it could support specialised local providers of inputs; ii) it would offer labour market pooling, that is, workers would be less likely to remain unemployed if their current employer did badly, and firms would be more likely to find available labour if they did well; and iii) it would facilitate the spread of information (the formation of a highly specialized labour force and the production of new ideas, both based on the accumulation of human capital and face-to-face communications) (Fujita, 2010, p.20-22).

The centripetal and centrifugal forces, described in the *Table 1*, responsible for the tension between concentration and dispersion structures, should be viewed as non-exhaustive, because they do not represent, in the words of the authors (Fujita, Krugman *et al.*, 1999), all the items contributing to agglomeration in the real world. They note that "(…) *it would be useful to carry out a more systematic exploration of the implications of our menu, to inquire into the behaviour of models in which multiple centripetal and centrifugal forces are operating, to ask how the predictions of those models depend on the relative importance of those forces* (1999, p.346)." The authors proceed arguing that "(…) *only by carrying out such an exploration will we be in a position to interpret the results of the obvious next step: empirical research.*" Thus, quantified models can play an important role in NEG, which, according to Fujita and Krugman (2003), should be theoretically consistent, that is, based on some mix of data and assumptions (though this does not mean a model fitted to actual data).

As has been mentioned throughout this section, economic interactions can be addressed at different spatial levels, for this reason NEG provides a kitbag of tools for analysis at these different levels[30]: i) international models, where the economy is divided into a set of discrete nations; ii) regional models or core and periphery models, where the economy is divided into a set of discrete regions; and iii) urban-system models, where the economy consists of a set of cities including their surroundings (Fujita, 2010). These three classes of models represent minor variations of the same basic modelling presented below (Fujita and Krugman, 2003).

Methodologically, the NEG model starts from a standpoint that space is homogenous, meaning that production activities are equally distributed over territory and consumers' preferences are the same at all locations. Such an assumption is made to control the effect that first nature may have on the distribution of the economic activities. To find and explain the economic mechanisms, which are responsible for agglomerations (second nature geography), the model focuses on the forces (centripetal and centrifugal) that can cause some asymmetric shock across sites and consequently generate an unevenness in the distribution of economic activities.

[30] See Fujita, Krugman and Venables (1999) for development of many of these models.

The NEG model assumes an economy space with two initially similar regions (vertical axis)[31] as a function of the transport cost. It is considered that region contains two types of labours: producers of the agricultural good (farmers) as the traditional sector, and producers of manufactured goods (the workers) as the modern sector. Workers are freely mobile between regions although farmers are immobile but distributed equally between two regions. Transportation costs are extremely expensive for agriculture commodities, whereas the international trade of manufactures involves transport cost (in an iceberg form – Paul Samuelson (1952) approach). The immobility of farmers, caused indirectly by the substantial cost of transporting agricultural commodities, is a centrifugal force, while workers are a centripetal force involving a circular causation, à la Myrdal (1957).

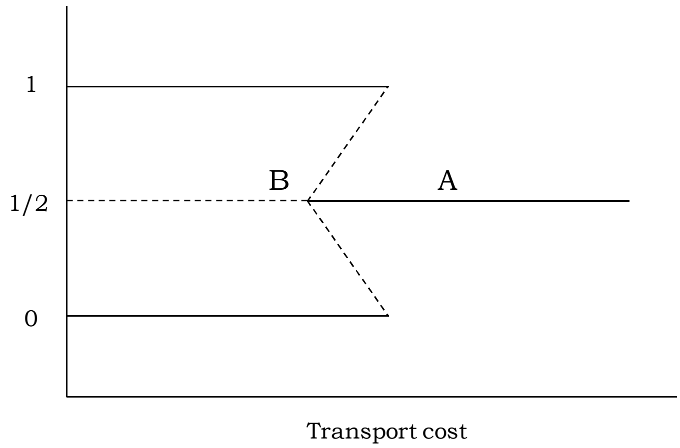


Figure 7 - Location of manufacturing in two regions: core periphery bifurcation[32]

Source: Fujita, Krugman et al. (1999)

Two different situations can occur (illustrated in *Figure 7*): i) manufacturing is equally distributed between regions and relatively little inter-regional mobility is expected in the case of high transport costs (the even dispersion of manufacturing is indeed consistent with a unique equilibrium with workers evenly divided between the regions; a situation represented by the in the point labelled A); or ii) all manufacturing workers are entirely concentrated in one region when trade costs are low (then there are multi equilibria); a situation represented by the in the point labelled B. This agglomeration of workers

[31] The Y-axis represents the proportion of mobile economic activity located in one of the two regions.

[32] Stable equilibria are represented by solid lines and unstable by broken lines.

generates a larger market where the location is advantageous for firms. The entry of firms bids up wages making an attractive place for new workers, not only because of the higher wages, but also for the improved access to consumer goods that this purchasing power can provide. In the first situation (high transportation costs) a decreasing number of other workers in the same region are expected because the wages that workers can earn depend mainly on the amount of local competition (see for more details Fujita, Krugman et al., 1999).

This illustrative example emphasises the typical dynamic of new geography models, resulting in a multiple equilibria situation (when the trade cost are low) *"(...) caused by a spontaneous self-organization of the economy into some kind of spatial structure, often one with very uneven distribution of activities among locations"* (Krugman, 1998, p.12).

The application of the NEG in the context of urban economics has been initiated by Fujita and Krugman (1995) and focused on two main features: the general equilibrium modelling of an entire spatial economy; and on the spatial distribution of cities. However, instead of considering an abstract, and given city, and a continuous two dimensional space, as von Thünen assumed, these models consider two discrete areas, where manufacturing may possibly occur in either, under certain conditions. Due to the centripetal force, resulting from a circular causation, scale economies may emerge (Fujita and Krugman, 2003).

III.2.5. Summary

Factors such as globalization, means of transport and communication contribute profoundly to transform urban shape over time. Global competition has contributed to change the economic relationships between firms and, as a result, how those firms are spatially organized. Quoting Walter Powell (1990), *cited in* Anas and Arnott *et al.* (1998), firms have developed new modes of interaction which are neither market nor hierarchy, but rather a network organizational form. These phenomena have directly impacted on the way that urban structures are organized. Cities have expanded beyond their territorial boundaries and central business district's (CBD) and are no more the unique centre of a city. The recent phase of urban transformation is thus characterised by the existence of several centres with different levels of importance among the space, caused essentially by economies of agglomeration and externalities[33], leading to a more polycentric process of urbanization.

The notion of economies of agglomeration and the concept of externalities, taking place at different spatial scales, makes the process of understanding the spatial structures more complex, both in terms of spatial heterogeneity and spatial dependence (spillovers). If it is considered that these urban spatial patterns can be permanently changed, the difficulty of delimiting and understanding urban spatial units and its interactions increases.

Despite these difficulties, urban economics, which has its main focus on the urban land use location theory, makes important contributions to an understanding of spatial structure in urban spaces. This sub-field of the economy emerged in opposition to the neoclassical economic theory which considers a set of simplifying assumptions to avoid the consideration of space in their models. This argument was underlined by Walter Isard (1949, p.477) when in 1949 he accused economics of taking place in a "*wonderland of no dimensions*".

Spatial dimension is considered an important element in urban economic issues; however, assuming that there are no economies of scale and that the territory is isotropic and has a flat surface, models can be considered unrealistic, since spatial concentration and specialisation of the economic activities are observed. This argument does not sustain the idea that urban location models should be a reliable picture of the real world and all achievable variables have to be measured. Abstraction and flexibility are important aspects in the urban location theory.

---

[33] There is in the literature a distinction between two different types of externalities (see Glaeser, Kallal *et al.*, 1992; Abdel-Rahman and Anas, 2004): Jacobs externalities, and Marshall-Arrow-Romer (MAR) externalities. Both are related to knowledge spillovers between firms. The main difference is that the MAR (or localization) externalities arise from knowledge transfer within an industry, between firms that belong to the same industry, whereby the Jacobs (or urbanization) externalities arise from transfers between industries, among firms of different industries. Jacobs (1969), unlike MAR, believes that the most important knowledge transfers come from outside the core industry. In empirical work by Glaeser, Kallal *et al.* (1992) it is argued that cities with a diversified industrial base (subject to Jacobs externalities) may have a faster growth rate than specialized urban areas. However, other empirical studies (Henderson, Kuncoro *et al.*, 1995) suggest that MAR externalities are more important in traditional industries, while Jacobs externalities play an important role in modern high-technology industries. The lack of consensus on this issue illustrates that the theoretical literature on city systems has not yet provided a clear explanation regarding the impact of specialization or diversification on the rate of growth of a city (Abdel-Rahman and Anas, 2004).

The example, described by Krugman (1999), highlighting the dichotomy between the rigour and the loss of information, is a excellent illustration of the advantages and disadvantages of both aspects. The author uses the particular case of the production of maps, particularly the evolution of the maps of Africa, but the example can be generalised to other geographical features. In the fifteenth century the African continent maps were quite inaccurate regarding distances and coast lines, nevertheless, they were drawn with a lot of information on the features of the land, in the interior of such limits. Based on explorer reports (sometimes second or third hand travellers reports) the location and description of rivers, cities, resources and even imaginary creatures (men with their mouths in their stomachs) were represented. Because of the evolution of the technical tools there is no possible comparison between maps produced at that time and the maps which started to be produced in the eighteenth century, where coastlines were precisely reproduced and the quality of information used to make cartography got gradually better. On the other hand the interior of the continent emptied out, that is, was left blank, losing in a certain way their three-dimensional perspective.

Therefore, conclusion from the example described above is that simplification and the capacity of generalization are important aspects when one wants to analyse relevant aspects of a complex world.

In the context of urban economic thinking, it is fundamental to identify and clarify significant patterns and interdependences in a particular context. Of course the price of such generalization is that, at some point, it is not possible to explain all processes in detail. So, what should a "*space theory*" be [using the terminology of Isard (1956)]? Two different perspectives can be cited. One viewpoint is given by Isard (1956), saying that space theory is "(…) *conceived as embracing the total spatial array of economic activities, with attention paid to the geographic distribution of inputs and outputs and the geographic variations in prices and costs*." The other standpoint is given by Fujita (2010, p.2) pointing out that space theory aims, "(…) *either in a descriptive or normative context, to explain the geographical distribution of all agents in a given location space.*"

It was beyond the scope of this section to deal extensively with all location paradigms in the field of urban economics. Nevertheless, important urban economic location theories have been presented in chronological terms. To organize such information several taxonomies could be used depending on the different problems that each model intends to solve. Of course the subject is always the same: each model tries to contribute with some theoretical and empirical approaches to an understanding of spatial patterns and the interdependences of agents, for both industries and households, over space, in order to provide an appropriate framework to answer the question: where and why are particular economic activities located in a given spatial system?

In summary, some of the difficulties involved in understanding urban spatial patterns in a globalizing world are resolved through general location models, helping to analyse the essential aspects of urban spatial structures (at various levels of spatial scale) in a unified manner. Building such model is crucial and, at same time, a challenging task.

III.3. The role of space in Urban Geography

As was seen in the Chapter III.2, the way that space has been analysed has always changed. In the case of urban geography the situation is not different. In order to describe and interpret the various social contexts, different theories and philosophies of space have emerged, providing a distinctive view of the nature of geography.

Traditionally, the word geography is derived from Greek, *geo* referring to Earth and *graphy* meaning picture or writing. Literally interpreted, geography means to write or describe the world. However the subject of geography is much more complicated than the enumeration of capitals and maps (Spellman, 2010). The meaning of geography is continuously changing, and a huge number of definitions can be found in the literature: from the perspective of Hartshorne (1959) which saw geography as a *idiographic science*, giving emphasis to description, to the perspective of Yeates (1968) which saw geography as *a nomothetic science*, giving emphasis on explanation, many perspectives can be considered (Hubbard, Kitchin *et al.*, 2002). Thus, Hartshorne (1959, p.21) defines geography "(...) *as the science which is concerned to provide accurate, orderly, and rational description and interpretation of the variable character of the Earth's surface*", while Yeates (2001, p.1) defines geography saying that "(...) *can be regarded as a science concerned with the rational development, and testing, of theories that explain and predict the spatial distribution and location of various characteristics on the surface of the earth*". For a complete review about how geography has been variously theorised over time, see for instance, Holt-Jensen (1999), Kitchin and Tate (2000), Hubbard, Kitchin (2002), and Pacione (2005), among others.

Geography as a more general discipline assumed different philosophical perspectives over time and urban geography, as a small piece of the entire puzzle, was not immune to this transformation. Since the late 1970s the scope of urban geography has expanded rapidly, and geographers have approached the study of the city from different forms (Pacione, 2005). The same author refers eight main epistemological developments in urban geography: i) *environmentalism* has major concerns in the relationship between people and environment (relevant during the first half of the twentieth century); ii) *positivism* is characterised by the application of scientific method to test hypothesis, and thus to construct theories[34] (which gained prominence in the late 1950s with the development of the spatial analysis school)[35]; iii) *behaviouralism* gives emphasis to the role of cognitive processes to mediate the relationship between the urban environment and people's spatial behaviour (which emerged to overcome the shortcomings of spatial analysis); iv) *humanism* is characterised by using methodologies which explore people's subjective experience of the world views, as it considers that each individual is a determinant agent of change in the city; v) *structuralism* refers to the idea that explanations for observed phenomena must be analysed in terms of social, economic and political structures, rather than through empirical study of the phenomena alone; vi) *managerialism* is a process where social groups seek to maximise their benefits by restricting access to resources and opportunities; vii) *post-modernism* is characterised by the rejection of positivism and structuralism, giving emphasis to human difference, uniqueness and individual sensitises (emerging in the late 1980s and 1990s); and finally, viii) *moral philosophy* seeks to examine critically the moral bases of society, focusing on what should be rather than what is (representing an emergent perspective in urban geography).

Spatial aspects, and specifically space, considered the basic requirement for the existence of geography as a science, is not independent of the different theoretical perspective, described above. Its relevance is justified by diverse interpretations proposed in the theories and methodologies of urban geography, which appear more or less evident as the object of analysis.

According to Aase (1994) it is possible to argue, ontologically, in favour of all perspectives of understanding space; problems emerge when the various concepts of space are applied to the analyses of real world phenomena. For this reason the author (1994) suggests that our energy should be spent on

---

[34] Concepts such as distance decay were also introduced in the study of urban phenomena: meaning the attenuation of a pattern or process over distance (Pacione, 2005)

[35] Despite the fact that positivism was evident previously in the work of Christaller (1933 [1966]) and Losch (1940 [1954]), as quoted in Pacione (2005).

understanding those contexts rather than searching for a real (once-and-for-all) definition of space. The question "*what is space*?" cannot be seen as having a philosophical or linguistic solution independent of anything else. In Harvey's words "(...) *space becomes whatever we make of it during the process of analysis*" (Harvey, 1973, p.13). Depending on the circumstances, space can become one or all together (Harvey, 2006).

This section will contribute to an understanding, expressed through a synthesis of meanings, of how geography theory views space, and is organized centred upon a multitude of conceptions of space in the seminal works of the following authors (geographers and philosophers): Ernest Curry (1996); David Harvey (1973); and Henri Lefebvre(1974 [1991]). Many other authors explored how different traditions have conceptualized space, for example, Soja (1989; 1996), Castells (1977; 1983), Haggett (1965; 2001) and Massey (1991; 1995; 2005), among others (see Hubbard, Kitchin *et al.*, 2002; Hubbard, Kitchin *et al.*, 2004; Pacione, 2005, for a extensive review of the most relevant thinkers on space). While not comprehensive, the main philosophical traditions in geographic thought and writings on space are overviewed.

The three aspects of space studied in this work, heterogeneity, spillovers and scale are highlighted and linked to the main ideas resulting from the urban geography theoretical framework.

III.3.1. Aristotle, Newton, Leibniz and Kant's space

According to Curry (1996, p.5 and 24), there are four different ways of thinking about space, which have gained popularity among scientists, philosophers and geography thinkers, in western thought. The first is static, hierarchical, and concrete space, codified by Aristotle. The author (Curry, 1996) considerers Aristotelian's space the space of every life, the space of discourses and of activities, seeing the world as a place where things belong here and not there, with real and clear hierarchies. The second, associated with Newton, is a kind of absolute grid, within which objects are located and events occur. It is the space of people's reflections, where they exist somewhere within a vast and unidirectionless space and perfectly indifferent to anything that people do. The third adopts the scientific outlook of Newton but assumes that the definition and understanding of space should consider the relationships among objects and events, and is found in Leibniz's work. In this notion of space it is attempted to define conceptual systems that will comprehend the world, viewed in terms of a set axioms and assumptions. And finally, the fourth, codified by Kant is related to the need for seeing space as a form imposed on the world by humans. It is by nature unitary, three dimensional and infinitive.

III.3.2. Harvey's tripartite division: absolute, relative and relational space

Another distinguished way to conceptualise geographical space is a tripartite division of absolute; relative and relational (cognitive). Harvey (2006) associates absolute space with the theories of Kant and Newton, and associates relative space with the name of Einstein, while relational space is often attributed to Leibniz. An important feature of the distinction between these three perspectives is the way in which space is conceptualised (Smith, 1990).

The absolute notion of space or Newtonian space is a concept used by most geographical analyses up until the 1970s (Shields, 1997). This view was developed by Newton in his "*Fundamental Principles of Natural Philosophy*" (1686 *cited in* Curry, 1996), who defined absolute space "(…) *in its own nature, without relation to anything external, remains always similar and immovable.*" Thus, Newtonian space is seen "(...) *as emptiness, as a universal receptacle in which objects and events occur* (…) *as a frame of reference, a co-ordinate system* (…) *within which all reality exists*" (Smith, 1990, p.95).

Absolute space is represented as pre-existing and fixed grid manageable with established measurements, and efficiently understood by a frame (Harvey, 2006a). The space is reduced to the essence of geometry and is represented with relative precision through the Euclidean geometry, by a system of topographical coordinates (clearly situated in a grid or a map). Thus, concepts of distances are fundamental in the notion of absolute space, usually expressed as a straight line to describe any phenomena, e.g.: measured in kilometers or latitude/longitude

(Harvey, 2000; 2006). This conception of space largely supports the work of positivistic and quantitative geographers (Hubbard, Kitchin et al., 2002). This group of geographers believes that there are pre-existing physical laws that can be scientifically measured with an x, y and z dimensional system, in which contents of space are unquestionably understood as being natural and given. Throughout the greater part of their existence geographers have represented spatial phenomena using models with Euclidean measure of distances, maps constructed on a Euclidean coordinate system and theories of spatial organization summarised cartographically and mathematically in Euclidean terms. The recognition of this Euclidean formula as the most appropriate away to represent horizontal connections presupposes the existence of an isotropic notion of space, in which the same geometric relations are holding in all parts of space (Golledge and Hubert, 1982). If space is viewed as absolute "(...) *it becomes a thing in itself with an existence independent of matter*" (Harvey, 1973p. 13).

During the 1960s spatial analysis provided a unifying theme for geography (Unwing, 1992 *cited in* Clifford, Holloway *et al.*, 2009) and reinforced the idea that geographers should "(...) *pay attention to spatial arrangement of the phenomena in an area and not so much to the phenomena themselves* (...) *spatial relations are the ones that matter in geography, and no others*" (Schaeffer, 1953, p.228). Schaeffer claims that objects are not more than objects and the search for laws takes place in other disciplines, not in geography. This perspective is also evident in Hartshorne's works (1959), where geography was about interpreting associations of objects relative to each other over space, commonly represented in maps. Thus, vital spatial dimensions have been neglected over the past few decades (see Clifford, Holloway *et al.*, 2009).

The absolute view of space was for so long accepted because it had been covertly built into Newtonian physics. As has been said before Isaac Newton conceived space to be an infinite container (with no boundaries) in which objects could be situated at any point. He argues that space just exists, independently of any other objects or facts and without relation to anything external (Scruton, 1996; Agnew, 2005). The legacy of Euclidean geometry, reinforced by a Newtonian view of spatiality, dominated geographical thought and practices in the first half of the 20th century, and has remained influential to date (Graham and Healey, 1999). However, the idea of reducing the world to an inflexible geometry is of very limited utility, and for this reason, absolute conception of space has been a sustained criticism. Three aspects were crucial to considering the pure Euclidean geometry perspective doubtful, in the nineteenth and twentieth century, as suggested by Davoudi and Strange (2009): i) the discovery of non-Euclidean, n-dimensional geometries; ii) the recognition of the impossibility in describing the geometry of the world without describing the forces that are at work in it; and finally iii) the acknowledgment of the importance of the dimension time to describe physical space.

The following *Figure 8* and *Figure 9* represent different perspectives of absolute space. At more mundane level it is the space of the location of the several agents acting in the territory (houses for people and firms for activities) and the territorial boundaries precisely defined (such as several administrative levels of organizations, urban plans etc.) as containers of social objects (Harvey, 2006).

Concerning scale, the absolute notion of space is by its nature a hierarchy of scales.

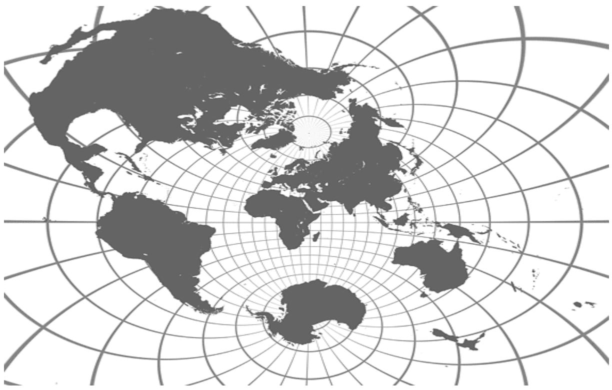


Figure 8 - The world represented in terms of absolute space

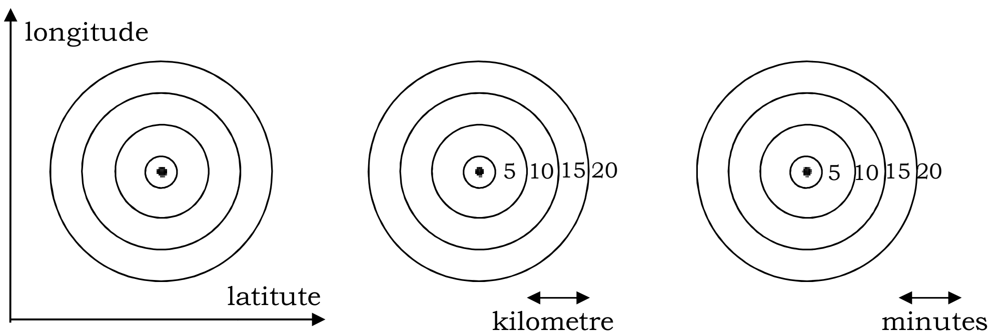


Figure 9 - Different representations of absolute space

Geographers quickly developed new ideas about distances, and as a result the notion of space itself. Non-Euclidean distances started to appear in the lexicon of analysis of the geographers and the notion of relative space emerged as a new perspective of understanding spatial phenomena.

In relative space, location and distances between different events are considered and measured in terms of some measures that express the phenomena under study. For example, time (accessibility, travel-time, etc.), costs (economic, social and environmental), model split (car, bicycle or bus), energy, topological relations and perceptual distances within a complex network (Couclelis, 1992; Holt-Jensen, 1999; Harvey, 2006). These non-exhaustive particular dimensions of measuring distances highlight the notion that the shortest distance between two points is not necessarily a physical distance. Putting it in a different way, movements of people, goods, services and information takes place in a relative space.

The change from absolute to relative space resulted in philosophical implications and was in this sense revolutionary. In the 1960s David Harvey (1969) and Peter Haggett (1965) contributed in a decisive way to a fundamental debate within this subject (Holt-Jensen, 1999). In Harvey's words (1973) relative space is viewed as a relationship between objects which exists only because objects exist relative to each other, *"(...) things do not exist outside of or prior to the processes, flows and relations that create, sustain, or undermine them"* (Harvey, 1996, p.49).

The relative notion of space is mainly associated with Einstein's space-time concept, arguing that all forms of measurement depended upon the frame of reference of the observer (Harvey, 2006). In the 19th century a shift occurred in the language of space, the idea remaining that it is impossible to analyse space without time. Space started to be assumed as being interdependent of objects and events, created and defined from the relations between them. The non-Euclidean geometries began to be a key issue in the relational view of space (explained below). However, this philosophical view is far older, when Leibniz suggested that "(...) *spatial properties are relational and position of any object is to be given in terms of its relation in the objects*" (Scruton, 1996, p.362). Thus, Gottfried Wilhelm von Leibniz (1646–1716) vigorously combated Newton's absolute view of space, in favour of a relative perspective, where space in his opinion must consist of the totality of spatial relations between objects (Scruton, 1996). The following example, described in Scruton (1981, p.75), illustrates clearly the Leibniz position: “(...) *if one asks for a definition of a point in space, Leibniz says, he can provide it by showing what it is for two objects to occupy the same point. Two objects occupy the same point in space if they stand in the same spatial relation to all other things.*” For Leibniz space is not a substance, instead it is a combination of observed relations of coexisting phenomena, derived from a series of observations of groups taken together.

In the mundane world, and keeping the illustration described by Harvey (2006), many of the inhabitants move on a daily basis across relative space into their sites of employment where they earn money that permits them to import back into the absolute space. The following *Figure 10* and *Figure 11* represent the principles of relative space.

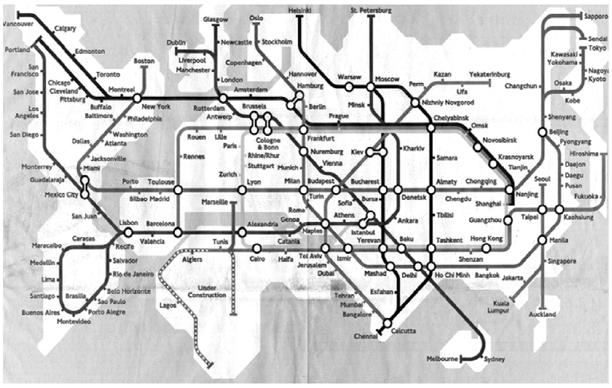


Figure 10 - The world represented in terms of relative space

Source: Abler, Adans et al., (1977, p.74)

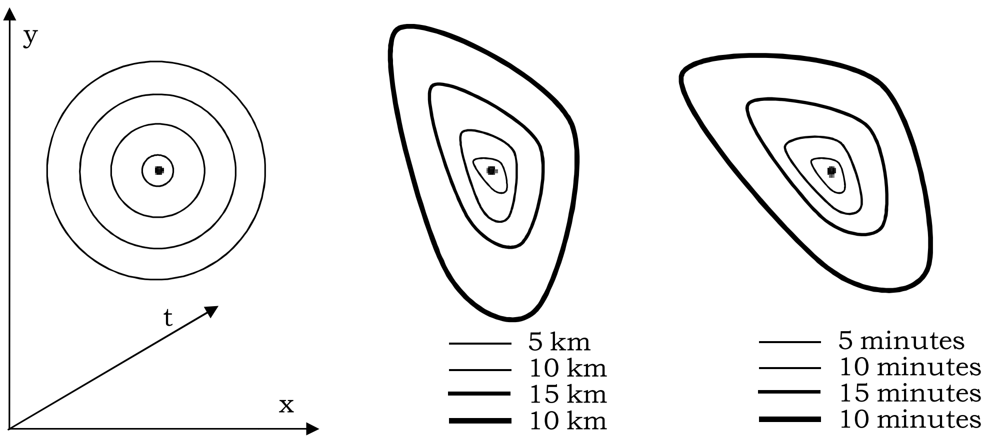


Figure 11 - Different representations relative space

So far the notion of absolute and relative space has been presented. In the former space becomes a thing in itself and exists independent of matter, while in the latter, space is understood as a relationship between objects and exists only because objects exist and relate to each other. However, Harvey (1973) considers a third perspective in which space can be understood as relative. The author called this relational space (cognitive), which is a "(...) *space regarded in the manner of Leibniz, as being contained in objects in the sense that an object can be said to exist only insofar as it contains and represents within itself relationships to other objects*" (Harvey, 1973, p.13). Relational vision of space is often associated with Leibniz[36]. He acknowledges that there is no such thing as space outside of the processes that define it (Harvey, 2006). The point here is that processes and objects do not occur or exist in space but define their own spatial frames through their relationships with one another. Quoting Harvey (2006, p.273 and 274), "(...) *space is embedded in or internal to processes",* that is, an event or a thing at given point "(...) *cannot be understood by appeal to what exists only at that point*", depending "(...) *upon everything else going around it.*"

A more relational conception of space coincides with the increasing consciousness of the role of the social in human geography in the early seventies. Because urban geography encompasses the complexity of social interaction, rather than an *a priori* and rigid spatial structure, the conception of relational space emerged to be more appropriate to characterise space itself. Thus relational space emphasises the analyses of how space is constituted and acquires different meanings through human interactions. In this sense, space is not a given neutral and passive geometry but rather is continuously constructed and produced through socio-spatial relations and experiences. It is a product of cultural, social, political and economic relations (Massey, 1995).

In Murdoch (2006, p.86) relational space is seen "(...) *as an undulating landscape in which the linkage established in networks draws some locations together while at the same time pushing others further apart.*" The relational (cognitive) space is defined and measured, not in terms of commonly used Euclidean geometries, instead, in terms of the values, feelings and perceptions about locations (neighbourhood, regions, countries etc.) and is embedded in people's intentions and actions (Holt-Jensen, 1999). A relational view of space

[36] The notion of relative and relational space emerges in opposition to the absolute space, for this reason Leibniz is associated with both.

considers individuals, not as autonomous subjects with individual preferences, but as formed within social contexts. Social networks are a common concept to describe relational space. For a further discussion see Healy's (1998) institutionalist approach which emphasises the importance of when and how being in a place matters to people. Thus, space is made not by (underlying) structures but by the diverse and complex system of intersection, which themselves are made by interdependences established between entities of several types (Massey, 1993).

The way to measure events within and between spaces turns out to be more challenging the closer its moves towards an urban context of relational space.

As in the case of relative space, it is impossible to dissociate the dimension time from the relational notion of space, in the sense that, the internal relations and the external influences can be incorporated into the processes through time (Harvey, 2006). As previously, the following *Figure 12* and *Figure 13* show how relational space can be represented.

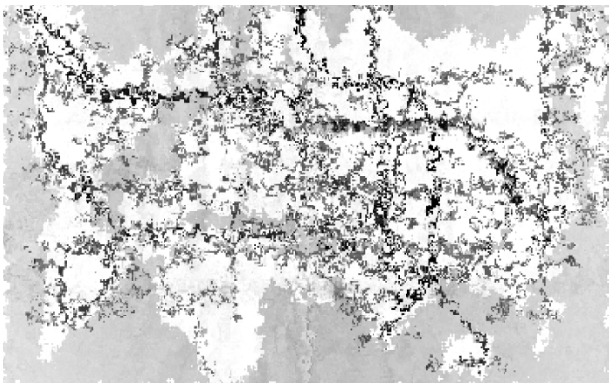


Figure 12 - The world represented in terms of relational space

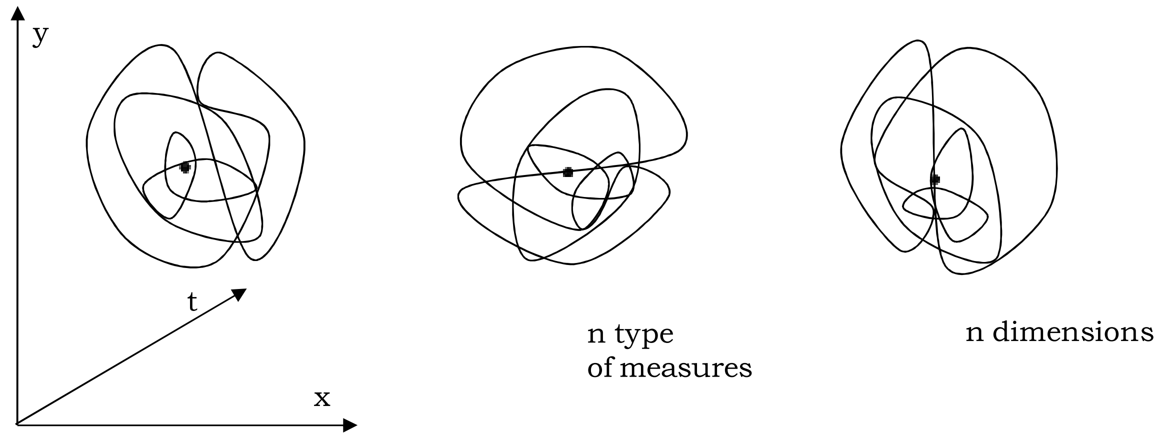


Figure 13 - Different representations of relational space

Table 2 - Key concepts associated with absolute, relative and relational space

Absolute space	**Relative space**	**Relational space**
Bidimensional Euclidean geometry	Bidimensional non-Euclidean geometry	n dimensional non-Euclidean geometry
Independent of any matter	Referent to things and processes	Dependent on perceptions and feelings
As a container	Defined by things and processes	Places, territories, environments
Inventories and maps	Studies related to functions	Studies related to social and economic behaviour

In short, absolute space becomes an entity in itself with an existence independent of whatever objects and events or phenomena are considered individually (Harvey, 2006). That is, space is independent from matter. Geometrically, space is viewed as a container represented by Cartesian's notions (fixity and static) and lies outside of society. Both relative and relational space is socially produce by people, and thus is dynamic, fluid, pleated, twisted as a chain and unstable (Murdoch, 2006). Multiple geometries (Euclidean and non-Euclidean) can be used depending on the observer (Harvey, 2006). When space is viewed as relative, space exists only because objects and its relationships exist.

This argument is linked with the idea that, as mentioned by Harvey (1973)[37], the distance between two points is considered relative because it is strictly related with the perception of distances by the people. Finally, the relational space, another perspective of relative space, is not a container but something that exists dependent on the processes and substances that make it up. The relevant aspects that characterise the tripartite of absolute, relative and relational space are reported in *Table 2*.

In addition, a hierarchy can be clearly defined between the tripartite divisions of space described above. Harvey (2006) pointed out that relational space can include the relative and the absolute; relative space can include the absolute; and absolute space is just absolute.

III.3.3. Lefebvre's spatial triad: Spatial practice, representations of space and spaces of representation

The notion of considering space as socially produced is also clearly advocated in a central theme of the work of Henri Lefebvre (1974 [1991]). Henri Lefebvre, a French sociologist, intellectual and philosopher, provided a new epistemological framework through which space could be understood and interpreted within social theory (Soja, 1989). In his book *"*The Production of Space*"* (*Production de l'espace,* 1974), Lefebvre had the aim of bringing the philosophy and epistemology of mental space into discourse with real or empirical space (Parker, 2004). He considers that people are *"*(…) *confronted by an indefinite multitude of space, each one piled upon, or perhaps contained within, the next: geographical, economic, demographic, sociological, ecological, political, commercial, national, continental, global"* (Lefebvre, 1974 [1991], p.8). Descartes, Kant and Hegel were his major interlocutors, for whom the spatial dimension of human experience has been often neglected or mishandled (Parker, 2004).

[37] The work of David Harvey (1973) puts emphasis on the way in which particular forms of society create particular spaces.

Lefebvre introduced the difference between concrete (material and physical) and abstract (mental and geometric) space (Lefebvre, 1974 [1991]). The first is defined as being the space of gestures and journeys, of the body and of the memory, of the symbols and of the sense, the space of experiences. In short, it is the lived space, very distinct from standard geographical ways of viewing it, established by our perceptions of space. The second is the measurable space, corresponding to our conception of space. This is the abstract space of the drawing of the architects and urbanists, working like an instrumental space where passive geographic or empty geometric milieu comes out. Abstract space "(…) *reveals its oppressive and repressive capacities in relation to time (…) it rejects time as an abstraction*" (Lefebvre, 1979, p.287) and is divorced from the level of lived space and tends towards homogeneity (Lefebvre, 1974 [1991]; Elden, 2004). The author claims that on the ground this space is not innocent, acting as a bulldozer to enforce the goals of capitalism.

However, the author considers that for making progress in understanding space, a grasp of the concrete and abstract space is needed. In his book "*The production of space*" (Lefebvre, 1974 [1991]), the author detailed how the production of space (process of spatialization) is premised on three complementary levels, distinguishing his own tripartite division, which draws attention to: i) Spatial *space*; ii) *representations of space*; and iii) *space of representations* (Lefebvre, 1974 [1991]; Soja, 1996; Elden, 2004).

This first concept is strictly related to the material production of the necessities of everyday routines; it is the real space generated and used, through which the totality of social life is reproduced (houses, cites, roads etc.). It is related with abstract representations of space through the logic of maps and mathematics; and for this reason it is responsible for the (re) production of the city, making and unmaking it as a functioning urban system. Spatial practice influences the *where* of human endeavours and "(...) *is lived directly before it is conceptualised*" (Lefebvre, 1974 [1991], p.34). This perceived space (*perçu*) embodies the close association between daily reality and urban reality, where routes and networks link up the place set aside for work, private life and leisure (Lefebvre, 1974 [1991]; Elden, 2004).

The second notion in the dialectic of spatial terms is the representation of space or conceived space (*conçu*). Is the conceptualised space of scientists, planners, urbanists, politicians, engineers and certain artists which have a scientific approach, "(...) *all of, whom identify what is lived and what is perceived with what is conceived*" (Lefebvre, 1974 [1991], p.38). Here space serves to conceptualise and represent a world of tactile and sensual interaction with matter, offering an objective view of the way the city works, e.g., images, books (words), films or urban representations like plans, diagrams or maps. Thus, the representation of this space, the reality that surrounds us, includes abstract representations, such as, "(...) *Euclidean Geometry, social scientific discourses on urban and social behaviour, architectural and planning theories of the built environmental, the Quattrocentro theory of visual perspective, Einsteinian relativity theory, and philosophical theories of space of Leibniz, Kant and Hermann Weyl*" (Dimendberg, 1998, p.21). All of these abstract representations seek to symbolize under different regimes and ideologies the way that people live through their emotions and imaginations.

The space of representation (the third Lefebvre division of space) corresponds to social space or lived space (*vecu*), the space of human action and conflict and sensory phenomena. This space is less formal, produced and modified over time and invested with individual meanings and symbolism. In contrast to represented space, this is the space that is lived and felt by people. Thus, it is imbued with ideological and political content, with imaginations, fears, emotions, psychologies, fantasies and dreams. It is in such spaces that the dehumanizing tendencies produced by capitalist processes could be overcome, that is, proclaiming the power of people to produce their own space and create new forms of urban life (Lefebvre, 1974 [1991]).

Table 3 - Key concepts associated with Lefebvre's spatial triad

spatial practice	**representations of space**	**spaces of representation**
empirical	verbal and signs	non-verbal symbols or signs
perçu (perceived)	conçu (conceived)	véçu (lived)
maps	mental	images
physical and absolute	abstract conceptions and cognitive	social
materialism	idealism	materialism/idealism
real and everyday life	*savoir* (knowledge)	*connaissance* (symbolic)

Contrary to what happens with Harvey's conceptions of space, Lefebvre stated that these three forms of space are not isolated expressions of spatiality, and combined are responsible to produce a complex spatiality embroiled in a constant state of dialectical interaction, which can vary over time (Lefebvre, 1974 [1991]). For this reason it is not reasonable to consider this tripartite division of spatial terms as hierarchically organised, it is most appropriate to keep these three categories in dialectic tension. See *Table 3* where important concepts associated with Lefebvre's spatial triad are reported.

It is through the dialectical interaction of these different manifestations of space that Lefebvre claimed space as socially produced (Lefebvre, 1974 [1991]). As a consequence, the non-absolute understanding of space is not explicit in Lefebvre's epistemological framework. For him, space has come to mean different things to many people and is highly complex. Thus, it should be understood in a variety of ways (Crang and Thrift, 2000a; b).

III.3.4. Cassirer's tripartite distinction for understanding space: organic, perceptual and symbolic spaces

Another fundamental tripartite distinction of the notion of space is given by Cassirer (1944). Ernest Cassirer focuses on social space distinguishing three types of spatial experience i) organic; ii) perceptual and iii) symbolic spaces, each one arranged in a certain order, from the lower to the higher strata (Harvey, 1973; 2006). However he says that "(...) *rather than investigate the origin and development of perceptual space, we must analyse symbolic space*" (Cassirer, 1944, p.49).

The lowest stratum is the organic space (the space of action) and comprehends all those forms of spatial experience given biologically. The author considerers that every organism, even the lower adaption of organisms, requires certain processes of spatial orientation, perceptions and reactions to adapt themselves to the conditions of the environment in order to survive. It is the result of spatial experience registered through the particular uniqueness of sense perceptions of each species.

Perceptual space refers to the way that perceptions are processed neurologically and registered in the world of thought. It contains elements of all the different kinds of sense experiences, such as, optical, tactile, acoustic, olfactory, and other sensory factors. It is the world created through sensary experiences of the organism. For this reason Cassirer (1944) considers that organic space is the space experience given to the higher order species (such as, men, dogs and monkeys) which have a higher complex array of sensory capabilities.

Finally, symbolic space is abstract and responsible for producing distinguishing meanings, and it is "(...) *by a very complex and difficult process of thought*" that peoples (or animals) arrive "(...) *at the idea of abstract space*" (Cassirer, 1944, p.49). Aesthetic characteristics are crucial, which have no foundation in any physical or physiological form, but symbols for abstract relations (Cassirer, 1944; Harvey, 2006). These spaces are made social, and become places. Since space is symbolic, it is also abstract, in the sense that if one looks at a place, it can be visualized as geometric objects and mathematical relationships. In this sense, geometry is used to represent spatial relations without relating to a specific place.

The human being compensates for the deficiencies in both organic and perceptual space through a long process of development in his understanding of symbolic space, using signs with different levels of abstraction to express things in the world.

III.3.5. Summary

Space being a central concept in geography, has been interpreted in different manners and, like geography "(...) *has meant different things to different people at different times and in different places*" as quoted by Livingstone (1992). Furthermore, "(...) *space is neither absolute, relative or relational in itself, but it can become one or all simultaneously depending on the circumstances"* (Harvey, 1973, p.13). As has been shown above, space is entirely dependent on the context, but despite the diversity of perspectives, Harvey (2006) pointed out that the complication or major challenges of dealing with space arise from its continual modification, rather that the complexity of the notion of space itself.

The concept of space in geography, turns out to raise a set of difficult issues (Curry, 1996) and has always been controversial (Holt-Jensen, 1999).

In line with Harvey (2006) the proper way to conceptualise space is through human activity. The following examples describes this argument: a house located and precisely defined in its boundaries over space, as a physical entity, is situated in absolute understanding of space; the movement of people, goods, services and information (associated with expenditure of money, time and energy), encompassing sites of employment, leisure and residence, characterises and creates the relative space; and finally, neighbourhoods which are absorbers of the complex externalities, given different meanings (rational or irrational) and importance to the use of space emerging from the relational space. The same illustration could be made using other distinctive divisions of space, such as in Lefebvre or in Cassirer.

One of the main issues in urban geography, besides the understanding of space, is scale. Because different contexts occur at different spatial scales (and time) it is very difficult to incorporate all the crucial aspects of spatial activities into the same framework. It can be assumed that processes and phenomena which encompass a broader spatial scale usually change more slowly than the small urban scales where processes act more intensively. Thus, specific phenomena show distinct scale thresholds. Urban patterns which appear to be structured at one scale, may appear to be disordered at other scales (Miller, 1978); it is also true that different spatial processes can generate exactly the same urban spatial patterns. Spatial dependence may vary with scale, depending on the degree of spatial heterogeneity. Thus, generally the selected scale should be the one which maximizes the variability across space (Meentemeyer, 1989).

III.4. The role of space in Urban Planning

In a continuation of the previous two sections, seeking to understand how space has been analysed in disciplines which involve spatial studies, the context of urban planning is now addressed. As mentioned previously, the three principal urban studies disciplines are not independent from each other, and this aspect is underlined by Murdoch (2006) when he points out that, as a form of applied geography, planning plays an important role in formulating ideas about space, having the opportunity to put these ideas into practice.

The theories under consideration appeared in the post Second World War era. The second half of 20th century was marked by an unprecedented urbanisation process, with major historical transformations in the context of urban planning theory[38]. The relevance of this period is highlighted by Ward (2004, p.2 and 6), reinforcing the idea that planning was "(...) *a significant thought somewhat marginal influence before 1939 (…) a series of radical reformist ideas about changing and improving the city*". Nevertheless, more general scientific theories and philosophies in which planning theory is supported, and presented here, are not restricted to these dates, covering a rather longer period. In this regard, Friedmann (1987) addressed that the practice of planning, in the modern sense, began in the early decades of the last century, but, its ideological roots go back to the early nineteenth century.

The arguments presented during this chapter are neither strictly chronological in relation to the historical development of these theories per se, nor are they descriptive of all urban planning theories from the last sixty-five years. Instead, the main aspects of the different approaches to the question of how space has been analysed and understood in the theoretical work of urban

[38] Despite that during the mid and late of the 19th century, planning activity had as well a relevant role in controlling many economic and social changes occurred from the industrial revolution which were accompanied by a rapid population growth.

planning are emphasised. Furthermore, other dimensions of planning are not considered in this exposition, for example, regional planning, which may be influenced by other affairs. Many chronological overviews to aid the understanding of the diversity of philosophies, events and issues which lead to the present configuration of planning theory can be consulted in the many text books and papers of authors in the field (see, e.g.: Hall, 1975; Galloway and Mahayni, 1977; Friedmann, 1987; Hall, 2002; Campbell and Fainstein, 2003; Ward, 2004; Jenkins, 2007).

Over the recent history of urban planning, as has happened with other urban disciplines (urban economics and urban geography), the theory and practice have changed significantly, trying to follow the main challenges of the ongoing process of urbanization. Different urban planning approaches emerged in response to the range of forces operating on cities, with the main purpose of contributing to the introduction of some forms of regulation and promotion of the use of land. To deal with problems and opportunities, in this complex world, and to provide and improve quality of life of the communities, planning is widely recognized as an interdisciplinary discipline, where an approximation between theory and practice should be guaranteed.

Before proceeding with the framework of examining space in urban planning, clarifications about the definition of some concepts are presented below. The difficulty of delimiting urban planning starts, according to the literature (Hall, 2002, for instance), with the definition of an urban area and its multidisciplinary nature. Urban environments are in their essence complex systems. Multiple interrelated elements, such as social, cultural, economic and political, operate not in an empty container, but instead within a physical and built environmental with its constraints and opportunities. Therefore, urban planning is considered by Hall (1975) a special case of general planning, of a more limited and precise nature, which incorporates spatial representation. In turn, the word planning[39] is "(...) *given a bewildering variety of meanings"* (Banfield, 1959, p.361), but can be defined as "(...) *a special way of deciding and acting*" (Webber, 1973, p.74 *cited in* Camhis, 1979), which involves the

[39] Planning is also known as physical or spatial planning which is perhaps a more neutral and more precise term (Hall, 1975), which is defined in EU compendium of spatial systems and policies as follows: "*(...) methods used largely by public sector to influence the future development of activities in space* (CEC, 1997, p.24)".

application of scientific and technical knowledge (however crude) to policy making in order to solve the problems and achieve the goals of a social system, trying to preserve its strengths and opportunities (Faludi, 1973c; Alden and Morgan, 1974; Friedmann, 1987). The application of planning theory, in Archibugi's (2008, p.13) words, "(...) *must help planning to be comprehensive and consistent, (...) the most important aspect of the true skill of the planner".*

In summary, urban planning or city and regional planning, or simply planning, are essentially concerned with shaping the future, where better conditions for human life can be achieved (Ward, 2004; Hiller and Healy, 2008).

In all of these previous definitions, it is implicit that planning can take place at various scales. As quoted by Friedmann (1993, p.482) "(...) *planning is instituted at all levels of public decision making*" such as, local, regional, national or transnational. Even if we are talking about urban planning, as a narrow concept of planning, different scales may arise, from neighbourhood or community level to a larger geographic scale such as metropolitan areas (at this stage the threshold between urban and regional planning is not so obvious). As such, the context within which urban planning should try to accomplish its objectives can be understood.

Over the last sixty-five years, urban planning has permanently changed according to different established paradigms[40]. Hence, the fist paradigm is placed during the period from 1945 to the late 1960s and is designated as planning-by-design or comprehensive planning. This period was characterised by a technocratic, positivistic approach, largely influenced by architects and civil engineers; the key planning instruments being master plans and regional plans. The second paradigm encompasses the period between 1960 and 1970, known as system and rational planning*;* the key planning instruments here are structure plans. And finally, the third paradigm, planning-as-negotiation, where the key planning instruments are arenas for negotiation and dialog: *a*

[40] The concept of paradigm is linked with the name of Thomas Kuhn, which defines it as (Kuhn, 1962, p.10) "*a radical change in underlying beliefs or theory*", in other words, describes fundamental theoretical changes in the history of science and ideas, in which, once established, shapes the whole way a scientific community (Galloway and Mahayni, 1977). Kuhn's postulate that science was not steady; neither the scientific advancement is based on evolutionary nor cumulative acquisition of knowledge instead is a "*series of peaceful interludes punctuated by intellectually violent revolutions*", and in those revolutions "*one conceptual world view is replaced by another*". To better understand how Kuhn's discourses on the development of scientific thought help to explain the diversity of directions in planning literature, see Galloway and Mahayni (1977)

participatory approach is commonly used in planning processes. In addition, it should be mentioned that these traditions of planning are associated with three philosophical trends that exist in social sciences: i) positivism; ii) structuralism and iii) post-structuralism (post-modernism).

To return to the main question of this chapter, the fundamental aspects of the historical processes and contents, and their implications for the conceptualization of space are presented below. The remains of this chapter are organised under three main points, each one representing paradigms in non-socialist core countries that have emerged in the literature (Taylor, 1998; Jenkins, 2007), during the second half of the twentieth century.

III.4.1. Plan-making processes: Planning-by-design, Blueprints and the Positivism approach

Urban planning practice, in the early years of the post Second World War phase, had as its main purpose the creation of health, attractiveness, efficiency and safety in communities; the principal concern was physical urban environmental factors, mainly forms and patterns. The urban (town) planning was essentially an exercise in physical planning and design, viewed as a natural extension of architecture or civil engineering, involving identical kinds of spatial design abilities. The emphasis on urban *planning-as-design* necessarily involved the preparation of master plans, in which aesthetic aspects of urban space were a key concern. A prime task of these instruments of planning should be, in as detailed a fashion as possible, to conduct the future development of spatial configuration of land use and urban form, essentially based on the design of the entire physical layout of buildings and spaces. Plans were seen as blueprints for the future configuration of cities, produced usually by architects or engineers. This is hardly surprising, given that most built environment practitioners in the post-war period were architect-planners[41]. The aesthetics of urban form and design dominated the standard urban town (urban) planning literature in the

[41] According to Cherry (1974), *cited in* Taylor (1998), over the period of 1946 to 1956, 45% of the Associate Members of the UK Town Planning Institute were architects, as compared with the 22% of 'direct entry' planners, 14 % of engineers, 9% of surveyors, and 9% holding other first-degree qualifications, such as geography. This situation was reflected in other European countries.

post-war period and planning decisions were made largely on the basis of intuition (Taylor, 1998; Jenkins, 2007). It tended to impose an Euclidean order upon the organic forms of nature, where "(...) *the straight lines and right angles of orthogonal design were classic instances of an artificial, rational ordering of space*" (Friedmann, 1987, p.22).

The blueprint character of urban planning is well illustrated by the plans established at that time. Two examples, described by Taylor (1998), emphasised the primacy of the aesthetic aspects. The first, is related to the location of industries were deemed best placed far away from the residential areas, because there were considered unpleasant to live nearby (a very correct option in many cases). The other example is related to the planning of the new urban areas, where plans were laid out based on an orderly pattern of physically distinct neighbourhoods, all of which were generally the same size.

Two seminal works by Frederick Gibberd (1967) and Lewis Keeble (1952) summed up this physicalist design-based conception of urban planning. Both books were widely recommended to planning students in the English-speaking world throughout the 1950s and early 1960s (Taylor, 1999). As Lewis Keeble (1952, p.10) affirmed "(...) *the town ought to have a clear legible structure*" underlying the fact that the major uses of land should be clearly distinguished and provided separate zones. Also, Buchanan (1963) proposes that well planned cities should be orderly cellular structures of geographically distinct neighbourhoods or environmental areas. This was the central argument that most urban planners espoused for the ideal urban structure.

These ideas of exhausting zoned plans, specifying how particular sites were to be used, can be seen in its most influential form in utopian schemes for ideal cities, such as: in Arturo Soria Mata's plans (19th century) for La ciudad Lineal (in Spain)*;* in Theodor Fritsch's (1896) for Die Stadt der Zukunft (in German); in Georges Benoit-Lévy (1904) for La Cité Jardin (in France); and later on, in Le Corbusier's plans (1920s and 1930s) for the radiant city, in Frank Lloyd Wright's plans (in the 1930s) for Broadacre City, as has also been Ebenezer Howard's Garden City (Taylor, 1998; Davoudi and Strange, 2009). For these social visionaries, planning the city was a way of creating a peaceful path to real reform. Albeit, these utopian images were based on their creators' vision of how the future should look like, rather than a systematic scientific approach; it was as much an art as a science (Davoudi and Strange, 2009), a kind of "(...) *pseudo-science*" in Jacob's words (1969, p.16). Although, according Davoudi and Strange (2009), these urban planners were not considered physical planners their ideas had elements of physical determinism. As concerns space, it was seen as a neutral container, a blank canvas that is filled with human activity (Hubbard, Kitchin *et al.*, 2004).

These examples held sway during the years following the Second World War in Britain, but were not peculiar to Britain in the sense that it was a international movement, namely with a following in other European countries and in North America, as quoted by Taylor (1998) author goes further stating that the European concept of town (urban) planning has proved more durable than in Britain.

In the post-war era planning theorists assumed a general consensus concerning which values and ideas urban planning should embody. Principles of good urban planning, viewed as being in the public interest, were seen as self-evident (see, e.g., Keeble, 1952). Given this unitary view, adhered to by several planning scholars (Keeble, 1952; Hall, 1975; Taylor, 1998) the assignment of planners was simply a practical one of finding the technical instruments necessary to achieve given objectives, that is, blueprint plans where urban patterns of future developments were expressed. Thus, planning was seen as mainly a technical or practical exercise.

In short, the planning-by-design approach emerged as a response to population growth and the pressure for expansion in cities focusing essentially on the exact disposition of all land use. The main objective was often to limit the city growth, in a period which private land rights needed to be restricted by professionals (urban planners) who were seen as neutral experts. This tended to be very precise and was a product of the three-stage classical paradigm planning process of survey, analysis, and plan associated with Patrick Geddes. The considered pioneer town planner argued that planning had to start with a survey of the resources of such a natural region (Hall, 2002, p.147). However, these survey works were essentially related to physical aspects of the cities. Four main aspects characterise the Geddesian dictum of the *survey-analysis-plan* process of planning: first, planning processes terminated with the production of the plan; second, the outcome was one single plan rather than a number of alternative strategies (Taylor, 1998); third, planners themselves are guardians of public interest (Hall, 1975); and fourth, planning was based on the infallibility of experts, reinforcing the apolitical, technical nature of the process (Batty, 1979 *cited in* Davoudi and Strange, 2009). However, as stated by Taylor (1998) three main drawbacks are inherent to the process. Firstly, it remained unclear as to the usefulness of a survey, since as quoted by Hall (2002, p.324) "(...) *the goal was left implicit, to be defined intuitively by planners themselves who were seen as experts, apolitical and the guardian of public interests.*" The second shortcoming is related to the use of the term 'plan' in the singular, which implies that the outcome would be only one possible plan, rather than multiple and alternatives strategies. Thirdly, a *survey-analysis-plan* based approach implies that the process of planning would end with the production of the plan. Despite the rhetoric "(...) *plans and planning decisions were made largely on the basis of intuition or, rather, on the basis of simplistic aesthetic conceptions of urban form and layout which embodied physical determinist assumptions about how best to accommodate the diverse economic and social life of cities*" (Taylor, 1998, p.14).

Survey methods before planning are concepts that illustrate the positivism approach to the process of planning (Hall, 1975). Although the utopian visions of planning could be considered as a non-systematic scientific analysis, their creators envisage physical outcomes (limited to physical descriptions) with elements of positivist interpretation of space (Davoudi and Strange, 2009). The creation of some form of spatial ordering, usually expressed on urban maps, became "(...) *inseparable from an ongoing labour of seeking to tell the truth about the city*" by employing "(...) *mundane techniques of gathering, organisation, classification, and publication of information*" (Osbourne and Rose, 1999, p.739). In the map's geometrical and physical perspective "(...) *space is delineated, reduced to the clarity of the line"* where the elements (such as, streets, buildings, etc.) were "(...) *differentiated but composed of the same medium, that of an extreme form of geometrical space"*, allowing the observer to view the city as a unique entity (Joyce, 2003, p.54)".

This short review of the planning-by-design approach in urban planning thought and practice, which dominated the field in the post-war paradigm, helps to provide a framework to understand the major criticisms against this approach presented below. Many planning writers launched several criticisms in their writings (Jacobs, 1969; McLoughlin, 1969; Taylor, 1998; Hall, 2002; Jenkins, 2007) These can be summarised into five points.

Criticisms of planning-by-design approach

i) *Criticism of physical and morphological determinism*:

This criticism is strictly related to the rigidity of land use zoning plans, potentially unrelated to the non-materialist forces which shape urban development. Urban planners had a propensity to analyse cities and their problems only in physical and aesthetic perspectives. Social considerations were ignored, even if the plans indirectly reflected social, economic and other environmental concerns. Keeble (1952, p.1) describes the urban planning, in the first page of his planning textbook, as follows: "(...) *the art and science of ordering the use of land and the character and siting of buildings and communicative routes*." This view was one which emphasised the physical shape and urban development. The author considers that the urban planning deals "(...) *primarily with land, and is not economic, social or political planning*." The distinction between physical and social planning, as well economic is clearly pointed out. Although Lewis Keeble advanced that "(...) *though it is not social and economic planning, it may greatly assist in the realization of the aims of these other kinds of planning*", the statement is controversial and incongruent. It is implicit in this last citation that physical form of urban structures can affect or determine the quality of social or economic life, assuming that social, economic and political ends could be achieved by physical means. Cities are a form of social action and in the nineteenth-century a planning movement emerged for reasons of public health, generally regarded as being within the social remit. Also, the idea that urban planning is apolitical is questionable, in the sense that decisions about land use, for instance, inevitably involve making choices that positively and negatively affect interests of different people (Taylor, 1998).

ii) *Criticism of being rigid (inflexible)*:

This second criticism, related to the previous, underlies the fact that detailed plans for specific localities were not appropriate instruments for longer time horizons. Because of its rigidity, planning practice did not include a strategic component. Precisely delineated boundaries of zoned land use did not allow possible changes that could occur in the process of urban development. The focus was more often on the plan as a product, as a one-off exercise, rather than an ongoing process and its effects. This point is well underlined by Brown and McLoughlin (Brown, 1968; McLoughlin, 1969) when both these authors stressed the tension between the dynamism of urban functions and the fact that any plan attempted to specify its future form by freezing it into fixed land use zones. The results could surely lose relevance over time. The example described by Taylor (1998) is a good illustration of this problem: the frequently implicit negative view of urban growth led to an underestimation of future urban population. Implicitly, it is suggested here that urban planning should address a wider range of issues than matters of physical design and aesthetics.

iii) *Criticism of consensus and lack of participation*:

The third main criticism regards the consultation practice (or the lack of it). Urban planners were criticised for failing to involve the relevant agents in discussions of their plans. Reasons cited for this hiatus include on one hand, a failure to recognise the importance of value judgments by residents, and on the other hand, the poor institutional link between professional plan development, political interests, decision-making on city budgets, and other agencies involved in infrastructure and service provision (Devas 1993; Dwyer 1975; Lowder 1986, *cited in* Jenkins, 2007). As a consequence inhabitants wishes were not taken into account when elaborating neighbourhood plans with planning decisions just technical matters, made by experts, planners who assumed they knew best. In summary, this third point, related to the process of planning, criticised planners for their misunderstanding of planning and how it works in pratice (Davoudi and Strange, 2009).

iv) Criticism of simplicity and utopian perspective:

The inadequate understanding of how cities actually functioned was evident in the activities of post Second World War urban planners, and the substantive content of their work (Davoudi and Strange, 2009). Talyor (1998), quoting Jacob (1961), mentions that planners demonstrated a superficial understanding of the cities. The argument asserted by the author blaims an excessive concern with simplistic utopian visions, instead of trying to address real life problems of the cities. The lack of financial analysis and the unrealistic assumption of the economic costs of interventions support Jacob's position (1969, p.16), where at the beginning of the "*The Death and Life of Great American Cities*", she writes that "(...) c*ities are an immense laboratory of trial and error, failure and success, in city building and city design.*" There was incomplete information about existing and future developments. Urban planners were seen as architectural designers on a larger canvas (Taylor, 1998).

v) Criticism of separation and ordered view of urban patterns:

The standard principles of planning urban land use into separate and distinct homogeneous zones were also subject to criticisms. Jane Jacobs (1969, p. 23-24) was an apologist for the richness of urban activity and for the diversity and mixture of uses, not their neat separation into single function zones, which is a precondition of good city life, in the sense that it "(...) *is the need of cities for a most intricate and close grained diversity of uses that give each other constant mutual support, both economically and socially."* The author (1969) considers that urban diversity is resultant from the life sciences, in particular from the three main techniques for revealing organized complexity in dynamic systems; first, a need to analyse processes and their catalysts; second, a requirement for reasoning from the particular to the general rather than the reverse; and third, a need to look for outlier clues which reveal the way larger and more average processes are operating.

These aspects emphasised concerns for social complexities and for the dynamic and emergent qualities of cities.

III.4.2. Rational decision-making: systems and rational planning, structuralism and the modernism approach

As mentioned above, urban planning theory and practice, for almost 20 years following the post Second World War, was dominated by a concept of physical design. As a reaction to some critiques mentioned above, during the mid- to late 1960s, two distinct theories were responsible for a rupture and for a paradigm shift in planning thoughts[42]: the systems view of cities, and the view of planning as a rational process of decision-making[43]. This break with tradition was expressed in two ways: i) while the former focused on the object or the substance of planning, addressing the structure and the functioning of a city (the main aim was to improve the understanding of the problems which planning addresses, now seen as a system of interconnected part); ii) the latter was mainly concerned with the method and procedural component of planning itself, including its ideology, values, purposes and principles (Hightower, 1969). Andreas Faludi (1973a) distinguished and labelled these two subjects, in *theory-of-planning* and *theory-in-planning,* respectively. The distinction is controversial because one cannot be applied without understanding or considering the other, that is, planning process and content of planning policies should not be separated. Being usually considered together these two theories of planning are distinct from one another, the proof is that is possible to subscribe to one and not the other (Taylor, 1998).

The systems view of planning was developed by Norbert Wienner in 1948, derived from the science of cybernetics (Davoudi and Strange, 2009) and imported into planning by the work of Brian McLoughlin, where in his book,

---

[42] This paradigm shift is well represented by Andreas Faludi's seminal works, *Planning Theory* and *A Reader in Planning Theory* (Faludi, 1973a; Faludi, 1973b).

[43] Both of these planning thoughts originated from more general theory with had developed outside of the field of urban planning (see, e.g., Simon, 1945; 1960; Faludi, 1987)

"*Urban and Regional Planning: A Systems Approach"* (1969), conceptualised urban structures as complex systems, or indeed as sub-systems of an entire system. The focus was, besides the physical environment, the social and economic aspects, particularly, transportation planning and environmental qualities. These two concerns emerge clearly in the Buchanan report (1963), "*Traffic in Towns"*, which made significant contributions to the theory and methodology of planning.

The acknowledgment that urban systems are in permanent transformation leads one to assume that an *ex ante* understanding of the system is crucial, based on reliable information to define suitable strategies of development. This shift also recognised the necessity of involving professionals with skills in economics and social sciences rather than design (Taylor, 1998), an essential feature in the first paradigm in planning. Urban planning began to be conceived as a form of systems analysis, as a continuous process of control and monitoring of the urban system (Taylor, 1998). The empirical evidence required was "(...) *more analytical than descriptive and more sophisticated than the simple survey work advocated by Geddes*" (Davoudi, 2006, p.17).

Planners, at this stage, start to explore the principles of positivism to understand the complexity of systems beyond their physical description. In this way they pulled away from the simple descriptive physical survey (expressed by detailed maps and blueprints). An engineering-based component emerged with the use of mathematical techniques and data processing powers of computers (Hall, 2002). The development of computers, able to process large volumes of data, encouraged the systems theory in urban planning. This quantitative revolution, that took place in the 1960s, stemmed from a desire to make social sciences (particularly geography and planning) more rigorous and scientific rather than an art (Taylor, 1998; Davoudi and Strange, 2009). The essence of scientific methods, advocated by Karl Popper, who proposed that the "(...) *criterion for the scientific status of a theory is its falsifiability, or refutability, or testability*" (Popper, 1959, p.37), had begun to be adopted.

As noted above, rational decision-making planning is a process or procedural theory and not a substantive planning theory [using Faludi's (1973a) terms]. For this reason, only the process (methods and means of planning) is emphasised. The rational process model does not say anything about the substantive end or goals of urban planning. It can be considered, in words of Taylor (1998), merely a formal and instrumental model of reasoning and should fulfil two major conditions. The first premise is that planning decisions should be cautiously analysed rather than made by instinct. The second condition is related to the previous one, and implies the explicit clarification of the reasons for making planning decisions. In this context, urban planners should be able to clearly explain every stage of a planning process and the reasons for each option.

The increasing interest in corporate styles of management by governments, following the recommendations of the Bains report (1972) and, the increasing belief in the application of scientific principles[44] to policy making, contributed favourably to the generalised acceptance of the rational process view in the planning practice.

While being two different approaches to planning the system that emerged simultaneously, the views of the rational planning processes were often considered together, sharing some common assumptions about the nature of the world and social action. Both believe in people's capacity to improve the social and economic conditions (human well-being) on the bases of scientific or rational understanding of the environment. The two planning thoughts prevalent in the 1960s were closely associated with the rise of modernism[45] and its exuberance for science and technology (Giddens, 1994; Taylor, 1998; 1999).

The emergence of systems and rational process view of planning is explained, in part, as a logical response to the deficiencies and criticisms of the traditional planning-by-design. Four major differences summarise this shift in planning thought. First, once recognized that cities are set of distinct, interconnected and interdependent parts, it became clear that planners needed to understand the urban phenomena as system, determined by the structure of its parts and their relationships. Thus, a physical and morphological view of urban planning was substituted with the logic of analysing cities as a set of complex and interrelated systems in a permanent dynamic. This was one of Jane Jacobs's (1961) main criticism of traditional planning theory, that is, the lack of

[44] Note that, being scientific meant the use of quantitative analysis, otherwise it was not considered to be scientific (Taylor, 1998).
[45] However, the modernist faith in reason and science began with its roots in the European Enlightenment of the eighteenth century (Young, 1990).

understanding of the complex reality which planners were dealing with. It should be mentioned that a change to one part of the city will directly or indirectly affect some other part of the city. Second, while planners had a propensity to view planning predominantly in design and aesthetic terms, they were now to examine the urban phenomena in terms of its social life and economic activities. Rather than architectures it suggested that planners should be trained in analysing and understanding social and economic features. The third aspect is related to dynamism and changeability of the planning activity, which emerged in opposition to the master plan's end-state approach. Because the urban contexts were seen as a live, functioning thing, the need for more adaptable flexible plans and the idea of ongoing processes of monitoring, analysing and intervening in a fluid situation was required. The fourth and last divergence concerns the kind of skills and techniques considered appropriate to urban planning. If it is assumed that cities are a set of complex interrelated systems, not just physical and aesthetical elements, rigorously analytical, scientific methods of analysis can be imposed. The major shift in urban planning theory and practice was from the planner as a creative designer (an artist) to the planner as a scientific system analyst and rational decision-maker (Taylor, 1999). Of course these changes were not abrupt, in other words, did not involve the complete replacement of one view by another. In Talyor's (1998) words the revolution was not a wholescale and was carried out at two distinct levels (and speeds): at a larger scale, considered more appropriate to define strategic orientations and, at a local scale, for more immediate intervention, in which a long way is still missing.

The background of the theoretical and intellectual movement behind this approach, both in systems and rational planning, is structuralism, which refers to a rationalist approach to scientific knowledge (see e.g., Francis Bacon's first way of discovering truth in Davoudi and Strange, 2009). The task is to discover hidden structures, forces and laws of human behaviour (whether physiological or social) and reveal secrets of the natural orders which lay beyond all explicit perceptions (Hollis, 2003). Using metaphors strictly related to Euclidean structuralist theories "(...) *tends to enact and produce a Euclidean reality of discrete entities of different sizes contained within discrete and very often homogeneous social spaces*" (Law and Urry, 2004, p.398). Euclidean laws of geometry were considered, for this purpose, a very reasonable way of providing this access to reality, which can work as metaphors, associated with height, depth, size and density and frequently referred to as a homogeneous social space (Law and Urry, 2004; Murdoch, 2006; Davoudi and Strange, 2009). In the discipline of linguistics the structuralist approach emerged through the work of Ferdinand de Saussure, in the early years of the twentieth century (undertaken just prior to First World War). Saussure laid the foundations for a structuralist approach to the study of language, considering that the relationship between given words and given objects are purely arbitrary (Murdoch, 2006).

Within social sciences (anthropology), it is the French anthropologist Claude Levi-Strauss[46] (in the middle years of the twentieth century), whom is widely regarded as the father of structuralism (Kurzweil, 1980). He introduced structuralist principles to a wide audience. Lévi-Strauss' work ("The Elementary Structures of Kinship") demonstrates that there was original structure which was both universal and ahistorical (Lévi-Strauss, 1947), believing that the complex details of social and cultural life could be thus explained. In his work, this French anthropologist and ethnologist, as noted by Descombes (1994, p.115), structures work in terms of three metaphysical statuses: *"(…)* as natural causes,. as mechanisms which generate phenomena"; "(...) *as laws of spirit, as constants which the observation of cultural phenomena helps us to discern*"; and finally "(…) *as ideal rules, as intellectual models which agents could not follow if they did not have some understanding of them*"[47].

Structuralism extended into geography, planning and urban theory largely through structuralist Marxism, during the mid-1970s. Within this perspective, the city itself is a result of capitalist logic, in pursuit of profit, and planning is considered to be a part of the problem (Davoudi and Strange, 2009). The preface of Marx (1859) relates the perspective that in societal developments observation beyond the actions of individuals is highlighted, where it is said that "(…) *it is not the consciousness of men that determines their being, but, on the contrary, their social being that determines their consciousness*". In opposition to the Aristotelian's essentialism (Section III.3.1), such relational thinking was central to Marx, who insisted that materialism should be both dialectical and historical.

[46] Four basic procedures of structuralism are specified by Levi-Strauss: i) structural analysis examines unconscious infrastructures of cultural phenomena; ii) it regards the elements of infrastructures as relational, not as independent entities; iii), it pertains single-mindedly to systems; and iv), it propounds general laws accounting for the underlying organizing patterns of phenomena.

[47] For more developments see for example, the book *"The Savage Mind"* (Lévi-Strauss, 1964).

This dialectical reasoning (dating back to Greek philosophy) focused on analysing the relations between things rather than the things themselves (Henderson and Sheppard, 2006). Marx remained indifferent as to the question of space, nevertheless, Engels, in the mid-nineteenth-century, developed Marxist ideas to the spatial distribution of classes. Later on, after a century, seminal works applying a Marxist analysis to space and a structural reading of a city emerged (Davoudi and Strange, 2009). Among others, David Harvey (1973), Manuel Castells (1977), Doreen Massey (2004; 2005), and Henri Lefebvre (1974 [1991]) stand out. As has been explained in more detail in the previous Section III.3, in Harvey's (1973) words, space is socially produced and for this reason the question of "*what is space*?" should be replaced by the question "(...) *how is it that different human practices create and make use of different conceptualization of space?* (1973, p.13 and 14). For Castells (1977, p.124) there is no specific theory of space. According to him, "(...) *space, like time, is a physical quantity that tells us nothing about social relations*". Space becomes a reflection of social process, where is determined by structural laws acting beyond the social structures, which are a combination between activities and their locations. Massey (2005) argues as to the dichotomous understanding of the notion of space and place, suggesting that there is no reason to consider these two concepts as oppositional. Space in a structuralist perspective is frequently seen as abstract sited beyond the level of immediate human influence and understanding, and place, in contrast, as real, grounded, meaningful and lived. Massey considers space as "(...) *no more than the sum of relations and interconnections*" (2005, p.184), and place as "(...) *a product of relations which spread out way beyond it*" (2004, p.6). In sum, in Doreen Massey (1993), space is a complex intersection of a set of social, cultural, political, biological, economic, physical relations and power geometries, because different social groups and different individuals are placed in very distinct ways in relation to these flows and interconnections. It is the scale that differentiates these myriad of spatialities. Lefebvre (1974 [1991]) rejected the idea of absolute or abstract space claiming that "(...) *space considered in isolation is an empty abstraction*". Every society and every mode of production produces its own space, emphasising the notion of space as socially produced. Lefebvre's analysis of the history of space and place as socially constructed marks the transition from structuralism to a new way of theorising space associated with post-structuralism thinking (Davoudi and Strange, 2009).

As mentioned above this second paradigm had developed partly in response to the insufficiencies of its predecessor, however, it itself was confronted with similar critiques. Considering the criticisms of procedural planning, from the second wave of criticisms[48] which emerged in the 1970s two major points can be highlighted.

Criticisms of systems planning and the rational approach

i) *Criticism of lack attention to the content and substance*

This point is related to the lack of content or substance about how planning in practice operates (see, e.g.: Camhis, 1979). Procedural theory, expressed in Faludi's (1973b) work, is criticised to be essentially empty and vacuous giving no explanation for the products of planning. Jane Jacobs's (1961) main criticism relies on the lack of understanding of the complex reality and the lack of a firm theoretical foundation demonstrated by the practice of planning. As mentioned above, the rational and systems view of planning emerged, in part, in reaction to the earlier criticism of planning-by-design theory (for its lack of understanding of how urban systems functioned), although, systems and the rational process view of planning continued to be expressed in highly abstract conceptual terms. In addition, Friedman (1969) stated in unequivocal terms that the rational view of planning showed little understanding of the action end of planning, being a false top-down approach to planning. The expressed idea here is that urban planning and planners exists not only to understand the world, but more relevantly, to improve it. The impacts and implementation of planning principles and practices were not taken into great attention. This argument is stated by John Friedmann (1969, p.312 *cited in* Davoudi and Strange, 2009), criticising the top-down view of planning with little, if any, understanding of how plans were implemented and

[48] The first wave of criticisms emerged tackling the theory of planning-by-design (post-1945 planning theory).

the nature of their consequences. Quoting Scott and Rowies (1977, p.1116) there was "(…) *a definitive mismatch between the world of current planning theory* (…) *and the world of practical planning intervention*".

ii) *Criticism of faith in scientific methods*

The second aspect is related to the unreasonable belief (the optimism) that the environment (understood as a system of interconnected activities) should be studied and analysed through the application of scientific and rational methods of decision-making, for the overall purpose of enhancing human welfare. This statement leads Taylor (1998, p.66) to point out that the systems and rational process theories of planning, were considered "(…)*part of the heady modernist optimism of the 1960s*." Moreover, the premise of maximising utility (benefits, happiness, wellbeing, etc.) for a specific group of inhabitants as a whole, obtained by computers, used to model complex systems and applied in a logic of utilitarianism, as if the city were some kind of machine, has been subject of intense controversy (Taylor, 1998). Citing Scott and Rowies (1977, p.1116) planning theory "(…) *set itself the task of rationalising irrationanilties* (…) *bringing* (…) *a set of abstract, independent and transcendent norms*".

In short, the main differences between the first and the second wave of criticisms, as pointed out by Taylor (1998), is that the former criticise planners for their inadequate understanding of urban contexts (related with the substantive content of planning), while the latter criticised planners for their misunderstanding of planning and how it works on the ground (related with the process of planning).

III.4.3. Political decision-making: Planning-as-negotiation, post-structuralism and post-modernism approaches

In this sub-section, a third significant change in post war planning thought, which some authors have branded planning-as-negotiation, is presented. This new emergent paradigm in planning is not independent of the shifts taking place in western thought and culture, from modernism, structuralism orthodoxy to post-modernism and post-structuralism, which began to be highly influential in the social sciences and in particular in planning theory. Thus, post-modernism and post-structuralism planning theory is seen as representing a rupture with the analytical approach and scientific understanding of the phenomena. Broadly speaking, and quoting Young (1990), in the former prevalence is given to simplicity, order, uniformity and precision; and in the latter, aspects like complexity, diversity, difference, multiculturalism and pluralism are typically emphasised. Postmodernist conception of urban planning assumes that the understanding of people's experience of places is much more diverse than was implicit in many modern schemes.

Leonie Sandercock in his book "*Towards cosmopolis: planning for multicultural cities*" (1998), attempts to set directions for planning thought and practice in the post-modern era and advances two important distinctive features between the modernist (which the author censures) and post-modernist paradigm of urban planning (which the author approves). The first is related to the propensity for quantitative methods and scientific knowledge in planning (and social science in a general sense). Sandercock (1998) considers that despite the fact that means-ends rationality may still be a useful concept, a greater and more explicit confidence on practical knowledge is needed, which is derived from several kinds of experiences; grounded, contextual and intuitive knowledge. These kinds of knowledge are expressed through speeches, song, stories and various visual forms. The second distinction involves the way that overall public interest was captured. Instead of assuming the top-down state-directed model of planning, Sandercock (1998) supports the idea of community-based planning and recognise that planning "(...) *is no longer exclusively concerned with comprehensive, integrated and coordinate action, but more with negotiated, political and focused planning*" (1998, p.30). However, these arguments do not contest a scientific position, or provide judgement on the experience and knowledge of local communities.

Post-structuralism thought is premised on searching for multiplicity of meanings and assumptions (Davoudi and Strange, 2009) and characterised positivist and structuralist approaches as not being a representation of the truth (Murdoch, 2006). Despite the variety and disparity of post-structuralism threads (see e.g.: Olssen, 2003) all have in common their opposition to the idea that single underlying structures determine social action. As Massey (1991, p.28) notes: "*If it is now recognised that people have multiple identities then the same point can be made in relation to places*". Post-structuralists argue that the world is far more complex and subtle than has typically been conceived, a mixture of natural and social and human and non-human heterogeneous relations (Murdoch, 2006). This argument is in line with the one expressed by Jacobs (1969), giving preference to complexity in the city rather than the simplified order which modern urban planning theorists advocated, criticizing single zone options and comprehensive development. She points out (1969) that a standard concept of environmental quality, that pleases everyone, is a somewhat difficult task to achieve. In short, according to Murdoch (2006), post-structuralism affects the analysis and the perception of space in two ways: on the one hand, through new attention to differences in spatial identifications; and on the other hand, a new interest in processes of spatial emergence.

Since urban planning action can extensively influence a large amount of inhabitants, and since different groups of people (or even individuals) may have opposite perspectives and interests regarding how urban environment should be planned, the adoption of a purely technical and scientific approach can be inadequate (and has been the most scathing critique of the systems and rational process view of planning). Thus the introduction of mechanisms of participation in the lexicon of planning was a shift in the concept of planning (Taylor, 1998). It is still considered the dominant paradigm of planning theory.

The concern with participation in the planning activity was not entirely new, because in the master-planning periods great attention was given to the presentation of plans, where the main objective was to communicate proposals of the plan clearly and attractively. However, little attention was paid to the participation, seen as a process of dialogue, debate and negotiation (Taylor, 1998).

Thus, the third paradigm of planning considered includes participation in the planning process. In a bottom-up approach, the planner is seen as a kind of facilitator, where the exercise is no longer viewed as merely technical or scientific. The acknowledgement that planning decisions should be more participative emerged because of two reasons: first due to the idea that urban planners were not technocratic elites or expert professionals, and second, because urban planning could not be considered simply a rational process where the main mission was the requirement of efficient and meticulous data collection and analysis (Jenkins, 2007). It was acknowledged that the role of the urban planner should be as an advocate for those groups whose interests were poorly represented in the process of planning. In this sense, this view of planning is more meaningful and comprehensible to local communities. Acceptance of this position means, according to Davidoff, the "(...) *rejection of prescriptions for planning which would have the planner act solely as a technician*" (Davidoff, 2003 [1965], p.210).

However, the notion that public participation meant consultation was contested. Arnstein (1969, p.216) says that the idea of participation "(…) *is a little like eating spinach: no one is against it in principle because it is good for you.*" The author argues that public participation can mean different things: from "(…) *empty ritual of participation", in which citizens are at least informed about the planning action, to situations where citizens are consulted about the proposals* (...) *having real power to affect the outcome of the process*". In this situation public participants are enable to negotiate in trade-offs with traditional powerholders. Obviously this is a simplification, but it helps to illustrate the point that there are is vast range of levels of citizen participation[49]. The important aspect that

[49] Arnstein (1969, p.127) suggested a typology of eight levels of participation. In the level of a non-participation process there are two: 1) Manipulation and 2) Therapy. Then, 3) Informing and 4) Consultation are a kind of 'tokenism' that allow the have-nots to hear and to have a voice. 5) Placation is simply a higher level tokenism because the ground rules allow have-nots to advise, but

should be emphasised is that participation without re-distribution of power is an empty and frustrating process for the powerless (Arnstein, 1969). It is important to note that communication and negotiation is not the same thing and requires different skills, "(...) *effective communication would seem to be a necessary but not a sufficient condition of effective negotiation*" (Taylor, 1998, p.122).

Early on it was recognised that the interpersonal skills of communication and negotiation were required in the planning activity and two main challenges were essential (Habermas, 1979, *cited in* Taylor, 1998): the first was the capability to identify and establish contacts with other relevant actors necessary for the implementation of a plan or policy; the second, because different individuals or groups of agents have their own interests which do not always coincide with those of the public interests, planners and policy-makers must acquire the skill of negotiating (Taylor, 1998). The tradition that emerged of the urban planner was as someone able to identify and mediate between different agents involved, someone who acts as a translator of planning issues, not someone which as a superior technical expertise in making value-judgments about urban land development. Nonetheless, it should be noted that to be successful as a negotiator, urban planners would require some specialist substantive knowledge to bring the fundamental elements of planning decisions to the discussion (Taylor, 1998).

Hirt (2005) outline five related areas with which post-modern planning might be concerned: i) a growing interest in participatory planning; ii) the search for urbanity, urban identity and cultural specificities; iii) an appreciation of historicity of places and a return to traditional urban forms; iv) the primacy of mixing land use and flexible zoning; v) return to a human scale approach, urban compactness and high-density development.

Despite the advantages over its predecessors two major arguments summarise the critique in relation to this third planning paradigm.

---

retains the right to decide for the power holders. Citizens can enter into a 6) Partnership that enables them to negotiate and engage in trade-offs with traditional power-holders. At the topmost rungs, 7) Delegated Power and 8) Citizen Control; have-not citizens obtain the majority of decision-making seats, or full managerial power.

Criticisms of the negotiation approach to planning

i) *Criticism of consensus in decisions*

Participation was viewed as a process of involving people, a public consultation, rather than a public actively participative in decision making. The notion that participatory procedure was advised by a professional planner which had the decisive responsibility and influence to give final decisions, was contested by many planning writers. Planning decisions, which account for everyone's interests (privileging the views of all individuals) appears to be hard and even infeasible with no limit as to the range of possible interpretations for any given situation.

ii) *Criticism of non-neutrality rhetoric*

This criticism has been transposed from the previous paradigm (Planning-by-design). It seems evident that the existing perception of planning as neutral is arguable, because of two main reasons: first, because planning is naturally politically inspired, and second, by its nature the activity of planning creates an expectation of conflicts. In addition, there has been further recognition of the limitations of the mechanisms of representative democracy.

III.4.4. Summary

The purpose of this section was to explore the concept of space in post-WW II urban planning theories and in the practice of spatial planning. Urban planning has been transformed by society over time and several important changes in thinking occurred.

From the planner creative-designer to the planner as a manager and communicator, two significant changes in urban planning paradigms illustrate the diversity of philosophies which have led to the present configuration of planning theory (considering only the period since 1945): the first significant change took place in 1960s, when the urban planning design tradition (physical shape design) was transformed into a systems and process view of planning (a shift from art to science); the second change, occurring in the 1970s and 1980s, evolved from viewing the planner as a technical expert to viewing the planner as a kind of negotiator, communicator or facilitator, allowing several participants or stakeholders to express planning judgements. Planning should shift to public affairs and collect opinions of the masses representing different interests groups (a shift from technical experts to managers of expectations).

In general, it can be said that throughout history the major change in urban planning theory, and its relation with the perception of space, was the transformation from the physicalist or materialist notions of space (or topographical) to a more social or immaterial space (or topological). Thus, space can no longer be seen as simply a container, but something organised by and constituted of heterogeneous relations (Murdoch, 2006). Friedmann (1993) introduced what he called the non-Euclidian mode of planning, referring to the move into a non-Euclidian world of many space-time geographies. These multitudes of geographic spaces are a consequence of the complex system of spatial interactions that occurred across territories leading to a non-obvious organization of both horizontal (spatial heterogeneity) and vertical (spatial scale) spaces. This is a recognition of the fact that new and more appropriate approaches and methodologies to deal with space are required.

III.5. The influence of space on the housing market

Housing can be defined as *"(...) the stock of houses, apartments, and other shelters that provide the usual residences of persons, families, and households"* (Adams, 1987, p.515), but it is much more than this. It is a: i) *physical facility* unit, which provides shelter to its inhabitants and requires services supplied by governments, such as, water and sewerage; ii) *social or collective good*, in which a certain minimum standard of housing is needed and should be available for all families, regardless of their ability to pay; iii) *package of services*, which is related to its location (accessibility of jobs and amenities) and neighbourhood (parks, schools and social environmental) attributes; and iv) *economic good and sector*, which is exchanged in a market, is part of the fixed capital stock and produces benefit and utility (Bourne, 1981).

As an economic good, housing can be clearly distinguished from other goods and services. Five major differences make houses rather special and different from other goods (Bourne, 1981; Adams, 1984; Bramley, Bartlett *et al.*, 1995; Costa, 2010): i) the first is its *durability*, a building can last for many decades or even centuries and the land where the house is placed is permanent; second ii) being an *expensive investment* (the only good that is both a consumption good and an investment) is for most individuals the greatest asset in their lifetime (many of them have to request bank financing in which the availability of credit and interest rates are important aspects helping to define the potential demand); the third important distinction iii) is related with *immobility*, both location and intrinsic attributes are consumed jointly with the housing itself in a spatial fixity; fourth, iv) when the location is taken into account, no two houses are exactly equal in terms of cost, space, location, and neighbourhood, thus the *inhomogeneity* caused by its multi-dimensionality is another specific housing characteristic; the fifth particular characteristic v) is the *policy overlay*, meaning that housing is subjected to several institutional regulations imposed by government at various levels; and finally, vi) housing is an economic good where the interlink of spatial *externality effects* and its valorisation plays an important role, that is, rundown houses can generate certain adverse external effects (positive externalities result from well managed housing stock). There are of course, in the market, goods with some of these characteristics, however housing is unique in its combination.

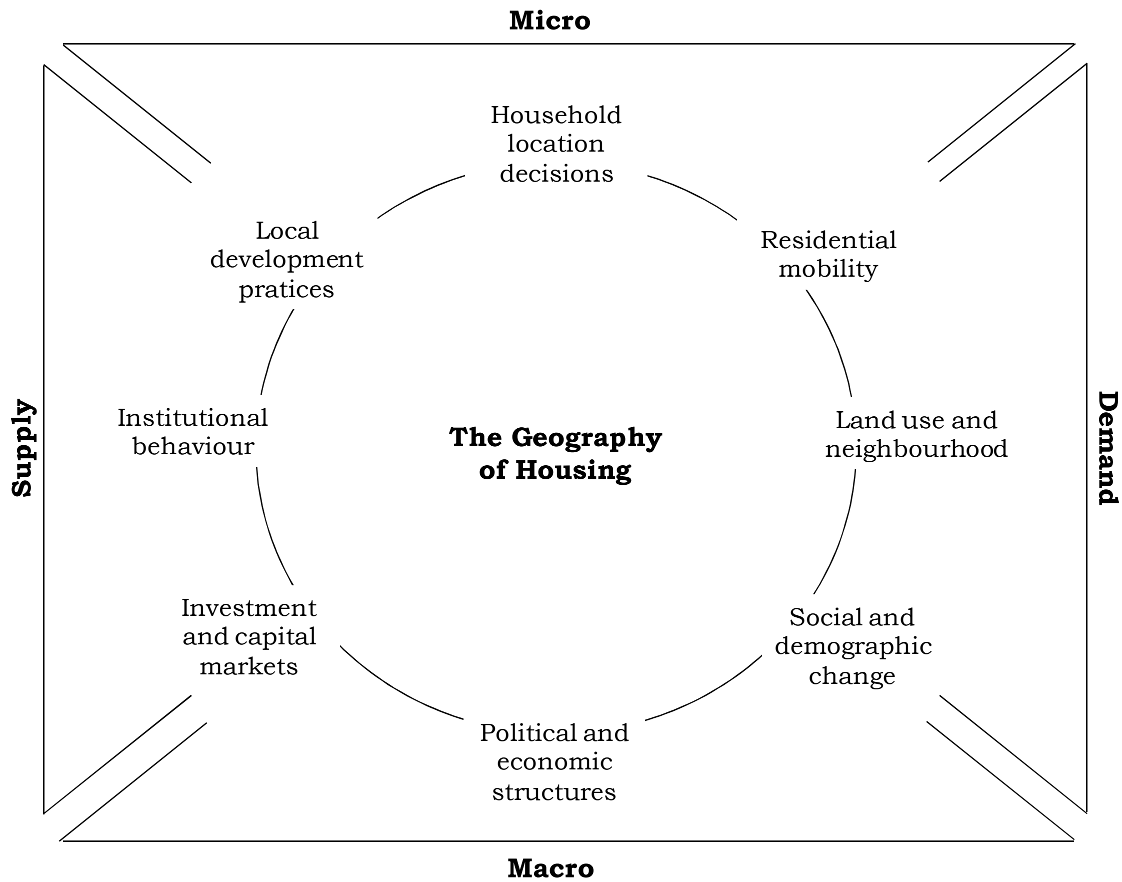


Figure 14 – The literature context for studies of the geography of housing

Source: Bourne (1981, p.10)

Figure 14 represents the various research areas that directly have contributed to an understanding of housing in the urban spatial context. These domains vary in terms of scale (macro and micro), in subject matter (demand, supply and policy) and in terms of philosophy and methodology. The study of housing comprises essentially two scales of analysis. As mentioned in the introduction of this thesis, the study of housing at a macro level, where concerns are more global, involves the analysis of the national economy, government, institutions and agencies, demographic changes, migration, social preferences etc.. At a micro level, the focus is on explaining the location decision of the individual, usually involving analysis of residential mobility or housing preferences for several intrinsic or extrinsic housing characteristics. In the case of demand, the most relevant areas range from residential location models and decision-making at local level to social context (values, attitudes) and political structures. On the supply side, the most relevant areas include the study of the national housing market and the different agents involved in it, as well as the various relationships established between them, causing specific patterns of land use and supply conditions of housing at the local level (see Bourne, 1981 for a syntesis of the diverse perspective on what housing is).

The function of housing is to provide more than just shelter, representing social status and position, economic wealth and power and emotional value (aspirations and personal identity) which are in turn influenced by a larger context. More broadly the housing market is responsible for promoting several forms and levels of socialization and segregation, and hence, can be used as an instrument to regulate and distribute the circuit of capital (allocation of scarcity rents related to Marxism theory).

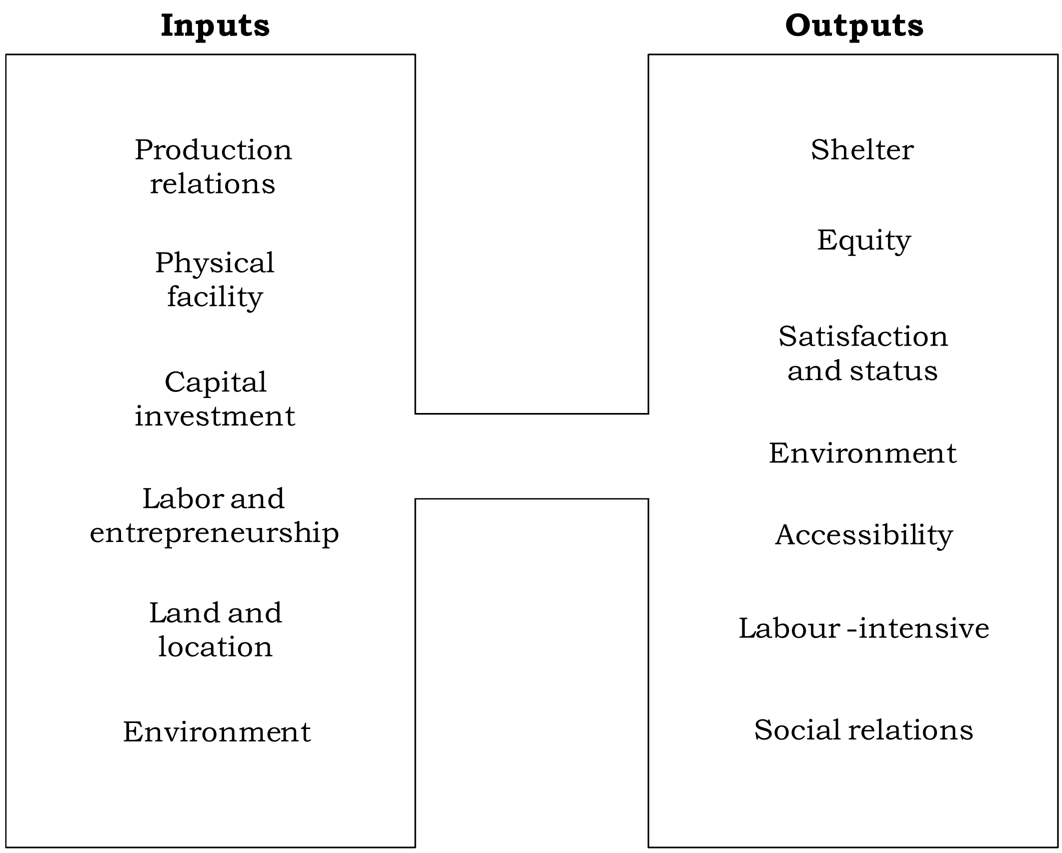


Figure 15 – Input and output representations of housing services

Source: Bourne (1981, p.15)

In *Figure 15* the variety of inputs and outputs representing services provided by housing are summarised in the form of a schematic flow diagram. Depending on the type of housing and on the location, different inputs are required and different services are delivered. According to Bourne (1981) the role of housing market agents (public or private) is to match these inputs and outputs.

Both characteristics and functions of housing are, directly or indirectly, closely associated with the notion of space and its distinct aspects (spatial heterogeneity, spatial dependences and spatial scale). Because qualitative differences in housings (use value), as well as quantitative price differences (transaction value) are not constant over space (where each local provides specific urban amenities), the relation between the housing market and how space is organised, both in terms of spatial patterns and spatial interactions, is a very important issue. These aspects (territorial expression and housing market) are reciprocally interlinked, in the sense that housing outcomes not only reflect, but also shape, social, economic and territorial differences. The acknowledgement that housing stock is impossible to detach from its location and neighbourhood context is clearly emphasised by Bourne (1981), who mentions that the relationships between housing and space are circular and cumulative, that is, housing is both cause and effect.

In short, space has a strict influence on housing and vice versa, underlined in the following five points: i) no two houses are identical, particularly when location is taken into account; ii) each housing location provides specific urban amenities; iii) the value of housing depends on location (at micro and macro scale) and this dependency changes over time; iv) mobility potential depends on housing location and on the accessibility to leisure and work places; and v) the way housing shows status depends on its internal characteristics and on its specific location.

The role of housing in the economy (at macro and micro level) and the importance of space (location) in the housing market, thus, justify a theoretical and methodological approach in which the most important and essential aspects of these interdependences are emphasised.

IV. Methodologies to analyse the housing market

Housing is a heterogeneous product whose price depends on the valuation of several determinants which vary over space and also over time. The large quantity of factors involved and the way they interact are important aspects in the analysis of housing markets, and appropriate methods to describe this phenomenon should be used.

This chapter contains a selective review of recent research on methodologies to support housing policies for long term planning and decision making processes, highlighting the capabilities and limitations in quantitative and qualitative analysis of the housing market, both in terms of spatial (dependence and heterogeneity) and temporal perspectives (time series analysis and foresight). In Marques and Castro (2010) and Marques, Castro *et al.* (Forthcoming) an overview is presented considering the potential application of these techniques to the issue of housing market. In short, these authors stated that, despite the complexity of the subject, the possible changes to the main drivers of the housing market and the assessment of its impacts should not be ignored.

IV.1. Spatial housing market analysis

The methodologies presented in this section are focused on the hedonic pricing model (described in equation 5), since it is flexible and, endowed with adequate spatial structure, useful for understanding many important issues related to housing markets. Under hedonic pricing models, the attributes that determine housing value are intrinsic characteristics, as well as characteristics associated with the external environment of a house and accessibility to the urban system where housing is located. In turn, both the hedonic prices of attributes and the stochastic error affect the overall value of each house and each depend on its location. This necessarily implies that space configures the housing market in a multifaceted and complex way.

Assuming that there is an equilibrium between the housing market supply and demand (and vacancy rates), the housing prices are typically represented in a reduced form by a hedonic pricing model (Maclennan, 1977; Malpezzi, 2003):

$$p = d + Hv + \varepsilon \qquad \text{Eq. 5}$$

Where p is a vector of *m* housing prices (typically in a logarithmic form); v is a vector of hedonic (or shadow) prices, reflecting the valuation of *n* housing attributes, whose choice is contingent on the available information and on the perceptions of individuals that make housing decisions[50]; H is the matrix containing *n* attributes for *m* dwellings, related to intrinsic characteristics and location[51]; d denotes the intercept; and ε is the vector that represents the stochastic error or noise. This linear formulation can be replaced by more complex specifications (see Section IV.1.1.3). Note that H can constitute either a

[50] Is the regression coefficient that corresponds to "(...) *the implicit prices of attributes and are revealed to economic agents from observed prices of houses and the specific amounts of characteristics associated with them*" (Rosen, 1974, p.34). In other words, the coefficients obtained in this regression are the marginal implicit prices of each attribute ($\partial P / \partial H_j$) which represents consumers' willingness to pay for an additional unit of each attribute.

[51] Typically, some attributes would be qualitative, so that they appear in the equation as dummy variables.

set of observable housing characteristics or a combination of observable variables through the use of suitable techniques such as factor analysis (see Sections VII.4.3 and VII.5.4).

In general, the vector υ of hedonic prices is not known. There are alternative methodologies to reveal these prices (described in the Section IV.1.1.2): i) the determination of revealed preferences using econometric estimation of a hedonic price function; ii) the estimation of a utility function, based on questionnaire results that determine the stated preferences between a series of attributes, which are quantified using the willingness to pay or accept the provision of certain attributes (WTP or WTA); iii) an indirect analysis where the purpose is to find relationships between v and any set of known variables or those determined by questionnaire through surveys on quality of life.

Equation 5 can be generalized to cases of spatial heterogeneity, where the vector υ becomes a V matrix whose *p* columns correspond to different housing markets (areas) with specific combinations of hedonic prices (see Section IV.1.2). In this case, it is also likely that the stochastic error reflects spatial autocorrelation, where the vector ε is replaced by the sum $\varepsilon' + \varepsilon_Z$, where ε_z reflects the interaction effect of market z with its surroundings (see Section IV1.3).

In addition to its spatial variation, the attributes of dwellings and their hedonic prices tend to vary over time. It is commonly assumed that the transaction price of a house at any specific period of time is determined not only by its structural and location attributes, but it is also subjected to the influence of price effects from prior sales within its surrounding area. The temporal perspective has not been empirically developed in this work, but it should be mentioned that the equation 5 can also be generalized to a dynamic context in the temporal domain. The corresponding literature on time series analysis and foresight techniques is vast, although generally not explicitly specialised to the housing market (this topic is developed in more detail in the Section IV.1.3).

Finally, it may be noted that, if the attribute matrix H or the vector of hedonic prices v (or both) possessed spatial and temporal patterns, they could have the potential to be highly descriptive and interpretative. The taxonomies associated with numerical techniques for factor analysis also have a vast potential for use.

Following this brief methodological presentation, methods and techniques useful for analysing the housing market are examined in greater detail in this chapter, which is organized as follows. First, a standard situation is considered, where the stochastic error is completely idiosyncratic, is independent of the attributes and location, and has white noise characteristics. Correspondingly, the hedonic prices are also constant throughout the considered territory (Section IV.1.1). In conceptual terms, this model is not endowed with spatial structure and is therefore highly unrealistic; however, this simple model serves as a reference case for further special analysis. Next, spatial heterogeneity is considered, where the shadow prices are allowed to vary over the territory (Section IV.1.2). Additionally, a model with spatial dependence is adopted, where the stochastic error is spatially related to neighbouring, and potentially distant locations (Section IV.1.3). Finally, methods to analyse temporal housing phenomena are presented in Section IV.2, both in terms of time series and foresight analysis.

IV.1.1. Estimation of hedonic prices under conditions of spatial homogeneity

In a housing market study, where the objective is to examine the determinants of property value, and where spatial structure is absent (neither spatial heterogeneity nor spatial dependence), three challenges are generally presented: i) the selection of attributes relevant for explaining house prices, that is, the definition of matrix H; ii) the determination of the influence of each attribute in the explanation of the price of housing, that is, the estimation of the vector υ; and finally iii) the identification of the functional form that best describes the relationships between the explanatory variables H and the housing price p.

IV.1.1.1. Selection of attributes

For the selection of attributes, a set of housing characteristics is considered, combining physical (structural features of buildings and lots) and location attributes (proximity to central locations and accessibility to goods and services) that may explain the property value of a dwelling.

There is a vast body of literature that uses the hedonic prices methodology. A hedonic price model gives information to explain the relationship between the observed price of goods and its characteristics, based on the hypothesis that goods are evaluated for their utility-bearing attributes or characteristics (Rosen, 1974). The most common application of the hedonic price approach is related to the willingness to pay for housing. Typically a measure of housing price is regressed against all housing attributes, using the best-fitting functional form, where the price of a house (dependent variable) depends on a wide range of attributes (independent variables).

There are literally hundreds of possible housing attributes that can be included as independent variables on the right hand side of the hedonic equation 5 (Malpezzi, 2003). Even the choice of the dependent variables to include on the left hand side of the hedonic models is not obvious. The alternative is choosing rent or value of the housing unit. As quoted by Malpezzi (2003) there is a confusion regarding the use of the term *housing price*, loosely used as a synonym for housing value, as a dependent variable in the hedonic model. Several measures of value can be used, combining: i) the owner or tenant estimates of the value of the unit, where this utility payment can be expressed in rent or sale; and ii) expressed in absolute (currency monetary unit) or relative terms (currency monetary unit per area).

Regarding the independent variables in general, and according to the literature (Malpezzi, 2003), they can be grouped into two main categories: i) structural characteristics of the property (intrinsic attributes) and ii) location or neighbourhood characteristics of the property (extrinsic attributes). The structural characteristics are the specific description of the dwelling itself, such as, housing dimension and additional facilities or special housing features. The location characteristics are the bundle of goods (amenities) that are purchased along with the physical characteristics. These variables include socio-economic characteristics of people living near the property, public services provided by the local authorities in the neighbourhood of the property, or the effect of other objects in the environment, etc..

According to Stull's (1975) classification (and Malpezzi, 2003), the housing price is a function of four categories. The first group is the *Physical housing attributes* (F). All attributes related to the physical structure of a house itself and its lot are included in F. According to Bajic (1985) these types of attributes are considered the primary contributors to its economic value, because they provide the greatest utility to the owners and are thought to be more tangible and precisely evaluated than other housing characteristics. Grether and Mieszkowski (1974) also provided evidence that the physical attributes of housing are the most critical factors in determining residential housing price. Relevant factors within this group might include type of housing, lot and dwelling size, number of rooms, living rooms, bedrooms, bathroom, garages, age of the building, conservation of the house, and existence of luxury items (see *Figure 16*). Typically housing value increases with the number of bedrooms, living area, total area and number of garages; and decreases with the age of the dwelling.

The second category considered by Stull (1975) is *Environmental and neighbourhood attributes* (E). This group represents all the attributes related to the surrounding of the house such as neighbourhood characteristics and socio-economic aspects of residents. The literature provides several examples (see next section) where this category of attributes is presented, usually related to environmental amenities, creating both positive (that increase the property value) and negative externalities (that decrease the property value). Included in the first bundle are: scenic views, safety, number and area of green spaces; and in the second the disamenities effects, that include variables such as crime rate, environmental pollution and environmental risks (noise, air or water pollution levels and powers lines).

The third category is defined as *Proximity to locational amenities' attributes* (L). It corresponds to the proximity and accessibility of the property (measured by distance, travel time or costs) to a specific facility or land use, whether desirable or undesirable. Accessibilities increase property value for the reason that they decrease transportation costs (Forrest, Glen *et al.*, 1996). However, accessibility might remove locational advantages when there is noise, air pollution, and traffic congestion in nearby transport routes or higher crime rates in nearby parks and recreation areas (Sanchez, 1993). Relevant amenities might be included (depending on the purpose of the study), considering proximity to workplaces, schools, shops, parks, central business district (CBD), train station, bus stops, major highways, etc..

The last category is *Public service characteristics' attributes* (S). This bundle of attributes includes property tax rates and number and quality of public services. There has been little research involving such variables. The main reason for this paucity of research, as stated by Goodman (1977) and Hwang (2003), is that many of these variables tend to be homogeneous within a community and for this reason there are no observable differences when examining property values in a single urban area.

Other external and more general aspects might be used in the explanation of the housing price. As mentioned before, housing values vary over time and space and for this reason some other factors are important (Millington, 1994): i) the international economic environment; ii) the national economy and financial situation; iii) the national and local government policies and planning controls; iv) the geographic factors (climatic conditions, topography); v) the communication services and accessibilities to local and regional services and amenities; vi) fashion (with respect to regions, towns, neighbourhoods, or to certain houses types or styles); and vii) the individual features of the property (design and architecture, functionality).

In summary, using the categories of Stull (1975), the matrix H that quantifies the attributes of a dwelling can be decomposed into a set of four sub-matrices, F, E, L and S, and time when this value is available.

$$[H] = \begin{bmatrix} F \\ E \\ L \\ S \end{bmatrix} \qquad \text{Eq. 6}$$

The main attributes that can affect the price of a dwelling are represented, in more detail, in *Figure 16*.

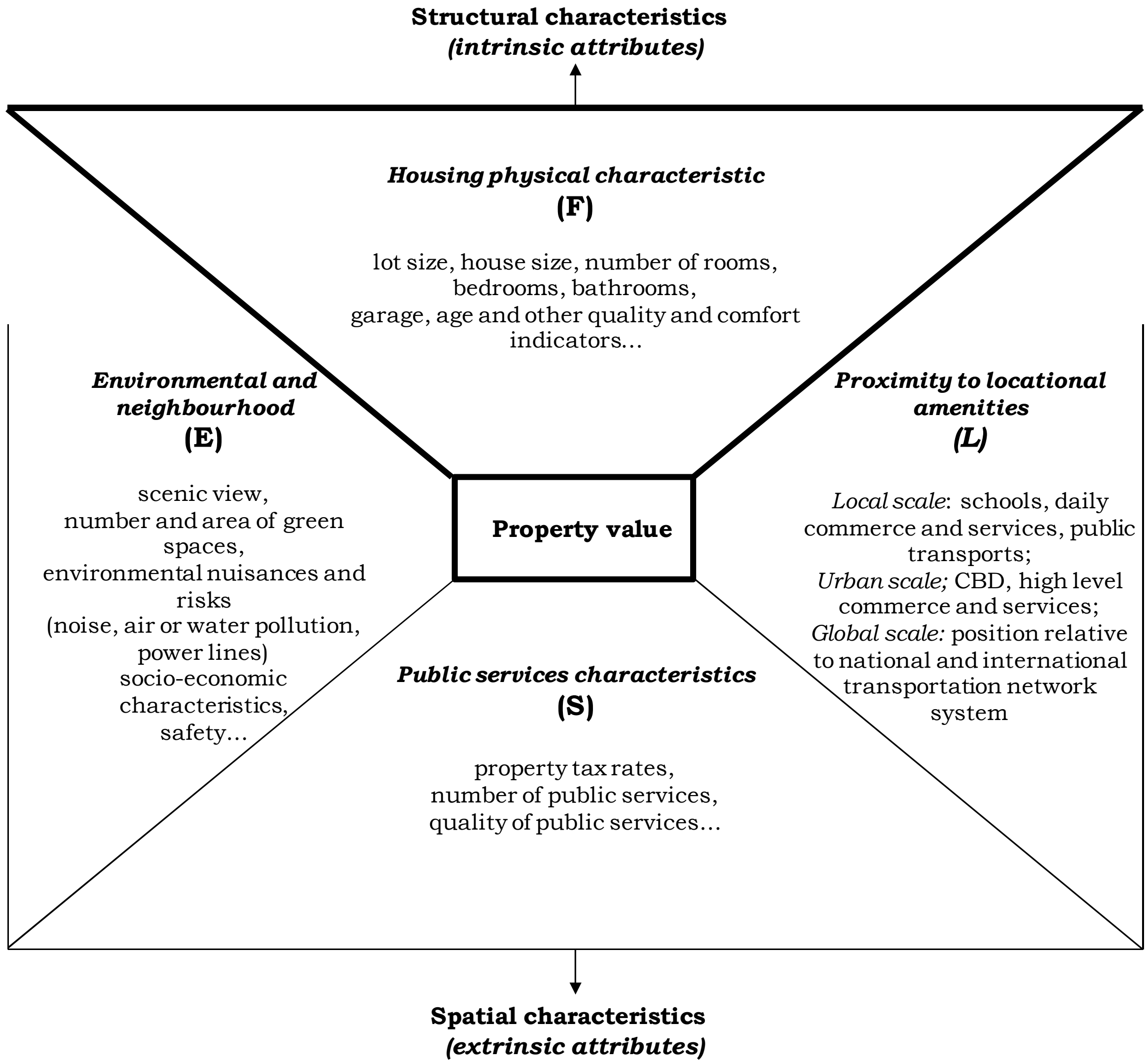


Figure 16 - Physical and locational attributes that could affect the property value

The list of attributes presented in *Figure 16* is not exhaustive, and should be considered as indicative, since it does no guide exists in the literature concerning which variables are important or imperative to include in hedonic analysis[52]. Therefore, the choice depends on several factors with the purpose of the study and the spatial scale used being just some examples. Nevertheless, Malpezzi (2003) argues that a complete data set in terms of physical characteristics should be included: i) rooms in the aggregate, by type (bedrooms, bathrooms, etc.); ii)

[52] As mentioned before, there is a wide range of physical housing attributes that can be included in the analysis. The experience shows that when a real estate agency (or other institutional organization) expands the type of indicators to be collected a huge amount of missing values appear, and only a few parameters are complete and are useful.

floor area of the unit; iii) structure type (single family, attached or detached)[53]; iv) type of heating and cooling systems; v) age of the unit; vi) major categories of structural materials, and quality of finish; and vii) other structural features, such as presence of basements, fireplaces, garages, etc. In terms of location, the characteristics should include: i) neighbourhood variables (an overall neighbourhood rating, quality of schools, socioeconomic characteristics of the neighbourhood); ii) distance variables (to the central business district, to sub-centres of employment, to shopping areas, to schools and to other important amenities). Date of data collection is also an important parameter that should be included in the analysis[54].

The inclusion of relevant variables is an important issue. It is essential to capture all explanatory variables that can contribute to a better understanding of the housing price variability. Many problems may occur: on one hand i) if relevant variables are excluded (omitted variables) the coefficients of the included variables will be biased and inconsistent, unless those variables are not correlated with the ones included, on the other hand ii) if irrelevant variables are included in the model and are correlated with the relevant ones (multicollinearity), the standard error of the regression coefficients will be higher. Multicollinearity is an important issue in multiple regression analysis. It exists whenever an independent variable is highly correlated with other independent variables (Morton, 1977; Hair, Black *et al.*, 2010). However, if the goal is simply to predict the value of the dependent variable from a set of independent variables this is not a problem. In this case, the predictions will still be accurate, and the overall R^2 (or adjusted R^2)[55] quantifies how well the independent variables explained the dependent variable. But, if the goal is to calculate and understand how the various independent variables impact the dependent variable, then multicollinearity will induce a biased understanding of their individual effects. For example, a high level of correlation may exist between the area of the dwelling and the number of rooms or between the number of bedrooms and bathrooms. The same explanation can be made for locational attributes (for example, administrative services and commerce are often located adjacent to one

[53] If multifamily it should identify the number of units in the structure and number of floors.
[54] Hocking (1976), Amemiya (1980) and Leamer (1978) are among useful guides to the actual selection of variables when theory provides little guide, *cited in* (Malpezzi, 2003).
[55] The adjusted R-squared value from multiple regression (R^2 adj) has the same meaning of the R-squared, although provides a more conservative estimate of explained variance, taking into account the number of independent variables associated with the regression equation.

another). So, if many intrinsic or extrinsic attributes occur in close proximity, it may be difficult to distinguish their individual effects.

The multicollinearity effect is detected by the computation of the correlation matrix. Loomis *et al.* (1997) suggested that any correlation higher than approximately 0.8 indicates collinearity. Another method to indentify the high correlation between independent variables is through the variance inflator factor (VIF) index. It provides a scale of measurement of the amount by which the variance of each regression coefficient is increased in relation to a situation in which all of the predictor variables are uncorrelated. Any VIF of 10 or more provides evidence of serious multicollinearity involving the corresponding independent variable (Cohen and Cohen, 1983). Some alternatives can be used to deal with high levels of dependency between explanatory variables. One option is to consider dummy variables for every possible combination of each independent variable. The drawback of this method is that it increases the number of independent variables and, consequently, can face degrees of freedom problems. Other options include the introduction of interaction terms between related variables or the use of principal component analysis; see Mark and Goldberg (1988) for a detailed review.

In summary, the formal relation between the market value of a property (usually measured by its sales price) and housing attributes can be written as follows:

$$P(H) = P(H_1, H_2, H_3, H_4, ..., H_i) \quad \text{Eq. 7}$$

Where, $P(H)$ is the observed property value and $H = (H_1, H_2, H_3, H_4, ..., H_i)$ is the bundle of housing attributes.

In other words, the general specification for a hedonic house price equation (equation 5), using Stull's (1975) categories, can be generalised as following:

$$p = f(F; E; L; S; v) + \varepsilon \quad \text{Eq. 8}$$

Where the p is the sale price of the house; F is the bundle of physical housing attributes; E represents the environmental and neighbourhood attributes; L is the proximity to locational amenities attributes; S translates the public services attributes; v is the regression coefficient, and ε is the error term.

IV.1.1.2. Estimation of hedonic prices

For the estimation of weights corresponding to each attribute (hedonic prices), several methods can be used, grouped into three categories: revealed preference methods, stated preference methods and indirect methods.

a) Revealed preferences

The method of revealed preferences determines the consumer's appreciation of the various attributes associated with housing, through observed prices in the market. The analysis of the housing market through revealed preferences is conducted using econometric models.

The basic principle of this approach follows the concept of hedonic prices. It was introduced by Andrew Court in the 1930s, for the automobile industry (Goodman and Thibodeau, 2007a)[56]. Court published the first article on hedonic price, in 1939, in order to assess the automobile price as a function of its different characteristics (such as horsepower, weight, and so on). A later, but still early and influential automobile application, was performed by Griliches (1961). There are some divergent opinions on the origin of the hedonic price approach. Triplet (2006) and Baranzini *et al.* (2008) argue that the origins of this methodology may possibly be found in previous works, for example in Waugh (1928), "(...) *who estimated a price-characteristics function on vegetables*"[57], and in Hass (1922), "(...) *who even earlier estimated land price-location functions*" (Triplett, 2006, p. 91). Wallace (1926) also used the hedonic price method for studying farmland and vegetables prices (Goodman and Thibodeau, 1998).

The hedonic methods are based on the principle that goods are not homogeneous and differ in number of attributes or characteristics. In its most basic form, hedonic prices result from a functional relationship between the price p of a heterogeneous good (e.g. a house) and its quality characteristics, defined by a vector H as expressed in the equation 5. Where $\boldsymbol{v}$ corresponds to the implicit price of each attribute and ε is a non-explained part of the price.

---

[56] For more detail on the history of hedonic prices principles and its applications see Goodman (1988) and Colwell and Dilmore (1999), among others.

[57] Frederick Waugh has regressed the price of different types of asparagus on their colour, diameter and homogeneity, with the goal of helping farmers in producing the quality demanded by the market; in Baranzini, Ramirez *et al.*, (2008).

As mentioned before, hedonic price methods, originally, were mainly used to describe the prices of non-spatial goods, such as automobiles, tires, and refrigerators. Later, in the 1970s, Griliches (1971) and Rosen (1974) improved hedonic modelling techniques for housing markets. These authors began a rich theoretical and empirical literature exploring the role of housing attributes in consumer decision making (Baranzini, Ramirez *et al.*, 2008).

The implementation of Rosen's hedonic framework assumes a perfectly competitive market, with several sellers and buyers, and that a variety of Z house attributes are available[58] "(...) *represented by a vector of coordinates* $z = (z_l, z_2, ..., z_n)$*, with* z_j *measuring the amount of the* j^{th} *characteristic contained in each good"*. Rosen's hedonic model considers that "(...) *both consumers and producers base their locational and quantity decisions on maximizing behaviour, and equilibrium prices are determined so that buyers and sellers are perfectly matched"* (Rosen, 1974, p.35). The market clearing function is created by an interaction between bid functions of buyers[59] and offer functions of sellers[60], where products are *"(…) completely described by numerical values of z and offer buyers distinct packages of characteristics (…) furthermore, existence of product differentiation implies that a wide variety of alternative packages are available"* (Rosen, 1974, p.35)

The theory of hedonic prices assumes that a good, in this case housing, should not be seen as a good but rather as a bundle of characteristics that match the household's utility function maximum. For this reason, many hedonic studies cite Lancaster's work. The classic paper by Kelvin Lancaster (1966) contributed to the development of a sophisticated branch of microeconomic theory in which utility is generated, not by a good per se, but by the characteristics of the goods (Malpezzi, 2003). More detailed reviews of the hedonic method and common estimation concerns in hedonic applications can be found in Palmquist *et al.* (2005) and Taylor (2008).

Since Rosen and Lancaster's seminal contributions, numerous empirical studies have used hedonic methods to evaluate the determinants of housing prices. The importance of this methodology is justified by its application in a variety of fields and for different purposes, namely:

---

[58] Instead of using H Rosen (1974) uses the notation Z.

[59] Reflects buyers' willingness to pay for an attribute of interest, under certain restrictions (e.g., their income and tastes).

[60] Reflects sellers' acceptable minimum unit prices for forsaking a bundle of housing attributes, taking into account producer costs and benefits.

- *Environmental issues*

Since the levels of environmental quality and risks are implicitly incorporated into the land market, the hedonic models have been used extensively to estimate willingness to pay for specific environmental amenities or improvements, that is, to analyse the variation of the value of a dwelling when located near a good or bad environmental amenity. It should be noted that most environmental goods are not traded on markets, so their evaluation occurs by revealed preferences methods, such as hedonic methods (Palmquist, 2002).

The first environmental application of hedonic techniques was concerned with air quality (Ridker and Henning, 1967), but since then many others studies have evaluated the impacts of several specific environmental externalities, both positive and negative. The first group includes studies that are focused on the assessment of hedonic values, derived from pleasant landscape, clean air, quiet and screening, green spaces and urban watersheds, as well as recreational activities. Tyrvainen (1997), Earl *et al.* (1998) and Acharya *et al.* (2001) provide evidence that depending on the particular view and location, willingness to pay for a good outdoor environment, with green space provision, proximity to parks, and views of green space and water, is quite high. In the second group, studies are included that use the hedonic price method to determine the monetary compensation for exposure to environmental disamenities, such as, water and noise pollution or distances from other locally undesired land uses. High power electric networks, waste incinerators and hazardous waste landfills are some infrastructures that are usually analysed [see, for example, Gayer *et al.* (2002) Deaton *et al.* (2004)]. These authors concluded that in situations where hazardous waste sites and heavy industrial activity are spatially correlated, urban residents are likely to be confronted with a portfolio of disamenities. An extended review of existing studies that have used the house price hedonic technique to estimate the prices that consumers are willing to pay or to accept (WTP or WTA) for environmental goods such as air quality, water quality, and distance from toxic or potentially toxic sites can be found in Boyle *et al.* (2001) and Baranzini (2008).

- *Urban infrastructures and public service access issues*

Urban theory shows that urban infrastructures and accessibilities play an important role in house price determination. The literature that seeks to investigate the impact of the construction of new infrastructures on nearby residential property values examines both the resulting improvement of spatial accessibility to the city (Adair, McGreal *et al.*, 2000; Mikelbank, 2005; Mathur, 2008), and the potential negative effects caused by the physical barriers created, or other inconveniences, such as, noise and congestion during the construction phase or even after it has been concluded (Bae, Jun *et al.*, 2003; Lin and Hwang, 2004). The hedonic price techniques are also well suited to obtain the willingness to pay for a more central location related to accessibilities or the presence of education, post office, fire and police protection facilities (e.g., Brasington, 2003).

- *Neighbourhood specific characteristics' issues*

Neighbourhood factors are also strong determinants of housing values. Substantial efforts have been made in the understanding of neighbourhood impacts on housing prices, specially the analysis of effect of racial, ethnic and socioeconomic differences, or other social groups, in housing prices. Follain and Malpezzi (1981), Galster (1992), Kiel and Zabel (1996) and Myers (2004) are among many contributions to this strand of literature. Even if the main objective is not the understanding of the impacts of each specific neighbourhood aspect, it is crucial, when testing for price differentials, to include indicators on the household and on the general submarket information, such as: socio-economic characteristics (e.g. income and unemployment), land-use or zoning characteristics (e.g. agricultural, residential or commercial land use) to guarantee lack of bias.

- *Identification and modelling of housing submarket issues*

An old issue in real estate economics is the existence and measurement of housing submarkets (Nelson, 2008). For this purpose hedonic prices are widely used with the main aim of estimating segmentation price differentials. Several authors have recognised that the importance of each attribute in housing value may vary considerably within a particularly housing market (Maclennan, Munro *et al.*, 1989; Maclennan and Tu, 1996). It is consensual that planning administrative areas and historically recognized neighbourhoods may not define boundaries that represent distinct levels of housing quality. Considering the housing markets as geographic areas, where the per unit price of housing services is constant, the hedonic prices approach has been applied to determine if the estimated regression coefficients vary by geographical location (Goodman and Thibodeau, 1998; Bourassa, Hamelink *et al.*, 1999; Bourassa, Hoesli *et al.*, 2003; Bourassa, Cantoni *et al.*, 2007). This issue will be developed in more detail in the Section IV.1.2.

- *House price indexes and housing quality issues*

Hedonic analysis has been widely used to construct house price indices. According to Malpezzi (2003), this field of research has been one of the first, and still the most important application of hedonic models, whether to develop time series, place-to-place, or panel data price indexes. Bailey *et al.* (1963), Case *et al.* (1991), Quigley (1995), Malpezzi (1998) and Bourassa *et al.* (2006) are studies that can be cited[61], among many other examples.

In addition, hedonic analysis has been widely used in other perspectives: i) to understand how selling conditions influence property prices, for instance, to model the impact of bargaining processes on house price (Song, 1995); ii) to compare the performance of hedonic models with estimates provided by experts or professionals[62] (Dodgson and Topham, 1990); and iii) to analyse urban sprawl, in order to understand peri-urbanisation, that is, the move of urban dwellers to near-city locations [see, e.g., Cavailhès, Peeters *et al.* (2004)]. For more information see Baranzini, Ramirez *et al.*, (2008) that gives a review of the general framework upon which hedonic analysis is built, and provides an overview of some implementation issues and recent developments.

In the Portuguese context, hedonic pricing approaches, besides the papers developed in the scope of this PhD research (Marques and Castro, 2007; Marques, Castro *et al.*, 2009; Castro, Marques *et al.*, 2011), some studies have been applied in different urban contexts and scales: Pinho (1992), Moreira (2000), Guimarães (2004), Catalão (2010) and Valente (2010), Reis (2011), among others. However, all these works by other authors only consider the standard hedonic method.

Despite the multiplicity of applications, there is a vast literature that considers hedonic models ineffective, since they are subject to a number of restrictive assumptions (Anselin, 1988; Anselin and Florax, 1995; LeSage and

---

[61] Pollakowski and Wachter (1991) and Kiel and Zabel (1997) present a number of indexes estimated from hedonic regressions, e.g., "simple hedonic"; "expanded simple hedonic"; "complex hedonic" and "expanded complex hedonic", but also based on other methods, such as repeat valuation models, and hybrid-models, based both on owner-provided values and sales prices.
[62] In order to study the effects of various market imperfections on house price.

Pace, 2009). Apart from the uncertainty of the choice of explanatory variables and functional specification of the model, the robustness of the traditional econometric estimation depends on restrictive assumptions that do not necessarily hold in a housing market context (Marques and Castro, 2007). This issue will be discussed further, later in the Section IV.1.3.

b) **Stated preferences**

As stated before, revealed preference methods are based on real markets for private commodities, where consumers reveal their willingness to pay by purchasing (or refusing to purchase) more or less units at different relative prices, other things remaining constant. In the case of the stated preferences method, a simulated market is constructed such that the fictitious consumers will accurately and truthfully state their willingness to pay for additional units of the commodity (Nelson, 2008). This second family of methods uses direct surveys of residents to assess their preferences for various attributes that characterize a dwelling. Also called contingent valuation methods, this type of approach simulates market scenarios rather than actual situations; therefore, they record stated or intended measures of willingness to pay under hypothetical conditions.

These preferences can be obtained by two basic methods: a holistic approach and an analytical approach.

The holistic approach is based on a simulation of the market that aims to learn, from the respondents, the willingness to pay for a given set of hypothetical dwellings, each corresponding to a different set of implicit attributes (H). In this case, the selection of attributes must be carefully undertaken to achieve a balance between the hypothetical number of dwellings, whose evaluation is queried from respondents, and the richness of information that can be extracted, both in terms of variety of attributes considered and correspondingly hedonic prices. One method for increasing the number of attributes without making the questionnaire too long and difficult to survey is to ask different people to assess housing that combining different sets of attributes. This, however, implies the assumption of homogeneity within the sample, which can sometimes be tenuous.

The effectiveness of the holistic approach depends on the ability of respondents to grasp the variation of attributes and assess the impact of this variation on price. Knowing if the results of impressionist evaluations of a large number of respondents become reliable is akin to evaluation of whether the law of large numbers is applicable, that is, whether the impressionist ratings are, or not, affected by systematic biases. Added to this specific problem is an issue common to all methods of determination of stated preferences: the commitment and sincerity of the respondents, and therefore, the reliability of their responses. See in Castro, Marques *et al.* (Forthcoming-a) an empirical application of this approach.

The analytical evaluation consists in obtaining directly, from each respondent, the hedonic price vector v. The idea is to ask about the willingness to pay for hypothetical housing, all other things being equal, varying gradually the quantity of an attribute or a small set of attributes. This alternative assumes that the isolated assessment of each attribute produces better estimates than the impressionistic view.

This approach has received much criticism based on an argument that is an extremely hypothetical exercise. Mundy *et al.* (1998) compared these methods to a monopoly money game arguing that there is no downside to overestimating willingness to pay in a hypothetical situation, and it remains unknown what portion of residents would be willing to pay using their own, real money.

c) **Indirect Methods**

Individual preferences can also be assessed by indirect methods that relate hedonic prices to the balance between the financial affordability of individuals for housing prices, and the ability of different attributes to satisfy their objective needs or subjective tastes. In turn, the information required for using these methods can be acquired through surveys, or in complicated cases through a combination of questionnaires allowing the analysis of different factors, such as accessibility to the urban centre and health effects caused by certain pollutants, etc..

Being focused not specifically on direct assessment of willingness to pay for a specific good, these methods provide results that do not correspond directly to demand, but can nevertheless be extremely useful in urban studies. The concept of quality of life (QoL) is particularly useful in this context, to evaluate a range of psychological and physiological factors that are responsible for transmitting the feelings of (or lack of) satisfaction caused by social and physical environment that surrounds residents (Biagi, Lambiri *et al.*, 2006). The concept of QoL is attracting particular attention within the academic community and policy makers and, as noted in Baker (2003, p.734), "(...) *the need to improve the quality of life is now a very common requirement* (...) *often arise in our lexicon and rhetoric*". The same argument is reinforced by Friedman (1997, p.12) when the author states that "(...) *the quality of life is a mundane concept that is daily in on people's minds daily*". However, the complexity and multidisciplinarity associated with the concept of QoL, serves to restrict its use as an effective analytical tool in urban planning and housing policy (Gomes, Marques *et al.*, 2008; Belbute, Marques *et al.*, 2009).

The choice of methodology for estimating the weight of each attribute, particularly between revealed and stated preference methods, has been widely discussed and is far from consensual. For this reason many authors choose to develop hybrid approaches in order to benefit from the relative advantages of each method (Timmermans, Molin *et al.*, 1994; Whitehead, Pattanayak *et al.*, 2008).

The following *Table 4* presents a summary description of the revealed and stated preferences methods, focusing on their main advantages and disadvantages.

Table 4 - Comparison between revealed and stated preferences - Summary description

Revealed preferences	**Stated preferences**
Based on actual market behaviour	Based on hypothetical scenarios
Cognitive congruence with behaviour of market demand	Risk of incongruence with market behaviour
Difficulty of measuring intangible attributes	Intangible attributes more easily incorporated
It is not possible to directly predict the answer to new alternatives	It is possible to assess new alternatives
Correlated attributes (problem avoided with the use of Factor Analysis)	Attributes not correlated (the design of the survey take this into account)
Limited number of attributes (restricted by data availability)	Unlimited number of attributes (the limit is the 'grasp' ability of the respondents)
A major source of errors is poor measurement of the attributes	The primary source of errors is poor understanding of the attributes by respondents

Source: based on Econometrics Laboratory of University of California at Berkeley (2000)

IV.1.1.3. Functional specification of the model

The functional form refers to the method in which the dependent variable (housing price) is correlated with the independent ones (explanatory attributes of a dwelling). Theory provides no guidance for the selection of a appropriate functional form for the hedonic housing price model, that is, hedonic literature does not specify *a priori* an ideal model to estimate the relationship between the selling price and the characteristics of the housing (Goodman, 1978; Malpezzi, 2003). Several different functional forms can be used, including the additive (linear form) and the multiplicative (logarithmic form) models. These hedonic functions mainly differ in whether the variable itself or its logarithm appears in the equation.

The basic difference between linear and non-linear form is that, the former requires independence on the explanatory variables chosen[63], while in the latter, implicit prices of characteristics are dependent upon the levels of other characteristics.

A more detailed overview of the functional forms most used in the hedonic price models is presented next.

a) Linear model

This is the simplest functional form for a hedonic equation. It is given by the following equation:

$$P_i = d + \sum_{j=1}^{J} \nu_j H_{ij} + \varepsilon_i \qquad \text{Eq. 9}$$

The hedonic prices are given by:

$$\frac{\partial P_i}{\partial H_{ij}} = \nu_j \qquad \text{Eq. 10}$$

Where, H_{ij} are the housing attributes. The υ_j (j = 1, ..., J) terms are unknown coefficients, or parameters, and may be interpreted as the marginal price contribution to a house's value (incremental increase or decrease). Those coefficients are constant for any house *i* and independent of the value of H_{ij} (are estimated from the sample data). The intercept, d, can be interpreted as the mean of the housing price when all H_{ij} variables are zero. Since a model based on a set of independent variables cannot predict exactly the observed values of P_i, it is necessary to introduce ε_i (denominated as the error or residual or stochastic component).

The linear model implies constant partial effects between house prices and housing characteristics, that is, the willingness to pay for an additional unit remains invariable and contributes the same value to the overall price[64], in other

[63] This assumption is considered unrealistic, especially in housing market.
[64] For example, the addition of one square meter of a room to a property is likely to be of greater value to a house whose existing size is 140 m^2 than it is to a house of 400 m^2. Or, the addition of

words, housing attributes are not subject to the rule of decreasing marginal utility. One way to avoid this problem without losing the advantages of linearity is to transform to logarithmic form all, or at least, some of the hedonic attributes (for example, age of buildings, floor area, etc.). There are many reasons to assume non-linearity in the relationship between the price and the variables that describe the attributes of housing (Malpezzi, 2003).

Despite this restriction, linear models are still in use because of the direct meaning of interpretation of the coefficients.

b) Semi-logarithmic (log-linear) and log-log models

A non-linear hedonic function is useful for recovering the underlying structural demand curve from estimates of the hedonic relationship (the reduced form). In the semi-logarithmic form, only the left hand side variable (the property prices) appears as a logarithm, the right hand side (explanatory) variables appear in their own value.

The semi-log approach is characterized by:

$$P_i = d\prod_{j=1}^{J}\exp(v_j H_{ij})\ \varepsilon_i \quad \text{Eq. 11}$$

Or,

$$\ln P_i = d + \sum_{j=1}^{J} v_j H_{ij} + \varepsilon_i \quad \text{Eq. 12}$$

The hedonic prices are given by:

$$\frac{\partial \ln P_i}{\partial H_{ij}} = v_j \quad \text{Eq. 13}$$

The semi-logarithmic equation takes into account the interaction among the independent variables, that is, the multiplicative effects of the attributes. However, because of the form of the exponential function, the dependent variable is more sensitive to variables with a larger variance in their values. Many authors

another bedroom to a house is likely to add more to the price of two bedroom house than it is five bedroom house.

argue that this specification usually fits the data better than does the conventional linear equation (Wooldridge, 1999; Malpezzi, 2003). The semi-log functional form has a number of advantages over the linear, namely: i) it reduces the likelihood of heteroskedasticity, which means that the variance of the unobservable error (conditional on the independent variable) is not homogeneous; ii) the dependent variable in logarithmic form narrows the range of the dependent variable by a significant amount, which makes estimates less sensitive to extreme problem points (or outliers) on the transformed variable; iii) it has two interpretations for the coefficients, not only as the implicit or hedonic price, but also as the approximate percentage change in the rent or values given a unit change in the independent variable[65]; iv) the variation in the value of a particular characteristic depends in part on the house's other characteristics, that is, it allows the values added to vary proportionally with the size and quality of the dwelling; and finally, v) it is possible to build specification flexibility into the right hand side, using the logarithms of some independent variables, using dummy (or indicator) variables (Malpezzi, 2003).

When the attributes are in logarithms, then the implicit prices (υ_j) measure the elasticities of prices with respect to each attribute, indicating the percentage of the price P_i that increases if the j^{th} characteristic H_{ij} changes by one percent. This model can be interpreted as a partial elasticity and is called double logarithmic, double-log or log-log transformation. Several other important advantages motivated Follain and Malpezzi (1980) to recommend the semi-log form. When a logarithmic transformation is used, the coefficients represent the percentage of increase or decrease in property value associated with one-unit change in the level of the independent variable.

Some authors, according to Malpezzi (2003), have recommended more flexible forms than the model specifications presented above, for example, a trans-log functional form.

---

[65] Nevertheless, Malpezzi (2003) was aware of the fact that the percentage interpretation is an approximation and it is not necessarily accurate for dummy variables. The author explained the potential bias based on the following example: "(...) *if the coefficient of a variable representing central air conditioning is 0,219, then adding it to a structure adds about 22 percent to its value or its rent*" Malpezzi (2003, p.21-22). However, the suitable value should be 24%, using the method given by Halvorsen and Palmquist (1980) *cit in* Malpezzi (2003). These authors show that a much better approximation of the percentage change is given by e^b-1, where b is the estimated coefficient and e is the base of natural logarithms [exp (0,219) – 1 = 24%].

c) Trans-log functional form

The trans-log approach is characterised by:

$$\ln P_i = d + \sum_{j=1}^{J} v_j \ln H_{ij} + \frac{1}{2}\sum_{j=1}^{J}\sum_{k=1}^{K} \gamma_{jk} \ln H_{ji} \ln H_{ki} + \varepsilon_i \qquad \text{Eq. 14}$$

Continuing with the potential functional form of a hedonic model, Maplezzi's work (2003) presented an even more general and flexible class of functions (compared to the previous models: linear, logarithmic and translog functions), where synergy effects are considered. These flexible forms are carefully developed by Box and Cox (1964).

d) Box-Cox transformation

Basically, the Box-Cox form provides a mean of generalizing the linear model and provides a statistical basis for choosing among different functional forms (Box and Cox, 1964).

According to Freeman (1993), Goodman (1978) was the first to apply the Box-Cox transformation of the dependent variable in a hedonic study. Goodman's approach was still somewhat limited in possibilities, since he did not consider alternative forms for the independent variables. Considering both transformation of the dependent variable and different transformations of each independent variable, the more general Box-Cox transformation, if linear, is often estimated according to:

$$P_i^{(\theta)} = d + \sum_{j=1}^{J} v_j H_{ij}^{(\phi)} + \varepsilon_i \qquad \text{Eq. 15}$$

Or if quadratic:

$$P_i^{(\theta)} = d + \sum_{j=1}^{J} v_j H_{ij}^{(\phi)} + \frac{1}{2}\sum_{j=1}^{J}\sum_{k=1}^{K} \gamma_{jk} H_{ij}^{(\phi)} H_{ki}^{(\phi)} + \varepsilon_i \qquad \text{Eq. 16}$$

Where the price P_i of a property is transformed through the parameter θ to:

$$P^{(\theta)} = \begin{cases} \dfrac{P^{\theta} - 1}{\theta} & (\theta \neq 0) \\ \ln P & (\theta = 0) \end{cases} \qquad \text{Eq. 17}$$

And the exogenous variables $H_{ji}^{(\theta)}$ with the parameter to ϕ[66]:

$$H^{(\phi)} = \begin{cases} \dfrac{H^{\phi} - 1}{\phi} & (\phi \neq 0) \\ \ln H & (\phi = 0) \end{cases} \qquad \text{Eq. 18}$$

Where:

P_i = Price of the house;

H_{ij} = Housing characteristic *j* of the house *i*;

Such a form is quite flexible, with parameters θ and ϕ limiting the functional form:

i) If θ and ϕ are both 1 and γ_{jk} are all identically zero, than the model is reduced to a linear form.

ii) If θ and ϕ approach zero and γ_{jk} are all identically zero, than the Box-Cox form is reduced to a logarithmic model.

iii) If θ and ϕ approach zero and but some γ_{jk} are non-zero, the Box-Cox form becomes the trans-logarithmic model.

This specification is suggested by several 1980s studies (for e.g., Cropper, Leland *et al.*, 1988), however, Cassel *et al.* (1985) note several drawbacks in using the Box-Cox model. According to the cited authors the large number of coefficients to be estimated may reduce the accuracy of any single coefficient. Another problem described is that the non-linear transformation results in complex and weighty estimates of slopes and elasticities. And finally, this functional form is not suited to any data set containing negative numbers.

As has been shown, many alternative models can be used in empirical research. The most frequent forms used in the hedonic model's literature are the linear and semi-log forms. More complex transformations are generally avoided, due to the difficulty in deducting inferences and interpretation of results. Nevertheless, the assumption that any of these specifications correctly describes the reality analysed is not *a priori* justified. Cropper (1988) carried out a large number of simulations to assess the sensitivity of the results to functional

66 Dummy variables cannot be transformed, as these can only assume the values 0 or 1, see Cassel and Mendelsohn (1985, p. 138).

specification and concluded that, when there are omitted variables[67], simpler functional forms such as linear or semi-log perform better than more complex forms. The functional form of the hedonic function is entirely an empirical matter and may be answered by the data itself. The coefficients of these equations are not comparable. However, the use of the determination coefficient (R^2), that is the part of the total variance that can be explained, allows some insight. According to Pace (1993) an ideal statistical method for real estate valuation would possess low specification error, robustness against outlying observations, superior post-sample predictive accuracy and known statistical properties.

It is in principle possible to adopt a generic non-linear specification and find, by econometric estimation (for example, polynomial or non-parametric), the best functional form. However, one needs to be wary of the issues relating to interpretation and the potential consequence of losing the intuitive meaning of the concept of hedonic price. Overall, functional form specification should be an object of careful analysis, including adequate consideration of relevant theory.

So far, the hypothesis of spatial homogeneity in the implicit prices has been assumed. The consequences of violation of this hypothesis are discussed in the following section.

IV.1.2. Estimation of hedonic prices in spatial heterogeneity conditions

Spatial homogeneity is a strong assumption in the hedonic housing price context; if not analysed conveniently can be a potential source of specification errors. Spatial heterogeneity occurs when a territorial segmentation exists in the housing market and, therefore, hedonic prices associated with different attributes are not constant over space. For example, it is reasonable to expect that households living in the centre of an urban area may value proximity to central facilities differently from those living on the periphery; likewise the implicit price of an additional bedroom in a leafy neighbourhood in the suburbs is likely to be different from that in the centre. Indeed, understanding such

[67] Or when proxy variables are used in the absence of a measure of the real variable.

variation in implicit prices over space is an important objective of studying the spatial structure of the housing market, by estimating hedonic pricing models.

In the context of the model expressed in equation 5, spatial heterogeneity implies that the vector of regression coefficients remains stable within specified subareas, but varies across different subareas. In other words, spatial heterogeneity is related to a lack of stability over the space, in both choice of explanatory variables H and their influence (v) on housing price (P).

The recognition of housing submarkets and the argument that caution should be exercised when interpreting the results of standard hedonic models has been identified early in the literature (e.g. Rapkin, Winnick *et al.*, 1953; Grigsby, 1963; Straszheim, 1974; 1975; Maclennan, 1977; Quigley, 1979). Despite the argument that housing submarkets should be adopted as a working framework, some ambiguity remains about how to deal with this issue. Watkins (2001) illustrates this difficulty suggesting five reasons to explain the failure of submarket models. The first argument, mentioned by the author, is the difficulty of defining a housing submarket, because a range of meanings can be adopted; second, even if theoretically there is consensus about the definition, there is little consensus about how submarkets should be identified in practice; third, there is a large spatial variability regarding how urban areas are examined, making comparison difficult between studies; the fourth reason is related to the previous one, that is, the variability of the time period from which market data are drawn, which affects the market condition; and finally, the variety of statistical tests used to analyse the existence of submarkets in different studies.

The definition of housing submarket areas has proved a difficult problem[68]. Maclennan *et al.* (1989) affirm this, saying that a housing market in most urban analysis can be considered as *"(...) a simple theoretical construct with no specific form and often it has no qualitative, temporal or spatial dimensions"*. Nevertheless, some definitions of a housing submarket can be found in the literature, which are useful for delimiting the concept.

William G. Grigsby, considered a pioneer on the study of neighbourhood changes, pointed out, in his book published in 1963, "*Housing Markets and Public Policy*", that submarkets are distinctive because houses within them are

---

[68] Sometimes a housing submarket is confused with the notion of neighbourhoods; however, a submarket is more than this, potentially being comprised of several neighbourhoods, across which hedonic prices have no significant differences (Tu, Sun *et al.*, 2007).

viewed as (more or less) perfect substitutes by the households. In this way, two dwellings are elements of the same submarket *"(…) if the degree of substitutability between them is sufficiently great to produce palpable and observable cross-relationships in respect to occupancy, sales, prices and rents, or in other words, whether the units compete with one another as alternatives for the demanders of housing space"* (Grigsby, 1963 p.34). In fact, Grigsby followed the theoretical framework of Rapkin and Winnick (1953) defining housing submarket as "(…) *the physical area within which all dwelling units are linked together in a chain of substitution",* considering that " (…) *every dwelling unit within a local housing market may be considered a substitute for every other unit" (Rapkin, Winnick et al., 1953, p.9-10).* More recently the same concept of substitutability has been followed by Goodman and Thibodeau (2007a, p.4) that considered housing markets *"(…) as geographic areas where the price per unit of housing quantity (defined using some index of housing characteristics) is constant".*

All of these definitions rely on the concepts of substitutability and equilibrium, however, according to Bourassa *et al.* (2003) and Bourassa, Hoesli *et al.*, (2003) the aim of defining submarkets should not be necessarily to define relatively homogeneous submarkets consisting of substitutable dwellings, but rather, to segment the market in a way that allows for more accurate estimates of house values. This argument is supported by the following example: as a market is segmented into smaller and smaller (and more homogeneous) submarkets, the hedonic prices are estimated less precisely due to the inverse relationship between sample size and standard errors. Also, as a market is segmented into more homogeneous submarkets, variability in the hedonic characteristics will decrease and, consequently, some variables will drop out of the equation. Based on this argument, in practice, too much homogeneity may not be a good thing.

In practice, submarkets can be analysed at three levels. The first is considered a macro scale approach, which includes works that adopt national areas (or at least large regions, or states) as the unit of analysis. Linneman (1981) and Struyk (1980) are some works that fall into this category. The second is a meso scale level, dealing with a regional/metropolitan approach, more or less coincident with the labour market[69] and comprises works like Malpezzi *et al.* (1980); Goodman and Thibodeau (1998; 2007a) and Fingleton (2008). Finally, the

[69] Malpezzi (2003) includes the regional level in the first category; however, regarding the European context the regional level is closer to the metropolitan than the national level.

third approach is a micro level analysis, which examines submarkets below the metropolitan area. Several works adopt this level as the unit of analysis to examine submarkets, for example: Kiel and Zabel (1996); Maclennan and Tu, (1996); Bourassa *et al.*, (1999) and Clapp and Wang (2006)[70].

There is a substantial literature that presents appropriate methods for defining housing markets. The common point of these approaches is the idea of finding areas in which the coefficients of hedonic price equations are similar. The question arise of how to analyse this similarity.

The early empirical works on submarkets tended to be segmented into two perspectives: those studies that adopt a supply side determinant, and those that focus on demand side determinants (Goodman and Thibodeau, 1998; 2007a). Determinants that may be included on the supply side are: housing characteristics (this dimension includes structural characteristics of dwellings); and neighbourhood characteristics (e.g., public education, public safety, status or racial discrimination). On the other hand, if the focus is on the demand side, the determinants are based on household incomes or other demographic and socioeconomic characteristics. In this case the identification of distinct subgroups of demand is considered crucial to assess preferences and views of the two previous dimensions.

Thus, all three aspects (structural characteristics, spatial characteristics and subgroups of demand) can be used, separately or interactively, to determine submarkets (Adair, Berry *et al.*, 1996; Maclennan and Tu, 1996). Most recent studies acknowledge that there are both spatial and non-spatial drivers of submarkets and so some form of joint estimation is used (Goodman, 1981; Adair, Berry *et al.*, 1996; Maclennan and Tu, 1996; Leishman, 2009).

The treatment of spatial heterogeneity is standard in the econometrics literature. It is commonly assessed with the use of Chow's F test, which examines whether the structural relationship between the dependent and explanatory variables is subject to some kind of change. Assuming the existence of spatial heterogeneity, there remains the problem of identifying and delimiting the various submarkets. This can be done informally, using *a priori* knowledge of the

---

[70] In this level housing markets are usually segmented by type of location (central city vs. suburb), or by housing quality level (structural characteristics), or by race or income level (distinct group of demand) (Malpezzi, 2003).

geographical area under study, or by employing analytical methods (Nelson, 2008).

Typically pre-existing geographic or administrative boundaries, such as, census track, zip code district, school district or local political jurisdictions, are used to define submarkets even if they are considered an inappropriate way to deal with the problem. The reason for using administrative boundaries in some empirical work is typically because of data constraints rather than because of any belief that they are the most appropriate defining concept of the housing market. Works by Straszheim (1975), Goodman (1981), Goetzmann and Spiegel (1997) and Brasington and Hite (2005) are some examples which used political boundaries to define submarkets. These procedures, based on *a priori* judgement to define submarkets, are subject to *a posteriori* validation. Hedonic regressions are computed separately for each submarket, and F tests then determine whether the resulting reduction in sum of squared residuals is significant. If it is, the submarkets are assumed to be appropriate (Goodman and Thibodeau, 1998). The criticism to this approach is that the housing segmentation is imposed rather than given by the data (modelled).

In spite of the traditional administrative boundaries, other methods can be applied: hedonic equations to identify submarkets, for instance. The principle is to use implicit prices of housing to identify areas with similar characteristics and then aggregate those areas into submarkets. A strategy commonly used is the application of dummy variables to describe each submarket in the hedonic specification, rather than estimating a separate hedonic equilibrium for each submarket (Castro, Marques *et al.*, 2011). Once the submarkets are defined, slope and intercept dummy variables for each submarket should be included in the model. Significant differences in slope across submarkets indicate spatial heterogeneity in implicit prices, in other words, the hedonic prices associated with different attributes vary in whole or in part across the submarkets. Likewise, statistical significance of the intercept dummies indicates unobserved heterogeneity in the fixed effects across different submarkets. The hedonic equation including dummy variables takes the form:

$$p = \sum_{z=1}^{Z} d_Z + H^* v^* + \varepsilon = \sum_{z=1}^{Z} d_Z + \sum_{j=1}^{n} \sum_{z=1}^{Z} v_{jz} h_j d_z + \varepsilon \qquad \text{Eq.19}$$

Here, d_z denote intercept dummies corresponding to each of the Z submarkets (z=1,...,Z), H^* the modified matrix of housing attributes where each characteristic interacts with every submarket dummy, and $\boldsymbol{v}^*$ the corresponding vector of hedonic prices, heterogeneous by submarkets. In matrix notation:

$$H^* = [d_1 H : \quad d_2 H : \quad \dots : \quad d_Z H]$$
$$v^* = (v_{11} \quad \dots \quad v_{n1} v_{12} \quad \dots \quad v_{nZ})' \qquad \text{Eq.20}$$

The main disadvantage of this method is the large number of implicit prices to be estimated, which requires large sample size. However, if there is lack of spatial homogeneity, there is no simpler alternative. To ignore such heterogeneity would result in specification (omitted variable) bias and a serious failure to account for spatial structure.

Examples of the use of hedonic approach to identify space heterogeneity can be found in: Goodman (1981), Maclennan and Tu (1996), Goodman and Thibodeau (Goodman and Thibodeau, 1998), Bourassa, *et al.*, (1999) and Bourassa *et al.* (2007).

It is also frequent the application of principal component analysis (PCA) and cluster analysis (CA), as a complement of hedonic approach to identify local market areas. The approaches using PCA and CA do not depend on *a priori* boundary definition, but rely on the structure of data. Factor scores obtained using principal components are used in cluster analysis to find groups of homogeneous observations, which result in submarkets that do not impose contiguity. These methods require, from the research, three major decisions: the clustering algorithm[71]; the clustering criterion[72] and the dissimilarity measures[73] (this is explored in more detailed in Section VII.3.5).

In these multivariate analyses some other techniques may be included: hierarchical models[74] (Goodman and Thibodeau, 1998; Raudenbush and Bryk,

---

[71] Two clustering algorithms are usually used: K-means and hierarchical clustering. Goodman and Thibodeau (1998) suggest that submarkets should not be imposed (K-means) but specified explicitly using a hierarchical approach.

[72] Four clustering criteria are usually used: single linkage, complete linkage; group average; and Ward's Methods.

[73] Five dissimilarity measures are usually used: binary variables; categorical variables; continuous variables; and mixed variables.

[74] The method, suggested by Goodman and Thibodeau (1998), has been adopted from the education and the evaluation literatures (technique using data for the Carrollton-Farmers Branch Independent School District - CFBISD) and the procedure: *"(...) starts by estimating a hierarchical model for two adjacent school zones. Then, if the coefficient associated with the submarket is significant, those school zones are considered to pertain to different submarkets. If, on the other hand, the coefficients are not significant, the two zones are merged. One by one, each school zone is*

2002; Goodman and Thibodeau, 2003), and mixtures of linear models[75] (Ugarte, Goicoa *et al.*, 2004). Spatial partitions based on socio-economic or environmental characteristics, as shown in Galster (1987), Hårsman *et al.* (1995) and Schnare *et al.* (1976; 1980) are simple methods that can be also used to define submarkets.

The non-parametric spatial statistical methods developed in Clapp *et al.* (2006), (Bhattacharjee and Jensen-Butler, 2005) and Bhattacharjee and Holly (2010a; b) are more sophisticated approaches to delineate submarkets. These methods consider residual spatial autocorrelation and are explored in the Sections VII.4.6 and VII.5.7.

The previous methodologies focused on the statistical techniques to determine housing submarkets, however it can be effected subjectively given expert knowledge, that is, delineated by real estate agents or appraisers (e.g.: Palm, 1978; Michaels and Smith, 1990).

An important issue discussed in some empirical submarket studies is the spatial adjacency of the housing submarket. Is it an important element that should be taken into consideration? If it is considered that demand does not confine a housing search to a delimitated spatial area, but, fundamentally, it is based on their incomes, then housing consumption decisions are based on similarly priced neighbourhoods located throughout a wide urban area (Goodman and Thibodeau, 2007a). Thus, for the empirical definition of submarkets there is no contiguity requirement, submarkets can be defined spatially or non-spatially, in line with the multidimensional Euclidean space sustained in this thesis.

The complexity of dealing with this issue increases if immutability of submarkets is assumed, that is, different submarkets can appear and disappear, and several scales can be used to analyse the phenomena. It is clear that empirically is it difficult to examine a submarket, because of it changeability.

---

added until all zones have been included. To avoid submarket definitions that are path dependent, sensitivity checks are included of how the final submarket definition depends on the starting point".(Anselin and Lozano-Gracia, 2008, p. 29).

[75] "(…) *provides a classification of the observation into groups (submarkets), and then estimates the parameters for the hedonic price equilibrium in each group. The data are allowed to determine the group structure and coefficients are estimated jointly. A linear mixed model with random effects is estimated by means of non-parametric maximum likelihood* (Anselin and Lozano-Gracia, 2008, p. 29).

In fact, many factors can affect the variability of housing submarkets. Even if it was easy and possible to consider an optimal spatial scale, changes in the social, demographic and economic characteristics[76] of the resident population or some urban amenity, have strong impacts in neighbourhood and can obviously change the implicit valuation of a housing characteristic. This idea is emphasised by Meen (2001) that states that housing markets are not independent phenomena. The author continues arguing that the validity of a macro scale analysis (he mentions the national level) must be based on some condition: first, all households (spatial areas) must behave in an identical manner; second, factors that affect demand must grow at the same rate (e.g. income, employment conditions, etc.); and finally, since land market plays an important role in any explanation of housing market behaviour, differences in land market conditions, both in terms of price and availability, should be considered in the explanation of house price variations at international and regional levels. Of course the problem can be extended to a local scale approach, as an urban area, where it is possible to find several conditions (exogenous and endogenous) that affect the changeability of housing markets.

Grigsby's, Baratz's *et al.*, (1987) housing submarket-based framework provides guidance for those who would like to formalise it. The authors pointed out that households frequently move to maximise the utility of their housing within and overall budget constraint and both *exogenous* and *endogenous factors* can influence their choice. Included in the group of *exogenous factors* are: demographic changes (in consumer expectations, in the number of households, in age, size and family composition of households); economic changes (in real incomes, in the relative cost of housing, changes in the location, amount and type of business investment); governmental interventions that affect housing supply and demand (land-use regulations, tax policies, public service delivery, public facilities, production of subsidised housing, federal transport policies, federal housing insurance policies); obsolescence (building, site, locational); and other changes (rates of new construction, transport and communications technologies). On the other hand, *endogenous factors* can be: negative externalities (crime, physical deterioration and abandoned housing, social

[76] For example, change in marital status; age of the head of the household; retirement; number of pre-school children; previous tenure; income and credit constraints; housing costs; job change; and employment status of any spouse (Meen, 2001).

deterioration) and changing expectations about future house-price appreciation (redlining, disinvestment by property owners).

The *Figure 17* shows the interdependency of some factors mentioned above (and different scales where those factors occur) that can influence the variability of housing markets. The referred figure, adopted from Grigsby *et al.* (1987, p.33), can be described as following: *"Changes in social and economic variables (1) cause households acting directly or through a system of housing suppliers and market intermediaries (2) to make different decisions regarding level of maintenance, upgrading, conversion, whether to move, new construction, boarding-up, and demolition (3), producing changes in dwelling and neighbourhood characteristics (4)".*

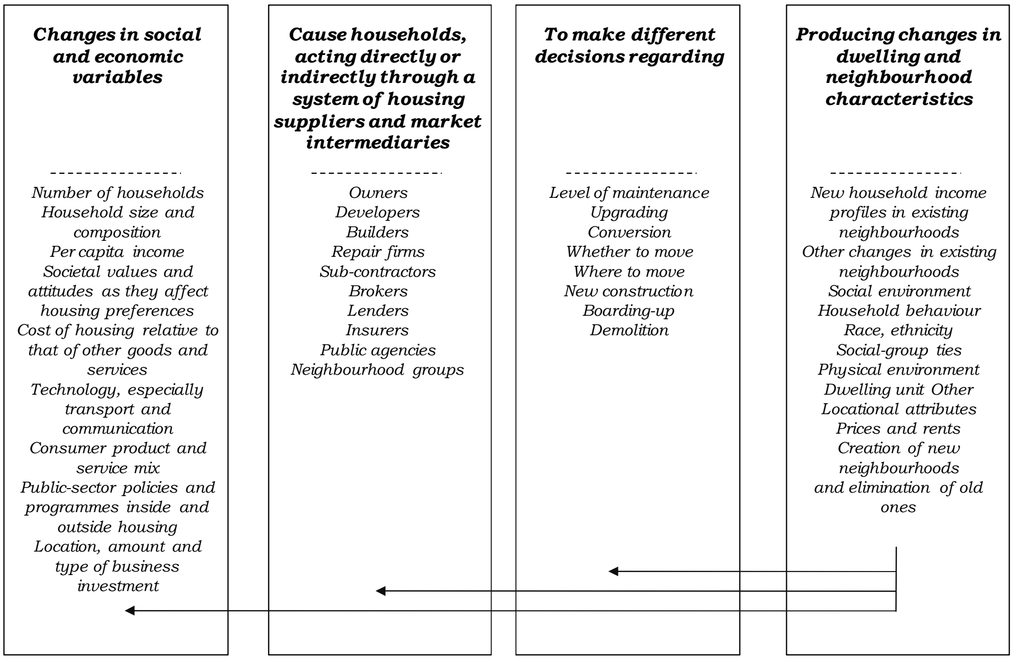


Figure 17 - Framework for analysing submarket change

Source: Grigsby et al. (1987, p. 31)

But, why is it important to define submarkets? There are several reasons, according to Goodman *et al.*, (2007b), to understand how urban areas are segmented: i) from a purely statistical perspective it increases the prediction accuracy of the estimated hedonic model; ii) analytically it enables researchers to better understand the spatial variation in housing prices; iii) an accurate assignment of properties to submarkets improves lenders' and investors' abilities to price the risk associated with financing homeownership; and finally, iv) providing submarket boundary information to housing consumers reduces their search costs.

In many cases adequate treatment of spatial heterogeneity could considerably reduce the presence of spatial dependence effects, even though the two problems are theoretically distinct (Can and Megbolugbe, 1997). Spatial dependence effects are analysed in the next section.

IV.1.3. Spatial dependence analyses

In addition to spatial heterogeneity (discussed above), the other important aspect of spatial structure of housing markets rests on spatial dependence. Spatial dependence refers to a collection of econometric models that explain why observations on any spatial feature tend to be correlated across space. In the context of a hedonic housing price model, such spatial autocorrelation can arise either from spatial diffusion or spillovers in the prices themselves, or from spatial dependence in the stochastic error term. In the former case, the price of each house is affected by housing prices in the neighbourhood, either because of price spillovers, or because these prices are correlated with omitted variables, which, in turn, are spatially correlated. In the second case, correlation arises from the fact that the omitted variables are spatially correlated, even if they are not necessarily correlated with the housing characteristics included in the hedonic model. In either case, ignoring the structure of spatial dependence typically leads to biased estimates of hedonic prices (and their standard errors).

The following example, adapted from Miron (1984), illustrates the problem of not considering the determinants of spatial dependence in prices in an urban area. Let us divide a hypothetical urban area into two submarkets, A and B, with higher housing prices in the first market (A).

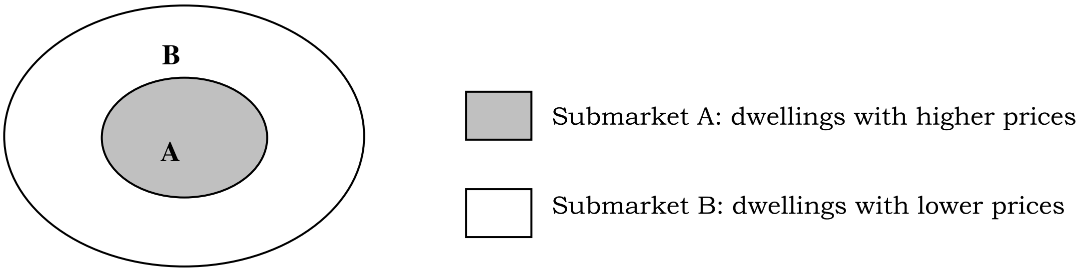


Figure 18 - Hypothetical urban area

The hedonic model of housing price in each of the submarkets can be expressed by the following equation, where it is erroneously assumed that the only reasons for differences in house prices between the two submarkets are the size of houses and the numbers of rooms, which are both larger in A:

$$p_i = v_T h_{iT} + v_N h_{iN} + \varepsilon_i \qquad \text{Eq. 21}$$

Here, h_{iT} represents the total area of the house and h_{iN} is the number of bedrooms. By not including variables such as accessibility to the centre (h_{iC}) or urban amenities (h_{iA}), the model is affected by two types of error. Since h_{iC} is higher in the submarket A, as are h_{iT} and h_{iN}, the included regressors are positively correlated with the omitted variable (h_{iC}), with the result that the estimated hedonic prices υ_T and υ_n are positively biased. Besides, such bias affects estimation of standard errors, which consequently are biased.

By contrast, if the attribute h_{iA} is not correlated with h_{iT} or h_{iN}, either positively or negatively, omission of the variable from the regression relationship will not lead to biased estimates of the hedonic price estimates $\boldsymbol{\upsilon_T}$ and $\boldsymbol{\upsilon_n}$. However, the stochastic error will be larger because of additional variation resulting from the exclusion of the omitted variable, and therefore the precision of predictions obtained from the estimated hedonic model will be relatively poor. Besides the above efficiency issue, and more importantly, the errors would typically be spatially related and examining this spatial pattern is in itself an important objective of understanding the spatial structure of the housing market[77].

[77] If the two missing variables are included in the model, it would nevertheless be ensured that the specification used was correct. Since those houses in submarket A tend to be better and more expensive, this could have a status effect that would make the houses located in A gradually more expensive than the hedonic function determines (including small ones). This would mean that submarket A would have a systematic positive error and the opposite would occur in B. If, on the other hand, for any purpose of trend, the submarket B becomes more attractive, this would imply a positive bias in B.

Methodologies that address the above spatial issues comprise an area of spatial econometrics. As such, spatial econometrics has emerged as a separate scientific domain (subfield of econometrics), to help correct for spatial autocorrelation (treatment of spatial interaction) and spatial heterogeneity (related with spatial structure) when they are in fact present in the data generating process (Paelinck and Klaassen, 1979; Anselin, 1988). To be more precise, Anselin (1988) stated that there are both substantive and pragmatic reasons for incorporating spatial effects into the specification of a hedonic house price model. Substantive reasons are related to the capture of either interaction effects, market heterogeneity, or both while pragmatic reasons refer to it as a nuisance, in that spatial autocorrelation (in omitted variables, or in unobserved externalities) and heterogeneities are relegated to the error term.

Spatial econometric models were first introduced by Whittle (1954) and later by Cliff and Ord (1973) but the first comprehensive attempt at outlining the field of spatial econometrics and its distinct methodology is associated with Jean Paelinck and Leo Klaassen in the reference work "*Spatial Econometrics*" (Paelinck and Klaassen, 1979), After that an extensive bibliography followed: see Anselin (1988), Anselin and Florax (1995), Le Sage and Pace (2009), among others, for excellent discussions of the state of the art in this area, including many exciting applications. A recent comprehensive review of the field can be found in Elhorst (2010) and Anselin (2010).

Paelinck and Klaassen (1979) start out by specifying five principles to guide the formulation of spatial econometric models: i) the role of spatial interdependence in spatial models; ii) the asymmetry in spatial relations; iii) the importance of explanatory factors located in other spaces; iv) the differentiation between ex-post and ex-ante interaction; and v) the explicit modelling of space. Based on these five principles, Anselin (1988, p.7) defines spatial econometrics as the "(...) *collection of methods and techniques that deal with the peculiarities caused by space in the statistical analysis of regional science models*". The spatial econometric model typically draws a sharp distinction between two different kinds of spatial effects: spatial heterogeneity, discussed in the previous section and spatial dependence or spatial autocorrelation (Anselin, 1988).

Spatial dependence occurs when observations at a given location depend on observations at other locations, and is therefore the main cause of spatial autocorrelation typically observed in housing markets, geography, regional studies, and literally all spatial phenomena. The pattern of spatial dependence is closely related to Waldo Tobler's first law of geography, which states: everything is related to everything else, but nearer things are more related than distant things (Tobler, 1970). This law and the consequent idea of distance decay configure the notion of spatial structure (patterns) and relative location (interaction).

Following this philosophical perspective, Anselin (1988) developed two alternative models of spatial dependence: the *spatial lag model* and the *spatial error model*, both including the traditional model defined in equation 5. In the former, prices of neighbouring houses, modelled by a spatial lag dependent variable, are perceived to have a direct effect on the price of an index house. In the latter, spatial dependence arises through autocorrelation in the spatial errors.

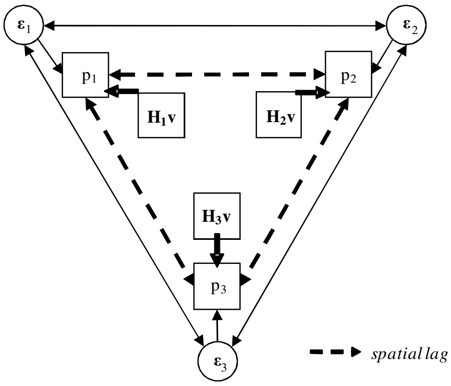
Both these models are described through a spatial weights matrix, generically denoted by W, which captures the interactions between neighbouring spatial units. With *m* spatial units, W is a square (*m* x *m*) matrix with zero diagonal elements, and the off-diagonal elements (or spatial weights) represent the strength of interaction between a pair of units.

The *Spatial Lag Dependence* (SLD) *model*, also called the spatial regression model, assumes that the dependent variable in each observation is correlated with the dependent variables of observations located in the neighbourhood (Anselin, 1988), that is, the price of a particular housing property is not only explained by its own associated attributes, but also by the prices of neighbouring properties (*Figure 19*). If neglected the results would be biased and inefficient.

Introducing the element of spatial lag dependence in equation 5, the following model is obtained:

$$p = \rho W_1 p + Hv + \varepsilon \qquad \text{Eq.22}$$

Where, W_1 is a spatial weights matrix that measures the interaction of neighbouring observations; ρW_{1p} is the spatial autoregressive component, which quantifies the mutual influence of housing prices in neighbouring areas (spatially lagged dependent variable), ρ is the spatial autoregressive coefficient[78], determined by the model, which captures the average influence of neighbouring units, and ε is the error (Meen, 2001).



Indices correspond to several submarkets

Figure 19 - Spatial lag dependence

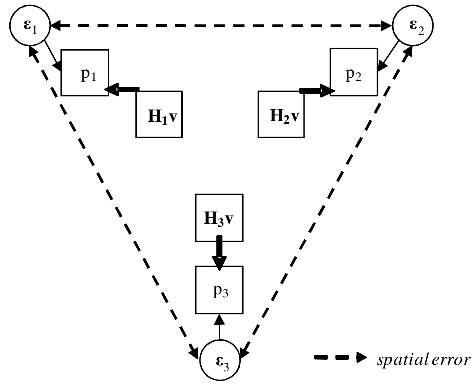
The *Spatial error dependence* (SED) model assumes that the error term at each location is correlated with the errors from nearby locations (*Figure 20*). Inefficient parameter estimates will result if such spatial autocorrelation is not accounted for. The solution is to incorporate spatial dependence through a spatial autoregressive error term in equation 5:

$$p = Hv + \varepsilon \qquad \text{Eq. 23}$$

$$\varepsilon = \lambda W_2 \varepsilon + \mu$$

Where, W_2 is a weights matrix similar to W_1 (usually assumed to be the same) $W_2\varepsilon$ is the spatial lagged error term; λ is the spatial error autoregressive coefficient; and μ denotes the vector of independent error terms (Anselin, 1988; Meen, 2001; LeSage and Pace, 2009).

[78] There are two interpretations for a significant spatial autoregressive coefficient: i) a contagion process or the presence of spatial spillovers; or ii) existence of a mismatch between the observed spatial unit and the true spatial scale of the process being studied (Anselin and Bera, 1998).



Indices correspond to several submarkets

Figure 20 - Spatial error dependence

Combining the spatial lag and spatial error models, a hybrid model can also be assumed for spatial dependence in the prices of housing, which is given by the following equation (Anselin, 1988):

$$p = \rho W_1 p + Hv + \lambda W_2 \varepsilon + u \qquad \text{Eq. 24}$$

$$u \sim N(0, \sigma^2)$$

In this case, if the model is correctly specified, the error term is no more than white noise. The SLD [$\rho W_1 p$] expression indicates that values of the dependent variable are related for reasons beyond sharing similar characteristics; moreover, SED [$\lambda W_2 \varepsilon$] measures the related residuals of neighbourhood properties. When λ and ρ are equal to zero, what remains is the standard model of equation 5 that can be estimated by ordinary least squares. The autoregressive coefficients in both cases are unknown and must therefore be estimated together with the regression coefficients.

The understanding of spatial linkage through the spatial weights matrix W can follow two different perspectives (described in more detail in the next two subsections): i) a parametric perspective where arbitrarily spatial weights matrices are chosen by research, that is, it is assumed that the structure of dependence between observations is known by the researcher[79] (Anselin and Le Sage common approach); and ii) a non-parametric perspective where unknown spatial weights are estimated (Bhattacharjee and Jensen-Butler, 2005; Bhattacharjee, Castro *et al.*, 2012)

[79] Usually it is based on spatial contiguity (*Queen* and *Rook* contiguity matrices) or distance decay (measured in meters).

IV.1.3.1. Known spatial weights matrix

The specification of spatial weights matrix plays an important role in the definition of the appropriate form of spatial model. To estimate the parameters of the equation 24 the specification of W is required. The spatial weights matrix, also called the connectivity matrix, specifies the degree of potential interaction between neighbouring locations and should be chosen to reflect the assumed spatial correlation of data. Any spatial model assumes that observations located in the neighbourhood provide information that is not captured directly by the simple explanatory variables. The weight matrix is, for this reason, essential for most economic markets since it represents spatial interaction, externalities, spillovers, etc..

Similarly to time series analysis (presented in Section IV.2.1), spatial stochastic processes are categorized as spatial autoregressive (SAR) and spatial moving average (SMA) processes. However, the analysis of how both spatial and time dependence operate being complex, there are substantial differences between the cross-sectional and time series contexts. Despite the similarities, spatial dependence is conceptually more difficult. In a time dependence context the effects occur in a longitudinal axis (time axis), that is, in one direction, the research assumes that earlier observations can influence later ones, but not the reverse. There is no corresponding concept in the spatial domain, especially when observations are located irregularly in space, while in the spatial context the direction of influence is not limited to one, but can occur in multiple directions. A spatial weight matrix is thereafter constructed as a proxy for the multiple dependencies between observations[80].

There are various ways to define a weight matrix: using contiguity or distances criteria. Spatial contiguity uses a binary representation as a frequent and simple approach. These types of weighting matrices are based on the decision of whether two houses are neighbours or not. In this case, elements of the matrix appear as $W_{ij}=1$ when house *i* and *j* are neighbours (spatially related), and $W_{ij}=0$, otherwise. The idea is to choose how many nearest neighbours can be

[80] It's equivalent to the covariance structure in time series.

considered neighbours and analyse which weight matrices best fit to the model. The matrix can be row-standardised as[81]:

$$W_{ij}^{s} = W_{ij} / \sum_{j=1}^{J} W_{ij} \qquad \text{Eq. 25}$$

The other method to define the weights matrix is based on some distance decay function, that is, on the distances between houses. A cut-off distance (ξ) is used for this purpose. If houses are located within this limit, the inverse distance ($1/d_{ij}$) between two houses is computed, otherwise, if a house lies beyond the cut-off distance a zero value is returned. In this situation, the weight matrix is defined as:

$$W_{ij} = 1/d_{ij}, \quad if \quad d_{ij} \leq \xi \qquad \text{Eq. 26}$$

$$W_{ij} = 0 \quad if \quad d_{ij} > \xi$$

A hypothetical urban housing market configuration (each unit corresponds to a square) is presented in *Figure 21*. Based on this example the definition of several weighting matrices is illustrated, both contiguity and distances relating to the reference point 6 are reported in *Table 5*.

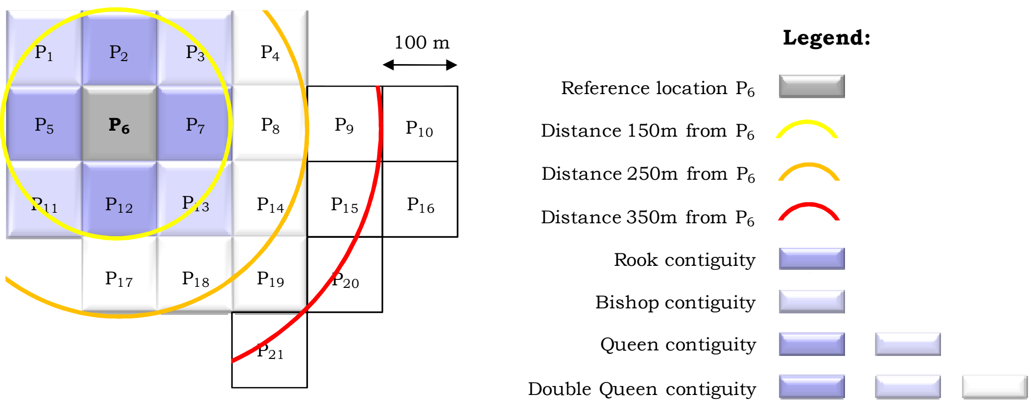


Figure 21 – A representation weighting matrices in a square geographic grid[82]

[81] When row-standardised the matrix may not be symmetric.
[82] The spatial configuration described in this example is of the very simplest form; other configurations can be used to represent a real situation.

Table 5 - Definition and examples of different weights matrices

Classification		Definition	Application
Contiguity[83]	Rook	W_{ij}=1 for cells that share a common side with the reference location *i*, W_{ij}=0 otherwise.	For reference location i=6, W_{ij}=1 for j=2, 5, 7 and 12, and W_{ij}=0 otherwise.
	Queen	W_{ij}=1 for cells that share a common side or vertex with the reference location i, W_{ij}=0 otherwise.	For reference location i=6, W_{ij}=1 for j=1, 2, 3, 5, 7, 11, 12 and 13, and W_{ij}=0 otherwise.
	Bishop	W_{ij}=1 for cells that share a common vertex with the reference location i, W_{ij}=0 otherwise.	For reference location i=6, W_{ij}=1 for j=1, 3, 11 and 13, and W_{ij}=0 otherwise.
Distance[84]	d_{ij}=150m	W_{ij}=1 for cells that are located at a distance less than 150 metres from the reference location i, W_{ij}=0 otherwise.	For reference location i=6, W_{ij}=1 for j=1, 2, 3, 5, 7, 11, 12 and 13, and W_{ij}=0 otherwise.
	d_{ij}=250m	W_{ij}=1 for cells that are located at a distance less than 250 metres from the reference location i, W_{ij}=0 otherwise.	For reference location i=6, W_{ij}=1 for j=1, 2, 3, 4, 5, 7, 8, 11, 12, 13, 14, 17 and 18, and W_{ij}=0 otherwise.
	d_{ij}=350m	W_{ij}=1 for cells that are located at a distance less than 350 metres from the reference location i, W_{ij}=0 otherwise.	For reference location i=6, W_{ij}=1 for j=1, 2, 3, 4, 5, 7, 8, 9, 11, 12, 13, 14, 15, 17, 18 and 19, and W_{ij}=0 otherwise.

Different types of weights matrices, described in the *Figure 21*, can be computed based on a normalised or non-normalised row structure. The following *Table 6* shows the procedure involved:

[83] Other variations of spatially contiguous neighbours can be defined: for example, lengths of shared borders divided by the perimeter, constrained weights for an observation equal to some constant, or *m* nearest neighbours with decay (LeSage, 1999).

[84] There are several ways to define a physical distance function: beside the bandwidth as the m^{th} nearest neighbour, the distance approach can be expressed in the table as inverse distances raised to some power, a variation on the *n* nearest in the neighbourhood, or ranked distances (LeSage, 1999).

Table 6 - Normalized and non-normalized weight matrix

Contiguity	Not row-normalised	W_{ij}=1 if location *i* is spatially related to location *j*, and W_{ij}=0, otherwise.
	Row-normalised	$W'_{ij}=1/\sum_j W_{ij}$ if location *i* is spatially related to location *j*, and W'_{ij}=0, otherwise.
Distance	Not row-normalised	$W_{ij}=1/d_{ij}$ if location *i* is spatially related to location *j*, and W_{ij}=0, otherwise.
	Row-normalised	$W'_{ij}= W_{ij}/\sum_j W_{ij}$ if location *i* is spatially related to location *j*, and W'_{ij}=0, otherwise ($W_{ij}=1/d_{ij}$)

In *Figure 22* and *Figure 23* the matrices W reflecting first-order rook's contiguity relations for five regions are presented. Note that a second, third or *n* order contiguity can be computed but in this case they rely on the distance of the shared border. There are different ways to define weights matrices[85]; for a good discussion of these issues, see Appendix 1 of Kelejian and Robinson (1995) and LeSage (1999).

[85] 1) Spatially contiguous neighbours, 2) inverse distances raised to some power, 3) lengths of shared borders divided by the perimeter, 4) bandwidth as the m^{th} nearest neighbour distance, 5) ranked distances, 6) constrained weights for an observation equal to some constant, 7) all centroids within distance *d*, 8) *m* nearest neighbours, and 9) *m* nearest neighbours with decay (LeSage, 1999).

W_{ij} =

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}	P_{13}	P_{14}	P_{15}	P_{16}	P_{17}	P_{18}	P_{19}	P_{20}	P_{21}
P_1	0	**1**	0	0	**1**	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P_2	**1**	0	**1**	0	**1**	**1**	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P_3	0	**1**	0	**1**	0	**1**	**1**	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0
P_4	0	0	**1**	0	0	0	**1**	**1**	**1**	0	0	0	0	0	0	0	0	0	0	0	0
P_5	**1**	**1**	0	0	0	**1**	0	0	0	0	**1**	**1**	0	0	0	0	0	0	0	0	0
P_6	**1**	**1**	**1**	0	**1**	0	**1**	0	0	0	**1**	**1**	**1**	0	0	0	0	0	0	0	0
P_7	0	**1**	**1**	**1**	0	**1**	0	**1**	0	0	0	**1**	**1**	**1**	0	0	0	0	0	0	0
P_8	0	0	**1**	**1**	0	0	**1**	0	**1**	0	0	0	**1**	**1**	**1**	0	0	0	0	0	0
P_9	0	0	0	**1**	0	0	0	**1**	0	**1**	0	0	0	**1**	**1**	**1**	0	0	0	0	0
P_{10}	0	0	0	0	0	0	0	0	**1**	0	0	0	0	0	**1**	**1**	0	0	0	0	0
P_{11}	0	0	0	0	**1**	**1**	0	0	0	0	0	**1**	0	0	0	0	**1**	0	0	0	0
P_{12}	0	0	0	0	**1**	**1**	**1**	0	0	0	**1**	0	**1**	0	0	0	**1**	**1**	0	0	0
P_{13}	0	0	0	0	0	**1**	**1**	**1**	0	0	0	**1**	0	**1**	0	0	**1**	**1**	**1**	0	0
P_{14}	0	0	0	0	0	0	**1**	**1**	**1**	0	0	0	**1**	0	**1**	0	0	**1**	**1**	**1**	0
P_{15}	0	0	0	0	0	0	0	**1**	**1**	**1**	0	0	0	**1**	0	**1**	0	0	**1**	**1**	0
P_{16}	0	0	0	0	0	0	0	0	**1**	**1**	0	0	0	0	**1**	0	0	0	0	**1**	0
P_{17}	0	0	0	0	0	0	0	0	0	0	**1**	**1**	**1**	0	0	0	0	**1**	0	0	0
P_{18}	0	0	0	0	0	0	0	0	0	0	0	**1**	**1**	**1**	0	0	**1**	0	**1**	0	0
P_{19}	0	0	0	0	0	0	0	0	0	0	0	0	**1**	**1**	**1**	0	0	**1**	0	**1**	**1**
P_{20}	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	**1**	**1**	0	0	**1**	0	**1**
P_{21}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**1**	**1**	0

Figure 22 –Queen contiguity weight matrix not row-normalised

W'_{ij} =

	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}	P_{13}	P_{14}	P_{15}	P_{16}	P_{17}	P_{18}	P_{19}	P_{20}	P_{21}
P_1	0	**0.2**	0	0	**0.2**	**0.1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P_2	**0.3**	0	**0.2**	0	**0.2**	**0.1**	**0.1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P_3	0	**0.2**	0	**0.3**	0	**0.1**	**0.1**	**0.1**	0	0	0	0	0	0	0	0	0	0	0	0	0
P_4	0	0	**0.2**	0	0	0	**0.1**	**0.1**	**0.2**	0	0	0	0	0	0	0	0	0	0	0	0
P_5	**0.3**	**0.2**	0	0	0	**0.1**	0	0	0	0	**0.3**	**0.1**	0	0	0	0	0	0	0	0	0
P_6	**0.3**	**0.2**	**0.2**	0	**0.2**	0	**0.1**	0	0	0	**0.3**	**0.1**	**0.1**	0	0	0	0	0	0	0	0
P_7	0	**0.2**	**0.2**	**0.3**	0	**0.1**	0	**0.1**	0	0	0	**0.1**	**0.1**	**0.1**	0	0	0	0	0	0	0
P_8	0	0	**0.2**	**0.3**	0	0	**0.1**	0	**0.2**	0	0	0	**0.1**	**0.1**	**0.1**	0	0	0	0	0	0
P_9	0	0	0	**0.3**	0	0	0	**0.1**	0	**0.3**	0	0	0	**0.1**	**0.1**	**0.3**	0	0	0	0	0
P_{10}	0	0	0	0	0	0	0	0	**0.2**	0	0	0	0	0	**0.1**	**0.3**	0	0	0	0	0
P_{11}	0	0	0	0	**0.2**	**0.1**	0	0	0	0	0	**0.1**	0	0	0	0	**0.3**	0	0	0	0
P_{12}	0	0	0	0	**0.2**	**0.1**	**0.1**	0	0	0	**0.3**	0	**0.1**	0	0	0	**0.3**	**0.2**	0	0	0
P_{13}	0	0	0	0	0	**0.1**	**0.1**	**0.1**	0	0	0	**0.1**	0	**0.1**	0	0	**0.3**	**0.2**	**0.2**	0	0
P_{14}	0	0	0	0	0	0	**0.1**	**0.1**	**0.2**	0	0	0	**0.1**	0	**0.1**	0	0	**0.2**	**0.2**	**0.2**	0
P_{15}	0	0	0	0	0	0	0	**0.1**	**0.2**	**0.3**	0	0	0	**0.1**	0	**0.3**	0	0	**0.2**	**0.2**	0
P_{16}	0	0	0	0	0	0	0	0	**0.2**	**0.3**	0	0	0	0	**0.1**	0	0	0	0	**0.2**	0
P_{17}	0	0	0	0	0	0	0	0	0	0	**0.3**	**0.1**	**0.1**	0	0	0	0	**0.2**	0	0	0
P_{18}	0	0	0	0	0	0	0	0	0	0	0	**0.1**	**0.1**	**0.1**	0	0	**0.3**	0	**0.2**	0	0
P_{19}	0	0	0	0	0	0	0	0	0	0	0	0	**0.1**	**0.1**	**0.1**	0	0	**0.2**	0	**0.2**	**0.5**
P_{20}	0	0	0	0	0	0	0	0	0	0	0	0	0	**0.1**	**0.1**	**0.3**	0	0	**0.2**	0	**0.5**
P_{21}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**0.2**	**0.2**	0

Figure 23 – Queen contiguity weight matrix row-normalised

The numbers represented in bold capture the notion of connectiviness between the 21 elements, which can be houses or housing submarkets.

The choice of a weights matrix is not obvious, however Griffith (1995) identifies the three questions that should be addressed in the specification of a weights matrix: i) whether the selected specification of weights matrix make any practical difference in results; ii) in what ways does misspecification of a weights matrix influence results; and finally, iii) what, if any, rules exist to guide specification of a weights matrix for a given study area?

The common practice to define an *a priori* weights matrix is to specify different versions of a spatial weights matrix (W_1 or W_2, or both) and then use goodness-of-fit statistics to choose the model that best represents the data. The "*Akaike information criteria*" is an indicator typically used to choose the appropriate W (Akaike, 1974; LeSage and Pace, 2009). Different methods are available in current econometric packages; see, for example, LeSage (1999) and Anselin (2005).

IV.1.3.2. Unknown spatial weight matrix

In the previous section a standard approach to define W based on contiguity and distance measures has been discussed. It is assumed that structure of interactions between observations is known and defined *ex ante*.

However, at the same time as the spatial weights characterise spatial dependence in useful ways, their measurement has an important effect on the estimation of a spatial dependence model (Anselin, 2002). Measurement is typically based on an underlying notion of distance between cross section units. These differ widely across applications, depending not only on the specific economic context but also on availability of data. Further, in many applications, there are multiple possible choices and substantial uncertainty regarding the appropriate choice of distance measure. However, while the existing literature contains an implicit acknowledgment of these problems, most empirical studies treat spatial dependence in a superficial manner assuming inflexible patterns of spatial interaction in terms of known, fixed and arbitrary spatial weights matrices. The problem of choosing spatial weights is an important issue in housing markets where, apart from geographic distances, notions of socio-cultural distances and transportation costs and time can be very important.

Such ambiguity places importance on the emerging econometric literature on unknown spatial weight matrices (Bhattacharjee and Holly, 2009;

Bhattacharjee and Holly, 2011; Bhattacharjee and Jensen-Butler, 2011), where spatial weights are estimated under either structural constraints (such as the assumption of a symmetric spatial weights matrix) or moment conditions.

These approaches use data in the model to estimate W, rather than impose any *a priori* structure. Neighbour influences are defined not in the pure sense of physical proximity, but in terms of non-geographic characteristics, based on the notion of social, cultural and economic distances between individuals. It is considered that the relationship between one point and another is not necessarily a decreasing function of physical distance. Azomahou and Lahatte (2000) argue that, for example, households may pay more attention to other households which are in a same socio-economic situation rather than those in close physical proximity. Following this argument it can be said that the use of a spatial weighting matrix to measure the influence between observations based on geographic notion of distance may lead to incorrect conclusions and to a mis-specified model. The same idea is underlined by Bhattacharjee and Jensen-Butler (2005), stating that the use of arbitrary structures of weight matrices lead to substantial differences in the results.

In particular, in the context of a hedonic housing price model with pure cross-section data, Bhattacharjee, Castro *et al.* (2012) developed a methodology based on statistical factor analysis to estimate spatial interactions within and across housing submarkets under the structural assumption of symmetry. In short, the spatial weights matrix W is estimated using the residuals from a first-step regression.

The methodology is explained in more detail in the Chapter VI and applied to the housing market in Aveiro (Portugal) at two different spatial scales, in the Chapter VII.

IV.2. Temporal housing market analysis

Inference drawn from hedonic prices, using the analytical techniques described above, can be used either to identify the processes of adjustment to exogenous factors (such as policy measures and changes in purchasing power) and endogenous shocks (such as cyclical phenomena), or for the projection of trends in the market. However, the complexity, dynamics and volatility of the determinants of these trends, as well as the patterns of interaction between these factors pose serious limitations on the reliability of such projections. Thus, alternative techniques based on expert assessments, such as foresight analyses, are often more useful. The combination of these types of techniques involving a strong qualitative component, with analytical methods (with greater formal rigour), is a rarely explored area, but with a high potential for use in the study of the housing market.

In general, the study of the temporal perspective, or future analysis, implies two different types of methodologies: *analytical methods* and *foresight holistic methods*. The first group is based on time series analysis and its extrapolation to the future. On the other hand, foresight methods, whether or not based on the analysis of the past, seek to build either a feasible and desirable future or, assuming possible evolutionary scenarios, be they desirable or not, to anticipate answers to such scenarios. For example: changes in the macro-economic environment, evolution of construction techniques and housing design, as well as, changes in cultural values. Such issues would be addressed by using different types of foresight techniques. The use of these techniques, combining a strong qualitative component with analytical methods, is a relatively unexplored field but has a high potential for use in housing market studies (Marques and Castro, 2010; Marques, Castro *et al.*, Forthcoming).

The literature on time series analysis and foresight techniques is vast[86] although, in general, not explicitly linked to the housing market. A brief overview of the potential application of these techniques to housing market analysis is presented in the next two sections.

IV.2.1. Time series analysis

Time series analysis is concerned with the forecasting of variables which involve a relatively predictable process. Panel data is used to determine the time variation of the vector v. There is a wide range of time series techniques that can be used to describe and explain the evolution of hedonic prices (simultaneous or individual) as the result of a reaction to a deterministic set of variables or as a response to a sequence of stochastic shocks – ARIMA and State Space Time Series Analysis are some examples of techniques that can be used (Chatfield, 2000; Commandeur and Koopman, 2007).

A time series is defined as a set of quantitative observations of a variable, measured sequentially over time (Box and Jenkins, 1976). Generally, a time series can be considered as the combined result of deterministic and stochastic processes and is decomposed into four elements, or state variables[87] (Commandeur and Koopman, 2007):

- *Level (N)*: component equivalent to the intercept of a classic linear regression model. It can be time-invariant or vary under the influence of stochastic shocks. In the latter case, the level has only a local meaning and resembles the effect of a dummy variable in time.
- *Slope (D)*: component of the time series that evolves in a particular way deemed to be linear or possible to linearise (exponential, logarithmic, etc.). As before, the slope can be either time-invariant or subject to the influence of stochastic shocks. In this case, the

[86] General literature regarding foresight analysis can be found in Godet (1991) and in Voros (2003).
[87] According to the models of representation of the *state space* (State Space Time Series Analysis; see Commandeur and Koopman, 2007), it is considered that a time series is the result of a dynamic process defined by a set of non-observable variables, stochastic or deterministic, designated as *state variables*.

sequence of local trends is equivalent to a set of slope dummy variables[88].

For the sake of simplicity, the components N and D can be aggregated into a single Trend (T)

- *Seasonal component (S)*: describes the oscillatory variations of a temporal series, usually with a periodicity of one year. It is particularly important in the analysis of rental markets in tourist resorts or places where the demand for temporary residence is concentrated in specific times of the year[89].
- *Irregular component (I)*: corresponds to the error term which, if the time series is correctly specified, is only white noise; otherwise it includes the influence of the previous terms of the series. As discussed further on, this time dependence can be described either as the cumulative effect of oscillations in the state variables or as combined processes of serial autocorrelation and moving average effects, reflecting the sequence of stochastic shocks and movements back to the average situation

In short, a time series can be described by the equation 27, and a typical decomposition is shown in *Figure 24*.

$$Y_t = f(T_t, S_t, I_t) \qquad \text{Eq. 27}$$

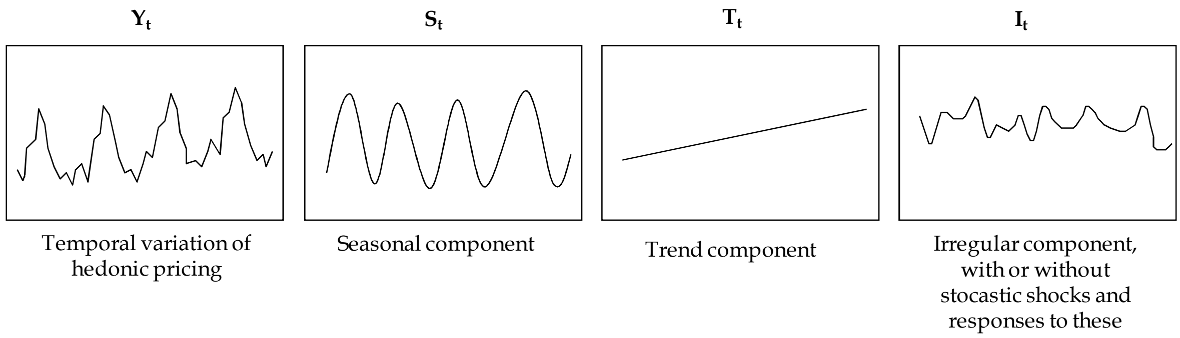


Figure 24 -Time series elements

[88] A time series with both constant level and tendency is correctly described by a linear process.
[89] Some authors distinguish between seasonal and periodic components, whereas the latter has variable periods normally longer than one year (Chatfield, 2000). Despite its importance for the study of the residential market, characterized by a succession of cycles of euphoria and implosion of speculative bubbles, this component is rarely considered in econometric models, given their structural complexity, coupled with the requirement for long time series.

The state space representation models (SSRM) analyze the equation 27, identifying the contribution of each of the state variables for the temporal evolution of the dependent variable y_t. These models: i) separate the deterministic element (constant) and the stochastic shocks which define the level and the slope of the time series, ii) identify the effect of stochastic shocks on the variation of the level and slope, and finally, iii) isolate the irregular component, which, according to the logic of these models, has a white noise distribution. SSRM models can also incorporate: i) explanatory variables which quantify the influence of any exogenous factor working either directly on y_t or through the change of state variables; ii) intervention (dummy) variables, identifying temporary or long term changes either in y_t or in state variables, resulting from the application of policy measures (for a full description of these models see Commandeur and Koopman, 2007).

The state space representation models can be applied to a single time series or to a multivariate analysis of a set of time series that evolve by common processes, where explanatory or interventional variables are somehow related to each other.

Whether as descriptive tools, or as forecasting techniques, SSRM models have a broad scope of application to the analysis of the residential market. These models identify both the processes underlying the development of combined or isolated hedonic prices, and the impacts on these processes arising from changes in social and economic factors or from the application of policies that directly or indirectly affect the housing market, e.g.: tax laws, town planning legislation, public interventions that change land prices, incentives to foster either house ownership or the rental market.

Combining flexibility and analytical rigor, SSRM models have the disadvantage of not considering the hypothesis that the effects of stochastic perturbations in the state variables weaken over time, rather than being permanently cumulative. This hypothesis is the basis of ARIMA models, based on the work of Box e Jenkins (1976), which have, however, the limitation (not present in SSRM models) of requiring that the time series under analysis are stationary (Commandeur and Koopman, 2007)[90]. Stochastic processes corresponding to random walks or likely to be described by polynomial series can

[90] It is said that a stochastic underlying a time series is stationary (second order) when its mean, variance, and auto-covariance are constant (Commandeur e Koopman, 2007).

be transformed into stationary processes through successive operations of differentiation, defining its degree of integration I(d), as the number of differentiations needed to become stationary (Pankratz, 1983).

Briefly, an ARIMA model shows how a variable can be related to other observations in the past[91]. Once differentiated, an operation that corresponds to the *I* of the acronym, the time series y_t is decomposed as follows:

$$y_t = C + \underbrace{\phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p}}_{AR(p)} + \underbrace{\lambda_0 u_t + \lambda_1 u_{t-1} + \lambda_2 u_{t-2} + ... + \lambda_q u_{t-q}}_{MA(q)} + a_t \qquad \text{Eq. 28}$$

The term *C* is a constant and a_t is white noise. The component AR(*p*) indicates that the time series is autoregressive of order *p* and describes how each element of the series is influenced by their previous values; the parameters ϕ_p are such that the influence of lagged values are attenuated over time, and the index *p* indicates the number of lagged values (usually years) that affect y_t[92]. Finally, the component of moving average of order *q*, MA(*q*), reflects the influence on y_t of the error terms of the elements of *t* previous series (stochastic shocks, u_{t-q}), being the parameters λ_q such that this influence is attenuated in time, while *q* is the number of lagged values[93]. In conclusion, an ARIMA (*p,d,q*) is a model that has *p* autoregressive terms, is integrated of order *d* and has *q* moving average error terms. The criterion of parsimony requires that, among all possible ARIMA series that describe a given stochastic process, the one that has the lowest *p* and *d* parameters will be chosen.

ARIMA models can be corrected to discount the effects of seasonality (SARIMA) and can be used to extrapolate the time series for the future, starting with the last value of the series that is available and, of course, considering that the future values of the white noise component (a_t) are zero (see *Figure 25*). It is also obvious that the extrapolated values should be submitted to *d* integrations, in order to produce a stochastic process consistent with the observed values.

---

[91] For a detailed analysis of this model see Box and Jenkins (1976); for a more accessible view see Pankratz (1983).

[92] Note that AR(*p*) is equivalent to the lagged spatial dependence of spatial econometric models.

[93] MA(*q*) is equivalent to the spatial errors of spatial econometric models

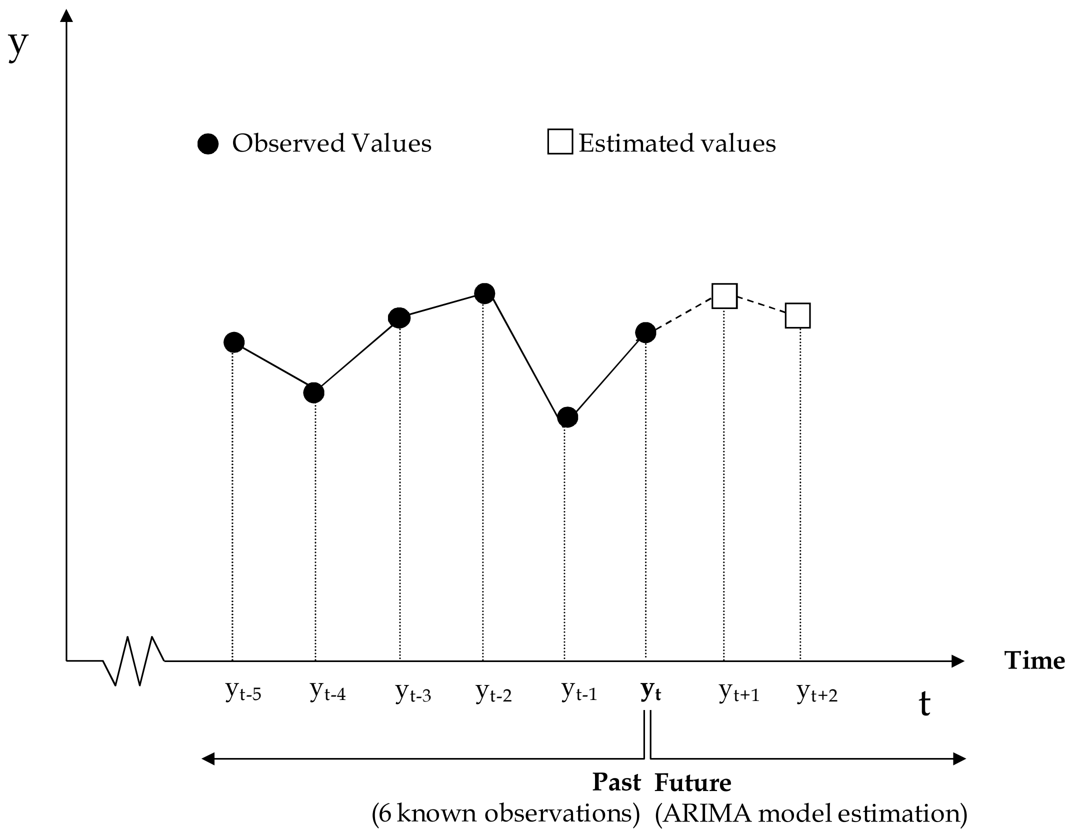


Figure 25 - Example of an ARIMA model

Source: based on Pankratz (1983)

IV.2.2. Holistic Foresight Methods

Despite the formal rigor, and the highly structured theoretical basis of the econometric models, they generate predictions of the evolution of the housing market which are affected by the difficulty in finding the required data and by the inability to deal with qualitative changes that influence, in a significant way, the underlying stochastic processes.

According to Johnson and Marcovitch (1994), until the 1950s, planning techniques were solely based on projections of recent trends over relatively short time horizons. However, the rapid pace of technological change, the increased competition at firm, regional and state levels, the unpredictability of social, political and economic systems, along with the generalization of the principles of strategic planning, led to the development of foresight techniques which were able to cope with the continuous transformation of reality. Such techniques seek to look at multiple and uncertain futures, aiming to inform decision making processes. Rather than making statistical inferences they determine probabilities, of a subjective nature, by consulting expert groups (Godet, 1997); rather than using analytical methods they seek the holistic understanding of phenomena and their multiple relations.

The expected quality of results arising from foresight techniques depends on several factors: the rigorous definition of the problems and issues to be addressed; the criteria for experts' selection; the clear description of the different subjects, and the quality of the discussion with experts; the existence of a tacit knowledge concerning the topic addressed, shared by the experts' panel.

Among the foresight techniques more commonly used are the qualitative *scenario analyses* and the quantitative *Delphi questionnaires*.

Briefly, scenario analyses are based on descriptions of alternative futures, constructed as combined outcomes of hypothetical variables which are exogenous to the strategic intervention domains but which strongly influence such domains and the desired development path (Fahey and Randall, 1998; Marques, Castro *et al.*, 2008). Compared with dynamic control theory, it can be said that the scenarios are distant points in the space of state exogenous variables (boundary conditions or restrictions) of the system under analysis; for each scenario and the corresponding restrictions, the experts define the programs most adequate to obtain the desired objectives; such programs correspond to the best combinations of control variables. In other words, scenario analysis consists of discussing the implications of each scenario and in identifying the best strategies to cope with the future developments described by the scenarios. According to Godet (1997) scenarios can consider those developments that are more probable or more extreme, closer to the expected trends or more contrasting and distorted. The diagram in *Figure 26* illustrates the different types of views in a foresight exercise.

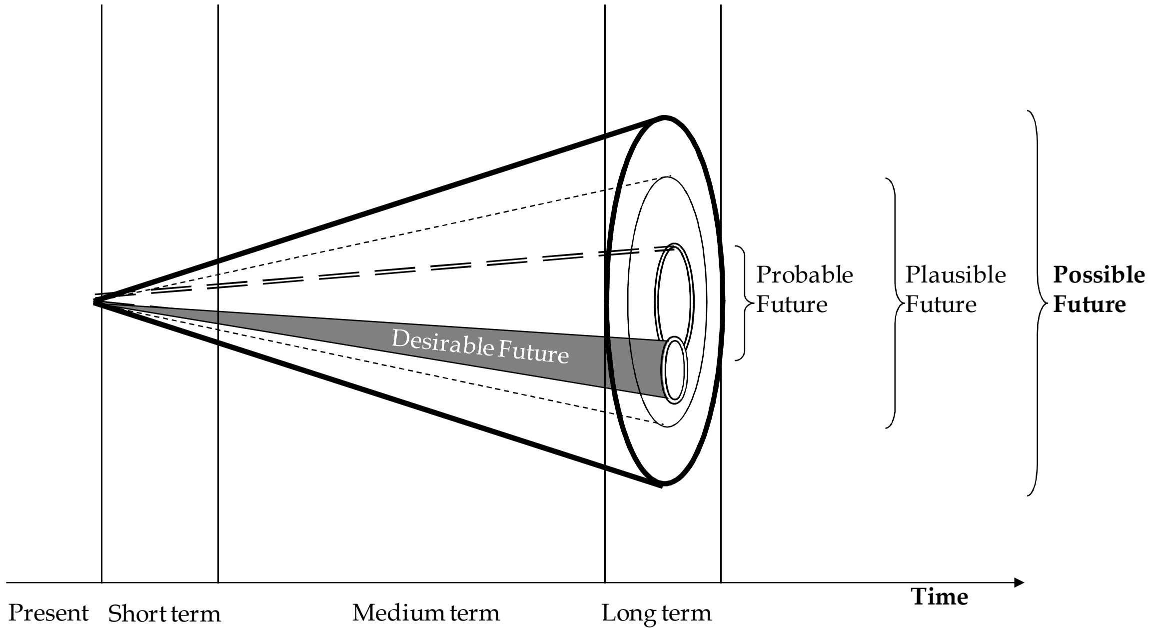


Figure 26 - Different visions of the future

Source: Adapted from Josep Voros, 2000

In the particular case of the housing market, the scenario analysis is a useful tool for the discussion of different strategies to be adopted by public authorities (municipalities, central government agencies) and private entities (civil construction, real estate firms) given the occurrence of various scenarios. For example, three hypothetical scenarios can be imagined, arising from the variation of three basic dimensions analysis: i) urban concentration, increasing transportation costs and, a great pressure for the rehabilitation of traditional buildings in inner cities; ii) urban concentration, increasing transportation costs, and a clear cut distinction between new and technologically advanced residential houses and traditional buildings, primarily for business and public service functions; iii) organised sprawl of the built space combined with a strong commitment to efficient public transport and inter-modality, fostered by a high demand for quiet green residential areas, where modernity and tradition are combined.

Delphi questionnaires are techniques which aim to collect expert opinions through objective and quantified questions related to: i) issues for which there is little or no empirical data; ii) future developments in which the simple extrapolation of trends is deemed insufficient or even impossible; iii) expected qualitative changes in the determinants of the future path (Cuhls, 2001; Gordeon and Pease, 2006; Marques, Castro *et al.*, 2008).

Delphi questionnaires typically require several rounds of questions, so that respondents can compare their answers with the average values for the group, and then reassess and possibly reconsider their answers. Theoretically, it is desirable that the process is repeated as many times as is needed to achieve consensus among the participants, which is often not possible due to logistic or time limitations [European Commission, (2002)]. The number of experts participating in the exercise is variable and should depend on their knowledge about the issues raised. The method involves individual and anonymous consultation in order to allow each expert to speak out freely and without constrains.

As suggested by Cardoso, Abiko *et al.*, (2005), there is no agreement on consensus as an objective of Delphi questionnaires. Some authors consider consensus as being the central objective of the process, while others admit this may possibly not occur in all questions. The above quoted authors developed Delphi questionnaires for residential construction in Brazil, presenting a detailed description of the methodology as well as of the data treatment and analysis.

The literature of these methodologies applied to the housing market is less extensive than desirable, but some examples of empirical application can be found in Basu and Schroeder (1977), Mulligan, Franklin *et al.* (2002) and Burke, Zakharov *et al.* (2005), applied to United States of America, Australia and the United Kingdom, respectively.

IV.3. Summary

In this chapter a summary of the techniques to analyse and understand the spatial nature of the housing market has been presented. The standard model used to assess the contribution of each of housing's characteristics, both physical and location, is based on the so called hedonic price models. Caution, however, should be exercised when devising hedonic models.

All regression models are based on a large number of assumptions such as the normality of errors, constant error variances or homocedasticity errors. Nevertheless errors should be independent from one another. These problems are considered the basic assumptions of classical regression methods, and if they are violated, the estimates of parameters provided may be inefficient and biased, resulting in incorrect confidence intervals. These problems may occur after the selection of the independent variables and the functional form of the equation.

Even considering that these fundamental assumptions were not violated there are other big challenges associated with the implementation of these standard hedonic models. Appropriate variables must be selected carefully and measured accurately.

The first exercise is to define attributes that should be included in the model, both in the dependent term (the variable used to measure the housing value can be a rent or sale) and dependent variable[94]. The decision to opt for one or the other of these can lead to different strategies for analysing the urban housing market and consequently to a distinct conclusions. The purpose of each study should be to support clearly the selection criteria. The choice of independent variables is more complex, not only because of the high the number of attributes involved in the explanation of house value, but because of the scarcity of information and missing indicators necessary to characterise all

[94] Note that this attribute can be measured in terms of a scalar, ordinal or dichotomic variable (using this strategy the regression model is called probit model)

dimensions of a dwelling. In this perspective, the size, quality and the conditions of a property are relatively easy to measure, but the measurement of environmental amenities is much more difficult to define. The reason for this argument is the following: even if technical measures are relatively easy to obtain (such as concentration of some air pollutant or peak noise level), it is not easy to be sure that those measures correspond to a households perception. The aggregate level of environmental indicators is an additional challenge, necessary to avoid bias in estimated coefficients.

The second big challenge is the choice of functional form for the hedonic equation relating all the attributes. The literature does not specify any particular model; additive or multiplicative forms can be used depending on the data that is used. Thus, the model should be chosen based on the accuracy of the model. However, it is clear that some variables are not linearly correlated with the price of a dwelling, for example: the age and area of a dwelling. The impact in the price of a variation in 10 m^2, in a small house, is not the same as in a big house; and the impact in the price of a house of a variation of two years in a new house is not the same as in a old one. In these particular situations the logarithm (log or ln) of age and area should be considered. The multiple coefficient regression (R^2) is a good marker to analyse different options.

Besides the difficulty of choosing the relevant independent variables and the model specification, three other specification problems can be found in hedonic price models: multicollinearity, structural heteroskedasticity, and spatial autocorrelation. Considerable effort has been expended in several scientific domains (geography, economy, statistics and regional and urban science in general) to develop diagnostics for detecting when these assumptions are violated and to define appropriate corrective actions.

The emerging spatial econometric issue has contributed to solve or to mitigate some of the mentioned problems. Spatial models began with the principle that proximity provides information about nearby observations that are not captured directly by the simple explanatory variable. According to the literature there are two distinctive ways in which spatial interaction is modelled in spatial regression analysis: spatial lag dependence (biased problem) and spatial error dependence (inefficient problem).

In both situations a weights matrix should be chosen to reproduce the spatial spillovers or spatial dependence between housing located in a specific surrounding. Besides its crucial importance, the definition of the weights matrix is the most non-consensual issue in spatial econometric analysis. The strategy of imposing an *a priori* structure of spatial dependence can not correspond to reality. Spatial contiguity or geometric distances to reflect the spatial interaction are frequent choices. The alternative is to consider the spatial weights matrix as an unknown symmetric matrix. The principle associated with this approach is that spatial interactions may be driven by other factors, such as, cultural, sociological and economic contexts. This is a controversial and unsolved issue that allows future advances in this area of the spatial econometrics literature. In Chapter VII, together with the empirical application, some contributions will be given for this particular topic.

The spatial segmentation or housing submarket is the other main feature of the spatial econometric domain. The spatial heterogeneity is an additional methodological concern in the explicit consideration of spatial effects and refers to the fact that the spatial econometric relationships may vary over space. The non-inclusion of this aspect in spatial model analysis can create problems because it is assumed that a unique constant relationship holds for the entire data sample.

Regarding the advantages of hedonic models, including spatial econometric models, with respect to the other valuation methods (for example, contingency evaluation[95]), the willingness to pay (WTP) is based on households' real willingness to pay for dwelling characteristics as revealed by the market, rather than households assessment of hypothetical alternatives from which they are supposed WTP is deduced. In addition, with the recent development of geographical information systems (GIS), statistical treatments provide a more reliable indicator for the hedonic approach. However, it should be noted that the hedonic price method, like all valuation techniques proposed in the literature, is a partial equilibrium approach.

[95] The contingent valuation method (CVM), is based on surveying individuals directly about how they would behave in a hypothetical market situation (Cummings and others 1986).

PART 3

New conceptual framework

V. How to analyse the role of space in the housing market

The main purpose of this chapter is to present a new conceptual framework[96] (the corresponding methodology is presented in the next chapter) to understand the role of space in urban housing markets. This framework pays special attention to the three distinct (but linked) aspects of spatial analysis: spatial dependence (or interactions or spillovers), spatial heterogeneity (or patterns of segmentation) and spatial scale, emphasising the perspective of multi-dimensional non-Euclidean geometric space.

From the literature (theoretical background presented in the Part 2) three main statements emerged:

i) From the *historical empirical evidence* (Chapter II), the idea that bi-dimensional Euclidean geometry may be not useful enough to explain the highly complex system of spatial urban structures, reflected by the sprawl of settlements.

ii) From the *urban studies disciplines* (Chapter III), the notion that the assumption of geometric space is not adequate to embrace the complexities of spatial structures. More specifically, in urban economics, some unrealistic assumptions, about heterogeneity and interactions, have been considered to

[96] Its application to the urban context of Aveiro/Ílhavo is presented in the chapter VII.

make models feasible and comprehensible. It should be noted that some spatial models, e.g., from New Economic Geography, already encompass some of this spatial complexities, however, these still suffer from lack of empirical opportunities to assess how several forces act together in the territory. In urban geography, emphasis is placed on the fact that many aspects and forms of reductionism and pure geometric understandings of space should be avoided, since space is considered to be socially produced. Finally, in the same line of reasoning, the apparent dematerialization of planning regained prominence in urban planning, where, the physical substance of the notion of space has, over time, been lost, becoming more immaterial.

iii) From the *spatial econometric* literature (Chapter IV), the fact that the choice of spatial weights is a key issue in the spatial models, and standard approaches which impose a theoretical and *a priori* pattern of spatial interaction, to represent and analyse spatial interactions between different submarkets or dwellings, may or may not correspond to the reality. Typically, spatial models are modelled as a function of geometric distances, both spatial contiguity and geographic distances.

These arguments, explained in more detail throughout this chapter, give good reasons to justify the development of an appropriate framework and corresponding methodologies to capture and analyse the intangibility of space, both, in terms of spatial heterogeneity and spatial spillovers.

The most important argument behind, and somewhat responsible, for the distinction between the notion of bi-dimensional Euclidean and multi-dimensional non-Euclidean space is the way that space is structured and analysed. In the former, and in line with Henri Poincare (1895) space is characterised as a rigid and immutable structure; the objects (individual or collective agents acting in the territory) are precisely distributed and organised based on pre-assumed logic and a well known metric. The space can be reduced to the essence of bi-dimensional Euclidean geometry, fitting with our geometric perception of space. The latter, refers to a more complex representation of space, where the location of objects are responsible to produce and transform space itsself. This notion of space does not necessarily conform to our physical concept of distance or contiguity, and hence, no fixed structures can be assumed; it is potentially an anisotropic, heterogeneous and multidimensional space.

The criteria for distinguishing theses two different notions of space are shown in the *Table 7*.

Table 7 - Bi-dimensional Euclidean and multi-dimensional non-Euclidean space

Axioms/specificities	**Bi-dimensional Euclidean space:** (Fits with our geometrical perception)	**Multi-dimensional Non-Euclidean space:** (Does not conform to our physical concept of distance)
Symmetric[97]	The distance from **a** to **b** is the same as that from **b** to **a** $d(\mathbf{a},\mathbf{b}) = d(\mathbf{b},\mathbf{a})$ for all $\mathbf{a},\mathbf{b} \in A$	The distance from **a** to **b** *is not necessarily* the same as that from **b** to **a**
Non-negativity[97]	d_{ij} must be zero or positive $d(\mathbf{a},\mathbf{b}) \geq 0$ for all $\mathbf{a},\mathbf{b} \in A$ and equality if and only if $\mathbf{a} = \mathbf{b}$ (a object is identical to itself)	d_{ij} can be positive or negative
Contiguity	Continuous: Geographical proximity where objects are in physical contact	Not continuous: Objects which are not connected (geographically)
Directional (in)dependencies	Isotropic: uniformity in all orientations	Anisotropic: properties are variant in relation to a particular direction (or not)
Dimensionality	Plane, two dimensional space	n dimensional space
Spatial patterns	Heterogeneity: precisely delimited by geographic boundaries	Heterogeneity: boundaries defined by social, economic, languages, cultural aspects and by other non-geographic dimensions
Fixity (spillovers)	Predetermined functional form (structure)	No fixed structures: unknown relationships are constructed according to information derived from the data

The new conceptual framework to understand the role of space in urban studies is presented below; resulted from the theoretical background (philosophies, facts and methods) and particularly from their gaps and contradictions.

The following *Figure 27* shows a schematic representation of the two different and dichotomous approaches of understanding space: a bi-dimensional Euclidean notion of distances (on the left hand side of the figure) and a multi-

[97] Space is a set of A in which a distance function or metric d exists.

dimensional non-Euclidean notion of space (on the right hand side of the figure). This conceptual framework resulted from the literature review but has been reorganised in a distinctive manner for the purpose of this work, playing a crucial role in the development of this research.

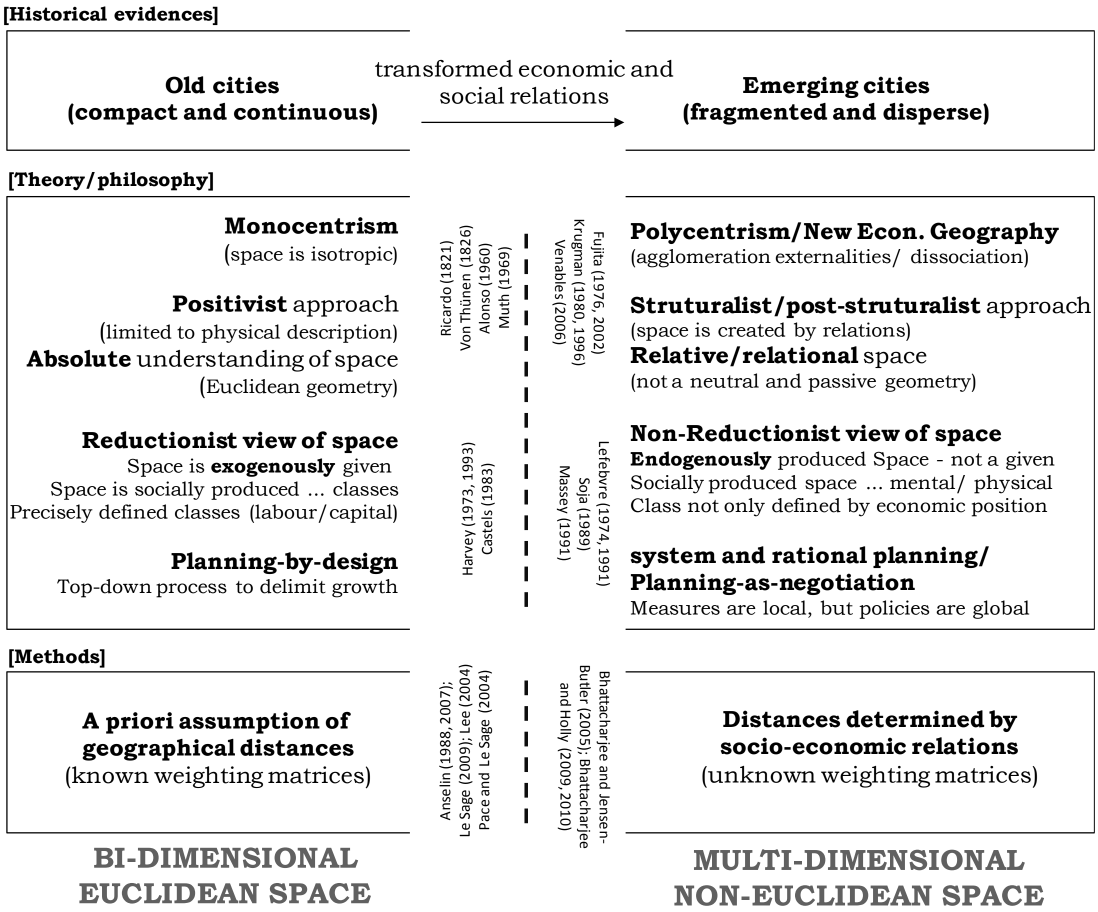


Figure 27 - Theoretical framework of the dichotomy between bi-dimensional Euclidean and multi-dimensional non-Euclidean notion of space

For each of these two different perspectives (bi-dimensional Euclidean and multi-dimensional non-Euclidean space) a more detailed interpretation follows, at three levels. First, focus on the historical overview about the evidence of urban growth (i), highlighting the main shifts and trends of the physical and social transformations of urban areas. Then, philosophical and theoretical perspectives in the urban studies are presented (ii), emphasising the major challenges of analysing urban patterns, and hence, the determinants of the spatial dynamics. As described previously, urban studies disciplines provide a variety of perspectives about the nature of space. Urban studies have long been an interdisciplinary field, in which several disciplines contribute to understanding the underlying social economic processes that produced urban space: as has been mentioned before, this research focused on three main disciplines: urban economic, urban geographic and urban planning; with the realisation that the delimitation of each of these domains is somewhat ambiguous and controversial. And finally, spatial econometric methods are presented (iii), to analyse the urban spatial structure and its interaction in the context of the housing market, subject to available methodological innovations. Note that, in the Section III of Chapter 2, a summary of the main techniques in the context of spatial econometric literature has been described and the lack of attention and insufficient importance attributed to the multi-dimensional non-Euclidean space has been highlighted.

V.1. The historical evidence of urban transformations

Spatial patterns of urban areas have changed, as a result of many combined factors (economic, technological, demographic, political and environmental), as has been described Chapter II (where the main urban transformations and the debate about the major challenges of the urban spatial development have been presented in more detail).

Urban sprawl is the latter stage of urbanization and was an inevitable phenomenon with its owns pros (the cost of a house is often lower) and cons (higher infrastructures costs, long commutes, air pollution, etc.)[98]. As a response of dependence on the automobile and to combat sprawl, New Urbanism (an

[98] See the project founded by the Portuguese scientific foundation (PTDC/AUR/64086/2006) "*Costs and Benefits of Urban Dispersion on a Local Scale*", where the main goal was to confront costs with benefits in different urban contexts (sprawl vs. compact city). Benefits were associated with the concept of quality of life, which changes from opinion group to opinion group. This concept was transformed into an algorithm which integrates this variability, based on the answers to a questionnaire given to the inhabitants of the two case study cities, Aveiro and Évora (Belbute, Marques *et al.*, 2009). Costs were divided into two different types: local public infrastructure and mobility related costs. (land consumption and other environmental externalities have been left for a later research opportunity)(CBOD, 2011).

emergent urban design movement), began to tout its validity and sustainability and advocate that development should be more compact, based on traditional urban forms, and neighbourhoods should be diverse in use and population (for a more detailed interpretation about principles of the New Urbanism, see the Charter of the New Urbanism, 1999).

A whole range of advantages and disadvantages of different types of urban growth are brought forward in the literature. Even if this takes different forms, the tendency has been for the evolution from the compact and continuous to fragmented and disperse types of occupation. Cities have expanded beyond their territorial boundaries and their CBDs are no more the unique centre of a city.

The existence of many sub-centres with different levels of importance among the spaces increases the complexity of analysing spatial patterns and spatial interactions. On one hand, spatial segmentation (housing submarkets) may not be clearly defined by traditional administrative boundaries, type of land use or any other socio-economic disaggregation; and on the other hand, spatial spillovers, that operate at different scales, might not be captured by a fixed and pre-assumed spatial weighting matrix (considered by the literature the most appropriate way to describe spatial interactions).

The mutability of spatial urban phenomena and the precise definition at which scale these various forces work, reinforce the need for analysing space with methodologies (spatial econometric methods) able to capture this complexity of space, characterised by flow and channels without any predetermined localised form. Thus, the particular importance of not reducing the space to the essence of geometry is emphasized, that is, represented with relative precision through a system geometric distances.

The following table summarises the main ideas of historical urban growth and its connection with spatial heterogeneity (SH), spatial interaction (SI) and spatial scale (SS).

Table 8 - Summary of historical evidences of urban growth and its connection with SH, SI and SS

Main transformations	▪ Traditional compact cities are evolving into large urban areas expanding beyond their own territorial limits. ▪ Spatial decentralization leads to suburbanisation and dispersion of households and activities. ▪ Central Business Districts (CBDs) are no more the unique centre of a city, many new other centralities or sub-centres emerge and change overtime.
Spatial heterogeneity	▪ Such dispersed urban growth does not result in a regular and standard structure, thus, spatial segmentation is not clearly defined by traditional administrative boundaries, type of land use or any other socio-economic disaggregation or standard concepts of neighbourhood.[99]
Spatial interaction	▪ The trends of increased accessibility (commuting travels between work, home and leisure) and the cheaper land prices available in more remote areas leads to a more interdependent and non-contiguious effects of interaction.
Spatial scale	▪ Because of the existence of many sub-centres with different levels of importance in the space (various forces at work) it is not possible to predetermine a specific scale to capture both spatial heterogeneity and spatial spillover. It is quite possible to have a spatial pattern centralized at one scale and dispersed at another.

[99] See Galster (2001) where the author explores the myriad of idiosyncrasies associated with neighbourhoods.

V.2. Philosophical and theoretical views

Urban studies theory makes an effort to follow the complexity of urban phenomena, but to some extent, models and approaches used to understand and represent this spatial reality are very simplified, based on unrealistic assumptions, or somewhat ambiguous because of the mismatch between the theory and empirics. In order to understand spatial patterns, within and between cities, and the consequent insights provided for future development, it is necessary to correspond with the relevant ideas in economic, geographic and planning fields of thought.

<u>The space in urban economics</u>

The core of urban economics is describing (positive theory) and interpreting (normative theory) the way that land has been occupied, both by households and firms. In line with what has been described above, a fundamental assumption of all spatial economic theories is that places with good accessibility are more attractive and have a higher market value compared to those located in peripheral locations. This fundamental assumption goes back to von Thünen (1826 [1966]) which is considered the first systematic approach of how spatial patterns in urban environments might emerge. His model, incorporating Ricardo's land rent concept, explains how the land is distributed in the case of free competition among farmers and landowners, arguing that the urban structures emerge from the aggregate outcome of the utility maximizing choices of individual consumers. Thus, bid rent helps to understand why various territorial agents (people and activities) pay a premium for places that have lower commuting costs. This model has been formalised in the urban context by Alonso (1964), and further extended by Muth (1961) and Mills (1972b). In the so-called AMM model of the monocentric city the isolated state has been replaced by the city centre, the farmers by the residents, and the agriculture land by residential areas. The land use monocentric model, considered the core of the scientific discipline of urban economics (Fujita, 1999), is a good example of how urban spatial structure can be conceptualised in a rigorous and very accurate manner. In its simplest form, the model describes the city, has a single, exogenously *a priori* determined city centre (CBD), as a circular residential area surrounding the CBD in which all activities are supposed to take place. Residents choose their location (organised in a radial system) rationally to minimize commuting cost to their workplace. As a consequence of this rational choice this model leads to a general equilibrium structure with concentric residential areas surrounding the CBD, where the less productive sectors (with a lower bid rent function) are located at the city periphery.

This monocentric model, and its extensions, is one of the most frequently used models in urban economics, and has been readily used to observe regularity throughout the world, for a long time. However, as mentioned above, the transformation of urban structures, characterised by the formation of several sub-centres, gradually led to the emergence of many polycentric, rather than monocentric, urban areas. Thus, permanent changes in the urban structures, lead the formalization of more realistic scenarios, fulfilling the limitation of the traditional monocentric models. The co-existence of two different and opposite type of forces, centripetal and centrifugal forces, as underlies New Economic Geography (Krugman, 1999), creates incentives to leave the traditional agglomeration centres and leads to the emergence of new forms of spatial organization. Because, economic opportunities are distributed among the territory, and household and firm's choices are based on a set of complex set of decisions (labour, residential, capital, leisure, etc.), the CBD is no more the key factor in organizing space. Increasing returns to scale, transport costs, and the movement of productive factors and consumers, are the analytical ingredients of NEG, and are considered essential to recognise the existence of spatial agglomerations, and implicitly, to understand and to explain emerging forms of urban structures and how they interact. This implies that agents (firms and the population in general) will prefer to locate where the market is more accessible. While these economic theories are quite comprehensive and are based on rigorous micro-economic foundations of geographical economics as well as modern tools of economic theory, they still suffer from a lack of empirical studies which asses this phenomenon (Fujita and Krugman, 2003). The collection of relevant data is an important and vital barrier since the compilation of historical data at low spatial disaggregation (cities, which do not have any administrative correspondence) proves to be complicated.

Thus, the notion of economies of agglomeration and the concept of externalities, well explained in New Economic Geography (Fujita, Krugman *et al.*, 1999; Krugman, 1999), occurred at different scales, making the process of understanding the spatial patterns and spillovers more complex. If the notion that space is not immutable (some centres can change and others can appear) is introduced, the difficulty of delimiting and understanding spatial determinantes increase. Moreover, because human beings tried to reduce the constraints of the first-nature features (intrinsic characteristics of a site which are exogenously given), thus generating second-nature logic (emerging as the outcome of economic agent's actions to improve the first, and thus, being endogenously produced), in many cases, independent from natural advantages, the understanding of the spatial distribution of economic activities and households becomes less clear and obvious. The reduction of spatial and physical barriers to trade and to circulation of people has, in fact, contributed to distances becoming less relevant to locational decisions.

The interaction between attractive (agglomeration) and repulsive (dispersion) forces, resulting in a more complex polycentric urban structure, and the endogenous notion of analysing space, in line with the social lived space of Lefebvre (explored below) justifies the development of methodologies beyond the pure notion of geometric distances. In other words, urban spatial structures, the interactions between and within them, and the urban scale where these facts are analysed, cannot be based on a pre-given or fixed platform for social and economic relations.

The following table summarises the main urban economic models, monocentric, polycentric and New Economic Geography (NEG), and its connection with spatial heterogeneity (SH), spatial interaction (SI) and spatial scale (SS).

Table 9 - Summary of urban economic aspects and its connection with SH, SI and SS

Main transformations	▪ Urban economic models have been developed to describe and explain spatial urban structures: from the simple monocentric model of New Economic Geography literature, passing through to the polycentric models. ▪ Some unrealistic assumptions (in terms of heterogeneity and interactions) have been considered to make models feasible and comprehensible.
Spatial heterogeneity	▪ Monocentric models: The space is considered isotropic with a very simple structure; the city has a single city centre. The spatial pattern is precisely defined by concentric uniform rings resulting from the gradient of land rents. The main goal of these models is to define different land use categories. ▪ Polycentric models: Agglomerative forces (externality effects) are considered responsible for the existence of several centralities which reflect a more complex spatial configuration. ▪ NEG: Spatial delimitation is not the key issue in this kind of models, instead, increasing returns to scale, transport costs, and the movement of productive factors and consumers are considered the analytical ingredients of NEG.
Spatial interaction	▪ Monocentric models: Very simple spatial interactions, between work and place of residence are considered, organized in a radial system (bidirectional relationship in a isotropic space between CBD and periphery). ▪ Polycentric models: The increasing decentralisation of economic activities leads to a more complex cross commuting logic between several centres (multidirectional relationship in a isotropic space between various CBD's and periphery). ▪ NEG: Circular and cumulative causations responsible for agglomerations creates mutiple equilibria of spatial interactions.

(cont)→

<u>Spatial scale</u>	▪ <u>Monocentric models</u>: The character of space is analysed at micro level but can be generalised for a larger scale (e.g.: the application of the Von Thünen model to explain Australian wheat farmers in the context of the world) ▪ <u>Polycentric models</u>: Since different externalities operate on several scales, more complex models appear. One simple scale might not be sufficient to capture the nature of spatial interactions and spatial patterns across space. ▪ <u>NEG</u>: Economic agglomerations occur at many geographical levels; therefore, NEG provides a kitbag of tools for analysis at three levels: international, national and urban system models.

<u>The space in urban geography</u>

As has been seen in the Chapter II.2, and summarised above, space has been understood in a variety of distinctive ways in the context of urban economics. In urban geography the scenario is not any different. Opposing views of an understanding space have been emphasised in the different areas of geographical thought. Initially, a dual perspective was considered. On the one hand, space is measured by proximity or distance processes, related with factual circumstances, such as, territorial border and racial segregation across the territory. Labels like *physical, objective, concrete* and *real space* can be used to characterise this type of space. On the other hand, space embraces mental constructions, imagined distances related to different cultures and surrounding environmental perceptions. In this case, *mental, subjective, abstract* and *cognitive* space, are the labels used.

The problem regarding this division is that different spatial dimensions cannot be treated as independent or separate from each other (see, Soja, 1989). Because of this, some authors, Henri Lefebvre (1974 [1991]) and David Harvey (1973; 1990; 2000; 2006), have suggested that a three-way division of space be considered. Both Lefebvre and Harvey developed and departed from the Marxist tradition in political geography but they have each developed a distinctive thinking approach for urban questions and how space is understood and produced.

Lefebvre developed a *unitary theory of space* suggesting three different categories of space, each one being socially produced, framed by: i) *spatial practice* (nature); ii) *mental space* (formal abstractions about space); and iii) *social space* (the space of human action and conflict and sensory phenomena)[100].

Evidently important and determinant insights, from Lefebvre's thought, emerged from this research: the first being that space involves not only a concept of physical dispositions, but is also socially lived and socially produced, making a clear distinction with standard geographical and geometrical views of space. Lefebvre conceives that space is not a container, but rather, a set of interactions between subjects, their actions and their environment, giving the notion that space depends both on physical and mental constructs. The second message is that, even if space may not change much, our perception of it does, becoming finer, subtler, more profound and differentiated (Elden, 2004). Both of these arguments provide support for the idea of bi-dimensional Euclidean and multi-dimensional non-Euclidean notions of space.

Many other geographers and philosophers have been inspired by Henri Lefebvre and by his notion of space, considered socially produced, meaningful and lived. The most prominent author is David Harvey, starting similarly from Marxist traditions, who developed Lefebvre's ideas further. In his essay "*Space as a Keyword"*, Harvey breaks down space into: i) *absolute*, ii) *relative*, and iii) *relational space*. Roughly speaking, *absolute space* is fixed and largely supports the work of positivistic and quantitative geographers that believe in pre-existing physical laws that can be scientifically measured based on the essence of Euclidean geometry. *Relative space* is a differentiated, rather than homogeneous, space and allows for multiple geometries resulting from different perceptions and interpretations of individuals. And finally in the *relational view of space*, the external influences within the context are internalised, implying application of the idea of internal relations (Harvey, 2006).

---

100 A more detailed explanation of why these three aspects of social space are not isolated expressions of spatiality is presented in the section II.3.

Measurement becomes more problematic the closer it moves toward a context of relative or relational space (associated with multi-dimensional non-Euclidean).

Harvey's seminal work also developed Marx's own analysis of spatialities of capitalism and capitalism's social formation and draw out implications for land use patterns and locational dynamics. More precisely, this Marxist geographer, stressed that production (and indeed all economic activities) are spatially organised, following the principle of the power of capital over labour and circulation of capital. Harvey remains firmly within the boundaries of traditional Marxism, assuming a reductionist position of how space is structured, and therefore assumes space as exogenously given, in line with the first of nature's spaces. The class reductionist view (also stressed by Castells, 1983) takes the fixed and axiomatic position that every individual belongs to a particular class (referring to a group of people sharing, at least potentially, a single political interest). The boundaries of such classes and fractions can be accurately defined and that people's interests follow from the class to which they belong. Therefore, in this view, class provides the basic criteria for action, and all groups and interests can be assigned to one of the two economic categories: capital and labour.

This reductionist and deterministic perspective of reducing social differences to classes has been the subject of intense criticism from other Marxist geographers, such as, Henri Lefebvre (1974 [1991]), Doreen Massey (Massey, 1991) and Edward Soja's (1989) seminal works. For these authors, the space is endogenously produced by the capitalist system. Hence, space is not a given, it is theorized as socially produced, being a part of a second nature, as suggested by Soja (1989). For instance Lefebvre (1974 [1991]) rejects the idea of absolute and abstract space arguing that every society produces its own space (non-reductionist perspective).

The non-reductionism (post-Marxist) approach considers that the behaviour of groups cannot be understood by a simple class analysis; because classes are no longer viewed solely as an economic position, an individual can belong to more than one group. Thus, in Lefebvre's view of space every society and every mode of production produces its own space (Lefebvre, 1974 [1991]), characterised by alliances between different such classes, which are no longer viewed exclusively as an economic position but as a group of people that share common political and cultural interests. Further, introduction of the concept of pressure groups (such as political parties) that may represent a variety of different class positions (not just capital and labour) imply that class interests can crystallise around forces other than the fundamental power relation between capital and labour. Indeed, in current civil societies, the concepts of class fraction or faction have been substituted by the terms strata, communities, groups, populations and interests.

As has been shown above, within the Marxian tradition there are many aspects of space, and plenty of controversies. On one side, the notion of considering space as precisely defined, is advocated in the work of Harvey and Castells. These Marxists thinkers consider space as socially produced and spatially organized, but because every economic agent belongs to a particular economic category (labour and capital) the boundaries of such classes, that is, the definition of spatial heterogeneity, can be accurately delimited, becoming much closer to the notion of geometric space. On the other side, space considered to be continuously produced through social and spatial relations means different things to different people; it can affect people in different ways.

Therefore, the analysis of spatial patterns (heterogeneity) and spatial interaction (dependence) should seek to embrace the complexity and uncertainties of the space socially produced. As quoted Whitehead, (2003), new patterns of social mobility, economic integration and cultural interaction, contribute to a dislocated, spatial interpretation of space, from its geographic coordinates (it can move from one scale to another), underlying the fact that, many aspects and forms of reductionism and geometric understandings of space should be avoided.

The following table summarises the main urban economic models, monocentric, polycentric and New Economic Geography (NEG), and its connection with spatial heterogeneity (SH), spatial interaction (SI) and spatial scale (SS).

Table 10 - Summary of urban geography thoughts and its connection with SH, SI and SS

Main transformations	▪ Opposing views of an understanding of space have been emphasised. ▪ From an absolute notion of space to a relational space (passing through a relative notion of space). ▪ From the reductionist to a non-reductionist view of space.
Spatial heterogeneity	▪ Absolute space: Space is reduced to the essence of geometry, and as a result, boundaries and spatial segmentation can be precisely defined. ▪ Relative space/reductionist: Space is socially structured in terms of classes (labour and capital - Harvey's reductionist view of understanding space), therefore, boundaries of such fractions can be accurately defined. ▪ Relational space/ non-reductionist: Space is socially produced and spatially organised as a product of social, political and economic process (Lefebvre's non-reductionist view of analysing space), therefore, boundaries cannot be precisely defined because of the alliance between classes – groups of people that share common political and cultural interests.
Spatial interaction	▪ Absolute space: Pre-existing physical laws can be scientifically measured and the contents of space are unquestionably understood as being natural and given. ▪ Relative space/reductionist: Because space is a set of social, political and economic relations the analysis of spatial spillover is far too simplistic but reasonably possible: reducing social differences to classes (space is exogenously given). ▪ Relational space/non-reductionist: This space is not a container, but rather, a set of interactions between subjects, their actions and their environments. The behaviour of groups cannot be understood by a simple class analysis; because an individual can belong to more than one group (space is endogenously produced).

(cont)→

Spatial scale	▪ Absolute space: Historically, the neighbourhood has provided an important unit of analysis of urban space. ▪ Relative space/reductionist: Since various spaces emerge in different social contexts, neighbourhoods are being dislocated from their geographic coordinates. ▪ Relational space/non-reductionist: Contradiction between global integration and territorial redifferentiation that leads to a generalised explosion of spaces in which the relations among all geographical scales are consciously rearranged and reterritorialised (Lefebvre, 1979).

The space in urban planning

Three different paradigms of planning approaches, essential to characterise the role of space, during the twentieth century, can be defined (Jenkins, 2007): i) plan-making, ii) rational decision-making; iii) political decision-making.

For almost 20 years following the Second World War, not only in Britain but also in much of Europe and in North America (westerns democracies) the idea prevailed that planning theory and practice were to be viewed essentially as an exercise in physical design, "(...) *as an art, albeit (again like architecture) an applied or practical art in which utilitarian or functional requirements had to be accommodated* (Taylor, 1998, p.159)". The first paradigm, planning-by-design relied on the idea, of representing the segmentation of space by different types of use (zoning) and activities. The main goal was often to limit the city growth, and for this purpose master plans and regional plans were used. This is considered a very linear view because it is limited to physical descriptions of space. The role of planners was seen as not only being expert and apolitical, but also imaginative and visionary, able to take a creative leap from the analysis of surveys to the making of plans (Davoudi and Strange, 2009). These concepts illustrate the positivist approach to the process of planning. For example, although the utopian visions of the *Garden City* (created by Ebenezer Howard) and *La Ville Radieuse* (conceived by Le Corbusier) envisage physical outcomes, there are also elements of positivist interpretations of space (Davoudi and Strange, 2009). In this particular context, space was seen as a neutral container, a blank canvas that is filled with human activity (Hubbard, Kitchin *et al.*, 2004). Some other characteristics of this planning process can be mentioned as follows: planning processes ended with the production of one single plan rather than a number of alternative strategies (Taylor, 1998); planners themselves have been seen as the guardians of public interest (Hall, 2002); and based on the infallibility of experts, reinforcing the apolitical, technical nature of the process (Batty, 1979). This perspective of the plan-making process is noticeably different from its successor, rational planning, mainly into two aspects. First, because planning-by-design was limited to physical descriptions, and second the role of planners was seen as being expert, imaginative, visionary and apolitical, able to interpret all his ideas and put it into a plan (Davoudi and Strange, 2009).

In the late of 1960s a conceptually distinct view of planning (called second paradigm) emerged representing a rupture with the previous tradition. A systems and rational-process view of planning appeared in a reaction against the rigidity and limited scope of detailed master and regional plans. The former (the systemic) was based on a view of the object, whereas the latter (the rational) was concerned with the process of planning itself. There are evident differences between these two planning theories; nevertheless, both views represented a rupture with the established design-based view. Roughly speaking, this second paradigm of thought refers to a rationalist approach and structural plans were the key planning instruments necessary to understand and plan urban areas as sets of systems (transport, economic, etc.). The fragmented and segmented growth of cities required the definition of strategic frameworks to understand the complexity of forces acting in the space. This is considered within social sciences as being structuralist thinking, which tries to identify hidden laws and forces, beyond powers of observation, seen as determinant of human behaviour (Hollis, 2003; Davoudi and Strange, 2009). In line with the non-reductionisnt view of space, where space is endogenously given, this perspective assumes that the social and psychological are important factors to take into account.

The third paradigm (mainly in the 1990s), and current trend in practices, emerged from the idea that planning is a political decision-making process and is based on a participatory approach. The role of the planner is reduced to that of facilitator legitimising actions by state and civil society (Dear and Flusty, 2002 *cited in* Davoudi and Strange, 2009). The search for place identity and cultural specificities (people-centered) and the primacy of mixed and flexible land use, are outlined by Hirt (2005), as the main concerns of this post-structuralism (post-modernism) thinking of planning. The space is not made by structures, hence by relations, emphasising the link between social and spatial relations. Planners advocating this model focus on existing unequal relations and distribution of power, opportunity and resources, and their goal is to work towards structural transformation of these systemic inequalities through empowerment. Different spaces are no more seen as hierarchical (moving from a macro scale, both at national and sub-national level, to an urban local approach) but as nodes in relational settings. It is the significance and composition of relations that define the scale (Murdoch, 2006).

Chronologically major trends in urban planning practice occurred. Starting from a positive view (the desired to create a spatial order) planning rapidly evolved to a post-structuralism perspective, focusing on space and place as socially produced. Consequently, the understanding of space has also clearly changed over time, in the sense that the idea of dematerialization of planning regained prominence, that is, the apparent physical substance of the notion of space has diminished, becoming more immaterial. Again, the notion of multi-dimensional non-Euclidean distances to access spatial interactions is underlined by the structures determined by social actions.

The following table summarises the main ideas beyond the urban planning practices, from the first to the third paradigms of planning, and its connection with spatial heterogeneity (SH), spatial interaction (SI) and spatial scale (SS).

Table 11 - Summary of urban planning thoughts and its connection with SH, SI and SS

Main transformations	▪ Chronologically major trends in urban planning practice occurred. ▪ Starting from a positive view (the desired to create a spatial order) planning rapidly evolved to a post-structuralism perspective, focusing on space and place as socially produced.
Spatial heterogeneity	▪ Plan-making process/Planning-by-design/positivism: The segmentation of space into different types of use (zoning of land use) and delimitation of the growth of the city are restricted to physical descriptions of space. ▪ Rational decision-making/system planning/structuralism: The spatial segmentation is important to define interactions in the urban system but delimitations of these spaces are not considered relevant. ▪ Political decision-making/planning as negotiation/post-structuralism: The spatial segmentation is perceived as a complex system of relations, where their limits are not precisely defined.
Spatial interaction	▪ Plan-making process/Planning-by-design/positivism: No spatial interactions are considered. Planning practices ended with the production of the plan showing how land use and activities are distributed over space. ▪ Rational decision-making/system planning/structuralism: In reaction to the rigidity and limited scope of master and regional plans, space started to be analysed as a set of interdependent systems or structures. ▪ Political decision-making/planning as negotiation/post-structuralism: The space is resultant from structures but by social relations, emphasising the intangible factors of space.

(cont)→

Spatial scale	▪ Plan-making process/Planning-by-design/positivism: The macro scale approach is emphasised and the local specificities were not taken into account. ▪ Rational decision-making/system planning/structuralism: The search for hidden laws and forces for analysing cities as integrated systems leads to the consideration of a multitude of scales. ▪ Political decision-making/planning as negotiation/post-structuralism: Different spaces are no more seen as hierarchical and ambiguous, bridging the hiatus between where the policies take place and where the measures should be applied.

V.3. Spatial econometric methods to analyse space

In the spatial econometric literature, two aspects are mentioned as important in analysing and understanding the determinants of space in housing markets, spatial heterogeneity and spatial spillover. Spatial heterogeneity (or spatial pattern), which is related to structural differences between housing markets, refers to the idea that different hedonic housing amenities may be valued differently in different locations and in different housing submarkets. On the other hand, spatial dependence (or spatial interactions), are connected to the spillovers that may exist across and within submarkets, that is, the degree by which price increases (or decreases) in a given submarket is influenced by that that occurs in other submarkets, and also other properties within the same submarket.

A third important aspect of space, not directly tackled in the spatial econometric literature is that of spatial scale. Since, both spatial heterogeneity and spatial dependence can move from one scale to another, assuming different configurations, the definition of an appropriate scale should be considered essential to capture the real meaning of the urban spatial structure. Spatial scale, being a set of interlinked hierarchies (as quoted by Smith, 1992), is important to define and characterise submarkets; and in turn, submarkets are a key issue for assessing the existence and the direction of spatial spillovers, particularly, when analysed in terms of multi-dimensional non-Euclidean notions of space.

Spatial scale, not being so much purely an econometric issue (it appears as a negligible factor), but rather an important empirical question, has been widely discussed in the literature of urban economics, for example, in Malpezzi (2003). The definition of the most appropriate territorial level (in terms of disaggregation and scope) to capture the relevant aspects of spatial patterns and spatial interactions is a key issue.

The common practice, widely stressed in the spatial econometrics literature in spite of problems of heterogeneity and dependence, of representing the spatial interactions using a weight matrix (W) *a priori* defined, modelled by functions of distance or contiguity (Anselin and Griffith, 1988; LeSage and Pace, 2009), has often been expressed as inadequate. The main reason is that spatial dependence (spillovers or interactions) may be driven by non-tangible factors or simply by multi-dimensional non-Euclidean notions of space (widely emphasised throughout this chapter), not captured by the traditional approach.

Thus, in Chapter VI a methodology is presented that allows treating the two most commonly discussed features of spatial data in terms of a multi-dimensional non-Euclidean perspective, mainly spatial dependence effects where pre-assumed and fixed spatial weighting matrices are not assumed.

VI. Methodological approach to analyse multi-dimensional non-Euclidean space in housing markets

Since, spatial interactions may be driven by other intangible factors, the choice of a spatial weights matrix based purely on bi-dimensional Euclidean distances might be inappropriate. Thus, in line with a multi-dimensional non-Euclidean notion of space, stressed in Chapter IV, a methodology to estimate an unknown spatial weights matrix is presented and discussed below in further detail. The proposed approach is based on factor analysis and on a multi-dimensional non-Euclidean notion of space to understand spatial hedonic models. This method combines spatial hedonic analysis based on orthogonal factors with a method for extracting inferences from an unknown spatial weights matrix under the structural constraint of symmetric spatial weights. This methodology is first developed by Bhattacharjee and Jensen-Butler (2005) for panel data and applied in the context of this work using cross section data (see also Bhattacharjee and Jensen-Butler, 2011; Bhattacharjee, Castro *et al.*, 2012).

Thus, the estimation of an unknown spatial weights matrix using cross section data is based on the spatial autocovariance matrix of residuals across the submarkets, and proceeds as follows:

1. *Fix a spatial scale* and identify the important housing hedonic characteristics:

    a. Collect data on *n* characteristics, combining both location and housing attributes (H_n)

b. Define the main dimensions of housing attributes through statistical factor analysis (F_k), where k<n
c. Build a hedonic model using factors: Factor based Hedonic Pricing Model

2. *Spatial Heterogeneity* across submarkets:
   a. Identify housing submarkets (Z_i) at the given spatial scale, aggregating the smallest units of analysis by two methods:
      - Aggregation based on expert knowledge of submarkets
      - Aggregation based on cluster analysis
   b. Compute factor based hedonic pricing model for each submarket and compare the hedonic coefficients across the submarkets
3. *Spatial Spillovers* between housings and submarkets:
   a. Spatial dependence and *a priori* spatial weights
      i. Global indicators of spatial autocorrelation
      ii. Spatial error and spatial lag models
   b. Compute the spatial autocovariance matrix of residuals across the submarkets (Z_i)
      i. Choose a property (H_i) in one submarket; ex.: market 1
      ii. Identify another house in each of the other submarkets (Zi where i ≠ 1) which is close to this property, that is, holds similar characteristics (the smallest Euclidean distances of the factors [$min \sum(FH_1\text{-}FH_j)$])
      iii. Use the residuals from the matched houses to compute the spatial cross-market autocovariance matrix
   c. Estimate unknown spatial weights:
      i. Cross-submarket spatial weights are estimated from the matrix of spatial autocovariances using the methodology developed in Bhattacharjee and Jensen-Butler (2005; 2011), under the assumption of symmetric spatial weights.
      ii. Spatial econometric models are estimated by maximum likelihood (ML) assuming different within-submarket spatial weights.
4. *Fix different spatial scale, return to 1.*

VI.1. Housing hedonic characteristics

As described in Chapter IV, based on the seminal work of Lancaster (1966) and Rosen (1974), an hedonic pricing model is one that decomposes the price of an item into separate components that determine the price. In the specific application of housing, dwelling unit values (or proxies such as prices or rents) are regressed on a bundle of characteristics of the unit that determine that rent or value. The hedonic regression assumes that the determinants of a unit's value are known:

$$P = f(F, E, L, S, T) \qquad \text{Eq. 29}$$

Where, P denotes the value of the house (price, or price per unit area), and F, E, L, S and T denote respectively, structural characteristics of the dwelling; environmental and neighbourhood characteristics; location within the market; other characteristics (access to utilities and public services, such as clean water supply, electricity, central heating, etc.); and the time (date, month) when value is observed.

Estimating the hedonic price function using a collection of observed housing values and dwelling unit characteristics, yields a set of implicit prices for housing characteristics that are essentially willingness-to-pay estimates. This allows analysis of various upgrading scenarios, targeted to specific subgroups, defined either by socio-economic characteristics or by location. Thus, the model facilitates an understanding of residential location, and therefore urban structure, and provides valuable input towards urban planning and housing policy.

As mentioned before, theory provides no guidance as to the functional form appropriate for hedonic regression. However, a non-linear hedonic function is useful for recovering the underlying structural demand curve from estimates of the hedonic relationship (the reduced form). Arguments like the above, as well as several other important advantages, motivated Follain and Malpezzi (1980) to recommend the semi-log form[101]; for further discussion and applications, see Follain and Malpezzi (1980) and Malpezzi (2003).

Several housing characteristics can be included on the right hand side of a hedonic regression model. Unfortunately, coefficient estimates are not robust in the face of the omitted variables problem (Butler, 1982; Ozanne and Malpezzi, 1985). However, the same correlation between omitted and included variables that biases individual coefficient estimates often aids better prediction from a sparse model (Malpezzi, 2003).

This feature of the hedonic pricing model enhances the possibility of exploiting the factor structure to obtain parsimonious estimates and improved predictions. Several studies, beginning with Kain and Quigley (1970) and Archer and Wilkinson (1973) have taken this approach, and Davies (1974) has combined factor analysis with the regression approach. In a critical review, Maclennan (1977) suggests that the extracted factors may reflect statistical properties rather than behavioural collections of housing characteristics.

In this methodology, a hybrid approach, combining factor analysis with regression is proposed. Initially, a small collection of leading factors from a large number of potential hedonic characteristics, are identified[102]. At the second stage, factor values for all properties are predicted, including those for which some hedonic characteristics are missing, and these predicted factors are used to estimate the hedonic regression model.

The use of the cross section factor model in this context has several advantages. First building a hedonic model based on a small number of factors, rather than a large collection of housing characteristics, leads to a parsimonious model with more precise estimation, offering as a result a better interpretation of the regression coefficients. More importantly, in including all potential economic factors affecting prices, the factor based regression model is less susceptible to

[101] Where, logarithm of price per square meter of living space is regressed on logarithm of house area, conditioned by several other hedonic housing characteristics.

[102] This methodology is applied to two datasets in Part 4 of this thesis and the results show that the (orthogonal) factors are clearly identified with interpretable collections of housing characteristics, such as structural dimensions, access to utilities, centrality and access to local services.

the omitted variables problem. Second, hedonic regression based on factors allows the unique possibility of addressing missing value problems, where factors can be predicted (imputed) using the information available on only a selection of included characteristics, under the assumption that the missing data are allocated randomly across the properties and conditional on the values of observed features. This leads to considerably larger sample sizes for estimation of the hedonic model, with clear benefits of improved precision of the estimates. Third, the approach based on orthogonal factors is not subject to multicollinearity, and could therefore contribute to higher efficiency, which in turn can lead to better prediction of housing prices. Finally, and most importantly, the orthogonality of factors is crucial for the proposed methodology, both to analyse spatial heterogeneity (in submarkets with a low number of degrees of freedom this allows the capture of the main dimensions of housing characteristics), and spatial dependence (for estimating the unknown symmetric spatial weights matrix, based on the spatial autovariance matrix). See Marques and Castro (2007) and Bhattacharjee, Castro *et al.* (2012) for further details of the methodology and application.

VI.2. Spatial heterogeneity across submarkets

The conceptual notion behind spatial submarkets discussed above implies that the price determining (hedonic) mechanism can be heterogeneous over space (for further discussion see Section IV.1.2). Such spatial heterogeneity can originate from characteristics of the demand and supply of factors, institutional barriers or discrimination, each of which can cause differentials across neighbourhoods in the way that housing attributes are valued by consumers and house prices are determined (Anselin, Lozano-Gracia *et al.*, 2010).

The standard urban model in the Alonso-Muth-Mills tradition predicts a generally declining pattern of prices with distance from the centre of the city, though there may be spatial variation in relative preference for centrality. Other models based on localised amenities or multiple centres imply a stronger impact of access to local amenities. Like distances, the implicit prices for dwelling characteristics and size may also vary spatially, reflecting either supply constraints or residential sorting. Follain and Malpezzi (1980), Adair, McGreal *et al.*, (2000) and Soderberg and Janssen (2001), among others, have examined intra-urban variation in the price of housing using hedonic models.

As explored in more detailed in Chapter VII, and in line with the literature presented in the Section IV.1.2, three different strategies are considered to delineate housing submarkets: i) the use of traditional administrative boundaries, ii) identification of criteria, defined *ex ante*, considering several urban dimensions (demography, history, morphology, socio-economic), and iii) cluster analysis.

Hedonic pricing model coefficients, using factors, are allowed to vary across submarkets, and this estimated variation is used to infer residential neighbourhood choice and urban spatial structure.

VI.3. Spatial spillovers between housing and submarkets

In contrast to spatial heterogeneity, spatial dependence leads to spatial autocorrelation, implying that prices of nearby houses tend to be more similar than those of houses that are farther apart. Likewise, the average price of houses in nearby or related submarkets may be correlated more strongly. A common explanation for spatial autocorrelation is spatial spillovers or other forms of contagion effects. However, incorrectly modelled spatial heterogeneity, measurement problems in explanatory variables, omitted variables, and unmodelled features that show a clear spatial pattern can also lead to spatial autocorrelation (Anselin and Griffith, 1988). Recent empirical literature has addressed issues of bias and loss of efficiency that can result when spatial effects are ignored in the estimation of hedonic models,[103] and the use of spatial econometric models to address spatial autocorrelation is becoming increasingly standard.[104]

The usual approach to representing spatial interactions is to define a spatial weights matrix, denoted W, which typically represents a theoretical and *a priori* characterisation of the nature and strength of spatial interactions between different submarkets or dwellings.[105] These spatial weights represent patterns of diffusion of prices and unobservables over space, and thereby provide a meaningful and easily interpretable representation of spatial interaction (spatial autocorrelation). The spatial weights are typically modelled as functions of geographic or economic distance. The distance between two spatial units reflects their proximity with respect to prices or unobservables, so that the spatial

[103] See, for example, Pace and LeSage (2004) and Anselin and Lozano-Gracia (2008).
[104] For representative applications using different hedonic models in a spatial econometric setting, see Basu and Thibodeau (1998) and Anselin, Lozano-Gracia *et al.* (2010).
[105] For a setting with n spatial units under study, W is an $n \times n$ matrix with zero diagonal elements. The off-diagonal elements are typically inversely proportional to the distance between a pair of units, so that spillovers between a pair of units that are farther apart are also lower.

interaction between a set of units (dwellings) can be represented as a function of the economic distances between them.

Given a particular choice of the spatial weights matrix, there are two important and distinct ways in which spatial interaction is modelled in spatial regression analysis, the spatial lag model and the spatial error model. In the former, the hedonic regression includes as an additional regressor the spatial lag of the dependent variable (p), in this case price, represented by Wp, and the regression errors (ε) are completely idiosyncratic. By contrast, in the latter case, the regression errors are spatially dependent on their spatial lag, $W\varepsilon$.

The implications of spatial interaction on estimation of these two models are different. In the spatial lag model, the endogenous spatial lag implies that OLS estimates that do not account for spatial interaction would be biased, while in the spatial error model, they will be unbiased but inefficient. Though different in interpretation, the above two models are very difficult to distinguish empirically (Anselin, 1999; 2002). In line with current practise in the area of spatial econometrics, the hedonic pricing model under the spatial error assumption is estimated first. Next, to judge whether endogenous spatial lags are relevant, a test for spatial lag dependence is performed by nesting the spatial error model within a hybrid model incorporating both spatial lag and spatial error dependence; for more discussion on sequential model selection in the spatial context, see Born and Breitung (2009).

The choice of appropriate spatial weights is a central component of spatial models as it imposes *a priori* a structure of spatial dependence, which may or may not correspond to reality. Further, the accuracy of these measures affects profoundly the estimation of spatial dependence models (Anselin, 2002; Fingleton, 2003). Spatial contiguity or suitable functions of geographic distances are frequent choices. However, spatial data may be anisotropic, where spatial autocorrelation is a function of both distance and the direction separating points in space (Simon, 1997; Gillen, Thibodeau *et al.*, 2001). Further, spatial interactions may be driven by other factors, such as trade weights, transport cost and travel time. The choice typically differs widely across applications, depending not only on the specific economic context but also on availability of data. The problem of choosing spatial weights is a key issue in many applications.

Given the above ambiguities regarding measurement of spatial weights, and in line with the notion of multi-dimensional non-Euclidean space potentially driven by factors different from geographic distance or contiguity, the spatial weights matrix (W) is considered as an unknown symmetric matrix with zero diagonal elements. Further, spatial interactions can be potentially negative, often implying segmented housing markets or asynchronous housing cycles. Based on a given definition of urban submarkets (or a fixed set of spatial locations) and panel data on these spatial units, Bhattacharjee and Jensen-Butler (2005) and Bhattacharjee and Holly (2009; 2011) have developed several methods to estimate the spatial weights matrix between the submarkets.[106] Here, the panel estimation methodology is extended in Bhattacharjee and Jensen-Butler (2005) under the structural assumption of symmetric spatial weights for a purely cross-section setting.

Specifically, a cross-section factor regression model is considered where each housing property i (i = 1,..., n) belongs to a unique submarket M_j, that is, $i \in M_j$. The price of this property p_i depends linearly on a vector of unobserved orthogonal factors F_i of housing and locational hedonic characteristics, where the effect of the factors potentially varies across the submarkets. The corresponding regression error ε_i is uncorrelated with the factors, but may be spatially related to the errors for other houses through an unknown spatial weights matrix, W. In other words, a spatial error model is considered with a cross-section heterogeneous factor structure across the submarkets, where the effects of the factors are potentially different across submarkets and there may be submarket specific fixed effects:

$$p_i = \beta_{0k} + \beta_k' f_i + \varepsilon_i, \qquad i = 1,2,\ldots,n, \quad i \in M_K$$

$$\varepsilon = \lambda W \varepsilon + \upsilon, \qquad \upsilon_i \sim N\left(0, \sigma_k^{\ 2}\right) \text{ independent,} \qquad \text{Eq. 30}$$

Here $\varepsilon = \left(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n\right)'$, the vector of the random errors, has a spatial error structure with an unknown spatial weights matrix W_{nxn} having zero diagonal

[106] See Bhattacharjee and Holly (2011) for a review and discussion, as well as an application to network interactions in a monetary policy committee setting. Note that spatial weight matrix (W) is estimated, as well the coefficient λ. The autocovariance matrix represents a transformation of W and λ simultaneously.

elements, and the zero mean Gaussian idiosyncratic errors (υ_i's) are potentially heteroscedastic across submarkets but independent over the cross-section and uncorrelated with the random factors F_i. The equation described above is a simplified version of the cross-section factor model with heterogeneous group effects discussed in Andrews (2005), with additional Gaussian assumptions. These distributional assumptions are useful in this case for drawing inferences by maximum likelihood on the intra-submarket spatial weights.

The spatial weights matrix W is the row-standardised version of W^o_{ij}, which is assumed to be symmetric and have a block-structure as follows:

$$W_{ij}^0 = \begin{cases} 0 & \text{if} \quad i = j \\ \omega_0 & \text{if} \quad i,j \in M_k, k = 1,\dots,J \\ \omega_{kl} & \text{if} \quad i \in M_k, j \in M_l, M_k \cap M_l = \phi. \end{cases} \qquad \text{Eq. 31}$$

The above weights matrix is unknown and quite general, allowing for unknown but fixed spatial weights between properties in the same submarket, and similarly unknown spatial weights between properties in any pair of submarkets.[107] For identification in the reduced form, it is required that $(I - \lambda W)$ is non-singular. Further, following Bhattacharjee and Jensen-Butler (2005), it is assumed symmetric spatial weights within and between submarkets; a standard assumption in the spatial econometrics literature. However, spatial weights are allowed to be negative.

In this methodology, a spatial error model is assumed. Because of endogeneity, estimating the spatial weights matrix under the spatial lag model is a very difficult problem. However, it is possible to perform specification tests against the spatial lag model under the assumption that the same spatial weights matrix W describes both spatial lag dependence and error dependence. For this purpose, the above spatial error model is nested within the following model that includes both spatial lag and spatial error, with different autoregressive coefficients:

---

[107] Note that, since the spatial weights matrix is unknown, it is necessary to row-standardize W to enable identification of both W and the autoregressive parameter (λ). The assumption that the intra-submarket spatial weight is the same across all submarkets is not necessary, but retained here for computational simplicity.

$$p = \rho W p + \beta' F + \varepsilon,$$
$$\varepsilon = \lambda W \varepsilon + \upsilon, \quad \upsilon_i \sim N(0, \sigma_j^2) \text{ independent,} \qquad \text{Eq. 32}$$

Borg and Breitung (2009) propose a regression based test, where as a first stage, the spatial error model is estimated. At the second stage, the test evaluates whether there is any residual spatial dependence that can be explained by spatial lag effects. This test is used to verify whether the spatial error model is adequate for our empirical applications. The test is simple to apply and has several advantages over standard LM tests; see Born and Breitung (2009) for further details.

As discussed before, the main methodological contribution is to estimate unknown spatial weights within a factor-based cross-section spatial error model. Next, a description of the estimation methodology is presented, in two steps: first, the cross-market spatial interaction matrix W^* (defined in equation below); and second, the cross-submarket spatial autocovariance matrix Γ.

In the panel data setting, the methodology in Bhattacharjee and Jensen-Butler (2005) is based on a given consistent estimator for the underlying hedonic regression model with spatial errors. Based on residuals from the above estimation, a consistent estimator $\hat{\Gamma}$ is first obtained for the $J \times J$ cross-submarket spatial autocovariance matrix

$$\Gamma = (I - W^*)^{-1} \Sigma (I - W^*),$$

$$W^* = \lambda \begin{bmatrix} 0 & \omega_{12} & \cdots & \omega_{1J} \\ \omega_{12} & 0 & \cdots & \omega_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{1J} & \omega_{2J} & \cdots & 0 \end{bmatrix}, \qquad \text{Eq. 33}$$

$$\text{and} \quad \Sigma = diag\left[\sigma_1^2, \quad \sigma_2^2, \quad \ldots \quad, \sigma_J^2\right]$$

Let us assume that such a consistent estimator for $\hat{\Gamma}$ has been obtained. Bhattacharjee and Jensen-Butler (2005) show how this estimator $\hat{\Gamma}$ can then be used to estimate the unknown cross-submarket spatial weights matrix W^*.

Without any structural constraints on the weights matrix, the estimation problem is only partially identified, up to an orthogonal transformation of interactions. Specifically, they show that the matrix

$$V = \left(I - W^*\right).diag\left[\frac{1}{\sigma_1}, \quad \frac{1}{\sigma_2}, \quad \ldots \quad , \frac{1}{\sigma_J}\right]$$

is consistently estimated, up to an arbitrary orthogonal transformation, by

$$\hat{\Gamma}^{-1/2} = \hat{E}.\hat{\Lambda}^{-1/2}.\hat{E}',$$

where $\hat{E}$ and $\hat{\Lambda}$ contain the eigenvectors and eigenvalues respectively of the estimated spatial autocovariance matrix $\hat{\Gamma}$.[108] In other words, $\hat{\Gamma}^{-1/2}$ is a consistent estimator of VT for some unknown square orthogonal matrix T. Since T is an arbitrary orthogonal matrix, it has $J(J-1)/2$ free elements. Hence, the spatial weights matrix W^* can be precisely estimated only under additional structural constraints. Symmetry of the spatial weights matrix constitutes one set of valid identifying restrictions,[109] which is the structural assumption assumed.

Under the symmetry assumption, Bhattacharjee and Jensen-Butler (2005) describe inference methods and an algorithm for estimating the unknown spatial weights matrix. Estimation requires application of the "*gradient projection*" algorithm (Jennrich, 2001) which optimises an objective function over the group of orthogonal transformations of a given matrix; standard errors are obtained using the bootstrap.

This method can be applied to the spatial hedonic pricing model provided an initial consistent estimator can be found for the cross-submarket spatial autocovariance matrix Γ. In this work, a maximum likelihood method to estimate this autocovariance matrix is proposed. Estimation is based on the factor-model and Gaussian error assumptions. Since there is a unique relation between Γ

---

108 Here, $A^{1/2}$ denotes the symmetric square root of a positive definite matrix A, and $A^{-1/2}$ denotes its inverse. In other words, $A^{-1/2}$ has the same eigenvectors as A, but with the eigenvalues replaced by the reciprocal of the square root of the corresponding eigenvalues of A.

109 See Bhattacharjee and Holly (2011) for further discussion on partial identification and structural constraints in this context.

and the corresponding W^*, the estimates of cross-submarket spatial weights are therefore also maximum likelihood.

In the panel data setting, Bhattacharjee and Jensen-Butler (2005) estimated the underlying regression model and obtained residuals, and then estimated $\hat{\Gamma}$ as the simple sample covariance matrix of the cross-market residuals. This step was relatively simple because for each time period, there was a residual uniquely identified with each submarket.

In the current cross-section setting, the situation is more complex because *a priori* there is no natural way to associate a house in any one submarket with a corresponding house in any other submarket. For this matching problem, an analogy of the current cross-section factor model (see, for example Andrews, 2005) is used, with the multifactor error structure of cross-sectionally dependent panel data inherent in the common correlated effects methodology of Pesaran (2006).

In the common correlated effects approach (Pesaran 2006), linear combinations of unobserved common factors are approximated by cross-section averages of the dependent and explanatory variables, which are then included in the panel regression model in addition to the usual regressors. Clearly, these common factors can be alternatively modelled by a full set of time fixed effects.

Bhattacharjee and Jensen-Butler (2005) uses residuals across spatial units for the same time period to estimate the spatial error autocovariance matrix. The multifactor spatial error model provides a clear justification for this approach. Residuals for the same period are matched because the corresponding observations on different spatial units align perfectly along the dimension of the unobserved latent factors; in the panel data setting, the time specific common shocks. Taking this intuition to the pure cross-section setting, it is therefore natural to match housing property i in submarket M_j with the dwelling j in another submarket M_j that bears the closest correspondence in the vector of latent factors; in our case F_i and F_j. Thus, the proposed methodology proceeds as follows.

At the first stage a suitable set of orthogonal factors is estimated based on hedonic characteristics. Using these estimated factors, an estimation of the hedonic regression model is presented separately for each submarket, allowing for full spatial heterogeneity. Based on these submarket specific regression estimates, residuals for each property are obtained.

Properties across submarkets have been matched in the second step. Specifically, the residual for an index dwelling i in submarket M_j, is matched to the residual for that house j in submarket M_j that has the closest match in the vector of estimated factors; in other words,

$$j = \arg\min_{j^* \in M_j} \left(F_i - F_{j^*}\right)' \left(F_i - F_{j^*}\right) \qquad \text{Eq. 34}$$

In the third stage, based on matched residuals across the different submarkets, the cross-submarket spatial autovariance matrix is estimated simply by the sample covariance matrix $\hat{\Gamma}$. Finally, estimation of W^* follows using the Bhattacharjee and Jensen-Butler (2005) methodology outlined previously.

The assumption of the multifactor model is crucial for this estimation procedure. First, residuals from an estimated hedonic pricing model would be extremely susceptible to the problem of potential omitted variables. In practical terms, it is very difficult to avoid this problem, even if a large number of hedonic characteristics are included in the estimation. By contrast, in estimating the factor model, it is simpler to minimise this problem by including factors corresponding to all notional elements, that theory and past studies have identified as determinants of prices. Therefore, one can assume that what remains in the error is uncorrelated with the included factors.

Second, and as discussed above, the factor model is conceptually very closely aligned to the critical distinction between spatial strong and weak dependence, and therefore to the common correlated effects approach (Pesaran, 2006). Specifically, in the panel data setting, the theoretical justification for matching residuals corresponding to different spatial units for the same time period is that they match in the strong dependence (or, hidden factor) dimension.

Matching against estimated factors provides an exact conceptual counterpart in the cross-section factor model setting.[110]

Third, since estimated factors are orthogonal by construction, it is straightforward to match two properties in different submarkets by the inner product (sum of squares) of the vector of difference of their corresponding estimated factors.

Finally, under the assumptions of the factor model, $\hat{\Gamma}$ estimated as above is the maximum likelihood estimator of the spatial autocovariance matrix (see also Andrews, 2005), and therefore it is also the corresponding estimator of the cross-submarket symmetric spatial weights matrix.

Recent empirical econometric work has addressed the potential bias and loss of efficiency that can result when spatial effects are ignored in the estimation of hedonic models; see, for example, Pace and LeSage (2009) and Anselin and Lozano-Gracia (2008). Spatial patterns in the housing market arise from a combination of spatial heterogeneity and spatial dependence (Anselin, 1988). Additionally, choice of an appropriate spatial scale is important (Malpezzi, 2003). There follows a discussion of spatial issues in the construction of the hedonic pricing models, including all of the three above aspects of space.

[110] See Bhattacharjee and Holly (2011) for further discussion of the conceptual distinction between strong and weak dependence and their link with the spatial weights matrix.

VI.4. Spatial Scale

Definition of submarkets is important at both conceptual and empirical levels. Housing markets are local and diverse, and hedonic price estimation requires careful consideration of this issue (Malpezzi, 2003). The definition of submarkets in practice ranges from the national or regional scale, through metropolitan areas, to levels below the metropolitan level (see Section V.1.2).

Malpezzi (2003) argues that one reason the metropolitan area is appealing as the unit of analysis is that these areas are usually thought of as labour markets, which may therefore be approximately coincident with housing markets. On the other hand, submarkets below the metropolitan level can be segmented by location (central city/suburb), or by housing quality, or even by race or income levels. Such segmentation facilitates both understanding of residential neighbourhood choice and devising appropriate urban housing policy. However, the empirical literature does not suggest an unambiguous definition of a unique spatial scale.

In this work, the analysis has been conducted at two different spatial scales, both disaggregated to a relatively fine spatial level. In the first, administrative regions (parishes) within the city of Aveiro as submarkets are considered, and pool the suburban area together into a single submarket. This definition aids understanding of spatial heterogeneity and interactions within the urban area, but does not provide satisfactory analysis in terms of spillovers between the city and the suburban area. Second, analysis has been extended to a finer spatial scale in the suburban area, constructing submarkets with careful consideration to the principles of segmentation discussed above. This analysis reflects some advantages of using a flexible spatial scale, since forces of agglomeration and dispersion operate differently at different scales.

.

PART 4

Empirical Application

VII. The role of space in the urban housing market of Aveiro

In this chapter an empirical approach to analyse the housing market of Aveiro (Centro Region of Portugal), focusing on three aspects of spatial analyses (spatial heterogeneity, spatial dependence and spatial scale), is presented. The methodology is based on the philosophy explained in Chapter VI, and follows the recently emerging literature related to inferences from unknown spatial weights matrices (Bhattacharjee and Jensen-Butler, 2005; Bhattacharjee and Holly, 2010a; b) and a multi-dimensional non-Euclidean notion of space to understand spatial interaction. As has been mentioned previously, the main methodological innovation of this analysis is the development of a factor analysis hedonic model to estimate unknown spatial weights matrix.

The starting point is a hedonic model explaining the value of a dwelling (equation 5: Section IV.1.):

$$\ln p = f(H, v) + \varepsilon$$

Where: p is the vector of the logarithms of house prices; v is the vector of hedonic prices, reflecting the willingness-to-pay for attributes affecting house prices; H is the design matrix quantifying the attributes of dwellings (both intrinsic and location characteristics)[111]; and ε is the vector of regression errors.

Two different databases, covering two different spatial scales, are used in this empirical work. One of them corresponds to a smaller spatial scale (micro scale approach), which embraces 166 properties sold through one real estate

[111] Note that, for the quantification of the matrix *H*, an appropriate collection of aggregate indicators of attributes are considered, extracted by statistical factor analysis, instead of the original hedonic attributes.

agency (ERA Portugal), in 2007, located in the urban and suburban area of the municipality of Aveiro (Section VII.2.) The other, a wider dataset (macro scale approach), covers a larger spatial scale and has information for 12467 properties (after data cleaning). These properties are distributed by urban, suburban and rural areas of Aveiro, corresponding to two different municipalities, Aveiro and Ílhavo (Section VII.3.)

Thus, spatial heterogeneity, spatial spillover and spatial scale are analysed quantitatively as follows:

i) Spatial heterogeneity (or spatial patterns) is related to the market segmentation, where inhabitants in different housing submarkets potentially value the housing characteristics differently. In other words, the shadow prices (or slope parameters) estimated from the regression model (v) are not necessarily constant across space (z), leading to structural differences in various housing markets that are expressed as follows:

$$\ln p_z = f(H, v_z) + \varepsilon_z$$

Two different strategies are used to define housing submarkets:

- In the first dataset, market segmentation is based on the administrative boundaries of parishes, resulting in four housing submarkets: Vera Cruz, Gloria, Esgueira and the suburban area, encompassing the parishes of São Bernardo, Santa Joana and Aradas (Section VII.4.3).
- In the second analysis (for the larger dataset), covering a wider range of territory, two approaches reflecting several dimensions of market segmentation are considered: an inductive perspective, using *ex ante* criteria, and an analytical perspective, where a spatial clustering analysis is applied (Section VII.5.3). This analysis resulted in six, seven, eight and fifteen housing submarkets.

ii) Spatial dependence is associated with interactions between submarkets or houses within the same submarket, that is, when the hedonic price of a house in a particular location depends on other observations located nearby or elsewhere. In a hedonic model, the spatial dependence can be specified as follows (eq. 24:

$$\ln p = \rho W_1 p + Hv + \varepsilon$$
$$\varepsilon = \lambda W_2 \varepsilon + u$$

Where: W_1 and W_2 are spatial weights matrices, measuring the interaction between neighbouring sites, corresponding to spatial lag dependence and spatial error dependence respectively; ρ and λ are the estimated spatial autoregressive coefficients that capture the influence of the average unit located nearby; and u is the vector of idiosyncratic error terms (Anselin, 1988, 2005). The choice of spatial weights is a central issue in many applications of spatial interaction. The model is estimated using two distinct approaches, in both datasets:

- The first approach largely follows the traditional assumption of an *ad hoc* predefined matrix *W*, using distances and contiguity to measure the weights. Based on this assumed *W*, global tests for spatial autocorrelation (Moran's index and LISA indicator) and more specific tests of spatial autocorrelation, such as spatial error dependence (SED) and spatial lag dependence (SLD) are performed using the GEODA software (Anselin, 2005). The geometric notion of space, considering several specifications of pre-assumed matrices, both distances and contiguity matrices, is explored in the section VII.2.4 and VII.3.4, for smaller and for the larger dataset, respectively.
- The second approach adopts a multi-dimensional non-Euclidean view of space, and estimates the weight matrix under the structural constraint of symmetry. Instead of using a predefined matrix *W*, the unknown weight matrix is estimated using statistical inference methods, extending the panel data method in Bhattacharjee and Jensen-Butler (2011) to the pure cross-section setting; see Bhattacharjee and Holly (2011) for discussion of related models and methods in the panel data setting. The advantage of this method, when compared with the traditional approach, is that it does not make restrictive assumptions concerning the drivers of spatial dependence, providing unique opportunities for understanding the nature of interactions. This multi-dimensional non-Euclidean notion of distances, not related to any pre-assumed weights matrix, is applied to the smaller and larger dataset, in sections VII.2.5 and VII.3.5, respectively.

iii) Finally, the spatial scale is considered, which is closely related to the vertical spatiality of each of the aspects described above. The idea expressed here is that both spatial heterogeneity and spatial dependence might be strongly conditioned by the spatial scale at which the housing market is analysed. For this reason, an identical methodology is replicated to both datasets, providing interesting insights about the importance of spatial scale in the housing market study.

- The use of two databases, with submarkets defined at different spatial scales, allows the investigation of robustness of the results (in terms of spatial heterogeneity and spatial dependence), and therefore, the assessment of importance of scale itself.

The reminder of this chapter is organised as follows. First, an overview about how to produce an appropriate dataset to analyse an urban housing market is presented in the Section VII.1. Next, in Section VII.2, the urban area of Aveiro is described, highlighting the most important aspects of the spatial housing distribution. Section VII.3 focuses on the construction of new location attributes and a preliminary data analysis is also presented. The next two major sections of Chapter VII focus on the spatial econometric analysis of the housing market of Aveiro and, besides the study of spatial heterogeneity and spatial (dependence) spillovers across and between submarkets (described above), include: a standard hedonic pricing model using initial variables (both intrinsic and extrinsic attributes) and factors, provided by a factor analysis (VII.4.1 and VII.4.2 for the first small dataset; and VII.5.1 and VII.5.2 for the second and larger dataset).

VII.1. Databases and data collection

An important consideration highlighted in Chapter IV is that housing is not a homogenous asset and can be characterised by a bundle of attributes, both locational and physical. The availability of reliable housing data, in terms accuracy and relevance, allowing an appropriate analysis on the complexities of urban housing theme, is thus a paramount issue.

In general housing data are useful to all housing market participants for: i) academics or private researchers to conduct their research; ii) public housing authorities, policymakers, government or local authorities to support their policy decisions; iii) business agencies to support their investments; iv) households to support a rational purchase decision; and finally, v) all agents that are involved in numerous urban and housing issues.

The application of appropriate techniques and methodologies to achieve the intended purposes is a major challenge, however because those methods depend on the accuracy, consistency, and completeness of the initial information, a well structured dataset assumes a crucial role. A mismatch between data source, methods and objectives can subsequently produce unreliable results (Pollakowski, 1995a; b). Two examples are described by Pollakowski (1995a) to illustrate this problem.

The first example is related with the measurement of housing price indices, most of the time defined without any correction for heterogeneity of houses, leading to inconsistencies in results caused by infrequent sales for a specific kind of house. The typical approach to assess sales prices or rents (mostly when the average price of housing units sold in a specific city is available) is using medians or averages from recent transactions, and a comparison is made with prices of the preceding month or with the same month in the previous year. However, prices may rise because larger houses (or better quality houses, or those in more desirable locations) are being built or are sold in a situation of changing market conditions. The major advantage of this approach is that the data requirements are not very strict and for the most simplistic variant just the transaction values have to be known.

The second example is related to scale. Because housing conditions operate at multiple geographic scales, it is convenient to adopt appropriate databases to analyses the housing phenomena. Using a national or a regional housing database to study a particular metropolitan urban area may be quite misleading, housing markets behave in different manners depending on the scale. The same is valid if metropolitan databases are used to analyse municipal or neighbourhood housing events (Meese and Wallace, 1991). For this reason housing market conditions or more specifically neighbourhood conditions should be analysed over several scales, because conditions at one scale may significantly alter the effects of an intervention at another scale (as been suggested in this methodology).

An empirical study of housing can contemplate a multitude of aspects, and depending on the goals of the research each set of housing attributes can assume different levels of importance; and thus, a relevant and corresponding database should be used, for example:

i) A more *specific set of attributes* is required and adequate if the study focuses on the impact of a certain intrinsic housing characteristic on households' preferences.

ii) Databases including *geographic location* (property identification number, street address, zone, or latitude and longitude) must be used if the examination of numerous spatial housing market issues is an aspect to be analysed. The precise or approximate housing location is useful for three main reasons. First, hedonic analysis can incorporate substantial spatial detail, allowing the evaluation of all kinds of environmental assets by means of revealed preference theories. This assessment can be made by analysing the impact of various locational attributes (environmental; urban infrastructures and public services accessibilities; characteristics of specific neighbourhoods) on housing value, both in terms of amenities (e.g., green areas, supermarkets) and disamenities (e.g., waste sites, heavy industry, abandoned parcels). Second, location permits the matching of property parcels to other statistical information systems (for example, neighbourhood characteristics given by the census, such as, income, educational qualifications, and concentration of poverty), allowing the identification of housing needs, and addressing broader issues of residential segregation and poverty. And last, it is possible to extend the simple hedonic models to a spatial econometric perspective, both in term of spatial heterogeneity and spatial dependence effects. This extended analysis is useful for understanding how housing segmentation is spatially distributed and interconnected.

iii) *Identical data sources* are needed if differences in housing prices or in its specific characteristic among countries, regions or metropolitan areas are the objectives of the research. The problem is that data produced by different sources often need to be integrated, mainly at three different levels: spatially, boundaries and roads may be topologically inconsistent; thematically, different datasets may have diverse attributes or coding; and methodologically, for international comparison of urban development outcomes, methodological inconsistencies of collecting data are common. To characterise comparatively housing markets, in two territories or in two periods of time, some effort to harmonize data should be made.

iii) *General* but *adequate housing attributes* (over time and cross-sectionally) are required if the goal is to calculate a housing price index. The construction of housing price and quantity indexes by hedonic regressions can address many specific policy questions about housing markets, such as regulation and taxation (Malpezzi and Mayo, 1997). In this context a great detail on the initial information is not required, but the lack of certain locational attributes and structural information lead to construction biased housing price indices. Usually macro indicators collected by national statistical institutes are used.

Several types of data are available for analysing urban housing issues, at micro and macro levels. At an international level, databases are available to compare housing performance across countries (see for example World Bank and OECD programs). At national level, and in the particular case of Portugal, three different sources of housing data can be used.

First, the most commonly used, is statistical information provided by the *National Institute of Statistics* (INE), where very general housing indicators are collected every 10 years.

Another possibility is the Local Tax on Real Estate database (IMI - Imposto Municipal sobre Imóveis). This information belongs to Ministry of Finance and is focused on the value of urban and rural buildings (and their characteristics) located in the national territory. Detailed information of property attributes for all housing that has been sold since 2004, are available. This database has two major problems, one is the confidentiality of information (the government does not share all the descriptors of the property, only aggregate attributes), and the other constraint is that reported prices are significantly below market, both buyer and seller have big incentives to declare the minimum amount possible for the transaction. To assuage these disparities, values are corrected by local expert committees, which estimate in loco the value of houses[112]. However, assessments are typically not performed every year, which creates a lag in recording quality adjustments for specific properties. In addition, this method does not guarantee homogeneity of appreciation.

The Real Estate Agencies (national or local level) and Real Estate Agent Associations (constructors) are other options (perhaps the most appropriate). Usually, each real estate agency (individually or organised in a group) have their own housing database resulting from their business. Nevertheless, one drawback of using this kind of database is that the transacting properties are not representative of the properties being measured, that is, houses sold or rented in any period may not be representative of the housing stock and do not necessarily reflect the composition of the housing stock. Some dwelling types are transacted more often that others and some do not transact at all. Another limitation is with regards to the harmonization of the information. Because each agency operates in a specific area of jurisdiction, only covering a part of the territory (influence area: urban or suburban areas), and is focused on a specific market segment (rental or sale; new or used), the availability of information may vary across

[112] See Decreto-Lei n.º 287/2003 12th of November.

space. The exaggeration of the qualitative aspects of structural characteristics to value the product is an additional negative aspect of these databases. The information gathered by loan associations and mortgage banks can be also useful for the purpose of housing analyses, however, two constraints can be enumerated, one is that housing values are truncated at maximum price, because of the upper limit on conventional mortgage lending amounts; and also properties transacted that are not subject to a mortgage are not included.

A non-exhaustive list of sources for housing information has been presented above; however, as has been mentioned, almost all of the several types of sources suffer from a lack of information. According Pollakowski (1995a) each database can diverge in the following types of features: i) measure of property value (transition price, asking price or estimated values); ii) quality of housing characteristics (number of data items: structural characteristics, seller and buyer characteristics, financing, and length of time on market); iii) location details (address, census track, city, suburban or rural); iv) completeness; v) representativeness (sample size); and vi) length of time covered and frequency (time series to analyse the housing market dynamic).

Thus, a household survey of a representative sample of dwellings is a way to adjust for some the problems mentioned above; however, the high costs of gathering this information can be decisively limiting.

VII.2. Description of the unit of analysis

The urban area analysed in this empirical work is located in the Centro Region of Portugal and includes two municipalities, Aveiro and Ílhavo. The former has a total area of 200 km^2 and a total population of 72866, of which 76% live in the city; the latter has an area of 75km^2 and 37162 inhabitants, of which 69% live in the city (Census 2001). The city of Aveiro together with neighbouring Ílhavo, make one conurbation with a population of 110000 inhabitants that makes it one of the most important by population density in the Centro Region. The population of the housing market studied in this work is distributed over 18 parishes, 14 belonging to Aveiro (with an average area of 1411 ha) and 4 to Ílhavo (with an average area of 1837 ha) (see Figure 28).

The analysis of the spatial distribution of the population and housings, presented in Table 12 and Table 13, show that urban area with only four parishes (Glória, Vera Cruz, Esgueira and Ílhavo) is the most populated, with 41,6% of inhabitants and 43.8% of the total of dwelling; the suburban area, with 9 parishes, has 22.1% of the total population and 18.9% of the total dwellings; and finally, the rural area, with 4 parishes, has 36.3% and 37.3% of the population and of the dwellings, respectively.

The empirical analysis for the micro-scale approach (smaller dataset) only considers 6 parishes of the urban and suburban municipality of Aveiro. As it can be seen in the Table 12 and Table 13 these parishes represent, in terms of classic family dwellings, 73.2% of the municipality of Aveiro and 47.5% of the entire zone. In terms of population and housing variation this area is considered the most dynamic area in the territory under study. The spatial distribution of the properties is presented in *Figure* 30, where each house is indicated by a dot.

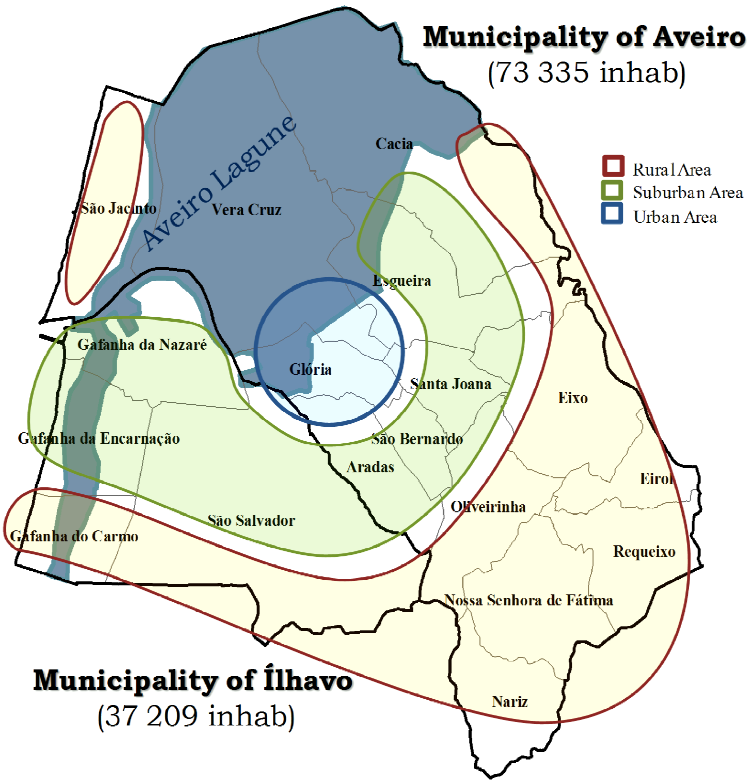


Figure 28 - Municipalities of Aveiro and Ílhavo

Table 12 - Population of Aveiro and Ílhavo

Geographic areas	**Population 2011**		**Variation 1991-2001**	**Variation 2001-2011**
Portugal	10555853		5,0	1.9
Centro Region	2327026		4.0	-0.9
Aveiro housing market	117026	**100.0**	**10.9**	**5.9**
Aveiro	**78463**		**10.4**	**7.0**
Aradas	9151	7.8	-11.3	20.0
Cacia	7399	6.3	7.3	5.6
Eirol	752	0.6	23.0	-3.7
Eixo	5533	4.7	40.1	5.3
Esgueira	13432	11.5	12.2	9.5
Glória	9053	7.7	8.9	-8.7
Nariz	1421	1.2	13.5	-3.1
Oliveirinha	4814	4.1	12.0	0.7
Requeixo	1234	1.1	0.9	3.0
São Bernardo	5018	4.3	23.1	23.0
São Jacinto	996	0.9	3.4	-2.0
Vera Cruz	9644	8.2	22.6	11.5
Santa Joana	8097	6.9	6.3	9.0
Nossa Senhora de Fátima	1919	1.6	3.4	2.6
Ilhavo	**38563**		**12.0**	**3.6**
Gafanha do Carmo	1754	1.5	11.9	15.3
Gafanha da Encarnação	5481	4.7	-2.5	11.7
Gafanha da Nazaré	14730	12.6	20.5	5.1
Ílhavo (São Salvador)	16598	14.2	10,2	-1.0

Table 13 - Classic family dwellings spatial distribution

Geographic areas		Dwellings 2011		Variation 1991-2001	Variation 2001-2011
Portugal		5879845		20.7	16.3
Centro Region		1450268		16.7	15.6
Aveiro housing market		62617	**100.0**	27.5	21.5
Aveiro		**40683**		**26.3**	**24.0**
	Aradas	4899	7.8	11.5	45.2
	Cacia	3194	5.1	13.8	21.3
	Eirol	342	0.5	23.2	7.2
	Eixo	2518	4.0	44.8	20.7
	Esgueira	6446	10.3	28.4	23.2
	Glória	6266	10.0	30.3	9.6
	Nariz	656	1.0	11.3	9.0
	Oliveirinha	2135	3.4	16.8	19.1
	Requeixo	563	0.9	3.0	16.8
	São Bernardo	2286	3.7	37.7	48.5
	São Jacinto	637	1.0	28.6	22.0
	Vera Cruz	6390	10.2	46.1	31.1
	Santa Joana	3516	5.6	21.0	21.7
	Nossa Senhora de Fátima	835	1.3	9.4	10.6
Ilhavo		**21934**		**29.6**	**17.2**
	Gafanha do Carmo	921	1.5	25.0	27.7
	Gafanha da Encarnação	3845	6.1	22.6	18.3
	Gafanha da Nazaré	8817	14.1	42.5	16.8
	Ílhavo (São Salvador)	8351	13.3	21.6	16.1

VII.3. Construction of the location attributes

According to the literature and in line with what has been described in the *Figure 16*, three broad categories are usually used to assess household's perception of the neighbourhood characteristics: i) accessibility or proximity to some spatially related element such as, employment centres or environmental (dis)amenities (airport, water surface, park etc.); ii) socioeconomic status of the neighbourhood, such as, racial composition, household income and level of education; and iii) municipal services such as, school quality, police protection and taxes.

Despite the relevance of the socioeconomic status and municipal services measurements, to explain variations of housing values, only location attributes are considered in this thesis. Most attempts to estimate the implicit market price of socioeconomic attributes were not statistically significant.

Several constraints emerged in the use of socioeconomic statistical data. First, Census Bureau data (National Statistical Institute of Portugal) has not enough spatial disaggregation to adequately represent the complexity of the local urban housing phenomena, the specificities of its people, housing and sites. The argument stated by Goodman (1977) underlines the importance of using suitable data. The author argues that a neighbourhood should be defined as a small urban area within which residents receive and perceive a common set of socioeconomic effects and services. Second, these types of indicators are highly correlated with each other. For example, educational rates are highly correlated with income; and the distance from the CBD with the population density, only to cite some examples. Of course there are many strategies to avoid the potential problem of multicollinearity, but after the computation of numerous models, the use of socioeconomic variables in the hedonic model were not justified, in other words, explained capacity of the hedonic models did increase significantly.

Additionally, the use of data from 2001 (the last available data) is also questionable; in one decade, many structural changes at a local scale can occur.

Given the great diversity of indicators and information associated with the housing market, the selection of attributes considered relevant to explain house prices is crucial for real estate agents, urban planners, as well as policy makers that, directly or indirectly, are involved in this issue.

Physical attributes are restricted by their availability in the existing datasets. It is obvious that is not possible to update, for large amount of data, missing physical values or attributes for a set of housing. However, for location attributes the story is quite different. Using the property location data (Geographic Information System – GIS coordinates) and information about the available urban amenities (equipments, services, place of interests, etc.), it is possible to construct new variables reflecting the distance of a given house to a set of amenities offered in its neighbourhood or city. The service provided by the website *SAPO MAPAS* (http://mapas.sapo.pt/) has been used to extract the exact location of relevant services and equipments available in Aveiro. As a result, 1050 points are obtained (*Table 15*), classified into 12 categories. The *Table 14* represents the distribution of amenities, by parishes and by large initial categories.

Because both levels of disaggregation have been considered inappropriate for the analysis (one is too detailed and the other is too aggregated), a new intermediate classification has been created, described in the *Figure 29* (22 categories).

Table 14 - Distribution of the location attributes (aggregate categories) by parishes

	Commerce	Culture	Education	Health	Industrial zone	Leisure	Mobility	Park and Garden	Sport	Tourism	Utilities Type A	Utilities Type B	Total
Aradas	4	2	13	3		8			5	7		12	**54**
Cacia	1		8	2		3	2		4	8	1	10	**39**
Eirol			2	1			1		1	2		1	**8**
Eixo	1	1	6	3	1	1	4		3	9		4	**33**
Esgueira	26	1	12	4	1	12	1		6	14	2	24	**103**
Gafanha da Encarnação	1		10	3		25			8	2	2	6	**57**
Gafanha da Nazaré	4	2	14	5		32	1		5	12	4	14	**93**
Gafanha do Carmo			2	1					1	2		2	**8**
Glória	12	6	18	7	1	50	5		14	33	17	25	**188**
Nariz			1	1					2	3		1	**8**
Nossa Senhora de Fátima			3	1	2				2	3		1	**12**
Oliveirinha	1		9	2		3	1		3	6		4	**29**
Requeixo			2	1			1		1	6		2	**13**
Santa Joana	1	1	7	3	1	1			2	4	8	4	**32**
São Bernardo			3	2		1			1	5		10	**22**
São Jacinto	1		2	1		5	1		4	2	1	3	**20**
São Salvador	5	2	20	5	1	14			9	16	10	12	**94**
Vera Cruz	15	6	6	8		62	3		11	31	19	34	**195**
Total	**72**	**22**	**138**	**53**	**7**	**217**	**20**	**41**	**82**	**165**	**64**	**169**	**1050**

Table 15 - Location amenities attributes by categories

	Commerce	Culture	Education	Health	Industrial zone	Leisure	Mobility	Park and Garden	Sport	Tourism	Utilities Type A	Utilities Type B	Total
Banks												72	72
Bars and pubs						41							41
Primary and secondary schools			74										74
Libraries		7											7
Petrol station												31	31
Fire Station											4		4
Soccer fields									13				13
Tennis Courts									5				5
Notaries											5		5
Typical and relevant houses										33			33
Cash & Carry	4												4
Castles and forts										1			1
Cathedrals and Basilicas										1			1
Cultural Centers		4											4
Health centers				5									5
Equestrian centers									3				3
Cinemas		3											3
Fitness circuits									1				1
Municipalities											6		6
Sports complexes									14				14
Registries											3		3
Bandstands		1											1
Post offices												15	15
Nautical Sports									4				4
Extreme Sports									1				1
Discos						5							5
Municipal enterprises											3		3
Vocational schools			2										2
Artistic and technical schools			5										5
Internet Spaces						4							4
Railway Stations							9						9
Football Stadiums									2				2
Health Extensions				14									14
Pharmacies				28									28
Lighthouses										1			1
Fairs	7												7
Finances											4		4
Fountains and aqueducts										3			3
Police - G.N.R.											9		9
Commercial galleries	10												10
Art Galleries		4											4
Gyms									3				3
Shopping malls	4												4
Large specialty stores	19												19
Health clubs									7				7
Hypermarkets	3												3
Hospitals				2									2
Hotels										13			13
Wi-Fi hotspots						13							13
Churches, chapels and shrines										26			26
Inatel Hotels										1			1
Government Institutions											4		4
Gardens								5					5
Parish councils												19	19
Law courts											5		5
Picnic sites								12					12
Marinas									1				1
Markets	7												7
Mini Golf									1				1
Viewpoints										1			1
Cash machines												30	30
Museums		1								9			10
Other schools			2										2
Police P.S.P.											3		3
Chemists				4									4
Parks								5					5
Camping										4			4
Exhibition Parks					1								1
Business parks					1								1
Industrial parks					5								5
Playgrounds								19					19
Pillories and cruises						3							3
Guesthouse						13							13
Swimmingpools									9				9
Municipal Police											2		2
Police											12		12
Historic Bridges										2			2
Tourist offices						5							5
Youth hostels						1							1
Taxi services							8						8
Sea Beaches						6							6
Kindergardens			51										51
Civil Protection											2		2
Sports complex									18				18
Restaurants						126							126
S.E.F. (Immigration and border services)											2		2
Theaters and shows		2											2
Social security												2	2
Religious symbolism										1			1
Superior Schools			4										4
Supermarkets	18												18
Catholic churches										69			69
Boat Terminals							3						3
Total	**72**	**22**	**138**	**53**	**7**	**217**	**20**	**41**	**82**	**165**	**64**	**169**	**1050**

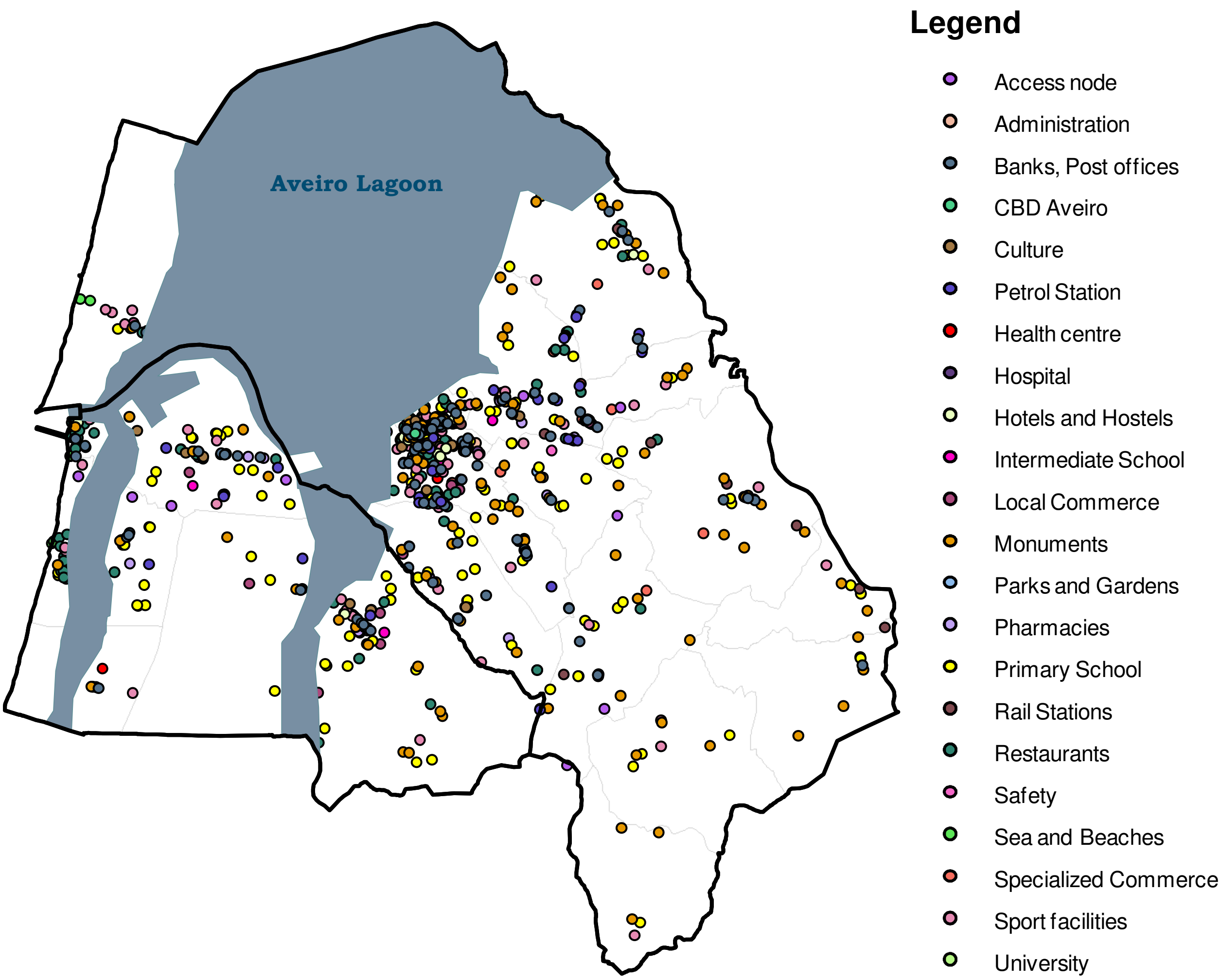


Figure 29 - Distribution of the local amenity attributes

Different strategies to define the location attributes are adopted, in both datasets. While for the smaller database, housing is geocoded based on its addresses, in the second bigger dataset, zones are used to identify an approximate housing location in space.

The literature (e.g.: Falk and Abler, 1980) identifies three types of geographical distances: *global distances*, *effort distances* and *metaphorical distances*, which are strictly related with the notion of absolute, relative and relational space, respectively, expressed in more detail in the Section II.3. The first type of distances is based on traditional notions of physical or geographical distances, called global distances, which are measured by counting the number of abstract units of length between places (e.g., kilometres). The second type of distance is related to the effort that one individual expends in moving from one place to another. The number of effort units can be measured using travel time, monetary units (cost of fuel and tolls), psychological stress caused by the level of traffic congestion, speed limit, quality of a street or natural barriers (e.g.: rivers, lagoons, mountains). A non-linear approach could be also be used considering a distance-decay function (Shaeffer, 1953). The third type measures distances with regard to social contacts between places and individuals, corresponding mainly to the process of space cognition (e.g.: frequency of shopping trips, telephone calls, cultural exchanges etc.).

In this empirical work, two different approaches are assumed, both in line with Falk and Abler's (1980) global distances: one is the use of straight-line distances (minimum distances as shown in the *Figure 30*) to measure the distance between two parcel centroids, and the other is the use of the potentials, computed as shown below.

Being a medium size city within the Portuguese urban context, Aveiro has, in general, quite good accessibilities, where the time spent in transit is not significant. For this reason the absolute notion of space is considered to be a good proxy for assessing the accessibility between places. Note that a more complex calculation could be done, such as use of the road network or other relative notions of space (time or costs), however, the assumption of linear distance as a measure revealed itself to be a good option. Some forms of abstraction and simplification are necessary if one wants to model the impact of perceived accessibility by a household.

The geographic distances are computed according to the classification illustrated in Figure 30 [intensity of use (occasional or intensive); scale and level of influence (local, parish or global); and type of measure (nominal, ordinal or scalar)] and shown in the *Table 19*.

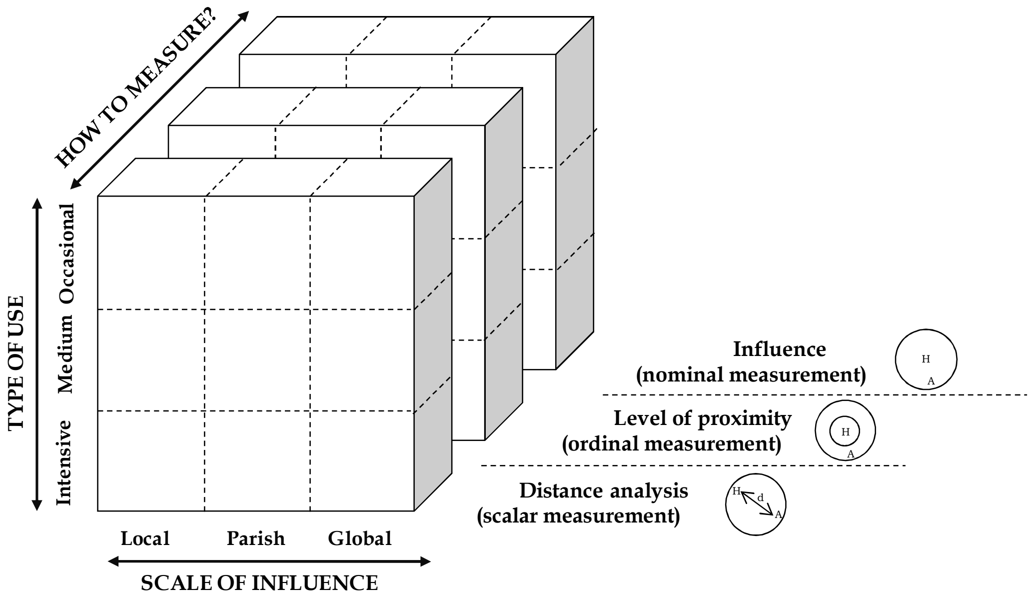


Figure 30 - Three dimensions to build location attributes

The potential [P_i(S)], corresponding to the accessibility indices of urban amenities, is given by a set of services of type (S_j) in a specific location point (i), and is computed using the following equation:

$$P_i(S) = \sum_{j=1}^{n} \frac{S_j}{d_{ij}} \qquad \text{Eq. 35}$$

Where,

$P_i(S)$ is the potential (accessibility index) of the amenity type S;

S_j represents location amenities (S_1, S_2, S_3, …, S_j);

d_{ij} is the distance between locations i and j (from house i to amenity j).

Thus, some location attributes are defined as minimum distances to services such as high schools or pharmacies and others are defined as gravity type measures of potential, generated by distances to services like restaurants, sport centres or public administration offices.

Table 16 - Type of measures considered in the analysis

Location attributes	Number	Type of measure	Scale	Use
Access node	10	Distance (minimum)	Global	Medium
Administration	12	Distance (potential)	Global	Occasional
Banks, ATMs, Post offices	117	Distance (potential)	Local	Medium
CBD Aveiro	1	Distance (minimum)	Global	Medium
Local commerce	43	Distance (minimum)	Local	Intensive
Specialised commerce	29	Distance (potential)	Global	Occasional
Culture	25	Distance (potential)	Parish	Medium
Petrol Stations	31	Distance (minimum)	Global	Medium
Health centres	5	Distance (minimum)	Parish	Occasional
Hospitals	2	Distance (minimum)	Global	Occasional
Hotels and Hostels	13	Distance (potential)	Global	Occasional
Intermediate School	8	Distance (minimum)	Parish	Intensive
Monuments	129	Distance (potential)	Global	Occasional
Parks and Gardens	10	Distance (minimum)	Parish	Medium
Pharmacies	32	Distance (minimum)	Local	Medium
Primary & nursery schools	117	Distance (minimum)	Local	Intensive
Rail stations	9	Distance (minimum)	Global	Intensive
Restaurants	140	Distance (potential)	Parish	Medium
Safety	2	Distance (minimum)	Local	Intensive
Sea and beaches	6	Influence	Global	Occasional
Sport facilities	56	Distance (potential)	Parish	Medium
Universities	1	Distance (minimum)	Global	Intensive
Total	**798**			

In addition, for the larger dataset (macro scale approach), two more aggregated levels of centrality are considered for characterizing the location attributes (*Figure 31*). The first level of centrality is a unique single point identified by the traditional CBD of city of Aveiro (Ponte Praça). The other, a lower level of centrality, designated by centrality level 2, corresponds to eight different locations, obtained by the following methodology:

- First, all urban amenities that can be important adding value (positive externality) or depreciation (negative externality) of a dwelling are identified (as described above in the *Table 15* and illustrated in the right hand side of the map presented in *Figure 31*).
- Second, the geographical distances from each abstract point in space (with a distance of 100 meters) to the several equipments and services available within the city are calculated through the use of GIS. The result is an isopotential for each facility identified in a gray gradient in *Figure 31*.

- The third and final step is the identification of the higher level of potential in each part of the territory (relative maximum), leading to eight new points of centrality (level 2), identified as the red areas in the *Figure 31*.

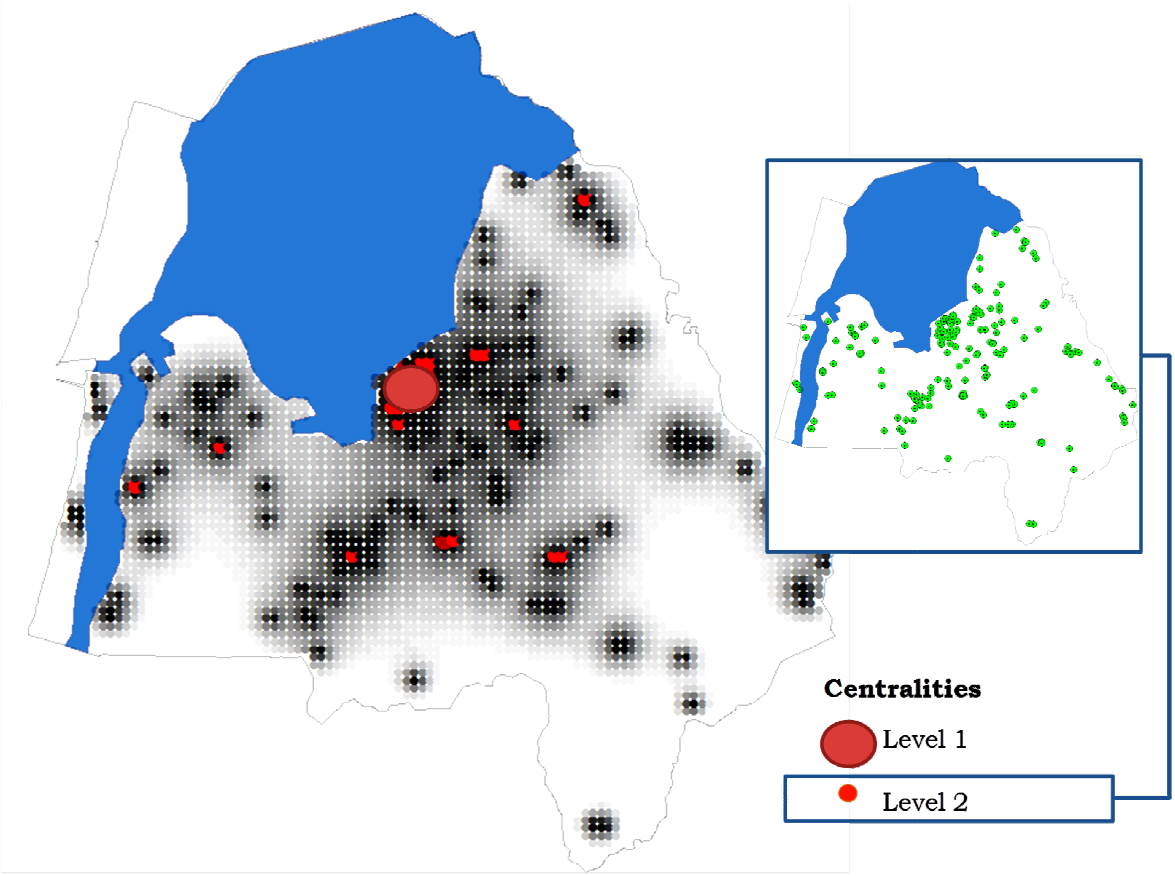
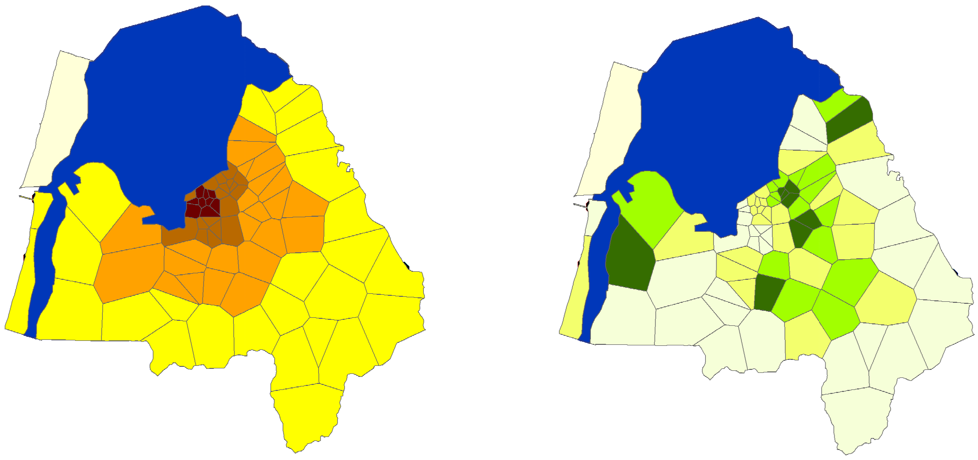


Figure 31 - Spatial attributes considered in the construction of location variables

After the identification of these two levels of centrality, the distances from each property to these points are calculated. Since the second level of centrality (Level 2) consists of eight remote locations, the closest point is considered. The different level of accessibility for each of these two levels of centrality is shown below in *Figure 32*.



Centrality level 1 Centrality level 2

Figure 32 - Centrality levels 1 and 2

Descriptive statistics for the final samples of 166 and 12467 observations are provided in *Table 17* and *Table 18*, as well as the type of measures used in each location attribute. Note that, for the macro scale approach, an additional dummy variable (beaches) is used because this larger dataset covers the entire study area. For the small dataset this attribute is not used because all properties are relatively equidistant to the beaches.

Table 17 - Descriptive statistics of location variables – micro scale approach

		Units	N	Min	Max	Mean	Std. Deviation
Location characteristics							
Ln	Central Amenities	(Min. Dist.-meters)	166	4.51	8.58	7.19	0.74
Ln	Local Amenities	(Min. Dist.-meters)	166	8.35	9.26	8.72	0.17
Ln	CBD Aveiro	(Min. Dist.-meters)	166	5.54	8.63	7.30	0.68
Ln	Local Commerce	(Min. Dist.-meters)	166	3.49	7.96	6.14	0.93
Ln	Primary Schools	(Min. Dist.-meters)	166	3.16	6.76	5.48	0.69
Ln	High Schools	(Min. Dist.-meters)	166	3.14	8.23	6.39	0.95
Ln	University	(Min. Dist.-meters)	166	6.06	8.70	7.49	0.58
Ln	Hospital	(Min. Dist.-meters)	166	4.96	8.37	7.08	0.62
Ln	Health Centres	(Min. Dist.-meters)	166	5.32	8.60	7.31	0.66
Ln	Pharmacies	(Min. Dist.-meters)	166	3.39	7.83	5.86	0.88
Ln	Parks and Gardens	(Min. Dist.-meters)	166	5.17	8.20	6.81	0.72
Ln	Rail Station	(Min. Dist.-meters)	166	4.88	8.21	6.90	0.70
Ln	Access Node	(Min. Dist.-meters)	166	5.41	8.31	7.19	0.51
Ln	Petrol Stations	(Min. Dist.-meters)	166	2.08	7.67	6.07	0.95
Ln	Police	(Min. Dist.-meters)	166	3.57	8.41	7.11	0.67
P	Administration	(Potencial)	166	5.49	9.09	6.89	0.72
P	Culture	(Potencial)	166	6.04	8.66	7.19	0.50
P	Specialised Comm.	(Potencial)	166	6.56	8.75	7.71	0.43
P	Restaurants	(Potencial)	166	7.80	10.15	8.90	0.54
P	Hotels and hostels	(Potencial)	166	5.48	8.15	6.72	0.65
P	Monuments	(Potencial)	166	7.95	10.90	8.71	0.48
P	Banks, ATMs, Post	(Potencial)	166	7.87	10.19	8.85	0.47
P	Sports	(Potencial)	166	7.04	8.81	7.88	0.38

d=dummy variable; ln=distances in logarithms; p=gravitational potential in logarithms

Table 18 - Descriptive statistics of location variables – macro scale approach

		Units	N	Min	Max	Mean	Std. Deviation
Ln	Central Amenities	(Min. Dist.-meters)	12467	5.42	11.97	8.02	0.83
Ln	Local Amenities	(Min. Dist.-meters)	12467	5.04	11.95	7.33	0.63
Ln	CBD Aveiro	(Min. Dist.-meters)	12467	5.23	11.98	8.08	0.80
Ln	Local Commerce	(Min. Dist.-meters)	12467	4.07	9.16	6.58	1.15
Ln	Primary Schools	(Min. Dist.-meters)	12467	3.65	7.59	5.60	0.83
Ln	Intermediate Schools	(Min. Dist.-meters)	12467	4.38	8.80	6.57	1.01
Ln	University	(Min. Dist.-meters)	12467	5.46	9.38	8.12	0.63
Ln	Hospital	(Min. Dist.-meters)	12467	5.39	9.34	7.84	0.88
Ln	Health Centres	(Min. Dist.-meters)	12467	4.78	9.16	7.15	0.87
Ln	Pharmacies	(Min. Dist.-meters)	12467	3.60	8.61	5.99	0.95
Ln	Parks and Gardens	(Min. Dist.-meters)	12467	3.97	8.84	7.04	0.95
Ln	Rail Station	(Min. Dist.-meters)	12467	4.41	9.22	7.55	0.99
Ln	Access Node	(Min. Dist.-meters)	12467	5.96	8.62	7.47	0.54
Ln	Petrol Stations	(Min. Dist.-meters)	12467	3.37	8.79	6.53	0.96
Ln	Police	(Min. Dist.-meters)	12467	5.39	11.97	7.84	0.81
p	Administration	(Potencial)	12467	2.02	8.71	6.28	1.10
p	Culture	(Potencial)	12467	5.24	8.05	6.46	0.69
p	Specialised Commerce	(Potencial)	12467	5.31	8.50	6.59	0.72
p	Restaurants	(Potencial)	12467	6.92	10.12	8.44	0.64
p	Hotels and hostels	(Potencial)	12467	5.79	9.41	7.25	0.69
p	Monuments	(Potencial)	12467	7.37	9.90	8.35	0.45
p	Banks, ATMs, Post	(Potencial)	12467	6.64	9.80	8.41	0.68
p	Sports	(Potencial)	12467	6.39	8.54	7.53	0.44
d	Sea/Beaches	(Yes=1; No=0)	12467	0.00	1.00	0.07	0.25

d=dummy variable; ln=distances in logarithms; p=gravitational potential in logarithms

The descriptive statistics presented in *Table 17* and *Table 18* reflect large variation in the location attributes across both datasets. On average, houses are located at 1.5 and 3.2 km from the CBD; and the maximum distance is 5.6 and 16 km, in the smaller and in the larger database, respectively.

VII.4. **Housing market of Aveiro:** Micro scale approach

Several variants of hedonic pricing models, applied to the urban housing market of Aveiro, are examined and presented in this section, using both initial variables and factors (obtained from factor analysis) as independent variables. In addition, spatial heterogeneity and spatial dependence are analysed.

The dataset used in this empirical work includes 166 available properties sold through one of the leading real estate agencies in Aveiro[113] in 2007[114], encompassing only the urban and suburban parishes of Aveiro, which corresponds to 6 of the 14 parishes of the municipality of Aveiro. Thus, in this analysis the rural parishes of Aveiro and all the parishes of the municipality of Ílhavo are not included.

The spatial distribution of the properties is presented in Figure 33, where each house is indentified by a dot. This housing database can be criticised because it is too small, however, the availability of: i) exact locations for each house (allowing the computation of accurate distances to several urban amenities, services and equipments), ii) detailed information for intrinsic characteristics (allowing a greater specification of the hedonic attributes); and finally, iii) the real price sale price (note that several studies consider the listing price, as it is considered in the dataset presented in Section VII.5), make this database very valuable. Moreover, the small size of this database is quite useful because it enables an expeditious estimate, with spatial econometric models using different alternatives of spatial weights and subsequent adequacy tests.

---

[113] *Era Aveiro* is the name of the real estate agency.
[114] Because all observations are related with the year of 2007, housing prices are not discounted.

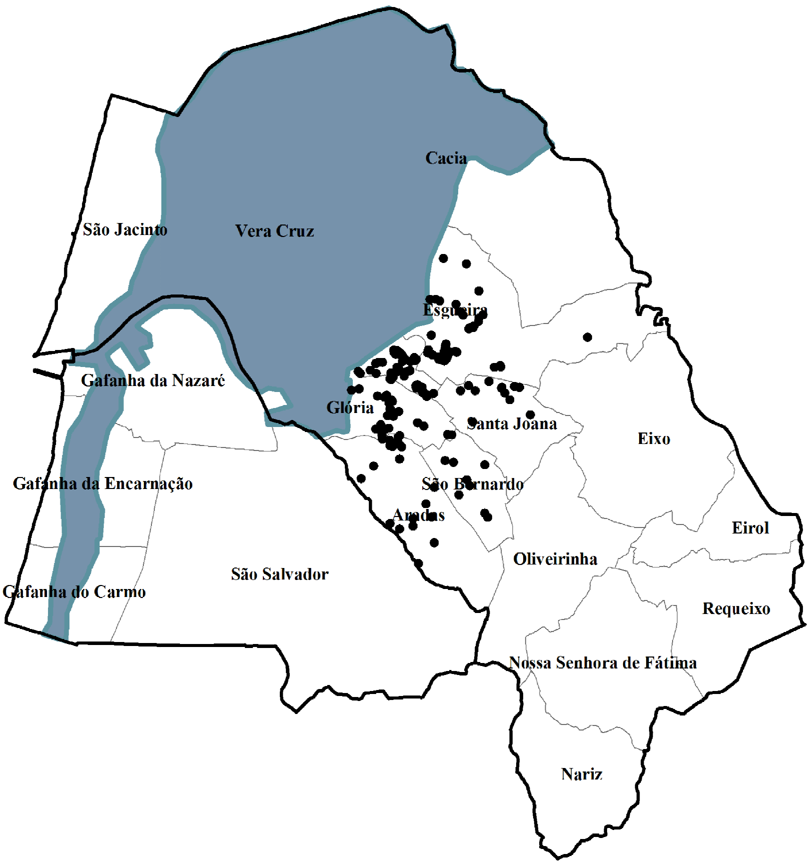


Figure 33 - Location of housing sample 1

VII.4.1. Explanatory analysis of the initial variables

A short descriptive analysis and statistical inference are presented in this section provide a general idea about the initial distribution and dependences of the data.

This dataset contains sales price of houses as well as other house characteristics, as described in *Table 20*. The data covers: houses (12.3%) and flats (87.7%), both new (11.8%) and used (88.2%) buildings, located in different urban and suburban areas of the municipality of Aveiro. Despite the fact that the choice of independent variables is somewhat limited by data confidentiality issues, the housing attributes collected are representative of physical characteristics of a dwelling.

Regarding the spatial distribution of houses, it can be said that the sample is relatively representative of the total amount of houses available in each parish, as can be shown in the *Table 19*.

Table 19 - Population and density of housing sample

Parishes	Population		Density	Housing Sample	
Aradas	7628	15%	854	22	13,3%
Esgueira	12262	24%	691	42	25,3%
Glória	9917	19%	1445	27	16,3%
Santa Joana	8652	17%	225	11	6,6%
São Bernardo	4079	8%	1037	9	5,4%
Vera Cruz	8652	17%	1273	55	33,1%
Total	**51190**	**100%**	**368***	**166**	**100%**

*including the area of Aveiro Lagoon

As has been mentioned before several dependent variables can be used in hedonic price models to characterise the value of a house. In this particular situation the transaction price per square meter (€/m2) is considered, transformed by its logarithm to have a more scale neutral, normalised measure. Some other independent variables (including total area), treated as explanatory variables are also normalised (those marked with *ln* in the *Table 20*) to obtain a better interpretation of the results.

Table 20 - Descriptive statistics of variables of intrinsic variables

		Units	N	Min	Max	Mean	Std. Deviation
Internal physical characteristics							
d	Type	(House=1, Flat=0)	166	0.00	1.00	1.13	0.34
d	Duplex[115]	(Yes=1; No=0)	162	0.00	1.00	1.20	0.40
d	Balcony	(Yes=1; No=0)	166	0.00	1.00	0.19	0.40
d	Terrace	(Yes=1; No=0)	166	0.00	1.00	0.10	0.30
d	Garage space	(Yes=1; No=0)	166	0.00	1.00	0.59	0.49
d	CATV (Cable Television))	(Yes=1; No=0)	166	0.00	1.00	0.26	0.44
d	Gas (natural)	(Yes=1; No=0)	166	0.00	1.00	0.38	0.49
	Number of bedrooms	(Number)	165	1.00	5.00	2.32	0.84
d	Preservation	(Used=1, New=0)	165	0.00	1.00	0.88	0.32
	Floors	(Number)	166	1.00	12.00	3.46	2.16
ln	Kitchen area	(m^2)	139	1.70	3.21	2.52	1.35
ln	Living room area	(m^2)	147	2.12	3.35	2.55	1.02
ln	Price	(euros/m^2)	166	5.98	8.01	7.16	6.10
ln	Total area[116]	(m^2)	166	3.50	5.52	4.74	3.78

d=dummy variable; ln= in logarithms

115 An appartment with two floors.
116 Missing values for *total area* were imputed.

A brief analysis of the descriptive statistics suggests a large variation in the physical attributes across the sample: the average housing price per square meter and the area is 1224 euros/m^2 and 106.9 m^2; ranging between 395.4 and 3010.9 euros/m^2 (with a standard deviation of 1.4), and between 33.1 and 249.6 m^2 (with a standard deviation of 1.2), respectively. Regarding other housing characteristics: 13.3% are duplex; 19,3% have a balcony; 10.2% have a terrace; 59.0% have use of a garage; 25.9% Cable Television (CATV) access; and 38.0% natural gas infrastructure.

The variables presented in *Table 20* are directly used as explanatory variables to build a hedonic model for the housing market of Aveiro, and are also submitted to a factor analysis for use in subsequent analyses, including the construction of factor based hedonic housing pricing model.

The three figures presented below (*Figure 34*, *Figure 35* and *Figure 36*) report the box plot diagrams, where variation of housing prices (absolute and relative) is analysed according to three different types of attributes: location (urban and suburban area), type of dwelling and number of bedrooms. From the analysis of the graphs three conclusions can be underlined: i) the house prices are higher in urban areas, both in relative and absolute terms; ii) in this sample, flats are more expensive than houses because flats are typically located in urban areas and houses in suburban areas; and finally, iii) there is a non-linear relationship between the number of bedrooms and the price per square meter, suggesting the use of logarithmic functions in the hedonic models.

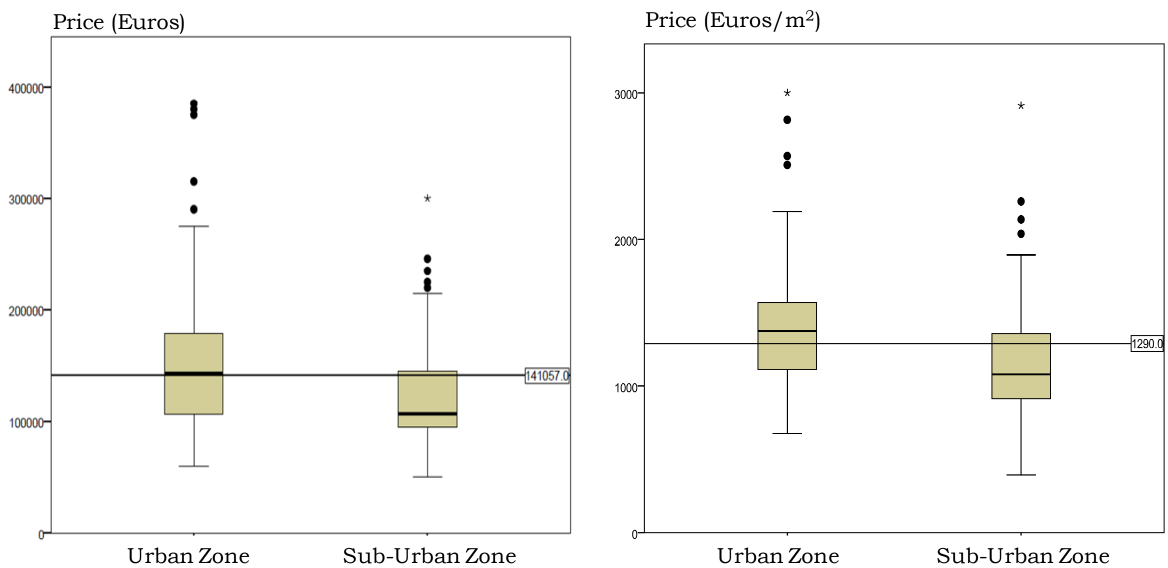


Figure 34 - Relation between price and location

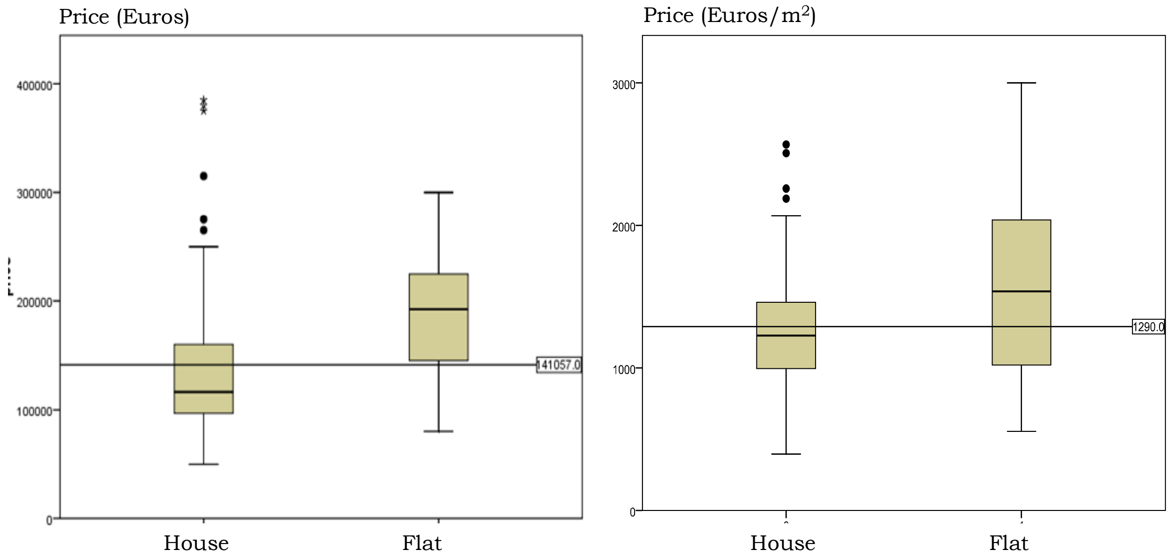


Figure 35 - Relation between price and housing type

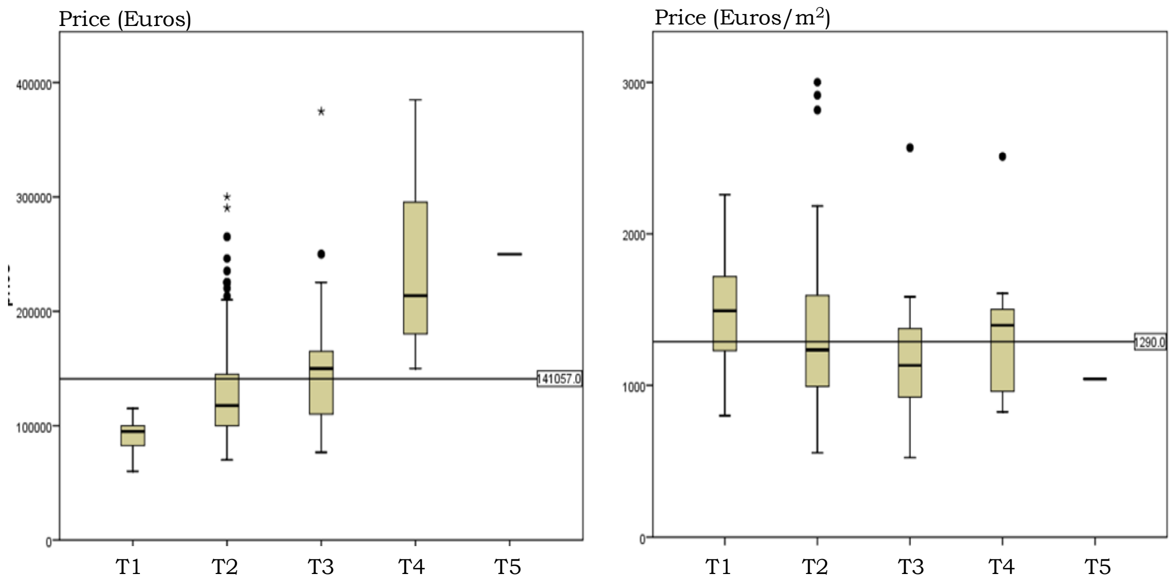


Figure 36 - Relation between price and number of bedrooms

In order to explore dependences between variables some statistical tests are computed. The results are presented in more detail in Appendix 1.

Because the variables are not normally distributed[117], as shown in the *Table 34*, three different non-parametric techniques[118] are applied in the analysis

[117] The null hypothesis for the test of normality states that the actual distribution of the variable is equal to the expected distribution, i.e., the variable is normally distributed. Since the probability associated with the test of normality is lower than 5%, the null hypothesis is rejected and it can be concluded that the variables reported in the Table 34 are not normally distributed.

[118] Non-parametric tests are used to overcome the underlying assumption of normality in parametric tests.

of the dependence between variables: i) Mann-Whitney[119] test is used to test whether the two independent samples (or groups) of observations are drawn from the same or identical distributions (based on their medians); ii) Kruskal-Wallis test[120], which is an extension of the Mann-Whitney test, is used for independent variables with two or more categories and an ordinal dependent variable or when the independent variable is not normally distributed; and, iii) Chi-Square (χ^2) test and contingence tables are used when two nominal variables are analysed. This last non-parametric test determines whether there is a significant difference between the expected frequencies and the observed frequencies in one or more categories, giving insights about the dependence (or not) between two categorical variables. For all the statistical tests presented below the level of significance considered to reject the null hypotheses is 5%. For more details see a textbook in statistics, such as, Hall (2009) among others.

Table 21 - Tests of Normality

	Kolmogorov-Smirnov		
	Statistic	df	Sig.
Kitchen area	0.112	134	0.000
Living room area	0.162	134	0.000
Price (€/m^2)	0.079	134	0.041
Price (€)	0.143	134	0.000
Number of bedrooms	0.262	134	0.000
Floors	0.272	134	0.000

a) Analysing differences between the different territories (parishes)

For the analyses of the dependency between different parishes (six groups) and the remaining intrinsic housing attributes, two different non-parametric statistical tests have been computed: Kruskal-Wallis when variables are scalars; and the Chi-square test when variables are nominal.

The Kruskal-Wallis results, presented in the *Table A.3* of the Appendix 1, suggest that there are significant differences in the price and in the number of bedrooms, across the considered parishes. For example, Vera Cruz (the urban

[119] A non-parametric equivalent to the independent sample *t*-test and is used when you do not assume that the dependent variable is a normally distributed interval variable.
[120] It is the non-parametric version of the one-way analysis of variance ANOVA test.

centre of Aveiro) is the parish where housing is most expensive (both in absolute and in relative value) and where the number of rooms is largest. Other differences can be seen in the *Table A.1* and *A.2*. In all the other variables (kitchen area, living room area, total area and floors) the differences between parishes are not statistically significant (at the 5% level).

Considering the dependency between nominal variables, there are several conclusions to be drawn from the results presented in the *Tables A.1.x* (*x=4,... , 19*). First, the dwelling type is statistically different between parishes, in the sense that houses are predominantly located in Aradas, Esgueira and Santa Joana (suburban zones); and flats in the two urban parishes (Glória and Vera Cruz). The second message is that, duplex dwellings are more common in Aradas, Esgueira and Santa Joana; and less so in Vera Cruz. Third, the existence of balconies in dwellings is higher in Esgueira, and lower in Santa Joana and Vera Cruz. Finally, the portion of housing with natural gas in Glória and São Bernardo is higher than expected, while in Vera Cruz it is lower. For the remaining attributes (terrace, garage space, CATV and preservation) the differences across parishes are not statistically significant.

b) Analysing differences between housing type (house or flat)

Regarding the Mann-Whitney test (presented in *Table A.1.21* of the *Appendix 1*) all the variables (with the exception of the price measure per square meter) present significant differences between the types of housing. As shown in the *Table A.1.20*, areas and prices (absolute value) are higher in houses, while the number of bedrooms and floors are higher in flats.

Regarding the Chi-square test presented in the appendix in *Table A.1.x* (*x*=22,... , 35), it can be said that the level of preservation (used or new), the existence of terrace, the existence of natural gas and provision of a garage are independent of the type of dwelling, in other words, being a house or a flat does not influence the existence of these four attributes. A different situation occurs with the duplex variable (more common in houses and less in flats), balcony (more common in houses and less in flats), and CATV (less common in houses and more in flats), where for each of these three attributes there exists a significant level of dependence (*p*-values <0.05).

c) Analysing differences between level of preservation (new or used)

The level of preservation is only associated with the price variable (absolute and relative), meaning that new dwellings are more expensive than the used ones (no one else would have expected another result as logical). No other variables are associated with the housing preservation variable, meaning that being new or used is not dependent on attributes of kitchen area, living room area, floor, number of the bedrooms, type of house, duplex, balcony, terrace, garage space, CATV and natural gas (all the details results are presented in *Appendix 1* in the *Tables A.1.x* (*x*=36,... , 48).

VII.4.2. Hedonic analysis for Housing market of Aveiro and Ílhavo

A hedonic model for the urban and suburban area of the municipality of Aveiro, estimated with ordinary least square assumptions, is presented in the *Table 22*. The results of estimating hedonic the model are relevant in analyzing the relationship between a single dependent variable (logarithm of the house price per m^2) and several explanatory variables, representing both physical and locational characteristics. These regressions allow the measurement of the impact that changes in each explanatory variable cause on the dependent variable, in other words, the relevance of each attribute in explaining the housing price. However, it is important to acknowledge that the price at which the housing is sold does not reveal it exact value. As mentioned by some authors (e.g.: Gatzlaff and Haurin, 1998; Smith, 2000) the transaction price may reflect other relevant factors, such as, uncommon marketing times, unusual financing, atypical buyer and seller motivations, and information asymmetries, among others. Even considering a single year (as considered in this analysis) houses were not sold necessarily on the same date (the year runs from the 1st of January to 31th of December 2007). This time disparity can be an additional constraint.

The hedonic model is described as follows:

$$P = \mathrm{F} \times \mathrm{L}$$ Eq. 36

or

$$\ln P_{(€/m^2)} = \alpha_1 \ln \mathrm{N} + \alpha_2 \ln \mathrm{A} + \sum_{k=1}^{8} \alpha_{F_k} d_{F_k} + \sum_{j=1}^{24} \alpha_{Lj} D_{Lj}$$ Eq. 37

Where:

Alfa (α) are regression coefficients of the housing attributes measuring the importance of each independent variable in the explanation of price.

P is the price

The normalised measure of price per square meter has been used to get a more scale neutral dependent variable.

F are intrinsic housing characteristics:

$$F = \mathrm{N}^{\alpha_1} . \mathrm{A}^{\alpha_2} . \sum_{k=1}^{8} d_{F_k}^{\alpha_{F_k}}$$ Eq. 38

N - Number of bedrooms

A - Area (square meters)

d_{F_k} - Dummy variable for the existence of several F_i physical attributes described in the *Table 20* (F_k=1,...,8); d_{F1}=Type of houses; d_{F2}=Duplex; d_{F3}=Balcony; d_{F4}=Terrace; d_{F5}=Garage space; d_{F6}=CATV; d_{F7}=Natural gas; and d_{F8}=Preservation.

L are housing location attributes:

$$L = \sum_{j=1}^{24} D_{Lj}^{\alpha_{Lj}}$$ Eq. 39

D_{Lj} - Distances to the location attributes described in the *Table 18* (L_j=1, ..., 24). All the location attributes are transformed into logarithmic form.

Table 22 - Regression coefficients for the hedonic model, using initial variables: Enter method

	Coefficients				Collinearity Statistics	
	Unstard.	Strand.	t	Elasticities (%)	Tolerance	VIF
(Constant)	10.653		1.71			
dType	0.344***	0.330	4.185	41.10	0.441	2.267
dBalcony	0.093*	0.124	1.845	9.70	0.608	1.643
dTerrace	0.055	0.059	0.902	5.70	0.638	1.567
dGarage space	0.216***	0.340	4.974	24.10	0.589	1.697
dCATV	-0.062	-0.093	-1.259	-6.00	0.508	1.967
dGas (natural)	0.029	0.045	0.621	2.90	0.513	1.948
Number of bedrooms	0.150***	0.385	3.938	16.20	0.288	3.476
dPreservation	-0.287***	-0.285	-4.573	-24.90	0.707	1.414
Floors	0.005	0.035	0.539	0.50	0.638	1.568
lnKitchen area	0.080	0.077	1.064	8.30	0.529	1.891
ln Living room area	0.320***	0.200	3.000	37.70	0.619	1.617
lnTotal area	-0.868***	-1.027	-10.057	-58.00	0.264	3.789
lnCentral Amenities	0.019	0.048	0.197	1.90	0.046	21.832
lnLocal Amenities	0.181	0.100	0.421	19.80	0.049	20.513
lnCBD Aveiro	-0.078	-0.174	-0.315	-7.50	0.009	110.339
lnLocal Commerce	0.057	0.174	0.968	5.90	0.085	11.756
lnPrimary Schools	-0.069	-0.149	-1.628	-6.70	0.328	3.048
lnInterm. Schools	0.039	0.122	0.938	4.00	0.163	6.144
lnUniversity	-0.242	-0.452	-0.881	-21.50	0.010	95.897
lnHospital	0.168	0.347	1.631	18.30	0.061	16.445
lnHealth Centres	0.054	0.118	0.436	5.50	0.038	26.507
lnPharmacies	-0.003	-0.008	-0.062	-0.30	0.167	5.988
lnParks and Gardens	-0.181	-0.417	-1.371	-16.60	0.030	33.566
lnRail Station	-0.013	-0.030	-0.175	-1.30	0.091	11.000
lnAccess Node	0.065	0.103	0.810	6.70	0.172	5.819
lnPetrol Station	-0.037	-0.110	-1.037	-3.60	0.243	4.121
lnPolice	0.059	0.133	0.849	6.10	0.112	8.949
lnAdministration	0.136	0.331	1.370	14.60	0.047	21.239
lnCulture	-0.030	-0.049	-0.123	-3.00	0.017	58.541
lnSpecial. Commerce	-0.030	-0.044	-0.182	-3.00	0.046	21.583
lnRestaurants	0.107	0.192	0.493	11.30	0.018	54.918
lnHotels and hostels	0.044	0.094	0.209	4.50	0.014	73.835
lnMonuments	-0.145	-0.237	-1.651	-13.50	0.133	7.510
lnBanks, ATMs, Post	-0.139	-0.218	-1.117	-13.00	0.072	13.804
lnSports	-0.064	-0.079	-0.319	-6.20	0.045	22.282
Number of Observations				**166**		
R				**85.6**		
Adjusted R-squared				**63.7**		

*** significant at the 1% level/** significant at the 5% level/* significant at the 10% level

The model presented in *Table 22* is expressed with both unstandardized and standardized values. The unstandardized coefficients (alfa coefficients), in the second column, describe the estimated change in the dependent variable for a unit change of the independent variables. The standardized coefficients, also known as beta coefficients, described in the third column, measure the relative impact on the dependent variable of a change in one standard deviation in either variable. In some situations these beta coefficients are more useful for interpretation purposes, because the problem of dealing with different units of measurement is eliminated. The coefficient of correlation is also presented (R) and determination (R^2): the former, measures the strength of the association between the dependent variable and the independent variables; and the latter measures the proportion of the variance of the dependent variable that is explained by the independent variables (see Hair, Black *et al.*, 2010).

In terms of overall explanatory power, the general hedonic model explains approximately 63% of the total variation of the housing price (€/m^2), with a high level significance.

The results of the regression coefficients for both physical and location attributes of dwellings have the expected signs and significance. However, it is expected that, for some variables, when controlling for other characteristics, the level of significance is not statistically significant. For instance, location within the city clearly does not affect significantly house prices (*p*-values > 0.825), as shown by the weak significant value of the variable 'distance to CBD'. This indicates that once the type and the dimension of house are taken into account, the distance to the CBD of Aveiro does not affect price very much. Note that in the more urban areas properties are typically flats and smaller when compared to the periphery, as is shown in the explanatory analysis presented in the previous section. This aspect is also clearly underlined in the multicollinearity analysis presented below. None of the locational variables are statistically significant (*p-value* > 0.10), which leads to the conclusion of spatial heterogeneity, corroborated by the results in the Section VII.4.4. An additional inference drawn by this result is that log prices are closely related to intrinsic house characteristics, but not strongly related to distances to urban facilities.

The standardized regression coefficients suggest that total area of the dwelling, type of house, garage space and number of rooms, have the largest impact on the assessment of housing prices. Since the dependent variable is the logarithm of price per square meter, the negative coefficients of the total area variable suggest that property values, measured in relative terms (euros/m^2), are higher when the dimension of dwelling is lower. Regarding the other two variables related to the size of dwellings, the coefficients are directly and positive correlated.

Two types of analyses are conducted to verify if the assumptions of the classical regression are upheld: multicollinearity and heteroskedasticity.

The existence of multicollinearity makes it difficult to interpret the impact of any single variable, due to the high correlation with a set of other explanatory variables. Note that in extreme cases of collinearity an independent variable is perfectly predicted by another independent variable (or more than one). To test for multicollinearity, in general, the tolerance level and Variance Inflation Factor (VIF) are computed. Tolerance values below 0.1 and VIF values above 10 suggest a multicollinearity problem (Hair, Black *et al.*, 2010). In the sixth and seventh columns of the *Table 22* the tolerance level and VIF value are shown for each variable, and indicate that the data have significant problems of collinearity. For this reason a stepwise method (forward addition)[121] is applied and presented in the *Table 23*.

Residual plot analysis is conducted to test for heteroskedasticity. Heteroskedasticity occurs when the error term appears to have common variance over a range of predictor variables. The existence of heteroskedasticity, and, unlike the case of multicollinearity, is critical to the proper application of many multiple regression techniques. Most cases of heteroskedasticity are a result of non-normality in one or more variables (Hair, Black *et al.*, 2010).

[121] This method of selecting variables starts by selecting the best predictor of the dependent variable and additional independent variables are selected in terms of the incremental explanatory power they can add to the regression model (Hair, Black *et al.*, 2010).

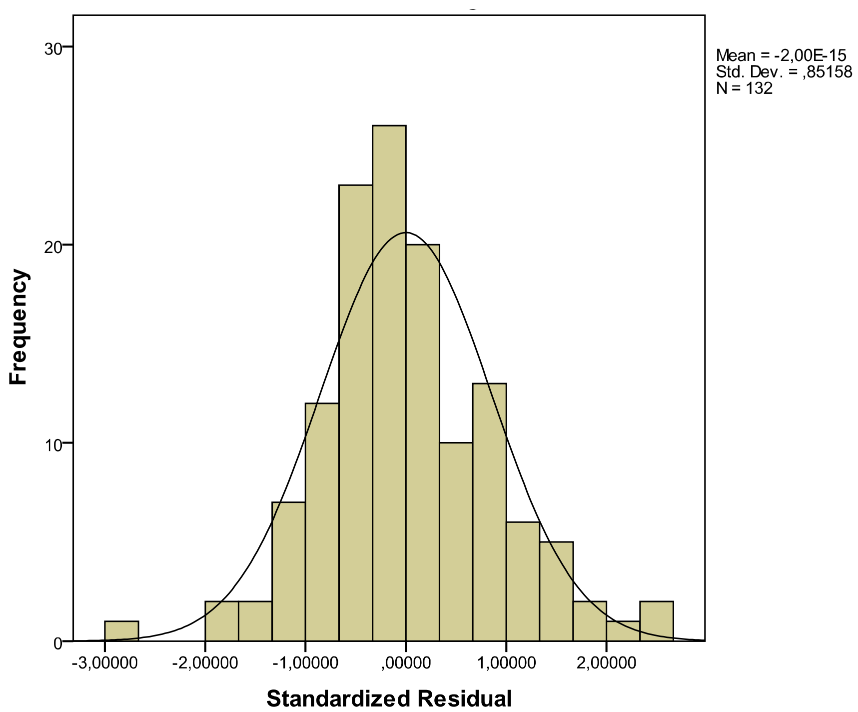


Figure 37 - Histogram of standardized residuals

Residuals appear to be approximately normally distributed are shown in *Figure 37*, meaning that the error term has equal variance and is independent across observations.

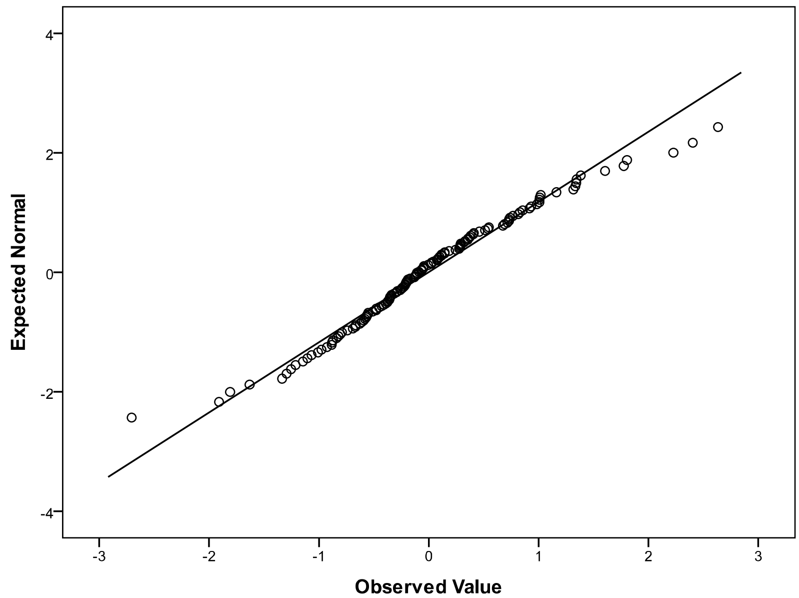


Figure 38 - Standardized normal P-P plot of the residuals

In *Figure 38* the standardized normal P-P plot of the residuals is presented, where the observed cumulative proportion is plotted against the expected cumulative proportion. As shown the points are located close to the straight line, signifying that the residuals are approximately normal distributed.

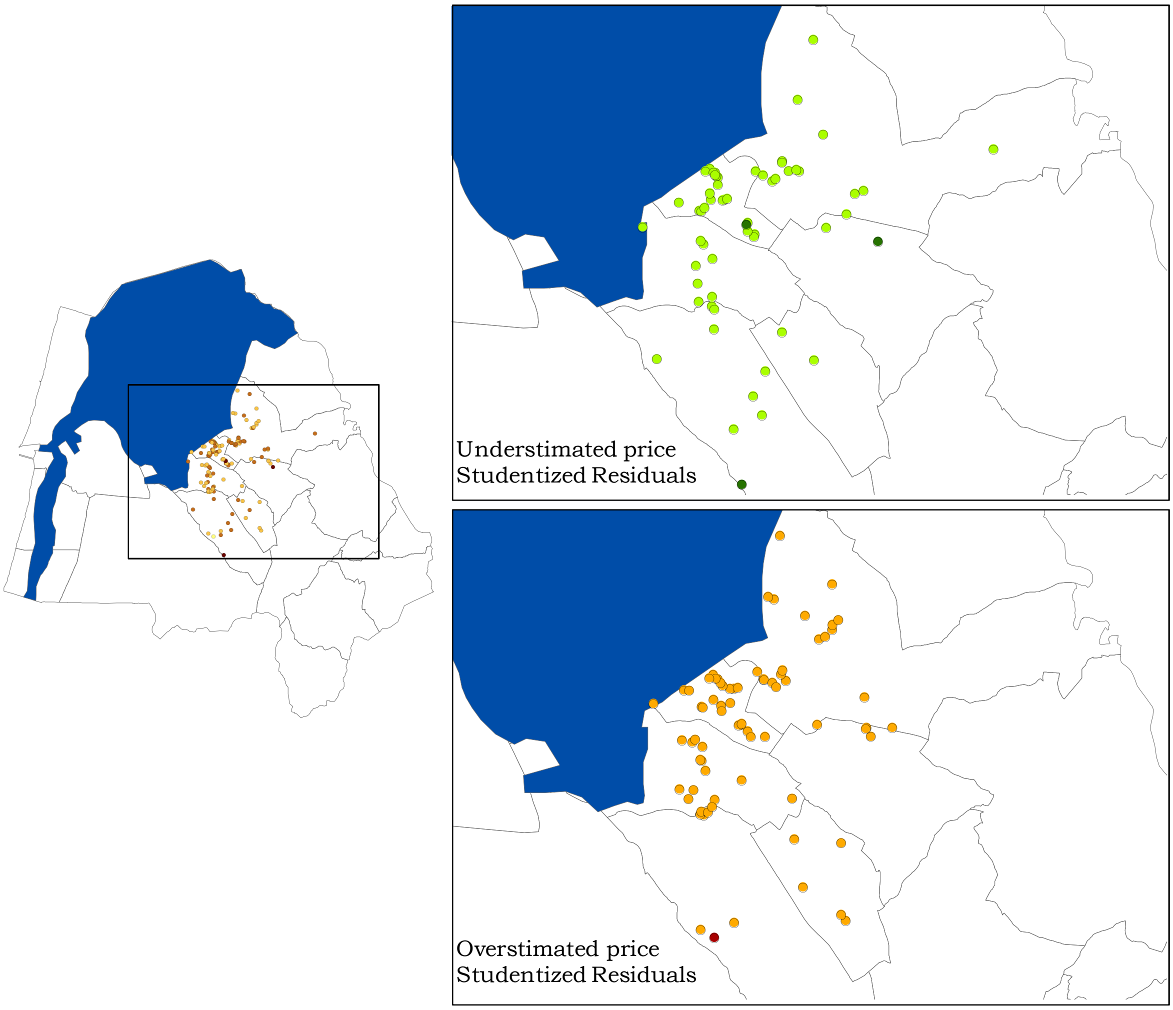


Figure 39 - Spatial analysis of error residuals

Figure 39 displays the spatial distribution patterns of standardized errors, positive and negative separately, of the model presented in the *Table 22*. A detailed analysis shows that a greater cluster of positive errors (thus an overestimated housing price) is observable in the urban parishes (Vera Cruz and Esgueira), while the negative errors are usually located in suburban areas.

A stepwise regression method is computed and is presented in *Table 23* to reduce omitted variable bias and avoid multicollinearity, as underlined previously.

Table 23 - Regression coefficients for the hedonic model, using initial variables: Stepwise method

	Coefficients		Elasticities	Euros	t
	Unstand. B	Stand. Beta			
(Constant)	10.438***				27.585
lnTotal area	-0.852***	-1.005	-57.3%	31349	-10.886
Garage space	0.228***	0.358	25.6%	-30658	6.417
Preservation	-0.288***	-0.287	-25.0%	-16550	-5.381
lnCBD Aveiro	-0.145***	-0.324	-13.5%	56013	-5.583
Type	0.376***	0.363	45.6%	20304	5.459
Number of bedrooms	0.153***	0.396	16.5%	34911	4.482
lnLiving room area	0.250***	0.156	28.4%	21056	2.797
lnKitchen area	0.158**	0.151	17.1%	11394	2.515
Balcony	0.089**	0.119	9.3%	31349	2.123
Number of Observations		**133**			
R		**81,9%**			
Adjusted R squared		**64,6%**			

*** significant at the 1% level/** significant at the 5% level/* significant at the 10% level

Contrary, to the previous results (when the enter method or simultaneous regression is used), the distance to the CBD is retained in the regression with a high level of significance. A lack of significance would be expected in the remaining location attributes, given the segregation that exists in the housing market (as explored below).

Although a full discussion of the individual coefficients would be very exhaustive, some inferences are explained as an illustration. Since a logarithmic specification is used, the estimated coefficients may be interpreted as elasticities. Note that, when the dependent variable is defined in a logarithmic form, the coefficient of dummy variable estimated should be transformed as follows:

$$100 * (e^{bi} - 1)\% \quad \text{Eq. 40}$$

Where, bi is the coefficient of the dummy variable i.

The results show, for the significant dummy variables[122] of the model presented in the *Table 23*, *ceteris paribus*, that: the provision of a garage adds 25,6% to a house value (an average of 32350€, approximately); a new house compared to a used one is 25% more expensive (30630€); the price of a house is, on average, 45% higher than a flat (55880€); the existence of a balcony increases the value of the property by 9.3% (11394€); and finally, an increase in the number of rooms increases sales price by 16.5% (20238€). The level of significance of the coefficients for the logarithmic of the distance variable suggests that the slope of price is much steeper as the location approaches the centre of the city. As expected, the logarithm of the distance to the CBD is negatively related to the sale price, indicating that moving away from the CBD, the value of houses are lower, 13,5% for each kilometre (16523€).

In sum, the hedonic model for house prices, in the urban and suburban area of Aveiro, is expressed by the following equation:

$$\begin{aligned} \boldsymbol{l}nP_{(€/m_2)} = 10.43 &- 0.85\, lnTotal_{\,area} + 0.228\, garage - 0.288\, conservation \\ &- 0.155\, CBD\ Aveiro + 0.376\, house + 0.153 bedrooms \\ &+ 0.250 lnLiving\ room_{area} + 0.158 lnKitchen_{area} \\ &+ 0.09 balcony \end{aligned} \qquad \text{Eq. 41}$$

For further information about this dataset see Marques and Castro (2007) where different specifications of the functional forms were tested. Note that the dataset is not exactly the same because different criteria for cleaning data are used.

VII.4.3. Factor based hedonic pricing model

Principal components analysis is used to extract a reduced set of orthogonal factors from the original housing attributes (explanatory variables).

[122] Dummy variables were used to represent subgroups in the entire sample.

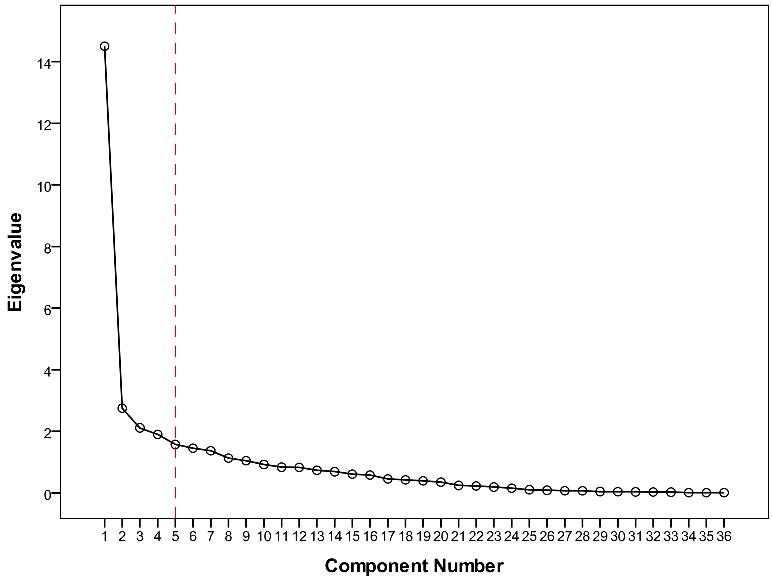


Figure 40 - Eigenvalue plot for scree test

The scree plot[123] (presented in *Figure 61*) suggests that five factors are appropriate to retain in the analysis, when considering the changes in eingenvalues. These five factors are then re-estimated by a rotated orthogonal varimax procedure. Taken together, the five factors explain 63.4% of the variance of all data.

The extracted factor loadings[124] are reported in *Table 24*; for visual clarity, the estimated loadings below the standard cut-off of 0.35 are excluded from the table. Based on these loadings, the predicted factor scores are computed for use in subsequent analysis: used as explanatory variables in the multiple regressions and spatial econometric analysis.

[123] The scree test is used to identify the optimum number of factors that can be extracted
[124] Loadings are the correlations between the variables and factors.

Table 24 - Factor loadings for the market of Aveiro

	Attributes	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Location attributes	Culture	-0.953				
	Restaurants	-0.940				
	University	0.930				
	Hotels and hostels	-0.923				
	Central Amenities	0.921				
	Sports	-0.919				
	CBD Aveiro	0.912				
	Parks and Gardens	0.876				
	Banks, ATMs, Post	-0.860				
	Local Amenities	0.839	-0.413			
	Monuments	-0.809				
	Local Commerce	0.790				
	Hospital	0.788				
	Administration	-0.784	-0.416			
	Health Centres	0.778				
	High Schools	0.733				
	Pharmacies	0.640	0.367			
	Police	0.580	0.426			
	Petrol Stations	0.397			0.374	
	Primary Schools	0.391				
	Specialised Commerce	-0.473	-0.814			
	Railway Station		0.785			
	Access Node		0.593			
Physical attributes	Gas (natural)			0.740		
	CATV			0.736		
	Floors			0.585		
	Type (House=1, Flat=0)			-0.473		
	Duplex					
	Total area				0.794	
	Number of bedrooms				0.749	
	Living room area					0.630
	Garage space					0.575
	Terrace					0.478
	Balcony					0.434
	Kitchen area					0.432
	Preservation (Used=1, New=0)					-0.362
	Total Variance Explained	**37.60%**	**8.21%**	**6.48%**	**5.65%**	**5.45%**

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization (absolute value > 0.35).

The explanatory variables included in the factorial analysis are organized into 5 factors and provide a clear interpretation in terms of the housing characteristics, with both physical and locational attributes. In general, factor 1 and 2 are related to attributes of location, whereas factors 3, 4 and 5 represent the intrinsic characteristics of dwellings. The characteristics of these factors and their contribution to overall variation are discussed below.

The first factor is responsible for 37.6% of the explained variance and arises from several indicators of centrality related to the city centre. The loadings are positive for characteristics measured in minimum distances, and negative on those measured in gravitational potential (the higher the score, the lower the centrality). This factor is denominated as *access to city centre.*

Factor 2 has an explanatory capacity of 8.2% in the total variance. This factor also describes centrality (as in factor 1) but in this case it is related to spatial elements, such as, shopping malls, railway stations, hypermarkets or motorway connections (the higher the score, the lower the centrality). The major difference between factor 1 and 2 is that the latter includes attributes located in a more dispersed way, namely, accessibility to other facilities with a low centrality. This factor is denominated as *access to local amenities.*

By contrast, factors 3, 4 and 5 represent the internal characteristics of dwellings. The third factor, with explanatory capacity of 6.5%, is related to a combination of attributes which, in the particular case of Aveiro, interact strongly with each other: being a flat or a detached house, being connected to natural gas and CATV infrastructure[125] (high values of the factor correspond to flats with gas and Cable TV). It aims to highlight whether or not homes have access to these types of services. This factor is denominated as *infrastructure access*.

With an explanatory capacity of 5.7%, the fourth factor combines, housing size with number of rooms that characterise the available space within the house. This factor is denominated *housing dimension.*

[125] Flats tend to be located in areas with high residential density which generate scale economies for the provision of this type of infrastructures.

The fifth and final factor is very similar to factor 4. Both these factors contain intrinsic attributes of housing; however these characteristics are not exclusively related with dimension. The variables included in this factor refer to additional elements, such as, the area of living room and kitchen or the existence of garage. This factor explains 5.3% of total variance and is denominated as *additional desirable features.*

VII.4.4. Housing submarket in Aveiro – spatial heterogeneity analysis

To determine whether the implicit prices differ across property types, the Ordinary Least Squares (OLS) model for each submarket is estimated. The OLS regression is used to examine the distribution of the unit price of the households in the municipality of Aveiro as a function of the set of explanatory attributes. The predicted orthogonal factors obtained above, including imputations for missing values, are used as the explanatory variables. In a total of 166 housing observations in the dataset only 118 had complete data. The possibility of imputation for missing values in the factors is considered an advantage of the factor based approach taken in this work. The total area of the house is also included as an additional regressor[126].

The regression models are estimated for the full sample and for each submarket defined by boundaries of administrative areas (parishes)[127]: submarket 1 includes São Bernardo, Aradas and Santa Joana; submarket 2 is Esgueira; submarket 3 includes Glória; and submarket 4 is Vera Cruz. The last two submarkets (Glória and Vera Cruz) are the most central areas encompassing the CBD of Aveiro, Glória being mostly residential while Vera Cruz is both residential and service oriented. Esgueira is partly urban and partly suburban.

[126] Since the dependent variable is logarithm of price per unit area, the coefficient on this regressor (βs) is expected to lie between zero and negative unity (0 and -1), with the interpretation that $1+\beta s$ is the price elasticity of house area.
[127] There are not sufficient degrees of freedom to define smaller submarkets

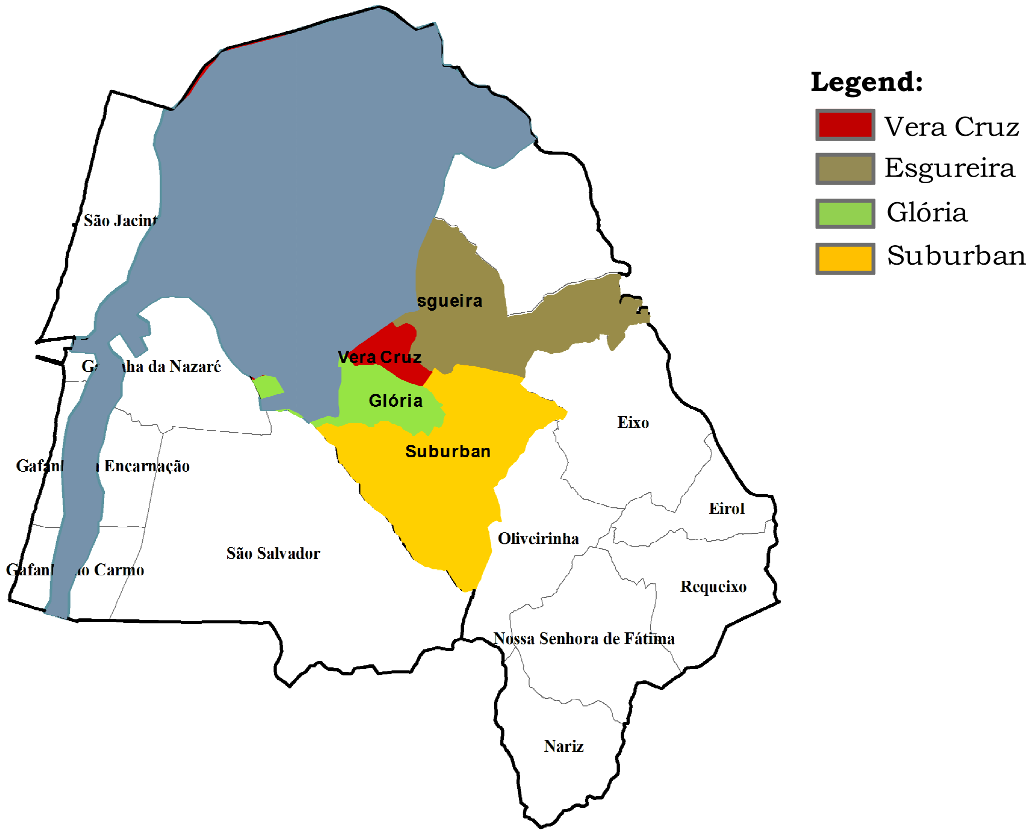


Figure 41 - Submarkets for Aveiro (micro scale approach)

The basic idea of this spatial disaggregation is to interpret spatial heterogeneity across the different sub-areas of Aveiro. For each model, the dependent variable is the natural log and the explanatory variable the score factors resulting from the factor analysis presented above.

The models presented below have the following specification;

$$\ln P_{h(€/m^2)} = \alpha_{h1} \ln A_h + \sum_{i=1}^{5} \alpha_{hi} F_{hi} \qquad \text{Eq. 42}$$

Where:

$\ln P_{h(€/m^2)}$ is the house price measured as a logarithm of euros per square metre in each submarket h (h=1, ..., 4)

$\ln A_h$ is the logarithm of the total housing area measured in m² in each submarket h

F_{hi} are factor scores (i=1,...5)

The estimated hedonic models using factors, summarized in *Table 25*, are parsimonious and offer good scope for interpretation, both in terms of individual coefficients (shadow prices) and their variation across the submarkets.

Table 25 - The estimated coefficients of the hedonic model using the factors

	Aggregate model *(All submarkets)*	**Submarket 1** *(Suburban)*	**Submarket 2** *(Esgueira)*	**Submarket 3** *(Glória)*	**Submarket 4** *(Vera Cruz)*
Constant	**11.49**	**12.05**	**10.22**	**10.64**	**11.34**
	(28.64)*	*(10.90)****	*(11.18)****	*(13.93)****	*(11.43)****
Log Total area	**-0.94**	-1.05	-0.70	-0.71	-0.90
	(-10.93)*	*(-4.66)****	*(-3.51)****	*(-4.39)****	*(-4.19)****
Factor 1	**-0.06**	-0.03	0.01	0.09	-0.23
(Access to city centre)	***(-3.76)****	*(-0.59)*	*(0.18)*	*(-1.58)*	*(-1.36)*
Factor 2	**0.00**	-0.03	-0.06	-0.06	0.26
(Access to local amenities)	***(-0.13)***	*(-0.77)*	*(-1.23)*	*(-1.22)*	*(1.49)*
Factor 3	**-0.05**	-0.09	-0.07	-0.03	0.02
(Infrastructure access)	***(-3.17)****	*(-2.14)***	*(-2.17)***	*(-0.83)*	*(0.31)*
Factor 4	**0.20**	0.26	0.05	0.16	0.16
(Housing dimension)	***(6.49)****	*(2.25)***	*(-0.52)*	*(2.68)***	*(1.63)*
Factor 5	**0.21**	0.26	0.27	0.15	0.19
(Additional desirable features)	***(10.92)****	*(4.49)****	*(8.79)****	*(4.57)****	*(3.65)****
Number of obs.	**166**	**42**	**42**	**27**	**55**
Adj R-squared	**0.583**	**0.587**	**0.736**	**0.587**	**0.332**

*** significant at the 1% level/** significant at the 5% level/* significant at the 10% level

The comparison of the regression coefficients estimates for each submarket reveals instability of the parameter estimates across equations. The statistical significance, as well as the relative magnitudes and signs of the coefficients, vary considerably across the submarket models. This is not unexpected, since housing is a heterogeneous good, and therefore the differences in the housing prices within each submarket should reflect characteristics of the property and its surrounding neighbourhood. Thus, there is strong evidence of spatial heterogeneity.

In the estimated overall model, and for the four submarkets, the explained variation (in terms of adjusted R^2) is quite high and all the regressors are highly significant, with the exception of factor 2 (*access to local amenities*). Independent variables explain 58.3% of the price variance. The sign of the factor coefficients are consistent with expectations. Housing dimension (total area) is negatively correlated with the sales price per square meter. Since the dependent variable is logarithm of price per square meter, the price per unit is lower when the area is higher, and the coefficient takes a value between zero and minus one. Additionally, the price per square meter decreases with area and with distance to the CBD (factor 1: access to city centre) and increases with factor 4 (housing space) and factor 5 (additional desirable features).

In general, and in line with the hedonic model which used initial variables, the most significant factors in the explanation of price (€/m2) are related to the intrinsic housing attributes. However, location and accessibilities also play an important role in the housing price, highlighted by the high level of significance of the factor 1. The first factor combines variables that measure the distances to facilities with a high level of centrality (e.g. culture, hotels schools, etc.) and for this reason the price per square meter of dwelling increases with their proximity (higher distances from the city centre lowers the selling price). The negative coefficient in the factor 3 (*infrastructure access*) implies a single unit house is preferable even if it implies absence of CATV or natural gas infrastructure. Because the contribution of factor 4 (*housing space*) is controlled for area, the positive sign of the coefficient means that the higher the number of rooms the higher the price. The relatively low price elasticity of living room area, about 6 percent, conceals heterogeneity across submarkets. Factor 4 has the least explanatory power, apparently indicating that buyers have excluded specialized commerce, railway station and access nodes in determining willingness-to-pay for a flat or a house. Factor 5 is related with other housing characteristics and is positively correlated with price. This means that some structural attributes, namely, garage space, terrace or balcony, contribute positively to a housing value.

The estimated models for each submarket have a relatively good explanatory power (R^2_{adj} = 0.583 for the total submarket, R^2_{adj} = 0.587 for submarket 1, R^2_{adj} = 0.736 for submarket 2, R^2_{adj} = 0.587 for submarket 3 and R^2adj = 0.332 for submarket 4). All of the individual submarket models are statistically significant.

As has been mentioned above, substantial spatial heterogeneity is observed across the 4 submarkets in terms of shadow prices for different factors, as well as the price elasticity of the total area. Analysis by submarkets shows important and interesting differences in the explanatory factors across the different parts of the city. First of all there is a substantial contrast between Vera Cruz (submarket 4) and the other parishes, showing that the traditional core of the city has a distinctive housing market. Looking at each explanatory variable it can be seen that the effect of total area is similar and highly significant everywhere, stronger in Glória and Esgueira, and weaker in the Vera Cruz and the suburban area.

The coefficient of factor 1 shows that distances to the CBD are not statistically significant in any submarket[128] but highly significant in the aggregated model; this means that distance to CBD discriminates the four areas but is not important to discriminate houses inside each submarket. Factor 2 (*access to local amenities*) is generally not significant, showing that centralities related to this factor do not provide any marginal value. This means that in Aveiro, proximity to shopping malls or hypermarkets does not increase the value of properties. Factor 3 is only significant in submarket 2 (Esgueira) and 3 (Glória). The significant negative coefficient for the factor 3 (*type of dwelling*) for the suburban area and Esgueira implies that detached houses are more valued even if they lack infrastructure facilities (like CATV or natural gas), for the more central Vera Cruz and Glória such an effect is not observed. The effect of factor 4 is the most heterogeneous across the submarkets, indicating that the importance ascribed to the number of rooms differs from area to area. Factor 5 provides similar results everywhere, in other words, the additional desirable features (living room and kitchen area and the garage space) attract a similar premium in all the 4 submarkets.

Forecast performance of the various models by cross-validation analysis has been evaluated, that is, by comparing each observation against the predicted value based on leave-one-out sample estimates omitting the index dwelling. In line with arguments in Malpezzi (2003), the factor based model generated better

[128] Given the small sample sizes in each submarket, it is not surprising that many regression coefficients are not statistically significant. The estimates indicate that, despite small sample sizes, it is important to allow for spatial heterogeneity. Further, the limitation of sample size is counterbalanced by the benefits of estimating spatial econometric models (spatial error and spatial lag models) by maximum likelihood, which is almost computationally impossible on large datasets.

predictions compared to a model with a full set of hedonic characteristics. The estimated factor hedonic model without imputed factors has a cross-validation mean squared error (MSE) that is 16 percent lower than that of a model with full hedonics included. The cross-validation MSE using predicted factors is 30 percent higher, but based on a substantially larger sample of 166 observations. On the whole, the factor based hedonic model has good predictive performance.

VII.4.5. Spatial interaction effects using a known spatial weight matrix

The results presented in this section attempt to analyse the existence of spatial autocorrelation in housing prices of Aveiro. There are several tool packages with classical estimation routines to deal with the problems of spatial dependence, and thus, build a proper model specification. In this empirical analysis GEODA software have been used.

Two different types of tests have been computed. The first group of results are global indicators which provide a preliminary analysis of the spatial autocorrelation phenomenon. Also called "*Exploratory Spatial Data Analysis*" (ESDA), this consists of a set of techniques to describe spatial distribution, identify atypical locations, and discover patterns of spatial association (Anselin and Bao, 1997). Usually a couple of global tests are used to analyse spatial autocorrelation, such as, Moran's index and Local Indicator of Spatial Association (LISA) (Anselin, 2005)[129]. Both tests consider the overall data patterns. The second group of results seek to analyse the spatial dependence in a more effective way, both in terms of the spatial lag and spatial error.

Thus, before estimating the hedonic price models with spatial effects, a check is made to see whether properties with similar square meter prices are spatially more clustered than normally expected, using Moran's I test. This test for the presence of spatial dependence is given by:

[129] Another global test for spatial autocorrelation is Geary's C (Anselin and Rey, 2008).

$$I = \frac{n}{S} \frac{\sum_i \sum_j w_{ij}(x_i - \mu)(x_j - \mu)}{\sum_i (x_i - \mu)^2} \qquad \text{Eq. 43}$$

Where: n is the number of observations; x_i and x_j are the observed prices (€/m²) in the location i and j (with mean μ); and S is a constant given by the sum of all weights (w_{ij}):

$$S = \sum_i \sum_j w_{ij} \qquad \text{Eq. 44}$$

As shown, the above statistic depends significantly on the chosen spatial weights matrix. The specification of spatial weights matrix W plays an important role in spatial models; however, the choice of spatial weights is often arbitrary and determined subjectively by the researcher, and there is usually very little formal evidence supporting such choices (Anselin, 2002).

Table 26 - Moran's I test for 7 spatial weights matrices

Distances (meters)/ **contiguity**	**Square metre price** (€/m²)
d100	0.1669
d500	0.0952
d1000	0.0954
d1500	0.1001
d3000	-0.0533
d5000	0.2263
Queen/Rook	0.1032

To ensure the robustness of the choice of the spatial weights matrix, several specifications are explored: binary weights based on distances between housings, ranging from the centre 100 (d100), 500 (d500), 1000 (d1000), 1500 (d1500), 3000 (d3000) and 5000 (d5000) meters, as well as Rook and Queen Contiguity matrices. The estimates of Moran's I statistic for these seven different specifications are reported in *Table 26*.

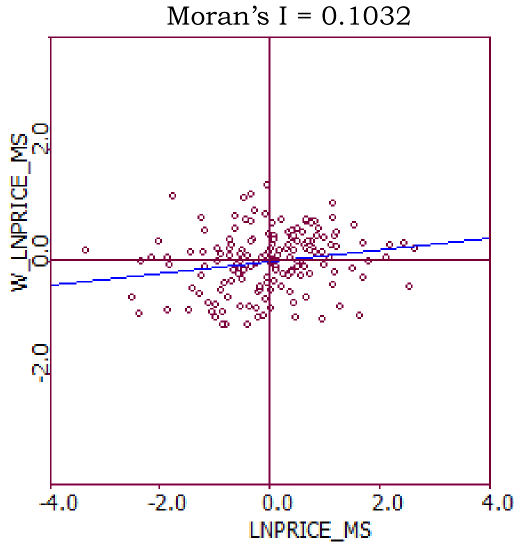


Figure 42 - Moran scatter plot for residuals (contiguity weight matrix)

Considering a contiguity weight matrix (*Queen*) the Moran's index is 0.1032 (visually illustrated in *Figure 78*). This value is a global spatial autocorrelation coefficient that varies between -1 and 1. Being positive and significantly different from zero, the above measures of contiguity or distance, evidence the existence of spatial autocorrelation of housing prices (square meter price). In this particular case, and based on the above measures of contiguity or distance, Moran's statistics tend to be positive but are not significant, showing little evidence of spatial autocorrelation in housing prices, that is, geographically adjacent observations have little or no influence on the property price.

The four quadrants in the figure provide a classification of different types of spatial autocorrelation: high-high (upper right) or low-low (lower left) for positive spatial autocorrelation; and high-low (lower right) or low-high (upper left), for negative spatial autocorrelation. A positive spatial autocorrelation means that high (low) values in a current location are surrounded by high (low) values in neighbouring observations. The slope of the regression line is Moran's I statistic, listed in blue at the top of the graph. While Moran's I index is useful for detecting the presence of spatial autocorrelation, it does not indicate the precise structure of spatial interactions (Anselin, 2005).

The possibility of existent spatial association is schematically present in the map of *Figure 43*, through the LISA indicator. It decomposes Moran's I into contributions for each location identifying areas that differ significantly from those expected under the null hypothesis that there is no association between the value observed at a location and the values observed at nearby sites (Anselin, 1995). LISA can be define as an indicator that accomplish two purposes: i) to detect significant patterns of local spatial association; and ii) to assess the extent to which the global pattern of association is reflected uniformly throughout the data set (Anselin, 1995).

This test statistic is defined as:

$$I_i = y_i \sum_{j=1}^{n} w_{ij} y_j \qquad \text{Eq. 45}$$

Where, y_i is variable observed at location I and w_{ij} are elements in a distance-based weights matrix.

The interpretation of this indicator is similar to that of Moran's I, that is, four types of association can be found: High-High and Low-Low, if the association is positive; and High-Low and Low-High, if the association is negative. The results are shown in *Figure 43*.

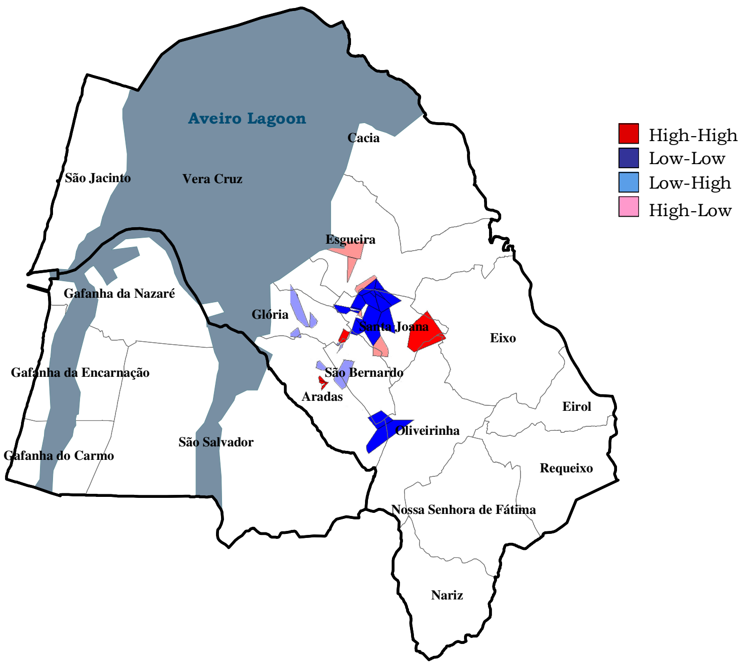


Figure 43 - LISA indicator

The second group of spatial analysis tests includes a series of Lagrange Multipliers (LM) tests, for both spatial lag dependence and spatial error dependence. The first step is to estimate an OLS initial model in order to calculate the residuals. Theses residuals are then used to test the hypothesis of no spatial dependence caused by spatially autoregressive errors (LM-error test) or by omitted spatial lag (LM-lag test). When the hypothesis cannot be rejected the results from the OLS model may be used, in other words, there is an absense of spatial dependence (Anselin, 1988). On the other hand, if the null hypotheses, for one or both tests are rejected a new model should be estimated. According to the literature (and as shown in the *Figure 44*) if both tests are rejected (LM-error and LM-lag tests) the proper model should be the most significant; in the case that only one test is rejected or significant (LM-error or LM-lag tests) a Spatial Lag Model or a Spatial Error Model, should be estimated.

The *Table 27* reports, in addition to OLS (a model without spatial effects), the estimation (by maximum likelihood - ML) of alternative spatial regression models and investigates whether a spatial error dependence (SED) or a spatial lag dependence (SLD) model is appropriate. Thus, the second column represents the traditional model and includes Moran's I statistic and Lagrange Multiplier (LM) tests to identify the need for a spatial model. The results are presented for a spatial weights matrix based on Rook and Queen Contiguity (other specifications of spatial weights have similar results).

Table 27 - OLS, SLD and SED model estimates

Variables	**No spatial dependence (OLS)**	**Spatial lag model (ML estimation) Coefficient**	**Spatial error model (ML estimation)**
Constant	11.49 (28.64)***	11.31 (14.66)***	11.55 (29.28)***
log Total area	-0.94 (-10.93)***	-0.94 (-11.18)***	-0.95 (-11.26)***
Factor 1	-0.06 (-3.76)***	-0.06 (-3.28)***	-0.06 (-3.42)***
Factor 2	-0.00 (-0.13)	-0.00 (-0.17)	-0.00 (-0.13)
Factor 3	-0.05 (-3.17)***	-0.05 (-3.18)***	-0.05 (-3.08)***
Factor 4	0.20 (6.49)***	0.20 (6.66)***	0.21 (6.60)***
Factor 5	0.21 (10.92)***	0.21 (11.06)***	0.22 (11.19)***
Lagrange Multiplier (lag)	0.08 (*p-value* 0.77)		
Robust LM (lag)	0.27 (*p-value* 0.61)		
Lagrange Multiplier (error)	0.67 (*p-value* 0.41)		
Robust LM (error)	0.86 (*p-value* 0.35)		
Lagrange Multiplier	0.94 (*p-value* 0.63)		
Number of observations	**166**	**166**	**166**
R^2	**0.598**	**0.598**	**0.600**
Log likelihood	20.404	20.442	20.753
Lag coefficient(Rho)		0.026 (*p-value* 0.78)	
Lag coefficient (Lambda)			0.109 (*p-value* 0.37)

*t-/z-statistics in parentheses; ***significant at the 1% level/ **significant at the 5% level/ significant at the 10% level*

Like Moran's I statistic, there is no evidence of spatial dependence. This is despite the fact that spatial heterogeneity in these estimates has not been accounted for (a feature that can contribute to spatial dependence). Neither the LM-error nor the LM-Lag models are significant. The null hypothesis of both tests, which is the lack of spatial dependence, cannot be rejected at the 5 percent significance level. If the tests indicate significant spatial dependence, this needs to be modelled using the appropriate spatial dependence model. However, in this case, the null hypothesis of spatial dependence cannot be rejected; both ρ and λ have a level of significance equal to 0.41 and 0.77 (not significant), respectively. Therefore, dependence is either absent or not related to the geographical notions of distances and contiguity considered in the above seven specifications. This highlights an important limitation of spatial econometric methods for studying hedonic pricing models. Specifically, while the choice of appropriate spatial weights is a central component of spatial models, interaction between observation units often cannot be either precisely measured or in other ways explained by observed measures of distance. In other words, treating spatial dependence as the outcome of spillover processes which are dependent on previously known, fixed and arbitrary spatial weights matrices cannot be an adequate procedure.

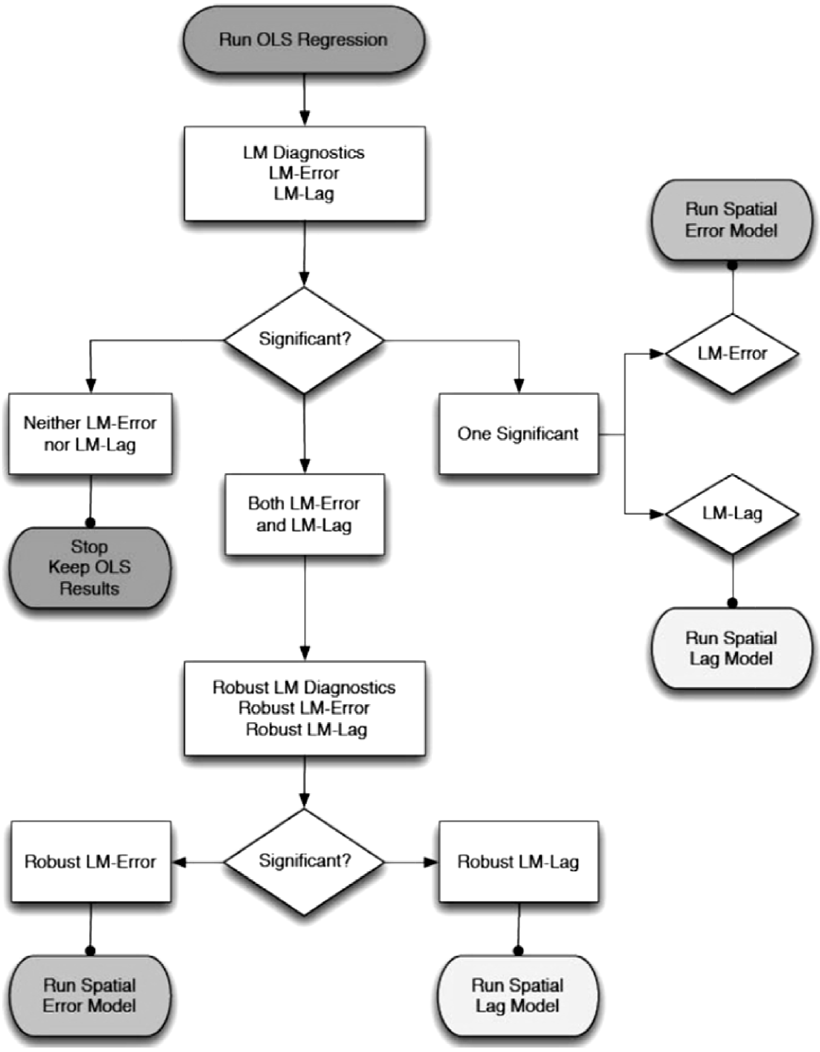


Figure 44 - Decision tree for the choice of the best model

VII.4.6. Spatial interaction effects using an unknown spatial weights matrix

As discussed before, this thesis extends an emerging area of research that takes a non-parametric view of the nature and strength of spatial diffusion and cross section interaction. Moving away from the usual practice of an *ex ante* definition of spatial interactions, the unknown spatial weights matrix is estimated and the result is consistent with an observed pattern of spatial dependence, representing a significant departure from the usual practice of assuming *a priori* the nature of spatial interactions. Once these interactions have been estimated, they can be subjected to interpretation in order to identify the true nature of spatial dependence. Specifically, as proposed in Chapter VI, an extension of the Bhattacharjee and Jensen-Butler (2005) estimator to the pure cross-section setting is used to obtain estimates of a symmetric spatial weights matrix under a spatial error model. The symmetry assumption adopted in this work is in line with the traditional practice in housing studies, and is a natural consequence of defining spatial weights based on distances.

Table 28 - Spatial error autocovariance matrix across submarkets

Submarkets	Submarket 1 (Suburb)	Submarket 2 (Esgueira)	Submarket 3 (Glória)	Submarket 4 (Vera Cruz)
1 (Suburb)	0.057			
2 (Esgueira)	–0.042	0.033		
3 (Glória)	0.085	0.142	0.050	
4 (Vera Cruz)	–0.150	0.031	–0.079	0.045

The first step is to estimate the spatial autocovariance matrix of residuals across the four submarkets. As discussed in Chapter VI, residuals are used across the four submarkets, matched by factors, to construct the cross-submarket error spatial autocovariance and autocorrelation matrix *(Table 28*). In contrast to the results above, based on *a priori* fixed spatial weights, significant spatial autocorrelation can be observed between some submarkets, but this autocorrelation is not strongly linked to contiguity (traditional geography). Next, the estimation method presented in Bhattacharjee and Jensen-Butler (2005) is used. A matrix of symmetric spatial interactions (spatial weights matrix W_2) consistent with the above autocovariance matrix is estimated and is reported in *Table 29*.

Table 29 - Symmetric spatial interaction matrix

Submarkets	Submarket 1 (Suburb)	Submarket 2 (Esgueira)	Submarket 3 (Glória)	Submarket 4 (Vera Cruz)
1 (Suburb)	0.00			
2 (Esgueira)	–0.024	0.00		
3 (Glória)	0.041***	0.074***	0.00	
4 (Vera Cruz)	–0.072***	0.017	–0.037	0.00

**** significant at the 1% level/ ** significant at the 5% level/ * significant at the 10% level*

The *Table 29* reports the corresponding estimated symmetric spatial weights matrix for cross-submarket interactions and, like the corresponding autocorrelations (*Table 28*), are not related to any pre-assumed notions of geography or contiguity. However, some spatial interactions implied by the estimates are significant, and offer interesting insights into the nature of spatial dependence across the submarkets of Aveiro. Results are consistent with the spatial structure of Aveiro, showing that Vera Cruz has a highly significant negative interaction with the suburban submarket, while Glória has a highly significant positive interaction with both the suburban submarket and Esgueira.

These observations can be explained by the urban geography of Aveiro. Vera Cruz represents a distinct housing market in the CBD of Aveiro and draws its housing demand from a population quite different from the inhabitants in large detached houses in the suburban area. Such segmented markets imply that negative spatial interactions are likely between these two submarkets. On the other hand, Glória and Esgueira are largely residential submarkets close to the centre and are likely to offer positive spillovers, and likewise for Glória and the suburban area which are contiguous.

VII.5. Housing market of urban and suburban area of Aveiro: Macro scale approach

The empirical analysis presented in this section is extended to the housing market of the suburban and rural area of the municipalities of Aveiro and Ílhavo. The database covers a more heterogeneous area, when compared with the previous dataset, which enables a richer interpretation, both in terms of spatial heterogeneity and spatial dependence. The organization of this section follows the structure of the previous one, that is: an explanatory analysis of data, the definition of an hedonic model using initial variables (a descriptive and an explicative approach) and factors (provided by a factor analysis), spatial heterogeneity analysis (using an inductive and a analytical approach), and spatial dependence analysis (using known and unknown spatial weights matrices). An additional section is presented before these topics describing the data cleaning process of the dataset.

The housing market of Aveiro, analysed by a macro scale approach, includes the 14 parishes of the municipality of Aveiro and the 4 parishes of the municipality of Ílhavo. The population is distributed over three main areas: i) the rural area, has approximately 12% of the population of Aveiro and Ílhavo, and includes the parishes of Cacia, Eirol, Eixo, Gafanha da Encarnação, Gafanha do Carmo, Nariz, Nossa Senhora de Fátima, Oliveirinha, Requeixo and São Jacinto; ii) the suburban area, with 55% of the population contains six parishes, Aradas, Esgueira, Santa Joana, São Bernardo, São Salvador and Gafanha da Nazaré; and finally, iii) the urban area, which is the urban city centre of the Aveiro region and encompasses just two parishes (Vera Cruz and Glória), having 33% of the population.

The database used for this empirical work is provided by the firm Janela Digital S.A., which owns and manages the real estate portal database CASA

SAPO. This portal created in 2000 "(...) *is the largest site in Portugal of real estate diffusion. Currently, its database has a portfolio with more than 1400 real estate agencies and about 500000 properties, distributed throughout the country*" (http://www.casa.sapo.pt). Since 2000 (from October 2000 to March 2010) this database collated about 4 million records of properties available for transaction in Portugal during the last decade, covering all the national territory, and hence, concentrates valuable information for understanding the Portuguese real estate market[130]. *Figure 45* and *Figure 46* describe the distribution of properties in the Aveiro and Ílhavo (in absolute number and percentage) by parish available in the portal and the evolution of recorded additions during the studied time period, respectively.

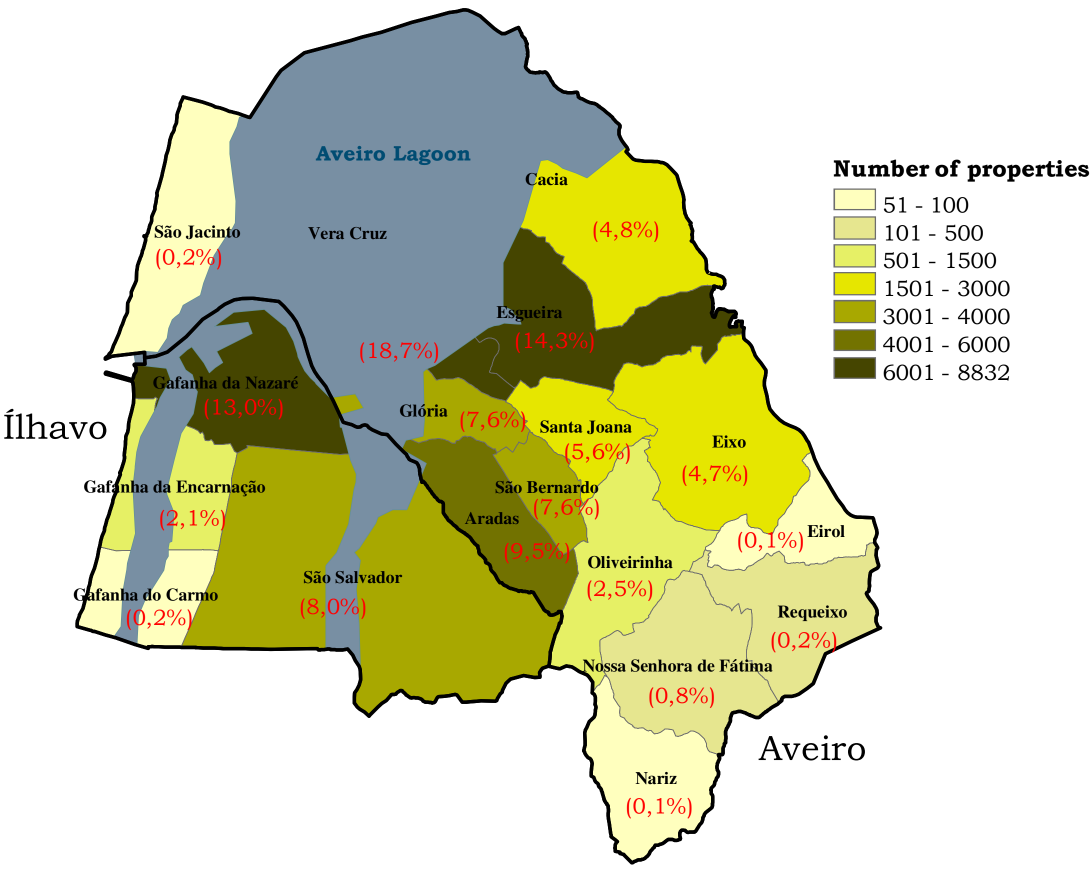


Figure 45 - Distribution of houses by parish in 2001 in Aveiro and Ílhavo

130 House advertising is mainly placed by real estate agencies and less often by the owners themselves. The portal information can be freely accessed by any potential buyer.

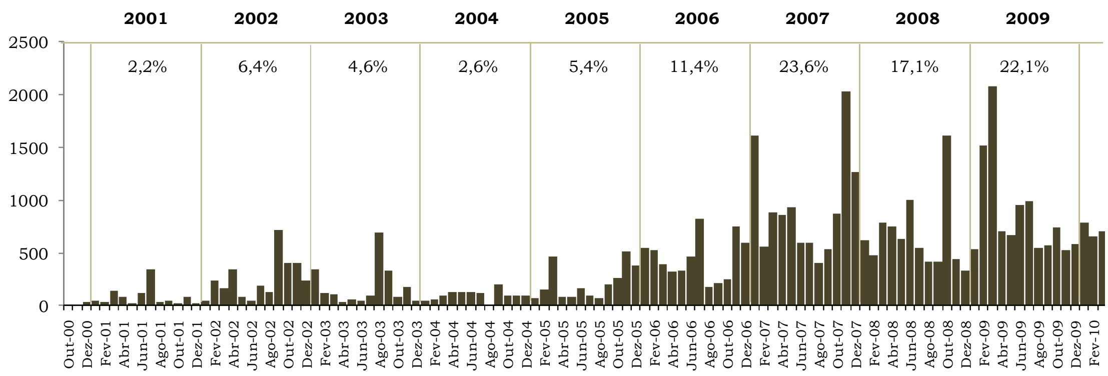


Figure 46 - Number of properties added to the database between 2000 and 2010

In a preliminary assessment of the data quality, four major challenges to further statistical analysis were identified in this dataset. The first aspect is the inconsistency of some data, resulting from aberrant values filed in the dataset by real estate agents and owners. A second problem is the existence of missing values. This, situation is particularly critical for the observations without location information, such as, zone or parish, making the spatial econometric analysis of those properties impossible. A third challenge is the construction of new attributes from the initial variables, both combining existing attributes and extracting useful information (housing characteristics) from the description field (where people introduce a free textual description to describe properties). Finally, a fourth major challenge is duplicate entries. Many records have exactly the same information in all fields (approximately 26%), caused essentially by the lack of exclusivity in placing housing market advertisements. Real estate agents are allowed to introduce the same property more than once, with the same characteristics. These records are not identifiable because of privacy protection rules. However, two observations with the same information do not mean that is the same property, it can be several houses located in a single building with identical types of attributes. The option taken therefore is not to remove any duplicate cases[131].

In order to overcome some of these problems, data cleaning processes are used based on the procedures described in the next section.

131 Additional explanatory analyses have been done removing the duplicates where the main inferences did not change. See for example: (Batista, Castillo *et al.*, 2011)

VII.5.1. Data cleaning and creation of new variables

This study, as mentioned above, considered information for the municipalities of Aveiro and Ílhavo, in a total of 56570 housings. With the purpose of this work being the analysis of residential housing, other types of properties (shops, garages, farms, land, etc.) not included in this category, were removed. After this filtering procedure the database is reduced to 47188 observations. Initial variables included in the *Casa Sapo* database and a summary of the more relevant descriptive statistics (before data cleaning) are reported *Table 30*.

Table 30 - Initial variables included in the initial database

Code	Descriptive analyses
Nature (of the property)	83.4% are housing; 16.6% are other properties (such as lots, garages, commercial spaces etc.)
Type (Type of dwelling)	1-Flat in a House (0.6%); 2-Flat (72.9%); 3-House (16.5%); 4-Semidetached house (1.8%); 5-Terraced house (4.7%); 6-Detached house (3.4%); 7-Old House (0.1%)[132].
Typology (Number of bedrooms)	T0 (1.8%); T1 (13.8%); T2- (34.0%); T3 (27.4%); T4- (19.4%); T5 (2.4,%); T6 (0.2,%); >T6 (38 cases)
Municipality (where the housing is located)	Aveiro (76.7%); Ílhavo (23.3%)
Parish (where the housing is located)	18 parishes (14 in Aveiro and 4 in Ílhavo) number of houses in each parish is presented in *Figure 45*
Zone (where the housing is located)	103 zones (number of houses in each zone is presented in *Figure 48*)
Business (Type of business transaction)	For sale (90.1%); For rent (9.1%)[133]
Preservation (Level of preservation)	1-New (27.0%); 2-Under construction/project (19.8%); 3-Restored (0.2%); 4-Used (39,0); 5-To recover (13.9%).
Price (in euros)	Mean (149558); Standard Deviation (806778); Min (1); Max (99450000)
Area (net)	Mean (203); Standard Deviation (2730); Min (0); Max (22420)
Area (all)	Mean (183); Standard Deviation (652); Min (0); Max (52019)

(cont)→

132 In the database 24.5% are houses and 75.5% are flats.

133 Other types of arrangement exist in the initial database, but were not significant in terms of value: 29 cases (e.g.: auction; transfer; exchange)

Area (land)	Mean (787); Standard Deviation (2742); Min (0); Max (50000)
Age (of construction)	Mean (1987); Standard Deviation (175); Min (1); Max (2012)[134]
Description (free text)	Variable is string type and is a free text field where agents (individuals or firms) characterise the property
Date of entrance (When the property was added to the database (day/month and year))	Date between 20/Oct/2000 and 20/Mar/2010 (evolution of number of properties added to the database is presented in *Figure 46*)
Date of exit (When the housing was removed from the database (day/month and year))	Date between Oct/2000 and Mar/2010 (Time on market - TOM of property by year is presented in *Figure 51)*

A brief analysis of the results (in Table 30) shows some evident irregularities and problems detected in the data, namely, anomalous outliers and a big amount of missing values (the initial database has only 133 cases without blank cells). The records with no information for sale prices and those who could not be geocoded by micro zones or parishes are deleted.

Three main categories of variables are used to describe each property. The first group of variables is related to *physical or intrinsic housing attributes*, which characterises each dwelling in terms of: type of dwelling, number of bedrooms, preservation, price, age of construction and area (gross, net and land). A second group of variables addresses the approximate location of houses. Thus, *housing location* is described by: municipality, parish or zone. And finally, the *description field* which is a free text field where real estate advertisers write in their own words all the information that they consider relevant for a creating a good business opportunity. A more detailed description of the main transformation on each of these groups of variables is presented below.

i) *Physical or intrinsic housing attributes*

This group of variables characterises each property considering its basic physical structures, and are, as already mentioned, quantitatively and qualitatively limited by its availability (given by real estate agents and individual owners). *Figure 47* summarises the initial variables considered in this analysis and some of the transformations that are made.

[134] It seems evident in this field that confusion arises between the age of the property and the year of construction.

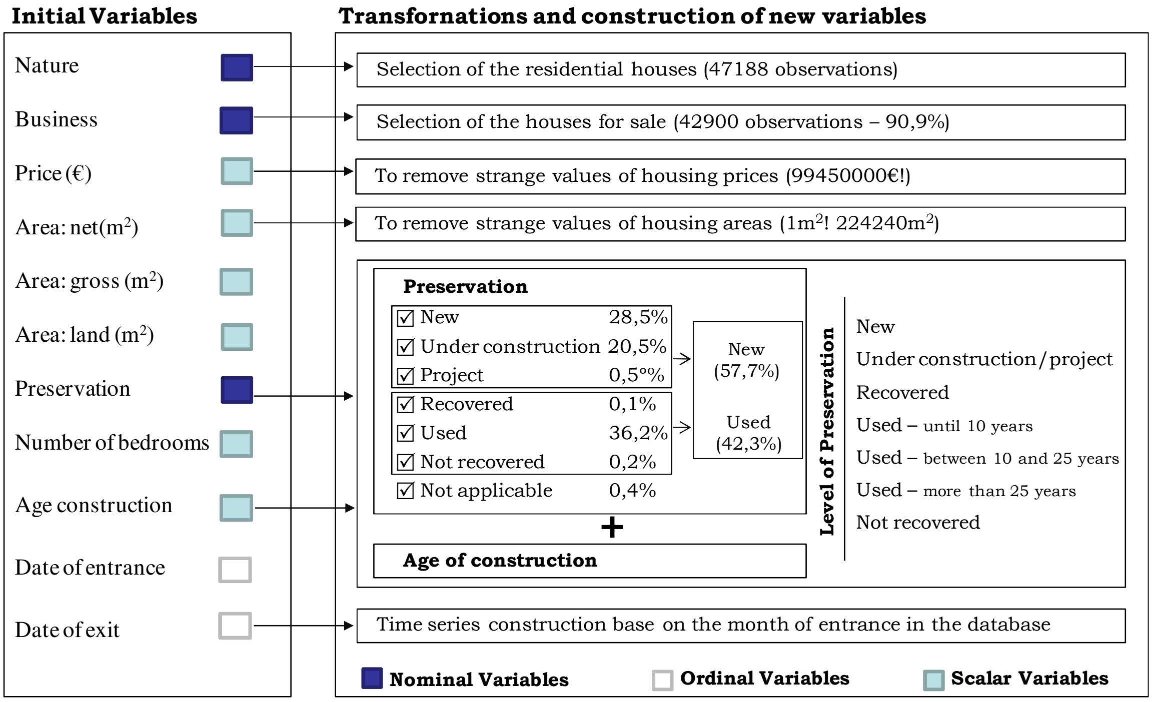


Figure 47 - Transformation of the initial variables

Besides some filter procedures, to exclude all properties other than dwellings (garages, pharmacies, hotels, rooms, shops, land, etc.) and those properties that were advertised for renting purposes, some major transformations are carried out to the variables for preservation and age of construction. These two variables are merged into one unique attribute, named 'level of preservation'. The criteria used for this purpose is the following:

- New (blank cells in the preservation field and age equal to 0 or 1)
- Under construction/In project phase (blank cells in the preservation field and age less than 0 [-3,-2,-1])
- Restored (if preservation field holds a value of new and the age is greater than 5 years)
- Used less than 10 years (blank cells in the preservation field and age between 2 and 9)
- Used between 10 and 25 years (blank cells in the preservation field and age between 10 and 24)
- Used more than 25 years (blank cells in the preservation field and age greater than 25)
- Dwelling to renovate

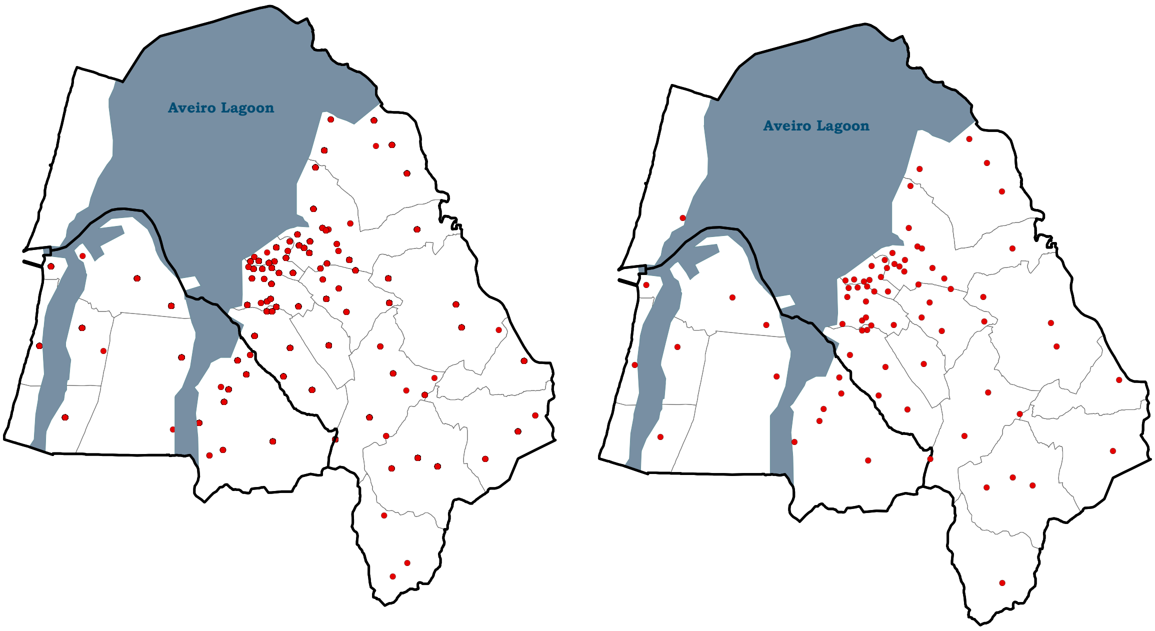
For the remaining numerical variables (areas and price) erroneous values have been removed and inconsistencies have been corrected.

ii) *Housing location*

Three types of variables are included in the database to identify the place where a specific property is located: two of them related to administrative boundaries (municipalities and parishes), and one, representing homogeneous territories, smaller than parishes, which in some cases are identified as residential neighbourhoods, but in others cases match the designation of parish or a more abstract location, such as a centre or beaches, for instance. These more disaggregated units of analysis are designated by zones (see *Figure 48*).

Property georeferentiation is done using the field micro zone. Because of confidentiality problems, the exact position of observations within each zone in unknown, which means that all observations located within a specific zone are attributed a unique coordinate, and thus the same accessibility to several urban amenities.

In this process of georeferentiation it has been decided to remove the following cases: i) those observations without any zone identification; ii) those observations with a very widespread designation of zone that prove difficult to identify as a point in space (e.g. centre); and iii) those observations where the number of cases is less than 10 (zones considered relatively homogeneous were merged). On the other hand some imputation is carried out. Those situations with missing values in the zone field, belonging to rural and identical parishes, assume the name of the parish in the zone. This procedure is applied in the following parishes: Nariz, Requeixo, Eirol, Gafanha do Carmo and S. Jacinto. The result of georeferentiation, following these assumptions, is illustrated in *Figure 48.* These procedures reduce the number of zones from 103 to 76.



103 initial zones 76 zones after data cleaning

Figure 48 - Georeferentiation of zones

Given a set of points in the plane, the associated set of regions surrounding these points, are obtained by the spatial Thiessen polygons or Voronoi diagrams (Figure 49). This method is used to define the approximate limits of each zone.

Figure 49 - Thiessen Polygons of the zone

After filtering the dataset for housing and for sale purpose, and after some procedures of data cleaning, the distribution of 12467 dwellings by each zone is as presented in the following *Table 31*.

Table 31 - Number of housings by zone

Micro Zones	Cases	Micro Zones	Cases	Micro Zones	Cases
Agras	2	Eirol	6	Paco	92
Agras do Norte	11	Eixo	281	Patela	223
Alagoas	526	Escolas	113	Povoa do Paco	97
Alboi	89	Esgueira	51	Povoa do Valado	3
Aradas	32	Estacao	35	Quinta da Bela Vista	119
Av. Dr Lourenco Peixinho	36	Eucalipto	189	Quinta do Cruzeiro	233
Azenha de Baixo	102	Feira de Marco	121	Quinta do Gato	4
Azurva	563	Forca	635	Quinta do Loureiro	129
Bairro de Santiago	43	Forum	2	Quinta do Picado	139
Bairro do Liceu	169	Gafanha D'aquem	82	Quintas	127
Barra	705	Gafanha da Encarnacao	258	Requeixo	41
Barrocas	975	Gafanha da Nazare	1444	Ribas	44
Beira Mar	70	Gafanha do Carmo	40	Rossio	100
Bonsucesso	193	Glicinias	151	Santiago	35
Cabo Luis/Qta das Acacias	3	Granja de Baixo	13	Sao Bernardo	1143
Cacia	13	Gulbenkian	2	Sao Jacinto	28
Cale da Vila	63	Mamodeiro	29	Sarrazola	235
Cancela	29	Mario Sacramento	102	Sol Posto/Presa	3
Carramona	109	Mataducos	152	Taboeira	69
Centro (Ilhavo)	1158	Moitinhos	26	Verdemilho	233
Centro de Congressos	6	Nariz	37	Viaduto	2
Cidadela/Qta Sto Antonio	7	Nossa Senhora de Fatima	94	Vila Jovem / Santiago	21
Cilhas	17	Oita	7	Vilar	217
Costa do Valado	118	Olho d'agua	7	Viso/Caiao	17
Costa Nova	100	Olho D'agua	2	Vista Alegre	36
Coutada/Medela	37	Oliveirinha	22		

iii) *Description field*

So far, some outliers were removed and some transformations in the data have been made. In addition, other variables are created from the description field. In this field, sellers could include descriptive information (sometimes with spelling mistakes) considered interesting for advertising the property; however, this does not means that the house has such specific attributes, instead real estate advertisers considered the attribute relevant to describe the dwelling. This field is a free text attribute, where additional information about housing particularities could be addressed (such as, existence of a terrace, balcony, garage, etc.) or other more subjective descriptions (such as, nice location, close to the centre etc.).

Words that appear more frequently in the free text field (and generated by software available in http://www.wordle.net/) are highlighted in *Figure 50*. These words are in Portuguese because it would be unreasonable to translate all the information written is this field. The bigger the font used, more frequent is the word in the database.

Figure 50 - Word clouds for the housing characteristics in the field description

To extract and transform descriptive and textual information contained in the *Casa Sapo* database into computationally manageable integer data the following criteria are used: i) if no description was found than a missing value was registered; ii) if the specific housing characteristic word appeared than a value of 1 was registered; and finally, iii) if the word did not appear than a 0 value was registered.

Apart from the mentioned problems, this field provided useful additional information about the property (14.5% of the cases have at least some description). Applying automated text processing to the free-text description of each advertised property in the registry, information regarding the existence of some important housing features in each property is extracted, creating a total of 13 binary (dummy) variables. The most important housing features are reported in Table 32. Note that, other attributes have been found (e.g.: whirlpool, storeroom, attic etc.) but were not considered in further applications.

Table 32 - Intrinsic attributes created from the description field

New variables	**Frequency**
Garage	51.0%
Central heating	33.8%
Balcony	30.8%
Fireplace	21.3%
Terrace	15.0%
Garage space	14.0%

Note: The distintion between Garage and Garage space is that a Garage is a closed individual space while a Garage space is a place in a shared area.

VII.5.2. Explanatory analysis of the initial variables

Following the same approach as with the previous small dataset, a short descriptive analysis and statistical inference overview is carried out for the most relevant initial variables, looking at the distribution and statistical dependences.

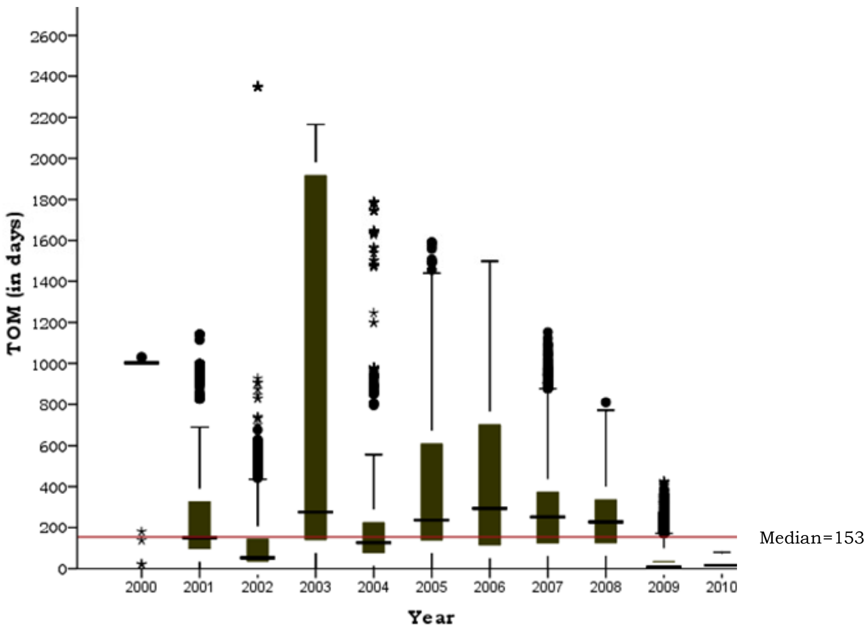


Figure 51 - Time on market (TOM) of properties, in days

A general idea of the time that properties are available in the *Casa Sapo* portal is reported in *Figure 51*. Note that the 'date of exit' does not mean that the house was sold, but simply means that the property was removed. The median time on market is 153 days with a tendency for this figure to decrease over the observation period. The year of 2003 is an outlier in the overall distribution where some houses remained on the site more than 2000 days (five and a half years).

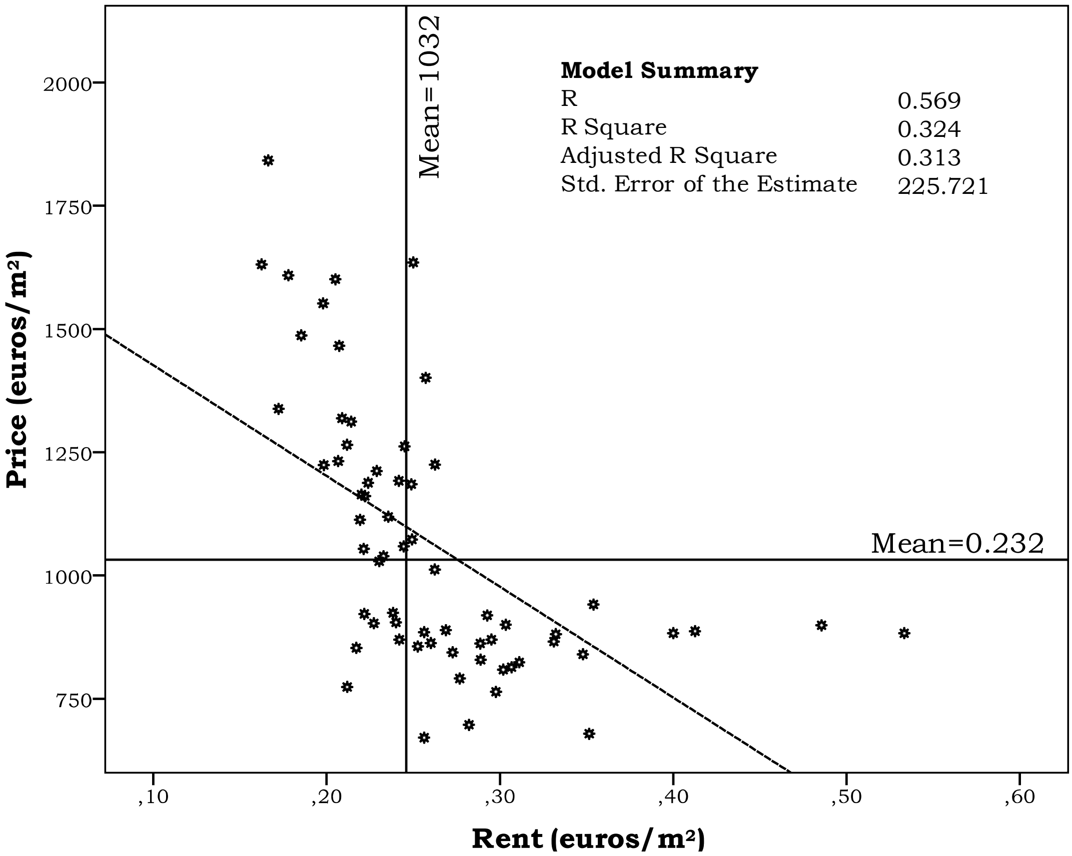


Figure 52 - Relation between sale prices and rents by zone

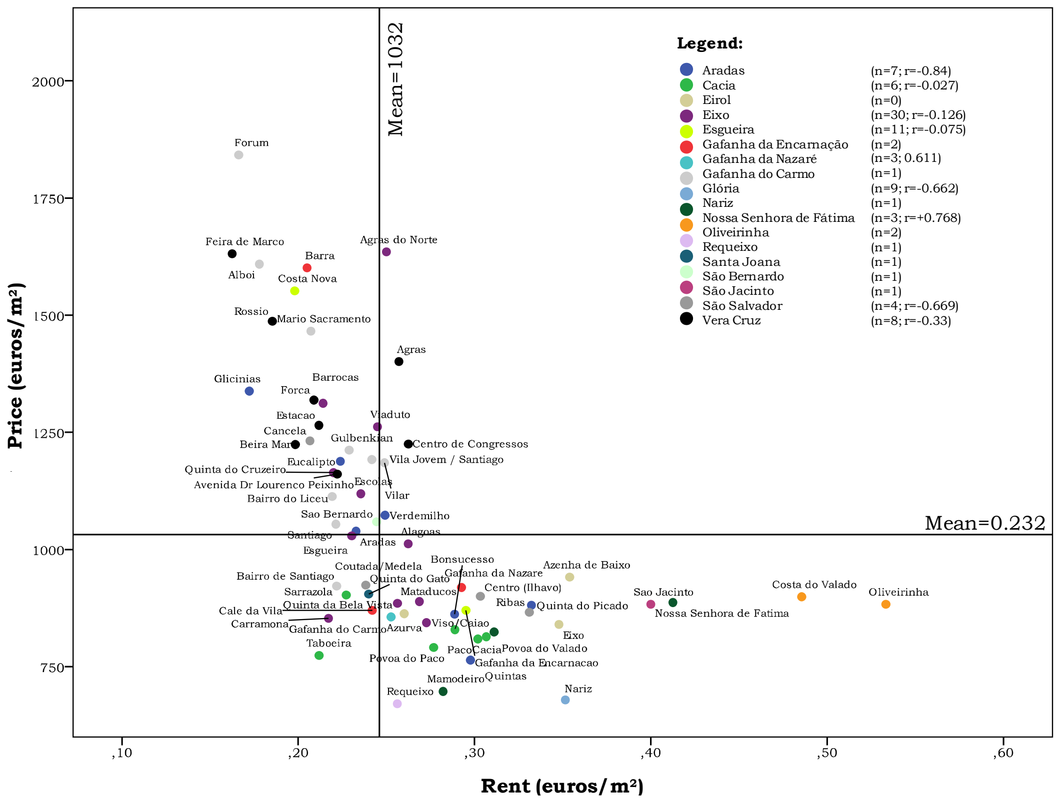


Figure 53 - Relation between price and rent values by zone and by parish

In the Figure 52 and Figure 53 the relation between the value of housing (euros/per square meter) and the rent are plotted, where each dot represents the average of values and rents in every zone. As shown in Figure 52, there is a negative linear[135] correlation (-0.57) between price (euros/m^2) and the rent (euros/m^2), meaning that, zones with high house price values have a corresponding low rental value, and vice-versa, or, in other words, zones with properties with a high rent are sold at a low price. In Figure 53 it is possible to see the same relationship between these two indicators but distributed by zone and parish. The scatterplot shows that, typically, dwellings located in urban zones have a higher price but lower rents, whereas higher rents appear in zones predominantly located in the outskirts of the city.

In *Table 33* the final physical housing attributes and corresponding descriptive statistics (means, standard deviation, minimum, and maximum) are reported. These variables are used as independent variables in an hedonic model of Aveiro and Ílhavo (jointly with the location attributes of *Table 19*), and as inputs for computing the factor analysis, which are consequently used as determinants for the study of spatial heterogeneity and spatial dependence.

An important difference in relation to Section VII.4 is that the price considered in this dataset is the list price, rather than sale price. Not incorporating the real housing market values would lead to some expected bias in the characterization of the real estate market, however, there is no reason to believe that the asking prices vary significantly across all properties, and thus interfering with the regression weights. In addition, the gap between listing and selling prices is compensated by including the logarithm of time on the market (in days) and the date (year) in the regression, fixed effects to help control for aggregate cyclical and political factors.

[135] The correlation seems to exhibit convexity, but a linear regression has been used to simplify.

Table 33 - Descriptive statistics of the intrinsic variables of houses for sale

		Units	N	Min	Max	Mean	Std. Deviation
Internal physical characteristics							
d	Type	(House=1, Flat=0)	12467	0.00	1.00	0.28	0.45
ln	Number of bedrooms	(Number)	12467	0.00	2.48	1.23	0.33
d	Duplex[136]	(Yes=1; No=0)	12467	0.00	1.00	0.12	0.33
d	Pres: New	(Yes=1; No=0)	12467	0.00	1.00	0.31	0.46
d	Pres: Under constr.	(Yes=1; No=0)	12467	0.00	1.00	0.25	0.43
d	Pres: Restored	(Yes=1; No=0)	12467	0.00	1.00	0.00	0.06
d	Pres: Used <10 years	(Yes=1; No=0)	12467	0.00	1.00	0.34	0.47
d	Pres: Used 10-25 years	(Yes=1; No=0)	12467	0.00	1.00	0.08	0.27
d	Pres: Used > 25 years	(Yes=1; No=0)	12467	0.00	1.00	0.01	0.11
d	Pres: Not restored	(Yes=1; No=0)	12467	0.00	1.00	0.00	0.03
ln	Price	(euros/m^2)	12467	5.19	8.65	7.03	5.91
ln	Total area	(m^2)	12467	3.00	6.40	5.01	4.38
ln	Time on the market	(Days)	12467	0.00	7.76	5.00	1.64
d	Balcony	(Yes=1; No=0)	12467	0.00	1.00	0.39	0.49
d	Terrace	(Yes=1; No=0)	12467	0.00	1.00	0.18	0.39
d	Garage space	(Yes=1; No=0)	12467	0.00	1.00	0.16	0.37
d	Garage	(Yes=1; No=0)	12467	0.00	1.00	0.64	0.48
d	Central heating	(Yes=1; No=0)	12467	0.00	1.00	0.43	0.50
d	Fireplace	(Yes=1; No=0)	12467	0.00	1.00	0.29	0.45

d=dummy variable; ln= in logarithms; p=gravitational potential

As occurred in the small dataset there are large variations in the physical attributes across data here. The average price (in euros per square meter) is 1126 euros and ranges from 178 up to 5714 euros, while average dimension per dwelling is 149 m^2. The smallest dwelling in the sample has a floor area of 20 m^2 while the largest has 600 m^2. These values are quite different from the other database, justified by the coverage level of the two samples. Because the larger dataset involves the rural part of the city of Aveiro, characterised by the existence of single houses (instead of flats), the average price is lower and the average area is higher. Regarding other housing characteristics, the distribution is the following: 28.4% are single houses and 71.6% are flats, 12.3% are duplex, 39.3% have balcony; 18.2% have a terrace; 16.1% have garage space; 63.8% have a garage; 43.3% have central heating; and 28.9% have a fireplace. In general, this dataset has identical proportions of intrinsic housing attributes when compared with the dataset analysed previously.

[136] A flat with two floors.

In the box plot diagrams presented below (in *Figure 53, Figure 54* and *Figure 55*) the housing price variability is shown (in absolute and relative terms), according to three indicators: location, type of dwelling and number of bedrooms. Without looking at all the details, it is important to note that the price of housing is higher in urban areas (value per m^2), but if we analyse the value of houses in absolute terms, this argument is not valid. The reason for this is partly explained by *Figure 55*, that is, the predominance of flats (typically smaller) in urban areas and houses (typically bigger) in suburban and rural areas. The last diagram (in *Figure 56*) shows a non-linear relationship between the size of the dwelling and the price per m^2, which is taken into consideration in the model specification (such variables should be transformed into logarithms, as happened in the previous analysis).

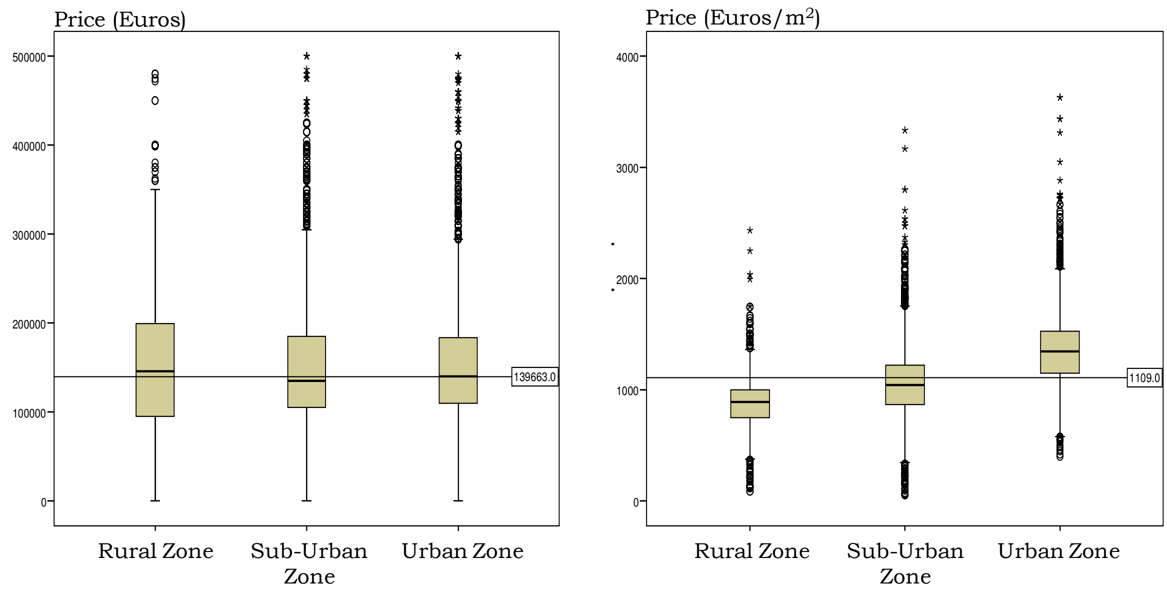


Figure 54 - Variability of prices in different locations

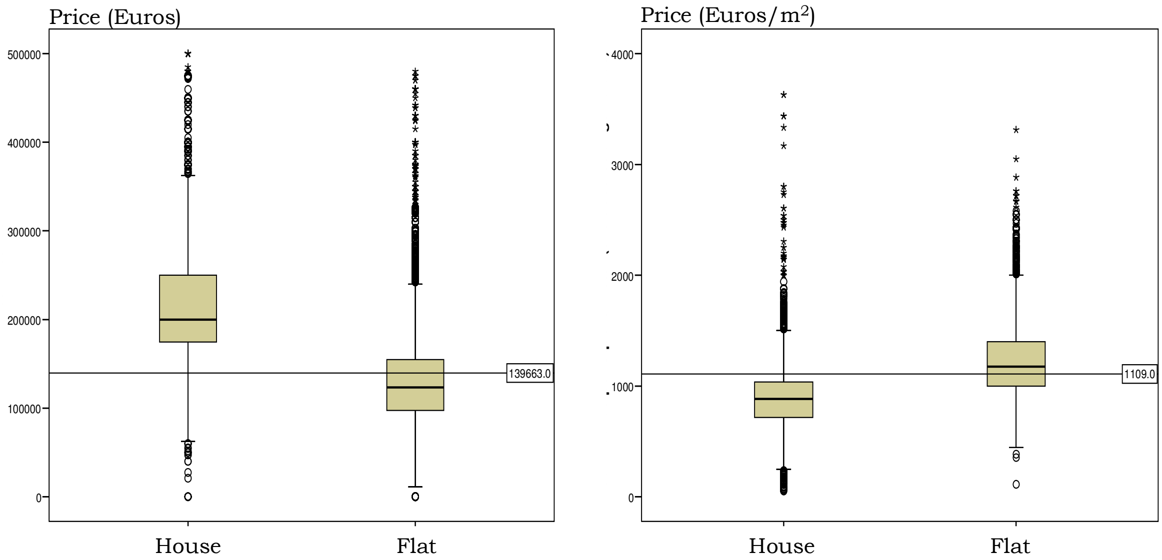


Figure 55 - Variability of prices by dwelling type

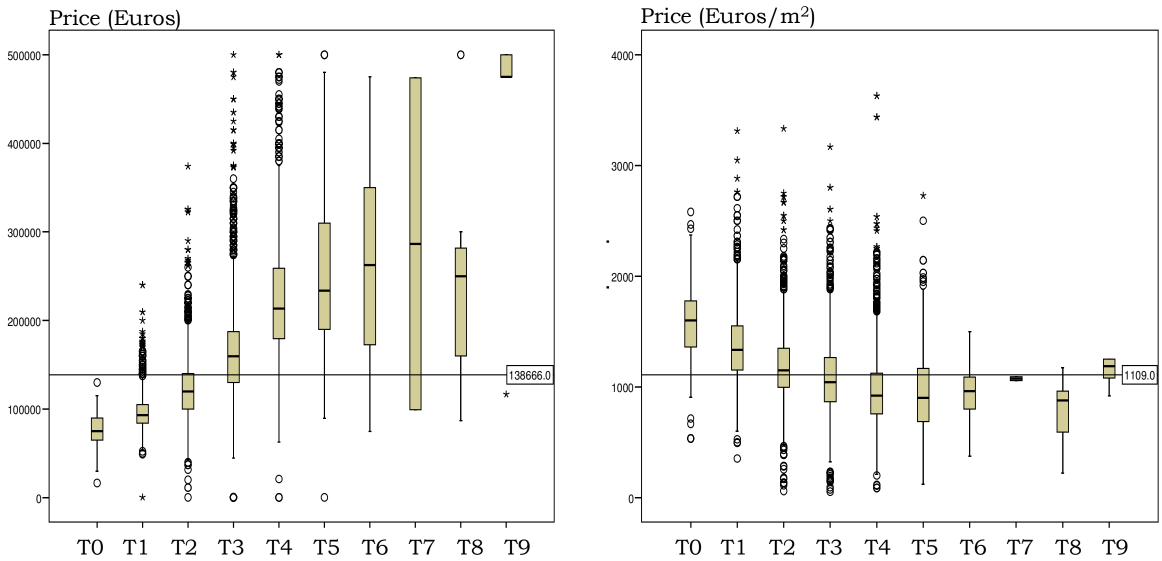


Figure 56 - Variability of prices by number of bedrooms

The following statistical analyses explore dependences between the variables. As mentioned in Section VII.4 the use of parametric or non-parametric statistical tests is conditioned by the nature of the variables. Being scalar (many of the statistical methods require the assumption that variables are normally distributed), ordinal or nominal (two or more than two groups) the appropriate tests should be applied.

Starting with the study of normality for scalar variables, and as shown in the *Table 34*, none of the variables are normally distributed. So, non-parametric tests are more appropriate for analysing the dependence between groups. The results are presented in more detail in Appendix 2.

Table 34 - Tests of Normality

	Kolmogorov-Smirnov[a]		
	Statistic	df	Sig.
Price_€	0.096	12467	0.000
Price_€sqrtm	0.077	12467	0.000
Area_sqrt	0.152	12467	0.000
Number of rooms	0.209	12467	0.000

a. Lilliefors Significance Correction

a) *Analysing differences between the different territories (urban, suburban and rural areas)*

Instead of considering parishes as the territorial units (as has been done in the smaller dataset), to analyse the territorial dependence, three aggregated areas have been considered: urban (encompassing the two parishes, Vera Cruz and Glória), suburban (encompassing the six parishes, Aradas, Esgueira, Santa Joana, São Bernardo, São Salvador and Gafanha da Nazaré), and rural (encompassing ten parishes, Cacia, Eirol, Eixo, Gafanha da Encarnação, Gafanha do Carmo, Nariz, Nossa Senhora de Fátima, Oliveirinha, Requeixo and São Jacinto).

In order to test whether differences exist between groups (territorial units) two different non-parametric statistical tests have been applied.

The *Kruskal-Wallis test* is used to understand whether groups are significantly different for the following characteristics: number of bedrooms, housing value (absolute and relative terms) and total area. All the results (*Table A2.2*) are statistically significant meaning that there are differences between the territories, in line with the information presented in *Figure 54*. The number of bedrooms, the prices (in euros) and the area (m^2) of housings are significantly lower in urban areas (Vera Cruz and Glória) and higher in rural parts of the study area, however the price per square meter shows an inverse relationship, that is, price is higher in urban areas (as reported in *Table A2.2*).

To compare territorial differences concerning the other nominal (categorical) data the *chi-square statistic* has been computed. This technique instead of using medians, means and variances, uses frequencies. All variables considered in this analysis are statistically significant, except that is for flats being duplex (see *Tables A2.4, A2.6, A2.8, A2.10, A2.12, A2.14, A2.16, A2.18* and *A2.20*). Thus, type of house, preservation, balcony, terrace, garage, central heating and fireplace are dependent on the spatial location of dwellings. For instance, in suburban areas there are no differences between observed and expected numbers of types of properties (both flats and houses), but in rural and urban areas big differences prevail; the urban part of the city is characterised by flats, and in the rural area houses are more frequent (as reported in *Tables A2.3, A2.5, A2.7, A2.9, A2.11, A2.13, A2.15, A2.17* and *A2.19*).

b) Analysing differences between housing type (house or flat)

The housing type attribute being a dummy variable with two categories, both non-parametric *Mann-Whitney tests* (for scalar variables) and *Chi-square tests* (for nominal variables) are computed. Concerning the first test (*Mann-Whitney*), as the *p-value* is lower than 5% (*Table A2.22*) for all variables, it can be concluded that there is enough statistical evidence to suggest that significant differences exist in the number of bedrooms, price and total area for houses and flats. The findings are mostly identical to what has been said before, that is, number of bedrooms, prices (in euros) and areas (m^2) of dwellings are significantly lower for flats (predominantly located in urban areas) and higher in single unit houses (located in rural areas). However, the price per square meter in urban areas is higher in flats, not because flats are more expensive than houses, but because they have a bigger bid rent (as reported in *Tables A2.21* and *A2.22*).

The second set of tests show that all values of the *Chi-square* are highly significant (*Tables A2.24, A2.26, A2.28, A2.30, A2.32, A2.34, A2.36* and *A2.38*), meaning that, single houses are characterised as new, having balconies, terraces garages, central heating and fireplaces; while garage space, and, obviously, duplex are more common in flats.

c) Analysing differences between level of preservation (new or used)

There is a significant dependence between the level of preservation and the scalar intrinsic housing attributes considered in the analysis. It not surprising that new dwellings are more expensive than the used ones, but a more interesting insight is the conclusion that new houses have are larger, both in terms of number of rooms and total area, as shown by *Tables A2.38* and *A2.39*.

The *Chi-square tests* (from *Table A2.41* to *Table A2.52*) show that the variables garage and duplex are not statistically significant, meaning that the level of preservation (new or used) is independent of the existence of these two attributes. However, balconies and terraces are characteristics of used dwellings, while garage space, central heating and fireplace characterise new properties. These variables are statistically significant at a level of 1%.

VII.5.3. Hedonic analysis for the housing market of Aveiro and Ílhavo

Hedonic pricing models using initial variables are presented in this section with the aim of characterising the housing market in Aveiro and Ílhavo. The models (both using initial variables and factors) consider a database with 12467 observations (from the total of 47188), which is the number obtained after data cleaning.

Thus, the general hedonic model includes a set of attributes which are organised according to three major categories:

$$P = \mathrm{F} \times \mathrm{L} \times \mathrm{T}$$

Eq. 46

Where:

P is the **price**

In order to get a more scale neutral dependent variable the normalised measure price of per square meter has been used. This measure facilitates comparison with other studies.

F are housing **intrinsic characteristics**[137]:

$$F = \mathrm{N}^{\alpha_1}.\mathrm{A}^{\alpha_2}.\mathrm{d}_{\mathrm{T}}^{\alpha_3}.\mathrm{d}_{\mathrm{D}}^{\alpha_4}.\sum_{i=1}^{7}\mathrm{d}_{\mathrm{C}_i}^{\alpha_{C_i}}.\sum_{k=1}^{6}\mathrm{d}_{\mathrm{F}_k}^{\alpha_{F_k}}$$

Eq. 47

N - Number of bedrooms

A - Area (square meters)

[137] A dummy variable that equals one if the property has the mentioned attribute and equals zero otherwise.

d_T - Housing type dummy variable (1=house; 0=flat)

d_D - Dummy variable for duplex flats (yes =1; no=1)

d_{C_i} - Level of preservation dummy variable (C_i=1,..., 7)

d_{C1}=New; d_{C2}=Under construction; d_{C3}=Restored; d_{C4}=Used10 (<10 years); d_{C5}=Used1025 (10-25 years); d_{C6}= Used25 (> 25 years); and d_{C7}=Not restored.

d_{F_k} - Other physical attributes (F_k=1,...,6)

d_{F1}=Balcony; d_{F2}=Terrace; d_{F3}=Garage space; d_{F3}=Garage; d_{F5}=Central heating; and d_{F6}=Fireplace

L are housing **Location attributes**:

$$L = \sum_{Z=1}^{76} d_{Z_j}^{\alpha_{Zj}} . \sum_{u=1}^{23} D_{d_u}^{\alpha_{du}} \qquad \text{Eq. 48}$$

d_{Z_j} - Dummy variables identifying zones (Z_j=1,...,76)

D_{d_u} - Distances to several urban amenities (d_u=1 and 23)

T is the **Time**:

$$T = \sum_{m=1}^{114} d_{T_m}^{\alpha_{Tm}} . T_{TOM}^{\alpha_5} \qquad \text{Eq. 49}$$

$d_{T_m}^{\alpha_m}$ - Dummy variable for time (monthly dummies) (T_m=1, ..., 114)

$T_{TOM}^{\alpha_5}$ - is the time on the market (TOM) measure in days

Alfa (α) are parameters which measure the relevant implicit marginal prices of the attributes.

Following the same strategy used in the model built for the small dataset (Section VII.4), some of the independent variables were transformed into their logarithm form, to account for decreasing scale effects (e.g. type, size and price).

Combining the variables described above, two multiplicative models are developed: i) a descriptive model and ii) an explicative model. Both describe elasticity effects of each attribute in the value of residential price, the difference remains in the type of locational variables used. The descriptive model considers dummy variables to identify the location of each dwelling (d_{Zj}), while the explicative model considers as location independent variables the effect of several urban amenities measured in distance or in potential (as described in *Table 19*). A third model is presented in the *Table 47*, where the estimated hedonic model is based on factors reported in *Table 40*.

i) Descriptive Model

This model is defined by physical characteristics of properties (F), dummy time variables (T) and location attributes measure with dummies in each zone, as shown in the following equation:

$$P_{€/m^2} = \underbrace{N^{\alpha_1}.A^{\alpha_2}.d_T^{\alpha_3}.d_D^{\alpha_4}.\sum_{i=1}^{7} d_{C_i}^{\alpha_{C_i}}.\sum_{k=1}^{6} d_{F_k}^{\alpha_{F_k}}}_{F}.\underbrace{\sum_{Z=1}^{76} d_{Z_j}^{\alpha_{Zj}}}_{L}.\underbrace{\sum_{m=1}^{114} d_{T_m}^{\alpha_{Tm}}.T_{TOM}^{\alpha_5}}_{T} \qquad \text{Eq. 50}$$

or

$$\ln P_{€/m^2} = \alpha_1 \ln N + \alpha_2 \ln A + \alpha_3 \ln d_T + \alpha_4 \ln d_D + \sum_{i=1}^{7} \alpha_{C_i} \ln d_{C_i} + \sum_{k=1}^{6} \alpha_{F_k} \ln d_{F_k} + \sum_{Z=1}^{76} \alpha_{Z_j} \ln d_{Z_j} + \sum_{m=1}^{114} \alpha_{T_m} \ln d_{T_m} + \ln T_{TOM}^{\alpha_5} \qquad \text{Eq. 51}$$

The results from the descriptive model are summarised in *Table 35*, regression coefficients (unstandardized and standardized) of the intrinsic characteristics of buildings are shown in *Table 36*, *Table 37*, and *Figure 57*.

Table 35 - Summary results of the regression model

R	**R^2**	**$R^2_{adjusted}$**	**Standard error**
0.834	0.696	0.691	0.175

As can be seen, this model is able to explain 69.1% (R^2_{adj}) of the variance in housing prices, that is, this explanatory power indicates that the value of dwellings, measured by euros per square meter, is reasonable explained by the independent variables.

The model includes 211 variables disaggregated into three large groups of indicators. The regression coefficients for each group are presented separately: i) housing physical attributes; ii) location attributes; and iii) time attributes.

a) Housing physical aspects (F)

Regression coefficients (unstandardized and standardized) of the intrinsic characteristics of buildings are shown in *Table 36*.

Table 36 - Regression coefficients of housing characteristics

	Coefficients		*t*	Elasticities (%)	Collinearity Statistics	
	Unstard.	Stard.			Tolerance	VIF
(Constant)	9.978***		59.42			
dType_house	0.146***	0.209	22.93	15.7%	0.30	3.35
LnTypology	0.312***	0.327	36.18	36.6%	0.30	3.29
Duplex	0.031***	0.032	5.69	3.1%	0.78	1.28
PreserNew	0.372***	0.547	6.28	45.0%	0.00	305.96
PreservConst	0.411***	0.563	6.94	50.8%	0.00	266.44
PreservRestorcov	0.293***	0.055	4.51	34.0%	0.17	6.00
PreservUsed10	0.261***	0.391	4.41	29.9%	0.00	317.22
PreservUsed1025	0.173***	0.151	2.92	18.9%	0.01	107.34
PreservUsed25	0.065	0.023	1.07	6.7%	0.05	18.30
Ln Total Area	-0.580***	-0.891	-83.81	-44.0%	0.22	4.57
Balcony	-0.004	-0.007	-1.20	0.2%	0.75	1.33
Terrace	0.026***	0.031	5.90	-32.7%	0.87	1.15
GarageSpace	-0.022***	-0.025	-4.42	-0.4%	0.76	1.32
Garage	0.033***	0.051	7.85	2.6%	0.59	1.69
CentralHeating	0.030***	0.048	8.16	-2.1%	0.72	1.39
Fireplace	-0.016***	-0.023	-3.82	3.4%	0.68	1.48

**** significant at the 1% level/ ** significant at the 5% level/ * significant at the 10% level*

Most of the variables are significant with the expected sign in their coefficients. Only two variables are not significant, the level of preservation for properties used more than 25 years (*PreservUsed25*) and the existence of a balcony.

For a more accurate analysis an indicator which reflects the relative value of each attribute in the price of a property type (the elasticity) is defined. The elasticities of each attribute are calculated using the exponential function of the regression coefficients of the model and are presented in the fifth column in the *Table 36*. The parameters indicate the increase in price (in percentage) when each of the characteristics exists (for scalar variables the units in which the attribute is measure should be considered). Thus, being new or under construction the property value is 45 and 51% higher respectively, compared with an un-restored property (the dummy used as reference in the preservation variables). The existence of one more bedroom increases the housing price (per meter square) by 36%, and the difference between houses and flats is on average (everything else being equal) 16%. The same interpretation can be made for the other intrinsic attributes.

Since some of the attributes associated with intrinsic characteristics (d_{Fk}) do not accurately reflect the presence or absence of the specific attribute, because they are listed for free by real estate agents, three negative coefficients do not give a clear picture. This occurs namely with the variables for a balcony, garage space and fireplace. In any case an explanation can be put forward: a house that has a fireplace and a garage space has a relative disadvantage in its price because it most likely will not have central heating or a full garage, which are more valuable aspects. A more detailed analysis of these indicators presents an opportunity for future work.

b) Location Aspect (L)

Regression coefficients for location attributes, considering dummy variables for each of 76 zones (descriptive model) is shown in *Table 37*. As has been done for intrinsic attributes, a relative assessment indicator for the model has been built for this group of attributes.

Table 37 - Regression coefficients of the location characteristics

Ranks	Zones	Coef.	Added Value
1	Fórum		62.4%
2	Barra	-0.191	45.0%
3	Costa Nova	-0.196	44.6%
4	Gulbenkian	-0.223	42.4%
5	Alboi	-0.261	39.4%
6	Agras do Norte	-0.262	39.3%
7	Rossio	-0.325	34.7%
8	Mario Sacramento	-0.364	31.9%
9	Glicinias	-0.386	30.3%
10	Barrocas	-0.393	29.9%
11	Agras	-0.396	29.7%
12	Cabo Luis Quinta das Acacias	-0.397	29.6%
13	Forca	-0.406	29.0%
14	Feira de Marco	-0.423	27.9%
15	Estacao	-0.431	27.4%
16	Viaduto	-0.434	27.2%
17	Bairro do Liceu	-0.440	26.8%
18	Beira Mar	-0.451	26.1%
19	Vila Jovem Santiago	-0.477	24.5%
20	Avenida Dr Lourenco Peixinho	-0.480	24.2%
21	Eucalipto	-0.486	23.9%
22	Vilar	-0.489	23.7%
23	Centro de Congressos	-0.511	22.4%
24	Patela	-0.511	22.4%
25	Quinta do Cruzeiro	-0.518	22.0%
26	Verdemilho	-0.519	21.9%
27	Sao Bernardo	-0.544	20.4%
28	Azenha de Baixo	-0.547	20.2%
29	Sao Jacinto	-0.550	20.1%
30	Coutada Medela	-0.570	19.0%
31	Cancela	-0.573	18.7%
32	Aradas	-0.576	18.6%
33	Sol Posto Presa	-0.576	18.6%
34	Escolas	-0.578	18.5%
35	Esgueira	-0.602	17.1%
36	Oita	-0.620	16.2%
37	Alagoas	-0.624	16.0%
38	Quinta do Gato	-0.632	15.5%
39	Mataducos	-0.638	15.2%

(continue)→

Ranks	Zones	Coef.	Added Value
40	Cilhas	-0.641	15.0%
41	Gafanha d'Aquem	-0.643	14.9%
42	Quinta do Picado	-0.645	14.8%
43	Ribas	-0.648	14.7%
44	Santiago	-0.659	14.1%
45	Nossa Senhora de Fatima	-0.669	13.6%
46	Carramona	-0.669	13.6%
47	Viso/Caiao	-0.677	13.2%
48	Centro Ilhavo	-0.689	12.6%
49	Sarrazola	-0.691	12.5%
50	Paco	-0.691	12.5%
51	Gafanha da Nazare	-0.694	12.4%
52	Oliveirinha	-0.696	12.2%
53	Povoado Valado	-0.698	12.1%
54	Bonsucesso	-0.709	11.6%
55	Gafanha da Encarnacao	-0.723	10.9%
56	Vista Alegre	-0.727	10.7%
57	Costa do Valado	-0.729	10.6%
58	Taboeira	-0.735	10.4%
59	Caleda Vila	-0.742	10.0%
60	Azurva	-0.744	9.9%
61	Quinta da Bela Vista	-0.747	9.8%
62	Cacia	-0.748	9.7%
63	Moitinhos	-0.748	9.7%
64	Eixo	-0.749	9.7%
65	Quintado Loureiro	-0.758	9.3%
66	Bairrode Santiago	-0.762	9.1%
67	Gafanha do Carmo	-0.771	8.6%
68	Mamodeiro	-0.777	8.4%
69	Cidadela Quinta de Santo Antonio	-0.779	8.3%
70	Povoa do Paco	-0.783	8.1%
71	Quintãs	-0.801	7.3%
72	Requeixo	-0.833	5.8%
73	Granja de Baixo	-0.845	5.3%
74	Eirol	-0.904	2.9%
75	Olho d’Água	-0.935	1.6%
76	Nariz	-0.978	0.0%

**** significant at the 1% level/ ** significant at the 5% level/ * significant at the 10% level*

Thus, considering the zone Nariz as the reference point (where housing prices per square meter are lowest) the values of the remaining zones, when compared to Nariz, are reported in *Table 37* and represented on the map in *Figure 57*. As seen in both representations, considering constant all physical attributes, the price of housing in the area of Forum is 63.3% more expensive than in the area of Nariz. In general, urban areas are more expensive (as expected) as can be seen by the degrees of shading on the map. However, some areas stand out as exceptions, notably that of Oita. These disparities can be justified by the existence of properties with a high level of degradation in these areas. Beaches (both Costa Nova and Barra) also emerge as highly valued in terms of property prices.

The drawbacks of such descriptive models is that they do not provide the value of each location and neighbourhood attributes, only aggregate values explained by a dummy variable. However, useful information is provided about the value of land (linked with the location coefficient used in IMI).

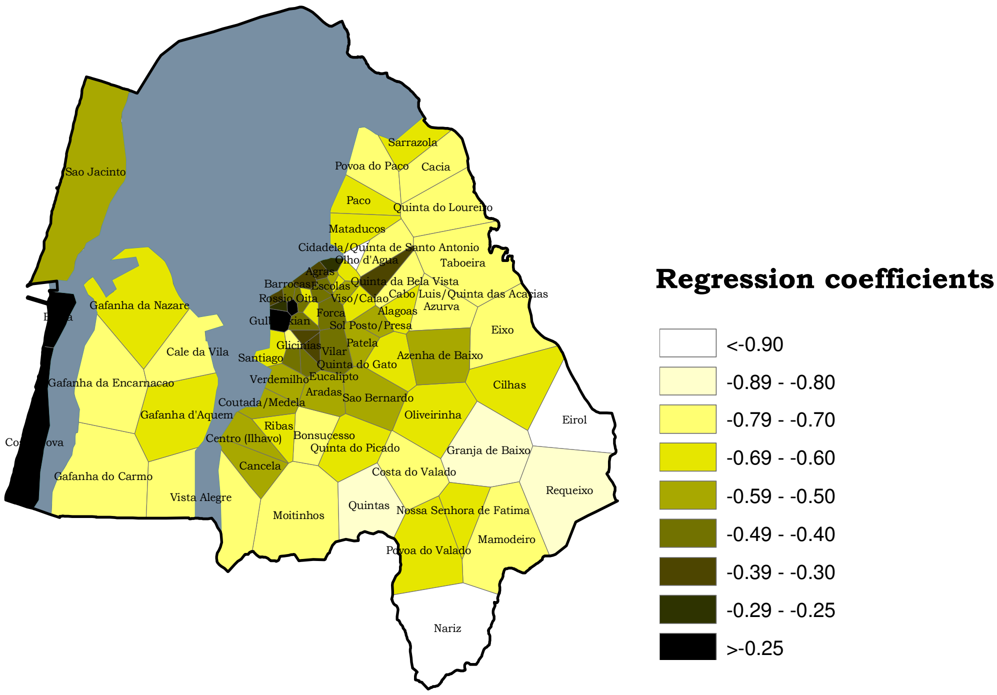


Figure 57 - Location coefficients (zoning)

c) Time aspects (T)

The graph presented in *Figure 58* represents the evolution of the regression coefficients for dummy time variables. In this case 114 dummy variables are considered representing each month, between October 2000 and February 2010.

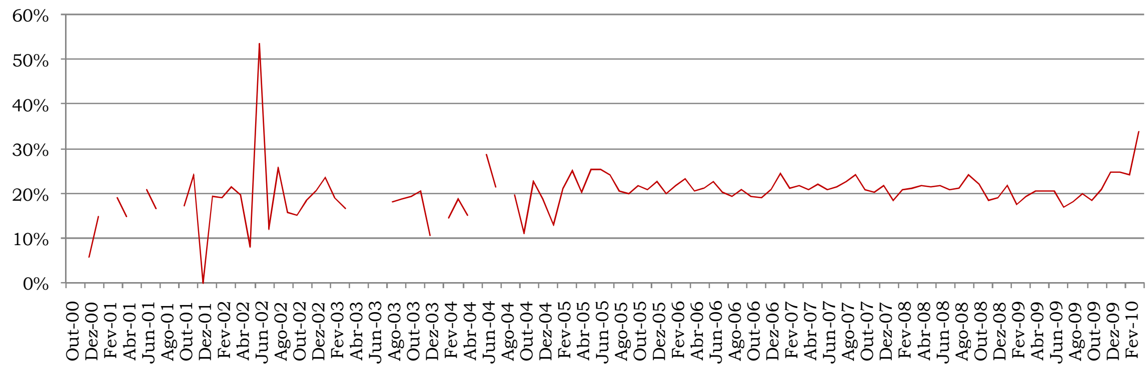


Figure 58 - Time coefficients (monthly dummy variables)

From the graph of *Figure 58* it emerges that considering the month of December 2001 as a reference (the lowest regression coefficient value), housing in June 2002 cost 50% more. The vertical axis of the graph represents the increase in the property value, in percentage, in each month when compared with the reference month. An analysis of the trend of these values shows an increase of house prices of approximately 0.06% per month, which corresponds to an annual growth of 0.79%.

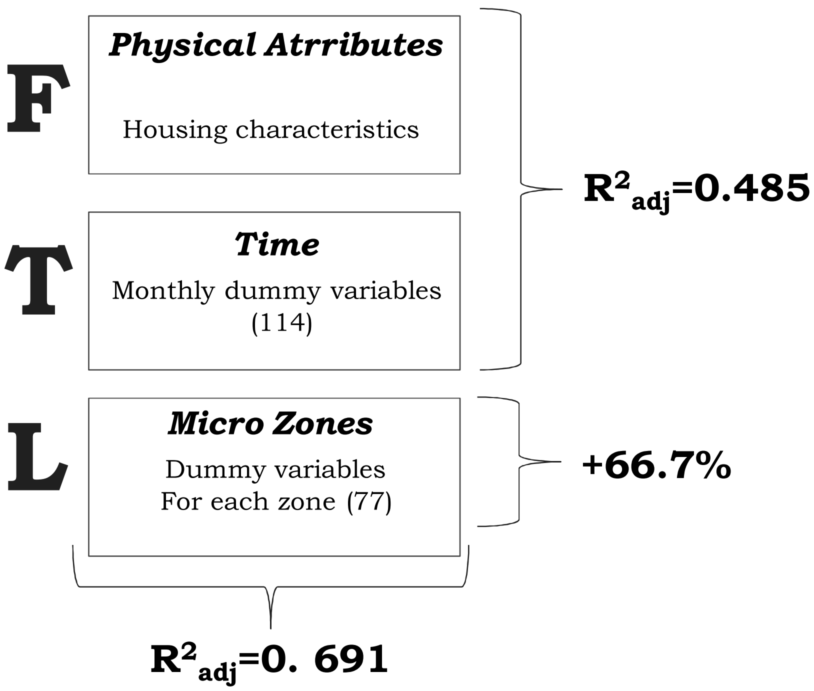


Figure 59 - Explanatory capacity of each component for the descriptive model

In summary, for the descriptive model presented previously, the physical aspects of housing and the time factor is responsible for explaining the variability of the price in 48.5% of a total of 69.1%. What is also relevant is that the spatial component (location attributes) explains about 66.7% of the remaining value, calculated by the following equation:

$$R_{i+1}^2 = \frac{R_{i+1}^2 - R_i^2}{1 - R_i^2} \qquad \text{Eq. 52}$$

ii) Explicative model for houses for sale

Physical characteristics of properties (F) and dummy time (T) variables do not change in this model. Only location variables, dummy variables for each zone, are replaced by accessibility to several urban amenities, as shown in *Table 39*. Thus, this model is developed using the following equation:

$$P_{€/m^2} = \underbrace{\mathrm{N}^{\alpha_1}.\mathrm{A}^{\alpha_2}.\mathrm{d}_{\mathrm{T}}^{\alpha_3}.\mathrm{d}_{\mathrm{D}}^{\alpha_4}.\sum_{\mathrm{i}=1}^{7}\mathrm{d}_{\mathrm{C_i}}^{\alpha_{C_i}}.\sum_{\mathrm{k}=1}^{6}\mathrm{d}_{\mathrm{F_k}}^{\alpha_{F_k}}}_{F}.\underbrace{\sum_{\mathrm{u}=1}^{23}\mathrm{D}_{\mathrm{d_u}}^{\alpha_{du}}}_{L}.\underbrace{\sum \mathrm{d}_{\mathrm{T_m}}^{\alpha_{\mathrm{Tm}}}.T_{\mathrm{TOM}}^{\alpha_5}}_{T} \qquad \text{Eq. 53}$$

or

$$\ln P_{€/m^2} = \alpha_1 \ln \mathrm{N} + \alpha_2 \ln \mathrm{A} + \alpha_3 \ln \mathrm{d}_{\mathrm{T}} + \alpha_4 \ln \mathrm{d}_{\mathrm{D}} + \sum_{\mathrm{i}=1}^{7} \alpha_{C_{\mathrm{i}}} \ln \mathrm{d}_{\mathrm{C_i}} + $$
$$+ \sum_{\mathrm{k}=1}^{6} \alpha_{F_{\mathrm{k}}} \ln \mathrm{d}_{\mathrm{F_k}} + \sum_{\mathrm{u}=1}^{23} \alpha_{\mathrm{d_u}} \ln \mathrm{D}_{\mathrm{d_u}} + \sum_{m=1}^{114} \alpha_{\mathrm{T_m}} \ln \mathrm{d}_{\mathrm{T_m}} + \ln T_{\mathrm{TOM}}^{\alpha_5} \qquad \text{Eq. 54}$$

Table 38 - Summary results of the regression model

R	**R^2**	**$R^2_{ajusted}$**	**Standard error**
0.822[a]	0.676	0.673	0.1806

In *Table 39* the coefficients for location attributes of this explanatory model are presented. The coefficients for intrinsic (F) and time attributes (T) are not shown as there is no significant change when compared to the previous model.

Table 39 - Regression coefficients for the location characteristics

		Coefficients		*t*	Elasticities
		Unstand.	Stand.		(%)
	(Constant)	8.947***		33.38	
D	lnCentral Amenities	-0.053***	-0.140	-4.44	6.3
D	lnLocal Amenities	-0.010**	-0.019	-2.20	2.0
D	lnCBD Aveiro	-0.034**	-0.087	-2.11	4.4
D	ln Local Commerce	0.033***	0.119	5.34	-2.3
D	lnPrimary Schools	-0.008*	-0.020	-1.74	1.8
D	lnIntermediate Schools	0.017***	0.056	3.89	-0.7
D	lnUniversity	-0.036**	-0.071	-2.30	4.6
D	lnHospital	-0.035***	-0.097	-4.04	4.5
D	lnHealth Centres	-0.001	-0.003	-0.19	1.1
D	lnPharmacies	0.001	0.004	0.25	0.9
D	lnParks and Gardens	0.012**	0.037	2.27	-0.2
D	lnRail Station	0.004	0.014	0.65	0.6
D	lnAccess node	0.054***	0.092	9.13	-4.4
D	lnPetrol Stations	0.014***	0.044	3.44	-0.4
D	lnPolice	-0.015*	-0.038	-1.80	2.5
p	lnAdministration	-0.006	-0.020	-1.13	1.6
p	lnCulture	0.019***	0.041	3.67	-0.9
p	LnSpecialised Commerce	-0.035***	-0.079	-2.85	4.5
p	lnRestaurants	0.108***	0.219	6.49	-9.8
p	lnHotels and hostels	0.011	0.024	0.77	-0.1
p	lnMonuments	0.021	0.029	1.49	-1.1
p	lnBanks, ATMs, Post	0.045***	0.098	4.67	-3.5
p	lnSports	-0.065***	-0.091	-5.44	7.5
d	Sea Beaches	0.563***	0.445	19.63	-55.3

*D=distances in metres; p=gravitational potential; d=dummy variable;*** significant at the 1% level/ ** significant at the 5% level/ * significant at the 10% level*

A general analysis of this model in *Table 39* shows that the replacement of dummy variables for each zone by location variables representing various centralities does not decrease the level of explanatory capacity of the model (69.1 to 67.3%). It would be expected that the inclusion of dummy variables for each zone, instead of distances or potential related to several urban amenities, would lead to a greater explanatory power; however, it is not possible through this approach to explain the determinants of space in which the residential property is valued. For example, properties located at the beaches are worth 55% more (note that this attribute is a dummy variable) and the square meter price of properties decreases on average 6.3% when the distance to the city centre of Aveiro (centrality level 1) increases 1%.

Figure 60 shows the results for each dimension considered in the model. In this case the location component is also crucial for the explanation of the residential value and the proportions do not differ too much when compared with the descriptive model.

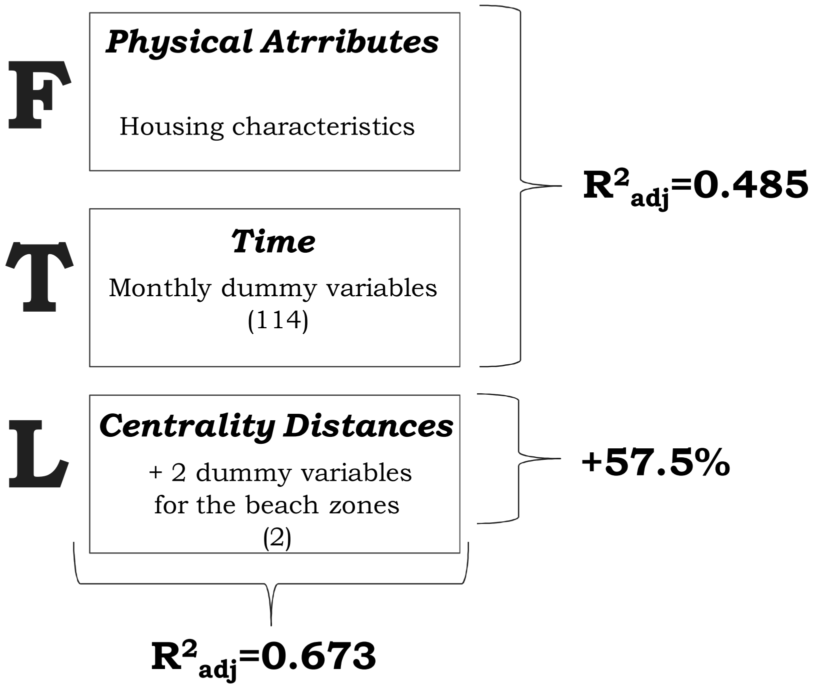


Figure 60 - Explanatory capacity of each component for the explicative model

VII.5.4. Factor based hedonic pricing model

Following an identical methodology, factor analysis with orthogonal varimax rotation is applied in this dataset and, as seen in *Table 40,* the resulting five leading factors (suggested by the scree plot presented in *Figure 61*) bring into focus the main housing characteristics related to behavioural patterns.

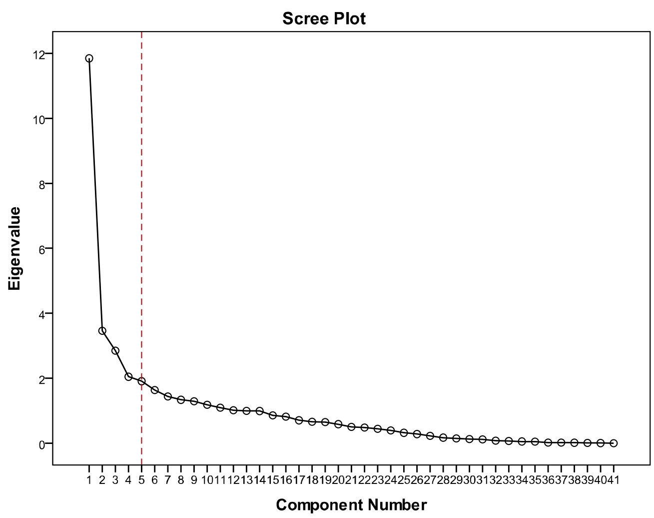


Figure 61 - Eigenvalue plot for scree test

Table 40 - Factor loadings for housing market of Aveiro and Ílhavo

	Attributes	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Location attributes	Specialised Commerce	-0.924				
	Centrality, Central Amenities	0.913				
	CBD Aveiro	0.907				
	Monuments	-0.889				
	Hospital	0.853				
	University	0.851	0.368			
	Hotels and Hostels	-0.844		0.443		
	Sports	-0.819	-0.376			
	Police	0.818				
	Culture	-0.752				
	Restaurants	-0.702		0.548		
	Rail Station	0.646		0.521		
	Access Node	0.460				
	Health Centres		0.878			
	Parks and Gardens		0.858			
	Banks, ATMs, Post	-0.421	-0.759			
	Administration	-0.563	-0.601			
	Petrol Stations	0.432	0.520			
	Intermediate Schools	0.494	0.518			
	Pharmacies	0.363	0.399			
	Sea/Beaches			0.849		
	Local Commerce		0.390	-0.785		
	Primary Schools	0.373		0.690		
	Centrality, Local Amenities					

(cont)→

Physical attributes	Pres:Used building, 10-25 years					
	Total area				0.815	
	Type (House=1; Flat=0)	0.353			0.759	
	Number of rooms				0.753	
	Pres:Used building, less than 10 years				-0.446	
	Pres:Under construction					
	Pres:New building					
	Garage					0.779
	Balcony					0.614
	Central Heating					0.575
	Fireplace					0.458
	Garage space					0.427
	Terrace					
	Duplex					
	Pres: Used building, more than 25 years					
	Pres: Restored					
	Pres: Not restored					
	Total Variance Explained	**25.02%**	**10.10%**	**8.03%**	**5.88%**	**4.91%**

Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization (absolute value > 0.35).

The explanatory variables included in the factorial analysis are organized into 5 factors. In general, factor 1, 2 and 3 are related to attributes of location; whereas factors 4 and 5 represent the intrinsic characteristics of dwellings. The characteristics of these factors and their contribution to overall variation are discussed below.

i) **Factor 1**: *Access to the centre or central amenities*

The first factor is responsible for 25.02% of explained variance and includes variables related to distances to tertiary facilities and services. This factor corresponds to location attributes essentially associated with a high level of centrality. However, as shown in *Figure 62* some of these urban amenities are distributed throughout the study area, for example: monuments (chapels and churches), banks (cash machines) and some sports facilities, while being more prominent in the urban centre.

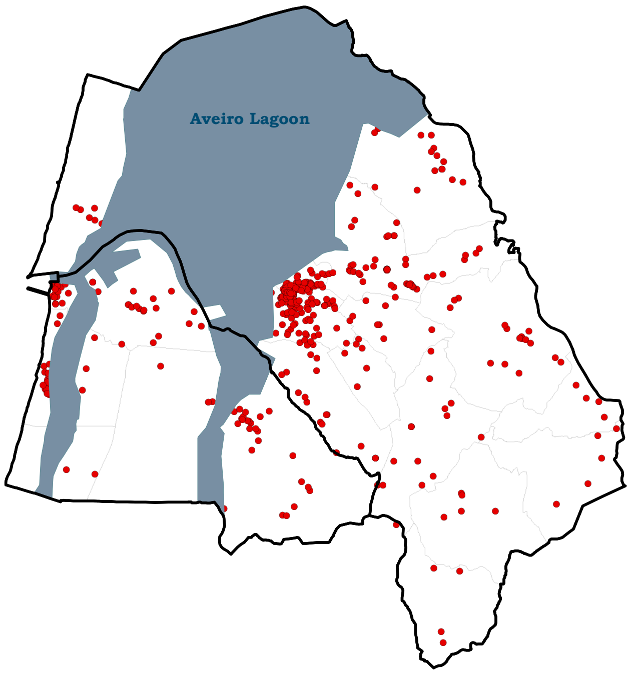


Figure 62 - Distribution of urban amenities presented in the factor 1 (Access to the centre)

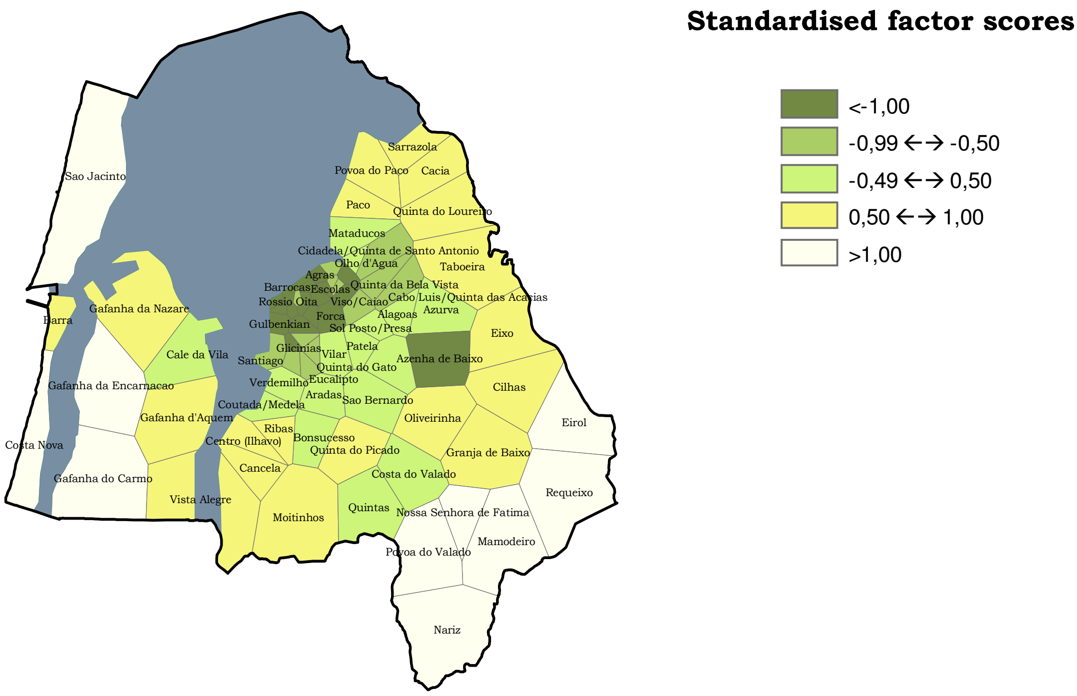


Figure 63 - Accessibility to central amenities in each zone

ii) **Factor 2**: *Health centres, parks and gardens or access to local amenities*

Factor 2 has an explanatory capacity of 10.10% of the total variance. As in factor 1, this factor is associated with variables that express distance, however, it differs from the previous one because it refers to location in a more dispersed way (see *Figure 64*). As such this includes access to local services and amenities (health centres, parks/gardens, etc.), also implying proximity to the traditional local centres within the area under study. It should be noted that the consolidation of a single urban area corresponding to the municipalities of Aveiro and Ílhavo was built on a territory previously organised as a set of small urban and rural clusters, each with its own small scale provision of services. Factor 2 reflects the proximity to such local centres.

Figure 64 - Distribution of urban amenities presented in the factor 2 (Health centres, parks and gardens)

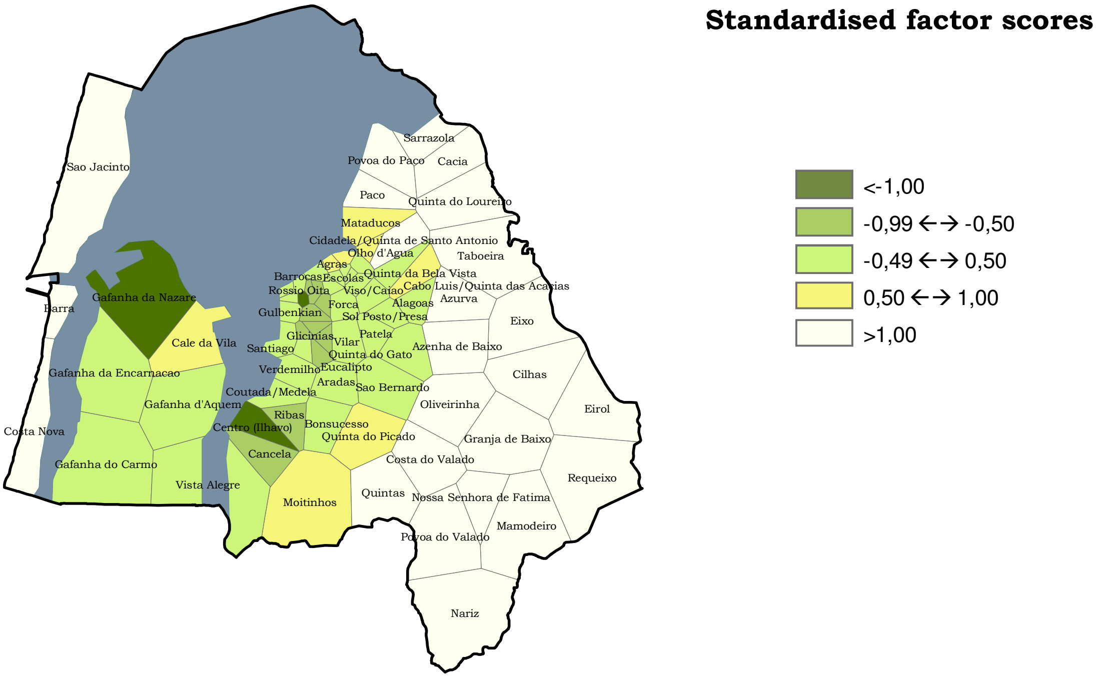


Figure 65 - Accessibility to local amenities in each zone

iii) **Factor 3**: *Access to beaches, schools and local commerce*

The third factor, with explanatory capacity of 8.03% includes three location attributes: beaches, primary schools and local commerce. The reason why these variables have opposite signs is the fact that on average the beach area has lower accessibility to the other amenities presented in this factor (local commerce and schools).

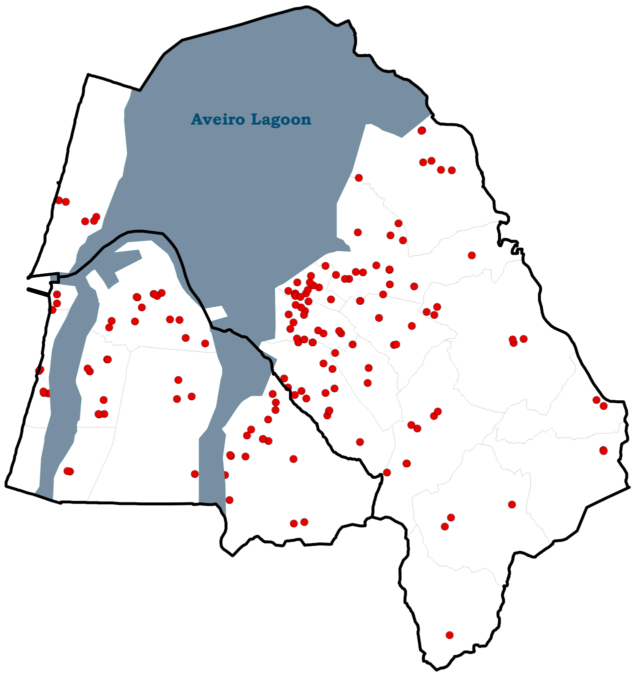


Figure 66 - Distribution of urban amenities presented in the factor 3 (access to beaches, schools and local commerce)

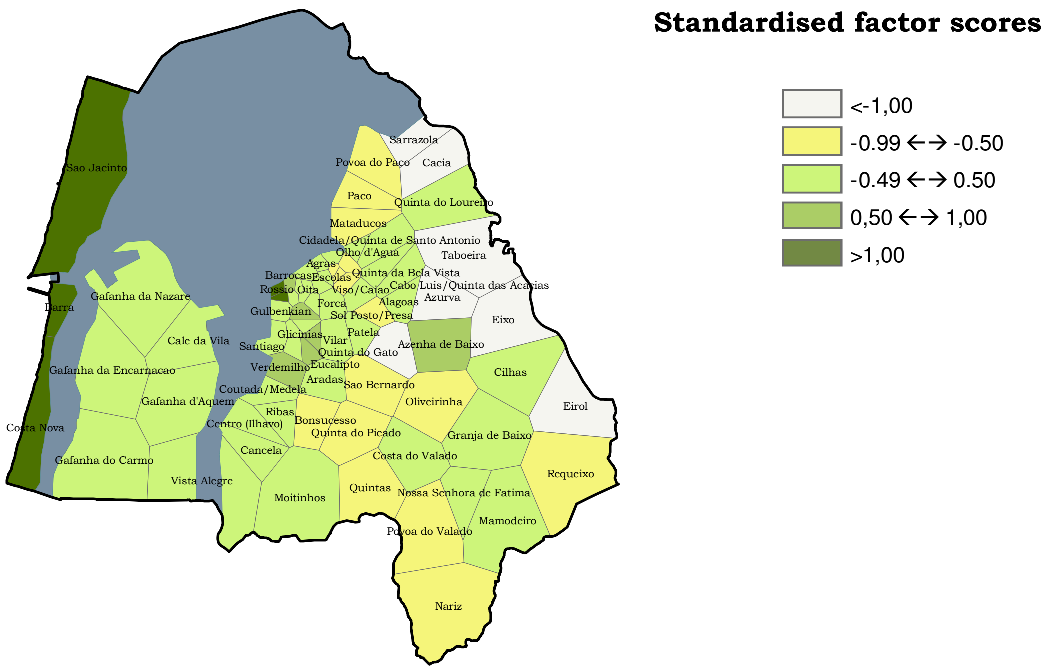


Figure 67 - Accessibility to beaches, schools and local commerce in each zone

iv) **Factor 4**: *Housing dimension*

The fourth factor, with an explanatory capacity of 5.88%, includes essentially variables related to the housing dimension (such as total area and number of rooms) and type of dwellings, single houses versus flats.

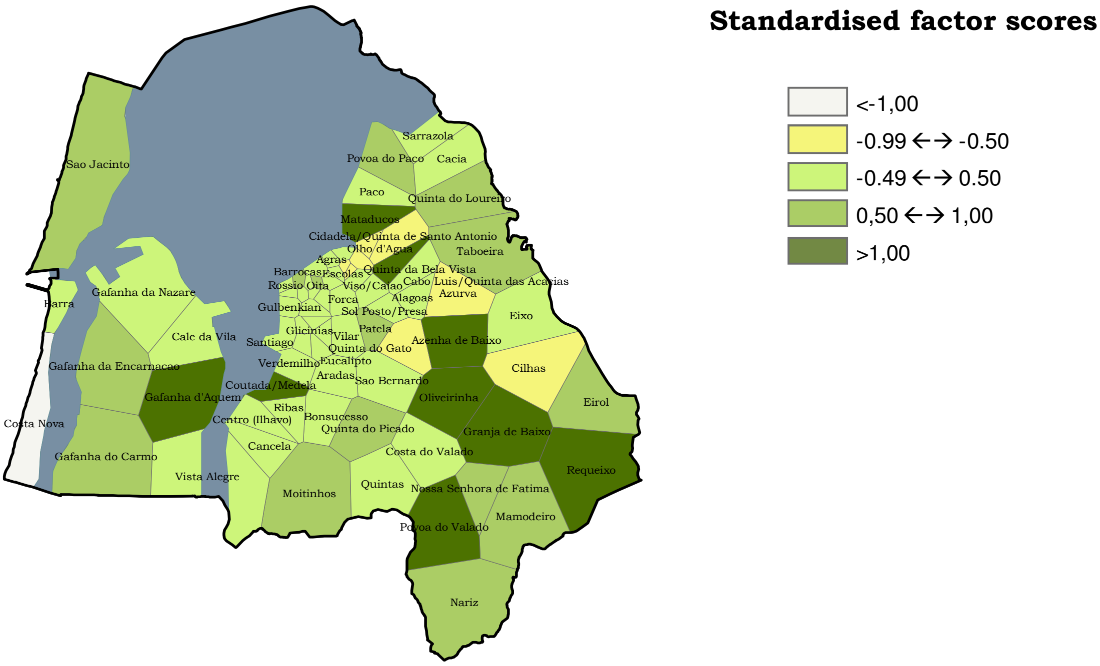


Figure 68 - Housing dimension distributes by zone

v) **Factor 5**: *Additional desirable features*

The fifth and final factor is very similar to factor 4. Both of these factors contain intrinsic attributes of housing; however these characteristics are not exclusively related with dimension. These variables are associated with size and use of space within the property along with other housing characteristics. The factor includes all housing characteristics included in the initial data: existence of garage, garage space, central heating, balcony, terrace and duplex. This factor explains 4.91% of total variance and is designated as *additional desirable features*.

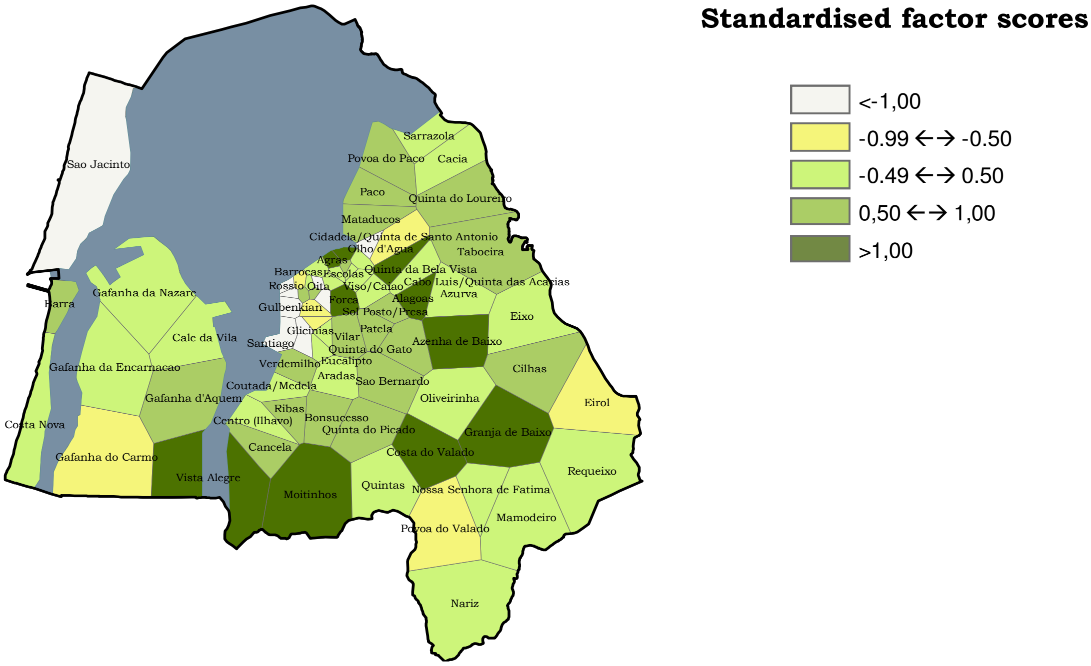


Figure 69 - Additional housing facilities distributed by zone

VII.5.5. Housing submarket in Aveiro and Ílhavo - spatial heterogeneity analysis

In this section housing submarkets for Aveiro and Ílhavo are defined and analysed. As quoted by Grigsby, Baratz *et al.*, (1987) a submarket is defined as "(...) *a set of dwellings that are reasonably close substitutes for one another, but relatively poor substitutes for dwellings in other submarkets*". From this definition the notion emerges that no contiguity of submarkets is required, thus, it is reasonable to include non-contiguous houses or zones into the same submarket.

As mentioned in the Chapter IV many approaches can be used to delimit housing submarkets, typically defined in terms of: physical housing characteristics of the dwellings, geographical aspects, socioeconomic characteristics of the neighbourhood, local real estate agents experts or pre-existing geographical boundaries.

The option of not considering the individual zones directly as submarkets is of interest due to the reduced number of observations in each unit of analysis. Two different approaches are applied in order to assess the robustness of the results: i) inductive perspective; and ii) analytical perspective. The first approach consists of defining *ex ante* criteria based on the empirical knowledge of the case study, while the second approach uses a spatial clustering analysis (allowing an identification of the patterns in the data highlighting their similarities and differences).

The following example, extracted from Goodman and Thibodeau (2007a), illustrates the two perspectives mentioned above.

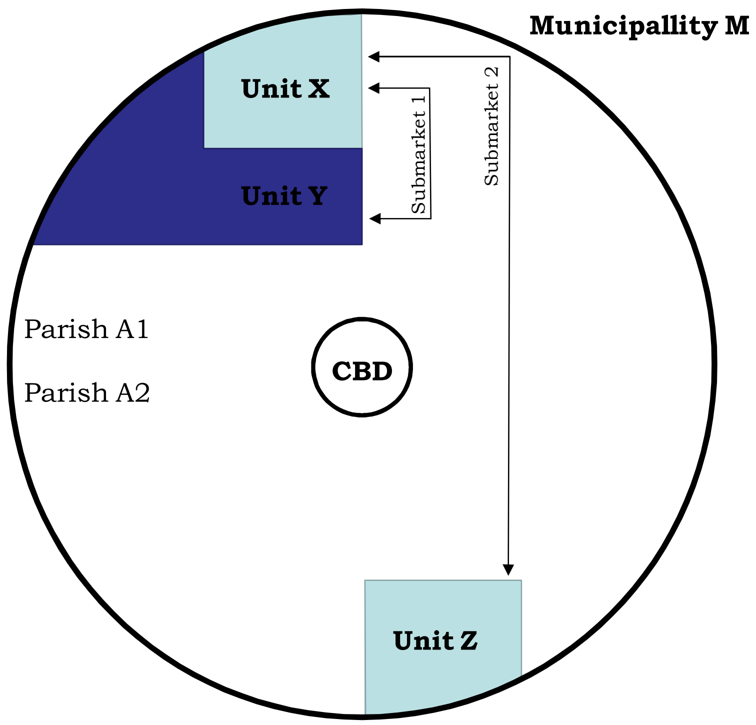


Figure 70 - Alternative characterizations of Submarkets

(Based on Goodman and Thibodeau (2007a, p.9 and 10)

The *Figure 70* represents three dwelling units, or zones, (X, Y, and Z), distributed across two parishes, and located at a specific distance from the CBD. Considering this situation, what should determine whether X is grouped with Y or with Z? Intuitively, it should be expected that because X is spatially close and located in the same parish (Parish A1) as the unit Y, both should be grouped together. However, if housing characteristics (even located within the same municipality) and households are stratified by income in Parish A1, it could very well be that X is more appropriately grouped with unit Z, located in Parish A2, even if it is located in the diametrically opposite direction. Thus, if price of housing of X is like Z, then X belongs in the same submarket as Z, even though Z is not close spatially.

i) inductive perspective

The criteria to delimit Aveiro and Ílhavo submarkets are:

- *Urban structure and demographic characteristics*: urban and rural areas are well differentiated in terms of both spatial urban structure (concentrated vs. sprawl) and socio-economic characteristics of the population (e.g. population density, population growth rate, education level, etc.). Thus, this is the second criteria to disaggregate the territory into bigger zones: urban, suburban and rural. Because the suburban area is a mixture of various realities some differentiations are considered according to the criteria described above.

- *Settlement historical growth*: different zones are considered considering different periods of urban development, inside of the core of Aveiro and in the suburbs.
- *Special features*: zones belonging to the beach area, harbour and related industries are also defined as a separate submarket.

Using these criteria a spatial disaggregation with seven housing submarkets (shown in *Figure 71*) is considered.

The seven housing submarkets for Aveiro and Ílhavo are described as follows:

In the municipality of Aveiro:

- *City of Aveiro*: includes the old core settlement of Aveiro, including the administrative and service centre, as well as high density housing. This area has a higher concentration of more affluent residents. It is considered the central business district (CBD) of the city. This submarket comprises the university campus, which is a subarea with a high concentration of students, with a specific housing dynamic, mainly concerning the prominence of a rental housing market. A social housing neighbourhood (Bairro de Santiago) is also located in this area, which is the most important of its type in the neighbourhood within the urban context of Aveiro and Ílhavo. These two particular areas are not considered separately because of degrees of freedom problems.
- *Suburban Type A*: a group of small areas not very far away from Aveiro inner-city. New planned residential areas dominate, being either blocks of flats or clusters of detached houses; these areas attract people coming from the *city of Aveiro* and looking for more affordable house prices. Traditional social groups, comprised of people owning a small agricultural property and working either in manufacturing or in low skill service jobs have been gradually substituted by the urban inhabitants referred to above. Additionally, the area of Esgueira is included in this submarket. It is an urban parish in the

municipality of Aveiro and is the second most important urban centre of this area. In the past Esgueira was the head of the old council. Nowadays, this area is located in a transition zone between an urban and rural context and is characterised by a mixture of socio-economic experiences and urban contexts. Thus, this submarket is characterised by a mix of old suburb with scattered new urban development. In general this area has good accesses to the main urban system.

- *Suburban Type B*: a combination of isolated new houses or blocks, typical of *Suburban type A*, with old rural settlements. The proportion of urban incomers, relative to traditional social groups, is lower than in *Suburban Type A*.
- *Suburban Type C*: Similar to *Suburban type B* but with a higher proportion of old rural settlements and traditional social groups this submarket corresponds essentially to the parishes located in the inner city of Aveiro. The rural areas of Aveiro and Ílhavo are characterised by low population and housing densities, and poor demographics and economic dynamics.

<u>In the municipality of Ílhavo:</u>

- *City of Ílhavo*: the administrative centre of a separate municipality and corresponds to the second centre of the urban agglomeration of Aveiro and the main urban centre of the municipality of Ílhavo.
- *Gafanhas*: a mixture of residential and industrial areas, including also the most important port of the Centro Region (Porto de Aveiro). The residential market combines houses located in older and consolidated settlements with detached houses spread in semi-urban areas. There is a marked predominance of working class and lower middle class residents. It is a city of the municipality of Ílhavo (the second urban centre of the municipality of Ílhavo and the third for the whole study area) with a high level of urbanization in the last decades.
- *Beaches*: an area with a high population density, corresponding to a strip of land stretching between the sea and the lagoon. Most houses are either second residences or used for rent in the high season. This submarket includes the zones located along the Atlantic coast, namely: Costa Nova, Barra and São Jacinto.

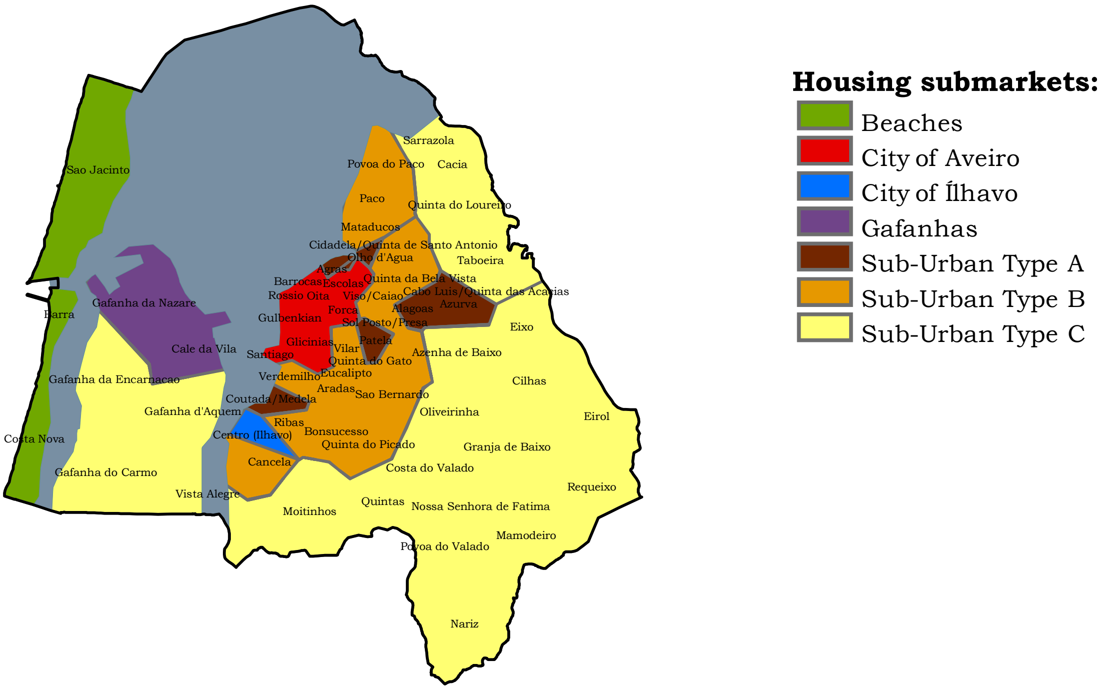


Figure 71 - Housing submarkets defined ex ante

A summary of descriptive statistics for the most important quantitative attributes (which is auto explicative) for each submarket is reported in *Table 41*.

Table 41 - Descriptive analysis for the seven housing submarkets

Submarkets		**N**	**Minimum**	**Maximum**	**Mean**	**Std. Dev.**
CBD Aveiro	Price (€)		30000	695000	143211.3	61979.6
	Price (€/m²)		353.8	5714.3	1316.8	330.9
	Total Areas (m²)		20	530	112.8	48.2
	Valid cases	3296				
CBD Ílhavo	Price (€)		28000	415000	149633.1	59788.3
	Price (€/m²)		208.3	2766.7	958.5	235.5
	Total Areas (m²)		40	600	164.1	76.4
	Valid cases	1188				
Gafanhas	Price (€)		35000	900000	141155.1	55595.6
	Price (€/m²)		202.7	2857.1	972.9	302.0
	Total Areas (m²)		35	520	155.3	70.2
	Valid cases	1765				
Suburban Type A	Price (€)		35200	750000	143685.0	68724.7
	Price (€/m²)		283.3	2473.2	1008.7	238.4
	Total Areas (m²)		42	593	151.1	83.7
	Valid cases	1421				
Suburban Type B	Price (€)		54000	675000	162729.1	69459.0
	Price (€/m²)		380.1	3167.9	1080.5	266.1
	Total Areas (m²)		38	600	162.2	86.4
	Valid cases	2480				
Suburban Type C	Price (€)		35700	900000	171863.6	72551.3
	Price (€/m²)		178.6	2909.7	905.9	264.1
	Total Areas (m²)		36	600	204.5	94.1
	Valid cases	1512				
Beaches	Price (€)		67500	748197	184199.4	73733.8
	Price (€/m²)		555.6	5272.7	1690.4	504.8
	Total Areas (m²)		40	540	117.4	62.2
	Valid cases	805				

The estimated hedonic model with spatial heterogeneity based on factors is reported in *Table 42*.

Table 42 - The estimated coefficients of the hedonic model using the factors

	Aggregate model	**CBD Aveiro**	**CBD Ílhavo**	**Gafanhas**	**Suburban Type A**	**Suburban Type B**	**Suburban Type C**	**Beaches**
Constant	**9.890**	9.786	10.638	10.560	10.567	10.016	10.375	15.122
	(236.87)***	(101,66)***	(55,36)***	(72,19)***	(86,53)***	(115,73)***	(89,63)***	(-16,56)***
Log total area	**-0.598**	-0.571	-0.685	-0.761	-0.762	-0.614	-0.693	-0.871
	(-70.79)***	(-30,14)***	(-22,20)***	(-29,30)***	(-29,83)***	(-34,61)***	(-29,28)***	(-25,66)***
Log TOM	**0.005**	0.006	0.016	0.003	0.011	0.004	-0.003	-0.007
	(3.69)***	(2,10)**	(3,99)***	(0,98)	(3,19)***	(1,30)	(-0,90)	(-1,53)
Factor 1	**-0.043**	-0.036	-0.164	0.099	-0.144	-0.025	0.001	-1.761
(Access to city centre)	**(-19.77)*****	(-3,65)***	(-1,57)	(2,29)**	-(6,34)***	(-2,27)**	(0,13)	-4,46***
Factor 2	**0.027**	0.010	0.180	0.042	-0.079	-0.098	-0.029	-0.146
(Health Centre, Parks and Gardens)	**(14.65)*****	(0,97)	(6,19)***	(2,04)**	(-7,06)***	(-7,58)***	(-2,17)**	(-0,84)
Factor 3	**0.077**	-0.016	-0.214	0.015	-0.120	-0.016	-0.005	-0.745
(Beaches, schools and local commerce)	**(38.21)*****	(-1,62)	(-2,78)***	(0,32)	(-4,31)***	(-1,29)	(-0,51)	(-5,48)***
Factor 4	**0.150**	0.199	0.217	0.209	0.242	0.162	0.171	0.211
(Housing dimension)	**(40.12)*****	(19,51)***	(15,64)***	(21,25)***	(20,65)***	(20,01)***	(15,60)***	(7,34)***
Factor 5	**0.043**	0.061	0.044	0.028	0.038	0.025	0.019	-0.002
(Additional desirable features)	**(21.34)*****	(15,13)***	(6,17)***	(5,15)***	(7,21)***	(5,92)***	(3,28)***	(-0,09)
Time fixed effects	**YES**	YES	YES	YES	YES	YES	YES	YES
Number of obs.	**12467**	**3296**	**1188**	**1765**	**1421**	**2480**	**1512**	**805**
Adjusted R Square	**0.572**	**0.359**	**0.459**	**0.483**	**0.557**	**0.498**	**0.484**	**0.557**

**** significant at the 1% level/ ** significant at the 5% level/ * significant at the 10% level*

The results show substantial heterogeneity across the submarkets, that is, hedonic coefficients for structural and neighbourhood characteristics are not constant across the several spatial units. Specifically, similar attributes are valued differently at different locations. The reason is clearly explained by the particularities of each context and locations, there are facilities related with centrality and others located in the sprawl areas. Several important observations follow.

The general hedonic model, including the seven submarkets, explains 57.2% of the total variation in the housing price (€/m2). The main inference drawn is that log prices are closely related to housing intrinsic characteristics and also strongly related to distances to urban facilities. All the coefficients are statistically significant at the 1% level and the estimated parameters have the expected sign.

Rather than presenting an exhaustive description of *Table 42*, some important messages are highlighted.

First, the submarket *Beaches* is quite a distinct housing market from the others, given that the estimated factor prices are very different from the rest of the submarkets.

Second, and in particular, the price elasticity of house area is the least for *Beaches* (0.129) and highest for the inner city of Aveiro (0.429). This implies that the size of houses designed for holidays and weekend purposes is not particularly valued, while the demand in the most affluent area (CBD of Aveiro) is considerably more sensitive to size.

Third, while the general model shows that prices increase with access to city centre, there is a large variation across the different submarkets. In the *CBD of Aveiro* or suburban areas close to the city, the negative value attached to poor access to the city centre is highly significant, while access is most valuable in the Beaches, which is the submarket located farthest from the centre. The same does not apply for the more remote *Suburban Type C* or submarkets such as *Ílhavo* or *Gafanhas*. This is explained by the different social profiles of inhabitants in these areas.

Fourth, access to local facilities has a heterogeneous effect on prices. However, by contrast to access to the centre, it is valued significantly, with the expected signs, in *Suburban Type A, B* and *C* submarkets. This means that proximity to local centres is prized in the suburban areas, but not in *CBD Aveiro*, and even negatively valued in *Ílhavo* and *Gafanhas*. This is because local centres in the more urbanised locations tend to produce negative externalities such as noise or a lack of parking space, while in the suburban areas they tend to be associated with better urban layouts and access to local amenities, different from that of unqualified suburban sprawl.

Fifth, additional facilities such as a garage, balcony and central heating are positively valued with high significance everywhere, except in *Beaches*, where such attributes do not matter.

Finally, living space is positively valued, and in largely equal measure, across all the seven submarkets.

In general, spatial heterogeneity is in line with the urban geography of Aveiro and reflects the dynamics of urban development, and its analysis is important to understand the spatial nature of the urban housing market and to provide guidelines for urban planning and housing policy. The unique character of the housing submarket in the *Beaches* is related to its evolution as the destination for second homes and rental properties for holiday-makers. Likewise, the high price elasticity for the housing area in the centre of Aveiro reflects scarcity rents. In turn, this shortage of housing space in the centre has led to migration from the city to the suburban areas, which have larger price sensitivities to access. It would thus appear that further development of quality housing and good local amenities and access to the centre would make the suburban areas both affordable and desirable for the urban population.

ii) Analytical perspective

While the approach applied above relies on an inductive perspective to delimit submarkets, focusing on ad hoc criteria, the following approach uses a cluster analysis underlying structures of data. Many authors used this technique to define submarkets, e.g.: Maclennan and Tu, (1996), Bourassa, Hamelink *et al.*, (1999) among others.

Cluster analysis is an analytical multivariate technique for developing meaningful subgroups of individuals with similar patterns on a defined set of characteristics or variables included in the analysis (see Hair, Black *et al.*, 2010, for a review). Thus, a first decision, before applying this taxonomy, is to define which variables are important to characterise and delimit submarkets.

Watkins (2001) and Bourassa, Cantoni *et al.*, (2008) provide a detailed review of the alternative approaches for characterizing and delineating housing submarkets. In Watkins' (2001) work, the author examines three different alternative approaches: i) *spatially stratified housing submarkets*, where geographic and locations attributes are key aspects to delimit housing segmentation[138]; ii) *structural characteristics submarkets*, which are based on the similarity of intrinsic housing characteristics[139]; and iii) *nested spatial/structural submarket*[140], which is a hybrid definition that nests dwelling characteristics based submarkets within spatially defined submarkets. The main conclusion of the study is that the nested model provided the best empirical approach for delineating submarkets.

The housing segmentation analysis, used in this empirical work, is based on the *nested spatial/structural submarket* approach and developed as follows. First, factor analysis is used to extract a reduced set of orthogonal factors from the original variables. Thus, factor scores (not variable values) are used for clustering purposes. These standardising scores are especially important if variables are measured on different scales, which is the case (see Hair, Black *et al.*, 2010). Second, cluster analysis is used on the principal components to determine the most appropriate composition of housing submarkets. Third, hedonic regressions are performed for each of the submarkets and compared.

In applying cluster analysis three major steps are:

1. Clustering algorithm;
2. Clustering criterion;
3. Dissimilarity measure.

---

[138] This approach could easily be made based on the classes (zoning) defined in Figure 57, where each class corresponds to a specific submarket, spatially defined.
[139] This approach has been applied in Castro, Marques *et al.* (2011), where cluster analysis was used on the principal components to determine the most appropriate composition of housing submarkets.
[140] As the following quote from Maclennan and Tu (1996, p.395) shows, it is possible to find different situations in some places within cities: it may be "*(...) that excess demands will relate to dwelling types and that submarkets will be sectoral; (…)that excess demand is evenly spread by sector but that particular locational combinations are difficult to replicate; and (…), both sectoral and spatial submarkets may both exist at the same time.*"

There are numerous ways in which clusters can be formed and several cluster algorithms exist in the literature using hierarchical and non-hierarchical methods (see Hair, Black *et al.*, 2010, for a review). Non-hierarchical cluster methods are often referred to as k-means and work by partitioning the data into a user-specified number of clusters and then iteratively reassigning observations until some numeric goal related to cluster distinctiveness is met (Hair, Black *et al.*, 2010); thus, it involves *a priori* knowledge of the number of clusters to be created. The drawback of this method is that it gives a simple or flat partition, with a single set of clusters. Further research in the application of these clustering procedures is undertaken by Bourassa, Hoesli *et al.* (2003). The authors use several segmentation approaches including principal component analysis and a k-means clustering procedure to delimit housing submarkets. Hierarchical clustering does not require *a priori* knowledge of the number of clusters and a tree-like structure (called dendogram) is constructed to see which clusters (housing zones) are grouped at which iteration, and thus the relationship among them. If it starts with individual observations (agglomerative method), at each successive iteration, two groups with the shortest distance are merged together based on a predetermined distance measure. But it can also start with every case being a cluster. In this case, in each following step one of the groups is divided into two. In both approaches, once a cluster is formed, it cannot be split. This can be considered a disadvantage because undesirable early combinations may persist throughout the analysis and lead to feasible results. Goodman and Thibodeau (1998) used this method and conclude that hierarchical models provide a useful framework for delineating housing submarkets. In this particular empirical study an agglomerative hierarchical clustering is applied since it is one of the most straightforward methods and probably the most widely used among the hierarchical methods (Hair, Black *et al.*, 2010).

Most common statistical packages (e.g., SPSS) use five different agglomerative methods: single linkage (nearest neighbour approach), complete linkage (furthest neighbour), average linkage, Ward's method, and centroid method. In depth reviews about the specificities of clustering criteria are well presented in the statistical book of multivariate analysis, Hair, Black *et al.* (2010), among others.

Another important issue to consider in cluster analysis is the type of distances that should be used to assess the similarity or dissimilarity between individual properties or zones. Euclidean distance is the most commonly used distance measure of similarity between two objects (Hair, Black *et al.*, 2010). It is a measure of the length of a straight line drawn between two objects when represented graphically.

After experimenting with alternative clustering criterion and dissimilarity measures, cases are clustered into five groups by means of Ward's method and Euclidean distance. In short, Ward’s method is distinct from all other methods because it is based on an analysis of variance approach to evaluate the distances between clusters, that is, minimizing the sum of squares of any two (hypothetical) clusters that can be formed at each step (Hair, Black *et al.*, 2010).

Choosing the number of clusters in hierarchical clustering means choosing at which level the dendrogram should be cut. As can be seen in the *Figure 72*, eight is the most efficient number of clusters (see *Table 43*), in the sense that it minimizes the variability within clusters and maximizes the variability between clusters. Analysing the dendogram the existence of two distinct submarkets is immediately clear. However, because price prediction accuracy increases for greater levels of disaggregation, as shown by Thibodeau (2003), a more detailed segmentation is selected.

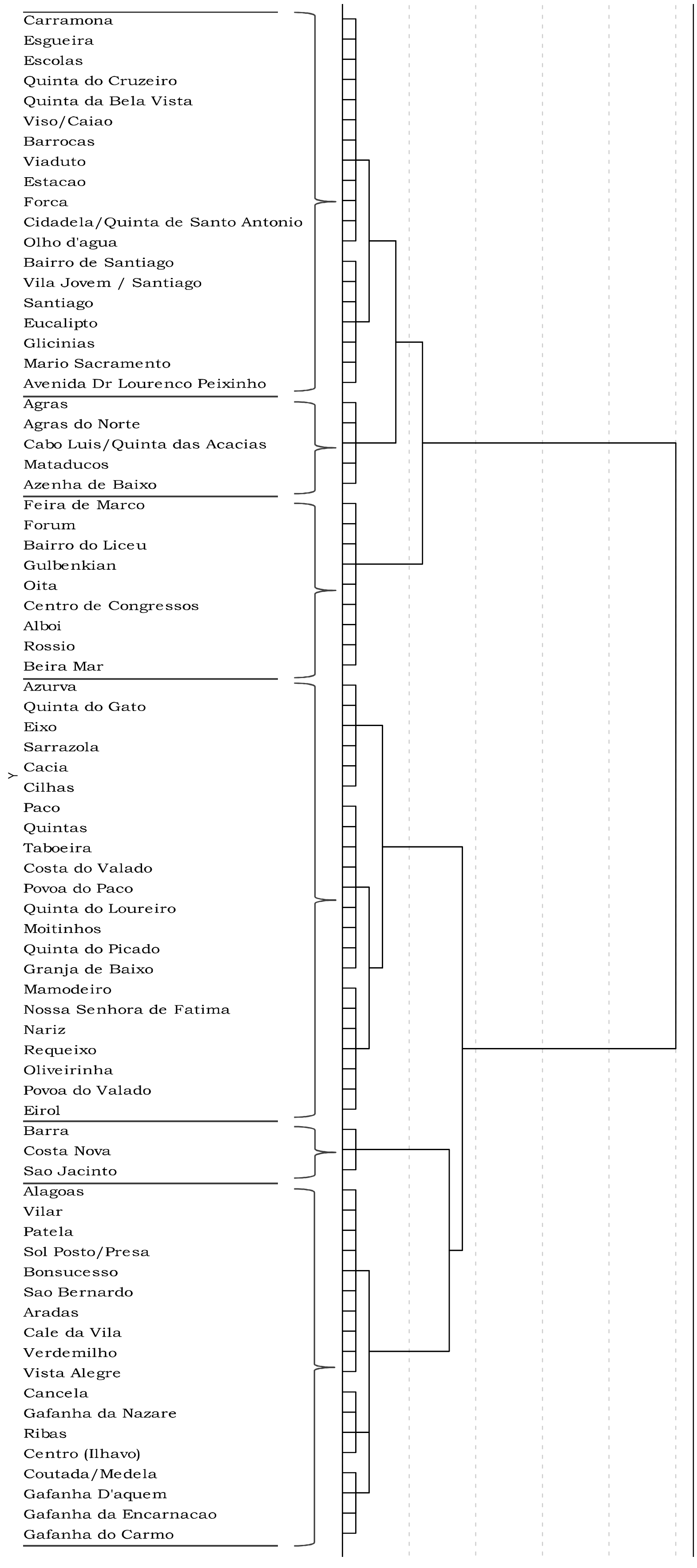


Figure 72 - Dendogram using Ward's linkage

Table 43 - Factor scores by zone and clusters formed

Zones	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Clusters
Agras	-1.09	0.57	-0.34	0.27	0.59	1
Agras do Norte	-0.97	0.63	-0.02	0.25	0.59	1
Azenha de Baixo	-1.35	1.10	0.59	1.84	0.38	1
Cabo Luis/Qta Acacias	-0.55	0.43	-0.16	1.27	1.09	1
Mataducos	-0.22	1.00	-0.67	1.13	0.16	1
Alagoas	-0.16	-0.43	-0.35	-0.30	0.41	2
Aradas	0.07	0.40	-0.47	0.15	-0.21	2
Bonsucesso	0.47	0.22	-0.63	0.09	0.10	2
Cale da Vila	0.31	0.76	-0.09	0.06	-0.19	2
Patela	-0.10	-0.26	-0.40	0.69	0.19	2
Sao Bernardo	0.17	0.41	-0.53	-0.21	0.20	2
Sol Posto/Presa	-0.30	-0.20	-0.93	0.22	0.22	2
Verdemilho	-0.22	0.33	0.51	0.41	0.10	2
Vilar	-0.47	0.03	-0.24	0.01	0.13	2
Vista Alegre	0.99	0.27	0.29	0.12	0.59	2
Alboi	-2.70	-0.22	1.06	0.07	-1.09	3
Bairro do Liceu	-1.71	-0.94	0.68	0.15	-0.52	3
Beira Mar	-2.56	-0.18	0.89	0.54	-0.57	3
Centro de Congressos	-1.62	-0.74	0.10	0.31	-1.47	3
Feira de Marco	-2.24	-0.56	0.44	0.28	0.18	3
Forum	-1.81	-1.30	-0.05	0.26	0.07	3
Gulbenkian	-2.12	-0.36	0.27	0.41	-1.30	3
Oita	-1.98	-0.76	0.41	0.55	-1.05	3
Rossio	-2.19	-0.10	1.28	0.12	-1.13	3
Av. Dr Lourenco Peixinho	-0.96	-0.66	0.00	-0.47	-0.09	4
Bairro de Santiago	-1.02	-0.65	0.26	-0.22	-1.20	4
Eucalipto	-0.94	-0.60	0.51	-0.20	-0.19	4
Glicinias	-0.86	-0.62	0.64	0.19	-0.33	4
Mario Sacramento	-0.95	-0.82	0.43	-0.34	-0.61	4
Santiago	-0.72	-0.02	0.15	-0.41	-0.84	4
Vila Jovem / Santiago	-0.63	-0.50	0.29	-0.42	-1.01	4
Barrocas	-1.46	0.08	-0.12	-0.07	-1.01	4
Carramona	-1.01	0.27	-0.63	-0.51	-1.01	4
Cidadela/Qta Sto Antonio	-0.54	1.05	0.12	-0.59	-1.01	4
Escolas	-1.12	0.42	-0.82	-0.47	-1.01	4
Esgueira	-0.95	0.31	-0.51	-0.50	-1.01	4
Estacao	-1.68	0.15	0.23	-0.29	-1.01	4
Forca	-1.49	-0.46	-0.45	-0.23	-1.01	4
Olho d'Agua	-0.20	0.89	-0.10	-0.76	-1.01	4
Quinta da Bela Vista	-0.61	0.50	-0.40	-0.40	-1.01	4
Quinta do Cruzeiro	-0.91	0.36	-0.53	-0.56	-1.01	4
Viaduto	-1.24	0.35	-0.17	-0.23	-1.01	4
Viso/Caiao	-0.69	0.41	-0.27	0.04	-1.01	4

(Cont)→

Zones	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Clusters
Azurva	0.27	1.11	-1.02	-0.55	-0.37	5
Cacia	0.68	1.19	-1.03	0.02	-0.08	5
Cilhas	0.98	1.73	-0.46	-0.66	0.17	5
Eixo	0.82	1.57	-1.08	-0.19	0.00	5
Quinta do Gato	0.08	0.30	-1.08	-0.78	0.14	5
Sarrazola	0.92	1.52	-1.42	-0.22	-0.07	5
Barra	0.51	1.17	3.41	-0.19	0.21	6
Costa Nova	1.20	1.06	1.75	-1.05	-0.15	6
Sao Jacinto	2.34	1.32	2.34	0.71	-0.91	6
Cancela	0.85	-0.63	0.18	-0.05	0.09	7
Centro (Ilhavo)	0.87	-1.94	-0.11	0.14	-0.14	7
Coutada/Medela	0.40	-0.20	-0.21	1.32	-0.29	7
Gafanha da Encarnacao	1.20	-0.27	-0.15	0.63	-0.03	7
Gafanha da Nazare	0.96	-1.04	0.30	-0.17	0.01	7
Gafanha d'Aquem	0.67	0.02	0.14	1.10	0.11	7
Gafanha do Carmo	1.93	0.01	-0.20	0.54	-0.59	7
Ribas	0.67	-0.56	-0.23	-0.46	0.06	7
Costa do Valado	0.22	1.53	-0.39	0.31	0.45	8
Granja de Baixo	0.80	1.62	-0.22	1.30	0.49	8
Moitinhos	0.89	0.97	-0.16	0.79	0.40	8
Paco	0.65	1.15	-0.64	0.46	0.06	8
Povoa do Paco	0.84	1.21	-0.70	0.75	0.10	8
Quinta do Loureiro	0.84	1.23	-0.46	0.83	0.22	8
Quinta do Picado	0.55	0.69	-0.57	0.68	0.12	8
Quintas	0.42	1.34	-0.76	0.33	0.01	8
Taboeira	0.53	1.42	-1.01	0.88	0.22	8
Eirol	1.37	1.94	-1.59	0.81	-0.77	8
Mamodeiro	1.29	1.89	-0.47	0.64	-0.11	8
Nariz	1.95	1.91	-0.66	0.72	-0.29	8
Nossa Senhora de Fatima	1.28	1.69	-0.44	0.86	-0.14	8
Oliveirinha	0.72	1.16	-0.92	1.31	-0.33	8
Povoa do Valado	1.02	1.90	-0.75	1.31	-0.55	8
Requeixo	1.80	1.44	-0.70	1.10	-0.19	8

The results are mapped out in *Figure 73*, and, as expected, they form generally well defined spatial areas.

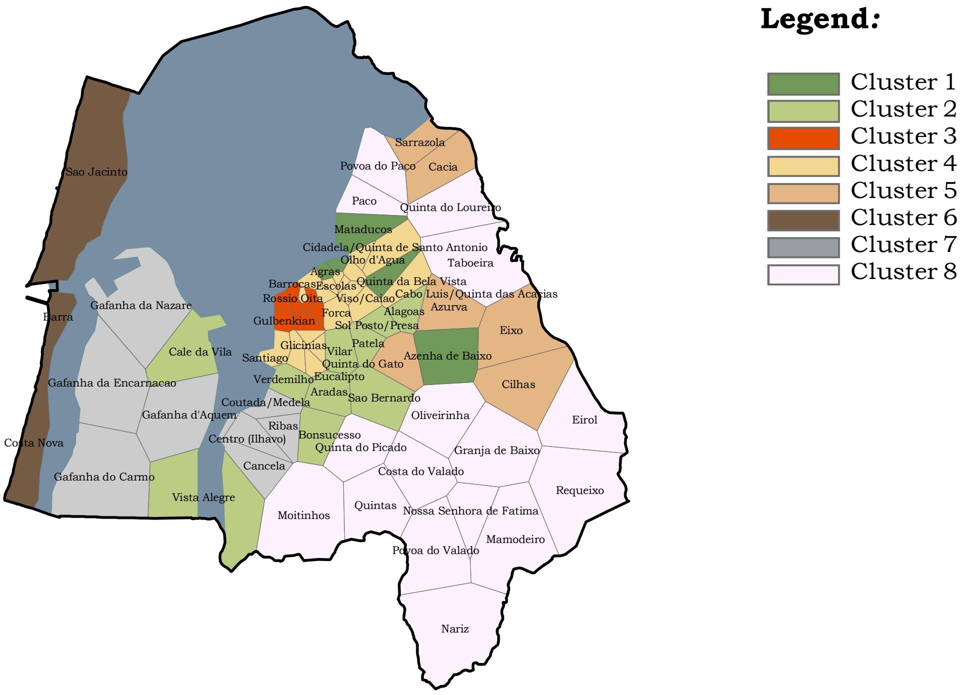


Figure 73 - Housing submarkets defined ex post

Table 44 reports the importance of five factors in each cluster (positive and negative) justifying the reason for its composition.

Table 44 - Clusters classification

	Factor 1 *Distance to the centre or central amenities*	**Factor 2** *Distance to local amenities*	**Factor 3** *Distance to beaches, schools and local commerce*	**Factor 4** *Housing dimension and single houses*	**Factor 5** *Additional desirable features*
C1	--	++	-	++	++
C2		+	--	+	+
C3	--	--	++	++	--
C4	--		++	--	--
C5	++	++	--	--	
C6	++	++	++	-	+
C7	++	-	-	+	
C8	++	++	--	++	

Legend: (++) more than 90% of positive scores; (+) between 80 and 90% of positive scores; (--) more than 90% of negative scores; (-) between 80 and 90% of negative scores.

Analysing *Table 44* the dichotomy between two groups of submarkets is evident: one group (C1, C3, and C4) is located close to the city centre where central amenities are abundant; and the other group is placed more distant from the city centre (C5, C6, C7, and C8). The spatial contiguity is verified in all clusters but not exclusively. Looking in more detail at each submarket the main characteristics are emphasised.

- C1: This set of zones belongs to a submarket characterised by a high level of centrality to the city centre (see *Figure 63*), but are also quite far from local amenities (see *Figure 65*). Regarding the physical attributes, this submarket has typically big properties with garages, balconies and central heating (housing features).
- C2: This second group of zones is well served in terms of local services (see *Figure 65*) and score factors for physical characteristics of properties are positive, meaning that houses located in this submarket are bigger and have more housing facilities than the average.
- C3: This cluster (submarket) corresponds to the *CBD of Aveiro* and is a distinguishable cluster because it has good accessibilities (both local and central) but bad housing facilities (almost all scores of the factor 5 are negative). The high positive values in the scores of factor 4 do not mean that properties are big, but that there are there few small houses. Note that variable housing type is included in this factor.
- C4: This submarket encompasses zones which typically are considered to be the urban city centre of Aveiro. This submarket has good accessibilities to beaches, schools and local commerce, has small housing, typically flats with less garages, balconies, terraces and other housing features than the average.
- C5: The fifth housing submarket belongs to the group which has bad access to the central and local facilities. Additionally, the access to beaches, schools and local commerce is not good, when compared with the average of housing submarkets for Aveiro.

- C6: This submarket is very similar with the previous submarket 5. The differences lie in the easy access to beaches, schools and local commerce. In this cluster, dwellings are typically flats (or small houses) with good additional facilities.
- C7: Distant to central amenities but high level of accessibility to local amenities, beaches, schools and local commerce are aspects that describe this housing submarket. Regarding the intrinsic housing attributes this submarket has traditionally larger single houses.
- C8: This last cluster has positive scores in almost all factors, except factor 3. It means that this housing submarket is distant from central amenities and houses are more spacious than the average (see *Figure 69*). The negative sign for factor 3 is not because is close to the beaches but reflects the access to schools and local commerce, as shown in the *Figure 66*.

Table 45 and *Table 46* present the adjusted R-squared of each model and a summary of descriptive statistics for each submarket in this housing spatial disaggregation. The interpretation of regression coefficients relies on the same conclusions described above.

Table 45 - Descriptive analysis for the eight housing submarkets

Clusters		**N**	**Minimum**	**Maximum**	**Mean**	**Std. Dev.**
C1	Price (€)		69000.0	750000.0	223096.1	92443.2
	Price (€/m²)		431.0	2142.9	998.6	315.3
	Total Area (m²)		70.0	600.0	238.6	96.9
	Valid Cases	**270**				
C2	Price (€)		54000.0	675000.0	157829.5	66439.4
	Price (€/m²)		333.3	3167.9	1112.0	248.9
	Total Area (m²)		37.5	550.0	151.1	78.8
	Valid Cases	**2669**				
C3	Price (€)		30000.0	600000.0	134774.9	58787.7
	Price (€/m²)		421.9	3312.5	1465.2	418.8
	Total Area (m²)		20.0	476.0	100.4	53.5
	Valid Cases	**566**				
C4	Price (€)		39400.0	695000.0	143189.6	62062.2
	Price (€/m²)		353.8	5714.3	1267.8	305.8
	Total Area (m²)		30.0	530.0	116.0	48.1
	Valid Cases	**2882**				
C5	Price (€)		35200.0	900000.0	133250.8	62154.3
	Price (€/m²)		283.3	2438.7	931.9	244.8
	Total Area (m²)		42.2	600.0	154.7	86.8
	Valid Cases	**1113**				
C6	Price (€)		61000.0	748196.8	185030.6	73883.2
	Price (€/m²)		509.3	5272.7	1667.2	515.0
	Total Area (m²)		40.0	540.0	120.9	65.7
	Valid Cases	**833**				
C7	Price (€)		28000.0	900000.0	147366.7	59478.0
	Price (€/m²)		202.7	2857.1	968.0	274.5
	Total Area (m²)		35.0	600.0	161.6	74.2
	Valid Cases	**3092**				
C8	Price (€)		50000.0	650000.0	178768.1	67694.1
	Price (€/m²)		178.6	2909.7	864.0	239.9
	Total Area (m²)		36.0	540.0	221.5	94.0
	Valid Cases	**1042**				

The models presented below have the following specification;

$$\ln P_{h(€/m^2)} = \alpha_{h1} \ln \mathrm{A}_{h} + \alpha_{h2} \ln \mathrm{T}_{hTOM} + \sum_{i=1}^{5} \alpha_{hi} \mathrm{F}_{hi} + \sum_{k=1}^{114} \alpha_{hk} \mathrm{d}_{hk} \qquad \text{Eq. 55}$$

Where:

$\ln P_{h(€/m^2)}$ is the housing price measured as a logarithm of euros per square metre in each submarket h (h=1, ..., 7)

$\ln A_h$ is logarithm of the housing total area measured in m² in each submarket h

$\ln T_{hTOM}$ - is the time on the market (TOM) measure as a logarithm of days for each submarket h

F_{hi} are factors scores (i=1,...5)

d_{hk} are the dummy variables for time (monthly dummies) (k=1,...114)

Table 46 - Model summary for the eight housing submarkets

Cluster 7	R	R Square	Adjusted R Square	Std. Error of the Estimate
C1	0.818	0.669	0.566	0.198
C2	0.736	0.542	0.526	0.161
C3	0.845	0.714	0.669	0.177
C4	0.657	0.432	0.414	0.184
C5	0.787	0.619	0.587	0.170
C6	0.790	0.625	0.587	0.201
C7	0.708	0.502	0.486	0.191
C8	0.728	0.530	0.489	0.203

In order to assess the price prediction accuracy defined by the two approaches (deductive and inductive), *Table 47* presents the comparison of the prediction accuracy measures for an aggregated prediction model. These models have an additional component which corresponds to the dummy variables of each submarket (d_j) and is defined by:

$$\ln P_{(€/m^2)} = \alpha_1 \ln A + \alpha_2 \ln T_{TOM} + \sum_{i=1}^{5} \alpha_i F_i + \sum_{k=1}^{114} \alpha_k d_k + + \sum_{j=1}^{n} \alpha_j d_j \quad \text{Eq. 56}$$

Table 47 - Price prediction accuracy for three proposed housing submarket models

	Model 1	**Model 2**	**Model 3**	**Model 4.1**	**Model 4.2**
(Constant)	9.893***	12.067***	10.215***	10.219***	10.073***
Ln Total Area (m^2)	-0.603***	-0.793***	-0.666***	-0.665***	-0.657***
lnTOM	0.004***	0.003**	0.005***	0.004***	0.004***
F1 (*Access to the city centre*)	-0.042***	-0.357***	-0.023***	-0.089***	0.036***
F2 (*Health centre, Parks and Gardens*)	0.028***	0.702***	-0.033***	0.018***	0.018***
F3 (B*eaches, Schools and local Com.*)	0.077***	-0.843***	-0.032***	-0.047***	-0.026***
F4 (*Housing dimension*)	0.156***	0.264***	0.192***	0.193***	0.187***
F5 (*Additional desirable features*)	0.047***	0.071***	0.044***	0.041***	0.047***
Time fixed effects			YES		
Number of obs.			12467		
Adjusted R Square	0.578	0.683	0.607	0.614	0.635
Number of Submarket (n)	1	76	7	8	6

**** significant at the 1% level/ ** significant at the 5% level/ * significant at the 10% level*

Model 1: General model for the entire study area without any housing segmentation

Model 2: Model 1 + dummy variables for 76 zones (Figure 48)

Model 3: Model 1 + dummy variables for 7 housing submarkets defined ex ante (Figure 71)

Model 4.1: Model 1 + dummy variables for 8 housing submarkets defined ex post using factors (Figure 73)

Model 4.2: Model 1 + dummy variables for 6 housing submarkets defined ex post, applying a cluster analysis to the location coefficients presented in *Table 37* (see *Figure 84*)

VII.5.6. Spatial interaction effects using a known spatial weight matrix

In this section, the macro scale approach (housing market of Aveiro and Ílhavo) is first examined for spatial autocorrelation by Moran's I test statistic (Moran, 1948) and by the LISA test (local indicators of spatial association), presented in *Figure 78, Figure 79, Figure 80* and *Figure 81* (see Anselin, 2005). These two measures of spatial association give an overall idea about the spatial pattern in the data. Next, an OLS regression is computed, in order to evaluate whether a spatial autocorrelation component has been internalized by the regression parameters or not. If not, the spatial lag (related with interactions between houses in space) or spatial error term (related to the correlation with the error terms of observations located nearby) should be embedded. If neglected the estimated parameter would be biased and inefficient (Anselin and Bera, 1998; LeSage and Pace, 2009).

As mentioned during this work the choice of spatial weights matrix (which is used to explain the spatial relationship between observations, in this case between zones), plays a crucial role in capturing spatial dependence. Therefore, as has been done in the previous analysis (Section V.II.4), seven weights matrices are tested for this dataset using the robust Lagrange Multiplier (LM) test, both for spatial lag and error specifications. All matrices used are based on distances between observations within ranges from 0 to 250 (d250), 500 (d500), 1000 (d1000), 1500 (d1500), 2000 (d2000) and 5000 (d5000) meters, and on contiguity (*Queen* and *Rook*). Note that since the sample of this dataset involves a larger territorial area, the range of distances used in this analysis is not comparable with the smaller dataset.

Because this dataset does not have the spatial georeferentiation of individual properties (instead zones are used where a set of different properties are included - *Figure 48*), the procedure used to build the spatial weights matrix is different from that commonly followed. Weights matrices based on zones, identifying zones in the same range of influence (75 x 75 matrix)[141], should be computed in order to match the set of properties in each zone *Figure 75*. A simple example for a *Rook* contiguity matrix with five zones (Paço, Póvoa do Paço, Cacia, Quinta do Loureiro and Sarrazola) is shown in *Figure 74* to illustrate the method.

Figure 74 - Neighbourhood of the zone Póvoa do Paço (Rook contiguity matrix)

[141] São Jacinto has been excluded in this spatial dependence analysis because it is considered that it has no spatial relation with the adjacent zones. To get there a ferry boat is primarily used for tourism purposes. The alternative is 40 kilometers by road via the municipalities of Estarreja and Murtosa.

		Póvoa do Paço			Sarrazola			Cacia			Paço			Quinta Loureiro			Other Zones			
		1	...	97	1	...	235	1	...	11	1	...	92	1	...	129	...	...	...	Σ=12467
Póvoa do Paço	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	
	...	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	
	97	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	
Sarrazola	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	
	...	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	
	235	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	
Cacia	1	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	0	0	0	
	...	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	0	0	0	
	11	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	0	0	0	
Paço	1	1	1	1	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0	
	...	1	1	1	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0	
	92	1	1	1	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0	
Quinta Loureiro	1	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	0	0	0	
	...	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	0	0	0	
	129	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	0	0	0	
Other Zones	...	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	
	...	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	
	...	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	
	Σ=12467																			

Figure 75 - Illustration of the spatial weights matrix for five zones

The elements in each cell of the weights matrix (1 if contiguity, or distance measures if the criteria is the distance) are set to 1 for all observations within the specified level of influence (contiguity or distance), and 0 otherwise. To obtain a row standardized matrix each element of the matrix should be divided by its row sum. *Figure 77* gives a general perspective of the spatial weights by zone before matching the respective 12467 properties (in *Figure 76* codes of each zone are reported).

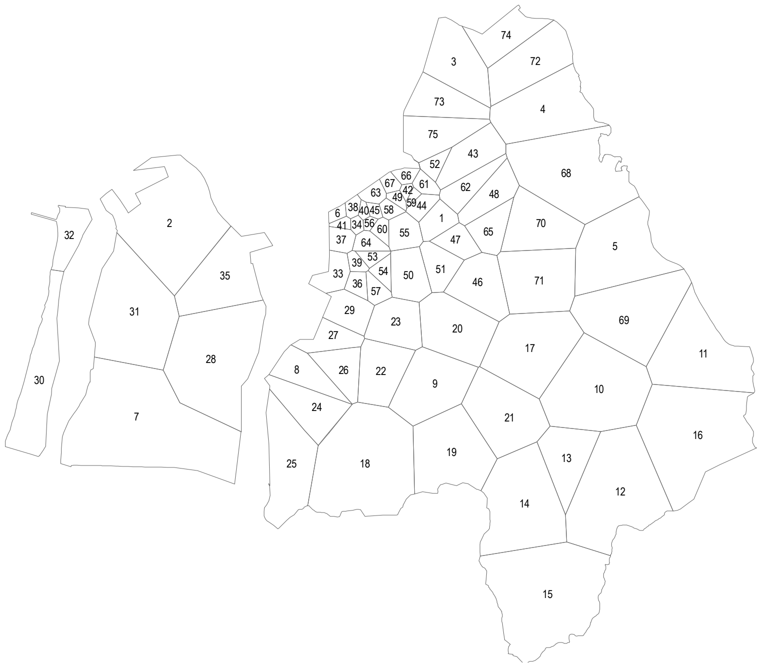


Figure 76 - Codes of the 75 zones used in spatial weight matrix

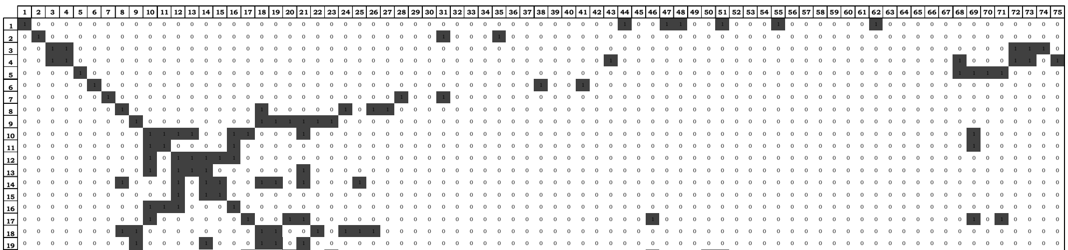


Figure 77 - Spatial weights matrix for 75 zones before being row standardized

The results from the Moran's I tests for the seven weights matrices are presented in *Figure 78*. In this analysis the use of distances to compute the spatial weights matrix may not be the most appropriate. The range of distances is very distinct between zones in different parts of the city. For example, in the urban part of the city the average distance between zones is 400 meters while in the rural area the value increases to 2000 metres; the shortest distance is 175 meters. Considering this argument, contiguity distances should be used.

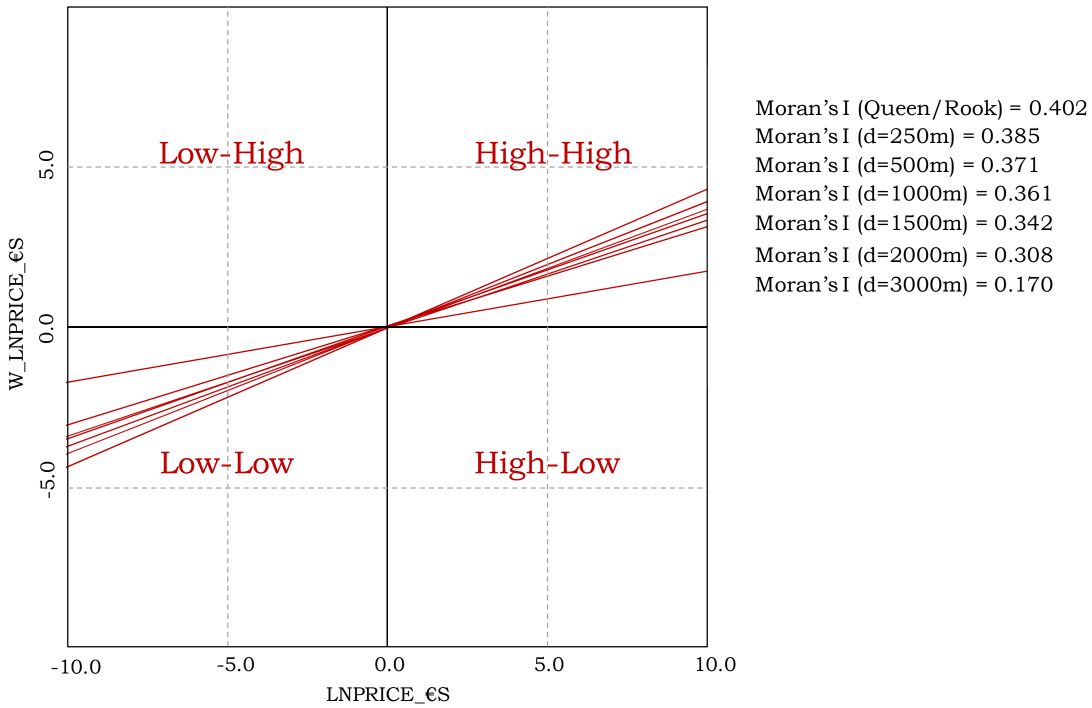


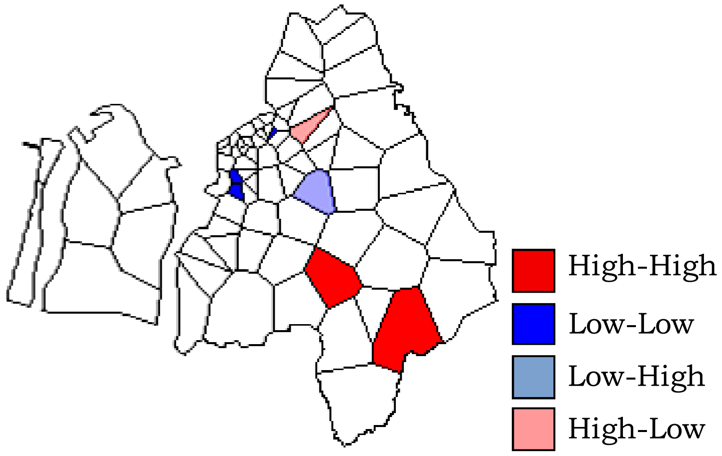
Figure 78 - Moran's I test and Moran scatter plot for 7 weighting matrices

As shown in *Figure 78*, the Moran's I value is 0.402, for the *Queen* Contiguity weight matrix, which is considered to be a high degree of spatial autocorrelation. For the other weights matrices the Moran's I value is also positive and significant, but decreases with the increased bandwidth, which is quite reasonable. This positive spatial association for the dependent variable means that high (low) values in property prices per square metre (measured in logarithms) in a location is surrounded by a large (small) value of the same variable[142]. In other words, adjacent houses tend to have a more similar price per

142 The opposite reasoning applies if negative spatial association exists, i.e., when the value of a price is large (small) in a location, it is small (large) in neighboring locations.

square meter than expected. The following figures represent the Moran's I and the indicator LISA for three variables (price measured in euros; price measured in euros per m^2 and area) and for three different situations: i) considering the average values for each zone (75); ii) a random samples of 481 cases computed because of the software limitations to compute Lambda and Rho (see *Table 48*); and ii) the entire dataset with 12467 observations.

Price [ln (euros)]



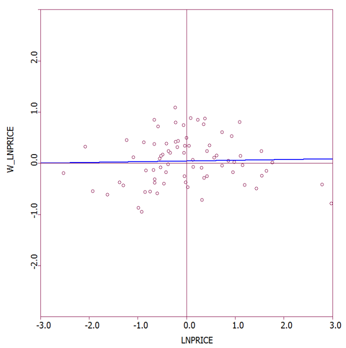
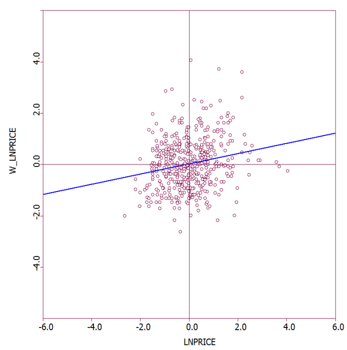
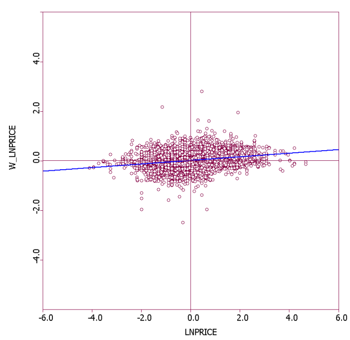
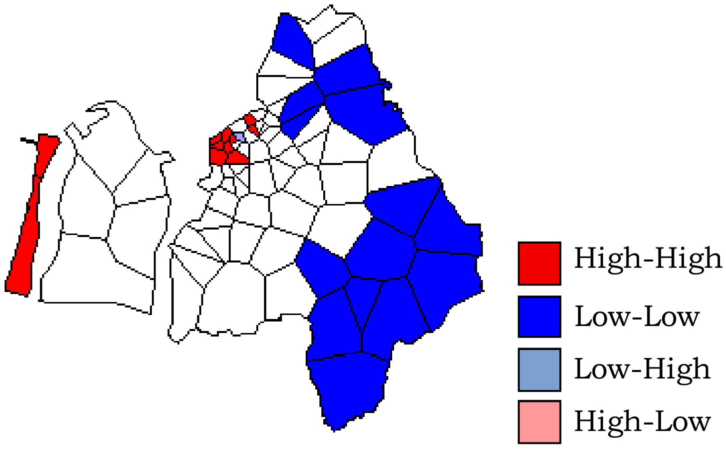
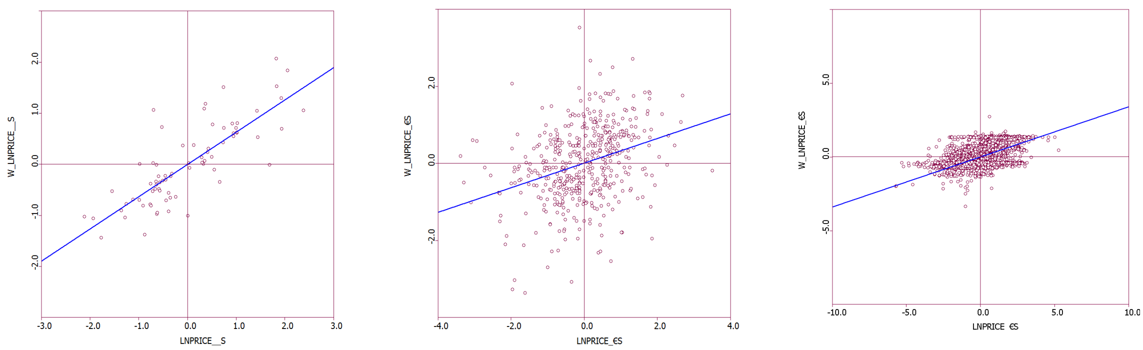
Moran's I=0.0145	*Moran's I=0.1999*	*Moran's I=0.0722*
		
75 zones	Random sample (481)	Dataset (12467)

Figure 79 - Moran's I test and the indicator LISA for the variable price [ln (euros)]

Price [ln(euros/m^2)]



Moran's I=0.6321	*Moran's I=0.3201*	*Moran's I=0.3201*



75 zones	Random sample (481)	Dataset (12467)

Figure 80 - Moran's I test and the indicator LISA for the variable price [ln(euros/m2)]

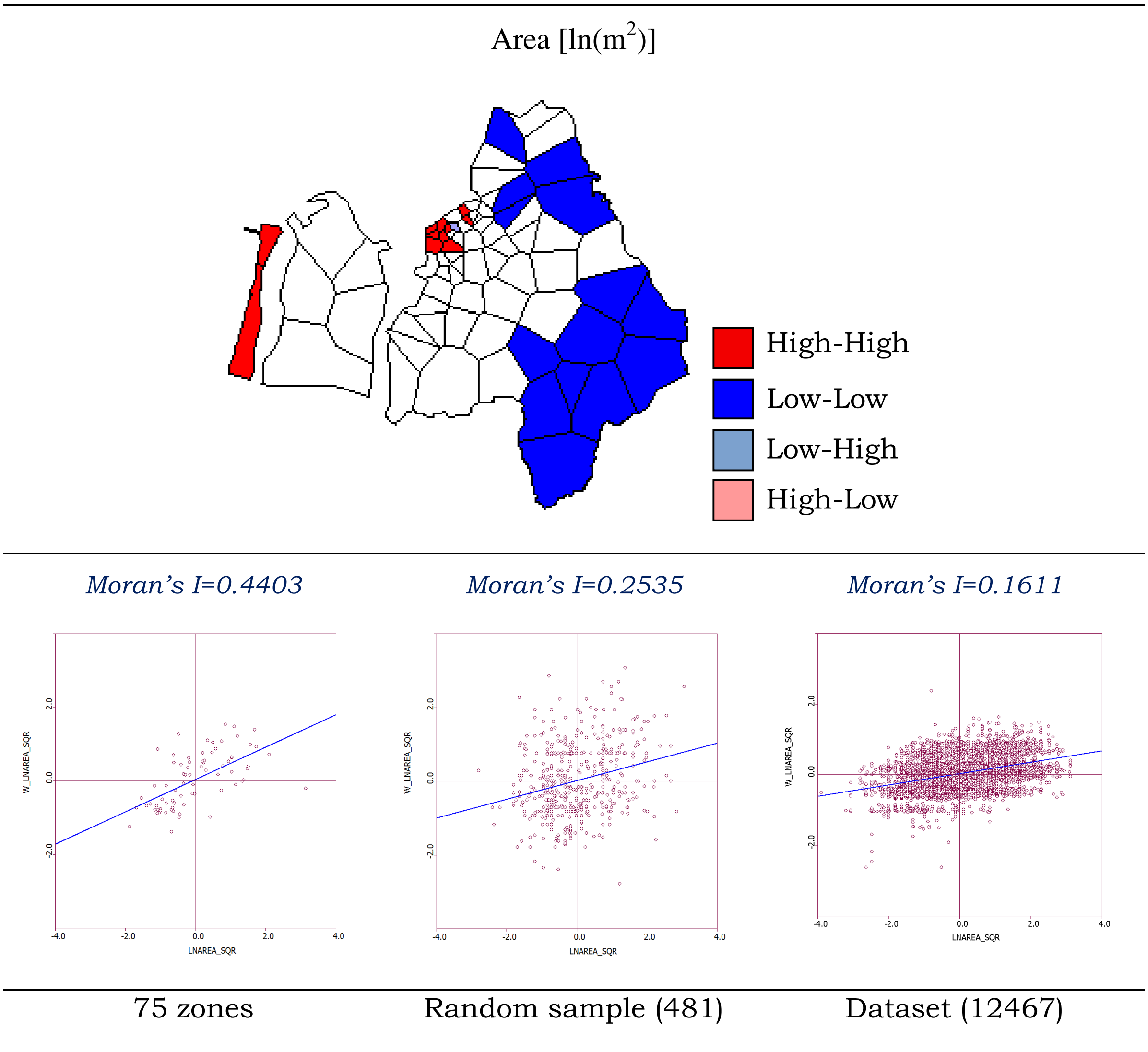


Figure 81 - Moran's I test and the indicator LISA for the variable area [ln(m2)]

As from the previous three figures (*Figure 79*, *Figure 80* and *Figure 81*), there is evidence of spatial autocorrelation, both in prices and in dwelling sizes. The areas illustrated in red (mainly in the inner city of Aveiro - *Figure 80*) have a high price (€/m²) for their dwellings and at the same time have in their neighbourhood properties with a high value. The interpretation is similar for the blue dots, but the dwellings prices are low, corresponding essentially to the areas located in the suburbs of Aveiro and Ílhavo with some exceptions. The urban centre of Ílhavo, being in the suburbs of Aveiro, has elevated prices, and therefore is in quadrant

High-Low (*Figure 78*). A final note concerns the properties with low prices located in a high-value area, such as is the case of Oita.

The results presented above (*Table 27*) are essentially descriptive. The next step is usually to examine further and resolve spatial dependence in a regression analysis.

Table 48 - OLS, spatial lag and spatial error model estimates

Independent Variables	**Global model**	**Zones 75**	**10 x Sample 481**
		Coefficients	
Constant	9.876 *(237.801)******	9.255 *(14.240)******	9.743 *(40.513)******
ln Total area	-0.600 *(-71.185)******	-0.481 *(-3.754)******	-0.584 *(-15.631******
ln TOM	0.006 *(5.634)******	0.018 *(0.622)*	0.015 *(-0.467)******
Factor 1	-0.044 *(-20.403)******	-0.045 *(-2.096)******	-0.043 *(-5.747)******
Factor 2	0.028 *(14.837)******	-0.032 *(-1.424)****	0.018 *(0.464)******
Factor 3	0.076 *(38.030)******	0.078 *(2.800)*	0.075 *(5.136)******
Factor 4	0.154 *(40.476)******	0.089 *(1.514)******	0.152 *(4.913)******
Factor 5	0.041 *(21.121)******	0.164 *(4.352)******	0.061 *(4.650)******
Lagrange Multiplier (lag)	Incomputable coefficients	9.239 *(p-value 0.002)*	6.713 (*p-value 0.027)*
Robust LM (lag)		18.504 *(p-value 0.000)*	1.639 (*p-value 0.373)*
Lagrange Multiplier (error)		0.075 (*p-value 0.784)*	5.626 (*p-value 0.033)*
Robust LM (error)		9.341 *(p-value 0.002)*	0.552 (*p-value 0.624)*
Lagrange Multiplier		18.580 *(p-value 0.000)*	7.265 (*p-value 0.068)*
N.º of observ.	12467	75	481
Adj. R-squared	**0.568**	**0.689**	**0.535**
Log likelihood		47.468	43.477
Lag coefficient(Rho)		0.429 (*p-value* 0.000)	0.007 (*p-value* 0.650)
Lag coefficient (Lambda)		0.241 (*p-value* 0.103)	-0.005 (*p-value* 0.960)

*t-/z-statistics in parentheses; ***significant at the 1% level/ **significant at the 5% level/ significant at the 10% level*

The results of the model presented below are obtained with the *Geoda software*. Due to operational constraints the initial sample was substantially reduced, both in terms of number of observations and in terms of dimensionality of the sample (factors are used). The results obtained after these changes are quite consistent with the general models, both the Ordinary Least Squares (OLS) and competing spatial dependence models (spatial lag and spatial error).

An apparent contradiction is presented by the coefficients of Moran (positive signal in *Figure 78*) and Lambda (negative signal in *Table 27*). It would generally not be expected that these two coefficients had opposite signs. One possible reason is the fact that the same sample has not been used for the analytical model, but the Moran test is also positive for all 10 samples of 481 observations. A more careful analysis shows that there is a considerable homogeneity in prices that is reflected by the positive autocorrelation of the Moran index. Moreover, within each zone this homogeneity, when controlled for the characteristics of the property, disappears, and therefore, the analytical model of the spatial value of lambda appears negative.

VII.5.7. Spatial interaction effects using a unknown spatial weight matrix

As mentioned in the Section VII.4, the use of a spatial weighting matrix based on the geographic notion of distance between housings may lead to incorrect interpretations, since the strength and type of relationships are not necessarily a decreasing function of physical distance. For instance, a set of dwellings located very close together in terms of physical proximity may be spatially less dependent than another set of dwellings that are not close together. This concern suggests that the notion of neighbourhood, and consequently the definition of a spatial weights matrix, seem to be more complex than the consideration of a physical space. Thus, in this section the neighbours are not *a priori* assumed neither defined in terms of a pure sense of physical proximity between two geographic points.

The methodology explained in the Section VI (and already applied in the Section VII.4, for the small dataset) is applied for the three different housing submarkets defined above (see *Table 47* - Model 3, 4.1 and 4.2). However, a detailed explanation is done only for one spatial housing submarket, the one presented in *Figure 71*. The results of the spatial weights matrix are presented in *Table 49* and *Figure 82* and described below. For the other two housing segmentations (defined by cluster analysis) only the results of spatial interaction across submarkets are shown (see *Figure 83* and *Figure 84*).

Table 49 - Symmetric spatial interaction matrix across seven submarkets

Submarkets	**CBD Aveiro**	**CBD Ílhavo**	**Gafanhas**	**Suburban Type A**	**Suburban Type B**	**Suburban Type C**	**Beaches**
CBD Aveiro	**0.00**						
CBD Ílhavo	0.0231^{**}	**0.00**					
Gafanhas	–0.0089	0.0521^{***}	**0.00**				
Sub. Type A	0.0415^{***}	0.0495^{***}	-0.0725^{***}	**0.00**			
Sub. Type B	-0.0190^{***}	0.0047	-0.0404^{***}	0.0189^{***}	**0.00**		
Sub. Type C	0.0227^{***}	0.0984^{***}	0.0263^{**}	-0.0309^{**}	0.0427^{***}	**0.00**	
Beaches	0.0674^{***}	0.0012	0.0328^{**}	0.0062	0.0274^{**}	0.0406^{***}	**0.00**

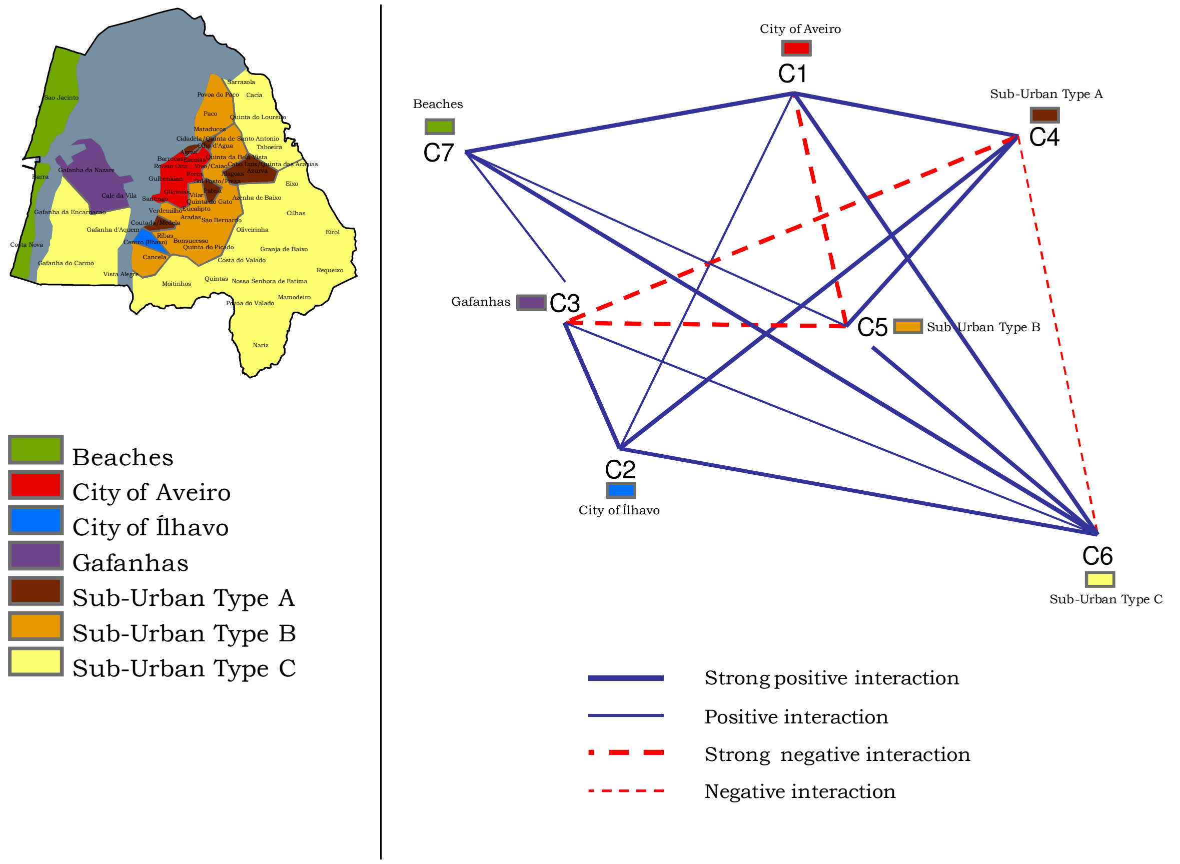


Figure 82 - Spatial interaction across seven submarkets

As expected, contiguity or distance explains a number of the significant positive spatial weights across submarkets in Aveiro. These include: spatial weights between *Beaches*, *Gafanhas* and *Suburban Type C*; and between *Suburban Type A* and *Suburban Type C* on the one hand and *CBD Aveiro*, *CBD Ílhavo* and *Suburban Type B* on the other.

However, the spatial weights between some pairs of contiguous regions are not statistically significant or even negative (for example, between *CBD Aveiro* and *Suburban Type B*), and some other significant weights related to non-contiguous regions. In other words, many significant spatial weights appear to be driven by reasons other than geographic distance or contiguity. Specifically for some of these submarkets, positive spillovers appear to be related to a combination of the core-periphery relationship and socio-cultural distances. Examples include: *CBD Aveiro* and *CBD Ílhavo*; *Beaches* and *CBD Aveiro*; and *CBD Ílhavo* and *Gafanhas*.

Finally, *Table 49* indicates significant negative spatial interactions between *CBD Aveiro* and *Suburban Type B*, and between *Suburban Type A* and *Suburban Type C*. Apparently, both of these are related to market segmentation, where each submarket is attractive to different segments of the population.

Spatial interactions considering the submarkets defined by the cluster analysis is presented in *Figure 83* and *Figure 84*.

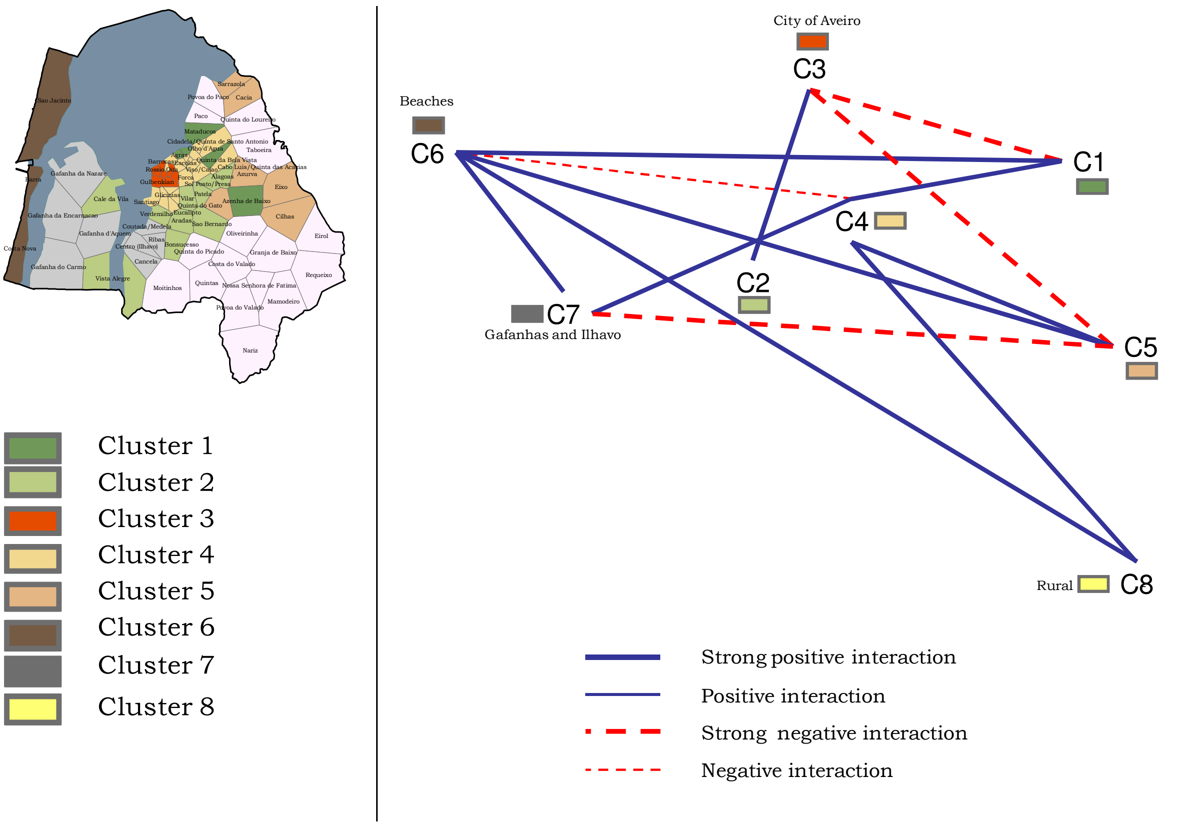


Figure 83 - Spatial interaction across eight submarkets

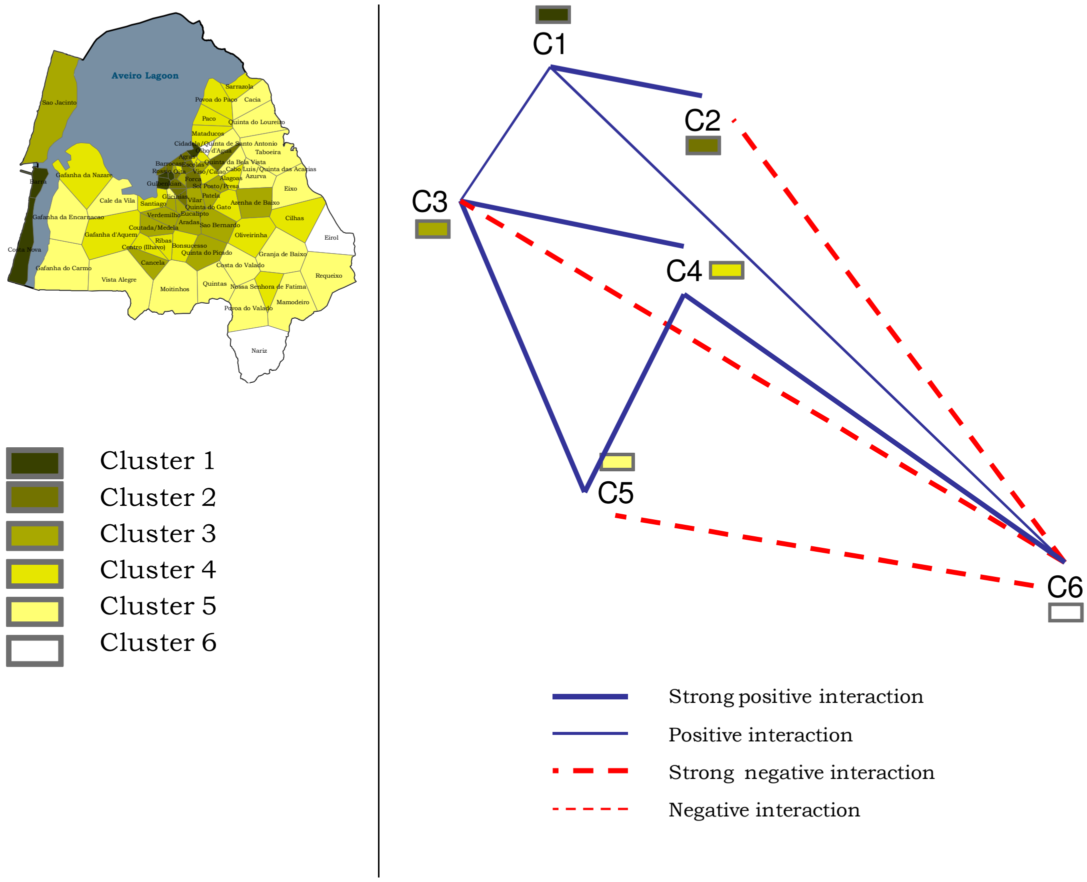


Figure 84 - Spatial interaction across six submarkets

Admittedly, some of the above explanations are tentative, and would require further research to confirm and interpret. However, what is clearly shown is that the spatial weights matrix, estimated based on this methodology, combined with the analysis of spatial heterogeneity, provides a very rich information set which can be the basis for detailed analysis and help uncover the causes underlying the observed spatial patterns[143].

Finally, the above analysis at a larger spatial scale, in combination with previous analysis (based on central parishes), provides some insights about the importance of spatial scale. Largely focusing on the urban scale, the previous analyses provided useful inferences with regard to spatial heterogeneity and interactions across parishes. However, understanding of spillovers between the urban and suburban parishes is somewhat limited by the fact that the suburban area contains a heterogeneous mix of neighbourhoods. This issue is addressed in

[143] Paradoxically, the main reason mitigating against more formal analysis of spatial structure using the estimated spatial weights matrix is large sample size. Specifically, current methods do not allow for ML based inferences in spatial econometric models when sample size is large. Suitable methodology for large sample applications, based perhaps on regularisation or subsampling, is planned for the future

the current analysis by dividing the suburban area into various notional submarkets that segregate the varieties of living space (Lefebvre, 1974 [1991]) in a more useful way. In this larger spatial scale too, very interesting inferences are drawn relating to spatial heterogeneity and interactions.

The spatial structure of the urban agglomeration of Aveiro is also prominent in the analysis of spatial interaction based on the estimated cross-submarket symmetric spatial weights matrix (Table 49). The first striking conclusion is that spatial interaction is significant for 17 out of 21 cells of the matrix. The main drivers of spatial interactions are common patterns of response to stochastic shocks; if for example, houses with particular characteristics (very big living rooms and terraces) become fashionable for given social groups, positive interactions between places with similar social structures are expected to be obtained and negative interactions for places where contrasting social groups dominate. On the other hand, temporary fashions affecting all types of houses are expected to generate an overall pattern of positive interaction. For example, if in a given year the size of the kitchen tends to increase in value, those houses sold in that year which have big kitchens will have positive error terms in all submarkets; conversely, houses with small kitchens will have negative error terms. Though this effect cannot be observed through time fixed effects, which only control for inflation, it creates a pattern of positive interaction in almost all cells. Does the pattern of spatial interactions reflect this general sensitivity to short term fashions? Do the few cases where negative interactions are detected reflect market segmentation? The development of such interpretation is beyond the scope of the thesis. What this thesis clearly shows is that the spatial interaction matrix, calculated according to the methodology presented above, combined with the analysis of spatial heterogeneity, provides a very rich set of information which can be the basis for detailed analysis and for disclosing the causes underlying the observed spatial patterns.

This highlights the fact that, with regard to the study of housing submarkets, a single scale may not always be adequate (Whitehead, 2003).

VII.6. Summary

The purpose of this chapter has been to improve the understanding of house price determination in a medium size city in Portugal. A hedonic pricing theory to estimate prices based on housing attributes in the city of Aveiro is used, focusing on the treatment of spatial heterogeneity and spatial dependence in the data. Several important findings emerge from this analysis.

First, there is substantial spatial heterogeneity across land units (independent of scale and number of housing submarkets) in terms of physical and location characteristics.

However, and second, this empirical study of the spatial error and spatial lag models, based on several standard specifications of distances and contiguity lead to the conclusion that spatial dependence is absent.

Third, evidence is found of important spatial dependence patterns, when spatial weights are allowed to be relatively free, and unrelated to traditional geographies based on distances and contiguity.

Fourth, important and interesting pricing patterns are identified by this study. Specifically, although location is important, property attributes are determinants of price formation. The square meter price of houses in Aveiro is determined by conventional variables: dimension of housing (total area, kitchen area and number of bedrooms), characteristics of housing (preservation, existence of a garage) and location (distance to the central business district). Single room houses are significantly more expensive per square meter in comparison with larger ones. Accessibility to the centre influences the price in a positive way as well. Further, important evidence of spatial heterogeneity is observed – effect of major factors, including total area and distance to the centre vary in interesting ways across the various subareas.

Fifth, the patterns of spatial interaction offer new insights into the nature of spatial diffusion of prices within the urban geography. The inferences drawn are important for related urban policy.

On the flip side, lack of data for some potentially important housing physical characteristics, such as quality and comfort attributes, somewhat limit the applicability and usefulness of our results. Further considerations of these and other related factors are retained for future work.

The empirical analysis presented presented in this fourth Part is a good test for the non-parametric spatial interaction model.

PART 5

Conclusions and Recommendations

VIII. Final remarks

The main findings of this thesis and its implications for urban housing policy are presented in Section VIII.1 following a brief summary of this research, highlighting the methodology and major challenges associated with the work. In Section VIII.2, the guidelines for further work are presented, looking to overcome some limitations found during this research, and giving some insight into new advances that can be made in the understanding of space in housing markets.

The main purpose of this research is to explore the role of space in urban housing markets. To achieve this main, goal a new framework, its corresponding methodology and an empirical application to an urban context is developed and presented. Three distinctive aspects for understanding the importance of space in the context of urban housing structures are considered: spatial heterogeneity, spatial dependence and spatial scale.

From the literature review (theoretical background stated in Part 2) the idea clearly emerged that: i) housing spatial patterns (heterogeneity, spillover and scale) may be driven by other intangible factors, and ii) objects (material and immaterial) deform space and change the action of the field of forces (extend, shrink or even annihilate distances), making complex territorial patterns. Therefore, the choice of methodologies based purely on bi-dimensional Euclidean space (which does not necessarily conform to our physical conception of space) might be inappropriate. In short:

i) Analysing urban growth, reflected by the sprawl of the settlements, there provides evidence that bi-dimensional Euclidean geometries might not be adequate to embrace the complexities associated with the new forms of urban structures.

ii) Analyzing the urban studies literature (urban economic, urban geography and urban planning) the fact is underlined that assumptions of bi-dimensional Euclidean space are not adequate to embrace these complexities, namely: in *urban economics* some unrealistic assumptions, about heterogeneity and interactions, have been adopted to make models feasible and comprehensible; in *urban geography* the perception that was revealed that many aspects and forms of reductionism and pure geometric understandings of space should be avoided, since space is considered to be socially produced; and finally, in *urban planning*, the apparent dematerialization of planning regained prominence, where, the physical substance of the notion of space has been lost over time.

iii) Analysing methodologies and techniques available in the field of spatial econometrics, it turns out that the usual approaches which impose a theoretical and *a priori* pattern of spatial interaction, might not correspond to reality when capturing the nature of spatial dependence. For instance, the uncertainty regarding the choice of the metric to measure spatial interactions (geographic distances, economic distances, socio-cultural distances and transport costs and time distances), being a key issue in determining spatial weights, is widely discussed in literature.

Thus, in line with what has been argued above, a methodology for understanding space in housing markets based on factor hedonic analysis and on the multi-dimensional non-Euclidean space is developed and applied to two different datasets.

One of the first steps of the empirical component of this dissertation was to collect data on housing sales for Aveiro and Ílhavo. The analysis considered a set of housing characteristics, physical and location, relevant to explaining the value of single-family property. The former are usually conditioned by their availability from real estate agencies, and the latter are built from the house position in space (precisely defined using property address or an approximate location using the zone where the property is located). This procedure seeks to identify urban amenities and disamenities which best represent the urban resident's preferences for housing. This type of data has been obtained from computer-based GIS maps. Once data are collected and compiled, the next step was to statistically estimate a function that relates property values to the property characteristics, aiming to investigate the households' level of preferences for specific types of housing facilities and amenities.

The difficulty of accessing appropriate information to characterise and to understand housing markets hinders an integrated (and at same time detailed) look at the diversification and specificities of housing supply and demand. Exhaustive databases and general models to analyse and describe the housing market are already available; however, the organization of information in decision support systems is still unsatisfactory. A good example is the dataset used in this research, provided by the real estate agency Casa Sapo. Created in 2001 by Janela Digital, it is the biggest portal for real estate advertisement at the national level; however, information is not organized to support policies and development strategies.

Among the huge variety of tools to deal with the issue of housing, taking various perspectives, are hedonic models. Based on revealed preference theory, these models take into account that the usefulness of a good derives from its properties or characteristics. When applied to housing, hedonic models consider that there is a set of characteristics, both physical (buildings and typological characteristics of the lot) or location (proximity effect to surrounding neighbourhoods and accessibility to goods and services) that can explain the asset value of a residential property. Thus, dwelling unit values (or proxies such as price or rents) are regressed on a bundle of characteristics of units that are most relevant in the explanation of the house price value. The results are a set of implicit prices for housing characteristics that are essentially willingness-to-pay estimates. There is another family of methods commonly used for the same purposes, known as stated preferences (not explored in this work). Through survey techniques an attempted is made to simulate hypothetical changes in the housing market and capture different individual perceptions and opinions about the aspects they value most in a house. The major advantage of using the hedonic models in this research, when compared to stated preferences, is that the former can be easily extended to spatial econometrics, where the most common aspects of space, heterogeneity and dependence, facilitate the understanding of residential location, and therefore urban structure, providing valuable input towards urban planning and housing policy.

As stated previously, a huge number of housing characteristics can be included on the right hand side of a hedonic regression equation. In this dissertation, a hybrid approach, combining factor analysis with regression, has been employed to obtain a small number of attributes, able to distinguish the main dimensions of housing characteristics. These predicted factor values are of crucial importance (apart from many other advantages explained throughout the dissertation) for estimating the unknown spatial weights matrices presented in Chapter IV.

VIII.1. Main findings and implications

Over the last decades the housing market has undergone profound transformations in societies, namely in terms of the dynamics of change in the demographic composition and social modification of the population. Factors such as population ageing, diversification of types of families, migration and, more generally, changes in lifestyles have led to new housing demands and requirements. It is thus expected that, to improve housing supply, urban policy makers, planners, economists, local tax appraisers and geographers, among other housing market agents, expend considerable effort in understanding the nature and characteristics of these changes, which require an effective knowledge of a complex set of phenomena: economic, social and territorial. Focusing particularly on housing spatial structures this thesis contributes to an analysis of the behaviour of the housing market, reflected by the willingness-to-pay for a set of housing attributes, where the location dimension assumes particular relevance. By comparing both spatial econometric results with those of GIS representations, new information is developed and presented, which can be crucial to understanding spatial urban housing market structures.

The analysis of space in the housing market being the main purpose of this dissertation, three distinct aspect of space have been considered: spatial heterogeneity, spatial dependence and spatial scale.

a) Main findings

This research contributes to fill a gap in an understanding of spatial aspects of the housing market and provide interesting insights into the relevance of multi-dimensional non-Euclidean notions of space to capture spatial interaction among housing submarkets.

In the literature, the difficulty in defining submarkets (heterogeneity) and understanding the relationship between them (spillovers) is broadly discussed, as well as methods to analyse these two aspects of spatial structure, however, there is no consensus as to the appropriate methodology and on the conceptual framework. As a contribution to understanding spatial structure in urban spaces, a new approach is presented, together with empirical analyses for the urban housing market of Aveiro, Portugal.

Regarding spatial heterogeneity three different approaches are used to define housing spatial segmentation, extensively discussed in the literature, namely: i) administrative boundaries, ii) expert knowledge, and iii) multivariate techniques (regression, factorial and cluster analysis) to aggregate units with similar regression coefficients in order to find groups where the principle of substitutability could be applied. The results show substantial spatial heterogeneity across housing submarkets, independent of the methodology used, that is, shadow prices and willingness-to-pay for different housing characteristics are different in the selected housing submarkets.

Regarding spatial dependence, a new methodology aiming to analyse spatial spillovers, assuming the notion of multi-dimensional non-Euclidean space, is developed and applied to the urban housing market of Aveiro. Rather than modelling spatial weights matrices as functions of any geographic or economic distance (being arbitrary predefined), it is assumed that spatial spillovers and interactions between observational units (housing submarkets) are unknown and can be estimated. It has been demonstrated that by not assuming any fixed and known pattern of spatial interactions (structure of spatial weights) it is possible to capture significant spatial spillovers, not detected by current traditional approaches. An interesting outcome of this methodology is the observation of positive and negative interactions that are meaningful in the specific spatial context, but are not always related to geographical distances or contiguity. Thus, an understanding of these potential drivers of spatial diffusion in residential value gives additional and useful information in understanding the role of space in urban housing markets.

In addition, the empirical work, using two distinct housing databases, applied to the urban housing market of Aveiro, highlights an important aspect of this analysis, that of spatial scale. Results show that research outcomes can vary significantly when spatial heterogeneity and spatial dependence are analysed at different spatial scales.

Despite the complexity and difficulties of analysing space, in the context of housing market, the fact that location is very important to determine the value of a property, both in terms of relative location (influence of the neighbourhoods) and absolute location (distances to certain elements, such as the city centre, beaches and local service centres) emerged from the empirical work.

a) Contributions and implications

The outcomes of this research are useful for housing urban policy for two main reasons: first, it allows an understanding of the way homeowners and renters bid for dwellings, emphasising the housing characteristics which are most important in explaining its value (relative or absolute terms). Comparing analyses across different attributes it is possible to capture how households assess several urban amenities (green areas, services, equipments, etc.), and therefore, the importance of space itself in the context of the bundle of housing characteristics. Secondly, it allows an understanding of the way that different spaces (housing submarkets) are connected together in a broader urban context. Because of the existence of spatial spillovers based on the multi-dimensional non-Euclidean notion of space (on unknown spatial weight matrix), it is important to keep in mind that any measures of regulations to stimulate or restrict the housing market activity, on a particular local scale, have impacts and interactions over space, not necessarily just in its geographical surroundings, but in a very complex multi-dimensional non-Euclidean spatial diffusion.

Depending of the type of housing agent the findings of this study can be useful in different settings. Five different categories of agents are considered:

i) Real estate agents and owners

The activity of these urban housing agents focuses essentially on an assessment of market conditions to maximize the sale value. Their concern is to ensure the highest volume of sales in the shortest possible time. To pursue their goals, both these intervenients, should be interested in the knowledge of household preferences (spatially distributed), in terms of which attributes should be emphasised when advertising a house, and the variability of housing price over time and over space. Thus, this information is useful for real estate agents and owners to support housing market prospection, activities of real estate mediation, and the definition of strategies to sell their properties.

ii) Promoters

This group includes, for example, construction companies, designers and building materials companies, which are interested in the investment value associated with different locations, that is, which housing physical attributes are critical to the value of housing in each geographic area. This assessment allows them, for example, to find more profitable construction solutions and investments that fulfil the wants of people. The analysis of spatial heterogeneity and spatial spillovers give valuable information about the possibility of alternative housing markets, when in a particular location the land price is extremely high or, simply, it is not possible to build.

iii) Public Institutions

Public institutions require information necessary for monitoring and defining strategies for territorial development. The outcomes produced during this research easily[144] extended to other spatial scales (metropolitan, regional or national level), and can be incorporated into different types of instruments of urban management and land policies to support, in an effective way, decision makers (not exclusively housing policy)[145] to implement territorial plans or other mechanisms of taxation policies. Public institutions responsible for housing policy development may wish to access multiple sets of indicators of the main

[144] It depends upon the data availability.
[145] For example, impact estimation (as a positive or negative externality) from investments in provision of some type of infrastructure in urban neighborhoods.

spatial and temporal housing market transformations (mapped geographically or not) to help decide prioritization of planning decisions. The identification of homogeneous housing areas to define pilot units for intervention and the possibility to assess the impact of a specific planning measure across space, is a paramount issue (e.g.: for policy planning purposes, an important issue concerns the estimation of the potential impacts, as positive or negative externalities, of any type of infrastructure).

iv) Users (Internet users)

Real-estate portal users are all interested in the acquisition of a property. A predictive tool providing confidence intervals for the value of property in each zone, depending on specific characteristics, and the evolution of house prices along a period of time provide useful information to someone that is searching for a house. Additionally, the results provided by spatial heterogeneity, and based on the principle of substitutability, offer alternative housing submarket solutions available in the urban context, not necessarily consistent with the generally perceived geographic notion of space.

VIII.2. Limitations and further work

The critical challenge for the success of this work is the ability to incorporate a set of attributes able to explain housing price and to distinguish the expected effects of spatial heterogeneity and dependence. The construction of a hedonic model has limitations at two levels: on one hand, the need to encompass a large set of attributes which describe the intrinsic housing characteristics, that cannot be more than that provided by real estate agents; on the other hand, the requirement that housing location analyses aspects of spatial differentiation, such as, socio-economic, urban design, and supply of equipment and services. Since the housing location is known (exact or approximate) additional information associated with urban characteristics of the environment can be used. The omission of relevant information at one of these levels has a considerable impact on the results, both in terms of explanatory power and consistency of the regression model.

Apart from the previously mentioned challenge, there are several limitations related to this research, essentially associated with data collection.

i) The explanation of the variability in housing prices across territorial units and over time, using hedonic models, may not be representative of the overall housing market, since they represent only a part of the transactions made in a given territory and in a given period. This inconsistency results from the fact that part of the transactions are completed without being advertised in any real estate agency (outside of real estate agent circuits), or some transactions may simply not be captured by the database analysed. But this point is only important if houses not included in these datasets have significant differences from those considered in the analyses. In the particular case of this research, and for the larger dataset, it is expected that, being a nationwide real estate agency the quantity of data available in such places and periods follows the housing and demographic dynamics.

ii) In this research, in the case of the larger dataset, the true transaction prices are not known; rather listing prices have been used. However, it is fair to assume that even if there is a significant difference between listing price and transaction price, it is expected to be relatively constant across all properties and would have no effect on any of the regression weights. Real estate agents estimate that the difference between asked price and the transaction price is about 10%. This estimation can be confirmed in future work.

iii) Distances to several urban amenities involved a high level of simplification. The distances and the measures of the potential (computed by a gravity model) were defined using a geographic information system, considering straight line distances, rather than the shortest street distance. In the particular case of this study this aspect is not problematic because the urban area of Aveiro and Ílhavo do not have strong accessibility constraints. However, it would be appropriate to test alternative measures of accessibilities, take into account, as much as possible, various traffic impedance and associated time costs, for instance.

iv) Quality and comfort of housing attributes were not included in the model because real estate agents usually do not collect this kind of information. Nevertheless, the use of dummy variables indicating the quality of the dwelling is important in explaining the values of housing. In this work the level of conservation (age and preservation) can give some insights into this particular housing dimension, but may be not sufficient. Thus, to improve the housing price model's explanatory power, as well as to avoid potential specification biases, a variety of quality and comfort attributes should be collected and incorporated into hedonic price regression models in future studies.

v) This research explored the impacts on the housing price of several socioeconomic variables at a level of disaggregation not consistent with the scale where urban phenomena impact. The results show that the impacts on housing price of several socioeconomic variables, considering parishes as the administrative level, were not significant. The boundaries of urban areas used in this approach are those determined by the census authorities and do not correspond with the local perception of a neighbourhood. To assess the effective impact on house prices, a more localised level for socioeconomic measurements should be considered. Even if it was possible to incorporate this statistical information into the housing database, only available from the census of 2001, it could be considered inappropriate to help represent the present urban context. In one decade, demographic, education and employment structures change significantly. Thus, the definition of zones should be consistent with the areas defined by the national statistical institute to incorporate external sources of information. If the exact position of housing is known, as was the case in the smaller dataset, the problem is partially solved and interesting information can be uncovered (e.g. walkability score).

vi) The limited size of the housing market, in the smaller dataset, only covering part of the territory (mainly suburban and rural areas) is another shortcoming of this work. A natural extension of this research would be data collection for a much larger and more extensive number of property sales including the entire urban area of Aveiro and Ílhavo. Despite the reduced number of observations included in this dataset, the fact that it includes the exact location of each house and the transaction price renders this database of relevant importance to this work.

vii) The duplication of cases without knowing if it is really a duplicate property or a dwelling in the same building, missing values and outliers are some other difficulties found in the data cleaning process. It is therefore crucial to incorporate in these larger databases methods and mechanisms to avoid this type of problems.

viii) To compute the autocovariance matrix and the spatial weight matrix using the methodology presented in this research (and based on Bhattacharjee and Jensen-Butler, 2005; Bhattacharjee and Holly, 2010; Bhattacharjee and Holly, 2011) matching residuals across submarkets are required. Thus, housing segmentation (and scale) is a key issue in the determination of spillover and should be investigated in more detailed way in future works. Different market segmentation approaches lead to very different price models.

ix) This work did not give particular attention to the analysis of the temporal aspect in housing market. The spatial econometric models presented and developed in this research proved to be useful for accurately describing house prices and the relevance of space in the preferences of households. However, the use of spatial econometric methodologies for forecasting exercises is affected by serious limitations in terms of data availability and stability of future trends. By contrast, some techniques developed in the social sciences analyse the future in a strategic perspective, but are not designed to directly produce results for decision support models. As has been mentioned in Chapter IV, such techniques can be divided into two main categories: i) scenario analysis, particularly helpful for the discussion and definition of strategies in uncertain situations, but not adequate to produce formal results; ii) Delphi surveys, which produce parameters, but which are not sensitive to exogenous contingent events. The research project DONUT (drivers of housing demand in the Portuguese urban system), funded by the Portuguese National Science Foundation, which has as its main goal the accurate description of the Portuguese housing market and the provision of a decision support tool to help policy makers anticipate and plan for future developments, has already started to develop and test methodologies which overcome some of the above limitations, by combining spatial econometric models with foresight techniques. Such methodologies are considered crucial for the construction of a decision support tool which will predict, with the highest possible accuracy, the factors determining property values. The importance of the capability to forecast future trends for effective housing and territorial planning policies is a key issue.

Despite the future work related with the limitation mentioned above, it is natural evolution to extend this work to a different urban context, with diverse characteristics, in terms of dimension and complexity, and to the entire national territory, where the estimation of spatial interaction between submarkets could provide an accurate description of the Portuguese urban system, to help decision makers.

Additionally, a network of researches across several countries and several scientific domains (planning, economic, geography, computational sciences, statistics, among others) allowing a interdisciplinary collaboration will bring substantial benefits for the understanding of housing phenomena from various points of view.

Several areas of future research can be suggested:

- Creation of a new database;
- Intelligent computational routines of georreferenciation;
- Spatial housing segmentation;
- Temporal housing markets;
- Historical evolution of the urban contexts.

Despite the significant developments in econometric techniques and, in particular, the techniques of spatial econometrics, there is still a vast potential for progress either in the techniques themselves, or in their application to the analysis of the housing market. Among several possible examples several deserve highlighting: i) the formulation of models to simultaneously examine spatial and temporal dependence; ii) the analysis of deterministic relations of spatial dependence, based on gravity models, combined with stochastic components of spatial lag or error; iii) analysis of the co-evolution of the housing market with a set of variables representing the economic, demographic and cultural dimensions. The combination of formal analytical methods with qualitative techniques of foresight is a field with potential to be explored, perhaps hindered by the undesirable division of social sciences, separating two groups with reduced willingness to co-operate: the world of subjectivity and qualitative thinking, and the world of objectivity of numbers and mathematical models, sometimes less objective than many would wish.

In parallel with developments in technical analysis, it is necessary to produce more and better information, detailed, reliable and georeferenced. As referred above, some of the existing information is fragmented and protected by entities that use it exclusively for their own ends, often unconscious of its vast potential for application. The interaction of the entities responsible for producing information with the researchers who add value to such information is increasingly necessary in a world where inter- and transdiciplinarity are decisive.

The end
of the beginning

References

Aase, T. (1994). *Symbolic Space: Representations of Space in Geography and Anthropology*. Geografiska Annaler, Series B, Human Geography **76**(1): 51-58

Abdel-Rahman, H. and Anas, A. (2004). *Theories of System of Cities*. *In* Handbook of Regional and Urban Economics. J. V. Henderson and T. Jacques-François, Elsevier. **4:** 2293-2339.

Abler, R., Adans, J. and Gould, P. (1977). *Spatial Organisation*, Prentice Hall, London.

Acharya, G. and Bennett, L. (2001). *Valuing Open Space and Land-Use Patterns in Urban Watersheds*. Journal of Real Estate Finance and Economics **22**(2-3): 221-237.

Adair, A., Berry, J. and McGreal, W. (1996). *Hedonic Modelling, Housing Submarkets and Residential Valuation*. Journal of Property Research **13**(1): 67-83.

Adair, A., McGreal, S., A., S., Cooper, J. and Ryley, T. (2000). *House Prices and Accessibility: The Testing of Relationships within the Belfast Urban Area*. Housing Studies **15**(5): 699-716.

Adams, J. (1984). *The Meaning of Housing Inamerica*. Annals of the Association of American Geographers **74**: 515-526.

Adams, J. (1987). *Housing America in the 1980's*, Russell Sage Foundation.

Agnew, G. (2005). *Space: Place*. *In* Space of Geographical Thought. P. Cloke and R. Johnston. London, SAGE Publication.

Akaike, H. (1974). *A New Look at Statistical Model Identification*. EEE Trans. Automatic Control **I**(19): 716-727.

Alden, I. and Morgan, R. (1974). *Regional Planning: A Comprehensive View*. London, Leonard Hills Books.

Alonso, W. (1964). *Location and Land Use*, Cambridge University Press.

Anas, A., Arnott, R. and Small, K. (1998). *Urban Spatial Structure*. Journal of Economic Literature **36**(3): 1426-1464.

Andrews, D. (2005). *Cross-Section Regression with Common Shocks*. Econometrica **73**: 1551-1585.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*, Dord-drecht: Kluwer Academic Publishers.

Anselin, L. (1995). *Local Indicators of Spatial Association – Lisa*. Geographical Analysis **27**: 93-115.

Anselin, L. (1999). *Spatial Econometrics*. *In* Companion to Theoretical Econometrics. B. Baltagi, Blackwell: Oxford**:** 310-330.

Anselin, L. (2002). *Under the Hood: Issues in the Specification and Interpretation of Spatial Regression Models*. Agricultural Economics **27**: 247-267.

Anselin, L. (2005). *Exploring Spatial Data with Geoda.Tm: A Workbook*.

Anselin, L. (2010). *Thirty Years of Spatial Econometrics*. Papers in Regional Science **89**: 3-25.

Anselin, L. and Bao, S. (1997). *Exploratory Spatial Data Analysis*. *In* Recent Developments in Spatial Analysis: Spatial Statistics, Behavioural Modelling and Computational Intelligence (Advances in Spatial Science). M. Fischer and A. Getis. Berlin, Springer**:** 35-60.

Anselin, L. and Bera, A. (1998). *Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics*. *In* Handbook of Applied Economic Statistics. A. Ullah and D. Giles. New York, Marcel Dekker**:** 237-289.

Anselin, L. and Florax, R. (1995). *New Directions in Spatial Econometrics*, Springer-Verlag, Berlin.
Anselin, L. and Griffith, D. (1988). *Do Spatial Effects Really Matter in Regression Analysis*. Papers, Regional Science Association **65**: 11-34.
Anselin, L. and Lozano-Gracia, N. (2008). *Spatial Hedonic Model*. GeoDa Center for Geospatial Analysis and Computation (Working Paper).
Anselin, L., Lozano-Gracia, N., Deichmann, U. and Lall, S. (2010). *Valuing Access to Water—a Spatial Hedonic Approach, with an Application to Bangalore, India*. Spatial Economic Analysis **5**(2): 161-179.
Anselin, L. and Rey, S. (2008). *Spatial Econometrics Foundations*.
Ara, S. (2007). *The Influence of Water Quality on the Demenad for Residential Development around Lake Erie*. *In* School of The Ohio State University, The Ohio State University. **PhD**.
Archer, C. and Wilkinson, R. (1973). *Measuring the Determinants of Relative House Prices*. Environment and Planning A **5**: 357-367.
Archibugi, F. (2008). *Planning Theory: From the Political Debate to the Methodological Reconstruction*, Springer.
Arnott, R. and McMillen, D. (2006). *A Companion to Urban Economics*, Blackwell Publishing.
Arnstein, S. (1969). *A Ladder of Citizen Participation*. Journal of the American Institute of Planners **35**: 216-224.
Azomahou, T. and Lahatte, A. (2000). *On the Inconsistency of the Ordinary Least Squares Estimator for Spatial Autoregressive Processes*. Working Paper BETA, Universite Louis Pasteur, Strasbourg, Department of Economics.
Backhouse, R. (2004). *Space in Economics: History of Economic Thought*. Universidade do Porto.
Bae, C., Jun, M. and Park, H. (2003). *The Impact of Seoul's Subway Line 5 on Residential Property Values*. Transport Policy **10**(2): 85-94.
Bailey, M., Muth, R. and Nourse, H. (1963). *A Regression Method for Real Estate Price Index Construction*. Journal of the American Statistical Association **48**(304): 933-942.
Bajic, V. (1985). *Housing Market Segmentation and Demand for Housing Attributes*
Journal of the American Real Estate and Urban Economics Association **13**: 58-75.
Baker, W. (2003). *Defining Quality of Life in Alabama: A Perceptual Community Based Definition for Local Leaders*. International Journal of Public Administration **26**(7): 733-751.
Banfield, E. (1959). *Ends and Means in Planning*. International Social Science Journal **XI**(3): 361-368.
Baranzini, A., Ramirez, J., Schaerer, C. and Thalmann, P. (2008). *Hedonic Methods in Housing Markets: Pricing Environmental Amenities and Segregation*, Springer.
Baranzini, A., Schaerer, C., Ramirez, J. and Thalmann, P. (2008). *Foreigners Pay Heigher Rents for the Same Quality of Housing in Geneve and Zurick*. Swiss Journal of Economics and Statistics (SJES) **144**(4): 703-730.
Basu, S. and Schroeder, R. (1977). *Incorporating Judgments in Sales Forecasts: Application of the Delphi Method at American Hoist and Derrick*. Interfaces **7**(3): 18-27.
Basu, S. and Thibodeau, T. (1998). *Analysis of Spatial Autocorrelation in House Prices*. The Journal of Real Estate Finance and Economics **17**(1): 61-85.
Batista, P., Castillo, G., Marques, J. and Castro, E. (2011). *Attribute Selection in Hedonic Pricing Modeling Applied to the Portuguese Urban Housing Market*. 15th Portuguese Conference on Artificial Intelligence; Lisbon, 10th to 13th October., Lisbon.
Batty, M. (1979). *On Planning Process*. *In* Resources and Planning. B. Goodall and A. Kirby. Oxford, Pergamon.
Beckmann, M. (1957). *On the Distribution of Rent and Residential Density in Cities*. *In* Paper presented at the Inter-Departmental Seminar on Mathematical Applications in the Social Sciences (Yale University, New Haven, CT).
Beckmann, M. and Thisse, J. (1987). *The Location of Production Activities*. *In* Handbook of Regional and Urban Economics. P. Nijkamp, Elsevier. **1:** 21-95.

Belbute, J., Marques, J., Gomes, C., Carvalho, J., Castro, E., Pinto, M., Gomes, P. and Pais, C. (2009). *Qual O Valor Da Qualidade De Vida? Um Contributo E Várias Conjecturas Metodológicas*. 15.º Congresso da APDR, Praia, Cabo Verde.
Bhattacharjee, A., Castro, E. and Marques, J. (2012). *Spatial Interactions in Hedonic Pricing Models: The Urban Housing Market of Aveiro, Portugal* Spatial Economic Analysis **7**(1): 133-167.
Bhattacharjee, A. and Holly, S. (2009). *Understanding Interactions in Social Networks and Committees*. Working Paper 1004, Centre for Dynamic Macroeconomic Analysis, University of St Andrews, UK.
Bhattacharjee, A. and Holly, S. (2010a). *Structural Interactions in Spatial Panels*. Centre for Dynamic Macroeconomic Analysis.
Bhattacharjee, A. and Holly, S. (2010b). *Understanding Interactions in Social Networks and Committees*. Faculty of Economics, University of Cambridge.
Bhattacharjee, A. and Holly, S. (2011). *Structural Interactions in Spatial Panels*. Empirical Economics **40**(Special issue on Panel Data Econometrics): 69-94.
Bhattacharjee, A. and Jensen-Butler, C. (2005). *Estimation of Spatial Weights Matrix in a Spatial Error Model, with an Application to Diffusion in Housing Demand*. Centre for Research into Industry, Enterprise, Finance and the Firm.
Bhattacharjee, A. and Jensen-Butler, C. (2011). *Estimation of the Spatial Weights Matrix under Structural Constraints*. Economic Studies Discussion Paper 254, Scottish Institute for Research in Economics, University of Dundee, UK.
Biagi, B., Lambiri, D. and Royuela, V. (2006). *Quality of Life in the Economic and Urban Economic Literature*. Working Paper CRENoS from Centre for North South Economic Research. University of Cagliari and Sassari, Sardinia.
Born, B. and Breitung, J. (2009). *Simple Regression Based Tests for Spatial Dependence*. Bonn Economics Discussion Papers, University of Bonn.
Bourassa, S., Cantoni, E. and Hoesli, M. (2007). *Spatial Dependence, Housing Submarkets, and House Price Prediction*. The Journal of Real Estate Finance and Economics **35**(2): 143-160.
Bourassa, S., Cantoni, E. and Hoesli, M. (2008). *Predicting House Prices with Spatial Dependence: Impacts of Alternative Submarket Definitions*. Swiss Finance Institute.
Bourassa, S., Hamelink, F., Hoesli, M. and MacGregor, B. (1999). *Defining Housing Submarkets*. Journal of Housing Economics **8**(2): 160-183.
Bourassa, S., Hoesli, M. and Peng, V. (2003). *Do Housing Submarkets Really Matter?* Journal of Housing Economics **12**(1): 12-28.
Bourassa, S., Hoesli, M. and Sun, J. (2006). *A Simple Alternative House Price Index Method*. Journal of Housing Economics **15**(1): 80-97.
Bourne, L. (1981). *The Geography of Housing*, Edward Arnold Ltd.
Box, G. and Cox, D. (1964). *An Analysis of Transformations*. Journal of the American Statistical Association **26**: 211-252.
Box, G. and Jenkins, G. (1976). *Time Series Analysis: Forecasting and Control*.
Boyle, K. and Kiel, K. (2001). *A Survey of House Price Hedonic Studies of the Impact of Environmental Externalities*. Journal of Real Estate Literature **9**(2): 117-144.
Bramley, G., Bartlett, W. and Lambert, C. (1995). *Planning, the Market and Private Housebuilding*, London, UCL Press Ltd.
Brasington, D. (2003). *The Supply of Public School Quality*. Economics of Education Review **22**(4): 367-377.
Brasington, D. and Hite, D. (2005). *Demand for Environmental Quality: A Spatial Hedonic Analysis*. Regional Science and Urban Economics **35**(1): 57-82.
Brown, M. (1968). *The Time Element in Planning*. Journal of the Town Planning Institute **54**: 109-213.
Brueckner, J. (2000). *Urban Sprawl: Diagnosis and Remedies*. International Regional Science Review **23**: 160-171.
Bruegmann, R. (2005). *Sprawl: A Compact History*, The University of Chicago Press.
Buchanan, C. (1963). *Traffic in Towns*. London, Penguin, Harmondsworth.
Burchell, R. (2000). *The Cost of Sprawl Revisited*, Rutgers University Press.
Burchell, R., Downs, A., McCann, B. and Mukherji, S. (2005). *Sprawl Costs: Economic Impacts of Uncheaked Development*, Island Press, USA.

Burke, T. and Zakharov, R. (2005). *Long-Term Housing Futures for Australia: Using Foresight to Explore Alternative Visions and Choices; Australian Housing and Urban Research*. Institute Swinburne-Monash Research Centre
Butler, R. (1982). *The Specification of Hedonic Indexes for Urban Housing*. Land Economics **58**(1): 96-108.
Camerer, C. and Loewenstein, G. (2004). *Behavioral Economics: Past, Present, Future*. *In* Advances in Behavioral Economics. C. Camerer, G. Loewenstein and M. Rabin, Princeton University Press.
Camerer, C., Loewenstein, G. and Rabin, M. (2004). *Advances in Behavioral Economics*, Princeton University Press.
Camhis, C. (1979). *Planning Theory and Philosophy*, Tavistock Publication.
Campbell, S. and Fainstein, S. (2003). *Readings in Planning Theory*, Wiley Blackwell.
Campbell, W. (1982). *Towards a Conservative Economics*. ModernAge **26**(1).
Can, A. and Megbolugbe, I. (1997). *Spatial Dependence and House Price Index Construction*. The Journal of Real Estate Finance and Economics **14**(1): 203-222.
Cardoso, L., Abiko, A., Haga, H., Inouye, K. and Gonçalves, O. (2005). *Future Prospective and Delphi Method: An Application in the House Building Supply Chain*. A. N. d. T. d. A. Construído.
Carruthers, J. and Gudmundur, F. (2002). *Fragmentation and Sprawl: Evidence from Interregional Analysis*. Growth and Change **33**: 312-340.
Case, B., Pollakowksi, H. and Wachter, S. (1991). *On Choosing among House Price Index Methodologies*. AREUEA Journal **19**(3): 286.
Cassel, E. and Mendelsohn, R. (1985). *The Choice of Functional Forms for Hedonic Price Equations: Comment*. Journal of Urban Economics **18**(2): 135-142.
Cassirer, E. (1944). *An Essay on Man: An Introduction to a Philosophy of Human Culture*, Yale University Press.
Castells, M. (1977). *The Urban Question: A Marxist Approach*. London, Edward Arnold.
Castells, M. (1983). *The City and the Grassroots*. Berkeley, University of California Press.
Castro, E., Marques, J. and Batista, P. (2011). *Estimação De Uma Superfície Hedónica De Preços Para O Mercado Habitacional Em Portugal*. 17º Conferência da APDR, Bragança/Zamora.
Castro, E., Marques, J., Gomes, P., Batista, P., Gomes, C., Pais, C. and Carvalho, J. (Forthcoming-a). *Benefícios Associáveis a Unidades Territoriais De Base*. *In* Ocupação Dispersa Custos E Benefícios, À Escala Local. J. Carvalho, 248-285.
Castro, E., Marques, J., Gomes, P., Batista, P., Gomes, C., Pais, C. and Carvalho, J. (Forthcoming-b). *Custos E Benefícios Associáveis a Utb-Padrão*. *In* Ocupação Dispersa Custos E Benefícios, À Escala Local. J. Carvalho**:** 286-279.
Catalão, A. (2010). *Estudo Do Mercado Imobiliário De Aveiro*. *In* Secção Autónoma de Ciências Sociais, Jurídicas e Políticas, Universidade de Aveiro. **Mestrado**.
Cavailhès, J., Peeters, D., Sékeris, E. and Thisse, J. (2004). *The Periurban City: Why to Live between the Suburbs and the Countryside*. Regional Science and Urban Economics **34**(6): 681-703.
CBOD (2011). *Relatório Final Do Projecto "Ocupação Dispersa Custos E Benefícios, À Escala Local"*. University of Aveiro (PTDC/AUR/64086/2006).
CEC (1990). *Green Paper on the Urban Environment*. European Union.
CEC (1997). *The Eu Compendium of Spatial Planning Systems and Policies*. European Union.
Chatfield, C. (2000). *Time-Series Forecasting*, Chapman & Hall.
Cheshire, P. (1999). *Trends in Sizes and Structures of Urban Areas*. *In* Handbook of Regional and Urban Economics. E. S. Mills, Elsevier. **3:** 1339-1373.
Cheshire, P., Summers, A. and Lanfranco, S. (1999). *Urban Change in the United States and Western Europe: Comparative Analysis and Policy*. Washington, D.C., Urban Institute Press.
Christaller, W. (1933 [1966]). *Die Zentealen Orde in Suddeutschland [Central Places in Southern Germany]*, Prentice-Hall: Englewood Cliffs, NJ.
Clapp, J. and Wang, Y. (2006). *Defining Neighborhood Boundaries: Are Census Tracts Obsolete?* Journal of Urban Economics **59**(2): 259-284.
Clifford, N., Holloway, S., Rice, P. and Valentine, G. (2009). *Key Concepts in Geography*, SAGE Publications Ltd.

Cohen, J. and Cohen, P. (1983). *Applied Multiple Regression: Correlation Analysis for the Behavioral Sciences*. Hillsdale, NJ, Erlbaum Associates.
Comission, E. (2002). *Hedonic Housing Prices without Characteristics: The Case of Case of New Multiunit House*. E. C. B.-E. Working Paper 117.
Commandeur, J. and Koopman, S. (2007). *An Introduction to State Space Time Series Analysis*, Oxford University Pess.
Committee, N. P. (2008). *Paul Krugman: Trade and Geography - Economies of Scale, Differentiated Products and Transport Costs*. Nobel Prize Committee
Correia-da-Silva, J. (2004). *Space in Economics – a Historical Perspective*. Universidade do Porto.
Costa, J. (2010). *Mercado Fundiário E Habitação*. *In* Compendio De Economia Regional: Teorias, Temáticas E Políticas (Volume 1). J. Costa, T. Dentinho and P. Nijkamp, Principia**:** 607-637.
Couclelis, H. (1992). *Location, Place, Region, and Space*. *In* Geography's Inner Worlds: Pervasive Themes in Contemporary American Geography. R. Abler, M. Marcus and J. Olson, New Brunswick: Rutgers, The State University of New Jersey Press
Crang, M. and Thrift, N. (2000a). *Introduction*. *In* Thinking Space. M. Crang and N. Thrift. London, Routledge.
Crang, M. and Thrift, N. (2000b). *Thinking Space*. London, Routledge.
Cropper, M., Leland, B. and McConnell, K. (1988). *On the Choice of Funtional Form for Hedonic Price Functions*. The Review of Economics and Statistics **70**(4): 668-675.
Cuhls, K. (2001). *Foresight with Delphi Surveys in Japan*. Technology Analysis & Strategic Management **13**(41).
Curry, M. (1996). *On Space and Spatial Practice in Contemporary Geography*. *In* Concepts in Human Geography. C. Early, K. Mathewson and M. Kenzer, Rowman and Littefield**:** 3-32.
Davidoff (2003 [1965]). *Advocacy and Pluralism in Planning*. *In* Readings in Planning Theory. S. Campbell and S. Fainstein, Blackwell Publishing Ltd.
Davies, G. (1974). *An Econometric Analysis of Residential Amenity*. Urban Studies **11**(2): 217-226.
Davoudi, S. (2002). *Polycentricity: What Does It Mean and How Is It Interpreted in the Esdp? In* Paper presented at the EURA Conference Urban and Spatial European Policies: Levels of Territorial Government. Turin 18-20 April 200.
Davoudi, S. (2006). *Evidence-Based Planning: Rhetoric or Reality?* disP 165 **42**(2): 14-24.
Davoudi, S. and Strange, I. (2009). *Space and Place in Twentieth Century Planning*. *In* Conceptions of Space and Place in Strategic Spatial Planning. S. Davoudi and I. Strange. London, Routledge (RTPI Library Series).
Deaton, B. and Hoehn, J. (2004). *Hedonic Analysis of Hazardous Waste Sites in the Presence of Other Urban Disamenities*. Environmental Science & Policy **7**(6): 499-508.
Descombes, V. (1994). *Is There an Objective Spirit*. *In* Philosophy in an Age of Pluralism: The Philosophy of Charles Taylor in Question J. Tully. Cambridge, Cambridge University Press**:** 97-118.
Dicken, P. and Lloyd, P. (1990). *Location in Space: Theoretical Perspectives in Economic Geography*, Prentice Hall.
Dimendberg, E. (1998). *Henri Lefebvre on Abstract Space*. *In* Philosophy and Geography Ii: The Production of Public Space. A. Light and J. Smith, Rowman and Littlefield Publishers, Inc.
DiPasquale, D. (1999). *Why Don't We Know More About Housing Supply?* Journal of Real Estate Finance and Economics **18**(1): 9-24.
Dixit, A. K. and Stiglitz, J. (1977). *Monopolistic Competition and Optimum Product Diversity*. American Economic Review **67**: 297-308.
Dodgson, J. and Topham, N. (1990). *Valuing Residential Properties with the Hedonic Method*. Housing Studies **5**(3): 209-213.
Duany, A., Plater-Zyberk, E. and Speck, J. (2000). *Suburban Nation: The Rise and the Decline of the American Dream*. New York, North Point Press.
Earl, D., Julia, L., L., A. and Greg, T. (1998). *Pricing Residential Amenities: The Value of a View*. The Journal of Real Estate Finance and Economics **16**(1): 55-73.

ECB (2002). *Hedonic Housing Prices without Characteristics: The Case of Case of New Multiunit House*. Working Paper n.º117. European Central Bank.
Edel, M. (1992). *Urban and Regional Economics-Marxist Perspectives*, Harwood Academic Publishers.
Elden, S. (2004). *Understanding Henri Lefebvre: Theory and the Possible*, Continuum.
Elhorst, J. P. (2010). *Applied Spatial Econometrics: Raising the Bar*. Spatial Economic Analysis **5**(1): 9-28.
ESDP (1999). *Towards Balanced and Sustainable Development of the Territory of the European Union*. E. S. D. Perspective:, European Commission.
Ewing, R. (1997). *Is Los Angeles-Style Sprawl Desirable?* Journal of the American Planning Association of Amercian Geographers **63**(1): 107-126.
Fahey, L. and Randall, R. (1998). *What Is Scenario Learning? In* Learning from the Future: Competitive Foresight Scenarios. L. a. R. Fahey, R. New York, John Wiley.
Falk, T. and Abler, R. (1980). *Intercommunications, Distance, and Geographical Theory*. Geografiska Annaler **B 62B**: 59-67.
Faludi, A. (1973a). *Planning Theory*. Oxford, Pergamon Press.
Faludi, A. (1973b). *A Reader in Planning Theory*. Oxford, Pergamon Press.
Faludi, A. (1973c). *What Is Planning Theory. In* A Reader in Planning Theory. A. Faludi. Oxford, Pergamon Press.
Fingleton, B. (2003). *Externalities, Economic Geography and Spatial Econometrics: Conceptual and Modeling Developments*. International Regional Science Review **26**(197-207).
Fingleton, B. (2008). *Housing Supply, Housing Demand, and Affordability*. Urban Study **45**(8): 1545-1563.
Follain, J. and Malpezzi, S. (1981). *Another Look at Racial Differences in Housing Prices*. Urban Studies **18**(2): 195-203.
Follain, J. R. and Malpezzi, S. (1980). *Dissecting Housing Value and Rent*. D. C. The Urban Institute: Washington.
Forrest, D., Glen, J. and Ward, R. (1996). *The Impact of a Light Rail System on the Structure of House Prices*. Journal of Transportation Economics and Policy **30**: 15-27.
Freeman, A. (1993). *The Measurement of Environmental and Resource Values: Theory and Methods*. Washington, DC, Resources for the Future.
Friedman, M. (1997). *Improving the Quality of Life: A Holistic Scientific Strategy*. New York, Praeger.
Friedmann, J. (1969). *Notes on Societal Action*. Journal of the American Institute of Planners **35**: 311-318.
Friedmann, J. (1987). *Planning in the Public Domain: From Knowledge to Action*, Princeton University Press.
Friedmann, J. (1993). *Toward a Non-Euclidean Mode of Planning*. Journal of the American Planning Association **59**(4): 482-485.
Fujita, M. (1988). *A Monopolistic Competition Model of Spatial Agglomeration: A Differentiated Product Approach*. Regional Science and Urban Economics **18**: 87-124.
Fujita, M. (1989). *Urban Economic Theory: Land Use and City Size*, Cambridge University Press.
Fujita, M. (1999). *Location and Space-Economy at Half a Century: Revisiting Professor Isard's Dream on the General Theory*. The Annals of Regional Science **33**(4): 371-381.
Fujita, M. (2010). *The Evolution of Spatial Economics: From Thünen to the New Economic Geography*. Japanese Economic Review **61**(1): 1-32.
Fujita, M. and Krugman, P. (1995). *When Is the Economy Monocentric? Von Thünen and Chamberlin Unified*. Regional Science and Urban Economics **25**: 265.
Fujita, M. and Krugman, P. (2003). *The New Economic Geography: Past, Present and the Future*. Economics of Governance **83**(1): 139-164.
Fujita, M., Krugman, P. and Venables, A. (1999). *The Spatial Economy: Cities, Regions, and International Trade*, The MIT Press.

Fujita, M. and Ogawa, H. (1982). *Multiple Equilibria and Structural Transition of Non-Monocentric Urban Configurations*. Regional Science and Urban Economics **12**(161-196).
Fujita, M. and Thisse, J. (2002). *Economic of Agglomeration: Cities, Industrial Location, and Regional Growth*, Cambridge University Press.
Galloway, T. G. and Mahayni, R. G. (1977). *Planning Theory in Retrospect: The Process of Paradigm Change*. Journal of the American Institute of Planners **43**(1): 62-71.
Galster, G. (1987). *Residential Segregation and Interracial Economic Disparities: A Simultaneous-Equations Approach*. Journal of Urban Economics **21**(1): 22-44.
Galster, G. (1992). *Research on Discrimination in Housing and Mortgage Markets: Assessment and Future Directions*. Housing Policy Debate **3**(2): 639-684.
Galster, G. (2001). *On the Nature of Neighbourhood*. Urban Studies **38**(12): 2111-2124.
Galster, G., Hanson, R., Ratcliffe, M., Wolman, H., Coleman, S. and Freihage, J. (2001). *Wrestling Sprawl to the Ground: Defining and Measuring an Elusive Concept*. Housing Policy Debate **12**(4): 681-717.
Gatzlaff, D. and Haurin, D. (1998). *Sample Selection and Biases in Local House Value Indices*. Journal of Urban Economics **43**(2): 199-222.
Gayer, T., Hamilton, J. T. and Viscusi, W. K. (2002). *The Market Value of Reducing Cancer Risk: Hedonic Housing Prices with Changing Information*. Southern Economic Journal **69**(2): 266-289.
Gibberd, F. (1967). *Town Design*, Architectural Press.
Giddens, A. (1994). *Beyond Left and Right: The Future of Radical Politics*. Oxford, Polity Press.
Gieryn, T. F. (2000). *A Space for Place in Sociology*. Annual Review of Sociology **26**(1): 463-496.
Gillen, K., Thibodeau, T. and Wachter, S. (2001). *Anisotropic Autocorrelation in House Prices*. Journal of Real Estate Finance and Economics **23**(1): 5-30.
Glaeser, E., Kallal, H., Scheinkman, J. and Shleifer, A. (1992). *Growth in Cities*. Journal of Political Economy **100**: 1126-1152.
Godet, M. (1991). *De L'anticipation À L'action. Manuel De Prospective Et De Stratégie*. Paris, Dunot.
Godet, M. (1997). *Manual De Prospectiva Estratégica: Da Antecipação À Acção*. Lisboa.
Goetzmann, W. and Spiegel, M. (1997). *A Spatial Model of Housing Returns and Neighborhood Substitutability*. Yale School of Management.
Golledge, R. G. and Hubert, L. J. (1982). *Some Comments on Non-Euclidean Mental Maps*. Environment and Planning A **14**(1): 107-118.
Gomes, C., Marques, J., Pinto, M., Castro, E., Gomes, P., Gomes, G. and Belbute, J. (2008). *A Multidimensionalidade Do Conceito De Qualidade De Vida". Sociedade E Território 42:* . Sociedade e Território **42**: 137-145.
Goodman, A. (1978). *Hedonic Prices, Price Indices and Housing Markets*. Journal of Urban Economics **5**: 471-484.
Goodman, A. (1981). *Housing Submarkets within Urban Areas: Definitions and Evidence.* Journal of Regional Science **21**: 175-185.
Goodman, A. and Thibodeau, G. (2007a). *The Spatial Proximity of Metropolitan Area Housing Submarkets*. Real Estate Economics **35**(2): 209-232.
Goodman, A. and Thibodeau, T. (1998). *Housing Market Segmentation*. Journal of Housing Economics **7**(2): 121-143.
Goodman, A. C. (1977). *Comparison of Block Group and Census Tract Data in a Hedonic Housing Price Model*. Land Economics **53**: 483-487.
Goodman, A. C. and Thibodeau, T. G. (2007b). *The Spatial Proximity of Metropolitan Area Housing Submarkets*. Real Estate Economics **35**(2): 209-232.
Goodman, C. and Thibodeau, G. (2003). *Housing Market Segmentation and Hedonic Prediction Accuracy*. Journal of Housing Economics **12**(3): 181-201.
Gordeon, T. and Pease, A. (2006). *Rt Delphi: An Efficient, "Round-Less" Almost Real Time Delphi Method*. Technological Forecasting & Social Change **73**: 321-333.
Gordon, P. and Harry, R. (1997). *Are Compact Cities a Desirable Planning Goal?* Journal of the American Planning Association **Winter**: 95-106.
Graham, S. and Healey, P. (1999). *Relational Concepts of Space and Place: Issues for Planning Theory and Practice*. European Planning Studies **7**(5): 623-646.

Grether, D. and Mieszkowski, P. (1974). *Determinants of Real Estate Values*. Journal of Urban Economics **1**: 127-146.
Griffith, D. (1995). *Some Guidelines for Specifying the Geographical Weights Matrix Contained in Spatial Statistical Models*.
Grigsby, W. (1963). *Housing Markets and Public Policy*, Philadelphia: University of Pennsylvania Press.
Grigsby, W., Baratz, M., Galster, G. and Maclennan, D. (1987). *The Dynamics of Neighborhood Change and Decline*. Oxford.
Guimarães, R. (2004). *Índice De Preços Hedónicos No Mercado Habitacional: Análise Exploratória No Quadro Do Sistema Estatístico Nacional*. *In* Faculdade de Economia, Universidade do Porto. **Mestrado**.
Haggett, P. (1965). *Locational Analysis in Human Geography*. London, Arnold.
Haggett, P. (2001). *Geography: A Global Synthesis*. Harlow, Prentice Hall.
Hair, J., Black, B., Babin, B., Anderson, R. and Tatham, R. (2010). *Multivariate Data Analysis*, Pearson Prentice Hall.
Hall, J. (2009). *A Guide to Doing Statistics in Second Language Research Using Spss* Taylor & Fancis.
Hall, P. (1966). *The World Cities*. London.
Hall, P. (1975). *Urban and Regional Planning*, Harmondsworth, Penguin.
Hall, P. (2002). *Cities of Tomorrow*, Blackwell Publishing.
Hårsman, B. and Quigley, J. (1995). *The Spatial Segregation of Ethnic and Demographic Groups: Comparative Evidence from Stockholm and San Francisco*. Journal of Urban Economics **37**(1): 1-16.
Hartshorne, R. (1959). *Perspective on the Nature of Geography*. Chicago, Rand McNally.
Harvey, D. (1969). *Justice, Nature and the Geography of Difference*. Oxford, Blackwell.
Harvey, D. (1973). *Social Justice and the City*, University of Georgia Press.
Harvey, D. (1990). *Between Space and Time: Reflections on the Geographical Imagination*. Annals of the Association of American Geographers **80**(3): 418-434.
Harvey, D. (1996). *Explanation in Geography*. London, Edward Arnold.
Harvey, D. (2000). *Spaces of Hope*. Edinburgh, Edinburgh University Press.
Harvey, D. (2006). *Space as a Keyword*. *In* David Harvey: A Critical Reader N. Castree and D. Gregory, Oxford: Blackwell**:** 270-293.
Healey, P. (1998). *Institutionalist Theory, Social Exclusion and Governance*. *In* Social Exclusion in European Cities. Processes, Experiences and Responses. A. Madanipour and G. Car, Regional Policy and Development, Regional Studie.
Henderson, G. and Sheppard, S. (2006). *Marx and the Spirit of Marx*. *In* Approaches to Human Geography. S. Aitken and G. Valentin, SAGE Publication.
Henderson, J., Kuncoro, A. and Turner, M. (1995). *Industrial Development in Cities*. Journal of Political Economy **103**(4): 1067-1090.
Henderson, J. V. (1974). *The Sizes and Types of Cities*. American Economic Review **64**: 640-656.
Hightower, H. (1969). *Planning Theory in Contemporary Professional Education*. Journal of the American Institute of Planners **35**(5): 326-329.
Hiller, J. and Healy, P. (2008). *Critical Essays in Planning Theory: Foundations of the Planning Enterprise*, Aldershot
Hirt, S. (2005). *Toward Postmodern Urbanism? Evolution of Planning in Cleveland, Ohio*. Journal of Planning Education and Research **25**(1): 27-42.
Hollis, M. (2003). *The Philosophy of Social Science: An Introduction*, Cambridge University Press.
Holt-Jensen, A. (1999). *Geography, History and Concepts*. London, Sage Publications Limited.
Hoover, E. and Giarratani, F. (1999). *An Introduction to Regional Economics*, Regional Research Institute.
Hotelling, H. (1929). *Stability in Competition*. The Economic Journal **39**: 41-57.
Hubbard, P., Kitchin, R., Bartley, B. and D, F. (2002). *Thinking Geographically: Space, Theory, and Contemporary Human Geography*, Continuum.
Hubbard, P., Kitchin, R. and Valentine, G. (2004). *Key Thinkers on Space and Place* London, SAGE Publication.

Hwang, S. (2003). *Environmental Amenities and Disamenities, and Housing Prices: Using Gis Techniques*. United States - Texas, Texas A&M University.
Isard, W. (1949). *The General Theory of Location and Space-Economy*. The Quarterly Journal of Economics **63**(4): 476-506.
Isard, W. (1956). *Location and Space-Economy*, Technology Press of Massachusetts Institute of Technology
Jacobs, J. (1961). *The Death and Life of Great American Cities*. Harmondsworth, Penguin
Jacobs, J. (1969). *The Economy of Cities*. New York, Vintage.
Jenkins, P. H. S., H. Wang, Y. (2007). *Planning and Housing in the Rapidly Urbanising World*, Routledge, London & New York.
Jenks, M., Burton, E. and Williams, K. (1996). *The Compact City. A Sustainable Urban Form*, Oxford Brookes University Press.
Jennrich, R. (2001). *A Simple General Procedure for Orthogonal Rotation*. Psychometrika **66**: 289-306.
Johnson, B. and Marcovitch, J. (1994). *Uses and Applications of Technology Futures on National Development: The Brazilian Experience*. Technological Forecasting and Social Change **45**(I30): 2-29.
Joyce, P. (2003). *The Rule of Freedom: Liberalism and the Modern City*. London, Verso.
Kain, J. and Quigley, J. (1970). *Measuring the Value of Housing Quality*. Journal of the American Statistical Association **65**(330): 532-548.
Kaldor, N. (1935). *Market Imperfection and Excess Capacity*. Economica **2**: 35-50.
Keeble, L. (1952). *Principles and Practice of Town and Country Planning*. London, Estates Gazette.
Kelejian, H. and Robinson, D. (1995). *Spatial Correlation: A Suggested Alternative to the Autoregressive Model. In* New Directions in Spatial Econometrics. L. Anselin and R. Florax. Berlin, Springer**:** 75-95.
Kiel, K. and Zabel, J. (1996). *House Price Differentials in U.S. Cities: Household and Neighborhood Racial Effects*. Journal of Housing Economics **5**(2): 143-165.
Kiel, K. and Zabel, J. (1997). *Evaluating the Usefulness of the American Housing Survey for Creating House Price Indices*. The Journal of Real Estate Finance and Economics **14**(1): 189-202.
Kitchin, R. and Tate, N. (2000). *Conducting Research in Human Geography*. Harlow, Prentice Hall.
Krugman, P. (1991). *Increasing Returns and Economic Geography*. Journal of Political Economy **99**: 483-499.
Krugman, P. (1998). *What's New About the New Economic Geography?* Oxford Review Economic Policy **14**(2).
Krugman, P. (1999). *Development, Geography, and Economic Theory*, The MIT Press Cambridge, Massachusetts London, England.
Krugman, P. R. (1993). *First Nature, Second Nature, and Metropolitan Location*. Journal of Regional Science **33**(2): 129-144.
Kuhn, T. (1962). *The Structure of Scientific Revolutions*. Chicago, University of Chicago Press.
Kurzweil, E. (1980). *The Age of Structuralism: Lévi-Strauss to Foucault*, Columbia University Press.
Lacour, C. and Puissant, S. (2007). *Re-Urbanity: Urbanising the Rural and Ruralising the Urban*. Environment and Planning A **39**(3): 728-747.
Lancaster, K. J. (1966). *A New Approach to Consumer Theory*. Journal of Political Economy **74**(2): 132-157.
Launhardt, W. (1885 [1993]). *Mathematische Begründung Der Volkswirtschaftslehre [Mathematical Principles of Economics]*, Leipzig: B.G. Teubner.
Law, J. and Urry, J. (2004). *Enacting the Social*. Economy and Society **33**: 390–410.
Lefebvre, H. (1974 [1991]). *The Production of Space*, Blackwell Publishing.
Lefebvre, H. (1979). *Space: Social Product and Use Value. In* Critical Sociology: European Perspectives. Freiburg. New York, Irvington Publishers**:** 285-295.
Leishman, C. (2009). *Spatial Change and the Structure of Urban Housing Sub-Markets*. Housing Studies **24**(5): 563 - 585.
LeSage, J. (1999). *The Theory and Practice of Spatial Econometrics*.
LeSage, J. and Pace, R. (2009). *Introduction to Spatial Econometrics*.

Lévi-Strauss, L. (1947). *Les Structures Élémentaires De La Parenté*. Berlin, Walter de Gruyter GmbH & Co.
Lévi-Strauss, L. (1964). *La Pensée Sauvage*. Paris, Pion.
Lin, J. and Hwang, C. (2004). *Analysis of Property Prices before and after the Opening of the Taipei Subway System*. The Annals of Regional Science **38**(4): 687-704.
Linneman, P. (1981). *The Demand for Residence Site Characteristics*. Journal of Urban Economics **9**(2): 129-148.
Livingstone, D. (1992). *A Brief History of Geography. In* The Student's Companion to Geography. A. Rogers, H. Viles and A. Goudie. Oxford, Blackwell.
Loomis, J. and Walsh, R. (1997). *Recreation Economic Decisions*, Venture Publishing, State College.
Lösh, A. (1940 [1954]). *Die Räumliche Ordnung Der Wirtschaf [the Economics of Location]*, Yale University Press.
MacDonald, J. and McMillan, M. (2010). *Urban Economics and Real Estate - Theory and Policy, 2nd Edition*, Wiley-Blackwell.
Maclennan, D. (1977). *Some Thoughts on the Nature and Purpose of House Price Studies*. Urban Studies **14**(1): 59-71.
Maclennan, D., Munro, M. and Wood, G. (1989). *Housing Choice and the Structure of Urban Housing Markets. In* Between State and Market: Housing in the Post-Industrial Era. B. Turner, J. Kemeny and L. Lundqvist, COthenburg, Almquist and Hicksell.
Maclennan, D. and Tu, Y. (1996). *Economic Perspectives on the Structure of Local Housing Systems*. Housing Studies **11**(3): 387-406.
Malpezzi, S. (2003). *Hedonic Pricing Model: A Selective and Applied Review. In* Housing Economics and Public Policy. T. O'Sullivan and K. Gibb, Blackwell Science Ltd.
Malpezzi, S., Gregory, C. and Richard, G. (1998). *New Place-to-Place Housing Price Indexes for U.S. Metropolitan Areas, and Their Determinants: An Application of Housing Indicators* . Real Estate Economics **26**(2): 235-275.
Malpezzi, S. and Guo, W. (2001). *Measuring Sprawl: Alternative Measures of Urban Form in U.S. Metropolitan Areas*. T. C. f. U. L. E. Research, The University of Wisconsin.
Malpezzi, S. and Mayo, S. K. (1997). *Housing and Urban Development Indicators: A Good Idea Whose Time Has Returned*. Real Estate Economics **25**(1): 1-11.
Malpezzi, S., Ozanne, L. and Thibodeau, T. (1980). *Characteristic Prices of Housing in 59 Smsas*. The Urban Institute, USA.
Malpezzi, S. a. and Maclennan, D. (2001). *The Long-Run Price Elasticity of Supply of New Residential Construction in the United States and the United Kingdom.* . Journal of Housing Economics **10**(3): 278-306.
Mark, J. and Goldberg, M. (1988). *Multiple Regression Analysis and Mass Assessement: A Review of Issues*. The Appraisal Journal **56**: 89-109.
Marques, J. and Castro, E. (2007). *Avaliação Hedónica Da Ocupação Urbana Residencial: Uma Análise Empírica Aplicada a Um Centro Urbano*. 13º Congresso da APDR, Açores.
Marques, J. and Castro, E. (2010). *Modelação Do Mercado De Habitação. In* Desafios Emergentes Para O Desenvolvimento Regional. J. Viegas and T. Dentinho, Principia**:** 257-286.
Marques, J., Castro, E. and Bhattacharjee, A. (2009). *A Localização Urbana Na Valorização Residencial: Modelos De Autocorrelação Espacial*. 15º Conferência da APDR, Cabo Verde.
Marques, J., Castro, E. and Bhattacharjee, A. (Forthcoming). *Modelling Housing Urban Market. In* Emerging Challenges for Regional Development. R. Capello and T. Dentinho.
Marques, J., Castro, E., Martins, J., Marques, M., Esteves, A. and Simão, R. (2008). *Exercício De Prospectiva Para a Região Centro: Análise De Cenários E Questionário Delphi*. Revista de Estudos Regionais **19** 111-131.
Marshall, A. (1890). *Principles of Economics*. London, Macmillan.
Marx, K. (1859). *A Contribution to the Critique of Political Economy*, Cornell University Library, Digital Collection.
Massey, D. (1991). *A Global Sense of Place*. Marxism Today **June**: 24–29.
Massey, D. (1992). *Politics and Space/Time*. New Left Review **196**: 65-84.

Massey, D. (1993). *Power Geometry and a Progressive Sense of Place. In* Mapping the Futures. J. Bird, B. Curtis, T. Putnam, G. Robertson and L. Tickner. London, Routiedge.
Massey, D. (1995). *The Conceptualisation of Place. In* A Place in the World: Places, Cultures and Globalisation. D. Massey and P. Jess. Buckingham, Open University Press.
Massey, D. (2004). *Geographies of Responsibility.* Geografiska Annaler **86B**(1): 5-18.
Massey, D. (2005). *For Space*, SAGE Publications Ltd.
Mathur, S. (2008). *Impact of Transportation and Other Jurisdictional-Level Infrastructure and Services on Housing Prices.* Journal of Urban Planning and Development-Asce **134**(1): 32-41.
Mayer, C. and Somerville, C. T. (2000). *Land Use Regulation and New Construction.* Regional Science and Urban Economics **30**(6): 639-662.
Mayo, S. K. (1981). *Theory and Estimation in the Economics of Housing Demand.* Journal of Urban Economics **10**(1): 95-116.
McCann, P. (2002). *Industrial Location Economics*, Edward Elgar Publishing Limited.
McDonald, J. and McMillan, M. (2010). *Urban Economics and Real Estate - Theory and Policy, 2nd Edition*, Wiley-Blackwell.
McLoughlin, J. B. (1969). *Urban and Regional Planning: A Systems Approach.* London, Faber & Faber.
McMaster, R. and Watkins, C. (1999). *The Economics of Housing: The Need for a New Approach. In* PRRES/AsRES/IRES Conference. Kuala Lumpur.
Meen, G. (2001). *Modeling Spatial Housing Markets: Theory, Analysis and Policy*, Springer.
Meentemeyer, V. (1989). *Geographical Perspectives of Space, Time, and Scale.* Landscape Ecology **3**(2/4): 163-173.
Meese, R. and Wallace, N. (1991). *Non-Parametric Estimation of Dynamic Hedonic Price Models and the Construction of Residential Housing Price Indices.* AREUEA Journal **19**(3): 308-332.
Michaels, R. and Smith, V. (1990). *Market Segmentation and Valuing Amenities with Hedonic Models: The Case of Hazardous Waste Sites.* Journal of Urban Economics **28**(2): 223-242.
Mikelbank, B. (2005). *Be Careful What You Wish For: The House Price Impact of Investments in Transportation Infrastructure.* Urban Affairs Review **41**(1): 20-46.
Miller, D. (1978). *The Factor of Scale: Ecosystem, Landscape Mosaic and Region. In* Sourcebook on the Environment. K. Hammond. Chicago, University of Chicago Press.
Millington, A. (1994). *An Introduction to Property Valuation.* Glasgow.
Mills, E. (1967). *An Aggregative Model of Resource Allocation in a Metropolitan Area.* American Economic Review **57**(197-210).
Mills, E. (1972a). *Studies in the Structure of the Urban Economy*, Johns Hopkins University Press.
Mills, E. (1972b). *Urban Economics*, Glenview, Illinois: Scott Foresman.
Miron, J. (1984). *Spatial Autocorrelation in Regression Analysis: A Beginner Guide. In* Spatial Statistics and Models. G. Gaile and C. Willmott, D Reidel Publishing company.
Moran, P. (1948). *The Interpretation on Statistical Maps.* Journal of the Royal Statistical Society **B**(10): 243-251.
Moreira, M. (2000). *A Dinâmica Pública Local E O Valor Da Habitação: Uma Aplicação À Área Metropolitana Do Porto. In* Faculdade de Engenharia, Universidade do Porto. **Mestrado**.
Mori, T. (2006). *Monocentric Versus Polycentric Models in Urban Economics.* Kier Discussion paper series, Kyoto institute of economic research(Discussion paper n.º611).
Morton, T. (1977). *Factor Analysis, Multicollinearity and Regression Appraisal Models.* The Appraisal Journal **Oct.**: 578–588.
Mulligan, G., Franklin, R. and Esparza, A. (2002). *Housing Prices in Tucson.* Urban Geography **23**: 446-470.

Mundy, B. and McLean, D. (1998). *Using the Contingent Value Approach for Natural Resource and Environmental Damage Applications*. The Appraisal Journal **56**(3): 290-297.
Murdoch, J. (2006). *Post-Structuralist Geography: A Guide to Relational Space*. London, SAGE Publication.
Muth, R. F. (1961). *The Spatial Structure of the Housing Market*. Papers and Proceedings of the Regional Science Association **7**: 207-220.
Myers, C. (2004). *Discrimination and Neighborhood Effects: Understanding Racial Differentials in Us Housing Prices*. Journal of Urban Economics **56**(2): 279-302.
Myrdal, G. (1957). *Economic Theory and under-Developed Regions*. London, Gerald Duckworth.
Nelson, A., Duncan, J., Mullen, C. and Bishop, K. (1995). *Growth Management: Principles and Practices*, Planners Press, APA, Chicago.
Nelson, J. (2008). *Hedonic Property Value Sutdies of Transportation Noise: Aircraft and Road Traffic In* Hedonic Methods in Housing Markets: Pricing Environmental Amenities and Segregation. Springer**:** 300.
Newman, P. and Kenworthy, J. (1989). *Cities and Automobile Dependence: A Sourcebook*, Aldershot: Gower Technical.
Ohlin, B. (1933 [1968]). *Interregional and International Trade* Harvard University Press.
Olssen, M. (2003). *Structuralism, Post-Structuralism, and Neo-Liberalism: Assessing Foucault's* Journal of Education Policy **18**(2): 189-2002.
Osbourne, T. and Rose, N. (1999). *Governing Cities: Notes on the Spatialisation of Virtue*. Environment and Planning D: Society and Space **17**: 737-760.
Ottaviano, G. and Thisse, J. (2004). *Agglomeration and Economic Geography*. Elsevier. **4:** 2563-2608.
Ottaviano, G. and Thisse, J. (2005). *New Economic Geography: What About the N?* Environment and Planning A **37**(10): 1707-1725.
Ozanne, L. and Malpezzi, S. (1985). *The Efficacy of Hedonic Estimation with the Annual Housing Survey Evidence from the Demand Experiment*. Journal of Economic and Social Measurement **13**: 153-172.
Pace, R. (1993). *Nonparametric Methods with Applications to Hedonic Models*. The Journal of Real Estate Finance and Economics **7**(3): 185-204.
Pace, R. and LeSage, J. (2004). *Spatial Econometrics and Real Estate*. Journal of Real Estate Finance and Economics **29**: 147-148.
Pacione, M. (2005). *Urban Geography: A Global Perspective*, Routledge.
Paelinck, J. and Klaassen, L. (1979). *Spatial Econometrics;* , Saxon House Farnborough.
Palander, T. (1935). *Beitrage Zur Standortstheorie*. Uppsala, Almqvist and Wiksell.
Palm, R. (1978). *Spatial Segmentation of the Urban Housing Market*. Journal of Economic Geography **54**: 210–221.
Palmquist, R. (2002). *Hedonic Models*. *In* Handbook of Environmental and Resource Economics. J. Bergh, Edward Elgar Publishing.
Palmquist, R., Maler, K. and Vincent, J. (2005). *Property Value Models*. *In* Handbook of Environmental Economics, Elsevier. **2:** 763-819.
Pankratz, A. (1983). *Forecasting with Univariate Box-Jenkins Model*, John Wiley: New York.
Parker, S. (2004). *Urban Theory and the Urban Experience*, Routledge.
Pesaran, M. (2006). *Estimation and Inference in Large Heterogenous Panels with Multifactor Error Structure*. Econometrica **74**: 967-1012.
Pinho, M. (1992). *Funções Hedónicas De Preços Para O Mercado Da Habitação: Aplicações Às Cidades Do Porto E De Aveiro*. *In* Faculdade de Economia, Universidade do Porto. **Mestrado**.
Poincaré, H. (1895). *L'espace Et La Geometrie*. Revue de métaphysic et de morale: 631-646.
Pollakowski, H. (1995a). *Data Sources for Measuring House Price Changes*. Journal of Housing Research **6**(3): 377-387.
Pollakowski, H. (1995b). *Housing Price Diffusion Patterns at Different Aggregation Levels: An Examination of Housing Market Efficiency*. Journal of Housing Research **8**(1): 107-124.

Pollakowski, H. and Wachter, S. (1991). *On Choosing among House Price Index Methodologies*. AREUEA Journal **19**(3): 286.

Popper, K. (1959). *The Logic of Scientific Discovery*. New York, Basic Books

Quigley, J. (1979). *What Have We Learned About Housing Markets? In* Current Issues in Urban Economics. P. Mieszkowski, M. Straszheim and J. Hopkins, Press**:** 388-426.

Quigley, J. (1995). *A Simple Hybrid Model for Estimating Real Estate Price Indexes*. Journal of Housing Economics **4**(1): 1-12.

Rapkin, C., Winnick, L. and Blank, D. (1953). *Housing Market Analysis. A Study of Theory and Methods*, Housing and Home Finance Agency. Division of Housing Research.

Raudenbush, S. and Bryk, A. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods (Advanced Quantitative Techniques in the Social Sciences)*, Sage Publications Inc.

Redding, S. J. (2010). *The Empirics of New Economic Geography*. Journal of Regional Science **50**(1): 297-311.

Reis, S. (2011). *O Impacto Do Metro Do Porto Na Valorização Imobiliária. In* Faculdade de Economia, Universidade do Porto. **Mestrado**.

Ricardo, D. (1821). *On the Principles of Political Economy and Taxation*, Library of Economics and Liberty.

Ridker, R. and Henning, J. (1967). *The Determinants of Residential Property Values with Special Reference to Air Pollution*. Review of Economics and Statistics **49**(2): 246-257.

Rosen, S. (1974). *Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition*. The Journal of Political Economy **82**(1): 35-55.

Rosenthal, S. (1999). *Residential Buildings and the Cost of Construction: New Evidence from the Housing Market*. Review of Economics and Statistics **81**: 288-302.

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

Sack, R. (1986). *Human Territoriality: Its Theory and History*, Cambridge University Press.

Samuelson, P. (1952). *The Transfer Problem and Transport Costs: The Terms of Trade When Impediments Are Absent*. Economic Journal **62**: 278-304.

Samuelson, P. (1983). *Thünen at Two Hundred*. Journal of Economic Literature **21**: 1468–1488.

Sanchez, T. (1993). *Measuring the Impact of Highways and Roads Onresidential Property Values with Gis: An Empirical Analysis of Atlanta. In* Third International Conference on Computers. R. Klosterman and S. French, Urban Planning and Management.

Sandercock, L. (1998). *Towards Cosmopolis : Planning for Multicultural Cities*, Chichester.

Schaeffer, F. (1953). *Exceptionalism in Geography: A Methodology Examination*. Association of Amercian Geographers **43**(3): 226–249.

Schnare, A. B. (1980). *Trends in Residential Segregation by Race: 1960-1970*. Journal of Urban Economics **7**(3): 293-301.

Schnare, A. B. and Struyk, R. J. (1976). *Segmentation in Urban Housing Markets*. Journal of Urban Economics **3**(2): 146-166.

Scruton, R. (1981). *A Short History of Modern Philosophy: From Descartes to Wittgenstein*, Routledge.

Scruton, R. (1996). *An Intelligent Person's Guide to Philosophy*. London, Duckworth.

Shaeffer, F. (1953). *Exceptionalism in Geography: A Methodology Examination*. Association of Amercian Geographers **43**(3): 226–249.

Shields, R. (1997). *Spatial Stress and Resistance: Social Meanings and Spatialisation. In* Space and Social Theory. G. Benko and U. Strohmayer. Oxford, Blackwell.

Simon, G. (1997). *An Angular Version of Spatial Correlations, with Exact Significance Tests*. Geographical Analysis **29**(3): 267-278.

Smith, A. (1776). *The Wealth of Nations*, Random House, Inc.

Smith, B. C. (2000). *Applying Models for Vertical Inequity in the Property Tax to a Non-Market Value State*. Journal of Real Estate Research **19**(3): 321-344.

Smith, L. B., Rosen, K. T. and Fallis, G. (1988). *Recent Developments in Economic Models of Housing Markets*. Journal of Economic Literature **26**(1): 29-64.

Smith, N. (1990). *Uneven Development: Nature, Capital and the Production of Space*. Oxford, Basil Blackwell.
Smith, N. (1992). *Geography, Difference and the Politics of Scale*. *In* Postmodernism and the Social Sciences. J. Doherty, E. Graham and M. Malek, Macmillan, Basingstoke 57-79.
Soderberg, B. and Janssen, C. (2001). *Estimating Distance Gradients for Apartment Properties*. Urban Studies **38**(1): 61-79.
Soja, E. (1989). *Post-Modern Geographies: The Reassertion of Space in Critical Social Theory*. London, Verso.
Soja, E. (1996). *Thirdspace*. Oxford, Blackwell.
Song, S. (1995). *Determinants of Bargaining Outcome in Single-Family Housing Transaction: An Empirical Examination*. Urban Studies **32**(3): 605-614.
Spellman, F. (2010). *Geography for Nongeographers*, Government Institutes.
Storey, D. (2001). *Territory: The Claiming of Space*, Pearson Educaton Limited.
Straszheim, M. (1974). *Hedonic Estimation of Housing Market Prices: A Further Comment*. The Review of Economics and Statistics **56**(3): 404-406.
Straszheim, M. (1975). *An Econometric Analyis of the Urban Housing Market* N. B. o. E. Research, Cambridge MA.
Struyk, R., Malpezzi, S. and Wann, F. (1980). *The Performance Funding System Inflation Factor: Analysis of Predictive Ability of Candidate Series and Models*. Urban Institute Working Paper.
Stull, W. (1975). *Community Environment, Zoning, and the Market Value of Single-Family Homes*. Journal of Law and Economics **18**(2): 535-570.
Taylor, L. (2008). *Theoretical Foundations and Empirical Developments in Hedonic Modeling*. *In* Hedonic Methods in Housing Markets: Pricing Environmental Amenities and Segregation, Springer.
Taylor, N. (1998). *Urban Planning Theory since 1945*, SAGE Publications Ltd.
Taylor, N. (1999). *Anglo-American Town Planning Theory since 1945: Three Significant Developments but No Paradigm Shifts*. Planning Perspectives **14**(4): 327-345.
Thibodeau, T. (2003). *Marking Single-Family Property Values to Market*. Real Estate Economics **3**(1): 1-22.
Timmermans, H., Molin, E. and Noortwijk, L. (1994). *Housing Choice Processes: Stated Versus Revealed Modelling Approaches*. Journal of Housing and the Built Environment **9**(3): 215-227.
Tobler, W. (1970). *A Computer Movie Simulating Urban Growth in the Detroit Region*. Economic Geography **46**(2): 234-240.
Triplett, J. (2006). *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes: Special Application to Information Technology Products*, OECD.
Tyrvainen, L. (1997). *The Amenity Value of the Urban Forest: An Application of the Hedonic Pricing Method*. Landscape and Urban Planning **37**(3-4): 211-222.
Ugarte, M., Goicoa, T. and Miltino, A. (2004). *Searching for Housing Submarkets Using Mixtures of Linear Models*. *In* Spatial and Spatiotemporal Econometrics. Advances in Econometrics. J. LeSage and R. Pace, Oxford: Elsevier.
Valente, M. (2010). *Tributação Do Património E Determinantes Económicas No Sucesso Político Dos Governantes*. *In* Escola de Economia e Gestão, Universidade do Minho. **Doutoramento**.
Venables, A. (1996). *Equilibrium Locations of Vertically Linked Industries*. International Economic Review **37**: 341-359.
Venables, A. (2005). *New Economic Geography*. *In* Palgrave Dictionary of Economics.
von Thünen, J. (1826 [1966]). *The Isolated State [Der Isolierte Staat]*. New York, Pergamon Press.
Voros, J. (2003). *A Generic Foresight Process Framework*. Foresight **5**(3): 10-21.
Ward, S. (2004). *Planning and Urban Change*. London, SAGE Publications Lda.
Watkins, C. (2001). *The Definition and Identification of Housing Submarket*. Environment and Planning A **33**(12): 2235-2253.
Weber, A. (1909 [1957]). *Theory of Location of Industries [Über Den Standort Der Industrie]*, University of Chicago Press
Weber, A. (1957 [1909]). *Theory of Location of Industries [Über Den Standort Der Industrie]*, University of Chicago Press

West-Pavlov, R. (2009). *Space in Theory: Kristeva, Foucault, Deleuze*. Amsterdam, Rodopi B.V.
Wheaton, W. C. (1999). *Real Estate "Cycles": Some Fundamentals.* Real Estate Economics **27**(2): 209-230.
Whitehead, J., Pattanayak, S., Houtven, G. and Gelso, B. (2008). *Combining Revealed and Stated Preference Data to Estimate the Nonmarket Value of Ecological Services: An Assessment of the State of the Science*. Journal of Economic Surveys **22**(8): 872-908.
Whitehead, M. (2003). *Love Thy Neighbourhood -- Rethinking the Politics of Scale and Walsall's Struggle for Neighbourhood Democracy*. Environment and Planning A **35**(2): 277-300.
Williams, K. (1999). *Urban Intensification Policies in England: Problems and Contradictions* Land Use Policy **16**(3): 167-178.
Wingo, L. (1961). *Transportation and Urban Land* Resources for the Future, Washington, DC.
Wooldridge, J. (1999). *Introductory Economics: A Modern Approach*, South-Western College Publishing.
Yeates, N. (2001). *Globalisation and Social Policy*, Thousand Oaks, CA, Sage.
Young, I. (1990). *Justice and the Politics of Difference*. Princeton, Princeton University Press.
Zabel, J. (2004). *The Demand for Housing Services*. Journal of Housing Economics **13**: 16-35.

Other references:

Banco de Portugal (Bank of Portugal) - BP:
http://www.bportugal.pt

Econometrics Laboratory, University of California at Berkeley:
http://elsa.berkeley.edu/

European Central Bank - ECB:
http://www.ecb.int

Eurostat -Estatísticas da União Europeia:
http://epp.eurostat.ec.europa.eu/

Instituto Nacional de Estatística (Statistics Portugal) - INE
http://www.ine.pt

Spatial Econometrics Toolbox for MATLAB:
http://www.spatial-econometrics.com/

The GeoDa Center for Geospatial Analysis and Computation:
http://geodacenter.asu.edu/

U.S. Bureau of Labour Statistics (USBLS):
http://www.bls.gov/

APPENDICES

Statistical inferences: sample 1

Statistical inferences: sample 2

APPENDIX 1

Statistical inferences: sample1

(Selected SPSS outputs)

1. DIFFERENCES BETWEEN THE TERRITORIES

(6 parishes)

...AND SCALAR VARIABLES (Kuskal Wallis tests)

(kitchen area, living room area, price in euros/m^2, price in euros, total area, floors and number of bedrooms)

Table A1. 1 – Non-parametric Kruskal Wallis ranks: parishes vs. (kitchen area, living room area, price in euros/m^2): (cont.)→

	parishes_codes	N	Mean Rank
Kitchen area	Aradas	20	62.95
	Esgueira	36	80.22
	Glória	20	51.00
	Santa Joana	10	65.65
	São Bernardo	8	65.25
	Vera Cruz	45	75.21
	Total	139	
Living room area	Aradas	19	80.11
	Esgueira	38	83.83
	Glória	27	67.28
	Santa Joana	8	41.06
	São Bernardo	8	64.19
	Vera Cruz	47	74.72
	Total	147	
Price (€/m^2)	Aradas	21	95.90
	Esgueira	42	53.90
	Glória	28	90.34
	Santa Joana	11	83.27
	São Bernardo	9	54.44
	Vera Cruz	55	102.68
	Total	166	

Table A1. 2 – Non-parametric Kruskal Wallis ranks: parishes vs. (price in euros, total area, floors and number of bedrooms): ←(cont.)

	parishes_codes	N	Mean Rank
Price (€)	Aradas	21	68.38
	Esgueira	42	73.98
	Glória	28	81.98
	Santa Joana	11	65.00
	São Bernardo	9	71.61
	Vera Cruz	55	102.96
Total area	Aradas	21	61.26
	Esgueira	42	93.42
	Glória	28	79.04
	Santa Joana	11	72.95
	São Bernardo	9	91.89
	Vera Cruz	55	87.43
Floors	Aradas	21	70.36
	Esgueira	42	80.69
	Glória	28	87.41
	Santa Joana	11	66.95
	São Bernardo	9	78.61
	Vera Cruz	55	92.78
Number of bedrooms	Aradas	21	58.50
	Esgueira	41	84.22
	Glória	28	79.75
	Santa Joana	11	72.05
	São Bernardo	9	71.06
	Vera Cruz	55	97.25

Table A1. 3 –Non-parametric Kruskal Wallis tests: parishes vs. (kitchen area. living room area. price in euros/m2. price in euros. total area. floors and number of bedrooms)

	Kitchen area	Living room	Price (€/m^2)	Price (€)	Total area	Floors	N bedrooms
Chi-Square	8.369	8.318	29.939	14.962	7.741	6.176	13.510
Asymp. Sig.	0.137	0.140	0.000	0.011	0.171	0.289	0.019

...AND NOMINAL VARIABLES (chi-square tests)

(type of house, preservation, duplex, balcony, terrace, garage space, CATV and natural gas)

Table A1. 4 - Crosstab: parishes vs. type of house

			Type (House=1. Flat=0)		
			0	1	Total
parish	Aradas	Count	17	4	21
		Expected Count	18.2	2.8	21.0
	Esgueira	Count	33	9	42
		Expected Count	36.4	5.6	42.0
	Glória	Count	27	1	28
		Expected Count	24.3	3.7	28.0
	Santa Joana	Count	7	4	11
		Expected Count	9.5	1.5	11.0
	São Bernardo	Count	8	1	9
		Expected Count	7.8	1.2	9.0
	Vera Cruz	Count	52	3	55
		Expected Count	47.7	7.3	55.0
Total		Count	144	22	166
		Expected Count	144.0	22.0	166.0

Table A1. 5 – Non-parametric Chi-Square Test: parishes vs. type of house

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	13.394	5	0.020

Table A1. 6 - Crosstab: parishes vs. preservation

			Preservation (Used=1. New=0)		
			0	1	Total
parish	Aradas	Count	3	18	21
		Expected Count	2.4	18.6	21.0
	Esgueira	Count	6	35	41
		Expected Count	4.7	36.3	41.0
	Glória	Count	3	25	28
		Expected Count	3.2	24.8	28.0
	Santa Joana	Count	0	11	11
		Expected Count	1.3	9.7	11.0
	São Bernardo	Count	1	8	9
		Expected Count	1.0	8.0	9.0
	Vera Cruz	Count	6	49	55
		Expected Count	6.3	48.7	55.0
Total		Count	19	146	165
		Expected Count	19.0	146.0	165.0

Table A1. 7 – Non-parametric Chi-Square Test: parishes vs. preservation

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	2.020	5	0.846

Table A1. 8 - Crosstab: parishes vs. duplex

			Duplex		
			0	1	Total
parish	Aradas	Count	17	4	21
		Expected Count	18.2	2.8	21.0
	Esgueira	Count	33	9	42
		Expected Count	36.4	5.6	42.0
	Glória	Count	27	1	28
		Expected Count	24.3	3.7	28.0
	Santa Joana	Count	7	4	11
		Expected Count	9.5	1.5	11.0
	São Bernardo	Count	8	1	9
		Expected Count	7.8	1.2	9.0
	Vera Cruz	Count	52	3	55
		Expected Count	47.7	7.3	55.0
Total		Count	144	22	166
		Expected Count	144.0	22.0	166.0

Table A1. 9 – Non-parametric Chi-Square Test: parishes vs. duplex

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	13.394	5	0.020

Table A1. 10 - Crosstab: parishes vs. balcony

			Balcony		
			0	1	Total
parish	Aradas	Count	16	5	21
		Expected Count	17.0	4.0	21.0
	Esgueira	Count	26	16	42
		Expected Count	33.9	8.1	42.0
	Glória	Count	22	6	28
		Expected Count	22.6	5.4	28.0
	Santa Joana	Count	11	0	11
		Expected Count	8.9	2.1	11.0
	São Bernardo	Count	8	1	9
		Expected Count	7.3	1.7	9.0
	Vera Cruz	Count	51	4	55
		Expected Count	44.4	10.6	55.0
Total		Count	134	32	166
		Expected Count	134.0	32.0	166.0

Table A1. 11 – Non-parametric Chi-Square Test: parishes vs. balcony

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	18.024	5	0.003

Table A1. 12 - Crosstab: parishes vs. terrace

			Terrace		
			0	1	Total
parish	Aradas	Count	18	3	21
		Expected Count	18.8	2.2	21.0
	Esgueira	Count	39	3	42
		Expected Count	37.7	4.3	42.0
	Glória	Count	24	4	28
		Expected Count	25.1	2.9	28.0
	Santa Joana	Count	9	2	11
		Expected Count	9.9	1.1	11.0
	São Bernardo	Count	8	1	9
		Expected Count	8.1	.9	9.0
	Vera Cruz	Count	51	4	55
		Expected Count	49.4	5.6	55.0
Total		Count	149	17	166
		Expected Count	149.0	17.0	166.0

Table A1. 13 – Non-parametric Chi-Square Test: parishes vs. terrace

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	2.600	5	0.761

Table A1. 14 - Crosstab: parishes vs. garage space

			Provision of garage		
			0	1	Total
parish	Aradas	Count	8	13	21
		Expected Count	8.6	12.4	21.0
	Esgueira	Count	24	18	42
		Expected Count	17.2	24.8	42.0
	Glória	Count	13	15	28
		Expected Count	11.5	16.5	28.0
	Santa Joana	Count	3	8	11
		Expected Count	4.5	6.5	11.0
	São Bernardo	Count	3	6	9
		Expected Count	3.7	5.3	9.0
	Vera Cruz	Count	17	38	55
		Expected Count	22.5	32.5	55.0
Total		Count	68	98	166
		Expected Count	68.0	98.0	166.0

Table A1. 15 – Non-parametric Chi-Square Test: parishes vs. garage space

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	8.332	5	0.139

Table A1. 16 - Crosstab: parishes vs. cable TV

			CATV		
			0	1	Total
parish	Aradas	Count	15	6	21
		Expected Count	15.6	5.4	21.0
	Esgueira	Count	31	11	42
		Expected Count	31.1	10.9	42.0
	Glória	Count	17	11	28
		Expected Count	20.7	7.3	28.0
	Santa Joana	Count	10	1	11
		Expected Count	8.2	2.8	11.0
	São Bernardo	Count	5	4	9
		Expected Count	6.7	2.3	9.0
	Vera Cruz	Count	45	10	55
		Expected Count	40.8	14.2	55.0
Total		Count	123	43	166
		Expected Count	123.0	43.0	166.0

Table A1. 17 – Non-parametric Chi-Square Test: parishes vs. CATV

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	7.633	5	0.178

Table A1. 18 - Crosstab: parishes vs. gas (natural)

			Gas (natural)		
			0	1	Total
parish	Aradas	Count	14	7	21
		Expected Count	13.0	8.0	21.0
	Esgueira	Count	26	16	42
		Expected Count	26.1	15.9	42.0
	Glória	Count	9	19	28
		Expected Count	17.4	10.6	28.0
	Santa Joana	Count	8	3	11
		Expected Count	6.8	4.2	11.0
	São Bernardo	Count	4	5	9
		Expected Count	5.6	3.4	9.0
	Vera Cruz	Count	42	13	55
		Expected Count	34.1	20.9	55.0
Total		Count	103	63	166
		Expected Count	103.0	63.0	166.0

Table A1. 19 – Non-parametric Chi-Square Test: parishes vs. gas (natural)

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	17.328	5	0.004

2. DIFFERENCES BETWEEN TYPE OF HOUSING

(house=1; flat=0)

...AND SCALAR VARIABLES (Mann-Whitney tests)

(kitchen area, living room area, price in euros/m^2, price in euros, total area, floors and number of bedrooms)

Table A1. 20 – Non-parametric Mann-Whitney ranks: housing type vs. (kitchen area, living room area, price in euros/m^2, price in euros, total area, floors and number of bedrooms)

	Type (House=1. Flat=0)	N	Mean Rank	Sum of Ranks
Kitchen area	0	123	65.59	8068.00
	1	16	103.88	1662.00
Living room area	0	132	71.33	9415.00
	1	15	97.53	1463.00
Price (€/m^2)	0	144	80.93	11653.50
	1	22	100.34	2207.50
Price (€)	0	144	77.78	11201.00
	1	22	120.91	2660.00
Total area	0	144	80.42	11580.00
	1	22	103.68	2281.00
Floors	0	144	88.18	12698.00
	1	22	52.86	1163.00
Number of bedrooms	0	143	85.92	12287.00
	1	22	64.00	1408.00

Table A1. 21 – Non-parametric Mann-Whitney tests: housing type vs. (kitchen area, living room area, price in euros/m^2, price in euros, total area, floors and number of bedrooms)

	Kitchen area	Living room	Price (€/m2)	Price (€)	Total area	N bedrooms	Floors
Mann-Whitney U	442.000	637.000	1213.500	761.000	1140.000	1155.000	910.000
Asymp. Sig. (2-tailed)	0.000	0.024	0.078	0.000	0.034	0.031	0.001

...AND NOMINAL VARIABLES (chi-square tests)

(preservation, duplex, balcony, terrace, garage space, CATV and natural gas)

Table A1. 22 - Crosstab: housing type vs. preservation

			Preservation (Used=1. New=0)		
			0	1	Total
Type (House=1. Flat=0)	0	Count	14	129	143
		Expected Count	16.5	126.5	143.0
	1	Count	5	17	22
		Expected Count	2.5	19.5	22.0
Total		Count	19	146	165
		Expected Count	19.0	146.0	165.0

Table A1. 23 – Non-parametric Chi-Square Test: housing type vs. preservation

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	3.132	1	0.077

Table A1. 24 - Crosstab: housing type vs. duplex

			Duplex		
			0	1	Total
Type (House=1. Flat=0)	0	Count	109	33	142
		Expected Count	113.1	28.9	142.0
	1	Count	20	0	20
		Expected Count	15.9	4.1	20.0
Total		Count	129	33	162
		Expected Count	129.0	33.0	162.0

Table A1. 25 – Non-parametric Chi-Square Test: housing type vs. duplex

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	5.837	1	0.016

Table A1. 26 - Crosstab: housing type vs. balcony

			Balcony		
			0	1	Total
Type (House=1. Flat=0)	0	Count	120	24	144
		Expected Count	116.2	27.8	144.0
	1	Count	14	8	22
		Expected Count	17.8	4.2	22.0
Total		Count	134	32	166
		Expected Count	134.0	32.0	166.0

Table A1. 27 – Non-parametric Chi-Square Test: housing type vs. balcony

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	4.758	1	0.029

Table A1. 28 - Crosstab: housing type vs. terrace

			Terrace		
			0	1	Total
Type (House=1. Flat=0)	0	Count	131	13	144
		Expected Count	129.3	14.7	144.0
	1	Count	18	4	22
		Expected Count	19.7	2.3	22.0
Total		Count	149	17	166
		Expected Count	149.0	17.0	166.0

Table A1. 29 – Non-parametric Chi-Square Test: housing type vs. terrace

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	1.740	1	0.187

Table A1. 30 - Crosstab: housing type vs. garage space

			Provision of garage		
			0	1	Total
Type (House=1. Flat=0)	0	Count	57	87	144
		Expected Count	59.0	85.0	144.0
	1	Count	11	11	22
		Expected Count	9.0	13.0	22.0
Total		Count	68	98	166
		Expected Count	68.0	98.0	166.0

Table A1. 31 – Non-parametric Chi-Square Test: housing type vs. garage space

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	0.856	1	0.355

Table A1. 32 - Crosstab: housing type vs. cable TV

			Cable TV		
			0	1	Total
Type (House=1. Flat=0)	0	Count	101	43	144
		Expected Count	106.7	37.3	144.0
	1	Count	22	0	22
		Expected Count	16.3	5.7	22.0
Total		Count	123	43	166
		Expected Count	123.0	43.0	166.0

Table A1. 33 – Non-parametric Chi-Square Test: housing type vs. cable TV

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	8.866	1	0.003

Table A1. 34 - Crosstab: housing type vs. gas (natural)

			Gas (natural)		
			0	1	Total
Type (House=1. Flat=0)	0	Count	82	62	144
		Expected Count	89.3	54.7	144.0
	1	Count	21	1	22
		Expected Count	13.7	8.3	22.0
Total		Count	103	63	166
		Expected Count	103.0	63.0	166.0

Table A1. 35 – Non-parametric Chi-Square Test: housing type vs. (natural)

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	12.019	1	0.001

3. DIFFERENCES BETWEEN LEVEL OF PRESERVATION

(new=1; used=0)

...AND SCALAR VARIABLES (Mann-Whitney tests)

(kitchen area, living room area, price in euros/m², price in euros, total area, floors and number of bedrooms)

Table A1. 36 –Non-parametric Mann-Whitney ranks: level of preservation vs. (kitchen area, living room area, price in euros/m², price in euros, total area, floors and number of bedrooms)

	Preservation (Used=1. New=0)	N	Mean Rank	Sum of Ranks
Kitchen area	0	14	82.21	1151.00
	1	124	68.06	8440.00
Living room area	0	16	92.53	1480.50
	1	130	71.16	9250.50
Price (€/m^2)	0	19	112.34	2134.50
	1	146	79.18	11560.50
Price (€)	0	19	125.68	2388.00
	1	146	77.45	11307.00
Total_area	0	19	100.32	1906.00
	1	146	80.75	11789.00
Number of bedrooms	0	19	93.18	1770.50
	1	145	81.10	11759.50
Floors	0	19	79.45	1509.50
	1	146	83.46	12185.50

Table A1. 37 –Non-parametric Mann-Whitney tests: level of preservation vs. (kitchen area, living room area, price in euros/m², price in euros, total area, floors and number of bedrooms)

	Kitchen	Living room	Price (€/m2)	Price (€)	Total area	N bedrooms	Floors
Mann-Whitney U	690.000	735.500	829.500	576.000	1058.000	1174.500	1319.500
Asymp. Sig. (2-tailed)	0.209	0.056	0.004	0.000	0.092	0.262	0.711

...AND NOMINAL VARIABLES (chi-square tests)

(duplex, balcony, terrace, garage space, Cable TV, central heating and fireplace)

Table A1. 38 - Crosstab: level of preservation vs. duplex

			Duplex		
			0	1	Total
Preservation (Used=1. New=0)	0	Count	15	4	19
		Expected Count	15.1	3.9	19.0
	1	Count	113	29	142
		Expected Count	112.9	29.1	142.0
Total		Count	128	33	161
		Expected Count	128.0	33.0	161.0

Table A1. 39 – Non-parametric Chi-Square Test: level of preservation vs. duplex

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	0.004	1	0.949

Table A1. 40 - Crosstab: level of preservation vs. balcony

			Balcony		
			0	1	Total
Preservation (Used=1. New=0)	0	Count	14	5	19
		Expected Count	15.4	3.6	19.0
	1	Count	120	26	146
		Expected Count	118.6	27.4	146.0
Total		Count	134	31	165
		Expected Count	134.0	31.0	165.0

Table A1. 41 – Non-parametric Chi-Square Test: level of preservation vs. balcony

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	0.798	1	0.372

Table A1. 42 - Crosstab: level of preservation vs. terrace

			Terrace		
			0	1	Total
Preservation (Used=1. New=0)	0	Count	15	4	19
		Expected Count	17.0	2.0	19.0
	1	Count	133	13	146
		Expected Count	131.0	15.0	146.0
Total		Count	148	17	165
		Expected Count	148.0	17.0	165.0

Table A1. 43 – Non-parametric Chi-Square Test: level of preservation vs. terrace

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	2.685	1	0.101

Table A1. 44 - Crosstab: level of preservation vs. garage space

			Provision of garage		
			0	1	Total
Preservation (Used=1. New=0)	0	Count	6	13	19
		Expected Count	7.7	11.3	19.0
	1	Count	61	85	146
		Expected Count	59.3	86.7	146.0
Total		Count	67	98	165
		Expected Count	67.0	98.0	165.0

Table A1. 45 – Non-parametric Chi-Square Test: level of preservation vs. garage space

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	0.726	1	0.394

Table A1. 46 - Crosstab: level of preservation vs. provision of cable TV

			Cable TV		
			0	1	Total
Preservation (Used=1. New=0)	0	Count	14	5	19
		Expected Count	14.0	5.0	19.0
	1	Count	108	38	146
		Expected Count	108.0	38.0	146.0
Total		Count	122	43	165
		Expected Count	122.0	43.0	165.0

Table A1. 47 – Non-parametric Chi-Square Test: level of preservation vs. CATV

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	0.001	1	0.979

Table A1. 48 - Crosstab: level of preservation vs. gas (natural)

			Gas (natural)		
			0	1	Total
Preservation (Used=1. New=0)	0	Count	14	5	19
		Expected Count	11.7	7.3	19.0
	1	Count	88	58	146
		Expected Count	90.3	55.7	146.0
Total		Count	102	63	165
		Expected Count	102.0	63.0	165.0

Table A1. 49 – Non-parametric Chi-Square Test: level of preservation vs. gas (natural)

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	1.281	1	0.258

APPENDIX 2

Statistical inferences: sample 2

(Selected SPSS outputs)

1. DIFFERENCES BETWEEN THE TERRITORIES

(urban, suburban and rural areas)

…AND SCALAR VARIABLES (Kuskal Wallis tests)

(number of bedrooms, price in euros, price in euros/m^2 and total area)

Table A2. 1 – Non-parametric Kruskal Wallis tests: territorial areas vs. (number of bedrooms, price in euros, price in euros/m^2 and total area)

	areas	**N**	**Mean Rank**
Number of room	Urban	2786	4921.92
	Suburban	7461	6313.96
	Rural	2220	7611.88
	Total	12467	
Price (€)	Urban	2786	5932.27
	Suburban	7461	6307.73
	Rural	2220	6364.85
	Total	12467	
Price (€/m^2)	Urban	2786	8810.00
	Suburban	7461	5932.54
	Rural	2220	4014.40
	Total	12467	
Total area	Urban	2786	4592.33
	Suburban	7461	6454.15
	Rural	2220	7554.35
	Total	12467	

Table A2. 2 –Non-parametric Kruskal Wallis tests: territorial areas vs. (number of bedrooms, price in euros, price in euros/m^2 and total area)

	Number of rooms	Price (€)	Price (€/m^2)	Total area
Chi-Square	757.351	25.653	2323.985	906.793
Asymp. Sig.	0.000	0.000	0.000	0.000

...AND NOMINAL VARIABLES (chi-square tests)

(type of house, preservation, duplex, balcony, terrace, garage space, garage, central heating and fireplace)

Table A2. 3 – Crosstab: territorial areas vs. type of house

			dType_house		
			0	1	Total
areas	Urban	Count	2726	60	2786
		Expected Count	1994.0	792.0	2786.0
	Suburban	Count	5199	2262	7461
		Expected Count	5340.1	2120.9	7461.0
	Rural	Count	998	1222	2220
		Expected Count	1588.9	631.1	2220.0
Total		Count	8923	3544	12467
		Expected Count	8923.0	3544.0	12467.0

Table A2. 4 – Non-parametric Chi-Square Test: territorial areas vs. type of house

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	1731.407	2	0.000

Table A2. 5 – Crosstab: territorial areas vs. preservation

			d1New		
			0	1	Total
territorial areas	Urban	Count	1185	1571	2756
		Expected Count	1213.5	1542.5	2756.0
	Suburban	Count	3195	3969	7164
		Expected Count	3154.5	4009.5	7164.0
	Rural	Count	918	1194	2112
		Expected Count	930.0	1182.0	2112.0
Total		Count	5298	6734	12032
		Expected Count	5298.0	6734.0	12032.0

Table A2. 6 – Non-parametric Chi-Square Test: territorial areas vs. preservation

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	2.404	2	0.301

Table A2. 7 – Crosstab: territorial areas vs. duplex

			Duplex		
			0	1	Total
areas	Urban	Count	2413	373	2786
		Expected Count	2442.3	343.7	2786.0
	Suburban	Count	6514	947	7461
		Expected Count	6540.6	920.4	7461.0
	Rural	Count	2002	218	2220
		Expected Count	1946.1	273.9	2220.0
Total		Count	10929	1538	12467
		Expected Count	10929.0	1538.0	12467.0

Table A2. 8 –Non-parametric Chi-Square Test: territorial areas vs. duplex

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	16.727	2	0.000

Table A2. 9 – Crosstab: territorial areas vs. balcony

			Balcony		
			0	1	Total
areas	Urban	Count	1813	973	2786
		Expected Count	1692.1	1093.9	2786.0
	Suburban	Count	4402	3059	7461
		Expected Count	4531.5	2929.5	7461.0
	Rural	Count	1357	863	2220
		Expected Count	1348.3	871.7	2220.0
Total		Count	7572	4895	12467
		Expected Count	7572.0	4895.0	12467.0

Table A2. 10 – Non-parametric Chi-Square Test: territorial areas vs. balcony

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	31.568	2	0.000

Table A2. 11 – Crosstab: territorial areas vs. terrace

			Terrace		
			0	1	Total
areas	Urban	Count	2494	292	2786
		Expected Count	2279.2	506.8	2786.0
	Suburban	Count	5928	1533	7461
		Expected Count	6103.7	1357.3	7461.0
	Rural	Count	1777	443	2220
		Expected Count	1816.1	403.9	2220.0
Total		Count	10199	2268	12467
		Expected Count	10199.0	2268.0	12467.0

Table A2. 12 – Non-parametric Chi-Square Test: territorial areas vs. terrace

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	143.745	2	0.000

Table A2. 13 – Crosstab: territorial areas vs. garage space

			Placeofgarage		
			0	1	Total
areas	Urban	Count	2094	692	2786
		Expected Count	2336.8	449.2	2786.0
	Suburban	Count	6343	1118	7461
		Expected Count	6258.1	1202.9	7461.0
	Rural	Count	2020	200	2220
		Expected Count	1862.1	357.9	2220.0
Total		Count	10457	2010	12467
		Expected Count	10457.0	2010.0	12467.0

Table A2. 14 – Non-parametric Chi-Square Test: territorial areas vs. garage space

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	246.720	2	0.000

Table A2. 15 – Crosstab: territorial areas vs. garage

			Garage		
			0	1	Total
areas	Urban	Count	1285	1501	2786
		Expected Count	1008.3	1777.7	2786.0
	Suburban	Count	2375	5086	7461
		Expected Count	2700.3	4760.7	7461.0
	Rural	Count	852	1368	2220
		Expected Count	803.5	1416.5	2220.0
Total		Count	4512	7955	12467
		Expected Count	4512.0	7955.0	12467.0

Table A2. 16 – Non-parametric Chi-Square Test: territorial areas vs. garage

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	185.000	2	0.000

Table A2. 17 – Crosstab: territorial areas vs. central heating

			CentralHeating		
			0	1	Total
areas	Urban	Count	1541	1245	2786
		Expected Count	1579.5	1206.5	2786.0
	Suburban	Count	4155	3306	7461
		Expected Count	4229.9	3231.1	7461.0
	Rural	Count	1372	848	2220
		Expected Count	1258.6	961.4	2220.0
Total		Count	7068	5399	12467
		Expected Count	7068.0	5399.0	12467.0

Table A2. 18 – Non-parametric Chi-Square Test: territorial areas vs. central heating

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	28.823	2	0.000

Table A2. 19 – Crosstab: territorial areas vs. fireplace

			Fireplace		
			0	1	Total
areas	Urban	Count	2327	459	2786
		Expected Count	1981.1	804.9	2786.0
	Suburban	Count	5131	2330	7461
		Expected Count	5305.3	2155.7	7461.0
	Rural	Count	1407	813	2220
		Expected Count	1578.6	641.4	2220.0
Total		Count	8865	3602	12467
		Expected Count	8865.0	3602.0	12467.0

Table A2. 20 – Non-parametric Chi-Square Test: territorial areas vs. fireplace

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	293.470	2	0.000

2. DIFFERENCES BETWEEN TYPE OF HOUSING

(house=1; flat=0)

...AND SCALAR VARIABLES (Mann-Whitney tests)

(number of bedrooms, price in euros, price in euros/m^2 and total area)

Table A2. 21 – Non-parametric Mann-Whitney ranks: housing type vs. (number of bedrooms, price in euros, price in euros/m^2 and total area)

	dType_house	N	Mean Rank	Sum of Ranks
Number of room	0	8923	4753.12	42412104.50
	1	3544	9962.52	35307173.50
	Total	12467		
Price (€)	0	8923	4885.76	43595593.50
	1	3544	9628.58	34123684.50
	Total	12467		
Price (€/m^2)	0	8923	7143.89	63744916.50
	1	3544	3943.10	13974361.50
	Total	12467		
Total area	0	8923	4644.59	41443645.00
	1	3544	10235.79	36275633.00
	Total	12467		

Table A2. 22 –Non-parametric Mann-Whitney tests: housing type vs. (number of bedrooms, price in euros, price in euros/m^2 and total area)

	Number of room	Price (€)	Price (€/m^2)	Total area
Mann-Whitney U	2597678.500	3781167.500	7692621.500	1629219.000
Asymp. Sig. (2-tailed)	0.000	0.000	0.000	0.000

...AND NOMINAL VARIABLES (chi-square tests)

(preservation, duplex, balcony, terrace, garage space, garage, central heating and fireplace)

Table A2. 23 – Crosstab: housing type vs. preservation

			d1New		
			0	1	Total
dType_house	0	Count	4232	4423	8655
		Expected Count	3811.0	4844.0	8655.0
	1	Count	1066	2311	3377
		Expected Count	1487.0	1890.0	3377.0
Total		Count	5298	6734	12032
		Expected Count	5298.0	6734.0	12032.0

Table A2. 24 – Non-parametric Chi-Square Test: housing type vs. preservation

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	296.042	1	0.000

Table A2. 25 – Crosstab: housing type vs. duplex

			Duplex		
			0	1	Total
dType_house	0	Count	7389	1534	8923
		Expected Count	7822.2	1100.8	8923.0
	1	Count	3540	4	3544
		Expected Count	3106.8	437.2	3544.0
Total		Count	10929	1538	12467
		Expected Count	10929.0	1538.0	12467.0

Table A2. 26 – Non-parametric Chi-Square Test: territorial areas vs. duplex

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	684.128	1	0.000

Table A2. 27 – Crosstab: housing type vs. balcony

			Balcony		
			0	1	Total
dType_house	0	Count	5520	3403	8923
		Expected Count	5419.5	3503.5	8923.0
	1	Count	2052	1492	3544
		Expected Count	2152.5	1391.5	3544.0
Total		Count	7572	4895	12467
		Expected Count	7572.0	4895.0	12467.0

Table A2. 28 – Non-parametric Chi-Square Test: housing type vs. balcony

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	16.696	1	0.000

Table A2. 29 – Crosstab: housing type vs. terrace

			Terrace		
			0	1	Total
dType_house	0	Count	7461	1462	8923
		Expected Count	7299.7	1623.3	8923.0
	1	Count	2738	806	3544
		Expected Count	2899.3	644.7	3544.0
Total		Count	10199	2268	12467
		Expected Count	10199.0	2268.0	12467.0

Table A2. 30 – Non-parametric Chi-Square Test: housing type vs. terrace

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	68.899	1	0.000

Table A2. 31 – Crosstab: housing type vs. garage space

			Placeofgarage		
			0	1	Total
dType_house	0	Count	7007	1916	8923
		Expected Count	7484.4	1438.6	8923.0
	1	Count	3450	94	3544
		Expected Count	2972.6	571.4	3544.0
Total		Count	10457	2010	12467
		Expected Count	10457.0	2010.0	12467.0

Table A2. 32 – Non-parametric Chi-Square Test: housing type vs. garage space

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	664.375	1	0.000

Table A2. 33 – Crosstab: housing type vs. garage

			Garage		
			0	1	Total
dType_house	0	Count	3309	5614	8923
		Expected Count	3229.4	5693.6	8923.0
	1	Count	1203	2341	3544
		Expected Count	1282.6	2261.4	3544.0
Total		Count	4512	7955	12467
		Expected Count	4512.0	7955.0	12467.0

Table A2. 34 – Non-parametric Chi-Square Test: housing type vs. garage

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	10.824	1	0.001

Table A2. 35 – Crosstab: housing type vs. central heating

			CentralHeating		
			0	1	Total
dType_house	0	Count	5291	3632	8923
		Expected Count	5058.8	3864.2	8923.0
	1	Count	1777	1767	3544
		Expected Count	2009.2	1534.8	3544.0
Total		Count	7068	5399	12467
		Expected Count	7068.0	5399.0	12467.0

Table A2. 36 – Non-parametric Chi-Square Test: housing type vs. central heating

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	86.593	1	0.000

Table A2. 37 – Crosstab: housing type vs. fireplace

			Fireplace		
			0	1	Total
dType_house	0	Count	6812	2111	8923
		Expected Count	6344.9	2578.1	8923.0
	1	Count	2053	1491	3544
		Expected Count	2520.1	1023.9	3544.0
Total		Count	8865	3602	12467
		Expected Count	8865.0	3602.0	12467.0

Table A2. 38 – Non-parametric Chi-Square Test: housing type vs. fireplace

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	418.601	1	0.000

3. DIFFERENCES BETWEEN LEVEL OF PRESERVATION

(new=1; used=0)

...AND SCALAR VARIABLES (Mann-Whitney tests)

(number of bedrooms, price in euros, price in euros/m^2 and total area)

Table A2. 39 – Non-parametric Mann-Whitney ranks: level of preservation vs. (number of bedrooms, price in euros, price in euros/m^2 and total area)

	d1New	N	Mean Rank	Sum of Ranks
Number of room	0	5298	5633.99	29848854.00
	1	6734	6317.44	42541674.00
	Total	12032		
Price (€)	0	5298	4655.62	24665470.50
	1	6734	7087.18	47725057.50
	Total	12032		
Price (€/m^2)	0	5298	5074.81	26886349.50
	1	6734	6757.38	45504178.50
	Total	12032		
Total area	0	5298	5536.50	29332373.00
	1	6734	6394.14	43058155.00
	Total	12032		

Table A2. 40 – Non-parametric Mann-Whitney tests : level of preservation vs. (number of bedrooms, price in euros, price in euros/m^2 and total area)

	Number of room	Price (€)	Price (€/m^2)	Total area
Mann-Whitney U	15811803.000	10628419.500	12849298.500	15295322.000
Asymp. Sig. (2-tailed)	0.000	0.000	0.000	0.000

...AND NOMINAL VARIABLES (chi-square tests)

(duplex, balcony, terrace, garage space, garage, central heating and fireplace)

Table A2. 41 – Crosstab: preservation vs. duplex

			Duplex		
			0	1	Total
d1New	0	Count	4646	652	5298
		Expected Count	4646.3	651.7	5298.0
	1	Count	5906	828	6734
		Expected Count	5905.7	828.3	6734.0
Total		Count	10552	1480	12032
		Expected Count	10552.0	1480.0	12032.0

Table A2. 42 – Non-parametric Chi-Square Test: preservation vs. duplex

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	0.000	1	0.986

Table A2. 43 – Crosstab: preservation vs. balcony

			Balcony		
			0	1	Total
d1New	0	Count	3008	2290	5298
		Expected Count	3256.6	2041.4	5298.0
	1	Count	4388	2346	6734
		Expected Count	4139.4	2594.6	6734.0
Total		Count	7396	4636	12032
		Expected Count	7396.0	4636.0	12032.0

Table A2. 44 – Non-parametric Chi-Square Test: preservation vs. balcony

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	88.037	1	0.000

Table A2. 45 – Crosstab: preservation vs. terrace

			Terrace		
			0	1	Total
d1New	0	Count	4297	1001	5298
		Expected Count	4353.1	944.9	5298.0
	1	Count	5589	1145	6734
		Expected Count	5532.9	1201.1	6734.0
Total		Count	9886	2146	12032
		Expected Count	9886.0	2146.0	12032.0

Table A2. 46 – Non-parametric Chi-Square Test: preservation vs. terrace

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	7.233	1	0.007

Table A2. 47 – Crosstab: preservation vs. garage space

			Placeofgarage		
			0	1	Total
d1New	0	Count	4495	803	5298
		Expected Count	4432.3	865.7	5298.0
	1	Count	5571	1163	6734
		Expected Count	5633.7	1100.3	6734.0
Total		Count	10066	1966	12032
		Expected Count	10066.0	1966.0	12032.0

Table A2. 48 – Non-parametric Chi-Square Test: preservation vs. garage space

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	9.693	1	0.002

Table A2. 49 – Crosstab: preservation vs. garage

			Garage		
			0	1	Total
d1New	0	Count	1954	3344	5298
		Expected Count	1925.1	3372.9	5298.0
	1	Count	2418	4316	6734
		Expected Count	2446.9	4287.1	6734.0
Total		Count	4372	7660	12032
		Expected Count	4372.0	7660.0	12032.0

Table A2. 50 – Non-parametric Chi-Square Test: preservation vs. garage

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	1.217	1	0.270

Table A2. 51 – Crosstab: preservation vs. central heating

			CentralHeating		
			0	1	Total
d1New	0	Count	3561	1737	5298
		Expected Count	2962.5	2335.5	5298.0
	1	Count	3167	3567	6734
		Expected Count	3765.5	2968.5	6734.0
Total		Count	6728	5304	12032
		Expected Count	6728.0	5304.0	12032.0

Table A2. 52 – Non-parametric Chi-Square Test: preservation vs. central heating

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	490.061	1	0.000

Table A2. 53 – Crosstab: preservation vs. fireplace

			Fireplace		
			0	1	Total
d1New	0	Count	3135	2163	5298
		Expected Count	3827.8	1470.2	5298.0
	1	Count	5558	1176	6734
		Expected Count	4865.2	1868.8	6734.0
Total		Count	8693	3339	12032
		Expected Count	8693.0	3339.0	12032.0

Table A2. 54 – Non-parametric Chi-Square Test: preservation vs. fireplace

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	807.231	1	0.000