UNIVERSITÀ DEGLI STUDI DI TORINO

The housing market as a complex system: may income segregation emerge in cities from microscopic interactions?

Tesi per la laurea magistrale in
Fisica dei Sistemi Complessi

Da sostenere in seduta pubblica il:
24/07/2015

Candidato:
Marco Pangallo

Relatore:
Prof. Pietro Terna

Correlatori:
Prof. Jean-Pierre Nadal
Prof. Annick Vignes

Controrelatore:
Prof. Michele Caselle

Acknowledgements

I would like to thank my supervisor Pietro Terna and my co-supervisors Jean-Pierre Nadal and Annick Vignes for their advice and support while preparing this thesis. In particular, I am grateful to Jean-Pierre Nadal for proposing me this topic and for arranging a six-months internship at the École des hautes études en sciences sociales (EHESS), in the laboratory Centre d'analyse et de mathématique sociales (CAMS), where the thesis was prepared. His point of view as a physicist has been essential in developing this work. Conversely, I thank Pietro Terna and Annick Vignes for their suggestions as economists. I also thank the examiner of this thesis, Michele Caselle.

I am grateful to all the participants at the seminar "Entre discret et continu" at the Institut des Systèmes Complexes de Paris Ile-de-France, where I had the chance to present this work. I thank in particular Andrea Tellini for organizing the seminar and Henri Berestycki for the stimulating discussion which followed. I also thank all the members of the CAMS who made my workplace a pleasant environment.

I acknowledge financial support from the Erasmus Traineeship Program.

Abstract (English)

The housing market as a complex system: may income segregation emerge in cities from microscopic interactions?

The allocation of people into the most productive areas, mostly the so-called alpha cities, is becoming an issue of central importance in the globalized world. The growing demand is inducing a soar in the real-estate prices, making it impossible for the poorest individuals to settle in the most important cities, with all the consequences in terms of lost opportunities that this phenomenon entails. We propose an Agent-Based Model (ABM) to explain the distribution of people with respect to their income and to determine whether the market mechanism implies residential income segregation. Our ABM is inspired by some stylized facts observed in a dataset recording more than 400,000 real-estate transactions occurred in Paris in the years 1990-2007, and by previous works, notably Gauvin *et al.* (2013), by which our model differs substantially on the market aspect. We actually give priority to provide our model with reasonable economic assumptions, at the cost of a greater complexity. However, we keep other features simple enough in order to allow for an analytical characterization of the steady state of the model in some specific cases, by using mean-field arguments and taking the continuum limit. Thanks to the mathematical solution we understand the effect of the (many) parameters and we prove formally that the agents self-organize into a stable state. Our model considers a city whose locations are characterized by a certain level of attractiveness, which evolves according to the social composition of the neighbourhood. The agents, heterogeneous in their income, compete for the available apartments: the price distribution emerges out of their interactions. This approach fits into the emerging field of Complexity Economics and aims at giving a concrete example of its great potential. We find under which conditions the markets clear and/or exhibit bubble behaviour. Our most important result is about the real-estate prices: the income distribution of the buyers determines the offset of the price distribution, the preferences of the incoming agents determine the steepness. Therefore, the prices rise everywhere in the city even though the rich are only willing to purchase an apartment in the center; when the buyers also take into account the social composition of the neighbourhood, we observe gentrification: the poorest are gradually segregated away from the city, but the other agents experience a higher level of social mixing. These results are confirmed by the trends in the data and highlight the urgent need for policies to prevent severe income segregation.

Keywords: housing market, agent-based model, complexity economics, segregation

Abstract (Italiano)

Il mercato immobiliare come sistema complesso: può la segregazione basata sul reddito emergere nelle città a partire da interazioni microscopiche?

La distribuzione della popolazione nelle zone più produttive, quali le cosiddette "città alpha", sta diventando un tema centrale nel mondo globalizzato. La domanda in crescita sta causando un rapido aumento dei prezzi degli immobili, rendendo impossibile per le fasce più povere stabilirsi nelle città di maggiore importanza, con tutto ciò che questo comporta in termini di opportunità perse. Proponiamo un modello ad agenti (ABM) per spiegare la distribuzione della popolazione rispetto al reddito e per determinare se il meccanismo del mercato implichi segregazione basata sul reddito. Il nostro ABM si ispira ad alcuni fatti stilizzati osservati in un dataset che riporta più di 400.000 transazioni nel mercato immobiliare avvenute a Parigi tra il 1990 e il 2007, e a lavori precedenti, in particolar modo a Gauvin *et al.* (2013), dal quale il nostro modello differisce in maniera significativa per quanto riguarda l'aspetto del mercato. La nostra priorità è di fornire al modello una base economica ragionevole, al costo di una maggiore complessità. Tuttavia, manteniamo altre assunzioni sufficientemente semplici in modo da permettere una caratterizzazione analitica dello stato stazionario del modello, tramite approssimazioni di campo medio e nel limite del continuo. Grazie alla soluzione matematica comprendiamo l'effetto dei (molti) parametri e dimostriamo formalmente che gli agenti si auto-organizzano in uno stato stabile. Il modello considera una città le cui zone sono caratterizzate da un certo livello di attrattività, che evolve a seconda della composizione sociale del quartiere. Gli agenti, eterogenei per la loro ricchezza, sono in competizione per gli appartamenti disponibili: la distribuzione dei prezzi emerge dalle loro interazioni. Il nostro approccio si inserisce nel campo emergente dell'economia della complessità, del cui grande potenziale ambisce a fornire un esempio concreto. Troviamo le condizioni da soddisfare affinché il mercato raggiunga l'equilibrio e/o mostri una bolla immobiliare. Il nostro risultato più importante riguarda i prezzi degli immobili: la distribuzione del reddito dei compratori determina il livello dei prezzi, le preferenze degli agenti in arrivo ne determinano la pendenza. Di conseguenza, i prezzi salgono dappertutto nella città anche qualora i ricchi vogliano solo comprare un appartamento nel centro; quando i compratori tengono anche in considerazione la composizione sociale del quartiere, osserviamo gentrificazione: i più poveri sono sempre più segregati dalla città, ma gli altri agenti beneficiano di una maggiore eterogeneità sociale. Questi risultati sono confermati dagli andamenti nei dati e sottolineano la necessità urgente di politiche che impediscano una forte segregazione basata sul reddito.

Parole chiave: Mercato immobiliare, modello ad agenti, economia della complessità, segregazione

Contents

Chapter 1

Introduction

The residential distribution of people with respect to their income is becoming a pressing issue as the cities become more and more important. Cities are "open-ended social reactors" (Bettencourt, 2013), where the density of interactions and job opportunities is coming to play an even more prominent role in the global economy. The relevance of the real-estate markets appeared to be declining until 40 years ago, but ever since the value of land has been on the rise in the most important cities of the world. Following some empirical insights and building on previous work (Gauvin *et al.*, 2013), we propose an Agent-Based Model (ABM) for the housing market in an artificial city. We keep the model complex enough to explain various aspects of the real-estate market, but simple enough to allow for an analytical resolution of some specific cases. Thanks to the mathematical characterization of the steady state we are able to identify the relevant parameters and to understand their effect. We reproduce a set of stylized facts and we are able to assess quantitatively the properties about the price distribution and the residential income segregation patterns, even though we cannot directly check the predictive power of our model through the data. Our work fits in the emerging field of *Complexity Economics* and aims at giving a concrete example of the great potential of that approach. For instance, by considering the heterogeneity of the agents we are able to study *gentrification* in a multidimensional way, and we find that this phenomenon drives away the poorest agents from the city, but it decreases the level of segregation between the other categories. Another interesting result is that when the real-estate prices rise because there are a lot of rich agents coming to the city, they rise everywhere even though the rich are only willing to buy an apartment in the center. This finding suggests that unless mechanisms to prevent segregation are put into place, the unavoidable outcome is that the poor are bound to be thrown out of the world cities, even from the outskirts.

In this chapter we present the motivation, the methodology and the main results of the thesis. We attempt to provide a self-contained explanation,

accessible to non specialists too. The rest of this thesis is organized as follows. In Chapter 2 we illustrate the main works which inspired our model, and we compare our approach with some related papers in the literature. In Chapter 3 we describe our database and we show the most important empirical results, both the ones which influenced the making of the model and the ones which are interesting per se. Chapter 4 is the core of the thesis: we describe the ABM, we present the results of the mathematical analysis and of the computer simulations, we discuss the originality of our assumptions and our conclusions. The intermediate outcomes before the development of the final version of our model are shortly shown in Chapter 5. We conclude and we examine some perspectives of this work in Chapter 6.

As *The Economist* writes[1]: "One of the main ways economies increase worker productivity, and thus grow richer, is through the reallocation of people and resources away from low-productivity segments to more efficient ones." World cities, such as San Francisco, New York, Paris and London, are efficient segments, since workers are more productive and there is a higher density of labour. The problem is that the high demand in the world cities boosts the rental and housing prices, making it less convenient to work there. Again from *The Economist*: "Many workers will take lower-paying jobs elsewhere because the income left over after paying for cheaper housing is more attractive. Labour ends up allocating itself toward low-productivity markets, and the whole economy suffers". The importance of cities has been highlighted in recent studies in the field of Complex Systems (Bettencourt, 2013; Batty, 2008; Louf and Barthelemy, 2013). These works focus on the scaling of city quantities: they find that production indicators such as GDP and number of patents scale superlinearly with respect to the population size, confirming the role and the importance of global cities. Here we look at cities from a different perspective, the one of the housing markets. We are interested in determining the distribution of the individuals with respect to their income (where do the rich, the middle-class, the poor settle?) and in explaining which factors influence it. We want to characterize the real-estate price distribution and to understand its implications for the residential income segregation. As it has already been mentioned, these issues are likely to affect the role of cities in the global economy through a negative feedback: the more important a city becomes, more people aim to settle there, the real-estate prices rise, the workers lose many advantages in staying there: therefore the city development is dampened if no countermeasures are taken.

An example of this trend is provided by Paris[2]. We observe in our dataset, recording more than 400.000 real estate transactions occurred in

[1]The paradox of soil, *The Economist*, April 4th 2015

[2]The project "Métropole du Grand Paris" can be seen as an attempt to face the aforementioned issues

Paris in the years 1990-2007, that in the period 2005-2007 a dramatic increase (a bubble) occurs in the prices. All the districts (the *arrondissement*) in Paris follow the trend, whether it is a rise or a fall in the prices, but the relative ranking varies. For instance the 16th arrondissement used to be the most expensive in 1990, but it was just the 7th most expensive in 2007. Moreover, in 1990 there were zones with low prices, whereas already by 2003 only few spots with low prices were surviving in the extreme North of Paris. Other interesting insights coming from the data concern the sellers: the more time lapses after the purchase of the apartment they are selling, the lower is the price of the transaction; if the apartment is sold in a bubble period, a positive profit is made; the fact that many apartments are bought and sold within a few years is a footprint of a speculative behaviour, which adds to the complexity of the housing market dynamics.

In order to build an ABM of the real-estate market, we base on the empirical observations above and on previous work, notably Gauvin *et al.* (2013) and Alonso *et al.* (1964). Our methodology is to give a robust economic ground to our model, and only subsequently to develop a software implementation and a mathematical description of it. This approach contrasts with the simple adaptation of a model coming from physics or from applied mathematics to a context in the social sciences, as in many works in the literature. For instance, Short *et al.* (2008) develop a diffusion model for burglaries, using the same differential equations as for chemotaxis. Nevertheless, a too simple model of the housing market would lose its main features, so we tune the complexity of our ABM in order to account in a realistic way for the decisions of the agents and for the market mechanism. An example is given by the reservation prices of the sellers[3]: they determine it by taking into account the average price of the real-estate transactions in their neighbourhood, the competition in the market (the number of potential buyers with respect to the apartments on sale) and the price they paid to purchase the apartment beforehand. Therefore, we need to keep track of the individual agents, with all the consequences in terms of computational effort and analytical treatment.

Yet, some assumptions are made to reduce the complexity of the model: in the author's opinion, ABMs should not be so complex that one would lose track of the causal relationships and of the effect of the parameters. For instance, the apartments are put on sale with an exogenous fixed probability. This is not the case in Gilbert *et al.* (2009), where the agents have to pay a mortgage and may lose their job, eventuality which forces them to put their apartment on sale: in this example it is not clear which factors determine the prices, as too many variables and feedbacks are considered. Therefore, we keep the model simple enough to allow for a mean-field mathematical

[3] A seller would not accept to sell his apartment at a lower price than his reservation price

solution of some specific cases in the steady state. Later[4], we perform computer simulations on the fully-fledged ABM in order to check the robustness of the mathematical results and we use the computational power of ABMs to get more general results. For instance, we find through the mathematical analysis that in a city where the intrinsic attractiveness of a neighbourhood decreases with its distance from the center, the real-estate price distribution gets flattened if the buyers take into account also the popularity of the neighbourhood (that is the social component of the attractiveness, i.e. the number of newcomers in this case). This outcome was not obvious: the social component could exacerbate the spatial asymmetry, making the center even more attractive, as it was indeed the case in Gauvin *et al.* (2013). This result is confirmed through the computer simulations. Moreover, whereas through the mathematical analysis at most two income categories can be considered, the simulations allow an arbitrary number of income categories, thus adding new insights: the poorest categories are segregated away from the city, but the rich and the middle-class experience a higher level of social mixing (see below).

Another important issue is whether it is possible to "take the model to the data". The ultimate goal of the models in the social sciences should be to help the policy makers in taking decisions. On the one hand, the qualitative insights that a logically coherent model can give are intrinsically useful. Think of the IS-LM model in macroeconomics, which accounts for an extremely stylized picture of an economic system; it prognosticates that an increase in government spending would help recovering from a recession, although it is not able to provide numeric estimates of its effect on GDP. On the other hand, modern models are expected to be able to make quantitative predictions. Sticking to the macroeconomics example, the underestimation by the traditional models of the recessive effect of the austerity measures in many European countries after the 2008 crisis is one of the factors which led to a severe unemployment and to a GDP plunge worse than expected. In order to avoid such mistakes, models should be calibrated through real data and be able to make reliable economic forecasts. This has not been possible so far for the model presented in this thesis, because of several variables hard to determine. For instance, in order to compute the intrinsic attractiveness of a neighbourhood in a real city many arbitrary decisions have to be taken, about the most important factors (e.g. distance from the center, public transports, amenities, etc.) and their relative weight. Moreover, the parameters determining the behaviour of the sellers depend on the situation of the housing market and are difficult to be estimated. On top of that, the high dimensionality of the parameter space enhances the

[4]Notice that although this is the logical timeline to find conclusive results, the actual research took great advantage by going back and forth from the exploratory simulations to the mathematical analysis

overfitting risk. We leave the calibration of the model for future work, and we focus here on another aspect: our model can quantitatively assess the effect of different settings, highlighting which ones have a dramatic impact on the market and which ones have only a limited effectiveness. This is a step ahead with respect to models such as the IS-LM: new mathematical and computational tools allow to consider much more complex models, thus studying different circumstances. For instance, we find that the effect on the real-estate prices of a small group of extremely rich buyers is one order of magnitude lower than that of a higher number of relatively affluent buyers.

A final methodological consideration is about Complexity Economics, a new research field born thanks to the pioneering work at the Santa Fe Institute (Anderson *et al.*, 1988). We compare it with Neoclassical Economics, the current mainstream approach, and we explain why it provides novel insights that standard economics cannot provide. From its foundations at the end of the XIX century, modern economic theory has always stuck to the same modelling paradigm: rational agents seek to maximize in isolation their utility given a budget constraint, rational firms maximize their profits given a cost constraint, a competitive (i.e. offer = demand) equilibrium is always reached under some technical conditions (Mas-Colell *et al.*, 1995). Neoclassical Economics achieved a great deal in understanding which conditions lead to certain patterns (such as equilibrium, Arthur 2014) and provided mathematical models for a number of stylized facts, such as marginalism and the "invisible hand". The basic framework has been extended to account for other phenomena: private information, externalities, public goods, monopoly and oligopoly. Despite these achievements, neoclassical economic theory did not really change from its foundations. Modern microeconomics and macroeconomics models are still based on full optimization and relax a few assumptions at a time[5]. Many scholars claim that Neoclassical Economics has not proven useful in foreseeing and dealing with the 2008 financial crisis (see, e.g., Krugman 2009). From the theory of business cycles, recessions were to be attributed to exogenous shocks (such as an increase in the price of oil), and this was not the case for the mortgage financial crisis and the subsequent depression in the real economy. As Trichet (2010) put it: "we felt abandoned by our conventional tools". Policymakers were actually guided by experience and common sense in facing the crisis and its aftermath. It has been argued (Kirman 2010b, Bouchaud 2008) that the crisis should lead to a scientific revolution in economics. And not only macroeconomics is concerned: experimental economics showed that

[5]From the author's point of view, the main reason of the success of this modelling approach is that it provides a logically coherent framework which is extremely flexible to adaptations in several economic context. It relieves the researchers from the daunting task of specifying and justifying the behaviour of the agents, since it is often just a matter of solving the model for the equilibrium resulting from the maximisation of an objective function

the foundations of microeconomics are systematically violated, and thus its assumptions seriously flawed (e.g. transitivity of preferences).

Complexity Economics treats the economy as a *complex evolving system.* Its keywords are interaction, self-organization, learning, heterogeneity, out-of-equilibrium, bounded rationality, interdependence and coordination. The agents are not fully rational, they just do not act against their own interest, usually they learn from their mistakes following a trial-and-error process and they form expectations adopting rule of thumbs (the rational expectations hypothesis is relaxed[6]). The agents are heterogeneous, as the assumption of the representative agent[7] is deemed to lose essential features of the economic behaviour. Actually, heterogeneous agents interact following non trivial contact patterns, such as power law and small world networks. Their actions are interdependent and the main issue is coordination: if it breaks down, sudden crises may happen. In the language of Complex Systems, the emergent behaviour of the agents may undergo phase transitions. Finally, models in Complexity Economics take into account the temporal dimension and may study out-of-equilibrium states: they question whether self-organization towards an equilibrium state always occurs or whether the economic dynamics may lead to a chaotic attractor. The models are based on reasonable assumptions and should lead to testable conclusions (Kirman, 2010a); they can be solved computationally (ABMs) and/or mathematically, using the "theory of complex systems" (techniques from disciplines such as statistical mechanics, dynamical systems and stochastic processes). This kind of mathematics is not based on optimization, but it is particularly suited to study the dynamics, the fluctuations and the link between the individual and the collective behaviour.

The description of Complexity Economics, admittedly, looks so far vague. This thesis aims at giving a concrete example of its great potential. The agents have here bounded rationality, the buyers do not indeed maximize their utility (they would all end up looking for an apartment in the same spot), but they choose probabilistically a neighbourhood on the basis of the utility they expect to have there. Thanks to this assumption, we find a reasonable result in the context of the housing markets: in most locations there is some form of social mixing, for any parameter setting (this was not the case in Alonso *et al.* (1964) and in the models it inspired). The heterogeneity of the agents, who are partitioned in income categories, let us study segregation in a multidimensional way: when some categories abandon

[6]In a nutshell, rational expectations mean that agents in the economy behave as if they knew the stochastic process, which regulates the prices in the future. In other words, they do not make systematic mistakes. Rational expectations were needed to make intertemporal models consistent, but it not clear whether a learning process would converge to them

[7]Many economic models assume a "representative household", a "representative firm", etc.

a location, the other ones may be better mixed in together. If we had used just two income categories we could not have got this result. Furthermore, we are able to study the process of relaxation to equilibrium: a perspective of this work could be to add random shocks to the local intrinsic attractiveness (to model the construction of modern buildings or the deterioration of a neighbourhood) and to compare the time scales of the housing market with those of the structural transformations in the city: we could study the out-of-equilibrium dynamics, which gives additional insights, more than just considering the steady states. Would a city ever be "in equilibrium"?

Summing up, we consider in this thesis a dynamic model of the housing market, with a similar approach as Gauvin *et al.* (2013), but substantially different for what concerns the economic assumptions. Our city is represented by a grid, where there is a fixed number of homogeneous apartments at each site. The home owners decide to put their apartments on sale with an exogenous fixed probability. They decide their reservation offer price by taking into account the market price[8], the competition in the market and the price they paid when they purchased the apartment. The buyers come from the outside of the city and are heterogeneous for their income. They choose where to look for an apartment on the basis of the attractiveness (which has an intrinsic component, due for instance to the presence of amenities, and a social component, depending on the social composition and the popularity of the neighbourhood) of the location and on the level of consumption they expect to afford, given that they have to pay for their accommodation, thus indirectly on the basis of the price they predict to pay. Buyers and sellers are matched through a continuous double auction (CDA)[9], which is modelled through an order book. Successful buyers settle at the desired location, successful sellers leave the city.

The main results of this thesis are summarized below:

- The spots with the highest intrinsic attractiveness (such as the center) are those with the highest real-estate transaction prices and the largest density of rich agents
- Some social mixing is preserved in most locations
- *Ceteris paribus*, the income distribution of the buyers determines the offset of the real-estate price distribution. This is not a trivial result, because if you force a small group of agents with a much higher income to buy a house just in the center, the prices raise as expected in the center, and by a lower but uniform amount in all other locations, independently of their distance from the center. However, a dramatic

[8]Here it is defined as the average price of the real-estate transactions in the neighbourhood

[9]Continuous double auctions are a way to match heterogeneous buyers and sellers, usually in financial markets. For an introduction to CDAs see Appendix A

increase in the prices is related to a larger share of relatively rich buyers more than to a few extremely rich agents. So, in a city with few super-rich and a vast majority of rather poor people the real-estate prices would be closer to the financial resources of the poor. Conversely, in a city inhabited by many fairly affluent individuals, the apartments would be more expensive and the poor could not afford purchasing an apartment anywhere in the city. Notice that the Gini index, a metrics that quantifies the accumulation of wealth by a small share of the population, and thus income inequality, may be similar in the two cases.

- *Ceteris paribus*, the preferences of the buyers determine the steepness of the real-estate price distribution. If agents do not take into account the social composition of the neighborhood (just considering the distance from the center), and they do not care of the prices, the gradient of the distribution is very steep: the prices are extremely high in the center and really low in the outskirts. As a result, the agents are really segregated (the "rich" live in the center, the "middle class" between the center and the periphery and the "poor" in the outskirts), but all categories afford living in the city. If, on the other hand, the agents take into account the popularity of a neighbourhood and look for a good value for money, the outcome is gentrification: the rich and the middle class choose to live also in the periphery, but the poorest abandon the city. The result is that the poorest categories are driven away, but the other ones experience a higher level of social mixing.

- Bubble behaviour may be observed in some locations. A bubble is due to agents who do not want to sell at a lower price than the one they paid; it occurs also when the buyers are constantly more than the sellers. In general, prices are "sticky downwards", i.e. they increase fast, but decrease slowly.

- There is no certainty that markets clear because of some frictions on the offer side (at a given time, not enough sellers may have decided to put their apartment on sale). From the demand side, when there is a strong preference for the center, there is excess supply in the outskirts and excess demand in the center. However, when agents also have a preference for peripheral neighbourhoods, because of their social composition or because they are less expensive, in most locations markets clear.

Chapter 2

Related literature

This chapter presents the most important works which have influenced the present thesis. In particular, in Section 2.1 we present the starting point of the thesis internship (Gauvin *et al.*, 2013). The authors modelled the housing market by focusing on agents heterogenous for their income, and on locations differentiated by their attractiveness. They addressed the issue of segregation from an income point of view: this was a novelty in the literature, since most works on segregation were based on the Schelling model (Schelling, 1971), which considers racial features and not income. The comparison between the assumptions and the results of that model with respect to the present one is performed after the presentation of the present one, in Section 4.4. In general, the thesis dealt with ameliorating Gauvin *et al.* (2013) from an economic point of view, by removing some assumptions and replacing them with other hypotheses which had better economic foundations. Nevertheless, the backbone of the present work is shared between the two models. One important add-on of the present model is the utility function (see Section 4.1.5), which comes from an adaptation of the so-called Alonso model (Alonso *et al.*, 1964), which is presented in detail in Section 2.2. The Alonso one is a model of residential location: the agents decide where to look for an apartment in a city on the basis of their preferences and their budget constraint, given that they work in the central business district (CBD), at the center of the city. They face a tradeoff between living in small apartments in the center, where the transport cost is lower and so the net income available for consumption is higher, and living in big apartments in the periphery, with less consumption of other goods but housing. An approach we tried to use to reproduce the results of Alonso (and its adaptation) from an agent-based model perspective is the same as that described in Lemoy *et al.* (2010), which is summarized in Section 2.3. The authors succeeded in discretizing the Alonso model and finding its main results. However, we had major problems in reproducing their results from a quantitative point of view, and we found some problems with their assump-

tions. The discussion about our attempts to reproduce the results of Lemoy *et al.* (2010) is presented in Section 5.2.1. Finally, we present in Section 2.4 other contributions which did not directly influence the present work, but which we contrast with our approach.

2.1 Housing market agent-based model

Gauvin *et al.* (2013) consider a dynamic model of the housing market, characterized by a "city", differentiated by a field of attractiveness, both "intrinsic" and "social", and by the agents heterogeneous for their income. Buyers coming to the city have to make decisions about where to look for an apartment, and sellers leaving the city have to sell their apartments. The spatial income distribution and the prices result from the interactions between the buyers and the sellers; the market mechanism determines the level of segregation, if it exists. Here we present the model and the results, which shall be contrasted with those in chapter 4. When the assumptions of Gauvin *et al.* (2013) are shared with the present model, we refer to Section 4.1 for further details. In Section 2.1.1 we present the model in Gauvin *et al.* (2013), in Section 2.1.2 we show the main results.

2.1.1 The model

The model considers both a temporal and a spatial aspect, features which are not common in many economic models (such as Alonso *et al.* (1964), see Section 2.2). Time is indexed by t, the city can be thought of as a grid of size L. The goods on the market are identical, N apartments are available at all locations. Γ buyers are coming any time step to the city, from an external "reservoir"; housed agents decide to put their apartment on sale if a Bernoulli trial with probability α is successful (in which case they become "sellers"). All these assumptions are shared with the present work, and are explained at lenght in Sections 4.1.1 and 4.1.2. All the agents are endowed with an income, which Gauvin *et al.* (2013) assume to be related to a "willingness-to-pay" (henceforth WTP), the maximum price the buyers are willing to pay for an apartment. The WTP determines also the offer prices. There are K income categories, indexed by $k \in \{0, ..., K-1\}$, and their WTPs are $P_k \leq P_{k+1}$. The agents are uniformly distributed between the WTP categories. Each location X is characterized by an intrinsic attractiveness $A^0(X)$; however, the attractiveness of location X is perceived by k-agents in different ways, taking into account the social composition of the neighbourhood. Since the social composition is time-dependent, the attractiveness as perceived by k-agents is $A_k(X, t)$. Gauvin *et al.* (2013) assume that the agents prefer to live with agents at least as rich as them (in line with empirical research, see the introduction of Gauvin *et al.* (2013),

Tiebout (1956) or Ioannides and Zabel (2003)). Thus, the attractiveness evolves:

$$A_k(X, t + \delta t) = A_k(X, t) + \omega \delta t \left(A^0(X) - A_k(X, t)\right)) + \epsilon \delta t \, v_{k_>}(X, t) \quad (2.1)$$

with $v_{k_>}(X, t) = \sum_{k' \geq k} v_{k'}(X, t)$. Here $v_k(X, t)$ is the number of successful k-buyers, which influences cumulatively the "quality" of a location. For more details on the concept of attractiveness, see Section 4.1.4.

On the market at location X and time t there is a number of k-buyers which depends probabilistically on the attractiveness of X. The probability to choose X is $\pi_k(X, t) = \frac{A_k(X,t)}{\sum_{X' \in \Omega} A_k(X',t)}$. The number of sellers depends mostly on the parameter α. The prices proposed by the buyers are P_k, those proposed by the sellers are given by $P_k^o(X, t) = P^0 + (1 - \exp(-\xi \bar{A}(X, t))) \, P_k$: they are modulated both by the WTP of the seller and by the intensity of the demand, which is here assumed to be proportional to the average attractiveness $\bar{A}(X, t) \equiv \frac{1}{K} \sum_{k=0}^{K-1} A_k(X, t)$. Then matching occurs following some matching rule, which is not detailed here, buyers settle in, sellers leave the city, the attractiveness is updated accordingly. At the beginning of next time step, the same dynamics is replicated.

A thing which has been omitted so far is that the initialization is such that the city is empty, and subsequently agents buy empty apartments. The problem is that the matching rule is so that buyers would choose to buy an apartment from a seller rather than an empty one which would cost less. This does not change qualitatively the behaviour of the model in the steady state, but is certainly problematic from an economic point of view.

2.1.2 Main results

Gauvin *et al.* (2013) show that only few effective parameters determine the behaviour of the model:

- $\widetilde{\xi} \equiv \xi A^0_{max}$, where A^0_{max} is the maximum intrinsic attractiveness. This parameter determines whether the system is in a "frozen state", i.e. sellers sell at a price that buyers cannot afford, and there are no transactions. This may be related to a bubble of the housing market.

- $\frac{\Gamma}{\alpha}$, i.e. the ratio between the incoming rate and the leaving rate. This determines whether the market is "saturated", i.e. there are more buyers than available apartments

- $\eta \equiv \frac{\epsilon \gamma}{\omega \langle A^0 \rangle}$ is the "social influence parameter", i.e. it quantifies the strength of the social influence, and it determines whether or not segregation occurs.

Gauvin *et al.* (2013) analyse the model through mathematical methods. This is possible because they assume a "non-saturated" equilibrium, where buyers who can afford goods at a given location always succeed in getting them. This assumption is due to the fact that unsuccessful buyers search again in the city at subsequent time steps, so it is difficult to compute the actual number of buyers, which determines $v_k(X,t)$. Gauvin *et al.* (2013) take the continuum limit and consider the set of partial differential equations:

$$\partial_t A_k(X,t) = \omega\,(A^0(X) - A_k(X,t)) + \epsilon\, v_{k_>}(X,t) \tag{2.2}$$

$$(1-\alpha)\partial_t u_k(X,t) = v_k(X,t) - \alpha\, u_k(X,t) \tag{2.3}$$

$$\partial_t \rho_k(X,t) = -\rho_k(X,t) \;+\; \frac{\gamma}{K}\, L^2\, \pi_k(X,t) \;+\; \pi_k(X,t) \sum_{X'\in\Omega} \bar{v}_k(X',t) \tag{2.4}$$

with $\rho_k(X,t) = v_k(X,t) + \bar{v}_k(X,t)$, where $\bar{v}_k(X,t)$ is the density (in the continuous limit "numbers" become densities) of unsuccessful buyers. In (2.3) $u_k(X,t)$ is the density of housed agents at location X and time t. Gauvin *et al.* (2013) consider only the stationary state of the model, such that all dynamical variables become constant in time. They obtain:

$$v_k^*(X) = \alpha u_k^*(X), \tag{2.5}$$

$$A_k^*(X) = A^0(X) + \frac{\epsilon}{\omega} v_{k_>}^*(X) \tag{2.6}$$

$$\rho_k^*(X) = \pi_k^*(X) \sum_{X'\in\Omega} \rho_k^*(X'). \tag{2.7}$$

They find that the mean density of k-agents is constant throughout the space. By denoting with $\Omega_k \equiv \{X \in \Omega | \bar{v}_k^*(X) = 0\}$ the set of locations where k-agents can afford an apartment, they define a *WTP threshold* $P_c^* = P_{\bar{k}}$ such that only agents with a WTP at least as high as $\bar{k}$ can buy a good anywhere in the city. Their main results concern the conditions to have segregation:

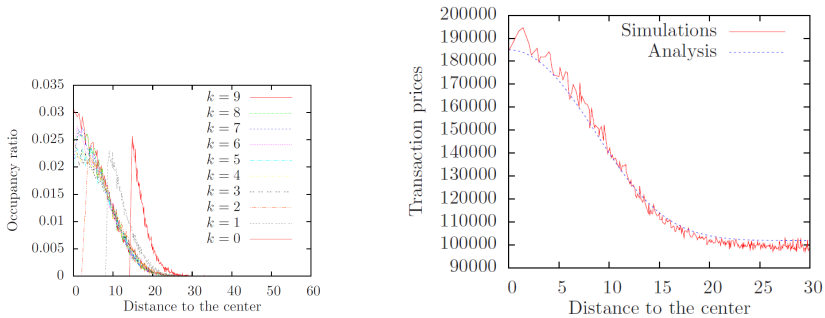
- If $\xi A^0_{max} > \log \frac{P_{K-1}}{P^0}$, the market is frozen, i.e. sellers sell at a higher price than the richest category can afford
- One can write the following equation for the WTP threshold:

  $$\log \frac{P_c^*}{P^0} - \widetilde{\xi} = \eta\, \frac{\widetilde{\xi}}{2} \left(1 - \left(\frac{P_c^* - P_0}{\Delta}\right)^2\right)$$

  It can be proven that P_c^* is an increasing function of η: if the effective social influence is strong enough, $P_c^* > P_0$ and at least category 0 is segregated from some location.

- At all other locations, some social mixing is preserved.

These results can be better understood with the numerical simulations performed in Gauvin *et al.* (2013). The authors take the intrinsic attractiveness to be a Gaussian decreasing from the center; the preference for the center is in line with the Alonso model. In Fig. 2.1 the main results are shown: the poorest categories are segregated away from the center (see Fig. 2.1a), i.e. the social influence is strong enough to raise the prices. As it can be seen in Fig. 2.1b the analytical results match well the results from the simulation.



(a) (b)

Figure 2.1: Density of agents and prices as a function from the distance from the center. At the left it is apparent that the poorest categories are segregated away from the center; at the right it is possible to see that the prices decrease from the center. From Gauvin *et al.* (2013)

A final interesting result in Gauvin *et al.* (2013) is the comparison with real data. The authors use part of the B.I.E.N. database, managed by the "Chambre des Notaires de Paris", which records 300000 real-estate transactions for Paris and the Ile-de-France region for the years 1990-2003. More details on the database are provided in Section 3. Gauvin *et al.* (2013) model the intrinsic attractiveness through the average price level in each "arrondissement" (see Figure 2.2). They compare the results from the simulations with the real data and find that the model reproduces general trends on price distribution and income segregation.

2.2 Alonso model

This presentation of the basic model of residential choice, which is the backbone of most neoclassical urban economic theory, follows mostly Fujita (1989) and is a slightly modified version of Alonso *et al.* (1964).

The pioneer of land use theory is Von Thünen, with his theory on the "isolated state" (Von Thünen, 1826). By managing an agricultural holding

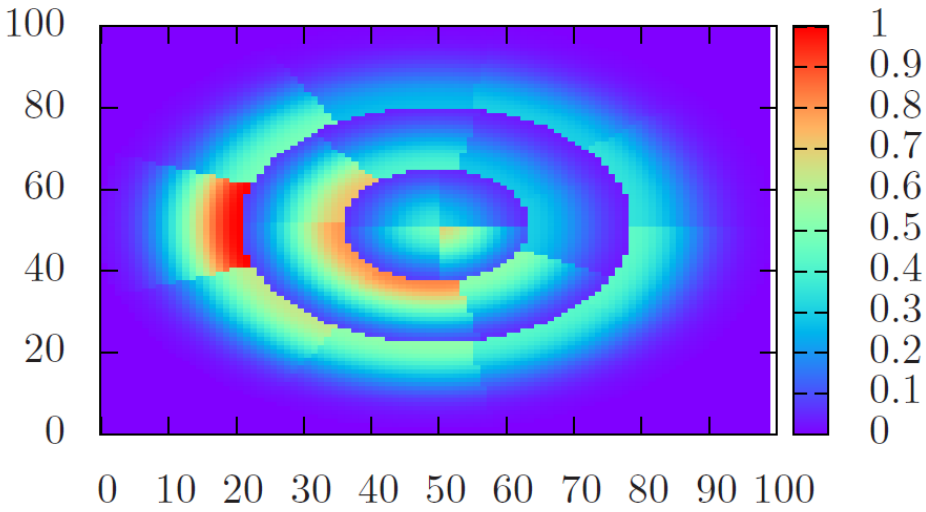


Figure 2.2: The city of Paris is divided in its 20 arrondissements. Only transaction prices in year 1994 are used. Then, the highest transaction prices were in the 16th arrondissement, which is assumed to be the one where the intrinsic attractiveness is the highest. The authors assume that the attractiveness decreases, per arrondissement, with the distance from the center. From Gauvin *et al.* (2013)

he had collected a lot of data which he could use to develop his theories. Von Thünen considered a city, which he assumed to be at the center of an "isolated state". Further from the city borders there is "wilderness". The space is uniform, i.e. there are no mountains, rivers, roads and the agricultural yield is the same everywhere. The inhabitants of the city are farmers who trade their commodities in a market located in the center: based on the distance of their land to the center, they have to travel every day to the market. The fundamental equation is:

$$R = Y(p - c) - YFm \tag{2.8}$$

In the above equation:

- R is the agricultural rent per unit surface
- Y is the agricultural yield per unit surface
- p is the market price per unit of commodity
- c is the production cost per unit of commodity
- F is the transport cost per unit of commodity per unit of distance. It depends on the type of commodity.

- m is the distance between the agricultural land and the center

The farmers are rational and produce the goods that maximize their profits. Hence, the city is organized in concentric "rings" (Von Thünen's rings): closest to the center are produced dairy products and perishable vegetables. Their hidden transport cost is extremely high, as they need to be sold quickly. The second ring is characterized by timber, which is heavy to transport and so cannot be produced far from the market. The third ring hosts non perishable crops, such as grain, which are lighter than wood. Finally, the fourth ring locates ranching: the animals need a lot of room and are extremely easy to transport. At the edge of the city it is not convenient anymore to produce anything: there is wilderness because the utility of colonizing that area is null.

Von Thünen theory anticipated the "marginal revolution" which would take place about fifty years later. For his contribution to neoclassical economic theory, Samuelson (1983) puts Von Thünen in the Pantheon of the great economists of all time, with Leon Walras, John Stuart Mill and Adam Smith.

As Fujita (1989) notices, standard microeconomic theory cannot be readily applied to a context in which economic agents choose between apartments. Actually, agents choose only one apartment, so their indifference curves have to be convex (see Fig. 2.3). Finding competitive equilibria with convex indifference curves may be impossible, so the housing market cannot be dealt with in a standard way (for further explanation on this point in the context of the present work see Section 4.1.7). Alonso *et al.* (1964) applied Von Thünen theory of land use to an urban context (urban land use theory). The city is essentially the same as the "isolated state", with a Central Business District (CBD) such that all inhabitants work at the center (the CBD is similar in spirit to the central market where the farmers in Von Thünen's model trade their commodities). Since the distance to the CBD is the only spatial feature of the Alonso model, the city can be considered as one-dimensional.

The agents are endowed with a utility function which depends on the lot size ("surface") of the apartment s and on the consumption of a composite good z. The agents face the following budget constraint:

$$Y - T(r) \geq z + R(r)s \tag{2.9}$$

In the above equation Y is their income, $T(r)$ is the transport cost to the center (with $T'(r) > 0$), the price of the composite good z is assumed without loss of generality to be unitary, $R(r)$ is the equilibrium price per unit surface at distance r, which will be determined from the equilibrium properties. One can define $I(r) = Y - T(r)$ as the net income. $U(z, s)$ is a utility function such that both good are essential, that is there are no corner

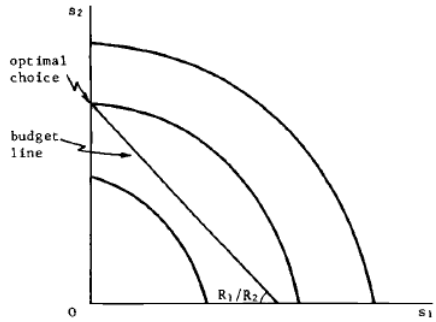


Figure 2.3: Indifference curves between the surface of two apartments. Since the agents choose to live only in one apartment, the curves have to be convex. From Fujita (1989)

solutions. A possible choice for the utility function is the log-linear:

$$U(z, s) = \alpha \log z + \beta \log s \tag{2.10}$$

Hence, agents maximize:

$$\max_{z,s} U(z, s) \text{ s.t. } Y - T(r) \geq z + R(r)s \tag{2.11}$$

This is called the basic model of residential choice. Agents face a tradeoff between living in small apartments in the center, where the transport cost is lower and so the net income available for consumption is higher, and living in big apartments in the periphery. Eq. (2.11) could be solved by simply considering the Lagrangian, but it is more instructive to use an alternative approach, that of bid-rent curves.

We define the bid-rent $\psi(r, u)$ as the maximum price an agent accepts to pay to live at distance r from the center to attain a utility level u. Formally:

$$\psi(r, u) = \max_{z,s} \{\frac{Y - T(r) - z}{s} | U(z, s) = u\} \tag{2.12}$$

The geometric meaning of $\psi(r, u)$ is the slope of the budget constraint when it is tangent to the indifference curve u. One can write the budget constraint (2.9) at the optimum as

$$Y - T(r) = z + Rs \tag{2.13}$$

where R is now a parameter (see Fig. 2.4 for an example with (2.10) as the utility function). We define $Z(s, u)$ as the function obtained inverting $U(z, s) = u$, and $S(r, u)$ the bid-max lot-size, i.e. the surface of the apartment bought at price $\psi(r, u)$. For general $U(z, s)$ one writes that the

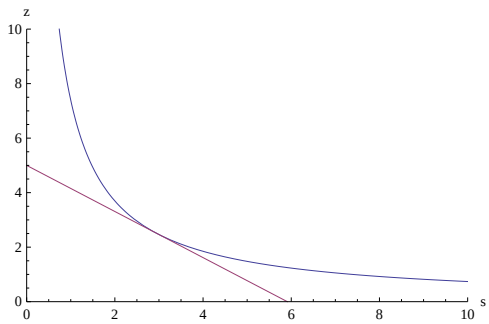


Figure 2.4: Bid-rent curve (in blue) and budget constraint (in purple), with R s.t. the curves are tangent. The parameter values are $\alpha = 0.5$, $\beta = 0.5$, $Y = 15$, $T(r) = 10$, $u = 1$. We choose (2.10) for the utility. The indifference curve has equation: $z^\alpha s^\beta = e^u$, i.e. $z(u,s) = e^{u/\alpha} s^{-\beta/\alpha}$. One can find R by equating the indifference curve with (2.13) and by equating their derivatives: $-\frac{\partial z(u,s)}{\partial s} = R$.

marginal rate of substitution $-\frac{\partial Z(s,u)}{\partial s}$ equals the slope of the indifference curve:

$$-\frac{\partial Z(s,u)}{\partial s} = \frac{Y - T(r) - Z(s,u)}{s} = \psi(r,u) \tag{2.14}$$

From the above equation it is possible to obtain $Z(s,u)$; from the indifference curve then one can find $S(r,u)$, and finally $\psi(r,u)$ from (2.14). Let us do it for the case of the log-linear in Eq. (2.10). One gets $Z(s,u) = e^{u/\alpha} s^{-\beta/\alpha}$, then $-\frac{\partial Z(s,u)}{\partial s} = \frac{\beta}{\alpha}\frac{z}{s}$ and so from (2.14) $\frac{\beta}{\alpha}\frac{z}{s} = \frac{Y-T(r)-z}{s}$. s cancels out and one gets $z = \alpha(Y - T(r))$, as it is standard in the case of Cobb-Douglas utility functions (the optimal consumption of a good is given by the weight given to that good in the utility function times the available income divided by the price of that good). Then $S(r,u) = \left(ze^{-u/\alpha}\right)^{-\alpha/\beta}$, i.e.

$$S(r,u) = \alpha^{-\alpha/\beta} \left(Y - T(r)\right)^{-\alpha/\beta} e^{u/\beta} \tag{2.15}$$

Finally, the bid-rent curve is:

$$\psi(r,u) = \beta\alpha^{\alpha/\beta} \left(Y - T(r)\right)^{1/\beta} e^{-u/\beta} \tag{2.16}$$

By using traditional microeconomics (a standard reference is Mas-Colell *et al.* (1995)) we can find general properties (i.e. for general $U(z,s)$ and $T(r)$) for the bid-rent curve and for the bid-max lot-size. Let us consider the maximization problem:

$$\max_{z,s} U(z,s) \text{ s.t. } z + Rs \leq I \tag{2.17}$$

In the above equation I is the net income and R is a parameter. Let us define the Marshiallian demand for land as $\hat{s}(R, I)$ and the indirect utility as $V(R, I)$. At the optimal level the following equations hold: $R = \psi(r, u)$, $I = Y - T(r)$. So we have:

$$S(r, u) = \hat{s}\left(\psi(r, u), Y - T(r)\right) \tag{2.18}$$

$$u = V\left(\psi(r, u), Y - T(r)\right) \tag{2.19}$$

Let us now consider the minimization problem:

$$\min_{z,s} z + Rs \text{ s.t. } U(z, s) = u \tag{2.20}$$

Now let us define $\tilde{s}(R, u)$ as the Hicksian demand and $E(r, u)$ as the expenditure function. Again, given that $R = \psi(r, u)$, $I = Y - T(r)$, we have:

$$S(r, u) = \tilde{s}\left(\psi(r, u), u\right) \tag{2.21}$$

$$Y - T(r) = E\left(\psi(r, u), u\right) \tag{2.22}$$

Finally, we can use the tools of microeconomics to prove some results. First of all, let us see how $S(r, u)$ and $\psi(r, u)$ vary as r changes. It can be proven that:

Proposition: The bid-rent decreases and the bid-max lot-size increases with the distance from the center, i.e. $\frac{\partial \psi(r,u)}{\partial r} < 0$ and $\frac{\partial S(r,u)}{\partial r} > 0$.

Proof: $\frac{\partial \psi(r,u)}{\partial r} = \frac{\partial}{\partial r} \max_s \frac{Y - T(r) - Z(S(r,u),u)}{S(r,u)}$. Thanks to the envelope theorem, one gets $\frac{\partial \psi(r,u)}{\partial r} = -\frac{T'(r)}{S(r,u)} < 0$. Furthermore from (2.21) $\frac{\partial S(r,u)}{\partial r} = \frac{\partial}{\partial r} \tilde{s}\left(\psi(r, u), u\right) = \frac{\partial \tilde{s}}{\partial R} \frac{\partial \psi}{\partial r} > 0$, since $\frac{\partial \psi}{\partial r} < 0$ and $\frac{\partial \tilde{s}}{\partial R} < 0$ thanks to the properties of the Hicksian demand.

Intuitively, the above proposition is due the fact that if the net income is reduced, as it happens farther from the center due to the transport cost, to keep the same level of utility the agents have to substitute land for general consumption.

Another useful property is that $V\left(\psi(r, u), Y - T(r)\right)$ is constant throughout the city: so the bid-rent curve $\psi(r, u)$ defines indifference curves in the urban space: all agents have the same utility but pay different prices.

It is interesting to consider two groups, or two income categories, with incomes Y_1 and Y_2, with utility levels u_1 and u_2. Their bid-rent functions will be $\psi_1(r, u_1)$ and $\psi_2(r, u_2)$. If one assumes that an English auction is organized to allocate the available apartments between the agents, only the agents with the highest bid-rent curve will be able to live at distance r. In order to have a spatial distribution with a group closer to the center and

another group farther from the center, the two curves need to intersect. This happens provided that one curve is steeper than the other one at the intersection x, i.e. if $-\frac{\partial \psi_i}{\partial r}\Big|_{r=x} > -\frac{\partial \psi_j}{\partial r}\Big|_{r=x}$ group i is closer to the center than group j. If one identifies category i with the "poor" with income Y_P and category j with the "rich" with income Y_R, $Y_R > Y_P$, one gets for the bid-max lot-size:

$$S_P(r, u_P) = \hat{s}\,(R, Y_P - T(r)) < \hat{s}\,(R, Y_R - T(r)) = S_R(r, u_R) \tag{2.23}$$

The above inequality follows from the properties of the Marshallian demand and from the assumption of the normality of land, as it is empirically supported (Fujita, 1989). Then:

$$-\frac{\partial \psi_P(r, u_P)}{\partial r} = \frac{T'(r)}{S_P(r, u_P)} > \frac{T'(r)}{S_R(r, u_R)} = -\frac{\partial \psi_R(r, u_R)}{\partial r} \tag{2.24}$$

The bid-rent curve for the poor is always steeper than that for the rich.

Summing up, the Alonso model considers a monocentric city with the poor agents living closer to the center, with small apartments and a lower transport cost, and rich agents living in the periphery, with bigger apartments and a higher transport cost. This asymmetry is due to the additivity of the transport cost. This is consistent with the empirical evidence in American cities such as Detroit. The "European" case with rich people living closer to the center can be obtained if one considers in the transport cost the "value of time", which is supposed to be higher for the rich (because the opportunity cost of spending time traveling is higher for the rich).

2.3 An Alonso-type agent-based model

Lemoy *et al.* (2010) build an agent-based model out of the standard Alonso model (more information about it is contained in the PhD thesis of Remy Lemoy, see Lemoy (2011)). Since the model described in Section 2.2 was not thought to be based on discrete individual interactions, they need to make many modifications to get the same results. This ABM is much different in spirit to that proposed by Gauvin *et al.* (2013) (explained in Section 2.1): individuals behave maximizing their utility function and the goal is to reproduce the results of the Alonso model. Lemoy *et al.* (2010) lie in between a methodology based on ABMs and a neoclassical one, as it will be apparent by the end of this section.

As in Gauvin *et al.* (2013) the space is a grid, and the cells can be inhabited by one or more agents, or be empty. In line with Alonso *et al.* (1964) they assume that empty locations are used for agriculture, since

the rent paid by the agents is not higher than the agricultural rent. They initialize the model by considering a fixed population made up of N agents, which can be split in two income groups. At the beginning, all prices are set to the agricultural rent: $p_0 = R_a$. All the agents occupy a surface which is the optimal one in the Alonso model:

$$s = \beta \frac{Y - tx}{p_0} \tag{2.25}$$

In the above equation all the quantities are defined as in Section 2.2. Once s is known, the agents compute their utility:

$$U = \alpha \log(Y - tx - p_0 s) + \beta \log s \tag{2.26}$$

At time step n one agent and one location are chosen randomly. The agent computes the surface she would get at that location:

$$s = \beta \frac{Y - tx}{p_n} \tag{2.27}$$

Here p_n is the price of the chosen location at time step n. Using (2.26) the agent computes the utility she would get at that location. If $\Delta U > 0$, the agent moves, proposing a bid:

$$p_{n+1} = p_n \left(1 + \epsilon \frac{s_{occ}}{s_{tot}} \frac{\Delta U}{U}\right) \tag{2.28}$$

In the above equation ϵ is a parameter determining the price increase, s_{occ} is the occupied surface at the chosen location, s_{tot} is the total surface. Eq. (2.28) means that prices increase when agents choose to go to that location. It should be noted that it may be that after the movement $s_{occ} + s > s_{tot}$ (here s is the surface occupied by the incoming agent). If this is the case Lemoy *et al.* (2010) assume that one agent living at that location is chosen randomly and sent to the "hotel": subsequently she will apply for another apartment at some location.

Finally, the authors assume that cells which are not completely full see their price decrease, according to the formula:

$$p_{n+1} = p_n - (p_n - 0.9 \cdot R_a) \frac{\Delta U}{T_p} \frac{s_{free}}{s_{tot}} \tag{2.29}$$

Here $s_{free} = s_{tot} - s_{occ}$. It is not clear how often prices decrease though: it could be that after all agents try to move the prices are decreased, or that this happens all the times agents try to move, or also that this happens only after each successful move. It is not specified in the paper. What is more, it never happens that a location is completely full, so prices decrease everywhere. There are some other problems with the model of Lemoy *et al.* (2010):

- Given the rule for the occupied surface of any agent $s = \beta \frac{Y - tx}{p_n}$ (which comes from the optimization problem) there is nothing that prevents s_{occ}, the occupied surface, to become greater than s_{tot}, when prices decrease (for instance because no other agent is going there).
- As a rule to decrease the price, they write use Eq. (2.29). Why $0.9 \times R_a$? This can make the price lower than the agricultural rent.
- In some locations where the price increases, s becomes smaller than 1, and so given the log-linear utility function $\log(s)$ gives a negative contribution. This never happens in standard economic theory because goods are countable. Also s should be like that? One has to be really careful with the choice of the parameters, since in most regimes one would get a negative utility!

The results of Lemoy *et al.* (2010) are summarized in Fig. 2.5. At the top the blue line represents the analytical curve $1/S(r, u)$, with $S(r, u)$ given by (2.15). The red dots come from the simulations, and fit remarkably the analytical curve. On the left there is the bid-rent curve $\psi(r, u)$, given by (2.16). On the right there is the optimal surface, given again by (2.15). In both cases the results from the simulations fit remarkably well the analytical predictions.

The authors consider also two income categories, the "rich" and the "poor": in line with the Alonso model, they find that the poor locate closer to the center, whereas the rich locate farther from the center. As with the one category case, rents decrease and surface increases as one moves away from the center.

Finally, there is the issue of stability. The authors claim that for some parameter values the system does not evolve towards a steady state where all the agents have the same utility (all the previous results from the simulations were taken in the steady state). Furthermore, it is not clear why the specific rules described in this section should approximate the Alonso solution, and whether another setting for the ABM could work as well.

2.4 Other papers

A very interesting contribution to the topic of segregation comes from Grauwin *et al.* (2009). The authors use tools from equilibrium statistical mechanics to solve analytically a simplified version of the Schelling model (Schelling, 1971). They link the individual behaviour, based on the utility function, to the global behaviour. They show that if there is no cooperation between the agents, coordination failures and the outcome is inefficient (Bouchaud, 2013).

They consider a city defined on a grid, composed of Q cells, where each cell hosts H apartments (the setting is the same as in Gauvin *et al.* (2013),

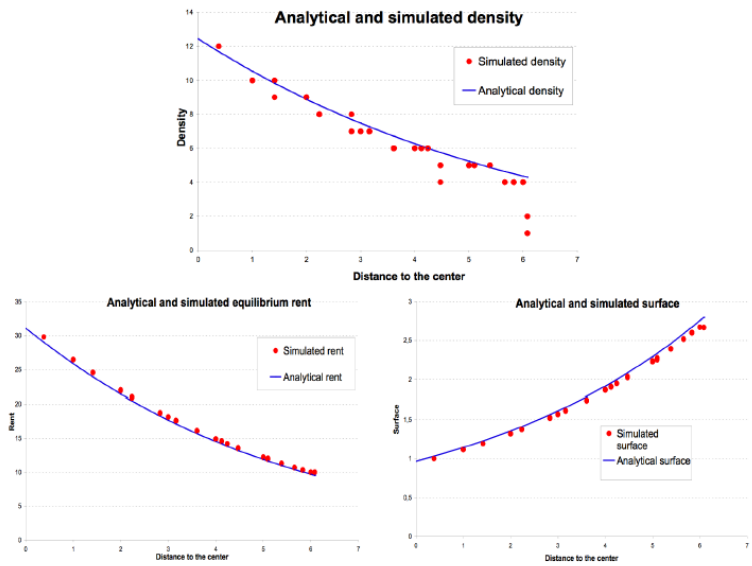


Figure 4.1: Comparison between the results of the agent-based model and the analytical results: density, rents and surface as functions of the distance to the center.

Figure 2.5: (*Top*): Density (number of agents divided by s_{tot}) as a function of the distance to the center. (*Left*): Average rent as a function of the distance to the center. (*Right*): Average surface as a function of the distance to the center. From Lemoy *et al.* (2010)

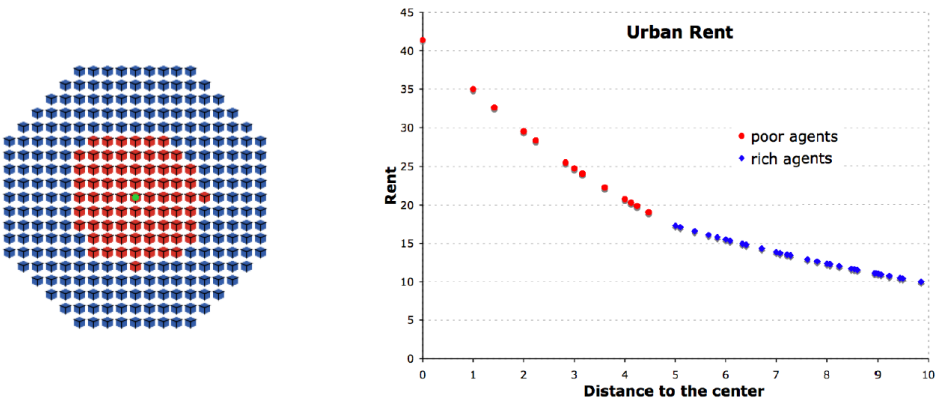


Figure 2.6: (*Left*): NetLogo's view of the spatial distribution of "rich" and "poor" agents. Here the poor agents are represented in red and the rich ones in blue (*Right*): Average rent as a function of the distance to the center. Here there is no comparison with the analytical results because it is not possible to solve the model with two categories. From Lemoy *et al.* (2010)

with $Q = L^2$ and $H = N$). The housed agents in cell q are $n_q \leq H$, $\rho_q = \frac{n_q}{H}$ is the density. As a first approximation, they just consider one kind of agents. The agents are described by their utility function $u(\rho_q)$: as in the Schelling model, agents are interested in how many like agents live in the same location. A configuration $x = \{\rho_q\}$ describes a coarse-grained approximation of the system, where only the densities in each cell are considered. The global utility is:

$$U(x) = H \sum_q \rho_q u(\rho_q) \tag{2.30}$$

Grauwin *et al.* (2009) use a link function to link the individual and the collective behaviour. In particular, they search for a *state function* $L(x)$ which has the property $\Delta u = \Delta L$: the individual difference in utility is reflected on a global level. The following function satisfied the above property:

$$L(x) = \sum_q \left(\sum_{m=0}^{n_q} u\left(\frac{m}{H}\right) \right) \tag{2.31}$$

They define the function:

$$C(x) = (1 - \alpha)L(x) + \alpha U(x) \tag{2.32}$$

The difference ΔC is the gain associated to a move by an agent: a weight $1 - \alpha$ is put on his individual utility (selfish behaviour) and a weight α is put on the global utility (altruistic behaviour). Thus α can be thought of as a cooperation parameter (the higher is α, the more the agents cooperate) or as a Pigouvian tax (Pigou, 2013), which internalizes the externalities. The authors consider a Markov Chain described by the transition probability:

$$W(x \to y) = \frac{1}{1 + e^{-\Delta C/T}} \tag{2.33}$$

In the above equation T is the noise parameter which represents the amount of noise in the decisions by the agents, as it is standard in the context of stochastic models. Since (2.33) satisfies detailed balance, it is possible to find the steady state probability, which is simply:

$$P(x) = \frac{1}{Z} e^{C(x)/T} \tag{2.34}$$

Using standard methods about phase separations, the authors find analytically (and validate through simulations) where it is located the phase transition between an homogeneous phase, with all cells sharing the same density ρ_0, and an asymmetric phase, with cells having densities ρ_1 and ρ_2. Since the utility function of the agents is chosen such that the highest utility level is attained for ρ_0, the highest utility is obtained in the homogeneous

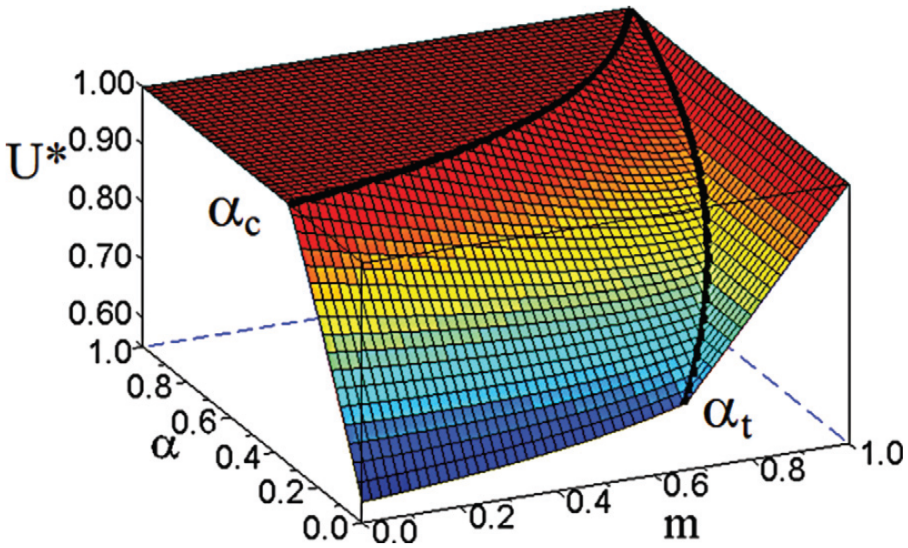


Figure 2.7: Utility as a function of the cooperation parameter α and of a parameter of the utility function m. Focusing on α, it is possible to see how for higher levels of cooperation the utility increases. In particular, for $\alpha = \alpha_c$, there is a phase transition which defines a phase of maximum utility, that is of optimal social mixing. From Grauwin *et al.* (2009)

phase (the non-segregated one. As it is shown in Fig. 2.7, if the cooperation parameter is not strong enough the utility level is not optimal, i.e. the outcome is not efficient.

An application of Grauwin *et al.* (2009) to an urban context is in Lemoy *et al.* (2011). The authors generalize Grauwin *et al.* (2009) to m categories and, by making a connection between utility and chemical potential, find a set of conditions that can be solved numerically. In a nutshell, they find that with two income categories, with one category richer than the other, provided that there is not too much noise the statistical mechanics approach coincides with the results from Alonso *et al.* (1964): the poor are located in the center, the rich are located in the periphery.

We did not decide to use the approach in Lemoy *et al.* (2011) for our model of the housing market mostly for one reason: the utility functions need to be symmetric, i.e. it must hold:

$$\frac{\partial u_{qi}}{\partial \rho_{qj}} = \frac{\partial u_{qj}}{\partial \rho_{qi}}, \forall i, j, i \neq j \tag{2.35}$$

This is not the case for the assumptions in Gauvin *et al.* (2013) (see Sections 2.1.1 and 4.1): the poorest category has a higher utility if the density of people from the richer categories is higher, but the reverse is not true. Furthermore, it is hard to solve in practice these equations for a general

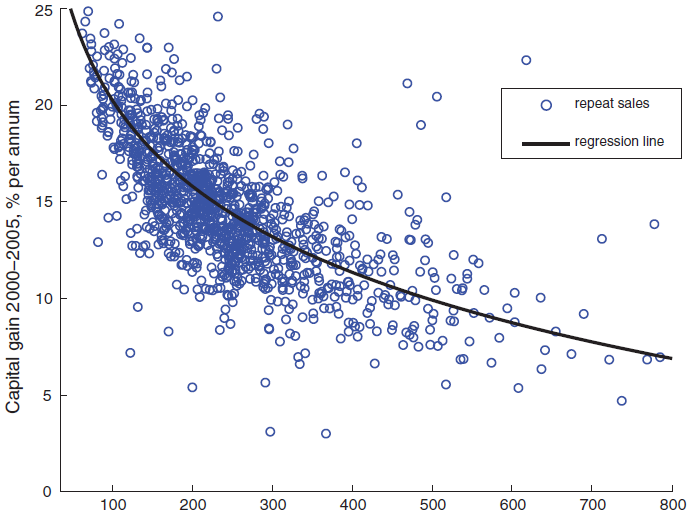


Figure 2.8: Capital gains (%) as a function of the houses' values. As it is apparent from this plot, houses whose value is lower were most appreciated in the period 2000-2005. From Landvoigt *et al.* (2015)

intrinsic attractiveness. Nevertheless, the approach of Grauwin *et al.* (2009) and Lemoy *et al.* (2011) is very interesting from a methodological point of view.

A very recent work on the housing market which had great impact and has been published on one of the most prestigious reviews in economics, the *American Economic Review*, is Landvoigt *et al.* (2015). The starting point is the empirical observation that, in the context of the housing market of San Diego, the houses which appreciated most were those with a lower absolute value, i.e. the percentage capital gains were higher for houses with a lower capital value (see Fig. 2.8).

Landvoigt *et al.* (2015) consider an *assignment model*: the assignment model is an optimization technique to match two groups. For instance, it is standard in the context of labour economics, where workers are matched with jobs. In this case, buyers are matched with houses of a certain quality, which is indexed by $h \in [0, 1]$. h plays the same role as the intrinsic attractiveness. The distribution of h is given by the cumulative distribution function (cdf) $G(h)$. The price function $p(h)$ is determined endogenously through the optimization. The agents are characterized by their wealth w (the cdf for the wealth is $F(w)$) and by their utility function over general consumption c and housing quality h, $u(c, h)$: notice that this utility function is almost the same as in the Alonso model. The individuals face a budget constraint: $c + p(h) \leq w$. A competitive equilibrium is defined as a

situation where markets clear, and is given by the condition:

$$\Pr\left(h^*(p,i) \leq h\right) = G(h) \tag{2.36}$$

In the above equation i denotes individual households. The meaning of (2.36) is that the demand of houses of level at least h is the same as the available density of such houses, given by $G(h)$. The assignment of houses to households is performed through an assignment function $w^*(h)$, such that:

$$F(w^*(h)) = G(h) \tag{2.37}$$

$w^*(h)$ assigns a house quality a wealth level. By writing $w^*(h) = F^{-1}\left(G(h)\right)$ it is possible to find analytically the assignment function. If F and G are chosen appropriately, the result can be obtained in closed form. Landvoigt *et al.* (2015) complicate this simple model by assuming that all agents face an intertemporal optimization problem and find numerically the solution. They calibrate the model with the data and show that the fit is good: this way, they characterize the housing market by taking into account individual characteristics, which they also study through econometric techniques.

Whereas this work is remarkable from many points of view, it neglects heterogeneities and interactions, focusing solely on a "representative agent" (for critics on the concept of the representative agent see Kirman (1992)), and it does not fully consider the spatial and temporal aspects, which are crucial to our work. Moreover, it assumes that individuals are rational and solve an intertemporal optimization problem, which is not consistent with much literature on behavioural economics. Finally, the large number of parameters lets the model potentially fit any dataset and the extremely difficult optimization problem prevents from fully understanding the process at work: these are common critiques to agent-based models, but also apply to this neoclassical economics case.

There are some agent-based models (ABMs) similar to Gauvin *et al.* (2013) and to the work presented in this thesis, and it is convenient to consider them.

Feitosa *et al.* (2008) propose an ABM which shows how segregation can emerge even if one considers the simplest setting with the minimal number of parameters. They start from an empty city where all locations are characterized by a uniform "quality" Q and by their distance to the center D. Agents, heterogeneous for their income, come to the city in a staggered fashion and decide where to settle. Their utility function is:

$$U(X) = \left(\frac{1}{D}\right)^{1-\alpha} Q^{1+\alpha} \tag{2.38}$$

The authors assume that Q increases as the richest agents go to location X. So Q assumes the same meaning as the social component of the attrac-

tiveness in Section 4.1.5, whereas $1/D$ can be thought of as the intrinsic component of the attractiveness, which is assumed in Gauvin *et al.* (2013) and in this thesis to be decreasing from the center. The main finding in Feitosa *et al.* (2008) is that for $\alpha < 0$ (more weight given to the distance from the center) the rich segregate in the center, whereas for $\alpha > 0$ the rich segregate in some area, a "hotspot", where a remarkable number of rich agents settled and increased the quality Q.

This ABM is much simpler than Gauvin *et al.* (2013) and the one proposed in the present thesis, but it is interesting because it highlights the stylized facts of segregation. Anyway, there is a major difference: Eq. (2.38) is such that the social component of the attractiveness is multiplicative, whereas in Gauvin *et al.* (2013) and in the present thesis it is additive. Thus, in Feitosa *et al.* (2008) the social component exacerbates the relative importance given to locations, whereas in Gauvin *et al.* (2013) the attractiveness field is flattened by the social component.

The approach of Gilbert *et al.* (2009) is in some sense opposite to Feitosa *et al.* (2008), in that the authors consider many variables and a large set of parameters in their ABM, which is much more complex than Gauvin *et al.* (2013) too. First of all, the number of dwellings at any location is not fixed: some houses can be built, other houses are demolished (either after a long period or if prices decrease too much). Moreover, whereas buyers come from the outside of the city at a constant rate, sellers are both those who leave the city and those who decide to relocate, based on the interest rates they have to pay on their mortgage. Gilbert *et al.* (2009) also consider inflation and estate agents, whom they call "realtors", who have an important role in determining the real-estate prices.

Gilbert *et al.* (2009) reproduce some stylized facts, such as the sticky downwards prices: prices raise very quickly with the increase of demand, but decrease slowly when the demand decreases, as the sellers just do not sell their houses waiting for the prices to increase again. Notice that this result (the "bubbles") is also reproduced in the present thesis, see Sections 4.1 and 4.2. It is also great that they incorporate the financial market in the agent-based model, as the interrelation with the housing market is undoubtedly important. The problem with Gilbert *et al.* (2009) is, in the author's opinion, that there are too many factors which affect the prices, and it is difficult to identify the causal relationships. The attempt in the present thesis has been to keep the model simple enough so that some specific situations could be solved analytically, guaranteeing a full understanding of the model dynamics. Another problem with Gilbert *et al.* (2009) is the quantity of parameters and "ad-hoc" decisions, which make it really difficult to evaluate its predicting power.

Another ABM which recently had a substantial impact is Filatova *et al.* (2009). This model shares some features with Lemoy *et al.* (2010), in that it starts from the Alonso model and defines interaction rules which make

the price distribution emerge. The authors consider a transportation cost which lets the agents with a net income Y, and endow the agents with a utility function which depends on the "amenities" and on the distance to the center. The buyers decide where to look for an apartment by choosing the location out of a set of randomly chosen locations where they would get the highest utility. Moreover, the buyers have a willingness-to-pay (WTP) which depends non-linearly on the utility of the chosen patch. On the other hand the sellers have a willingness-to-accept (WTA) which is set exogenously to the level of the "agricultural rent" times a constant. Both buyers and sellers determine the actual price taking into account the numbers of buyers and sellers in a negociation process.

The main results in Filatova *et al.* (2009) are the following:

- The authors reproduce the trend of the Alonso model, that is prices higher in the center and lower in the periphery. They also find that at the most distant locations there are no buyers and the price is fixed at the level of the agricultural rent.
- If the preference towards the lower distance from the center becomes weaker with respect to the amenities (which here are assumed uniform), the price distribution becomes flatter
- If there are more buyers (sellers) than sellers (buyers), the offset of the price distribution is lower (higher)

It is interesting to compare this model and these results with those in Chapter 4: the setting is similar, and the bargaining process is modelled through reservation prices and an updating mechanism. Also the results are quite similar (compare the above results with those in Section 4.3). Nevertheless, the setting in the model of the present thesis is more varied and accounts for more mechanisms, still being tractable to understand the causal relationships. For instance, the reservation offer price is determined taking into account the market situation and the price paid by the individual agent, whereas in Filatova *et al.* (2009) it is just set to the level of the agricultural rent times a constant.

Finally, it is worth mentioning the contribution from Short *et al.* (2008), which constitutes the starting point of Gauvin *et al.* (2013). The authors present a burglary model where the criminals have to choose where to rob among a set of sites located on a grid. Every site is characterized by a level of attractiveness: part of it is constant (intrinsic), another part comes from the repeat victimization effect (locations where a robbery just took place are more likely to undergo more criminal actions). The similarity with the social component of the attractiveness is apparent. Criminals decide to commit a burglary at a Poisson rate which is proportional to the attractiveness; if they do not act, they move somewhere else in the neighbourhood of the starting

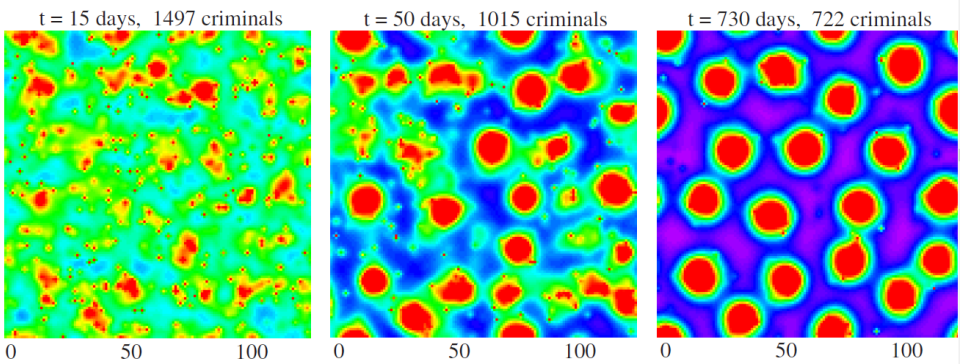


Figure 2.9: Emergence of hotspots of high criminality following the diffusion of burglaries. From Short *et al.* (2008)

site (local diffusion). The authors take the continuum limit and write the following differential equations:

$$\frac{\partial B}{\partial t} = \frac{\eta D}{z}\nabla^2 B - \omega B + \epsilon D \rho A \tag{2.39}$$

$$\frac{\partial \rho}{\partial t} = \frac{D}{z}\vec{\nabla}\cdot\left[\vec{\nabla}\rho - \frac{2\rho}{A}\vec{\nabla}A\right] - \rho A + \gamma \tag{2.40}$$

Let us explain the relevant terms: in (2.39) B is the varying component of the attractiveness. The first term on the right-hand side (RHS) is the diffusion term, the second term is the relaxation, the third is the reaction, which depends on the number of burglaries. (2.40) just shows the diffusion of the agents between neighbouring locations.

The authors consider a parallel with differential equations used in biology and are able to prove mathematically that a solution of (2.39) - (2.40) with the emergence of "hotspots" of high criminality can emerge in a wide parameter regime. In order to understand this, consider Fig. 2.9: at the starting point the attractiveness, and so the number of burglaries, is initialized randomly and is slightly higher in some locations; as time goes by, the spots where the attractiveness was higher start diffusing and a clear pattern emerges: it is an example of pattern formation.

Eq. (2.39) is very similar to (2.1), apart from the diffusion term. However, there is a major difference: in Gauvin *et al.* (2013) the unsuccessful buyers search for an apartment again at the city level (they may apply on the other side of the city), whereas in Short *et al.* (2008) the criminals search in the nearby locations. It would be an interesting extension of the present thesis to account for the diffusion of attractiveness, to see whether it would be enough to make hotspots of high prices emerge without imposing an intrinsic attractiveness structure.

Chapter 3

Empirical evidence

The purpose of this thesis is to make a model which accounts for stylized facts observed in real-world housing markets. Even though some of the assumptions in the model presented in Section 4.1 are based on plausibility criteria, others are grounded in the empirical observations from the data described in this chapter. The source is the database BIEN, managed by the Chambre des Notaires de Paris, recording more than 400000 real-estate transactions in the period 1990-2007. The issue is whether the model should be "taken to the data", i.e. whether it would be possible to fit the data with the predictions of the model. This is left for future work, but it should be noticed how the high dimensionality of the parameter space, some structural variables which cannot be known and the multidimensionality of the results make it hard to check quantitatively the explanatory power of the model.

The explanatory power of models with respect to empirical data is a common issue among all scientific disciplines. However, there are major differences between natural sciences and social sciences. In natural sciences, the theory should match perfectly the observations (for a graduate-level book on statistics and data analysis for physicists see Cowan (1998)), whereas in the social sciences the matching can only be approximate, and the phenomena may just be reproduced from a qualitative point of view. For instance, linear regression models in the social sciences assume the existence of "residuals" (an introductory text in econometrics is Stock and Watson (2003)), which is the part not explained by the statistical model. Moreover, the extreme difficulty of performing controlled experiments makes finding causal relationships extremely hard. It is easy to find correlations, but "correlation does not imply causation". Therefore, many techniques have been developed by econometricians to circumvent these difficulties, from Instrumental Variables to Fixed Effects in panel data (Arellano, 2003; Cameron and Trivedi, 2005). On the top of that, data have become "big", and some previous techniques in econometrics turned old-fashioned because of the sheer size of some datasets. The interrelation between "data science" (mostly machine

learning, see Abu-Mostafa *et al.* (2012)) and econometrics is investigated in Varian (2014), where the author advocates for a cross-fertilization between the two fields. The author gives a set of examples where machine learning techniques outperform econometric ones. For instance, he cites a study about the factors which influence who is approved for a mortgage, in order to test whether race was an important explanatory variable. By doing a regression, as it is standard in econometrics, Munnell *et al.* (1996) prove that a statistically significant negative coefficient exists on race about the probability to get a mortgage. Varian (2014) analyzes the data with a classification tree, and so is able to tell which factors are most important in determining whether the mortgage is approved. He finds that race shows up only marginally, whereas the biggest discriminant is "denied mortgage insurance". Another suggestion in Varian (2014) is to perform cross-validation to avoid over-fitting, a problem which is not considered in econometrics.

The goal of this chapter is to show descriptive statistics about the datasets (one for transactions in period 1990-2003, the other for 2005-2007, year 2004 is missing). The relevant variables are described in Section 3.1; the stylized facts which have been used to make the model are listed in Section 3.2; other interesting findings are presented in Section 3.3. Some parts of the code which has been used to analyse the data are shown in Appendix C. In particular, the parts related to the manipulation and the "cleaning" of the data have been shared in order to ensure full transparency, whereas the code to generate the plots has been omitted. The task of performing machine learning and/or econometrics on these data is left for future work.

3.1 Description of the database

The B.I.E.N database, recording real-estate transactions in Paris and in the Ile-de-France region year after year, is proprietary. Some years of the database have been purchased by Jean-Pierre Nadal and Annick Vignes thanks to two research projects directed by them. Because of this, there are two databases, one recording transactions 1990-2003, the other one for the period 2005-2007. Year 2004 is missing. The number of recorded real-estate transactions is 431975. Moreover, while the two databases are really similar for what concerns the most relevant variables, there are differences in denomination and some variables only exist in one database. The variables refer both to features of the apartments and to features of the traders. A short dictionary for the variables is provided in Table 3.1. In Appendix C a few comments about problems with the dataset are presented by variable, along with the code which has been used to address them. In what follows, we list the variables of the dataset, we explain their structure and provide a

short description[1], following the same order as in Table 3.1. The variables are:

- `REFE`: unique identifier. It is a natural number.
- `PRIX1`: price of the transaction. It is reported in euros and is a natural number.
- `SURFE1`: surface of the apartment which has been traded. The measure unit is m2 and it is a natural number.
- `prix.m2`: price per m2 of the transaction. It is reported in euros and is a rational number.
- `TYPAPP`: type of the apartment. It is a categorical variable with levels: "Standard Apartment", "Artist atelier", "Service room", "Duplex" , "Attic" , "Loft" , "Janitor's room" , "Studio" , "Triplex".
- `EPOQUE`: period of construction of the building where the apartment which is traded is situated. It is a categorical variable with levels: "Before 1850", "1850 - 1913", "1914 - 1947", "1948 - 1969", "1970 - 1980", "1981 - 1991","1992 - 2000", "From 2001".
- `ARRON`: "arrondissement" of the building where the apartment which is traded is situated. The arrondissement is an administrative entity in France, which may be translated in English as "district". It is a categorical variable with levels ranging from 1 to 20. Actually, even though it looks like a countable variable, it makes no sense to consider the cardinality of the arrondissement as it is not clearly related to the distance to the center.
- `X,Y`: geolocation of the building where the apartment which is traded is situated.
- `ANNEE`: year of the transaction. It is a natural number.
- `ANNEEMUTP`: year of the previous transaction for the same apartment. It is a natural number.
- `AGEB` (age of the buyer) and `AGEVE` (age of the seller): they are obtained respectively from `ANNEE - NAISSACQ` and from `ANNEE - NAISSVE`, where `NAISSACQ` and `NAISSVE` are respectively the years when the buyer and the seller were born. They both are natural numbers.

[1]Notice that the variables in the B.I.E.N. database which are not included in the data analysis in Sections 3.2 and 3.3 are not reported here. In the 1990-2003 database there is a total of 52 variables, in the 2005-2007 there are 45 variables

- `CSPAC` (social category of the buyers) and `CSPVE` (social category of the sellers). They are both categorical variables with levels: "artisans/shopkeepers/firm managers" , "executives/higher intellectual professions", "employees" , "workers" , "legal persons" , "middle-ranking professionals" , "retired" , "without activity" . It should be noticed how it is not possible to directly relate the "social category" with the income: for instance a firm manager may be such also for a very small and indebted firm.

- `DEPDA1` and `DEPDV`: they are respectively the "origin" of the buyers and of the seller. They are both categorical variables with levels: "DomTom" (Départments et Territoires d'Outre Mer, Overseas Departments and Territories), "Foreigner" , "Paris", "Petite Couronne" (the area around Paris), "Grande Couronne" (the area around the Petite Couronne), "Province".

- `PXMUTPREC` and `PXMUTPRECM2`: price and price per m2 of the previous transaction. They are cardinal variables, the first is an integer and the latter is a rational number.

- `NATIA` (nationality of the buyer) and `NATIV` (nationality of the seller): nationality of the traders. Most buyers and sellers are French, but a small fraction are foreigners.

3.2 Main insights

We focus here on some empirical results which have influenced the making of the model, in that they represent stylized facts which any model of a housing market should account for. However, it is not possible to observe in our data all the stylized facts, so some assumption of the model which will be presented in Section 4.1 come from plausibility assumptions or from other models, such as those described in Sections 2.1 and 2.2.

Figure 3.1 shows the average price per squared meter of the real-estate transactions in the Paris housing market as a function of the year. The data are disaggregated per arrondissement. It is possible to observe the temporal evolution of the average prices in different locations. As it can be readily noticed, prices decrease until 1998, then they start increasing much faster. Even though there is just one sequence of decrease-increase of the prices (it could be called a "business cycle"), one can see a typical feature of the housing markets, that is the sticky downwards prices, as highlighted by Gilbert *et al.* (2009) (see Section 2.4). Prices decrease slowly because people stay put in selling their apartments, whereas as soon as the prices start increasing people are more likely to put their apartments on sale. This pattern can be observed empirically in Figure 3.2, even though one should be cautious:

1990-2003	2005-2007	Explanation
REFE	REFE	Unique identifier
PRIX1	PRIX1	Price of the transaction
SURFE1	SURFE1	Surface of the apartment
/	prix.m2	Price per m2 of the transaction
TYPAPP_LIBEL	TYPAP	Type of the apartment
EPOQUE	REQ_EPOQU	Period of construction of the building
ARRON	BIARRON	Arrondissement of the building
X,Y	/	Geolocation of the building
ANNEE	ANNEE	Year of the transaction
ANNEEMUTP	DATMUTPREC	Year of the previous transaction
NAISSACQ	ANNAIS	Year when the buyer was born
NAISSVE	ANNAIS_VE	Year when the seller was born
CSPAC	/	Social category of the buyer
CSPVE	/	Social category of the seller
DEPDA1	/	Origin of the buyer
DEPDV	/	Origin of the seller
/	PXMUTPREC	Price of the previous transaction
/	PXMUTPRECM2	Price per m2 of the previous transaction
NATIA	CODNATIO	Country of origin of the buyer
NATIV	CODNATIO_VE	Country of origin of the seller

Table 3.1: Codes for the two databases and meanings of the variables. When a variable is only considered in one database (for reasons explained in Appendix C) the symbol "/" is used for the other database.

it may be that the office registering housing market transactions improved its efficiency or that a law forcing more transactions to be recorded was enforced. Another insight in Figure 3.1 is that when there is a global increase (or decrease) in prices in Paris, this happens in all the arrondissements. It is not a trivial result, it may well be that only the apartments in the center experience a "boom". However, this is never the case, as all the arrondissements follow the global trend. Finally, one can see that though following the trend, the ranking of the arrondissements with respect to the average price can vary. In particular, notice how the 16th arrondissement was in 1990 the most expensive one, whereas in 2007 it is just the sixth most expensive. This may be explained by swifts in preferences of the people. Further support for this hypothesis can be found in Figure 3.3. The maps of Paris are coloured by the average price of the transactions, which are geolocated as explained in Section 3.1. Through kernel techniques the authors of the maps average the prices over a small geographical area and are able to remove noise. The color scale assigns red colors to high prices and blue colors to low prices. Let us start from Figure 3.3a: at the beginning of the period spanned by the dataset the prices are generally high, as it can be argued by the red spot near the "Arc de Triomphe", on the left of the picture. However, some areas in the North and East of Paris are blue. In Figure 3.3b the pattern is similar, but the average level of the prices is lower (recall Figure 3.1). In Figure 3.3c the prices have increased again, but they are much more uniform: there are no peaks of high prices (no red spots), but blue spots only survive in the extreme North of Paris, an area peculiar for immigration patterns. In particular, the East of Paris started to undergo a process of gentrification. The "Bastille" case is often cited as an example where a previous working class area has changed as more and more middle-class and rich people decided to locate there. The reason is debated: some sociologists argue that this is due to the decision of many artists to live in the 11th arrondissement (around "Place de la Bastille"), others argue that it has been the construction of the "Opera Bastille" which started the gentrification process.

So far we only considered the location of the apartments and the price and the year of the transactions. Let us focus on some stylized facts from the point of view of the traders. Figure 3.4 refers to the difference in transaction price between two transactions for the same apartment. The data about the previous price are only available in the 2005-2007 database. Since we know from Figure 3.1 that those years are characterized by a housing market boom, the result in Figure 3.4a comes as no surprise: most of the people who sold their apartment in that period made a positive profit, even correcting for inflation. There is a peak of people who made a slightly negative profit, possibly those forced to sell their apartment, but the left tail of the distribution is almost risible with respect to the right tail. Figure 3.4b let us exploit the information about the year of the previous transaction. As it can be seen by the dip at a temporal distance of 15 years, the people

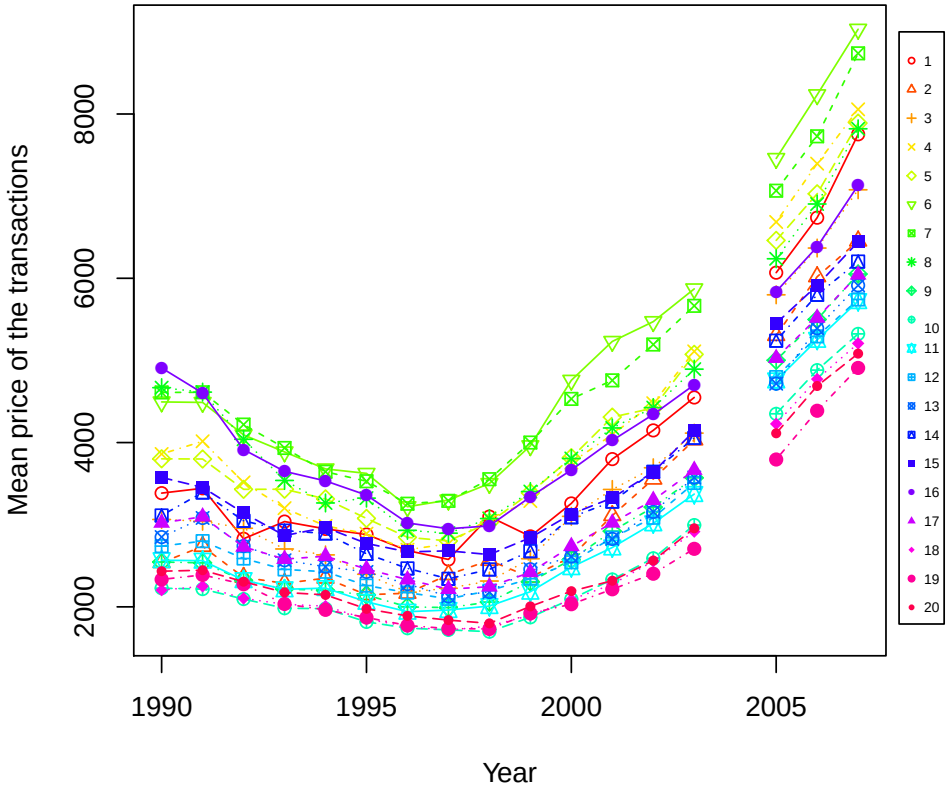


Figure 3.1: Average price per squared meter of the real-estate transactions in the Paris housing market as a function of the year. The data are disaggregated per arrondissement, allowing for a comparison of the intertemporal evolution of several areas in Paris. The following phenomena are apparent: prices decrease until they reach a dip in 1998, then they steadily increase; the global trend in prices is reflected in all the arrondissements, it never happens that one arrondissement behaves differently from the others; nevertheless, in few cases some arrondissements become less expensive compared to others. This is the case for instance of the 16th arrondissement.

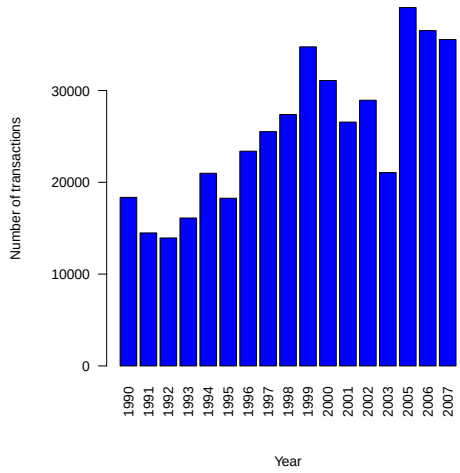


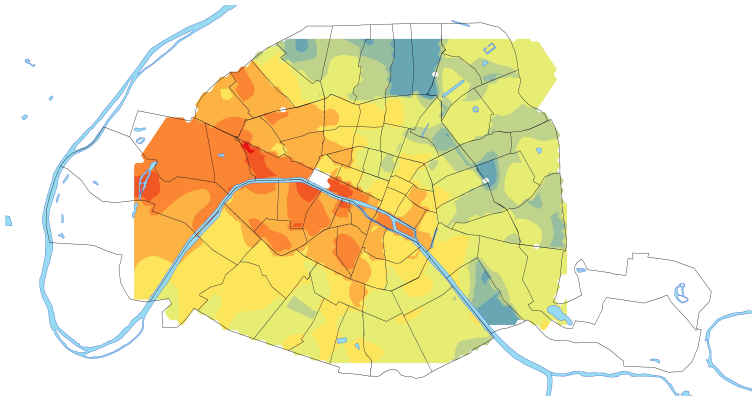
Figure 3.2: Number of real-estate transactions in the Paris housing market as a function of the year. The years when the highest number of transactions takes place are those when the prices rise most, compare with Figure 3.1. On the other hand, the initial dip in transactions is followed by a "bust" of the housing market.

who bought an apartment in the period when the prices were still high (i.e. 1990-1992) made on average a slightly positive profit, whereas those who bought the apartment in the period 1997-1999 made a much more positive profit.

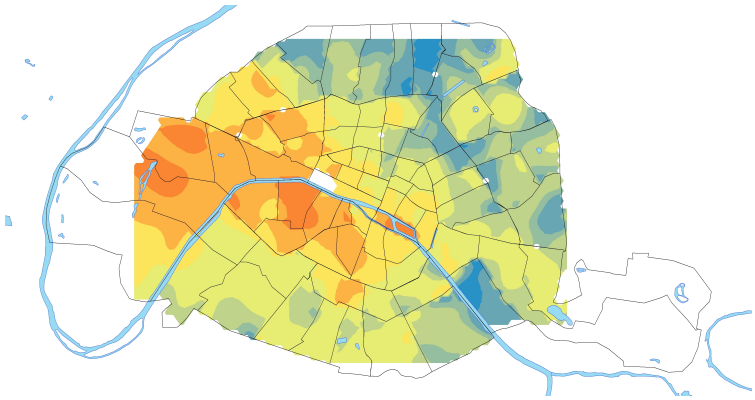
Another insight comes from Figure 3.5, where we take into account only the price of the current transaction and we consider in more detail the difference in years between subsequent transactions for the same apartment. In Figure 3.5a we show the distribution of the difference in years. Surprisingly, we find that the mode of the distribution is 1 year! This supports the statement that a significant part of the transactions are of speculative origin (the speculators in the housing market are called "marchands des biens" in French), as it is too of a short period for households to make a relocation decision. We also find that the mean is around 10 years, which is instead more of a sensible time between two apartment transactions for households. Figure 3.4b analyses another aspect, that is the correlation between the difference in years between the transactions and the average price per squared meter. A clear pattern emerges: the longer the time interval, the lower the price. We consider a linear fit:

$$P_i = a + b \cdot \Delta_i + u_i \tag{3.1}$$

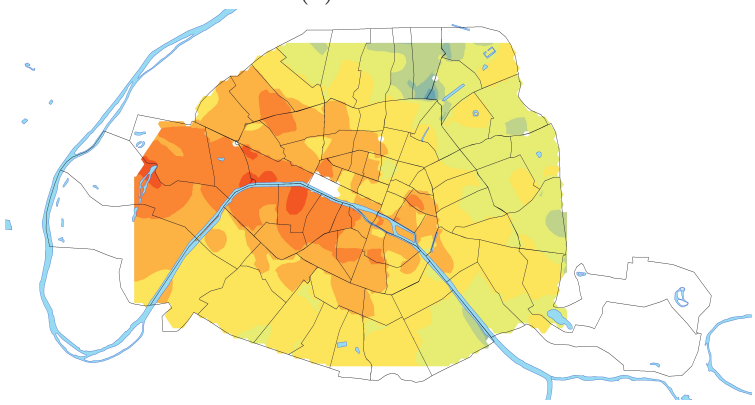
where P_i is the price, Δ_i is the difference in years between subsequent transactions for the same apartment and u_i is a residual, and get an $R^2 =$



(a) Year 1990



(b) Year 1998

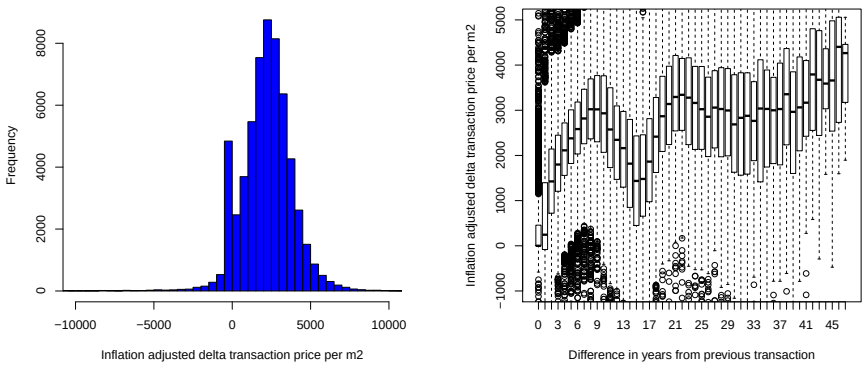


(c) Year 2003

Figure 3.3: Price distribution in Paris in some selected years, smoothed through kernel techniques. The color scale is constant throughout the years: it can be seen that prices in 1990 were high but there were sharp differences between most areas; in 1998 all prices decreased; in 2003 prices increased again, but became more uniform. The plots have been generated by a group of geographers working in the project DyXi, coordinated by J.P. Nadal.

0.859, with $a = (2676.04 \pm 3.85)$ €$/m^2$ and $b = (-10.93 \pm 0.30)$ €$/m^2$/year. The hypothesis that $b > 0$ is rejected at the 1% significance level.

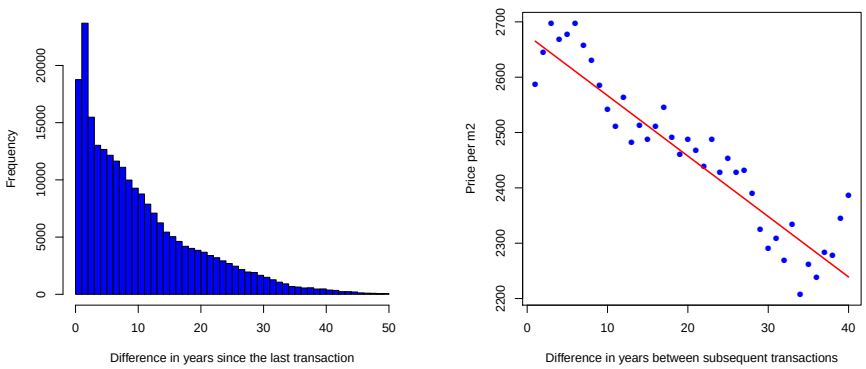
This supports an important hypothesis of the model described in Section 4.1: the sellers are more likely to decrease their price if a long time lapsed since they bought the apartment.



(a) (b)

Figure 3.4: (*Left*) Distribution of the difference in transaction price per m2 (corrected for inflation, we used data on inflation from the INSEE, the French national institute of statistics and economic studies) for apartments sold in the period 2005-2007. Most sellers made a positive profit. (*Right*) Box-plots for the distribution of the difference in transaction price per m2 (corrected for inflation) as a function of the difference in years between the transaction occurred in the period 2005-2007 and the former transaction for the same apartment. This plot is almost the specular version of that in Fig. 3.1

Let us conclude this Section by taking into account the nationality of the buyers: 359282 out of 431975 are French, and there are 48110 missing values. Thus the foreigners are the 5.69% of the buyers, which itself is not too of a low number. As it can be seen in Table 3.2 the price paid varies significantly depending on the country of origin of the buyer. In particular, buyers from countries whose GDP per capita is similar or higher than the French GDP per capita spend on average more than French people. On the other hand, buyers from countries whose GDP per capita is lower than the French GDP per capita spend on average less than French people. However, most of the buyers come from developed countries, so on average "foreigners" spend more than French people do. The pattern in the case of the sellers is similar, but notice that the number of foreign sellers is much less than the number of foreign buyers.



(a) (b)

Figure 3.5: (*Left*) Distribution of the difference in years between a transaction occurred in the period 1990-2007 and the former transaction for the same apartment. A cut-off of 40 years has been used. A substantial part of the distribution lies within a short difference in years. (*Right*) Average price per m2 as a function of the difference in years between a transaction occurred in the period 1990-2007 and the former transaction for the same apartment. It is apparent that apartments bought recently are sold at a higher price

Country buyer	Num	$\langle P_{buyer} \rangle$	Country seller	Num	$\langle P_{seller} \rangle$
France	359282	3437.07	France	374120	3457.60
Italy	3389	4657.47	Italy	1317	4207.36
United Kingdom	2111	4384.33	United Kingdom	968	4626.04
United States	1936	5082.82	United States	968	5025.21
Portugal	1828	2724.01	Algeria	855	2813.16
Algeria	1571	2490.39	Germany	751	4173.51
China	1348	2480.85	Spain	737	3464.52
Spain	1162	3834.96	Portugal	729	3418.99
Germany	1161	3621.13	Morocco	507	3343.83
Morocco	1092	3082.17	Switzerland	487	4076.79
Tunisia	830	2663.01	Belgium	432	4122.82
Belgium	617	4071.06	Tunisia	358	2944.53
Ireland	498	5442.46	Israel	288	3989.53
Switzerland	496	4002.06	Japan	287	4345.32
Japan	421	3850.42	Lebanon	254	4926.47
Iran	362	3586.35	Iran	227	4065.06
Israel	359	3116.44	China	207	3396.87
Lebanon	288	4205.83	Luxembourg	168	6034.48
Canada	286	4609.50	Canada	167	4306.63
Cambodia	273	2143.69	Greece	165	4197.72

Table 3.2: Countries of origin of the buyers and sellers, number and average price per nationality. The top 20 countries in terms of number of buyers/sellers are shown.

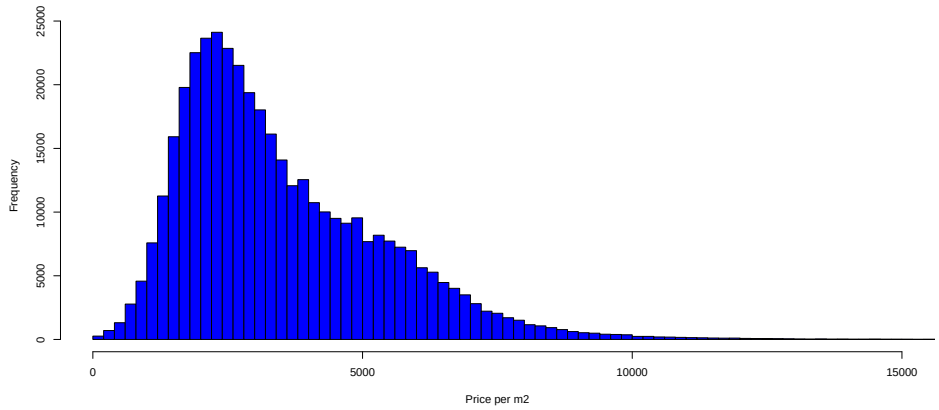
3.3 Related insights

We present in this section some complimentary empirical results. The number of variables makes it possible to consider an exponentially large number of combinations, so we focus on a few interesting cases. As dependent variable, we keep using the price, both the price per m2 and the total price.

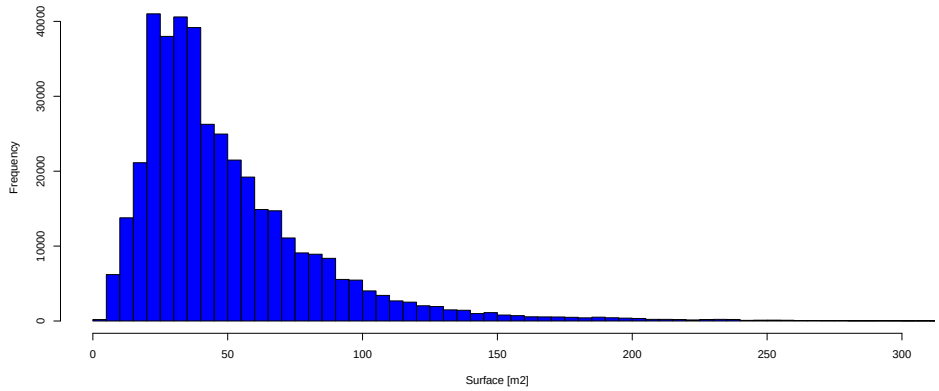
A preliminary result is given by the distribution of the price per m2 and of the surface of the apartments which have been traded in the whole period of the datasets (that is 1990-2007, with 2004 missing). As it can be seen in Figure 3.6 both distributions are asymmetric, with a right tail heavier than the left tail. The distributions are not heavy-tailed though, as the number of extremely expensive or really huge apartments is risible. A quantitative confirmation of what said is given by the mean and the median of the distributions: the mean of the price per m2 is 3500€$/m^2$, the median is 3024€$/m^2$. For what concerns the surface, the mean is $51m^2$, the median is $40m^2$.

Figure 3.7 shows the correlation between the age of buyers and sellers and the price they pay. In particular, Figures 3.7a and 3.7b show the correlation with price per m2, whereas Figures 3.7c and 3.7d refer to the total price. An apparent pattern is that when the price per m2 is considered, the price paid increases with the age of the buyer (as it is consistent, since they have a higher income and they can afford a higher price) but the price accepted by the seller decreases with the age of the seller: older sellers accept to sell at a lower price. An explanation for this empirical observation may be that old sellers bought their apartment more years in advance with respect to the year of the transaction, and so they accept to lower their price more easily. This would be in accord with the result in Figure 3.5b. Another interesting empirical observation is in Figure 3.7c, where it is clear the peak around 36 years old. If it is compared with Figure 3.7a it is possible to understand the reason, that is buyers buy bigger apartments: this can be explained as families with children who look for a more spacious apartment, possibly in a less expensive arrondissement. A thing which could be tested, and which we leave for future work, is whether the above hypothesis is true, by combining the variables on the arrondissement, the surface and the age of the buyer, and testing whether there is a positive coefficient on the combinations expensive arrondissement - small apartment - old/very young buyer or cheap arrondissement - big apartment - middle-aged buyer.

In Figure 3.8 we consider the social categories of the traders. It is remarkable how the vast majority of the buyers (see Figure 3.8a) are either "executives/higher intellectual professions" or "middle-ranking professionals". People whose jobs provide them with a lower income, such as "workers", "artisans/shopkeepers/firm managers" or "employees" are few with respect to the others. The price paid varies accordingly, with the "executives/higher intellectual professions" paying the highest price, as it can be seen in Figure



(a)



(b)

Figure 3.6: (*Top*) Distribution of the price per m2 for apartments sold in the period 1990-2007. (*Bottom*) Distribution of the surface of the apartments sold in the period 1990-2007.

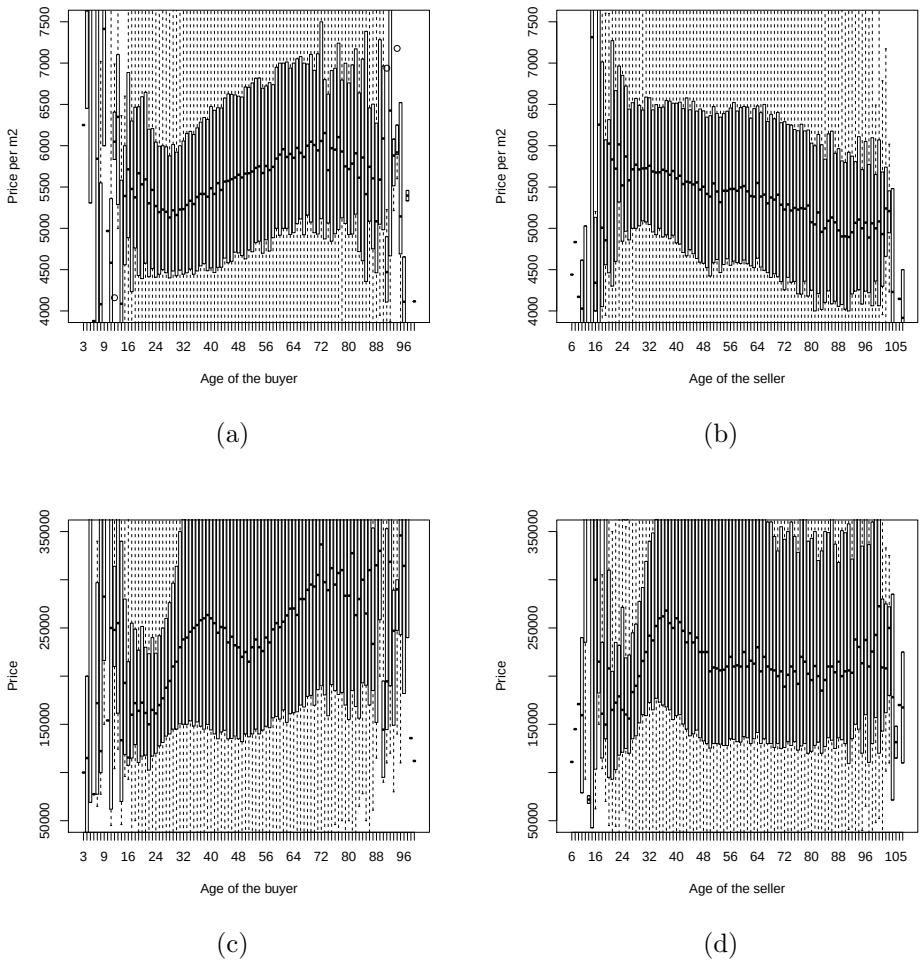


Figure 3.7: (*Left, Top*) Average price per m2 as a function of the age of the buyer for transactions occurred in the period 1990-2007. Until about 72 years old, the buyers pay a price increasing with age. (*Right,top*) Average price per m2 as a function of the age of the seller. The sellers pay a price.decreasing by age. (*Left, Bottom*) Average price as a function of the age of the buyer. It is apparent the peak around 40 years old, where the surface bought is substantially higher. (*Right, Bottom*) Average price as a function of the age of the seller.

3.8b. Let us turn to the sellers, Figures 3.8c and 3.8d: it is not surprising that there is a higher share of retired people, but is interesting to notice the high fraction of "legal persons", as if the institutions were selling some houses to the private sector. These may be council flats, but the price is not extremely low.

Figure 3.9 refers to the origin of the traders. With respect to the nationality of the traders in Table 3.2 it refers to local features, i.e. where the traders come from if they are not foreigners. It is interesting to notice that most buyers and most sellers (Figures 3.8a and 3.8c) come from Paris, even if it is not granted that they just relocate in Paris, it could just be that their residency was in Paris. Figures 3.8b and 3.8d show the average prices as a function of the origin of the buyers/sellers. Surprisingly, the people who pay most and sell their apartments at the highest prices are those from the DomTom (Départments et Territoires d'Outre Mer, Overseas Departments and Territories). They might be officials sent for a period in the DomTom who come back to Paris, but it is hard to interpret this result. This variable looks in general less related to the prices than other ones.

Figure 3.10 is about the period of construction of the building where the apartment which was traded is located. As most houses in Paris are from the period 1850-1913, the result in Figure 3.10a comes as expected. What is more interesting is the average price for the various levels of this categorical variable, shown in Figure 3.10b: surprisingly, the most expensive apartments are those in buildings constructed after 1980. It should be noticed that there is an ambiguity: it may be that a huge renovation for the building accounts as a reconstruction, even though this would not fully explain this result.

Finally, in Figure 3.11 we show the types of the apartments. Apart from the "standard apartment", there is a set of housing units with a different classification. The "service room", in French "chambre de service", used to be the room for servants of the richest families, located below the roof and without a private toilet. Now they are sold separately from the rest of the apartment, and they are usually used by students, as the renting prices for bigger apartments are extremely high. As it can be seen in Figure 3.11a they are a few with respect to the standard apartments. The prices for each type are shown in Figure 3.11b: duplex and triplex are on average more expensive than standard apartments, whereas service rooms are slightly cheaper.

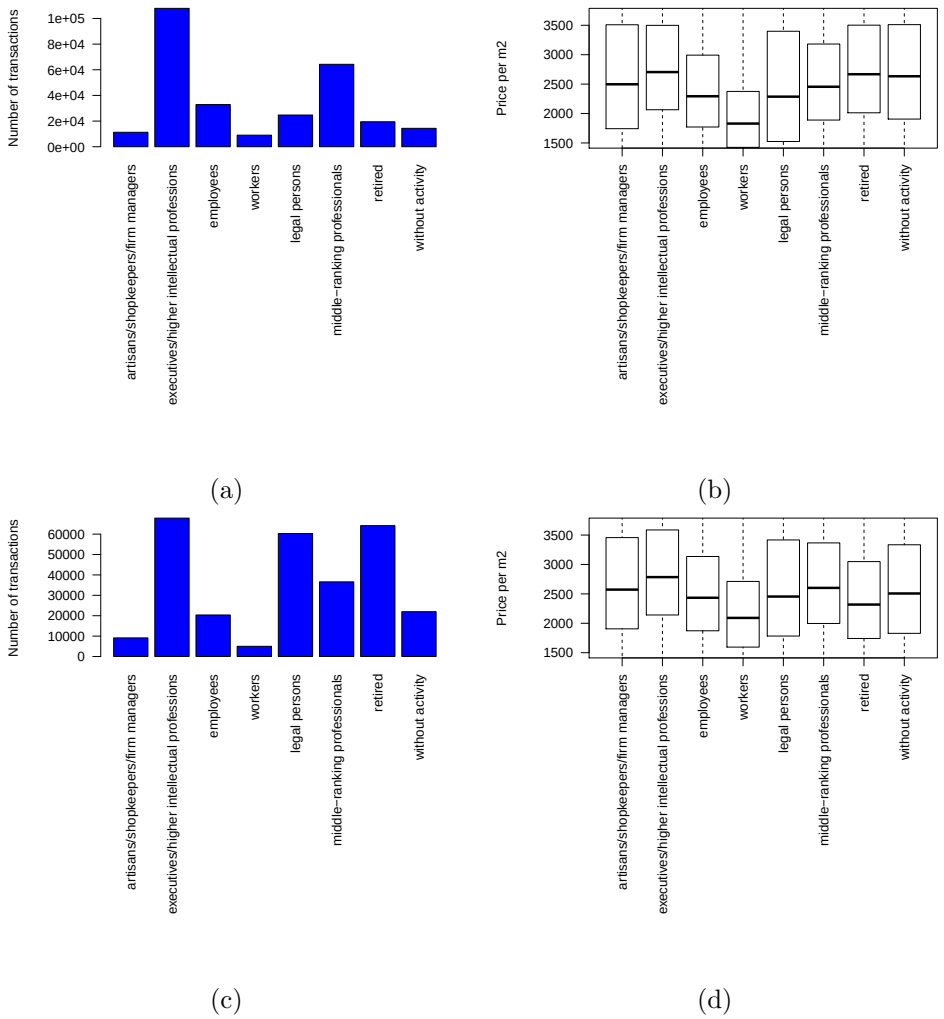


Figure 3.8: Social categories of the traders. (*Left, Top*) Frequencies of the social categories of the buyers. The mode is by far "executives/higher intellectual professions". (*Right,top*) Box-plots for the distribution of transaction prices per m2 as a function of the social categories of the buyers. The highest price is paid by the "executives/higher intellectual professions". (*Left, Bottom*) Frequencies of the social categories of the sellers. Again the mode is "executives/higher intellectual professions", but retired people and legal persons (i.e. legal entities) almost sold the same number of apartments. (*Right, Bottom*) Box-plots for the distribution of transaction prices per m2 as a function of the social categories of the sellers. The pattern is similar to the case of the buyers, but notice how retired people sell at a much lower price than the price they pay when they act as buyers. This is consistent with Fig. 3.7

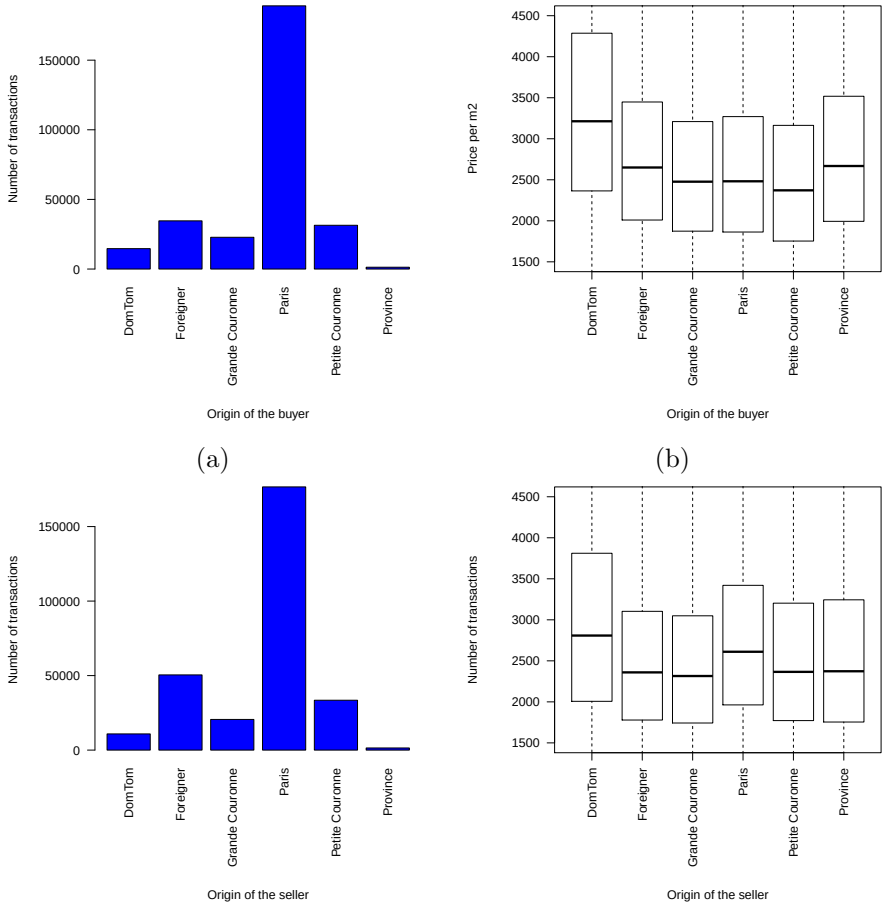
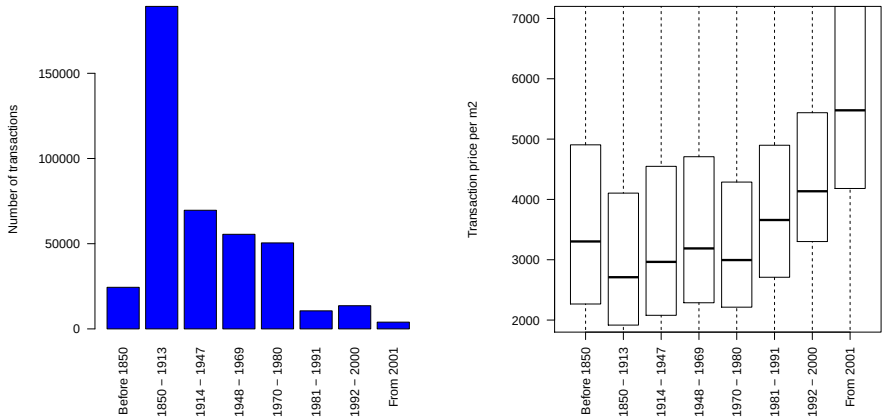
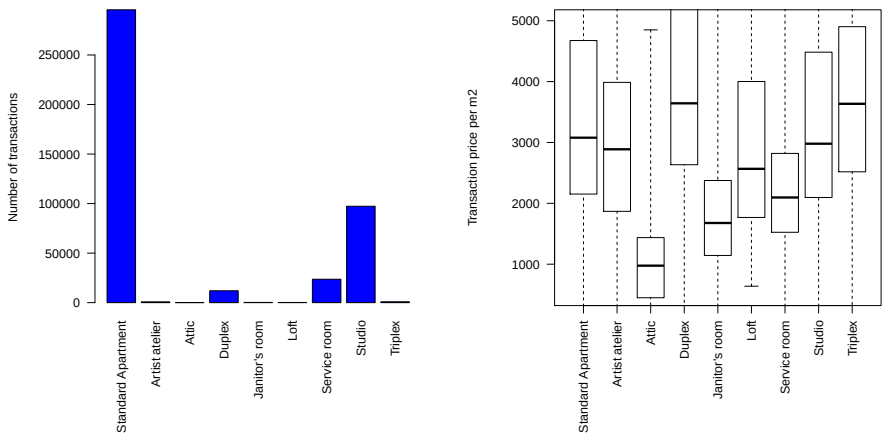


Figure 3.9: Origin of the traders. (*Left, Top*) Distribution of the origin of the buyer for apartments sold in the period 1990-2003. Most of the buyers come from Paris. (*Right, Top*) Box-plots for the distribution of the transaction price per m2 as a function of the origin of the buyers. The most expensive apartments are those bought by people coming from the DomTom (Départments et Territoires d'Outre Mer, Overseas Departments and Territories). (*Left,Bottom*) Distribution of the origin of the sellers. (*Right,Bottom*) Box-plots for the distribution of the transaction price per m2 as a function of the origin of the sellers.



(a) (b)

Figure 3.10: Period of construction of the building. (*Left*) Distribution of the period of construction of the building for apartments sold in the period 1990-2007. The mode of the distribution is: 1850-1913. (*Right*) Box-plots for the distribution of the transaction price per m2 as a function of the period of construction of the building. The most expensive apartments are those in buildings constructed after 2001.



(a) (b)

Figure 3.11: Type of the apartment. (*Left*) Distribution of the type of the apartment for apartments sold in the period 1990-2007. The mode of the distribution is: "standard apartment". (*Right*) Box-plots for the distribution of the transaction price per m2 as a function of the type of the apartment. The most expensive apartments are the "duplex", the cheapest are the "attics".

Chapter 4

Model and results

This chapter presents the original theoretical results of the thesis. The model which we consider is a modified version of Gauvin *et al.* (2013), influenced by some works described in Chapter 2 and by some empirical observations detailed in Chapter 3. The making of the model has been based on starting from sound economic considerations, and only subsequently developing a mathematical framework and a simulation software which were suited to it. This may look as an obvious statement, but it is not, as some models in "Econophysics" are adaptations to an economic context of models directly coming from physics (for a critical review on the achievements of econophysics see Gallegati *et al.* (2006) and the response McCauley (2006)). We also tried to keep the model simple enough at the level of individual agents, in order to keep it analytically tractable at least in part, but detailed enough to account for real features of the housing market. More details on the making of the model are provided in Appendix B, whereas more detailed methodological considerations have been presented in Chapter 1. The organization of this chapter is as following: in Section 4.1 we present the assumptions and the dynamics of the discrete agent-based model; in Section 4.2 we present the mathematical results which have been obtained by taking the continuum limit of the ABM, and which let us understand the mechanisms at work and the effect of the parameters; in Section 4.3 we employ the strength of the ABMs, that is simulating models which are not otherwise tractable. Finally, in Section 4.4 we summarize the outcomes and we compare them with other results existing in the literature, notably with those in Gauvin *et al.* (2013).

4.1 The agent-based model

We propose a dynamic model for the urban housing market. As in Gauvin *et al.* (2013) we consider a city whose locations are characterized by a certain level of attractiveness. Agents, who are heterogeneous in their income, compete for the available apartments. One of the biggest novelties with respect

to Gauvin *et al.* (2013) is that we model the bargaining process through a continuous double auction (order book). A short introduction to the order book is provided in Appendix A. The price distribution emerges out of the interactions of the agents: the market price is determined endogenously and depends mostly on the interplay between demand and supply intensities.

The assumptions of the discrete agent-based model are presented below[1]. When appropriate, we will compare them with the continuous version of the model, which will be presented in greater detail in Section 4.2.

4.1.1 Time, space and goods

- From Gauvin *et al.* (2013): [Time is discrete and indexed by t. The time increment is δt. In the numerical simulations in Sections 4.2 and 4.3, $\delta t = 1$ is the numerical time step. However, in the mathematical analysis in Section 4.2, the continuous time limit, $\delta t \to 0$, will be taken. The horizon is infinite.]

- In the numerical simulations [we consider a 'city' defined as a discrete set Ω of locations X uniformly distributed on a bounded open set $\widetilde{\Omega}$ in $\mathbb{R}^2$. The origin O is taken as the geographical center of the city, and we denote by $D(X)$ the Euclidean distance from the center to a location $X = (x, y)$. The total number of locations is $Card(\Omega) = L^2$, and the space is of linear size (diameter) $\mathcal{D} = aL$, where a gives the typical distance between two neighbouring locations]. In the mathematical analysis we take the typical distance $a \to 0$, and so we consider a continuum of locations. It will be convenient to characterize the city as a circle with radius $\bar{R}$ and the locations through their distance from the center, which we denote by r.

- [In the city, there is a total number $\mathcal{N}$ of goods (housing for sale) with identical intrinsic characteristics. The same number $N = \mathcal{N}/L^2$ of dwellings is available at each location X in Ω]. In the continuous limit, the density of apartments $n = N/a^2$ is uniform.

4.1.2 Agents

[At each period (given time t), there is a finite number of agents in the economy, who can be in one of the three following states: (1) buyer, (2) seller, (3) housed. We assume an infinite "reservoir" of agents outside the city. Agents in the reservoir are heterogeneous in their income - they are indiscernible except for their income category.

[1] To ease the comparison between the two models and to make it clear which assumptions are shared, some of the text presented below comes from Gauvin *et al.* (2013). The cited text is presented in squared brackets, and it always refers to the same paper.

From this reservoir, at each time step t a constant number $\Gamma\delta t$ of randomly-chosen agents arrive on the market.] Here we assume that buyers who were not successful in finding a house come back to the reservoir at the end of each period, and may come back to the city at the beginning of the subsequent period (it is not really important to keep track of the agents' identities). [At the same time, housed agents become sellers at a homogeneous rate α. Here we assume that they have to leave the city for exogenous reasons. So the goods available for sale at a given location are those put on the market by these agents plus, if any, those that have not yet sold. Then matching occurs between buyers and sellers at each location (see below, Section 4.1.7, for the detailed rule). Sellers who succeed in selling their good leave the market and return to the external reservoir.

Note that the total number of agents of each type - buyers, sellers and housed - are dynamical variables since they depend on the success rate of the previous period and on the inflow and outflow rates.]

4.1.3 Income categories

Agents are characterized by their income Y. A criticism may be that one should also consider the wealth, but we consider the price paid for an apartment as paid monthly, as it is actually the case when the payment comes from savings or from a mortgage. For simplicity, we consider a finite number K of income levels. Agents with the same income are denoted by k-agents, $k \in \{1, ..., K\}$, and have income Y_k. These incomes are ordered by increasing values, $Y_1 < Y_2 < ... < Y_K$. When the agent is acting as a buyer, his reservation demand price (i.e. the highest price he is willing to pay for an asset) is $\overline{P_k^d} = Y_k$. The actual price the agent will pay depends on the market mechanism, as explained in Section 4.1.7. Here we assume that income levels are separated by a constant Δ: $Y_k = Y_1 + (k-1)\Delta$. Note that once that Y_1, Δ and K are specified, all the income categories are determined. Differently from Gauvin *et al.* (2013) we do not assume that the income values are uniformly distributed in the population[2]. We denote the number of incoming agents in each income category by Γ_k, s.t. $\sum_k \Gamma_k = \Gamma$. Choosing appropriately Δ and $\{\Gamma_k\}$, any realistic income distribution may be modelled.

4.1.4 Attractiveness

[The attractiveness $A_k(X,t)$ of a location X at time t, for a k-agent, depends on both (1) an *intrinsic attractiveness*, $A^0(X)$, resulting from the

[2]The authors generalize formally their approach to an arbitrary income distribution (see Appendix A.4), but they do not perform the analysis on the effects of the distribution on the segregation pattern

location's intrinsic objective characteristics (e.g. local amenities), independent of the agent category, and (2) subjective characteristics which depend on the agent's social preferences, that is his preferences concerning the social characteristics of the neighborhood. This attractiveness matters], along with the average price of the dwellings at the same location (see Section 4.1.5), [in determining the inflow of buyers at X].

The intrinsic attractiveness

[The intrinsic attractiveness, $A^0(X)$, idiosyncratic to the position considered, is here assumed to be time-independent. This means that we consider the time scale of transactions to be much shorter than that involved in the transformation of amenities, which we do not take into account.]

Attractiveness dynamics

[The attractiveness $A_k(X,t)$ is a *subjective attractiveness*, whose value depends on the income of the agent looking at the location X. At each time step, new arrivals at the location X modify the social composition at this location, and the attractiveness evolves accordingly. Once the transactions between t and $t+\delta t$ have occurred (as described below), the attractiveness $A_k(X,t)$ of a location X as seen by a k-agent is updated according to:

$$A_k(X,t+\delta t) = A_k(X,t) + \omega\delta t\,(A^0(X) - A_k(X,t)) + \ \delta t\ \Phi_k(X,t) \quad (4.1)$$

where $\Phi_k(X,t)$ formalizes how the transactions at time t affect the attractiveness according to the social preferences of the agents of income category k. We assume $\Phi_k(X,t) = 0$ whenever no transaction occurred at time t: the time evolution (4.1) then implies that, when there is no transaction at a given location for a certain amount of time, the attractiveness relaxes towards its intrinsic value $A^0(X)$. In the present work we assume that the attractiveness of a location for an agent increases when more agents with similar or higher incomes are housed at this location. This is done by choosing $\Phi_k(X,t)$ proportional to the number of new buyers with income greater or equal to the agent's income. Hence we write:

$$A_k(X,t+\delta t) = A_k(X,t) + \omega\delta t\,(A^0(X) - A_k(X,t))) + \epsilon\delta t\, v_{k_>}(X,t) \quad (4.2)$$

with

$$v_{k_>}(X,t) = \sum_{k'\geq k} v_{k'}(X,t) \quad (4.3)$$

where $v_k(X,t)$ is the density of k-buyers in location X who complete a transaction at time t. Here and in all that follows, what we mean by *density* is a number (of agents) over the elementary surface area, a^2.] For dimensional consistency, also $A_k(X,t)$ is an attractiveness per unit area. [Note that the

level of attractiveness for a given income category depends on the intensity of the demand from agents of similar or higher category: the higher the demand, the higher the level of attractiveness.]

4.1.5 Utility function

The utility of the k-agents looking for an apartment at location X depends both on the attractiveness value $A_k(X,t)$ and on a composite good z, representing consumption of other goods but housing. As described in Section 4.1.7, the utility matters in determining the intensity of demand at location X. We assume that the utility function is a Cobb-Douglas with weight β on the attractiveness. Hence we have:

$$U_k(X,t) = z^{1-\beta} \left(A_k(X,t)s_0\right)^{\beta} \tag{4.4}$$

with $0 \leq \beta \leq 1$. Notice that the utility is dimensionless, which is why in Eq. (4.4) we multiply the attractiveness by s_0, the elementary surface of the homogeneous apartment.

The budget constraint of the k-agents is:

$$\tau z(t) + P(X,t) \leq Y_k \tag{4.5}$$

In Eq. (4.5) τ is the price of the consumption good and $P(X,t)$ is the market price, i.e. the average real-estate transaction price (the market price will be defined more clearly in Section 4.1.7), at location X and period t. At the optimal level, (4.5) holds with equality. Solving for z and replacing in Eq. (4.4) one gets the indirect utility

$$U_k(X,t) = \left(\frac{Y_k - P(X,t)}{\tau}\right)^{1-\beta} \left(A_k(X,t)s_0\right)^{\beta} \tag{4.6}$$

Without loss of generality, we can assume $s_0 = 1$ and $\tau = 1$. Anyway, one should bear in mind that $U_k(X,t)$ is dimensionless.

The number of k-buyers at location X will be proportional to the indirect utility, Eq. 4.6. Thus, agents will face a tradeoff between living in a very attractive location, where prices will presumably be higher (notice that the prices will be determined endogenously, as described in Section 4.1.7), and having a higher consumption level. Given the form of the utility function, if $\beta \neq 1$ and $P(X,t) \approx Y_k$, the marginal utility of z will be so high that k-buyers would look for a house where the prices are lower, to ensure a higher consumption level.

4.1.6 Reservation offer prices

When acting as sellers, the agents decide their reservation offer price (i.e. the lowest price they would accept for selling their apartment) by taking

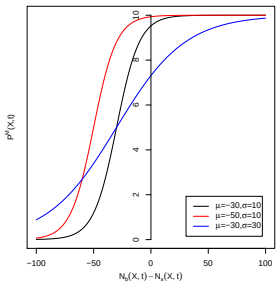


Figure 4.1: $P^M(X,t)$ for some parameter values. When $N = 100$, the parameter $\mu = 30$ means that if 30% of the apartments are on sale, $P^M(X,t)$ is halved: given that there is much competition on the offer side of the market, it is reasonable to think that the sellers have to decrease their reservation price to succeed in selling their apartment. When μ becomes more negative, more apartments have to be on sale for sellers to be convinced to decrease their reservation price, and the market somehow freezes, as it will be clear in Section 4.1.7. When σ increases the opposite effect is observed.

into account the price they paid for the apartment, the market price, and the relative demand and supply intensities.

First of all, we define the modified market price

$$P^M(X,t) = \frac{P(X, t-\delta t)}{1 + e^{-\frac{N_b(X,t) - N_s(X,t) - \mu}{\sigma}}} \tag{4.7}$$

Here $N_b(X,t)$ and $N_s(X,t)$ are respectively the number of buyers and sellers at location X. μ and σ are dimensionless parameters. The modified market price can be understood as the reservation offer price suggested by a real-estate agent to the seller, in order to ease the process of selling his apartment. The choice for the logistic in (4.7) means that, if $N_s(X,t) >> N_b(X,t)$, then $P^M(X,t) \to 0$: there is much competition on the sellers side, the only way to sell the apartment is to dramatically decrease the price. When $N_s(X,t) << N_b(X,t) + \mu$, $P^M(X,t) \approx P(X, t-\delta t)$: since there are enough buyers, the seller does not reduce the market price in taking his decision. The parameter μ is thus a threshold value which determines how many houses have to be on sale with respect to the number of buyers for prices to start to decrease; σ is the width of the logistic (Figure 4.1). Notice that σ determines the level of non-linearity in the price decision: if σ is high compared to N the price decrease starts smoothly as the number of buyers is not enough; in the reverse case, there is a sharp transition towards lower prices (see Section 4.2.1). Whether prices should decrease smoothly or whether they should decrease sharply once the threshold μ is reached depends on the features of the specific housing market.

The sellers sell at $P^M(X,t)$ only if it is higher than the price they paid when they bought the apartment: if they recently bought it, they do not want to incur in a loss, and they try to make a positive profit. Thus, the reservation offer price writes:

$$\underline{P_i^s}(X,t) = \max\left[P^M(X,t), P^M(X,t) + \lambda^{t-t_i}\left(P^i(X,t_i) - P^M(X,t)\right)\right] \tag{4.8}$$

Here $P^i(X,t_i)$ is the price agent i paid for the apartment at time t_i: as $t - t_i$ grows the reservation offer price relaxes towards $P^M(X,t)$. The relaxation speed is governed by λ: a value of λ close to one implies that the real estate market is undergoing a bubble, as the agents are speculating on the apartments in order to make a positive profit. Here we assume that λ is constant throughout the simulation and the agents. The underline notation emphasizes that the sellers do not want to sell at a price lower than the reservation price, which in this sense is a lower bound. Two remarks are helpful: differently from the buyer side, the sellers have to be treated individually, since it is needed to keep track of the price they paid; $\underline{P_i^s}(X,t)$ is the minimum price that agent i would accept for his apartment but, due to the bargaining process, the transaction price will generally be higher. Notice that, if the number of buyers is stable and sufficiently high, $P^M(X,t) \approx P(X,t-\delta t)$ and the prices at location X keep increasing: we have a bubble. However, because of the assumptions of the model, the prices cannot increase above Y_K.

4.1.7 Market dynamics: the matching

The initialization is as follows: we start at time $t = 0$ with a city fully occupied, i.e. with $M = N \cdot L^2$ agents in the economy. The attractiveness is set to its intrinsic value, the market price is set to an initial value P_0. [At each time step (between times t and $t+\delta t$), for each location X, the market mechanism ensures the matching between buyers and sellers.]

Demand and offer side

On the demand side. For each $k \in \{1, ..., K\}$, the total number of k-buyers in the city is $\Gamma_k \delta t$. [Each of these agents decides to visit one particular location. The probability $\pi_k(X,t)$ for a k-buyer to visit a given location X at time t is proportional] on the utility he would get at the location:

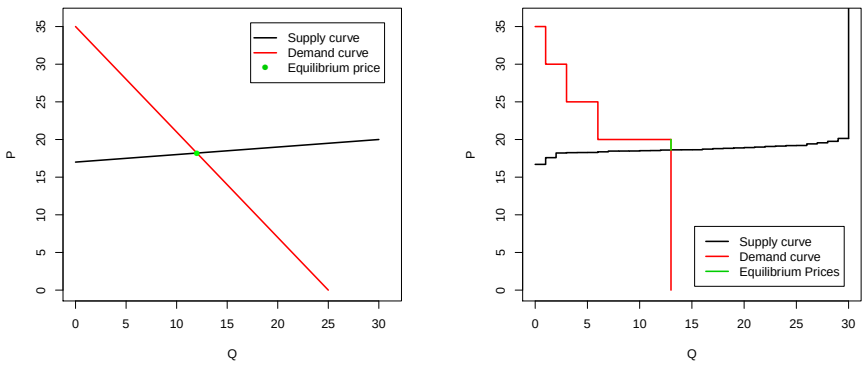
$$\pi_k(X,t) = \frac{U_k(X, t-\delta t)}{\sum_{X' \in \Omega} U_k(X', t-\delta t)} \tag{4.9}$$

[These decisions, being made in parallel by all the buyers, determine the demand at each location.]

The reason of the time lag in the above equation is that the buyers have to make decisions about where to look for a house according to the information available, so they consider past transaction prices. If $P(X,t) > P(X,t-\delta t)$ the market mechanism will however ensure that the budget constraint (4.5) is satisfied. If $P(X,t-\delta t) > Y_k$, we assume that the utility of the k-agents in choosing location X is null. The above rule can be seen as an example of bounded rationality. The agents are able to perceive where their utility would be higher, but they cannot fully optimize, i.e. choose with certainty the location where the utility is the highest. This can also be due to lack of information, exogenous reasons to choose a specific location and stochastic effects.

On the offer side. [At each location, the goods offered are (i) those put on sale by the current owner at some previous time step and not yet sold, if any; (ii) goods newly put on the market with probability α by the remaining housed agents.]

The matching



(a) (b)

Figure 4.2: Demand and supply curves in the continuous and discrete case. In Fig. 4.2b the actual values of the reservation prices in a numerical simulation have been taken. Notice that if the number of agents is large and the reservation prices are on a continuum, the curves in Fig. 4.2a are recovered, as it is almost the case for the supply curve in Fig. 4.2b.

The issue is to determine the market price from the reservation prices of the buyers and sellers. In Fig. 4.2a the traditional demand and supply curves are shown: buyers decide how much to consume given the price, sellers set the supply level at a certain price, the equilibrium price is such that there is no excess demand or excess supply. However, in a housing market each

agent is trading only one good. In Fig. 4.2b the discrete demand and supply curves for location X and time t are shown: at price P at least Q buyers (sellers) would buy (sell) the good which is traded, because their reservation prices are higher (lower). The equilibrium price is such that all possible transactions happen, i.e. at a higher price less buyers could afford the apartments on sale, at a lower price less sellers would accept to sell their apartment. Given this definition, there might be a set of equilibrium prices, as depicted in Fig. 4.2b. Nevertheless, when one thinks of the demand and supply curves in discrete terms and considers the bargaining process between buyers and sellers in the "short run" (that is, at time t, when the number of buyers and sellers is fixed), it is by no means obvious that all the transactions would happen at the equilibrium price(s). Moreover, if there are more buyers than sellers at any location at time t (or viceversa), there will always be excess demand (supply) at that location at time t, since the agents are trading just one good.

Therefore, it is needed to find a way to model a double auction, where buyers and sellers with several reservation prices meet and bargain. We use an order book which, though customarily used in financial markets, fits well to our case. See Appendix A for a short introduction to the concept of order book. The proposed prices of the buyers are called bids; those of the sellers are called asks. Bids and asks are put randomly in the order book; every time a new order comes, the highest bid is matched with the lowest ask, provided that it is higher. The transaction price is the average between the selected bid and ask, $P^i = (\text{Bid} + \text{Ask})/2$. There is a major difference with the order book as used in financial markets: in that situation, the market price is just the price of the last transaction. Traders know the prices of all the transactions, and take decisions accordingly. In a housing market traders only have an approximate idea of the prices of the transactions, so it makes sense to assume that the market price is the average of the transaction prices at location X at time t: $P(X,t) = \langle P^i(X,t) \rangle$.

In the mathematical analysis (Section 4.2) we use the simplest rule (i): the reservation prices are put directly in the order book as bids and asks. The market price, i.e. the average of the transaction prices, is generally higher than the equilibrium price, as the demand curve is steeper than the supply curve[3] (see Fig. 4.2b).

In the numerical simulations (Section 4.3), we test a more realistic rule (ii): through a repeated order book, buyers and sellers adaptively change the prices they propose. Buyers start bidding much less than their reservation price (they choose the minimum reservation price), sellers start asking much more (they choose the maximum). Denoting by j the round of the order

[3]Notice that the price of the last transaction will likely coincide with the equilibrium price, since the lowest bid and the highest ask are more likely to be matched last.

book we have, for the initial round at location X and time t:

$$\begin{aligned} P_k^d(j=0) &= \min\left[\underline{P_i^s}(X,t), \overline{P_k^d}(X,t)\right], \forall i,k \\ P_i^s(j=0) &= \max\left[\underline{P_i^s}(X,t), \overline{P_k^d}(X,t)\right], \forall i,k \end{aligned} \tag{4.10}$$

Notice again that the buyers can be dealt as k-agents, but the sellers have to be considered individually. The evolution for the proposed prices takes into account the ratio between the numbers of buyers and sellers: if there is much competition on the demand (supply) side, buyers (sellers) increase (decrease) their proposed price faster, converging quicker towards their reservation price. In formula:

$$\begin{aligned} P_k^d(j+1) &= P_k^d(j) + g(\zeta)\left[\overline{P_k^d} - P_k^d(j)\right], P_k^d(j) \leq \overline{P_k^d} \\ P_i^s(j+1) &= P_i^s(j) + g(1/\zeta)\left[\underline{P_i^s} - P_i^s(j)\right], P_i^s(j) \geq \underline{P_i^s} \end{aligned} \tag{4.11}$$

Here $\zeta = \frac{N_b(X,t)}{N_s(X,t)}$ is the ratio between the numbers of buyers and sellers and $g(\zeta)$ is an increasing function of its argument. To prevent the proposed price to oscillate around the reservation price, we take $g(\zeta) = \frac{\zeta}{1+\zeta}$, so that $0 \leq g(\zeta) < 1$. The larger $g(\zeta)$, the faster is the convergence towards the reservation price. A visual representation of this rule is provided in Figure 4.3.

4.2 Mathematical analysis

The agent-based model described in Section 4.1 has 15 parameters (see Table 4.1), plus the values $\{\Gamma_k\}$ (that is $K-1$ parameters more). Furthermore, it is needed to specify the intrinsic attractiveness $A^0(X)$. Even though some parameters apparently do not influence the behaviour of the system, it is necessary to give a mathematical description of the model to understand which control parameters matter the most, and if it possible to group them in fewer effective parameters, as it was done in Gauvin *et al.* (2013). The goal of this section is to show that, in spite of the great complexity of the agent-based model, its most important features can be analytically understood. When appropriate, we also compare the mathematical results with the outcomes of the simulations. We conclude this Section with a discussion about the relevant parameters, which dramatically reduce (see Section 4.2.5).

We average out stochastic effects by taking expected values. By neglecting fluctuations, we encounter problems in some cases (see Section 4.2.3),

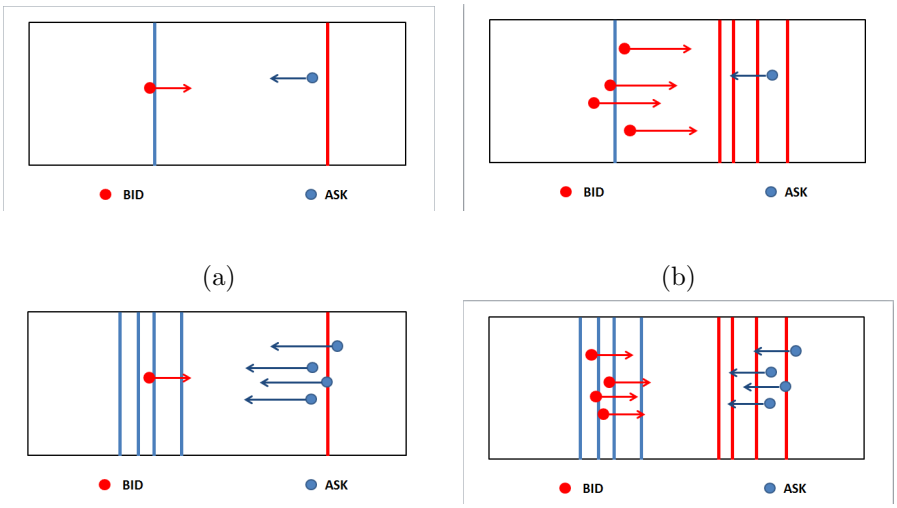


Figure 4.3: Solid points are bids and asks, moving rightward and leftward. Higher prices are on the right. The lines are the reservation prices. (*Top, Left*) One bid, one ask. The convergence speed is the same. (*Top, Right*) Four bid, one ask. The convergence speed is higher for bids, in accord with the intuition that high demand boosts prices. (*Bottom, Left*) One bid, four asks. The convergence speed is higher for asks, in accord with the intuition that a lot of unsold houses make prices decrease. (*Bottom, Right*) Four bids, four asks. The convergence speed is the same as in the first case for both buyers and sellers.

Symbol	Description
N	Number of apartments at each location
L^2	Number of locations
a	Distance between neighboring locations
δt	Time interval
K	Number of income categories
Y_1	Income of the lowest income category
Δ	Difference in income between 2 neighboring categories
β	Weight given to the attractiveness in the utility function
α	Probability for housed agents to become sellers
ϵ	Weight of the social attractiveness
ω	Weight of the relaxation towards the intrinsic attractiveness
λ	Discount factor for the reservation offer price
μ, σ	Threshold and steepness values for the modified market price
Γ	Total number of incoming agents each time step
Γ_k	Total number of incoming k-agents each time step

Table 4.1: Model parameters

as it is customary with mean-field approaches. Moreover, we assume continuous space and time ($a \to 0, \delta t \to 0$). As a first approximation, in line with Alonso *et al.* (1964), we model the intrinsic attractiveness as a function decreasing from the center of the city. What matters is only the distance to the center r. Notice that by choosing appropriately $A^0(X)$ every city structure, including polycentric cities, can easily be modeled. We take the city to be a circle with radius $\bar{R}$, which is chosen such that the area of the circle is the same as that of the square occupied by the grid defined in the discrete case. So $\bar{R}$ is determined from $(aL)^2 = \pi \bar{R}^2$. We assume the intrinsic attractiveness to be:

$$A^0(r) = \begin{cases} A^0_{max} e^{-\frac{r^2}{\bar{R}^2}}, & 0 < r < \bar{R} \\ 0, & r > \bar{R} \end{cases} \tag{4.12}$$

Having taken the continuum limit, we can formalize the evolution of the system through partial differential equations. The updating rule of the attractiveness $A_k(r,t)$ of a location at distance r from the center, seen by a k-agent, Eq. (4.2), gives:

$$\partial_t A_k(r,t) = \omega \left(A^0(r) - A_k(r,t) \right) + \epsilon \, v_{k_>}(r,t) \tag{4.13}$$

The market price at the same location is $P(r,t) = \left\langle P^i(r,t) \right\rangle$. The indirect utility (recall from Section 4.1.5 that $\tau = 1, s_0 = 1$, U is dimensionless)

is:

$$U_k(r,t) = (Y_k - P(r,t))^{1-\beta} (A_k(r,t))^{\beta} \tag{4.14}$$

We define the density of buyers at a location at distance r from the center as $n_b(r) = N_b(r)/a^2$, the density of sellers $n_s(r) = N_s(r)/a^2$. The uniform density of apartments is $n = N/a^2$, the flux of incoming k-agents is $\gamma_k = \Gamma_k/a^2$.

For the densities of buyers and sellers at a specific location at distance r from the center we have:

$$n_b(r,t) = \sum_k \gamma_k \frac{U_k(r,t)}{Z_k(t)} \tag{4.15}$$

$$n_s(r,t) = \bar{n}_s(r,t) + \alpha\,(n - \bar{n}_s(r,t)) \tag{4.16}$$

In (4.15) $Z_k(t) = \frac{2\pi}{a^2}\int_0^{\bar{R}} rU_k(r,t)dr$ is the normalization factor[4]. In (4.16) $\bar{n}_s(r,t)$ is the density of apartments which were already on sale at time t: the total density of apartments on sale is given by this density plus the density of apartments newly put on sale, that is $\alpha\,(n - \bar{n}_s(r,t))$. The time evolution of $\bar{n}_s(r,t)$ is:

$$\partial_t \bar{n}_s(r,t) = \alpha\,(n - \bar{n}_s(r,t)) - v(r,t) \tag{4.17}$$

Here $v(r,t) = \sum_k v_k(r,t)$, consistently with (4.3), is the density of successful buyers. Notice that when $v(r,t) = 0\ \forall t$, $\bar{n}_s(r,t) \to n$: all the apartments will be waiting to be sold. It is not trivial to relate $v(r,t)$ to $n_b(r,t)$, because the market mechanism may prevent some buyers to get the apartment: some simplifications are needed. In this section we study a set of specific cases where the analytical treatment is possible: in Section 4.2.1 we consider the simplest case, that is agents weigh only the attractiveness in their utility function (hence $\beta = 1$), there is no social component in the attractiveness ($\epsilon = 0$), there is only one category ($K = 1$). In Section 4.2.2 we relax the assumption that $\beta = 1$; in Section 4.2.3 we consider $\epsilon > 0$; in Section 4.2.4 we deal with 2 categories. Throughout this section we will use rule (i) of the order book (see Section 4.1.7) to ensure the matching, that is buyers (sellers) bid (ask) their reservation prices. The steady state is reached if all the independent variables of the model[5], $A_k(r,t), n_s(r,t), n_b(r,t), P(r,t)$, become constant in time. We denote their stationary values as: $A_k^\star(r), n_s^\star(r), n_b^\star(r), P^\star(r)$. Notice that there are several feedbacks in the model: the relative densities of buyers and sellers at

[4]From now on we will assume $a = 1$. It is actually equivalent for the continuum limit to consider $a \to 0$ or $L \to \infty$. What matters is that a is small compared to L. Even though we assume $a \to 0$ for the analysis, with values such as $a = 1$ and $L = 11$ the continuous approximation works well. Reminder: a ensures that Z is dimensionless

[5]The other variables can be written as a combination of those

location X influence its price and its attractiveness, which change the probability to visit location X, which itself determines the relative densities of buyers and sellers.

4.2.1 Baseline case

Here $\beta = 1, \epsilon = 0, K = 1$. We also set $A^0_{max} = 1$, since it is just a scale factor which cancels out[6] in (4.15). All the other parameters can vary. This is the simplest case because the density of buyers is fixed by the value of the intrinsic attractiveness:

$$n_b(r) = \gamma \frac{e^{-r^2/R^2}}{Z}, \;\; Z = \pi R^2 (1 - e^{-\bar{R}^2/R^2})$$

What is more, provided that $n_s(r,t) \geq n_b(r)$, all the buyers succeed in getting an apartment, since from Eq. (4.8) $\underline{P^s_i}(X,t) \leq Y$, so in Eq. 4.17 we have $v(r,t) = n_b(r)$. Replacing this in Eq. 4.17 we get:

$$\partial_t \bar{n}_s(r,t) = \alpha \left(n - \bar{n}_s(r,t)\right) - n_b(r) \tag{4.18}$$

Solving equation (4.18) we have:

$$\bar{n}_s(r,t) = n - \frac{1}{\alpha}\left[n_b(r) + \left(\alpha n - n_b(r)\right) e^{-\alpha t}\right] \tag{4.19}$$

From (4.16):

$$n_s(r,t) = n - \frac{1-\alpha}{\alpha}\left[n_b(r) + \left(\alpha n - n_b(r)\right) e^{-\alpha t}\right] \tag{4.20}$$

Let us consider some limiting cases of (4.20):

- $\lim_{t\to 0} n_s(r,t) = \alpha n$, consistent with the fact that at the beginning a fraction α of the n housed agents at distance r put their apartment on sale.
- $\lim_{t\to\infty} n_s(r,t) = n - \frac{1-\alpha}{\alpha} n_b(r)$

We can compute the modified market price as a function of the previous market price, i.e.

$$P^M(r,t) = \frac{P(r, t-\delta t)}{1 + e^{-\phi(r,t)}} \tag{4.21}$$

with $\phi(r,t) = \frac{n_b(r) - n_s(r,t) - \tilde{\mu}}{\tilde{\sigma}}$. Here $\tilde{\mu}$ and $\tilde{\sigma}$ are density parameters, i.e. $\tilde{\mu} = \mu/a^2$, $\tilde{\sigma} = \sigma/a^2$. For simplicity, hereafter we omit the tilde.

Now, the market price can be computed as an average between the reservation prices of the buyers and of the sellers. The buyers have simply reservation price Y; the computation in the case of the sellers is slightly more

[6] This is not true in Section 4.2.3, where we keep A^0_{max}

involved. If one assumes that prices do not increase too fast, the second term of the max in Eq. (4.8) will generally be larger (more precisely, this is true provided that $P^i(r,t_i) \geq P^M(r,t)$). Notice that this is always true in the steady state, as the denominator in Eq. (4.21) is greater than one. Thus, there exists a neighborhood of the stationary value $P^\star(r)$ in which:

$$\underline{P_i^s}(r,t) = P^M(r,t) + \lambda^{t-t_i}\left(P^i(r,t_i) - P^M(r,t)\right) \tag{4.22}$$

Moreover, the stochastic variable $t - t_i$ follows a geometric distribution with parameter α for the time at which the apartment is put on sale on the first time. Of course, it may take some time to succeed in selling the apartment (and it does happen so especially in the least attractive locations), so the geometric distribution gives a lower bound for the time at which the apartment is sold. Given these assumptions, the upper bound for the expected reservation offer price writes:

$$\mathbb{E}\{P_i^s\}(r,t) = P^M(r,t) + \frac{\alpha\lambda}{1-\lambda(1-\alpha)}\left[P(r,t-1/\alpha) - P^M(r,t)\right] \tag{4.23}$$

The constant term in the above equation is just $\mathbb{E}\{\lambda^k\}$, with k following a geometric distribution with parameter α. So:

$$\mathbb{E}\{\lambda^k\} = \sum_{k=0}^{\infty}\lambda^k(1-\alpha)^{k-1}\alpha = \alpha\lambda\sum_{k=0}^{\infty}\left(\lambda(1-\alpha)\right)^{k-1} = \frac{\alpha\lambda}{1-\lambda(1-\alpha)} \tag{4.24}$$

Because of the expected time lag $1/\alpha$ it is not possible in general to study the dynamics of (4.23). However, the stationary state can be studied by taking off temporal dependance.

Once the average offer price has been computed, the market price is just:

$$P(r,t) = \frac{Y + P^M(r,t)\frac{1-\lambda}{1-\lambda(1-\alpha)} + P(r,t-1/\alpha)\frac{\alpha\lambda}{1-\lambda(1-\alpha)}}{2} \tag{4.25}$$

By replacing (4.21) in the above equation and setting $\alpha = 1$ we get[7]:

$$\partial_t P_1(r,t) = -H_1(P_1(r,t) - H_0) \tag{4.26}$$

where $H_1 = \frac{(2-\lambda)(1+e^{-\phi})-(1-\lambda)}{2(1+e^{-\phi})} > 0$. Since H_1 is positive $\forall\phi(r,t)$, the fixed point $P^\star(r)$ is locally stable[8]. For general α the stationary market

[7]This is not realistic, as it would imply that agents put their house on sale immediately after they bought it. We just take it as a limiting case to study the stability

[8]The neighborhood is the set of times t s.t. $P(r,t-1) \geq P^M(r,t)$. With respect to (4.8) we replace the individual $P^i(r,t-1)$ by its expected value

price reads (this computation is just algebra, it is enough to replace (4.21) in (4.25) and take off the temporal dependence):

$$P^\star(r) = \frac{Y(1+e^{-\phi^\star(r)})(1-\lambda(1-\alpha))}{(2-2\lambda+\alpha\lambda)(1+e^{-\phi^\star(r)})-(1-\lambda)} \tag{4.27}$$

$$\phi^\star(r) = \frac{\frac{\gamma}{\alpha}\frac{e^{-r^2/R^2}}{\pi R^2(1-e^{-\bar{R}^2/R^2})} - n - \mu}{\sigma} \tag{4.28}$$

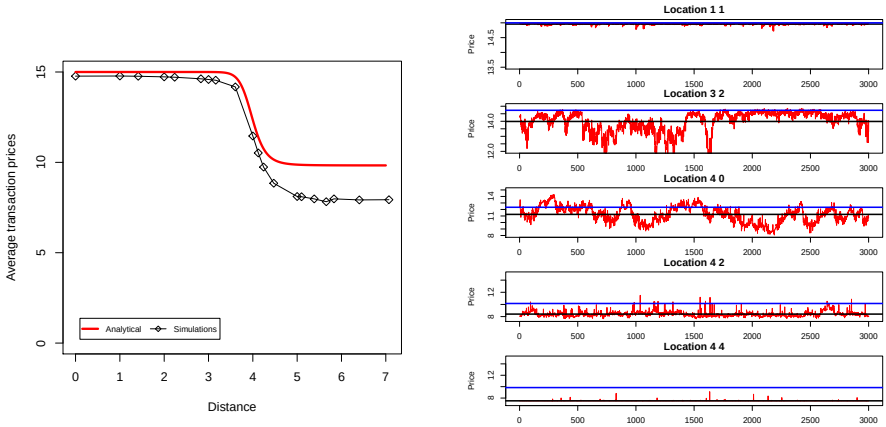
These equations express the price in closed form, so it is possible to study some limiting cases:

- $\lim_{\lambda\to 1} P^\star(r) = Y$. If the sellers do not accept to relax their reservation price towards $P^{M,\star}(r)$ the prices never decrease and settle to the income level.
- $\lim_{\phi\to\infty} P^\star(r) = Y$. If the density of buyers is constantly above that of sellers (taking into account μ and σ) the prices settle to the income level.
- $\lim_{\phi\to-\infty} P^\star(r) = Y\frac{1-\lambda(1-\alpha)}{2-2\lambda+\alpha\lambda}$. In the reverse case, prices do not go to zero since they result from a bargaining process where the buyers have reservation price Y.

Notice how in the first two cases the prices only increase, and are just limited by the constraint on the income level of the category. This is the main result of this section: if the sellers do not accept to decrease their price, or the flux of buyers is constant despite the increase in the price (as it is the case here, since buyers take decisions about where to look for an apartment only based on the attractiveness), prices just increase. Even though this cannot be directly seen in the steady state, it means that the location at distance r underwent a housing bubble which stopped only because the prices reached the income level Y, and the location is stuck in such situation.

Another interesting insight is that in (4.27) and (4.28) the parameters do not combine in effective parameters. The reason is due to the complexity of the model. For instance, one could think that the ratio $\frac{\gamma}{\alpha}$ should be an effective parameter. Nevertheless, α alone determines the time lag between the purchase of the apartment and the moment in which it is put on sale, so it determines the price independently.

Figure 4.4 shows the comparison between the analytical results and the simulations of the agent-based model. The analytical curve in both Figs. 4.4a and 4.4b is an upper bound for the simulated prices, and the spread increases with the distance from the center. This result was expected, since in the locations farthest from the center a long time might lapse until a seller succeeds in selling his apartment: the computation in (4.23) gives only an approximate result.



(a) (b)

Figure 4.4: (*Left*) Market price as a function of the distance from the center. The analytical prices are from (4.27), and they are compared with the results from the numerical simulations (carried out with rule (i) for the order book, as throughout in Section 4.2). Here and in all the plots that follow the simulated quantities are averaged over all the locations at a given distance from the center. The relevant parameter values other than β, K and ϵ are: $n = 100$, $L^2 = 121$, $Y_1 = 15$, $\alpha = 0.1$, $\lambda = 0.9$, $\mu = -30$, $\sigma = 10$, $\gamma = 1210$. (*Right*) Price evolution at some locations. The blue lines represent the analytical market price, the black lines are the average over the represented time interval, the red lines are sequences of market prices. The fluctuations are large in the transition region.

4.2.2 General utility

In this section we consider the case in which $0 \leq \beta \leq 1$, with $\epsilon = 0$ and $K = 1$. As explained in Section 4.1.5 the buyers face a tradeoff between very attractive locations and a high level of consumption, which is possible only if the prices of the apartments are not too high.

Since only the utility function is different, Eq. (4.27) is unchanged. Also (4.26) does not change, so the fixed point $P^\star(r)$ is still stable. On the other hand, Eq. (4.28) becomes:

$$\phi^\star(r) = \frac{\frac{\gamma}{\alpha}\frac{(Y-P^\star(r))^{1-\beta}e^{-\beta r^2/R^2}}{Z^\star} - n - \mu}{\sigma} \tag{4.29}$$

Here $Z^\star = 2\pi \int_0^{\bar{R}} w\,(Y - P^\star(w))^{1-\beta}\, e^{-\beta w^2/R^2} dw$.

It is still possible to study the case $\beta = 0$ in closed form. Since there is no space diversification anymore, it is:

$$P^\star = \frac{Y(1+e^{-\phi^\star})(1-\lambda(1-\alpha))}{(2-2\lambda+\alpha\lambda)(1+e^{-\phi^\star}) - (1-\lambda)} \tag{4.30}$$

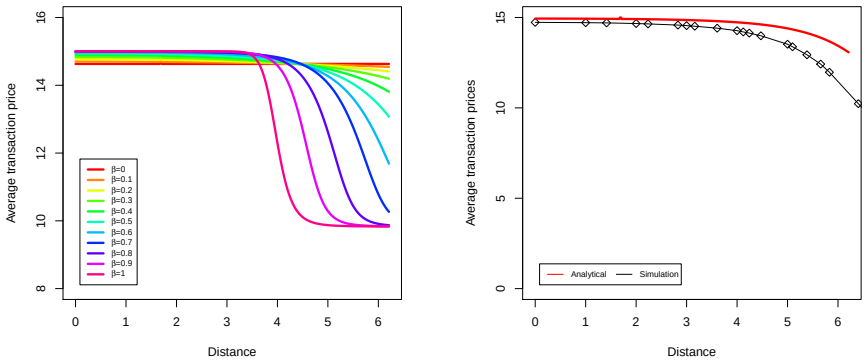
$$\phi^\star = \frac{\frac{\gamma}{\alpha}\frac{1}{\pi R^2} - n - \mu}{\sigma} \tag{4.31}$$

The case $0 < \beta < 1$ needs an iterative solution. We start from an arbitrary value for $Z^\star$ and solve (4.27) (replacing Eq. (4.29)) numerically for 10000 values $0 < r < \bar{R}$, computing $Z^\star$ with the trapezoidal method. We iterate this procedure until convergence for $Z^\star$ is reached. As it can be seen in Fig. 4.5a, for $0 \leq \beta \leq 0.5$ the prices are almost uniform across the city, $P^\star(r) \approx Y$. This means that there is little or almost none excess supply (if it were the case, the prices would decrease): if the agents take into account the price of the apartments in their indirect utility function, the markets clear more easily. In Figure 4.5b a comparison with the simulations is showed. Indeed, the very fact that there is not much excess supply implies that sellers succeed in selling their apartment shortly after they put it on sale, and so (4.23) gives a good estimate.

4.2.3 Social component

We consider $\epsilon > 0$, but still $\beta = 1$ and just one income category. The social component means here that there is a positive feedback in the attractiveness: the more agents choose a location, the more it becomes attractive. We denote $A_k(r,t) = A(r,t)$. Differently from Sections 4.2.1 and 4.2.2 we limit ourselves to the characterization of the steady state, without caring about the stability. In the steady state:

$$A^\star(r) = A^0(r) + \frac{\epsilon}{\omega} v^\star(r) \tag{4.32}$$



(a) (b)

Figure 4.5: (*Left*) Market price as a function of the distance from the center, for several values of β. The lower is β, the more prices are uniform. (*Right*) Comparison between the analytical solution and the numerical simulations for $\beta = 0.5$. All the other parameters are as in Fig. 4.4a. Even for the locations farthest from the center, the analytical curve lies close to the results from the simulations.

It is necessary to distinguish between two regions: the one closer to the center, that is for $r < r_c$, where r_c is a *critical radius*, such that $n_b^\star(r) \geq \alpha n = v^\star(r)$: all the apartments which are put on sale are immediately bought. The one farther from the center: for $r > r_c$, it is only $n_b^\star(r) = \gamma \frac{A^\star(r)}{Z} = v^\star(r)$. This gives a self consistent expression for the attractiveness: replacing in (4.32) we get $A^\star(r) = \frac{A^0(r)}{1-\frac{\epsilon}{\omega}\frac{\gamma s_0}{Z}}$. To make it clear that the denominator is dimensionless, we explicitly write s_0. Summing up, the attractiveness writes:

$$A^\star(r) = \begin{cases} A^0(r) + \frac{\epsilon}{\omega}\alpha n, & 0 < r < r_c \\ \frac{A^0(r)}{1-\frac{\epsilon}{\omega}\frac{\gamma s_0}{Z}}, & r_c < r < \bar{R} \\ 0, & r > \bar{R} \end{cases} \tag{4.33}$$

The computation of the normalization factor gives:

$$Z = 2\pi \int_0^{r_c} r A^\star(r) dr + 2\pi \int_{r_c}^{\bar{R}} r A^\star(r) dr =$$
$$= \pi R^2 A^0_{max} \left(1 - e^{-r_c^2/R^2}\right) + \pi r_c^2 \frac{\epsilon}{\omega}\alpha n + \pi R^2 A^0_{max} \left(e^{-r_c^2/R^2} - e^{-\bar{R}^2/R^2}\right) \frac{1}{1-\frac{\epsilon}{\omega}\frac{\gamma s_0}{Z}} \tag{4.34}$$

The critical radius can be computed from the boundary condition in Eq.

(4.33) and reads:

$$r_c = R\sqrt{\ln \frac{\gamma A^0_{max}}{\alpha n}\left(Z - \frac{\epsilon}{\omega}\gamma s_0\right)^{-1}} \tag{4.35}$$

Replacing (4.35) in (4.34) it is possible to compute Z numerically. There are two problems:

- It may well be that Z does not exist, for instance if $1 - \frac{\epsilon}{\omega}\frac{\gamma s_0}{Z} < 0$: one ends up with a negative argument for a logarithm. From the point of view of the model, it means that there is an inconsistency: for example, the flux of incoming agents γ is too strong, and all the locations get at least αn buyers, so the differentiation in zones $r < r_c$ and $r > r_c$ does not make sense.

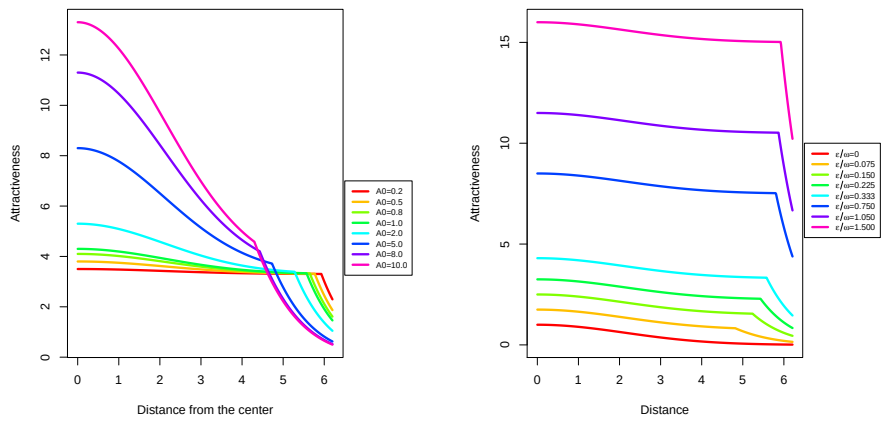
- Even with the parameters chosen consistently, there are usually two solutions for Z. However, one of them is such that $r_c > \bar{R}$, which is itself an inconsistency, so it can be neglected.

One thing which is apparent from (4.34) and (4.35) is that it is possible to consider $\frac{\epsilon}{\omega}$ as an effective parameter, since only this ratio appears in the formula. It is less trivial to consider the interplay between the other parameters. Here it is of particular interest the choice of A^0_{max} compared to that of ϵ/ω. As it can be seen in Figures 4.6 and 4.7a it is equivalent to keep A^0_{max} low for ϵ/ω fixed or ϵ/ω high for A^0_{max} fixed in order to have more space uniformity: the critical radius is closer to the edge of the city, $r_c \approx \bar{R}$. What it matters is the interplay between the broken translational symmetry caused by $A^0(r)$ and the homogenization effect due to the social component. This is the main result of this section: the social component of the attractiveness tends to level out the probability field over the city[9], as it gives the agents a reason to choose the locations where the intrinsic attractiveness is lower. Notice that this is true because the social component of the attractiveness is additive with respect to the intrinsic part, as we assume in this work (see Section 4.1.4).

The comparison with the numerical simulations, as it can be seen in Fig. 4.7b, is not as good as in the former cases[10]. In particular, the transition between the two regions appears smoother. This is due to stochastic effects, which should not be neglected. In Fig. 4.8a (results from the simulations) it is possible to see that there are $\alpha n = 10$ sellers only for $r \approx 0$, and then the number of sellers increases steadily. Actually, the probability that $n_b(r,t) \leq \alpha n$ is given by the cumulative of a binomial distribution with γ

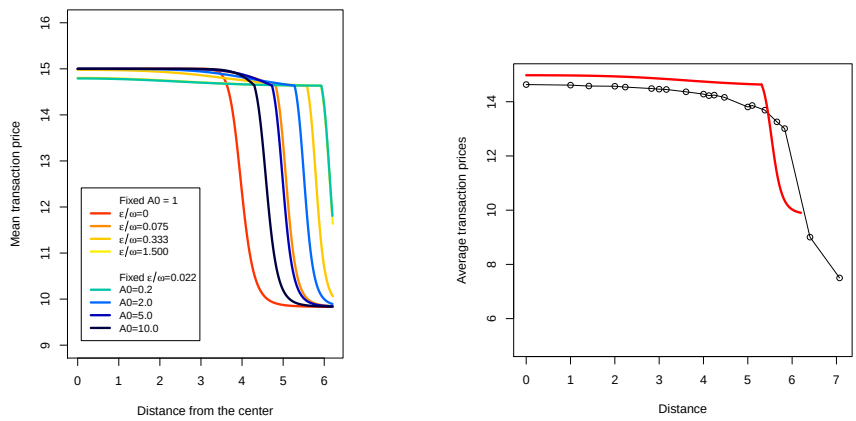
[9] Thus the outcome is the same as for small β, though for completely different reasons

[10] Notice however that the maximum distance from the center is higher in the case of the simulations, so to compare the location of the transitions one should rescale the distance



(a) (b)

Figure 4.6: (*Left*) Attractiveness in the steady state for several values of A^0_{max} (A0 in the plot), for $\epsilon = 0.022$ fixed. All the other parameters are as in Fig. 4.4a. If A^0_{max} is high the social effect plays almost no role, i.e. the shape of the attractiveness curve is the same as in the case without social influence. (*Right*) Attractiveness in the steady state for several values of ϵ, for $A^0_{max} = 1$ fixed. All other parameters are as in Fig. 4.4a. If ϵ is high, there are two effects: the offset of the attractiveness curve is higher, the transition is more apparent



(a) (b)

Figure 4.7: (*Left*) Interplay between A^0_{max} (A0 in the plot) and $\frac{\epsilon}{\omega}$. All the other parameters are as in Fig. 4.4a. (*Right*) Comparison between the analytical solution and the numerical simulations for $\epsilon = 0.022$, $\omega = 1/15$. All other parameters are as in Fig. 4.4a: with these parameter values $r_c = 5.31$. The decline in prices with distance from the center is less sharp in the case of the numerical simulations

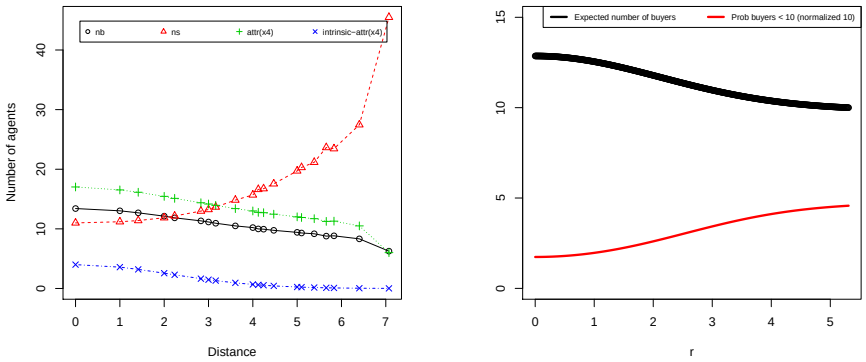
trials and parameter (see Fig. 4.8b):

$$\pi^\star(r) = \frac{A^0_{max} e^{-r^2/R^2} + \frac{\epsilon}{\omega}\alpha n}{Z} \tag{4.36}$$

This gives for the probability that $n_b(r,t) \leq \alpha n$, by assuming that αn and γ are integers[11] (for the parameter values in Figure 4.8a it is $\alpha n = 10$ and $\gamma = 1210$):

$$\mathbb{P}\left(n_b(r,t) \leq \alpha n\right) = \sum_{n_b(r,t)=0}^{\alpha n} \binom{\gamma}{n_b(r,t)} \left(\pi^\star(r)\right)^{n_b(r,t)} \left(1 - \pi^\star(r)\right)^{\gamma - n_b(r,t)} \tag{4.37}$$

If $n_b(r,t) \leq \alpha n$, on the next time step $\bar{n}_s(r, t+\delta t) > 0$, which invalidates at least for the simulations the argument presented at the beginning of this section. Nevertheless, the most important considerations still hold.



(a) (b)

Figure 4.8: (*Left*) The number of buyers, sellers, the total attractiveness (multiplied by 4) and the intrinsic attractiveness (multiplied by 4) as a function of the distance from the center. These figures are averaged over 1000 time steps. The parameters are as in Fig. 4.7b. It is not true that $n^\star_s(r) = \alpha n = 10, \forall r < r_c$. (*Right*) The expected number of buyers and the probability (normalized to 10) that less than 10 buyers show up, as a function of the distance from the center. Even though $\forall r < r_c, n^\star_b(r) \geq 10$, the probability that this is not true at time step t is almost 1/2 at $r \approx r_c$.

[11]The computation can be generalized to the case where they are not integers, by replacing the summation with an integral and by writing the binomial coefficient as a combination of Gamma functions

4.2.4 Two categories

We consider two income categories, defined by their income levels Y_1 and Y_2, $Y_2 = Y_1 + \Delta$. In the following we may denote 1-agents as "poor agents" and 2-agents as "rich agents", but this does not refer to any specific feature of the agents apart from their income. In order to understand the conditions for segregation (which will be quantified by the information entropy, see below) we focus on the simplest case, that is $\beta = 1$ and $\epsilon = 0$. However, as it will be clear by the end of this chapter, the same considerations as in Sections 4.2.2 and 4.2.3 still apply.

We determine the conditions under which there exists a critical radius r_c such that the density of housed 1-agents, $n_{h1}(r)$, is zero $\forall r < r_c$. In other words, the circle with center O and radius r_c is inhabited only by the 2-agents: there is total segregation. This happens if $P^\star(r) > Y_1, \forall r < r_c$. If the segregated region exists we can characterize it as in Section 4.2.1, so we can use Eq. (4.27) for the price (with $Y = Y_2$ and $\gamma = \gamma_2$). The condition for r_c is thus:

$$P^\star(r_c) = \frac{Y_2(1 + e^{-\phi^\star(r_c)})(1 - \lambda(1 - \alpha))}{(2 - 2\lambda + \alpha\lambda)(1 + e^{-\phi^\star(r_c)}) - (1 - \lambda)} = Y_1 \tag{4.38}$$

$$\phi^\star(r_c) = \frac{\frac{\gamma_2}{\alpha}\frac{e^{-r_c^2/R^2}}{\pi R^2(1-e^{-\bar{R}^2/R^2})} - n - \mu}{\sigma} \tag{4.39}$$

Rearranging (4.38) we get:

$$1 + e^{-\phi^\star(r_c)} = \frac{(1 - \lambda)Y_1}{Y_1(2 - 2\lambda + \alpha\lambda) - Y_2(1 - \lambda + \alpha\lambda)} \tag{4.40}$$

For r_c to exist, the RHS in the above equation must be greater than one. It turns out that this is always the case, provided that the denominator is positive. So the following inequality has to hold:

$$\frac{\Delta}{Y_1} < \frac{1 - \lambda}{1 - \lambda + \alpha\lambda} \tag{4.41}$$

Since $\frac{\Delta}{Y_1} > 0$, it is clear that if $\lambda \to 1$ the inequality is never satisfied. This suggests the interpretation for Eq. (4.41): if prices are not able to decrease enough even at the locations very far from the center, there is total segregation all across the city. Actually, the condition could be obtained from the limiting case $\phi \to -\infty$ in Eq. (4.27). This is an important result: if the sellers do not accept to decrease their price, only the category with the highest income will be able to settle in the city. Replacing (4.39) in (4.40) the condition for the existence of r_c becomes:

$$e^{-\frac{r_c^2}{R^2}} = \frac{\pi R^2(1 - e^{-\bar{R}^2/R^2})}{\gamma_2/\alpha}\left[n + \mu - \sigma \ln\left(\frac{\Delta/Y_1}{\frac{1-\lambda}{1-\lambda+\alpha\lambda} - \frac{\Delta}{Y_1}}\right)\right] \tag{4.42}$$

The RHS in the above equation has to be bounded between zero and one for r_c to exist. In the limiting case $r_c \to \infty$ the RHS goes to zero: a small value for the RHS means that total segregation occurs in a large part of the city. A negative value or a value such that $r_c > \bar{R}$ implies that the city is only inhabited by the rich agents. In the opposite case $r_c \to 0$, the RHS is close to one: if $RHS > 1$ there is no segregation, as it would be $r_c < 0$. Thus, the larger the RHS, the less the city is segregated. This is consistent with the parameters in the RHS: if the denominator in the logarithm is small (that is, (4.41) holds loosely) there is a significant negative contribution from the logarithm, which is balanced by n (if there are many available apartments it is less likely to have total segregation) and by μ: a small μ means that prices decrease even if there are slightly more sellers than buyers. An interesting insight in (4.42) is that only the ratio Δ/Y_1 matters: there is only one effective parameter in the incomes of the agents, that is the spread between income levels with respect to the lowest income level. Finally, a small value for γ_2 means that few rich agents are coming to the city: since they are those with the larger market power, their scarceness means that the market price is less likely to increase above Y_1. The result holds for any value of K: if the K-agents are not too many, they will not get completely segregated. This brings up the issue of what "too many" means quantitatively. One can define a critical flux of incoming rich agents γ_{2c} such that for $\gamma_2 \leq \gamma_{2c}$, $r_c = 0$: nowhere in the city there is total segregation. γ_{2c} is obtained by setting the RHS of (4.42) equal to one:

$$\gamma_{2c} = \alpha\pi R^2(1 - e^{-\bar{R}^2/R^2}) \left[n + \mu - \sigma \ln \left(\frac{\Delta/Y_1}{\frac{1-\lambda}{1-\lambda+\alpha\lambda} - \frac{\Delta}{Y_1}} \right) \right] \tag{4.43}$$

γ_{2c} defines a phase transition in the parameter space, between a region where total segregation does not occur anywhere and a region where the area $0 < r < r_c$ is only inhabited by the 2-agents. In order to characterize the phase transition, we need to define an order parameter for the segregation. We choose to use the information entropy: at distance r it is defined:

$$S(r) = -\nu_1(r) \log \nu_1(r) - \nu_2(r) \log \nu_2(r) \tag{4.44}$$

Here $\nu_1(r) = \frac{n_{h1}(r)}{n_{h1}(r)+n_{h2}(r)}$ and $\nu_2(r) = \frac{n_{h2}(r)}{n_{h1}(r)+n_{h2}(r)}$. $S(r)$ ranges between a minimum value for total segregation, $S(r) = 0$, s.t. only one category is living at distance r from the center, and a maximum value, $S(r) = \log 2$, s.t. $n_{h1}(r) = n_{h2}(r)$: there is perfect social mixing. As an order parameter, we define the information entropy averaged over the city:

$$S = \frac{2\pi}{\pi\bar{R}^2} \int_0^{\bar{R}} rS(r)dr \tag{4.45}$$

In order to actually compute the above equation, we need to compute the relative densities of housed 1 and 2-agents. For $r < r_c$, $n_{h1} = 0$ and so $S(r) = 0$; for $r > r_c$, since we are using rule (i) for the order book, there is no bias in favor of the rich agents, so the following equalities hold:

$$\frac{n_{b1}}{n_{b2}} = \frac{n_{h1}}{n_{h2}} = \frac{n_{s1}}{n_{s2}} \tag{4.46}$$

These also imply $\frac{n_{h1,2}}{n_h} = \frac{n_{b1,2}}{n_b}$: we can use equivalently the fractions of buyers and of housed agents in (4.44). The total number of buyers at distance r is simply:

$$n_b(r) = \gamma_2 \frac{e^{-r^2/R^2}}{Z} + \gamma_1 \frac{e^{-r^2/R^2}}{Z_1(r_c)} \tag{4.47}$$

where $Z_1(r_c) = 2\pi \int_{r_c}^{\bar{R}} r e^{-r^2/R^2} dr$ is the normalization factor for the poor agents, taking into account that the probability that they look for an apartment at locations at distance $r < r_c$ is null (see Section 4.1.7). We get:

$$\nu_1(r_c) = \frac{\gamma_1}{\gamma_1 + \gamma_2 \psi(r_c)}, \nu_2(r_c) = \frac{\gamma_2 \psi(r_c)}{\gamma_1 + \gamma_2 \psi(r_c)} \tag{4.48}$$

$$\psi(r_c) = \frac{e^{-r_c^2/R^2} - e^{-\bar{R}^2/R^2}}{1 - e^{-\bar{R}^2/R^2}} \tag{4.49}$$

$\psi(r_c)$ in (4.49) is the fraction of rich agents who do not live in the segregated area. The total entropy is:

$$S(r_c) = \frac{\bar{R}^2 - r_c^2}{\bar{R}^2} \left[-\nu_1(r_c) \log \nu_1(r_c) - \nu_2(r_c) \log \nu_2(r_c) \right] \tag{4.50}$$

where r_c is computed from (4.42) and $\nu_1(r_c)$ and $\nu_2(r_c)$ are defined in (4.48). For $\gamma_2 < \gamma_{2c}$, $r_c = 0$. As it can be seen in Fig. 4.9, the derivative of S is discontinuous at γ_{2c}: with respect to the information entropy, the phase transition is second-order[12]. In the inset of Fig. 4.9 the information entropy starts decreasing for $\gamma_2 > \gamma_{2c}$, so the phase transition is situated at the maximum level of social mixing. This is not true when $\gamma = 1210$, but this is simply because for high values of γ_2 the critical radius r_c increases less than linearly, and so the loss of surface in the area of social mixing (that is for $r_c < r < \bar{R}$) does not counterbalance the higher social heterogeneity in the same area (recall that most 2-agents live at $0 < r < r_c$).

[12]Formally, the order of the phase transition should be related to the free energy, yet it is not clear whether in our case it can be linked to the information entropy. We define the order of the phase transition on the basis of the kind of discontinuity of the order parameter. Notice that Gualdi *et al.* (2015) consider the unemployment rate as a function of the interest rate: they find that the unemployment jumps discontinuously to zero for a critical value of the interest rate. They classify this transition as first-order

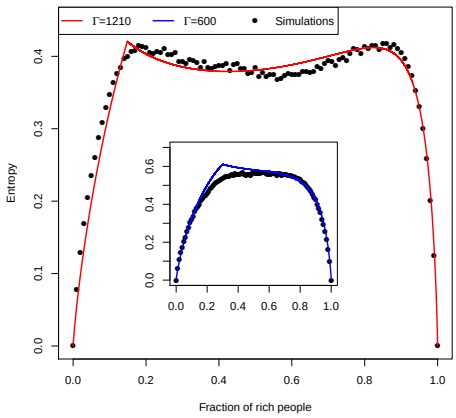


Figure 4.9: Entropy as a function of γ_2/γ, for two values of γ. The comparison with the results from the simulations is good, but the phase transition is not apparent in the simulations, mostly in the case $\gamma = 600$. Notice that γ_{2c} does not depend on γ.

We finally compute the prices for $r > r_c$, by generalizing the arguments given in Section 4.2.1. The most interesting finding is that there exists a region $r_c < r < \tilde{r}$ where the prices never settle to their stationary values: there are not enough 2-agents coming, so they cannot keep the market price $P(r,t) > Y_1$, but the intense flux of 1-agents coming to the locations at distance $r_c < r < \tilde{r}$ brings the price up until it overcomes Y_1 and it starts decreasing again, because the density of buyers dramatically drops. Notice that this phenomenon has no particular economic meaning, as it is simply a consequence of the simplification of this section, that is to consider just two income categories. Nevertheless, it helps explaining why the phase transition in Fig. 4.9 is not apparent in the simulations, especially for $\gamma = 600$.

We consider the market mechanism described in Section 4.1.7 with order book rule (i) and two categories, as it is the case for all locations at distance $r_c < r < \bar{R}$. On the demand side the density of buyers is given by Eq. (4.47); on the offer side, the ratio between 1 and 2-sellers is the same as that between 1 and 2-buyers (see Eq. (4.46)). The total density of sellers is given by (4.16) and (4.19), with the density of successful buyers in (4.16) given by (4.47). The densities of buyers and sellers determine $\phi(r,t)$. Here we first study the stability for an arbitrary $\phi(r,t) = \phi$ and setting $\alpha = 1$. Then we characterize the steady state, if it exists, and we show the comparison with the results from the simulations.

Stability

In the order book buyers are matched randomly with sellers, so it makes sense to aggregate all reservation offer prices in one expected reservation offer price:

$$\mathbb{E}\{P_i^s\}(r,t) = \frac{n_{b1}(r,t)\mathbb{E}\{P_{i1}^s\}(r,t) + n_{b2}(r,t)\mathbb{E}\{P_{i2}^s\}(r,t)}{n_b(r,t)} \tag{4.51}$$

where we have used (4.46) and we have considered a weighted average of the two categories. On the other hand, the reservation prices of the buyers are heterogeneous, and so it is needed to define two average transaction prices, according to the category of the buyer:

$$P_1(r,t) = \frac{Y_1 + \mathbb{E}\{P_i^s\}(r,t)}{2} \tag{4.52}$$

$$P_2(r,t) = \frac{Y_2 + \mathbb{E}\{P_i^s\}(r,t)}{2} \tag{4.53}$$

Since they differ just by an additive constant, we can consider just the transaction price of the 1-agents acting as buyers, and get the other price from:

$$P_2(r,t) = P_1(r,t) + \frac{Y_2 - Y_1}{2} \tag{4.54}$$

The market price is:

$$P(r,t) = \frac{n_{b1}(r,t)P_1(r,t) + n_{b2}(r,t)P_2(r,t)}{n_b(r,t)} \tag{4.55}$$

and the modified market price reads as usual:

$$P^M(r,t) = \frac{P(r,t-\delta t)}{1+e^{-\phi(r,t)}} \tag{4.56}$$

Following the same computation as in Section 4.2.1, in a neighborhood of the steady state $(P_1^\star(r), P_2^\star(r))$ where the second term of the max in Eq. (4.8) is greater than the first (see discussion in Section 4.2.1), we have for the time evolution of (4.52):

$$\partial_t P_1(r,t) = -L_1(P_1(r,t) - L_0) \tag{4.57}$$

where $L_1 = \frac{(2-\lambda)(1+e^{-\phi})-(1-\lambda)}{2(1+e^{-\phi})} > 0$, as in Eq. (4.26): the fixed point $P_1^\star(r)$ is locally stable, and so is $P_2^\star(r)$.

Steady state

For general α we find in the steady state:

$$P_1^\star(r) = \frac{Y_1(1+e^{-\phi^\star(r)})(1-\lambda(1-\alpha))}{(2-2\lambda+\alpha\lambda)(1+e^{-\phi^\star(r)})-(1-\lambda)} + \\ + \frac{n_{b2}}{n_b}\frac{Y_2-Y_1}{2}\frac{2(1+e^{-\phi^\star(r)})(1-\lambda+\alpha\lambda)+(1-\lambda)}{(1+e^{-\phi^\star(r)})(2-2\lambda+\alpha\lambda)-(1-\lambda)} \tag{4.58}$$

$$\phi^\star(r) = \frac{\frac{\gamma_2+\gamma_1/\psi(r_c)}{\alpha}\frac{e^{-r^2/R^2}}{\pi R^2(1-e^{-\bar{R}^2/R^2})} - n - \mu}{\sigma} \tag{4.59}$$

The first term in (4.58) is the same as in Eq. (4.27), with Y replaced by Y_1. The second term, which is always positive, involves an increase in the price, which is due to the 2-agents looking for an apartment at distance r. Indeed, if $n_{b2} = 0$ or $Y_2 = Y_1$ the second term vanishes. Notice again that it is the ratio Δ/Y_1 which matters[13]. We have for the market price:

$$P^\star(r) = P_1^\star(r) + \frac{n_{b2}}{n_b}\frac{Y_2-Y_1}{2} \tag{4.60}$$

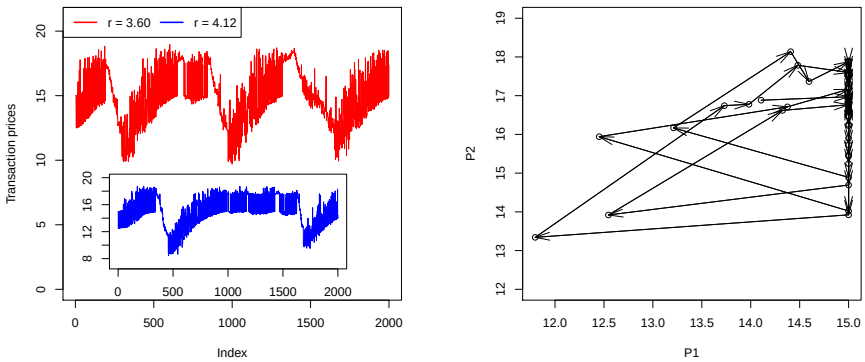
where we have used (4.54). We define $\tilde{r}$ as the distance from the center s.t. $P^\star(\tilde{r}) = Y_1$: in the region $r_c < r < \tilde{r}$ the prices never settle to their stationary value. This area is characterized by the fact that there are not enough 2-buyers to keep the market price steadily above Y_1, but there are too many 1-buyers to keep it steadily below Y_1. The phenomenon can be better understood looking at Fig. 4.10. In Fig. 4.10a the price evolution is shown[14]. For the location closer to the center the fluctuations are frequent: the number of 1-buyers is really high, so the price jumps immediately above Y_1. At the location farther from the center the frequency is lower: prices take more time to get to Y_1. In Fig. 4.10b one can see the reason why the stationary state is never reached: the fixed point $(P_1^\star, P_2^\star)$ does not lie in the half-plane defined by the condition $P_1 < Y_1$, so the dynamics cannot reach it and follows periodic cycles in the phase space instead. Figure 4.11 shows the comparison between the analytical results and the simulations: the mathematical analysis fails at identifying the distance where the prices start to drop and gives an overall higher price level. The behaviour of the prices is however qualitatively similar.

4.2.5 Effect of the parameters

Thanks to the analysis performed in this section we understand which parameters are the most relevant to determine the price distribution and the

[13] It can be seen more readily if one considers $P_1^\star(r)/Y_1$

[14] Notice that all the transaction prices are recorded, not just the average ones, which is why we have so many fluctuations



(a) (b)

Figure 4.10: (*Left*) Price sequence at location (3,2) (in red) and (4,1) (in blue). The parameters are as in Fig. 4.7b, apart from $K = 2$, $\gamma_1 = \gamma_2 = \gamma/2 = 605$. With these parameter values $r_c = 3.30$ and $\tilde{r} = 4.60$. (*Right*) Evolution of the average prices of subsequent time steps at location (3,2). We assume that $P_1(r,t) = Y_1$ whenever $P(r,t) > Y_1$ (formally, $P_1(r,t)$ is not defined, as no 1-buyers come to location r and complete a transaction, but this is useful for depiction). Here $\alpha = 1$,$\Delta = 3$,$Y_1 = 15$,$\lambda = 0.7$, the other parameters are as in the left picture. This choice was necessary to satisfy (4.41)

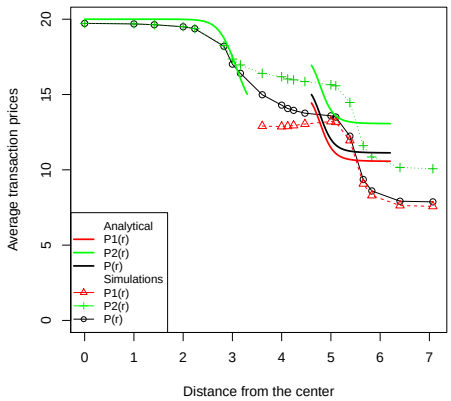


Figure 4.11: Average prices as a function of the distance from the center. The parameters are as in Figure 4.10

level of segregation in the agent-based model: The assertions in this section will be validated through fully-fledged simulations in Section 4.3.

- The discount factor λ, which determines how long it takes for the prices to start decreasing if there is not enough demand. A value $\lambda \to 1$ would imply that prices never decrease and settle at the highest income level, that is $P^\star(r) = Y_K, \forall r < \bar{R}$: there is a bubble in the real-estate market.
- The parameters, β and ϵ/ω which enter the utility function of the agents. A small value of β and a high value of ϵ/ω flatten the price distribution.
- The total number of incoming agents, γ, and the relative densities γ_k/γ. A strong demand from the "richest" agents implies a significant segregation. Only γ_K determines if there is total segregation anywhere (with just K-agents inhabiting a specific location).

All the other parameters matter marginally: a and δt only matter for technical reasons; K just gives how finely the income distribution is described; if the ratio Δ/Y_1 is large in absolute value (i.e. income levels are relatively spread, recall Eqs. (4.41) and (4.42)), there is slightly more segregation, but the income scale does not matter apart from a multiplicative constant in the price; if μ and σ are chosen reasonably (that is, $\mu < 0$ and σ relatively small compared to n) they do not influence qualitatively the behaviour of the model. In case μ is chosen to be large and negative, it has the same effect as $\lambda \to 1$, that is the prices do not decrease and settle at the highest income level (recall Eqs. (4.27) and (4.28)). For what concerns n and α, they can be rescaled with γ in Eqs. (4.28), (4.29) and (4.42), as it is apparent if one writes $\mu = \mu'\gamma$, $\sigma = \sigma'\gamma$ and $\gamma_2 = \gamma_2'\gamma$. Thus, it should be equivalent to consider high γ and small n, or high γ and small α. Nevertheless, notice that α and n do enter independently in Eq. (4.34), as the offset αn does not depend on γ. Furthermore, α determines the reservation offer price along with λ (a small α implies that the apartments are put on sale after a long time, so their price is more likely to decrease, for fixed λ), independently of γ. The fact that the parameters do not combine formally in effective parameter is an effect of the realistic assumptions of the model, but numerical simulations show that the effect of α, for fixed α/γ, is almost null. Finally, the intrinsic attractiveness $A^0(x, y)$ and L^2 (or equivalently $\bar{R}$) give the structure and the size of the city. For the choice in Eq. (4.12) R and $\bar{R}$ can be chosen with no loss in generality as a rescaled version of each other. Notice that this is also true with respect to β, as it can be seen from the exponential in Eq. (4.29): it is equivalent to consider a high β or a small R^2, or vice versa. A^0_{max} does not matter at all but in the case where the social component of the attractiveness is not null, where anyway it can be written in practice as a rescaled version of ϵ/ω (recall Fig. 4.7a).

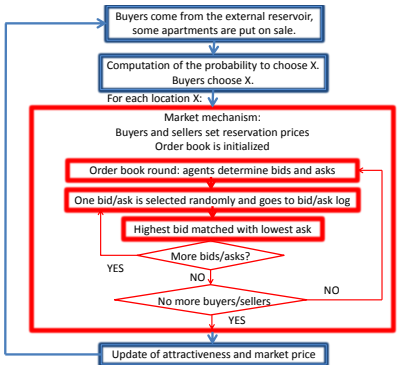


Figure 4.12: Flowchart representing the processes which occur at any time step for the agent-based model described in Section 4.1. Blue lines represent processes happening at the city level, red lines refer to processes at the location level.

4.3 Numerical simulations

In this section we perform some numerical simulations of the agent-based model described in Section 4.1. Figure 4.12 sums up the sequence of the actions occurring during a time interval, that is between t and $t + \delta t$. The code which has been used for the simulations is shown in Appendix D. We use the power of computer simulations to investigate some situations which are not manageable analytically. We use rule (ii) for the order book, which constitutes a bias in favour of the buyers from the richest categories, since they are the ones who increase their proposed price faster (see Eq. (4.11)). In Section 4.3.1 we perform numerical explorations on the most important parameters, as listed in Section 4.2.5; in Section 4.3.2 we verify the other assertions in Section 4.2.5, that is we check that the other parameters either do not matter much in determining the qualitative behaviour of the model or have the same effect of the important parameters, with some caveat.

4.3.1 Important parameters

We consider a larger number of categories: we set $K = 5$. Following the discussion in Section 4.2.5, we focus on $\{\gamma_k\}$,β and ϵ/ω. For comparison with the results of the previous section, we choose to model the intrinsic attractiveness through Eq. (4.12).

Figure 4.13 shows the spatial distribution of the k-agents for some variations on their relative frequencies. In Fig. 4.13a the income distribution is uniform over the values $\{Y_k\}$ (case (a)); in Fig. 4.13b the distribution is linear, decreasing for higher Y_ks (case (b)); in Fig. 4.13c we consider a Pareto distribution $f(Y) = \frac{C}{Y^{\theta+1}}$, whose parameter θ is set by the Gini coef-

ficient of France, which is taken to be $G = 0.32$ (data from the World Bank, year 2005). If one assumes that the distribution of income is a Pareto distribution[15], the equality $G = \frac{1}{2\theta-1}$ holds. Once θ has been determined (we find $\theta = 2.0625$), it is possible to compute all γ_k and then the normalization factor (case (c)). Case (a) is that with the highest income inequality, since the largest part of the available income is concentrated in the hands of the (many) agents from the richest category. It is apparent that when the share of the agents from the highest income levels decreases, "poorer" agents are able to settle closer to the center. Notice that 1-agents almost cannot find a dwelling in the city unless in Fig. 4.13c: the reason is simply that in the other cases 2-agents are too many with respect to 1-agents, and the same arguments as in Section 4.2.4 apply. Anyway, at most locations some social mixing between the other categories is usually preserved.

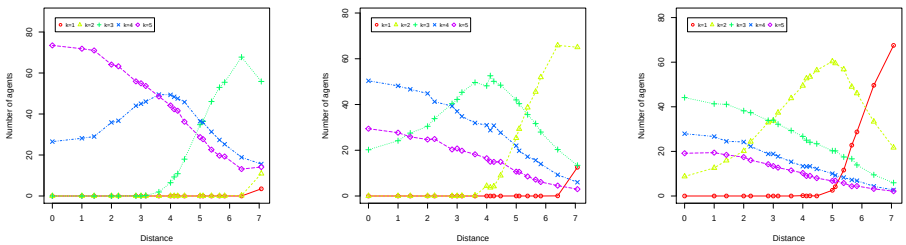
Figure 4.14 compares the distribution of the agents for some choices of β and ϵ (with ω fixed). The more the attractiveness field is flattened, the more the 1-2-3 agents distribute uniformly in the space: however, if this is the case, 5-agents face higher prices in all locations and almost disappear from the city. Actually, given the assumptions in Section 4.1.4, the social component of the attractiveness is additive, so the relative difference between the maximum and the minimum level of attractiveness is higher if there is no social component, $\epsilon = 0$. This enhanced space uniformity acts as a mechanism which prevents the richest agents to segregate in some areas, but drives away the poorest agents from the city: it is a case of *gentrification*. Nonetheless, 4 and 5-agents have a preference for the center, where presumably the prices will be higher. In order to quantify the global level of segregation, we use again the information entropy (4.44), discretized and generalized to K categories[16]. We compute it for the different cases in Fig. 4.14: in Fig. 4.14a $S = 0.93$, in Fig. 4.14b $S = 1.13$, in Fig. 4.14c $S = 1.29$. Interestingly, the entropy is higher when the poorest income category almost cannot live in the city, but the densities of the other income categories are more spatially uniform.

We finally study the price distribution with the same parameter settings as in Figs. 4.13 and 4.14. As it is shown in Fig. 4.15, the average prices decline smoothly from the center, without discontinuities as it was the case with one and two categories. The main result is that the income distribution $\{\gamma_k\}$ determines the offset of the price distribution, whereas the preferences of the agents, ϵ and β, determine the steepness. This is an unexpected result, since the income categories appreciate the locations in a different way, so one could expect that raising the frequency of the richest category would raise the prices mostly in the center. This is not the case even if we

[15]Notice however that the people looking for an apartment in a city are not generally a representative sample of the population

[16]Now it can range between 0 and $\log(K = 5) = 1.61$

consider a small number of agents from an additional category with a much higher income looking for apartments only in the center, such as some rich foreigners willing to buy an apartment in the center of Paris. More precisely, we perform computer simulations creating at the beginning of each time step $\gamma/100$ "foreigners", identical to the other agents apart from their income $Y = 100$ and for the way they choose a location to apply for an apartment: they pick randomly one of the locations in $\mathcal{E} = \{(x, y) \text{ s.t. } x^2 + y^2 \leq 4\}$. As one can see from the black line in Fig. 4.15, the foreigners increase substantially the price in $\mathcal{E}$, but surprisingly the prices raise uniformly all over the city. This is due to the fact that the 4 and 5-agents are less likely to choose a location in the center because of the higher price, so they have to search elsewhere, and this entails a cascade which influences the price even in the locations farthest from the center. This result is in accord with empirical evidence (see Fig. 3.1 and the discussion in Section 3.2) and has interesting policy implications.



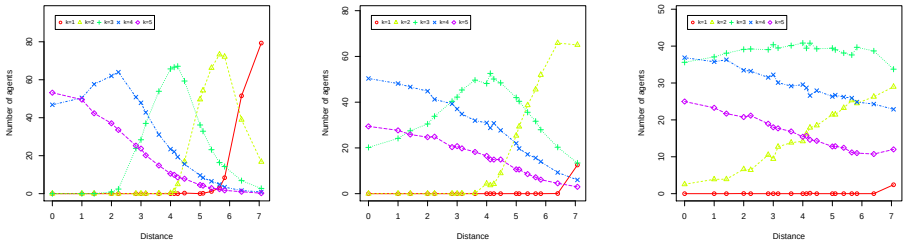
(a) (b) (c)

Figure 4.13: Number of agents as a function of the distance from the center. The color code is such that the agents of the lowest income categories are given warm colors, the ones from the highest are given cool colors. Here $K = 5$, $\gamma = 2000$ and the values $\{\gamma_k\}$ depend on the figure. All the other parameters are as in Fig. 4.5b. The higher the fraction of 4 and 5-agents, the more the other categories are segregated away from the center. (*Left*) $\gamma_k = \gamma/5, \forall k$ (case (a)) (*Center*) $\{\gamma_k\} = \{5, 4, 3, 2, 1\}\frac{\gamma}{15}$ (case (b)) (*Right*) $\{\gamma_k\} = \{0.55, 0.23, 0.12, 0.06, 0.04\}\gamma$ (case (c))

4.3.2 Other parameters

The purpose of this section is to validate through numerical simulations the assertions in Section 4.2.5, which were based on the mathematical analysis in Section 4.2.

In Figure 4.16 we consider the effect of λ and μ. Comparing Figures 4.16a and 4.13a it can be noticed how the effect of setting $\lambda = 0.99$ is almost to drive away of the city the 3-agents, with 1-agents totally driven away and



(a) (b) (c)

Figure 4.14: Number of agents as a function of the distance from the center. Here the values $\{\gamma_k\}$ are as in Fig. 4.13b, and the other parameters apart from β and ϵ are as in Fig. 4.13. As expected, lower values of β and values $\epsilon > 0$ flatten the distribution of the agents. (*Left*) $\beta = 1$, $\epsilon = 0$ (*Center*) $\beta = 0.5$, $\epsilon = 0$. Notice that the choice of the parameters is the same as in Fig. 4.13b (*Right*) $\beta = 0.5$, $\epsilon = 0.022$

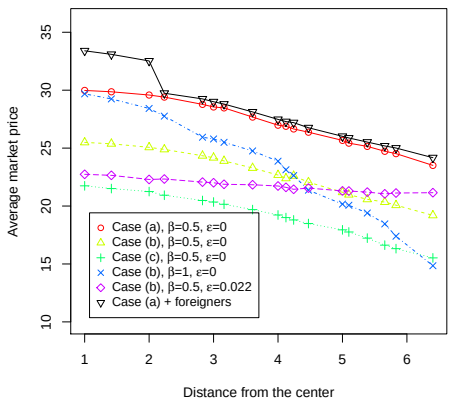


Figure 4.15: Average transaction prices as a function of the distance from the center. If one keeps the other parameters constant, the relative frequencies only change the offset of the price distribution; if one keeps the relative frequencies constant, $\beta = 1$ leads to a steeper price distribution and $\epsilon = 0.022$ flattens the price distribution. The inclusion of a small number of "foreigners" who search for an apartment only in the center rises the price level everywhere.

2-agents almost disappeared. This is consistent with the interpretation in Section 4.2.5: if the sellers do not decrease the price, it is impossible for the lowest income categories to afford an apartment anywhere in the city. A very similar pattern is noticeable in Figure 4.16b, where the fact that now $\mu = -50$ makes the process of price decrease much more difficult. The reverse case is in Figure 4.16c: now $\mu = -10$ and it is much easier for the prices to decrease, so a significant fraction of 3-agents does not live far from the center, and a considerable number of 2-agents is able to settle in the city.

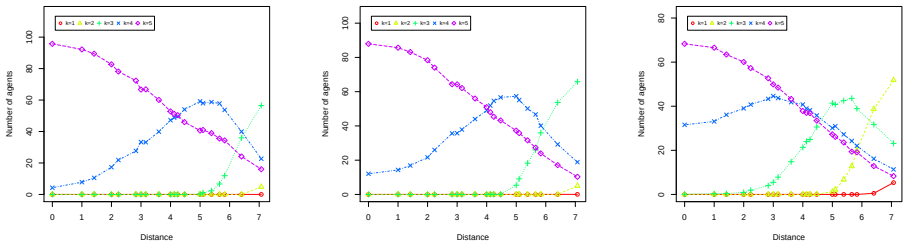
In Figure 4.17 we consider the effect of Y_1 and Δ. Figure 4.17b is almost identical to Figure 4.13a, as it should be the case, since the ratio Δ/Y_1 is the same. On the other hand, while keeping $\Delta = 5$ constant, we try to change Y_1 in order to see what happens when the ratio varies. In Figure 4.17a it is $Y_1 = 5$. Recall from Eq. (4.41) that a higher ratio implies higher segregation, and actually the 2-agents are almost segregated away from the city. The reverse case is presented in Figure 4.17c: now the ratio Δ/Y_1 is much smaller, the relative spacing between income categories is smaller, there is less segregation. Indeed, at distance 7 there is the highest number of 2-agents with respect to the other cases in Figure 4.17.

In Figure 4.18 we consider the effect of γ, α and n. We want to test if the assertion that n can be written as a rescaled version of γ holds. It is easy to see that this is the case, since Figure 4.18c, where the ratio γ/n is the same as in the other simulations, but $n = 50$, is almost identical to Figure 4.13b. It is more interesting to consider the effect of α. In both Figures 4.18a and 4.18b the ratio γ/α is the same as in the other simulations, but in the first case $\alpha = 0.2$, in the latter $\alpha = 0.05$. Recalling the discussion in Section 4.2.5, it can be noticed that a higher α means that the apartments are put on sale more frequently, so the prices have less time to decrease towards the modified market price, which is generally lower than the price paid by the seller. Indeed, there is more segregation in Figure 4.18a, with 1-agents almost disappeared from the city. Also the fraction of 3-agents in the center is lower with respect to both Figure 4.18b and 4.13b.

In Figure 4.19 we consider the effect of R, which determines the gradient of the intrinsic attractiveness field. Interestingly, Figure 4.19a is almost identical to Figure 4.14a, and Figure 4.19b is very similar to Figure 4.14c. Indeed, increasing or reducing R directly determines the steepness of the attractiveness field; the same effect is played by β (recall Eq. (4.29)) and ϵ/ω.

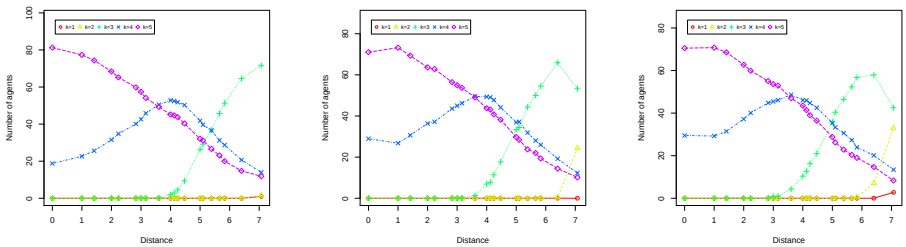
Finally, in Figure 4.20 we check if the ϵ and A^0_{max} parameters combine (whether only their ratio matters). This was confirmed in Figure 4.7a, but it was not clear from Eqs. (4.34) and (4.35). As it can be seen comparing Figures 4.20a and 4.20b, the assertion holds true, as the plots are almost identical. With respect to Figure 4.14c, the 3-agents can settle only farther from the center; this result was expected, since in this case the relative

strength of the social component with respect to the intrinsic attractiveness is lower.



(a) (b) (c)

Figure 4.16: Number of agents as a function of the distance from the center. Here we consider the same parameters as in Fig. 4.13a, apart from λ and μ. (*Left*) $\lambda = 0.99$, $\mu = -30$ (*Center*) $\lambda = 0.90$, $\mu = -50$. (*Right*) $\lambda = 0.90$, $\mu = -10$



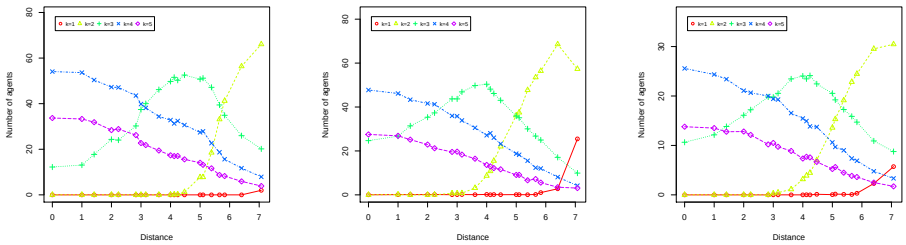
(a) (b) (c)

Figure 4.17: Number of agents as a function of the distance from the center. Here we consider the same parameters as in Fig. 4.13a, apart from Y_1 and Δ. (*Left*) $Y_1 = 5$, $\Delta = 5$ (*Center*) $Y_1 = 30$, $\Delta = 10$. (*Right*) $Y_1 = 30$, $\Delta = 5$

4.4 Discussion

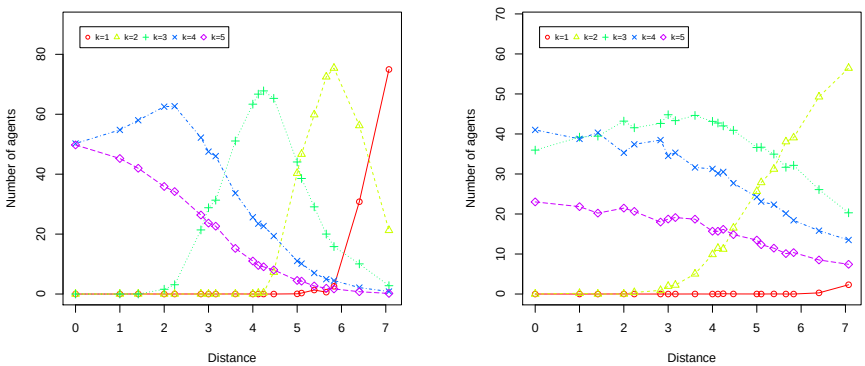
In this section we summarize the assumptions and the results presented in this chapter and we compare them with those in Gauvin *et al.* (2013). The comparison with other works in the literature has been done in Section 2.4.

Similarly to Gauvin *et al.* (2013), we consider a temporal and a spatial aspect, with an infinite sequence of time steps t and a grid of size L. On each site of the grid, N identical apartments are available to house the agents, who can be in the states of housed, seller or buyer. A small difference is that in the present thesis buyers who are not successful in finding an apartment are



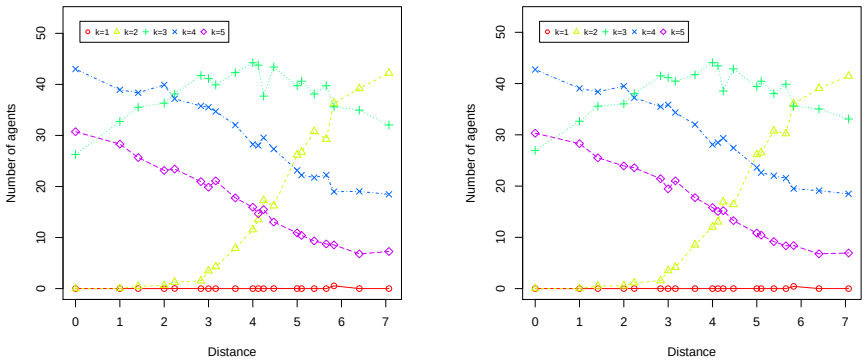
(a) (b) (c)

Figure 4.18: Number of agents as a function of the distance from the center. Here we consider the same parameters as in Fig. 4.13b, apart from γ, α and n. (*Left*) $\gamma = 4000$, $\alpha = 0.2$, $n = 100$ (*Center*) $\gamma = 1000$, $\alpha = 0.05$, $n = 100$ (*Right*) $\gamma = 2000$, $\alpha = 0.1$, $n = 50$



(a) (b)

Figure 4.19: Number of agents as a function of the distance from the center. Here we consider the same parameters as in Fig. 4.13b, apart from R. (*Left*) $R = 2$ (*Right*) $R = 4$



(a) (b)

Figure 4.20: Number of agents as a function of the distance from the center. Here we consider the same parameters as in Fig. 4.13b, apart from ϵ (with $\omega = 1/15$ fixed) and A^0_{max}. (*Left*) $\epsilon = 0.022$, $A^0_{max} = 5$ (*Right*) $\epsilon = 0.0045$, $A^0_{max} = 1$

removed from the simulation, allowing them to come back at the beginning of the subsequent time step with another cohort of agents. If it were not the case, thousands of agents from the lowest income category would hoard and substantially increase the computation effort without changing the price and segregation patterns (they would not be able to afford a dwelling where the price is higher than their income even if their number is soaring). Notice that in Gauvin *et al.* (2013) the assumption of "non-saturated equilibrium" was made to neglect the unsuccessful buyers, who were posing problems also from the point of view of the analytical treatment. Another difference with the previous work is that here we consider the income of the agents and not their willingness to pay. Indeed, here agents face a tradeoff between two goods (see below), so they spend their income on both; in Gauvin *et al.* (2013) the authors only considered housing as a good, so the willingness to pay (WTP) was simply the maximum price the k-buyers were willing to pay. In the present thesis the income distribution of the incoming agents is discretized in k-categories as in Gauvin *et al.* (2013), but here we consider in practice the effect of the income distribution on the segregation patterns, whereas in the previous work the authors considered the same income distribution throughout the paper, that is the buyers were uniformly distributed among the K WTP-categories (they generalize formally the result in Gauvin *et al.* (2013), Appendix A.4). The decision made by the buyers about the location where to apply for an apartment depends as in Gauvin *et al.* (2013) on the attractiveness, both intrinsic and social, but here the agents also weigh the general consumption they expect to afford at location X, so indirectly the

price they have to pay to purchase the apartment. The probability to choose location X is proportional to the utility they expect to get at location X, and not just on the attractiveness. This makes another difference with Gauvin *et al.* (2013): there the poorest categories were applying for an apartment in the most attractive locations along with the richest ones, not being able to settle there because of the cost; here each category has an optimal (most likely) location, depending on the price, and the sites where the price is higher than the income of a k-category are neglected by the k-agents. The major difference with Gauvin *et al.* (2013) is the choice for the reservation offer prices. In the previous work the rule was such that they depended on the WTP-category of the seller and on the average attractiveness, as the attractiveness was actually a proxy for the demand intensity; here the mechanism for the reservation offer prices is more complicated, as it tracks individual agents from the point of view of the price they paid; moreover, the offer prices at location X directly depend on the number of buyers and sellers and, most of all, take into account the market price at location X, which is more realistic from an economic point of view. Therefore, in the present thesis the prices are non-trivially related to the attractiveness, and this will deeply influence the results. The matching rule between the buyers and the sellers at location X and time t is also different with respect to Gauvin *et al.* (2013): there the authors considered two matching rules, that is buyers who can afford an apartment are selected randomly by the sellers, who themselves are chosen in a random order, or the sellers are treated in order of increasing offer price. Here we consider an order book at each location, where the traders directly bid or ask their reservation prices, or where they start from lower bids and higher asks and slowly converge towards their reservation prices following a bargaining process. Finally, a minor difference is about the initialization: in Gauvin *et al.* (2013) the starting point of the simulations was an empty city, so the authors had to disambiguate between empty and full apartments. This lead to some paradoxes, as the matching rule was such that the buyers would buy an apartment from the sellers, rather than purchasing an empty one which would cost less. In the present work we start from a full city, where all apartments are initially occupied; sellers cannot leave from the city until they succeed in selling their apartment.

Summing up, this work started from some weaknesses of Gauvin *et al.* (2013), which have been addressed as following (in order of importance):

- In Gauvin *et al.* (2013) the reservation offer prices were "ad-hoc". Now they are endogenously determined through the number of buyers and sellers, the market price and the price paid by the individual agents.

- In Gauvin *et al.* (2013) agents had no utility function. Now they have a Cobb-Douglas weighing consumption and total attractiveness.

- In Gauvin *et al.* (2013) the choices of the agents were only driven by the attractiveness. This means that the poorest agents were looking for an apartment in the center, where they certainly could not afford it. Here each category takes into account its economic possibilities, if $P(X,t) > Y_k$ the probability that k-agents go to X is null.

- In Gauvin *et al.* (2013) the matching rule between buyers and sellers was somehow arbitrary, whereas here the more elegant solution of the order book is used. Little changes from a practical point of view, as the matching is not the key element of the model (the reservation prices are).

- In Gauvin *et al.* (2013) buyers who were not successful in buying an apartment kept looking for another apartment until they eventually succeeded. In order to keep the problem tractable from an analytical point of view, most of the result were presented under the assumption of "non-saturated equilibrium". Here the unsuccessful buyers are cancelled from the simulation, under the assumption that they may come back on the subsequent time step (the identities of the agents are not influential if they are not sellers).

- In Gauvin *et al.* (2013) there was a problem with empty apartments. This model is initialized such that all locations are fully occupied, so there is no need to disambiguate between empty apartments and full ones.

Some results of the present thesis are similar to those in Gauvin *et al.* (2013), other ones are completely different. In what follows, we will present the results by explaining which assumptions of the present model imply them, and so why they are similar or different from Gauvin *et al.* (2013). Notice that other results come from the additional features of this model and so cannot be compared.

As in Gauvin *et al.* (2013) the locations with the highest intrinsic attractiveness are those with the highest prices, and they are mostly inhabited by the richest categories. Actually, even though the decisions of the buyers are more complicated in this model, the probability to choose a specific location still depends on the intrinsic attractiveness at that location. In Gauvin *et al.* (2013) the probability was just proportional to the attractiveness, here it depends non linearly on it. Another similarity is the fact that at most locations some social mixing is preserved. The reason is that the buyers from the richest categories are not enough to match the housing supply (for an appropriate choice of γ and α), so the prices at location X decrease and categories with a lower and lower income come to the same location until a dynamically stable price is reached (it does not need to be a competitive equilibrium, with demand equating supply, see below).

An important result of the present thesis is that the income distribution and the number of buyers determine the offset of the price distribution, whereas the preference parameters determine the steepness. In particular, the social component flattens the price distribution, and thus reduces the level of segregation. This result is the opposite as in Gauvin *et al.* (2013), but this is simply due to the assumptions of the models. In both Gauvin *et al.* (2013) and the present work the social component is additive, so it levels out the attractiveness field; however, in the previous work the higher attractiveness directly implied an increase in the prices, and thus a higher segregation; in the present work it just implies a demand distributed more uniformly, so lower prices in the center and higher prices in the periphery. It may be questioned whether it is the right choice to consider the social component of the attractiveness additively with respect to the intrinsic component, but the result in this thesis perfectly makes sense in the context of *gentrification*: buyers from the middle class and the higher income categories start liking a possibly peripheral neighbourhood, their preferences are more uniformly distributed over the city and this implies a price increase in places which used to be working-class areas. Recalling Figures 3.3a and 3.3c, this may be what is happening in Paris. It is hard to tell whether there is more overall segregation in this case or in a case where the richest categories only go to the center. In the latter case the poorest category is able to settle in the city, even though in the locations farthest from the center; in the first case it disappears since the most peripheral areas undergo a process of gentrification, but there is more mixing between the other categories all over the city. The metrics of the information entropy suggests that there is less segregation in the gentrificated case, because the higher social mixing between the other categories counterbalances the lacking contribution from the poorest category.

Coming back to the prices, it is remarkable that the income distribution of the buyers does not affect the steepness of the price distribution. If the share of the richest categories within the buyers is higher, that is there is more income inequality, the prices are higher all over the city. This acts as a mechanism that gets the middle-class (for a quantitative reference, one may think of rich agents as categories 4 and 5 in Section 4.3, middle-class as 3-agents, poor as 1 and 2-agents) away from the center, and possibly the poor away from the city. This result is confirmed when a small fraction of extremely rich agents looking for an apartment only in the center is introduced in the simulations: the prices rise uniformly everywhere. This finding has interesting policy implications, since it makes it clear that it cannot happen that some parts of the city are excluded from a global increase in the prices. This result was already clear in the empirical analysis, see Figure 3.1. Notice that complete segregation happens in the most attractive location when the share of agents from the highest category overcomes a threshold value, and when this happens there is a continuous phase transi-

tion in the parameter quantifying the global segregation (which becomes an order parameter), that is the derivative of that parameter with respect to the ratio of buyers from the richest category to the total number of buyers is discontinuous at the critical value.

Since the sellers take into account both the price they paid to get the apartment and the market price, a bubble behaviour is possible in this model. In particular, when the sellers do not accept to decrease their price or there is a constant flux of buyers coming to a location irrespective of its price, the location undergoes a period where prices soar until they reach the highest possible level, that is the income of the richest category Y_K. This is the feature that allows to identify in the steady state which locations experienced a bubble period. On the other hand, about μ apartments have to be on sale for prices to start to decrease, and the decrease process is in any case slower. This captures one stylized fact of the housing markets, that is prices are "sticky downwards", they increase steadily but decrease slowly.

Finally, it is worth considering when and under which conditions markets clear, i.e. when the number of buyers matches the number of sellers. Interestingly, the model shows some frictions which prevent the market in some locations to clear completely. The fact that on average only a fraction α of the apartments is put on sale at any time step t defines a threshold on the number of sellers at any location, so if there are constantly more buyers at that location the market never clears. An interesting extension of the model would be to endogenize the probability α to put the house on sale. On the buyers side, whenever the weight in the utility function β is given only to the attractiveness there is a substantial imbalance, with a huge fraction of the buyers applying in the center and a few buyers looking for an apartment in the outskirts. This implies a mismatch between buyers and sellers, and a competitive equilibrium is never reached. However, when buyers weigh equally consumption and housing ($\beta = 0.5$) the locations with high prices become less attractive, so supply and demand in most locations become balanced. This is a result which follows from the introduction of the utility function: if the buyers face a tradeoff, markets clear more easily.

The main results of this thesis are summarized below:

- The prices are higher and the richest categories settle where the intrinsic attractiveness is higher
- Some social mixing is preserved in most locations
- For fixed γ (number of incoming agents), β and ϵ/ω (preference parameters), the income distribution $\{\gamma_k\}$ determines the offset of the price distribution. This is not trivial, because even if you force some agents with a higher income (the "foreigners") to buy a house in the center (at locations at most at distance 2 for instance), all the prices out of this area raise uniformly. This is consistent with the data (when

prices rise in Paris, they rise uniformly in all arrondissment) and has interesting policy implications.

- For fixed γ and $\{\gamma_k\}$, the preference parameters determine the steepness of the price distribution. If $\beta = 1$ and $\epsilon = 0$, the agents only want to stay in the center and the price distribution is very steep. The result is that the agents are really segregated (the "rich" live in the center, the "middle class" between the center and the periphery and the "poor" in the periphery), but all categories afford living in the city. If, on the other hand, $\beta = 0.5$ and $\epsilon = 0.022$, we have gentrification: the rich and the middle class go also in the periphery, and the poors are driven away from the city. The result is that the middle class is mixed with the rich, but the poorest categories disappear. If one measures segregation through the information entropy, there is slighter more segregation in the gentrificated case.

- Bubble behaviour is possible. Here locations which undergo a bubble are those where the price in the steady state is set to the level of the highest income category. A bubble may be due to agents who do not want to sell at a price lower than they paid ($\lambda = 1$) or to a number of buyers constantly above that of sellers. Prices are "sticky downwards", i.e. they increase fast, but decrease slowly.

- Markets are never assured to clear because there are some frictions on the offer side. When $\beta = 1$ there is often excess supply or excess demand (some locations have more incoming agents than apartments on sale, for other locations there are almost no buyers). On the other hand, when $\beta = 0.5$, the agents also like locations which are not as attractive as in the center, but which have a lower price. So in most locations markets clear.

Chapter 5

Complementary results

The results shown in Chapter 4 are the final stage of a research process which took several directions. It started by reproducing the results in Gauvin *et al.* (2013) and exploring some variations on the original model. It developed towards the Alonso model, which we tried to simulate by following the approach in Lemoy *et al.* (2010). Since the results were not satisfactory, we came back to the original model introducing some aspects of the other works we had considered so far. We started considering a "closed city", i.e. a city where the number of agents was fixed since the beginning and agents were relocating within the city. The model was showing several problems, so we returned to the idea of an "open city", with an inflow of buyers and an outflow of sellers. After some trial-and-error we defined the model as in Section 4.1.

A more detailed description on the making of the model is provided in Appendix B. In this chapter we present the most interesting results about the explorations which led to the original results of this thesis. In Section 5.1 we mention a few results which have been obtained from the original model, that is Gauvin *et al.* (2013); in Section 5.2 we perform simulations of the Alonso model by following the approach of Lemoy *et al.* (2010): we try to reproduce their results and we simulate a simple adaptation of the Alonso model to our case. Finally, in Sections 5.3 and 5.4 we describe the results on the original model in its starting form, both in the case of the closed city and of the open city.

5.1 Explorations on the original model

The description of the model in Gauvin *et al.* (2013) was detailed enough to allow for a full implementation. Here we first show some results in accord with those in the paper, then we present the results of some explorations on the parameters and on the assumptions.

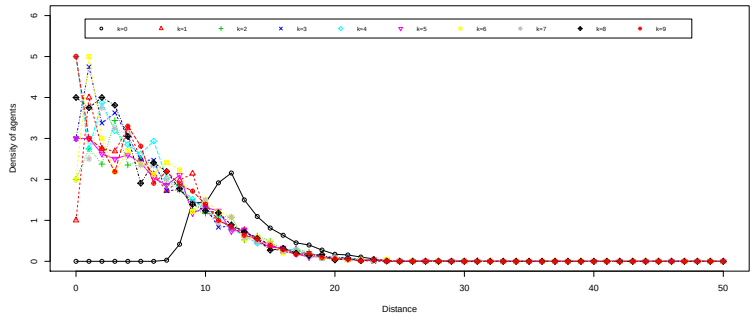
We tried changing γ, the rate of agents coming to the town (Figure

5.1). When $\gamma = 1000$ only the poorest category is segregated from the center of the town, whereas when $\gamma = 10000$ other categories cannot afford living in the center. Actually the higher flux of buyers rises the level of the attractiveness (in a linear way), and thus increases the prices everywhere (in a sub-linear way). It should be noticed that the transitions after any critical distance are not as sharp as those in the paper. The reason is probably in a different implementation of the matching rule, which in the paper was to sell to people who were at least as rich as the sellers. We also studied the case of the saturated equilibrium, which blurs the segregation pattern (Fig. 5.2a). In this case, the result is rather different than that in Gauvin *et al.* (2013), but we studied the saturated case by decreasing the number of goods to $N = 50$, while in Gauvin *et al.* (2013) the saturated case was simulated increasing the rate of incoming agents $\gamma = 20000$.

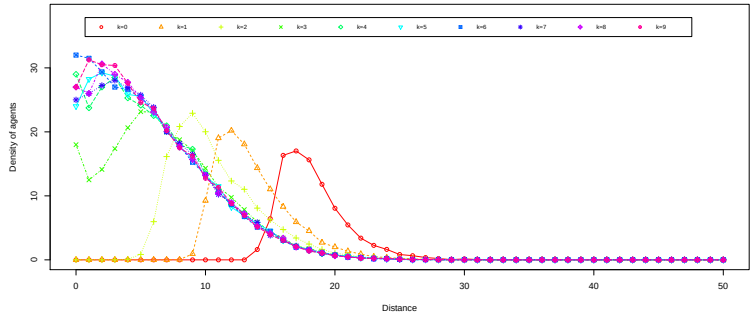
We explored the other matching rule of Gauvin *et al.* (2013), i.e. sellers are treated in order of increasing offer price. As it can be seen comparing Fig. 5.2b with Fig. 5.1b the results are not really different. Moreover, since it is reasonable that agents go to locations in which they expect they can afford a house (it is unlikely that the poorest agents would look for a house in the very center) we tested an adaptive rule to model the way agents learn to find the locations in which it is most likely they will be able to buy a house. We multiply $\pi_k(X,t)$ of Gauvin *et al.* (2013) by a multiplicative factor:

$$\pi_k^{'}(X,t) = \pi_k(X,t) \cdot \left(0.1 + \frac{0.9}{1 + e^{-0.8(d(X)-d_k)}}\right) \tag{5.1}$$

where $d(X)$ is the distance from the center of location X, and d_k is the distance learned by k-agents, computed by averaging over distances at which agents did not succeed in buying a home. The economic interpretation may be that the arriving agent has no information about the prices in the city, and asks to a random agent with the same income as him where to look for an apartment. As shown in Figure 5.2c the results were not really different from the baseline case, and they were extremely similar to the saturated case (Fig. 5.2a). Furthermore, we tried to turn off the social influence to check whether there would be segregation. As it can be seen in Figure 5.3 the result is ambiguous: the poorest category is segregated away from the center, but there is no WTP-threshold. Finally, we implemented an order-book rule for the matching between buyers and sellers. We almost did not find any difference from the point of view of segregation, but we found that the prices were on average higher.

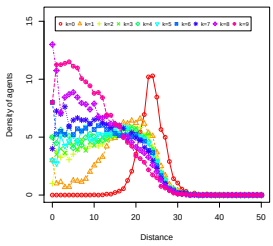


(a) $\gamma = 1000$

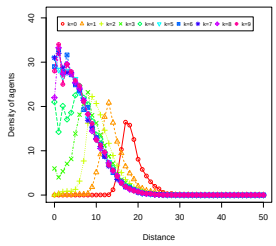


(b) $\gamma = 10000$

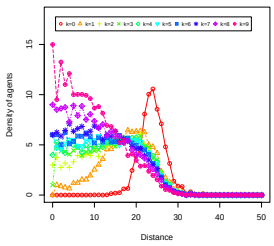
Figure 5.1: Number of agents as a function of the distance from the center in a 50x50 grid. All the plots in this section come from our implementation of the Agent-Based Model proposed by Gauvin *et al.* (2013)



(a) $N = 50$



(b) Ascending prices



(c) Adaptive agents

Figure 5.2: Experiments on the baseline model. (*Left*) Saturated regime (*Center*) Alternative matching rule, that is sellers are treated in order of increasing offer price (*Right*) Adaptive behaviour of the buyers

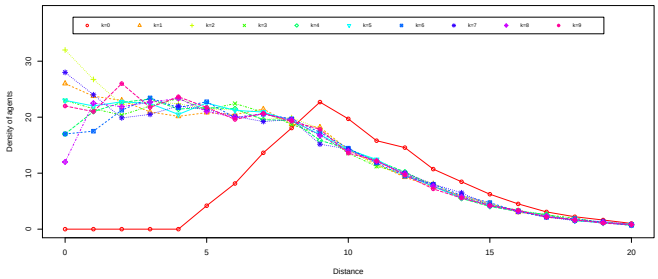


Figure 5.3: There is no social influence, that is $\epsilon = 0$.

5.2 Simulations on the Alonso model

The Alonso model (Alonso *et al.*, 1964), which has been presented in Section 2.2, is not discrete, does not allow for the heterogeneity and the interactions of the agents, does not have a temporal dimension, i.e. it does not explain how the equilibrium is reached. In order to make an agent-based model out of the Alonso model a lot of rules have to be specified, often without a specific justification. In this section we follow the approach of Lemoy *et al.* (2010) which, as specified in Section 2.3, has many limitations. The problems with this approach come about in Section 5.2.1, where we try to reproduce their results; in Section 5.2.2 we present the attempt to apply their model to a simple adaptation of the Alonso model, that is a model where the transport cost is replaced by the attractiveness.

5.2.1 Reproducibility of the results of Lemoy et al. (2010)

In addition to the specifications of Section 2.2, one needs to add some conditions to close the Alonso model. In particular, we need to impose the boundary condition on the edge of the city, that is the bid-rent is the same as the agricultural rent at the border of the city (landlords are indifferent between renting their land and using it for agriculture), and to specify that the total number of agents is such that all locations are fully occupied. An alternative approach, which we do not follow here, would be to allow for an open city; rather than specifying the number of agents, one specifies the utility outside the city U_{out}, and assumes that at equilibrium the utility in the city is the same as the utility outside $U = U_{out}$. Thus, the conditions read (recall Eq. (2.16)):

$$\psi(r_b, u) = \beta\alpha^{\alpha/\beta} \left(Y - T(r_b)\right)^{1/\beta} e^{-u/\beta} = R_a \tag{5.2}$$

$$\int_0^{r_b} 2\pi x \rho(x, u) dx = \frac{N}{s_{tot}} \tag{5.3}$$

In the above equations r_b is the radius of the city, N is the number of agents, all the other parameters are defined as in Section 2.2. In (5.3) it is $\rho(r,u) = \frac{1}{S(r,u)} = \alpha^{\alpha/\beta}\,(Y - T(r))^{\alpha/\beta}\,e^{-u/\beta}$. Solving (5.2) for $e^{-u/\beta}$ and replacing in (5.3) one gets:

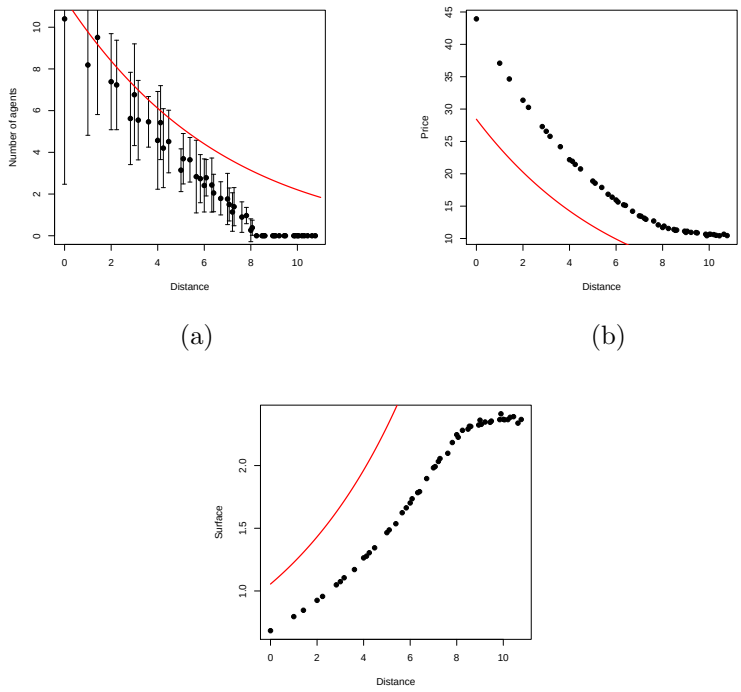
$$2\pi s_{tot}\frac{R_a}{\beta}\,(Y - T(r_b))^{-1/\beta}\int_0^{r_b} x\,(Y - T(x))^{\alpha/\beta}\,dx = N \tag{5.4}$$

If one specifies N and R_a in (5.4) it is possible to solve the above equation numerically for r_b. Then one gets the utility from (5.2):

$$u = -\beta\log(R_a/\beta) + \alpha\log\alpha + \log(Y - tr_b) \tag{5.5}$$

Lemoy *et al.* (2010) choose the parameters as follows: $\alpha = 0.9$, $\beta = 0.1$ (a disproportionate weight is given in the utility function to consumption rather than to housing), $Y = 300$, $t = 5$, $N = 700$, $R_a = 10$, $s_{tot} = 12$. From the point of view of the reproduction of the analytical results, the comparison with Lemoy *et al.* (2010) is perfect (compare Figure 2.5 with Figs. 5.4 and 5.5). It is different when it comes to the comparison with the results from the simulations. The problem is to specify the rule for the price decrease. A first attempt has been to decrease the prices in the non fully occupied locations after each movement try, i.e. after an agent and a patch were chosen randomly. The results were poor, as the prices were decreasing too often and were soon stabilizing everywhere to the level of the agricultural rent. A second attempt has been to decrease them only after all agents tried to move on average, i.e. after N movement tries. This lead to the results in Figure 5.4. Notice that in Figure 5.4a there was a lot of variability among the simulations, so it was needed to average the results over several simulations. The mismatch with the analytical results is apparent in Figs. 5.4b and 5.4c, where the results from the simulations are more stable. In particular, the average rent is systematically higher than in the analytical case, whereas the average surface is systematically lower. Notice however that the slope of the curves are similar. The reason is probably that the prices seldom decrease, so the offset is higher. Since the optimal surface is inversely proportional to the rent, it is clear that the average surface is lower. The third attempt to reproduce the results in Lemoy *et al.* (2010) has been to decrease the prices after each successful move, i.e. every time that an agent and a patch where the agent would get a higher utility were picked. As it can be seen in Figure 5.5, the comparison is very good. Notice that Figs. 5.4a and 5.5a are not so dissimilar: actually the total number of agents is the same whatever the rule to decrease the prices is, so the normalization is fixed. We also tested what would happen in the case of Fig. 5.5 if we let prices relax to R_a and not to $0.9 \cdot R_a$, as the authors in Lemoy *et al.* (2010) assume (see Eq. (2.29)). This assumption did not make any sense in the context of the Alonso model (if prices drop below R_a the landlord is better off using the

land for agriculture). The result can be seen in Fig. 5.6. The matching with the analytical result breaks down. So, not only it is needed a very specific rule for how often to decrease the price in order to get the matching with the analytical results, but it is also needed a wrong rule to let prices decrease fast enough to match the analytical equilibrium. In general, the problem is that to reproduce some results which do not come from a dynamic setting, one needs a lot of unjustified assumptions to change the offset of the rent and surface curves and match the analytical results.



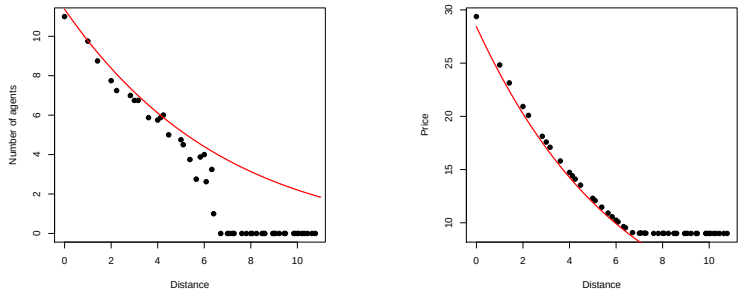
(a) (b)

(c)

Figure 5.4: Analytical results are red curves, results from the simulations are filled circles. On the x-axis is the distance from the center. In this picture we use the rule such that after on average all agents tried to move the prices decrease in the non fully occupied locations, following the rule in Eq. (2.29). (*Left*) Average number of agents. The error bars represent the standard deviation over 20 simulations (*Right*) Average rent (*Bottom*) Average surface

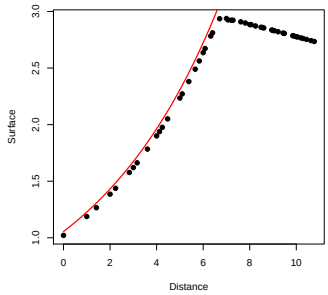
5.2.2 Adaptation to the attractiveness

In order to apply the Alonso model to the case studied in the present thesis, where the space is differentiated by the attractiveness, one can reformulate



(a)

(b)



(c)

Figure 5.5: Analytical results are red curves, results from the simulations are filled circles. On the x-axis is the distance from the center. In this picture we use the rule such that after each successful move the prices decrease in the non fully occupied locations, following the rule in Eq. (2.29). (*Left*) Average number of agents (*Right*) Average rent (*Bottom*) Average surface

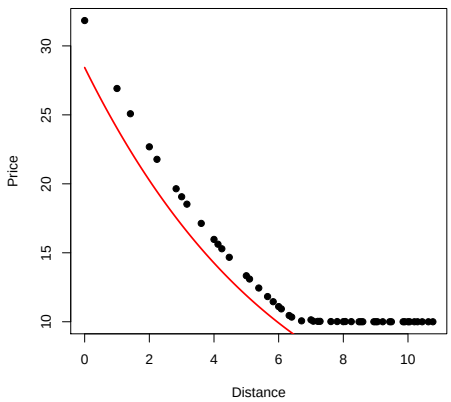


Figure 5.6: Average rent as a function of the distance to the center. In this picture we use the rule such that after each successful move the prices decrease in the non fully occupied locations, following the rule in Eq. (2.29) with $0.9R_a$ replaced by R_a.

Eq. (2.16) by setting the transport cost $t = 0$ and by considering for each location the attractiveness seen by a k-category $\mathcal{A}_k(r)$. The equation reads:

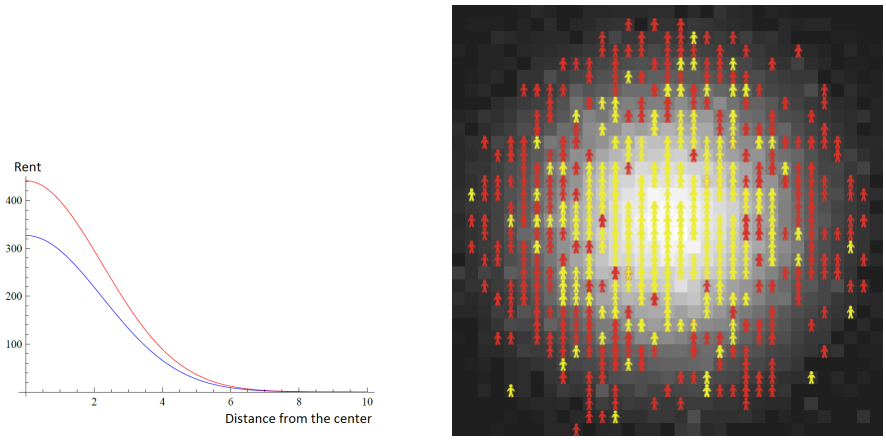
$$\psi(r, u) = \beta\alpha^{\alpha/\beta}e^{-u/\beta}Y^{1/\beta}\mathcal{A}_k(r) \tag{5.6}$$

Let us take $\mathcal{A}_k(r) = A^0(r) = A^0e^{-r^2/R^2}$, i.e. we just consider the intrinsic attractiveness. Let us now consider two categories, the poor with income Y_1 and the rich with income Y_2, $Y_1 < Y_2$. Since in Eq. (5.6) r enters only through $\mathcal{A}_k(r)$, the bid rent curves do not cross. It is not possible to find an analytical solution for the equilibrium utility with two groups. Anyway it can be maintained that the rich have a higher overall utility, $u_2 > u_1$. This means that the exponential in Eq. (5.6) is much lower for the rich, and overall the bid-rent curve of the poor is higher (Fig. 5.7a). This would mean that the poor would pay a higher price everywhere, thus no rich should be able to settle where the poor want to settle. The result of the simulations is the opposite, the rich stay in the center: see Fig. 5.7b. It is not clear how to interpret this result.

5.3 The closed city

The first formulation of the model presented in this thesis was a mixture of the models in Gauvin *et al.* (2013) and Lemoy *et al.* (2010). Here we present only the differences with the model in Section 4.1.

Every location has an intrinsic attractiveness $A^0(X)$, which we can take as a Gaussian decreasing from the center, and a social component $A_k(X) =$



(a) (b)

Figure 5.7: Results on the adaptation of the Alonso model to the attractiveness case (*Left*) Theoretical result. The red curve is for $Y = Y_1$; the blue curve is for $Y = Y_2$. The red curve for the poor is systematically higher (*Right*) Result from the simulations, which follow the model in Lemoy *et al.* (2010), adapted to the attractiveness. The yellow agents are the rich (with income $Y_2 = 20$), the red agents are the poor (with income $Y_1 = 15$). The rich locate closer to the center

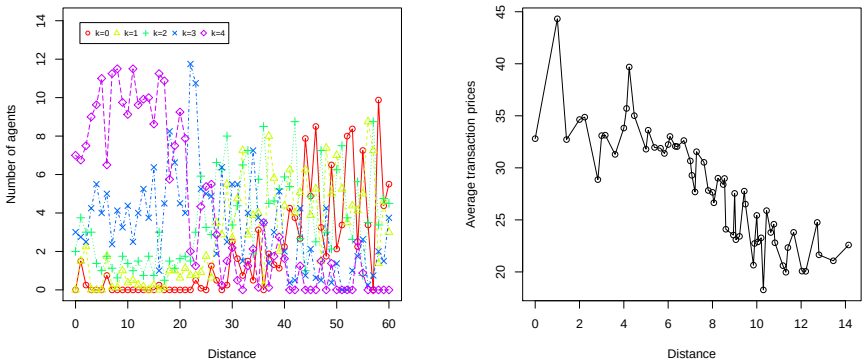
$\sum_{k'>k} v_{k'}$, where v_k is the density of k-agents. The total attractiveness of location X is: $\mathcal{A}_k(X) = A^0(X) + \eta A_k(X)$. Thus, we do not consider a dynamical equation for the evolution of the attractiveness. Every k-agent is endowed with a *wealth* Y_k. All the money which has not been spent on housing can be spent for the consumption good z. Agents, whose number is fixed from the beginning of the simulation (there is no inflow of buyers and outflow of sellers) move only if there is an increase in utility in doing so. If they see there are at least a fraction ϵ of locations at which the utility is higher, and if a Bernoulli random draw with probability γ is successful, they put their house on sale. Buyers at location X are people who were successful in selling their house (at the setup, a fraction ν of the agents are assumed to have just sold their house). Once a buyer settles at a new location, his wealth his reduced by the price he paid for the apartment.

There are several problems with this model:

- The number of agents looking for a house is fixed from the beginning of the simulation, it is always the initial fraction ν.
- If Y_i^k is wealth and not income, it is not true that the rest of wealth is spent on consumption. It is income which is spent on consumption. Considering both income and wealth would be too complicated, because one would not understand the causal relationships.

- Where do the agents that sold their house and are looking for a new house go? It cannot be the case that they wait until they have found a new house, because this would cause a cascade and nobody would move.

Despite all these shortcomings, the results which can be seen in Fig. 5.8 proved quite robust: there is segregation in the center, and the prices decrease towards the periphery. This model could be used to explain *gentrification*: after an initial random distribution, the poor go to the periphery with lowest mortgage rates, the rich move in to the center.



(a) (b)

Figure 5.8: Closed city model. (*Left*) Number of agents as a function of the distance from the center. Notice the huge variability. (*Right*) Average prices as a function of the distance from the center.

5.4 The open city

The model presented in this section is the most similar to that in Section 4.1, with two main differences: the rule to decrease the reservation offer prices, the assumption about the buyers who were not successful in finding an apartment.

The first version of the model was as following: reservation offer prices are determined through a discount factor λ: agents start from the purchase price, and then they decrease the reservation offer price every time step. The idea is that they really have to sell the house, so the reservation offer price could get close to zero. What is more, the number of buyers is taken to be approximately as the number of sellers, and unsuccessful buyers stay in the city until they eventually find an apartment.

As it can be seen in Figure 5.9 the system is not stable, as there are huge variations in the number of k-agents. What is more, the simulations

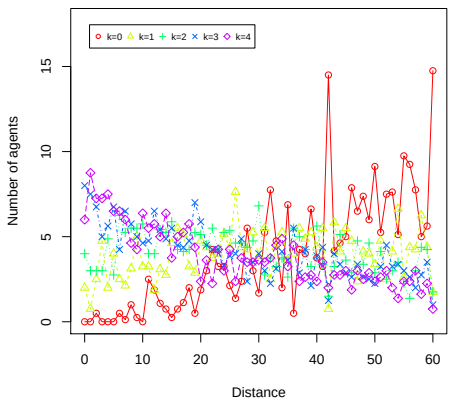


Figure 5.9: Number of k-agents as a function of the distance from the center. There is huge variability in the number of agents and the system is not stable

showed that there is huge price volatility, such as subsequent transactions happening at the lower and upper ends of the price interval. The reason is that the prices decrease too much and lack a reference level.

Therefore, we modified the rule for the reservation offer prices. Now

$$\overline{P_i^o} = \max\{P(X), P(X) + \lambda^{t-t_i}\left[P(X, t_i) - P(X)\right]\} \tag{5.7}$$

In the above equation $P(X)$ is the market price, and it is discounted by a factor λ if no transactions take place. This approach did not work. On the one hand, the fact that unsuccessful buyers are not replaced led to a situation where almost all buyers were from the lowest category (see Figure 5.10). On the other hand, when replacing unsuccessful buyers but sticking to (5.7) and to the rule that the price was discounted if there were no transactions led to a situation where only the agents from the richest category could afford an apartment, since it was enough that there was just one transaction for the prices never to decrease.

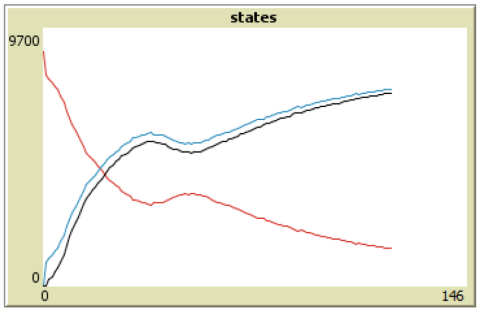


Figure 5.10: Number of agents in a certain state as a function of time: the black line stands for the buyers, the red line for the housed agents, the blue line for the sellers

Chapter 6

Conclusion

This thesis presents a research project about the residential distribution of people in cities. We are interested in understanding the factors that drive the residential choice of people with respect to their income and in explaining the real-estate price distribution and the segregation patterns. We propose a dynamic model for the urban housing market which accounts for a set of stylized facts we observed in our dataset, recording more than 400.000 real-estate transactions occurred in Paris in the years 1990-2007. The model considers a city whose locations are characterized by a certain level of attractiveness, which evolves according to the social composition of the neighbourhood. The agents, heterogeneous in their wealth, compete for the available apartments: the price distribution emerges out of their interactions. We specify no ad-hoc rules for the market price: we use an order book (double auction) to model in a realistic way the bargaining process between buyers and sellers. Using mean field arguments and continuum analysis, we are able to provide an analytical description of the steady state of a simplified version of the model, allowing for a better understanding of the space of parameters and of the stochastic effects of the agent-based model (ABM). Later, we perform computer simulations on the fully-fledged ABM and we obtain more general results. Our most interesting findings concern the price distribution: the income distribution of the buyers determines the offset, the preferences of the agents determine the steepness. We also find that the housing market displays easily bubble behaviour and that markets clear more easily when the buyers face a tradeoff between the most attractive locations and the level of consumption.

We started our investigation from a recent work on the housing market, that is Gauvin *et al.* (2013). The authors took from a burglary model (Short *et al.*, 2008) the concept of repeated victimization; they adapted it to the housing market by assuming that the locations with many newcomers become more attractive, drawing even more agents. Despite the great novelty of Gauvin *et al.* (2013), there were some problems concerning the

economic assumptions and we worked to fix them. We took inspiration from Alonso *et al.* (1964), a cornerstone of neoclassical urban economic theory, and we devoted part of the research project to an adaptation of the Alonso model (which considers the attractiveness and not the transport cost) and to an ABM directly related to it (Lemoy *et al.*, 2010). We considered other works on the housing market. For instance, Landvoigt *et al.* (2015) build a model to explain why the cheapest apartments were the ones showing the highest price increase during the housing market boom of San Diego. They use an assignment model, which does not really consider the interactions between buyers and sellers. Feitosa *et al.* (2008) and Filatova *et al.* (2009) make rather simple ABMs of the housing market, but they lack appropriate assumptions about the decisions of the sellers. On the other hand, Gilbert *et al.* (2009) makes an extremely complex ABM which endogenizes many variables, but we find it hard to interpret its results because the causality relations are not clear.

Some insights for the assumptions of our model come from some empirical facts we observed in our dataset. In particular, we find that the prices soared in Paris from year 1998, after a small drop. They increase in all the *arrondissements*, but in some neighbourhoods the rise is less apparent, so that the ranking of the most expensive arrondissements varies. This suggests that some mechanisms which change the attractiveness of some areas are into play. Moreover, the prices in 2003 are much more uniform than in 1990, suggesting that the preferences of the people became more spread. When one looks at the behaviour of the sellers, another interesting insight is apparent: the more time lapses between the purchase and the sale of the apartment, the less is the average transaction price. Other interesting observations which did not influence the behaviour of the model, but which are interesting in their own right, are the following: the older a buyer is, the higher is the price he pays, but the holder a seller is, the lower is the price at which he sells; the most expensive apartments are those in houses built recently, rather than in *Hausmanianns* buildings; the foreigners from countries as rich as France pay on average a higher price than French people, but those coming from poorer countries pay a lower price.

We use the above insights and some plausibility assumptions to make an ABM of the housing market. The artificial city of the model is a grid, where all sites host a fixed number of homogeneous apartments. The buyers come to the city from the outside and are differentiated by their income. Their residential choice depends on the attractiveness (which has both and a social and an intrinsic component) of the location and on the consumption of a composite good they expect to afford after they have paid for their accommodation, so indirectly on the price they expect to pay. The landlords decide to put on sale their apartment with an exogenous fixed probability. They decide their reservation offer price by taking into account the market price, the competition in the market and the price they paid when they

purchased the apartment. In particular, the more time lapses since the purchase of the apartment, the more their reservation offer price depends on the market situation. Buyers and sellers are matched through a continuous double auction (CDA), which is modelled through an order book. Successful buyers settle at the desired location, successful sellers leave the city.

The mathematical analysis is necessary because of the high dimensionality of the parameter space. We want to understand which control parameters matter the most, and if it possible to group them in fewer effective parameters. In spite of the great complexity of the agent-based model, its most important features can be analytically understood. We average out stochastic effects by taking expected values (mean field assumption). We also assume continuous space and time, so all "numbers" become densities and we can use calculus. In this work we model the intrinsic attractiveness as a Gaussian decreasing from the center. In the mathematical analysis we take the city as a circle with the same area as the square where the grid is inscribed. These assumptions are not enough to ensure an analytical treatment, so we have to study specific cases. First of all, we consider the simplest case, that is the agents weigh only the attractiveness in their utility function, there is no social component in the attractiveness, there is only one income category. We find that if the sellers do not decrease their reservation offer price from their purchase price, or if the flux of buyers is constant, the real-estate prices do not decrease and there is a bubble. Another result is that because of some frictions, the markets may not clear: in particular, the locations in the center have an excess demand, whereas the locations in the outskirts have an excess supply. The most interesting insight which we can obtain by relaxing the assumption that the agents only consider the attractiveness in their utility function is that the markets clear more easily. When we consider the social component of the attractiveness, we find that the real-estate price distribution is flattened, as the richest agents start appreciating also the locations farther from the center: this is a case of *gentrification*. Finally, if we consider two categories we discover which factors influence the segregation among them. One interesting result is that there is a second-order phase transition in the parameter quantifying the level of social mixing when the fraction of rich agents in the population reaches a critical value.

We also employ the power of computer simulations to shed some light on more general cases. Mostly, we consider 5 income categories and we investigate the distribution of the agents from each category. If the attractiveness field is flattened by the social component, this enhanced space uniformity acts as a mechanism which prevents the richest agents to segregate in some areas, but drives away the poorest agents from the city: it is a case of *gentrification*. Through the metrics of information entropy, we find that the overall segregation is lower in this case, because of the higher social mixing between the other categories but the poorest. Another important result is

that if the share of rich agent in the population is increased, the prices raise uniformly all over the city. Even if some extremely rich agents are forced to look for an apartment only in the center, the prices raise uniformly elsewhere too.

The findings in this thesis make it clear from a quantitative point of view that the unavoidable outcome of a constant flux of rich buyers (as it happens in world cities such as New York, Paris and London) is that the lower income categories are gradually displaced from the city. One possible solution is to enlarge the boundaries of the city and to improve the connections between the outskirts and the center. The other conceivable solution is to increase the number of housing units per location, e.g. by building skyscrapers rather than one-storey houses. As in Gauvin *et al.* (2013), subsidizing the poorest individuals would have counterproductive effects, since it would raise the prices. If the goal of the policy maker is to enhance the social mixing between the other categories but the poorest, the best way to do it is to give some incentive (direct or indirect) to the rich to settle even in the locations farther from the center.

The perspectives of this work are diverse. First, from an empirical point of view, it would be extremely interesting to analyse the data by combining techniques coming from econometrics and machine learning. For instance, it is intriguing to find the ranking of the most important factors affecting the real-estate prices. Standard econometrics techniques such as regression just tell which coefficients are significant. Yet, regression does not tell the relative importance of the explanatory variables. On the other hand, machine learning methods such as regression trees provide the ranking of the most discriminating variables. It would also be fascinating to find through unsupervised learning the amenities (e.g. monuments) in Paris: they should emerge from the data of the real-estate transactions. Moreover, it would be extremely interesting to check through the data if the prices exhibit diffusive behaviour, and possibly to develop a metrics testing diffusion. Finally, one could try to calibrate the model, with the caveats mentioned in the introduction. Second, from a theoretical point of view, a straightforward extension would be to allow for apartments of several sizes. This would allow a comparison between the model presented in this thesis and the Alonso model; notably, one could check whether the result that all the agents have the same utility (they are indifferent between staying in the center in a small apartment or staying in the periphery in a large one) holds. It would be interesting to study polycentric cities, heterogeneous preferences, frictions in the bargaining process. Finally, a natural development of this model would be to transform the equation governing the evolution of the attractiveness from a reaction to a reaction-diffusion equation, by adding a Laplacian linking the evolution of the attractiveness in the nearby locations. Probably one should consider local information, i.e. the agents which were unsuccessful in finding an apartment should keep looking for another one in the same

neighbourhood. Perhaps, one should simplify the model in the direction of Short *et al.* (2008). A reaction-diffusion model of the housing market may involve the emergence of hotspots of high prices which cannot be explained by the presence of amenities, as it is indeed the case in many cities in the world.

Appendix A

Outline on the order book

In most financial markets buyers and sellers are matched through a *continuous double auction* (CDA). Here we present a summary on the functioning of the CDA, which follows mostly from the introduction in Smith *et al.* (2003). For an overview of the models of CDAs, see the relevant section in Chakraborti *et al.* (2011): for a first attempt see Bak *et al.* (1997), for more advanced models see Cont and Bouchaud (2000) and Lux and Marchesi (1999).

The traders in a financial market can submit two types of *orders*: if they are willing to buy/sell some stocks as soon as possible they submit a *market order*, that is a quantity of stocks that they want to buy/sell at the *market price*. Here the market price is the best price currently available for the trader, i.e. the lowest price asked by a seller if the market order comes from a buyer, the highest price proposed by a buyer if the market order comes from a seller. These prices "asked" and "proposed" refer to another type of order, the *limit order*. When traders have no urgency in carrying out a transaction, the propose to buy (sell) a certain amount of stocks at a given price. The limit orders by the buyers are called *bids*, those by the sellers are called *asks*. Here it comes the key element of the CDA, that is the *(limit)order book*: limit orders have to be stored somewhere until a market order that "takes" them comes. Most financial markets keep records of bids and asks in two separate *logs*: there is a queue of limit orders that wait to be processed through the arrival of market orders. The first bid to be processed is that at the highest price, and we denote it as $b(t)$; the first ask to be processed is that at the lowest price, and we denote it by $a(t)$. $b(t)$ and $a(t)$ are called *inside quotes*. $s(t) = a(t) - b(t)$ is called *bid-ask spread*. It has to be $s(t) > 0$, because if it were not so the last market order, that is a bid higher than the lowest ask or an ask lower than the highest bid, would be processed instantaneously. The prices of bids and asks are not continuous variables, but they are multiples of quanta called *ticks*. The distance between the closest prices is called *tick size*. Notice that usually

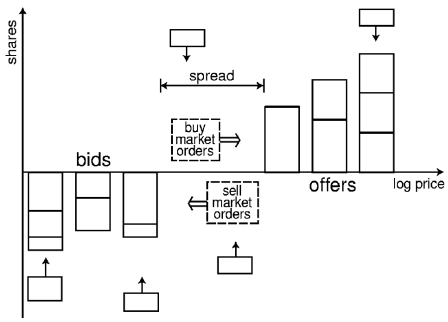


Figure A.0.1: Schematic view of the functioning of the order book. From Smith *et al.* (2003)

it is the logarithm of the price which is taken into account, but this is just a convention. Of course one can take the continuum limit for modelling. Let us denote by $n(p,t)$ the density (or number in the discrete case) of market orders at price p and time t: this function gives the *depth profile* of the order book. Both the *depth* $n(p,t)dp$ and the bid-ask spread give information about the *liquidity* of the market orders at price p. The *price impact* is the movement in the ask inside quote $a(t)$ due to a buy market order of ω stocks. Since the buyer gets the lowest asks, the new ask inside quote p' is given by:

$$\omega = \sum_{p=a(t)}^{p'} n(p,t)dp \tag{A-1}$$

The above equation can be solved for p': the price impact is defined formally as $p' - a(t)$. Let us consider for instance a buy market order of $\omega = 300$ stocks, and let us assume that the lowest asks standing in the ask log of the order book are the following market orders: sell 200 stocks at 60€and 300 stocks at 60.25€. The buyer will get all the stocks for 60€and 100 stocks for 60.25€. The price impact of his market order has been 0.25€.

Finally, it is worth mentioning for the purpose of the present thesis, the order book may be *one-shot* or *repeated*: in the first case the traders only have one chance to submit their limit orders, in the latter case there are several *trading periods* and the traders can modify their limit orders adaptively. In Section 4.2 the order book is one-shot, in Section 4.3 it is repeated: the detailed rule for the adaptive behaviour of the agents is given in Section 4.1.7. A schematic visualization of the functioning of the order book as described in this section is provided in Figure A.0.1.

Appendix B

The making of the model

The housing market model presented in this thesis began by the assumptions and the results of Gauvin *et al.* (2013) (hereafter "the original model"), but many contributions and a process of trial-and-error made it differ substantially from its starting point. The purpose of this section is to give an idea of the process which led to the model presented in Section 4.1.

We started from some weaknesses with the assumptions of Gauvin *et al.* (2013):

1. The reservation offer prices were "ad-hoc", that is they were determined according to a specific rule which was not grounded enough from the point of view of the economic interpretation.

2. The agents had no utility function. They were just acting according to the attractiveness of a specific location, they were facing no tradeoffs, and also the poorest agents were applying for apartments in the most expensive locations, where they certainly could not afford them.

3. The matching rule between buyers and sellers was somehow arbitrary, with buyers picked randomly from the sellers.

4. "Non-saturated equilibrium": it would not be realistic in a city as Paris.

5. Problem with the disambiguation between empty and full apartments.

6. In many locations markets do not clear. It should be possible for the city to evolve according to the demand.

7. The attractiveness could not decrease below $A^0(X)$, which is not realistic in case of the deterioration of the neighbourhood.

Only the last two items of the above list have not been addressed by the present thesis. We first reproduced the results of Gauvin *et al.* (2013)

by implementing the Agent-Based Model (ABM) described in the paper. We then used this implementation to perform some explorations: we used the order book as a matching rule and we simulated adaptive buyers who were deciding where to look for an apartment based on the success rate of like buyers. Moreover, in order to deal with the problem of the reservation offer prices, we tried to see what would happen if the sellers were trying to make a positive profit, leading to a frozen state where only the richest buyers could purchase an apartment anywhere. We also proposed a rule to let the attractiveness decrease below $A^0(X)$. These specific add-ons on the original model did not solve its weaknesses globally, so we agreed that we had to rethink its main assumptions. In particular, we focused on the economic aspects of the model, keeping the spatio-temporal structure, the heterogeneity of the agents and the attractiveness, both intrinsic and social, which characterizes the locations.

The starting point of most urban neoclassical economic theory is Alonso *et al.* (1964), so we referred to this model to look for new insights. Its main strength is the tradeoff between the position and the size of the apartments. We used the approach in Lemoy *et al.* (2010) to reproduce the results of the Alonso model from the point of view of the ABMs, but we encountered several difficulties and some problems with their work. We adapted the Alonso model to account for the attractiveness rather than for the transport cost, getting conditions on the number of agents which were implying a certain level of segregation. Unfortunately, given the problems with Lemoy *et al.* (2010) we did not have a framework to directly perform a computer simulation of the Alonso model. Therefore, we kept some aspects of the Alonso model, such as the utility function, but we decided to set up the ABM following mostly Gauvin *et al.* (2013).

A first attempt has been to consider a "closed city", as in Lemoy *et al.* (2010): the agents were initialized randomly and let free to interact until a dynamical equilibrium was reached, without an inflow or an outflow of other agents. The prices were determined endogenously through the order book. We had several problems from many point of view. On a technical side, the order book needed some specifications in order to prevent it producing negative prices or prices higher (lower) than the reservation prices of the buyers (sellers). The rules in Section 4.1.7 are the result of a long trial-and-error process. Moreover, the relation between income and wealth was ambiguous, the number of sellers was fixed, the moving rules in a full city were not clear.

Because of these problems, we decided to come back to the idea of an "open city", such that buyers come from an external reservoir and sellers decide exogenously to leave the city. We tried several rules to decrease the reservation offer prices in case the number of buyers was not enough. Some insights from the data were crucial in deciding the final rule, described in Section 4.1.6. We also tried several specifications for the evolution of the

prices and to deal with unsuccessful buyers. We found that if they were not removed from the simulation the buyers were almost only from the lowest income category.

After the model was determined in its final form, we succeeded in getting a mathematical description of some specific cases by making use of mean-field assumptions and thanks to the insights coming from the simulations. The interplay between the mathematical results and the simulations let us find results which are robust with respect to all parameter settings.

Appendix C

Data analysis code details

In order to ensure full replicability of the empirical results, all details about the "cleaning" and modifications on the data are reported here. Actually, there are differences in denomination between the 1990-2003 and the 2005-2007 databases, and some variables only exist in one database. The explanations for the variables are sometimes contradictory, and there are some apparent typos. The code which has been used for this part of the data analysis is presented in section C.1. Refer to table 3.1 for a short explanation about the meaning of the variables, and to Section 3.1 in general for some background. Here we list the problems with the variables:

- `REFE`: it is not clear if it refers to apartments or transactions. Moreover, out of the total of 431975 recorded transactions, 560 identifiers are repeated. It is not clear whether this supports the hypothesis that identifiers refer to the apartments, or whether it is just due to typos.
- `PRIX1`: there are 59 missing values, all of which come from the 2005-2007 database
- `SURFE1`: there are 32051 missing values, by taking into account both lacking records and surfaces set to zero
- `prix.m2`: it is not reported in the 1990-2003 database, but it is easily computed by taking `PRIX1/SURFE1`. There are 31704 missing values. It may seem odd that there are less missing values with respect to `SURFE1`, but the reason is simply that in the 2005-2007 database sometimes `prix.m2` is available even though `SURFE1` is not.
- `TYPAPP`: there are slightly different labels between the two datasets. In the 2005-2007 database there is an additional variable, "Studette", which we merge with "Studio". However, the significant lower number of "Chambre de Service" in the 2005-2007 database may suggest that in the 1990-2003 data "Studette" and "Chambre de Service" were

merged. Anyway, the fraction of "Studette" in the 2005-2007 database is not high. There are just 1310 missing values, only in the 2005-2007 database.

- `EPOQUE`: in the 2005-2007 database there is an additional label, "Z", which is not described. There are a few values with that label and we set them to NA (not available, used for missing values). There are 14548 missing values, all of which come from the 2005-2007 database.

- `ARRON`: there are only 739 missing values, all of which come from the 2005-2007 database. Some of the missing values have been recorded as "arrondissement 0", which does not exist.

- `ANNEE`: along with the year of the transaction, also the day and the months are provided. The problem is that no transactions are recorded but those happened between October and February, each year. Thus, only the information about the year seems reliable. In the 2005-2007 database the year is not disentangled from the other information, so a preliminary step is to separate the year from the day/month. The Python script which has been used to modify the csv file is provided in Section C.2. There are no missing values.

- `ANNEEMUTP`: there are 44102 missing values, but there is a problem with all the data recorded in the years 2002 and 2003. The dates are apparently wrong, ranging from year 100 to year 4000. Since it is unlikely that the last transactions for the same apartment occurred at the times of the Roman Empire, or will happen in the distant future, we wonder how the database was managed. By setting to NA these values the missing values become 91161, one fifth of the data.

- `AGEB` and `AGEVE` : there are 70414 missing values for `AGEB` and 113440 missing values for `AGEVE`. It should be noted however that there is a small fraction of buyers/sellers under 18 years old, the minimum age for French law to buy/sell an apartment. Anyway the number is risible in relative terms, they are probably typos.

- `CSPAC` and `CSPVE`: the definitions vary among the databases, so we decide to use only the database 1990-2003, which has more data. Two social categories are almost non-existent, so we set their values to NA. We end up with 36989 missing data for the buyers and 35519 for the sellers.

- `DEPDA1` and `DEPDV`: there are respectively 27074 and 27456 missing values. Because of the different definitions among the databases, we only take the data from 1990-2003.

- PXMUTPREC and PXMUTPRECM2: the price for the previous transaction is only available for the data recorded in the years 2005-2007. There are respectively 48503 and 51562 missing data.

C.1 R code

```
setwd("C:/Users/marco/Google Drive/Tesi/Data/Paris")

mydata <- read.table("./1990_2003_lgmt.txt", header=TRUE, sep=",",fill
    =TRUE,quote = "")
mydata1 <- read.table("./base 2005-2007_Paris-copia-1.txt", header=
    TRUE, sep=";",fill=TRUE,quote = "")

#Create new dataframe, with REFE numbers as first column

data = data.frame(c(mydata$REFE,mydata1$REFE))
colnames(data)="REFE"
nrow(mydata) + nrow(mydata1) == nrow(data) #Check
data$REFE[duplicated(data$REFE)]
#This gives that there are about 550 repeated reference number:
    identifiers
#are not unique

#Add price as a second column
data$PRIX = c(mydata$PRIX1,mydata1$PRIX1)
#Compute how many missing values
length(is.na(data$PRIX)[is.na(data$PRIX)==TRUE])
#Understand where they come from
length(is.na(mydata$PRIX1)[is.na(mydata$PRIX1)==TRUE])
length(is.na(mydata1$PRIX1)[is.na(mydata1$PRIX1)==TRUE])

#Add surface as a third column, setting to NA the zero values
data$SURF = c(mydata$SURFE1,mydata1$SURFE1)
data$SURF[data$SURF == 0] = NA
#Compute how many missing values
length(is.na(data$SURF)[is.na(data$SURF)==TRUE])
#Understand where they come from
length(is.na(mydata$SURFE1)[is.na(mydata$SURFE1)==TRUE])
length(is.na(mydata1$SURFE1)[is.na(mydata1$SURFE1)==TRUE])


#In 1990-2003 there is no price per squared meter: I need to compute
    it
#and to set as NA the values for which the surface is zero

mydata$PRIXM2 <- mydata$PRIX1 / mydata$SURFE1
#Replace inf values by NA
is.na(mydata) <- sapply(mydata, is.infinite)

#Add price per squared meter as 4th column
data$PRIXM2 = c(mydata$PRIXM2,mydata1$prix.m2)

#Compute how many missing values
length(is.na(data$PRIXM2)[is.na(data$PRIXM2)==TRUE])
#Understand where they come from
length(is.na(mydata$PRIXM2)[is.na(mydata$PRIXM2)==TRUE])
length(is.na(mydata1$prix.m2)[is.na(mydata1$prix.m2)==TRUE])

#Consider "TYPAPP_LIBEL" as 5th column
```

```
#Translate in english and make some modifications

levels(mydata$TYPAPP_LIBEL)[levels(mydata$TYPAPP_LIBEL)=="Appartement␣
    Standard"] <- "Standard␣Apartment"
levels(mydata$TYPAPP_LIBEL)[levels(mydata$TYPAPP_LIBEL)=="Grenier␣\'a␣
    am\'enager"] <- "Attic"
levels(mydata$TYPAPP_LIBEL)[levels(mydata$TYPAPP_LIBEL)=="Atelier␣d'
    Artiste"] <- "Artist␣atelier"
levels(mydata$TYPAPP_LIBEL)[levels(mydata$TYPAPP_LIBEL)=="Chambre␣de␣
    service"] <- "Service␣room"
levels(mydata$TYPAPP_LIBEL)[levels(mydata$TYPAPP_LIBEL)=="Logement␣de␣
    Gardien"] <- "Janitor's␣room"

levels(mydata1$TYPAPP_LIBEL)[levels(mydata1$TYPAPP_LIBEL)=="
    Appartement␣Standard"] <- "Standard␣Apartment"
levels(mydata1$TYPAPP_LIBEL)[levels(mydata1$TYPAPP_LIBEL)=="Grenier␣\'
    a␣am\'enager"] <- "Attic"
levels(mydata1$TYPAPP_LIBEL)[levels(mydata1$TYPAPP_LIBEL)=="Atelier␣d'
    artiste"] <- "Artist␣atelier"
levels(mydata1$TYPAPP_LIBEL)[levels(mydata1$TYPAPP_LIBEL)=="Chambre␣de
    ␣service"] <- "Service␣room"
levels(mydata1$TYPAPP_LIBEL)[levels(mydata1$TYPAPP_LIBEL)=="Logement␣
    de␣Gardien"] <- "Janitor's␣room"
levels(mydata1$TYPAPP_LIBEL)[levels(mydata1$TYPAPP_LIBEL)=="SU"] <- "
    Studio"
levels(mydata1$TYPAPP_LIBEL)[levels(mydata1$TYPAPP_LIBEL)=="␣"] <- NA

library(data.table)
data$TYPAPP = as.factor(c(as.character(mydata$TYPAPP_LIBEL),as.
    character(mydata1$TYPAPP_LIBEL)))
#Put standard apartment as first level: it will be useful
data$TYPAPP = relevel(data$TYPAPP,"Standard␣Apartment")

#Compute how many missing values
length(is.na(data$TYPAPP)[is.na(data$TYPAPP)==TRUE])
#Understand where they come from
length(is.na(mydata$TYPAPP_LIBEL)[is.na(mydata$TYPAPP_LIBEL)==TRUE])
length(is.na(mydata1$TYPAPP_LIBEL)[is.na(mydata1$TYPAPP_LIBEL)==TRUE])

#Consider "EPOQUE" (epoque) (9th column)

levels(mydata$EPOQUE)[levels(mydata$EPOQUE)=="avant␣1850"] <- "Before␣
    1850"
levels(mydata$EPOQUE)[levels(mydata$EPOQUE)=="1850␣\'a␣1913"] <- "1850
    ␣-␣1913"
levels(mydata$EPOQUE)[levels(mydata$EPOQUE)=="1914␣\'a␣1947"] <- "1914
    ␣-␣1947"
levels(mydata$EPOQUE)[levels(mydata$EPOQUE)=="1948␣\'a␣1969"] <- "1948
    ␣-␣1969"
levels(mydata$EPOQUE)[levels(mydata$EPOQUE)=="1970␣\'a␣1980"] <- "1970
    ␣-␣1980"
levels(mydata$EPOQUE)[levels(mydata$EPOQUE)=="1981␣\'a␣1991"] <- "1981
    ␣-␣1991"
levels(mydata$EPOQUE)[levels(mydata$EPOQUE)=="1992␣\'a␣2000"] <- "1992
    ␣-␣2000"
levels(mydata$EPOQUE)[levels(mydata$EPOQUE)=="\'a␣partir␣de␣2001"] <-
    "From␣2001"
mydata$EPOQUE <- relevel(mydata$EPOQUE,"Before␣1850")

levels(mydata1$EPOQUE)[levels(mydata1$EPOQUE)=="avant␣1850"] <- "
    Before␣1850"
```

```
levels(mydata1$EPOQUE)[levels(mydata1$EPOQUE)=="1850 \'a 1913"] <- "
    1850 - 1913"
levels(mydata1$EPOQUE)[levels(mydata1$EPOQUE)=="1914 \'a 1947"] <- "
    1914 - 1947"
levels(mydata1$EPOQUE)[levels(mydata1$EPOQUE)=="1948 \'a 1969"] <- "
    1948 - 1969"
levels(mydata1$EPOQUE)[levels(mydata1$EPOQUE)=="1970 \'a 1980"] <- "
    1970 - 1980"
levels(mydata1$EPOQUE)[levels(mydata1$EPOQUE)=="1981 \'a 1991"] <- "
    1981 - 1991"
levels(mydata1$EPOQUE)[levels(mydata1$EPOQUE)=="1992 \'a 2000"] <- "
    1992 - 2000"
levels(mydata1$EPOQUE)[levels(mydata1$EPOQUE)=="2001 \'a 2010"] <- "
    From 2001"
levels(mydata1$EPOQUE)[levels(mydata1$EPOQUE)=="Z"] <- NA
levels(mydata1$EPOQUE)[levels(mydata1$EPOQUE)=" "] <- NA
mydata1$EPOQUE <- relevel(mydata1$EPOQUE, "Before 1850")

data$EPOQUE = as.factor(c(as.character(mydata$EPOQUE),as.character(
    mydata1$EPOQUE)))
data$EPOQUE <- relevel(data$EPOQUE, "Before 1850")


#Compute how many missing values
length(is.na(data$EPOQUE)[is.na(data$EPOQUE)==TRUE])
#Understand where they come from
length(is.na(mydata$EPOQUE)[is.na(mydata$EPOQUE)==TRUE])
length(is.na(mydata1$EPOQUE)[is.na(mydata1$EPOQUE)==TRUE])

#Consider ARRON (10th column)

data$ARRON = as.factor(c(mydata$ARRON, mydata1$ARRON))
levels(data$ARRON)[levels(data$ARRON)=="0"] <- NA

#Compute how many missing values
length(is.na(data$ARRON)[is.na(data$ARRON)==TRUE])
#Understand where they come from
length(is.na(mydata$ARRON)[is.na(mydata$ARRON)==TRUE])
length(is.na(mydata1$ARRON)[is.na(mydata1$ARRON)==TRUE])


#Consider ANNEEMUTP (12th column)
data$ANNEEMUTP = c(mydata$ANNEEMUTP, mydata1$ANNEEMUTP)
data$ANNEEMUTP[data$ANNEEMUTP < 1898 | data$ANNEEMUTP > 2007] <- NA
data$ANNEEMUTP = factor(data$ANNEEMUTP)

#Compute how many missing values
length(is.na(data$ANNEEMUTP)[is.na(data$ANNEEMUTP)==TRUE])
#Understand where they come from
length(is.na(mydata$ANNEEMUTP)[is.na(mydata$ANNEEMUTP)==TRUE])
length(is.na(mydata1$ANNEEMUTP)[is.na(mydata1$ANNEEMUTP)==TRUE])


#Consider AGEB (13th column)
mydata$NAISSACQ[mydata$NAISSACQ==0]<-NA
mydata$AGEB = factor(mydata$ANNEE - mydata$NAISSACQ)
mydata1$AGEB = factor(mydata1$ANNEE - mydata1$NAISSACQ)
data$AGEB = as.factor(c(mydata$AGEB, mydata1$AGEB))

#Consider AGEVE (13th column)
mydata$NAISSVE[mydata$NAISSVE==0]<-NA
mydata$AGEVE = factor(mydata$ANNEE - mydata$NAISSVE)
```

```
mydata1$AGEVE = factor(mydata1$ANNEE - mydata1$NAISSVE)
data$AGEVE = as.factor(c(mydata$AGEVE, mydata1$AGEVE))

#Compute how many missing values
length(is.na(data$AGEB)[is.na(data$AGEB)==TRUE])
length(is.na(data$AGEVE)[is.na(data$AGEVE)==TRUE])
#Understand where they come from
length(is.na(mydata$AGEB)[is.na(mydata$AGEB)==TRUE])
length(is.na(mydata1$AGEB)[is.na(mydata1$AGEB)==TRUE])
length(is.na(mydata$AGEVE)[is.na(mydata$AGEVE)==TRUE])
length(is.na(mydata1$AGEVE)[is.na(mydata1$AGEVE)==TRUE])

#Consider CSPAC

levels(mydata$CSPAC)[levels(mydata$CSPAC)=="agriculteurs exploitants"]
    <- NA
levels(mydata$CSPAC)[levels(mydata$CSPAC)=="artisans/commerçants/chefs
    d'entreprise"] <- "artisans/shopkeepers/firm managers"
levels(mydata$CSPAC)[levels(mydata$CSPAC)=="cadres/ professions
    intellectuelles supérieures"] <- "executives/higher intellectual
    professions"
levels(mydata$CSPAC)[levels(mydata$CSPAC)=="retraités"] <- "retired"
levels(mydata$CSPAC)[levels(mydata$CSPAC)=="sans activité"] <- "
    without activity"
levels(mydata$CSPAC)[levels(mydata$CSPAC)=="employés"] <- "employees"
levels(mydata$CSPAC)[levels(mydata$CSPAC)=="marchand de biens"] <- NA
levels(mydata$CSPAC)[levels(mydata$CSPAC)=="non renseigne"] <- NA
levels(mydata$CSPAC)[levels(mydata$CSPAC)=="ouvriers"] <- "workers"
levels(mydata$CSPAC)[levels(mydata$CSPAC)=="professions intermédiaires
    "] <- "middle-ranking professionals"
levels(mydata$CSPAC)[levels(mydata$CSPAC)=="personnes morales"] <- "
    legal persons"

#Compute how many missing values
length(is.na(mydata$CSPAC)[is.na(mydata$CSPAC)==TRUE])


levels(mydata$CSPVE)[levels(mydata$CSPVE)=="agriculteurs exploitants"]
    <- NA
levels(mydata$CSPVE)[levels(mydata$CSPVE)=="artisans/commerçants/chefs
    d'entreprise"] <- "artisans/shopkeepers/firm managers"
levels(mydata$CSPVE)[levels(mydata$CSPVE)=="cadres/ professions
    intellectuelles supérieures"] <- "executives/higher intellectual
    professions"
levels(mydata$CSPVE)[levels(mydata$CSPVE)=="retraités"] <- "retired"
levels(mydata$CSPVE)[levels(mydata$CSPVE)=="sans activité"] <- "
    without activity"
levels(mydata$CSPVE)[levels(mydata$CSPVE)=="employés"] <- "employees"
levels(mydata$CSPVE)[levels(mydata$CSPVE)=="marchand de biens"] <- NA
levels(mydata$CSPVE)[levels(mydata$CSPVE)=="non renseigne"] <- NA
levels(mydata$CSPVE)[levels(mydata$CSPVE)=="ouvriers"] <- "workers"
levels(mydata$CSPVE)[levels(mydata$CSPVE)=="professions intermédiaires
    "] <- "middle-ranking professionals"
levels(mydata$CSPVE)[levels(mydata$CSPVE)=="personnes morales"] <- "
    legal persons"

#Compute how many missing values
length(is.na(mydata$CSPVE)[is.na(mydata$CSPVE)==TRUE])

#Consider DEPDA1
```

```
levels(mydata$DEPDA1)[levels(mydata$DEPDA1)=="Etranger"] <- "Foreigner
    "
levels(mydata$DEPDA1)[levels(mydata$DEPDA1)=="non renseigne"] <- NA
#Compute how many missing values
length(is.na(mydata$DEPDA1)[is.na(mydata$DEPDA1)==TRUE])

#Consider DEPDV

levels(mydata$DEPDV)[levels(mydata$DEPDV)=="Etranger"] <- "Foreigner"
levels(mydata$DEPDV)[levels(mydata$DEPDV)=="non renseigne"] <- NA
#Compute how many missing values
length(is.na(mydata$DEPDV)[is.na(mydata$DEPDV)==TRUE])

#Add a column: price of previous transaction per squared meter (only
    in 2005-2007 database)
mydata1$PXMUTPRECM2 <- mydata1$PXMUTPREC / mydata1$SURFE1
#Replace inf values by NA
is.na(mydata1) <- sapply(mydata1, is.infinite)

#Compute how many missing values
length(is.na(mydata1$PXMUTPREC)[is.na(mydata1$PXMUTPREC)==TRUE])
length(is.na(mydata1$PXMUTPRECM2)[is.na(mydata1$PXMUTPRECM2)==TRUE])


#Consider nationality of the buyer
data$NATIA= as.factor(c(as.character(mydata$NATIA),as.character(
    mydata1$NATIA)))
levels(mydata1$NATIA)[levels(mydata1$NATIA)=="  "] <- NA
summary(data$NATIA)

dt= data$NATIA=="F"  |
data$NATIA=="I"  |
data$NATIA=="GB"  |
data$NATIA=="USA"  |
data$NATIA=="P"  |
data$NATIA=="DZ"  |
data$NATIA=="RC"  |
data$NATIA=="E"  |
data$NATIA=="D"  |
data$NATIA=="MA"  |
data$NATIA=="TN"  |
data$NATIA=="B"  |
data$NATIA=="IRL"  |
data$NATIA=="CH"  |
data$NATIA=="J"  |
data$NATIA=="IR"  |
data$NATIA=="IL"  |
data$NATIA=="RL"  |
data$NATIA=="CDN"  |
data$NATIA=="K"

length(dt[dt==TRUE])


nomi=attributes(summary(data$NATIA[dt])[1:20])
freq=summary(data$NATIA[dt])[1:20]
a=tapply(data$PRIXM2[dt],data$NATIA[dt],mean,na.rm=TRUE)
a=a[!is.na(a)]
avgpr=a

#Consider nationality of the seller
```

```
data$NATIV= as.factor(c(as.character(mydata$NATIV),as.character(
    mydata1$NATIV)))
levels(mydata1$NATIV)[levels(mydata1$NATIV)==" "] <- NA
summary(data$NATIV)

dt1= data$NATIV=="F" |
data$NATIV=="I" |
data$NATIV=="GB" |
data$NATIV=="USA" |
data$NATIV=="DZ" |
data$NATIV=="D" |
data$NATIV=="E" |
data$NATIV=="P" |
data$NATIV=="MA" |
data$NATIV=="CH" |
data$NATIV=="B" |
data$NATIV=="TN" |
data$NATIV=="IL" |
data$NATIV=="J" |
data$NATIV=="RL" |
data$NATIV=="IR" |
data$NATIV=="RC" |
data$NATIV=="L" |
data$NATIV=="CDN" |
data$NATIV=="GR"

length(dt1[dt1==TRUE])


nomi1=attributes(summary(data$NATIV[dt1])[1:20])
freq1=summary(data$NATIV[dt1])[1:20]
a1=tapply(data$PRIXM2[dt1],data$NATIV[dt1],mean,na.rm=TRUE)
a1=a1[!is.na(a1)]
avgpr1=a1

miodataframe=data.frame(nomi,freq,avgpr,nomi1,freq1,avgpr1)
mytable <- xtable(miodataframe)
digits(mytable) <- c(0,2,3,0) #first digit is for counter,it must be
    included
print(mytable,include.rownames = FALSE)
```

C.2 Python code

```
fh=open('./base 2005-2007_Paris-copia.csv','r')
outfile = open('./base 2005-2007_Paris-copia-1.txt', '
    w')
line = fh.readline()
outfile.write(line)
listadate = list()
i = 0
for line in fh.readlines():
    s=line.strip().split(";")
    line = line.replace(s[3],s[3][-4:])
    outfile.write(line)
    i += 1
```

```
fh.close()
outfile.close()
```

Appendix D

Simulation code details

The purpose of this section is to show the code used for the numerical simulations in Section 4.3. We decided to use the programming language *NetLogo*[1] for this project. The reason of this choice is twofold. First, NetLogo's intuitive programming language is specifically designed for Agent Based Models and has an implicit object-oriented structure. So, it is possible to quickly modify the ABM in a modular way as some assumptions are changed. Second, the Graphical User Interface of NetLogo lets explore the results on a preliminary level, without the need to write input/output code. Another option we considered is Python: it is extremely versatile and intuitive, and it has an explicit object-oriented structure. It is perfect for ABMs where the behaviour of the agents is extremely complex (see, for instance, Terna and Taormina (2007), and the SLAPP project[2]). However, this is not the case for this project, where the complexity of the model arises from simple individual behaviour. Also C++ is commonly used for ABMs. Its nicest feature is its remarkable speed, which make it suitable to large scale simulations. Nevertheless, in the present work it is not needed to consider an extremely high number of agents and locations to get reliable results. Finally, Gama[3] is similar to NetLogo in many aspects, but it is mostly suitable to spatially explicit ABMs: it would be interesting to use it when considering empirical applications of this model.

In what follows, the code is presented with few comments: the extremely intuitive syntax of NetLogo is fairly similar to spoken language. Following the color code of NetLogo, comments are shown in grey. To keep the length reasonable, only the code necessary to reproduce the initialization and the dynamics described in Section 4.1 is shown. The parameters are as in Fig. 4.13b.

[1] `http://ccl.northwestern.edu/netlogo`
[2] whose latest version can be found at `https://github.com/terna/SLAPP`
[3] `https://github.com/gama-platform`

```
;these are the parameters and quantities useful on a global level
globals [number-of-apartments-per-location number-of-agents number-of-categories
     initial-price max-intrinsic-attractiveness R Y0 delta beta alpha gamma
     agents-list epsilon omega mu sigma lambda number-of-rounds]

;these quantities are useful at the level of each location
patches-own [modified-market-price intrinsic-attractiveness
     attractiveness-for-k-agents utility-for-k-agents probability-for-k-agents
     housed-k-agents-this-step market-price nb ns prices-of-transactions-this-step]

;these quantities are useful at the level of each agent
turtles-own [income income-category state purchase-price proposed-movement
     reservation-price proposed-price time-of-purchase]
;state = 0: buyer; state = 1: housed; state = 2: seller

to setup

  clear-all

  random-seed 1


  ;setup of the parameters

  ;parameters of the model
  set number-of-apartments-per-location 100
  set number-of-agents number-of-apartments-per-location * (count patches)
  set number-of-categories 5
  if number-of-agents mod number-of-categories != 0 [print "Attention: choose M, N
       and L s.t. M/numcategories is an entire number"]
  set initial-price 10
  set max-intrinsic-attractiveness 1
  set R 3
  set epsilon 0 ;weight of social effect
  set omega 1 / 15 ;time relaxation towards intrinsic attractiveness

  ;parameters of the agents
  set Y0 15 ; minimum income level
  set delta 5 ; difference in income between categories
  set beta 0.5 ; weight given to the attractiveness in the utility function
  set lambda 0.9 ;discount factor to determine the reservation price of the seller
  set mu (- number-of-apartments-per-location * 3 / 10) ;modified market price
       threshold
  set sigma number-of-apartments-per-location / 10 ;modified market price width

  ;parameters for agents coming and leaving
  set alpha 0.1 ;0.1 probability to put the apartment on sale
  set gamma 2000 ;number of incoming agents any time step
  set agents-list (list (5 / 15) (4 / 15) (3 / 15) (2 / 15) (1 / 15) )

  ;parameters for the order book
  set number-of-rounds 30 ;5 maximum number of trading periods
  ;this is just to prevent order book loops to be too long


  ;setup of the patches

  ask patches [
    set modified-market-price initial-price
    set market-price initial-price
    set intrinsic-attractiveness max-intrinsic-attractiveness * exp (-(pxcor ^ 2 +
         pycor ^ 2) /(R ^ 2))
```

```
    set housed-k-agents-this-step []
    set attractiveness-for-k-agents []
    set utility-for-k-agents []
    set probability-for-k-agents []
    let k 0
    repeat number-of-categories [
      set housed-k-agents-this-step lput 0 housed-k-agents-this-step
      set attractiveness-for-k-agents lput intrinsic-attractiveness
           attractiveness-for-k-agents
      set utility-for-k-agents lput ( compute-utility ( Y0 + k * delta )
           market-price (item k attractiveness-for-k-agents) ) utility-for-k-agents
      set probability-for-k-agents lput 0 probability-for-k-agents
      set k k + 1
      ]
  ]


  ;setup of the agents

  let k 0
  repeat number-of-categories [
  create-turtles number-of-agents / number-of-categories
                                [
                                 set purchase-price initial-price
                                 set reservation-price purchase-price
                                 move-to one-of patches with [count turtles-here with
                                      [state = 1] < number-of-apartments-per-location]
                                 set income Y0 + k * delta
                                 set income-category k
                                 set state 1]
  set k k + 1
  ]
  reset-ticks

end



to go

  ;create the agents at the beginning of each time step
  create-agents
  ;compute the probabilities for k-agents to visit any location
  compute-probabilities
  ;some agents decide to put their house on sale
  become-seller
  ;incoming agents search for an apartment with a probability proportional to the
      utility of going there
  choose-location
  ;matching happens through a continuos double auction
  do-continuos-double-auction
  ;updates attractiveness and computes "patch utility"
  update-attractiveness
  tick
  ;removes the turtles which were not successful in finding an apartment
  ask turtles with [state = 0] [die]

end
```

```
to become-seller

  ask turtles with [state = 1] [
    let u random-float 1
    if u < alpha [set state 2]
  ]

end



to create-agents

  let k 0
  repeat number-of-categories [
    create-turtles gamma * (item k agents-list) [
      set income Y0 + k * delta
      set income-category k
      set xcor min-pxcor
      set ycor min-pycor
      set state 0
    ]
    set k k + 1
  ]

end



to compute-probabilities

  let normalizationk []
  let k 0
  repeat number-of-categories [
    set normalizationk lput sum ( [item k utility-for-k-agents] of patches )
         normalizationk
    if item k normalizationk = 0 [set normalizationk replace-item k normalizationk 1]
    set k k + 1
  ]
  ask patches [
    set k 0
    repeat number-of-categories [
      set probability-for-k-agents replace-item k probability-for-k-agents ( item k
           utility-for-k-agents / ( item k normalizationk ) )
      set k k + 1
    ]
  ]

end



to choose-location

 let k 0
 let L sqrt (count patches)
 repeat number-of-categories [
   let ordered-probabilities-list [] ;from topleft to topright, until bottomleft to
        bottomright
   let i 0
   repeat (count patches) [
     set ordered-probabilities-list lput ([item k probability-for-k-agents] of patch
```

```
        ( min-pxcor + i mod L ) ( max-pycor - int (i / L) ) )
        ordered-probabilities-list
    set i i + 1
  ]
  let cumulated-list []
  set cumulated-list lput 0 cumulated-list
  set i 1
  repeat (length ordered-probabilities-list) [set cumulated-list lput (item (i - 1)
      ordered-probabilities-list + item (i - 1) cumulated-list) cumulated-list set
      i i + 1 ]
    ask turtles with [state = 0 and income-category = k] [ ;those looking for a new
        house
      let chosen-location-id inverse-transform cumulated-list
      set proposed-movement patch ( min-pxcor + chosen-location-id mod L ) (
          max-pycor - int (chosen-location-id / L) )
  ]
  set k k + 1
 ]

end



to do-continuos-double-auction

  ask patches [

   set prices-of-transactions-this-step []

   let buyers turtles with [proposed-movement = myself and state = 0]
   let sellers turtles-here with [state = 2]

   if not any? sellers or not any? buyers [
     update-market-price
     stop]

   set nb count buyers
   set ns count sellers

   set modified-market-price market-price * 1 / (1 + e ^ ( - ((nb - ns) - mu) /
       sigma ) )
   ask buyers [set reservation-price income]
   ask sellers [set reservation-price max list ([modified-market-price] of myself)
     ([modified-market-price] of myself + lambda ^ (ticks - time-of-purchase) *
         (purchase-price - [modified-market-price] of myself) )]

   let mean-res-offer-price mean [reservation-price] of sellers
   let max-price max [reservation-price] of (turtle-set buyers sellers)

   ask buyers [set proposed-price mean-res-offer-price]
   ask sellers [set proposed-price max-price]

   let count-rounds 0
   repeat number-of-rounds [

     let logB []
     let logS []

     if not any? sellers with [state = 2] or not any? buyers with [state = 0] or max
         [reservation-price] of buyers with [state = 0] < min [reservation-price]
         of sellers with [state = 2]
     [update-market-price stop]
```

```
    ask (turtle-set (buyers with [state = 0]) (sellers with [state = 2]) ) [

      ifelse state = 0 [
        let tmp[]
        set tmp lput proposed-price tmp
        set tmp lput who tmp
        set tmp lput [income-category] of self tmp
        set logB lput tmp logB
        set logB reverse sort-by [item 0 ?1 < item 0 ?2] logB
      ]

      [
        let tmp[]
        set tmp lput proposed-price tmp
        set tmp lput who tmp
        set tmp lput [income-category] of self tmp
        set logS lput tmp logS
        set logS sort-by [item 0 ?1 < item 0 ?2] logS
      ]

      if (not empty? logB and not empty? logS) and item 0 (item 0 logB) >= item 0
           (item 0 logS) [

        let transaction-price 0.5 * (item 0 (item 0 logS)) + 0.5 * (item 0 (item 0
             logB))

        let seller turtle (item 1 (item 0 logS))
        let buyer turtle (item 1 (item 0 logB))

        ask myself [
          set prices-of-transactions-this-step lput transaction-price
               prices-of-transactions-this-step
          set housed-k-agents-this-step replace-item ([income-category] of buyer)
               housed-k-agents-this-step (item ([income-category] of buyer)
               housed-k-agents-this-step + 1 )
        ]

        set logB but-first logB
        set logS but-first logS

        ask buyer [set state 1 move-to proposed-movement set purchase-price
             transaction-price set reservation-price purchase-price set
             time-of-purchase ticks]
        ask seller [die]
      ]

      let nb1 [nb] of myself
      let ns1 [ns] of myself
      ifelse state = 0 [set proposed-price proposed-price + (nb1 / ns1) / ( 1 +
           (nb1 / ns1)) * (reservation-price - proposed-price)]
                     [set proposed-price proposed-price + (ns1 / nb1) / ( 1 + (ns1
                          / nb1)) * (reservation-price - proposed-price)]

    ]
    set count-rounds count-rounds + 1
  ]

   update-market-price

 ]
```

```
end

to update-market-price ;to be run by a patch

  if prices-of-transactions-this-step != []
    [set market-price (mean prices-of-transactions-this-step) ]

end


to update-attractiveness

  ask patches [
    let k 0
    repeat number-of-categories [
      set attractiveness-for-k-agents replace-item k attractiveness-for-k-agents
           (item k attractiveness-for-k-agents +
       omega * (intrinsic-attractiveness - item k attractiveness-for-k-agents ) +
            epsilon * sum ( sublist housed-k-agents-this-step k ( (length
            housed-k-agents-this-step ) ) ) )
      set housed-k-agents-this-step replace-item k housed-k-agents-this-step 0
      set utility-for-k-agents replace-item k utility-for-k-agents ( compute-utility
           ( Y0 + k * delta ) market-price (item k attractiveness-for-k-agents) )
      set k k + 1
    ]

  ]

end



to-report compute-utility [Y PX Ak]

  ifelse Y - PX >= 0
    [ report (Y - PX) ^ (1 - beta) * Ak ^ beta ]
    [ report 0 ]

end


to-report inverse-transform [lista]

  let u random-float 1
  let l1 0
  let l2 length lista

  while [l2 - l1 > 1] [

    let l12 int ( ( l1 + l2 ) / 2 )
    ifelse u < ( item l12 lista ) [set l2 l12] [set l1 l12]

  ]

  report l1
end
```

Bibliography

Abu-Mostafa, Y. S., Magdon-Ismail, M. and Lin, H.-T. (2012). *Learning from data*. AMLBook.

Alonso, W. *et al.* (1964). *Location and land use. Toward a general theory of land rent.*. Cambridge, Mass.: Harvard Univ. Pr.

Anderson, P. W., Arrow, K. and Pines, D. (1988). *The economy as an evolving complex system*. Redwood City, Addison-Wesley Co.

Arellano, M. (2003). *Panel data econometrics*. In ≪OUP Catalogue≫.

Arthur, W. B. (2014). *Complexity economics: a different framework for economic thought*. Oxford University Press.

Bak, P., Paczuski, M. and Shubik, M. (1997). *Price variations in a stock market with many agents*. In ≪Physica A: Statistical Mechanics and its Applications≫, vol. 246(3), pp. 430–453.

Batty, M. (2008). *The size, scale, and shape of cities*. In ≪science≫, vol. 319(5864), pp. 769–771.

Bettencourt, L. M. (2013). *The origins of scaling in cities*. In ≪science≫, vol. 340(6139), pp. 1438–1441.

Bouchaud, J.-P. (2008). *Economics needs a scientific revolution*. In ≪Nature≫, vol. 455(7217), pp. 1181–1181.

— (2013). *Crises and collective socio-economic phenomena: simple models and challenges*. In ≪Journal of Statistical Physics≫, vol. 151(3-4), pp. 567–606.

Cameron, A. C. and Trivedi, P. K. (2005). *Microeconometrics: methods and applications*. Cambridge university press.

Chakraborti, A., Toke, I. M., Patriarca, M. and Abergel, F. (2011). *Econophysics review: II. Agent-based models*. In ≪Quantitative Finance≫, vol. 11(7), pp. 1013–1041.

Cont, R. and Bouchaud, J.-P. (2000). *Herd behavior and aggregate fluctuations in financial markets*. In ≪Macroeconomic dynamics≫, vol. 4(02), pp. 170–196.

Cowan, G. (1998). *Statistical data analysis*. Oxford University Press.

Feitosa, F. F., Reyes, J. and Zesk, W. (2008). *Spatial Patterns of Residential Segregation: A generative Model.*. In *GeoInfo*. pp. 157–162.

Filatova, T., Parker, D. and Van der Veen, A. (2009). *Agent-based urban land markets: agent's pricing behavior, land prices and urban land use change*. In ≪Journal of Artificial Societies and Social Simulation≫, vol. 12(1), p. 3.

Fujita, M. (1989). *Urban economic theory: land use and city size*. Cambridge university press.

Gallegati, M., Keen, S., Lux, T. and Ormerod, P. (2006). *Worrying trends in econophysics*. In ≪Physica A: Statistical Mechanics and its Applications≫, vol. 370(1), pp. 1–6.

Gauvin, L., Vignes, A. and Nadal, J.-P. (2013). *Modeling urban housing market dynamics: can the socio-spatial segregation preserve some social diversity?*. In ≪Journal of Economic Dynamics and Control≫, vol. 37(7), pp. 1300–1321.

Gilbert, N., Hawksworth, J. C. and Swinney, P. A. (2009). *An Agent-Based Model of the English Housing Market.*. In *AAAI Spring Symposium: Technosocial Predictive Analytics*. pp. 30–35.

Grauwin, S., Bertin, E., Lemoy, R. and Jensen, P. (2009). *Competition between collective and individual dynamics*. In ≪Proceedings of the National Academy of Sciences≫, vol. 106(49), pp. 20622–20626.

Gualdi, S., Tarzia, M., Zamponi, F. and Bouchaud, J.-P. (2015). *Tipping points in macroeconomic agent-based models*. In ≪Journal of Economic Dynamics and Control≫, vol. 50, pp. 29–61.

Ioannides, Y. M. and Zabel, J. E. (2003). *Neighbourhood effects and housing demand*. In ≪Journal of applied Econometrics≫, vol. 18(5), pp. 563–584.

Kirman, A. (2010a). *Complex economics: individual and collective rationality*. Routledge.

— (2010b). *The economic crisis is a crisis for economic theory*. In ≪CESifo Economic Studies≫, vol. 56(4), pp. 498–535.

Kirman, A. P. (1992). *Whom or what does the representative individual represent?*. In ≪The Journal of Economic Perspectives≫, pp. 117–136.

Krugman, P. (2009). *How did economists get it so wrong?*. In ≪New York Times≫, vol. 6(09).

Landvoigt, T., Piazzesi, M. and Schneider, M. (2015). *The Housing Market(s) of San Diego*. In ≪American Economic Review≫, vol. 105(4), pp. 1371-1407.

Lemoy, R. (2011). *The City as a Complex System. Statistical Physics and Agent-Based Simulations on Urban Models*. PhD Thesis, Université Lumière-Lyon II.

Lemoy, R., Bertin, E. and Jensen, P. (2011). *Socio-economic utility and chemical potential*. In ≪EPL (Europhysics Letters)≫, vol. 93(3), p. 38002.

Lemoy, R., Raux, C. and Jensen, P. (2010). *An agent-based model of residential patterns and social structure in urban areas*. In ≪Cybergeo: European Journal of Geography≫.

Louf, R. and Barthelemy, M. (2013). *Modeling the polycentric transition of cities*. In ≪Physical review letters≫, vol. 111(19), p. 198702.

Lux, T. and Marchesi, M. (1999). *Scaling and criticality in a stochastic multi-agent model of a financial market*. In ≪Nature≫, vol. 397(6719), pp. 498–500.

Mas-Colell, A., Whinston, M. D., Green, J. R. *et al.* (1995). *Microeconomic theory*, vol. 1. Oxford university press New York.

McCauley, J. L. (2006). *Response to "worrying trends in econophysics"*. In ≪Physica A: Statistical Mechanics and its Applications≫, vol. 371(2), pp. 601–609.

Munnell, A. H., Tootell, G. M., Browne, L. E. and McEneaney, J. (1996). *Mortgage lending in Boston: Interpreting HMDA data*. In ≪The American Economic Review≫, pp. 25–53.

Pigou, A. C. (2013). *The economics of welfare*. Palgrave Macmillan.

Samuelson, P. A. (1983). *Thünen at two hundred*. In ≪Journal of Economic Literature≫, pp. 1468–1488.

Schelling, T. C. (1971). *Dynamic models of segregation*. In ≪Journal of mathematical sociology≫, vol. 1(2), pp. 143–186.

Short, M. B., D'Orsogna, M. R., Pasour, V. B., Tita, G. E., Brantingham, P. J., Bertozzi, A. L. and Chayes, L. B. (2008). *A statistical model of criminal behavior*. In ≪Mathematical Models and Methods in Applied Sciences≫, vol. 18(supp01), pp. 1249–1267.

Smith, E., Farmer, J. D., Gillemot, L. s., Krishnamurthy, S. *et al.* (2003). *Statistical theory of the continuous double auction*. In ≪Quantitative finance≫, vol. 3(6), pp. 481–514.

Stock, J. H. and Watson, M. W. (2003). *Introduction to econometrics*, vol. 104. Addison Wesley Boston.

Terna, P. and Taormina, R. (2007). *Modelli di simulazione con agenti intelligenti: il sorprendente mondo dei camaleonti*. In ≪Sistemi intelligenti≫, vol. 19(3), pp. 391–426.

Tiebout, C. M. (1956). *A pure theory of local expenditures*. In ≪The journal of political economy≫, pp. 416–424.

Trichet, J.-C. (2010). *Reflections on the nature of monetary policy non-standard measures and finance theory*. In *opening address at the 6th ECB Central Banking Conference, Frankfurt am Main*, vol. 18.

Varian, H. R. (2014). *Big data: New tricks for econometrics*. In ≪The Journal of Economic Perspectives≫, vol. 28(2), pp. 3–27.

Von Thünen, J. (1826). *Der isolierte Staat*. In ≪Beziehung auf Landwirtschaft und Nationalökonomie≫.