



**Politecnico
di Torino**

Politecnico di Torino

Physics of Complex Systems

A.a. 2024/2025

Graduation Session October 2025

Modeling Gentrification Dynamics: An Agent-Based Approach to Segregation and Public Housing Policies

Supervisors:

Prof. Luca Dall'Asta
Dr. Vito D. P. Servedio

Candidate:

Flavio Brandoli

Abstract

Gentrification is a pressing and controversial phenomenon that is reshaping thousands of cities around the world. Although it offers the possibility of revitalizing degraded urban areas, its potential to displace long-term low-income residents poses a major threat. Many aspects of this phenomenon are still widely debated, including whether it results in greater social and income integration, and which policies could mitigate its harmful effects.

In this thesis, we propose an agent-based model (ABM) to study the impact of gentrification on urban segregation and to explore how public housing policies could affect its dynamics. The model simulates a stylized city populated by two classes of agents: low- and high-income residents. Agents move across the city according to a Schelling-like rule: a utility function defines their satisfaction level, which determines whether an agent is happy or unhappy. At each step, happy agents remain in place, while unhappy ones relocate. A developer agent drives the gentrification process by renovating properties in response to market signals. The model is initialized using real-world data from the United States. Although it is designed to represent a generic U.S. city rather than a specific one, it could be calibrated in future work to match a particular urban context and used for predictive purposes.

The model reproduces gentrification patterns that are consistent with observed trends in income and rent increases in real U.S. neighborhoods. According to our simulations, gentrification does not necessarily improve integration between income groups; from certain perspectives, it may even worsen it. The model also shows that even a small fraction of public housing units can be very effective in slowing down gentrification dynamics.

While using ABMs to model gentrification and segregation is not new, introducing an external agent as the explicit driver of the process, as well as incorporating housing policy interventions into the simulation, represent an innovative and underexplored approach.

Table of Contents

List of Figures	III
Introduction	1
1 Gentrification and Agent Based Modelling	3
1.1 Review of the literature debate on gentrification	3
1.1.1 Smith’s rent-gap theory	5
1.2 Schelling’s model of segregation	6
1.3 Models of gentrification	9
2 Our Modeling Scheme	13
2.1 A first toy agent-based model of gentrification	13
2.2 City Environment	20
2.3 Agents	23
2.4 Dynamics	24
2.4.1 Thermalization phase	25
2.4.2 Gentrification dynamics	28
2.5 Model validation	34
2.5.1 Real-world examples of gentrification	34
2.5.2 Simulations clustering	35
2.6 Introduction of public housing	42
3 Results	45
3.1 Mitigation effects of public housing	45
3.2 Segregation measures	49
3.2.1 Neighbor segregation index	50
3.2.2 Dissimilarity index	52
3.2.3 Theil index	56
Conclusions	61

A	63
B	64
Bibliography	66

List of Figures

1.1	The depreciation cycle of inner-city neighborhoods, from [11]. . . .	6
1.2	Comparison between the initial state of Schelling's model, where agents are randomly distributed on the grid (a) , and the stationary, visibly segregated state reached after few steps, where all the agents are happy (b) . (Vacancy Density: $\rho_0 = 0.02$, Tolerance Threshold: $f^* = 0.65$)	7
1.3	Trend of the neighbor segregation index (the fraction of neighbor links connecting two agents of the same color) averaged over 20 simulations that differ in the initial random configuration. The index significantly increases over the dynamics, meaning that the system evolves towards a more segregated state. (Vacancy Density: $\rho_0 = 0.02$, Tolerance Threshold: $f^* = 0.65$)	9
1.4	Housing age as a function of the distance from the center (x), from [25]. Comparison between three phases of the city expansion: (a) first phase, redevelopment has not occurred yet; (b) second phase, first round of redevelopment started; (c) third phase, second round of redevelopment started.	10
1.5	Plots from Mauro and Pappalardo's model [31]. (a) Initial spatial distribution of agents in Mauro and Pappalardo agent-based model. High-income agents predominantly occupy the city center, medium-income the inner areas, and low-income the periphery. (b,c) Moving out probability as a function of the agent's relative income percentile, respectively of low and medium-income classes.	12
2.1	Initial state of the city. High-income agents (red dots) are spread across the central quadrant, low-income ones (blue dots) are distributed in the four suburbs. The four corner quadrants are not part of the city. Imagine them as mountains.	14

2.2	Initial quadrant attributes in a specific simulation. Beauty: It is initially set to be high in the center and low in the suburbs; Cost of living: It is determined by the fraction of high income agents present; Density: It is initially fixed in the center and sampled from a Gaussian distribution in the suburbs.	15
2.3	Comparison between the thermalized state (a) and the final state (b) reached after the lower quadrant has undergone gentrification. .	17
2.4	Final quadrant properties. The central and the peripheral quadrants are now quite balanced in each property, while in the other peripheral areas the density increased.	18
2.5	Trends of the average satisfaction of the two groups during the gentrification dynamics. The trend is aggregate over 20 different simulations that differ in the initial random setting.	19
2.6	Plot of the model's environment. The three grids represent the same sketched city. The green patches (representing the amenities) and the black ones (representing the CBD) are the same in the three plots, what changes is the property of the housing patches displayed. (a) Age: The shade of orange represent the age of the dwellings, which, on average, decreases with the distance from the center. (b) Potential rent: The potential rent is higher for houses that are surrounded by a greater number of amenities. (c) Capitalized rent: The capitalized rent is a combination of the potential rent and the age of a house. This environment results in a city with cheaper houses in the inner-city, and some expensive hotspots in the suburbs.	20
2.7	(a) Capitalized rent as a function of the age for two values of the potential rent. (b) Rent-gap as a function of the age. At the same age, a higher potential rent is associated with a higher rent-gap. In both cases the confidence bands represent the variability due to the Gaussian noise.	23
2.8	Empirical distribution of annual household incomes, based on data from the 2013 American Community Survey (ACS). The distribution was used to assign income levels to agents in the model. Note that the income brackets in the source data have non-uniform sizes. . .	24
2.9	Income distributions. The gray dashed line shows the empirical data, while the blue line shows the sampled distribution from a single simulation run. The pink, dotted line represents the percentile threshold dividing the two agent classes.	25

2.10	Comparison between the initial state in which agents are randomly distributed on the grid (a) , and the stable state reached after few steps of thermalization (b) . Low-income agents are represented by the blue dots and high-income by the red dots. In the grid background it is displayed the capitalized rent of the housing patches.	26
2.11	Trend of the neighbor segregation index averaged over 20 simulations of the thermalization phase. The index increases over the dynamics, meaning that the system evolves towards a more segregated state.	27
2.12	Distributions of rents as a percentage of resident income. The gray dashed line shows the empirical data, while the blue line represents the aggregate distribution obtained by combining data from 20 simulations after the thermalization phase.	28
2.13	Comparison of the three housing patch properties before and after the gentrification dynamics. (a) Age: The age of redeveloped houses decreases, lightening them from orange to yellow. (b) Potential rent: Color yellow is associated with high potential rent, it increases because of spillover effects. (c) Capitalized rent: Rents of renewed houses and of their neighbors increase, changing color from purple to red.	32
2.14	Trends of the median of the rents and the median of the incomes in the inner-city for two different simulation. Seed 215646 (green) represents a successful gentrification case: both the indicators grow significantly; seed 4867 (red) is instead an unsuccessful example. .	33
2.15	An example of unsuccessful gentrification: few houses are renewed in the inner-city, while the investments in the suburbs are significant.	33
2.16	Trends of median rents (a) and incomes (b) in real U.S. neighborhoods from 2011 to 2020. The gentrifying areas (continuous lines) display a significant growth in both indicators, while the control neighborhoods (dashed lines) show little or no growth. The numbers in the legend represent the ZIP code identifying that area.	35
2.17	Trend of median inner-city rent and income from a single simulation. The 10 values shown are obtained by applying a rolling moving average over overlapping intervals of 40 steps each on the 112 simulated steps. All values are indexed to the first one. The dashed gray curves in the background represent the empirical data.	36
2.18	Clustering of the 600 simulations in the 4-parameter space, shown in all six possible pairwise 2D projections.	38
2.19	Trends of median rents (a) and incomes (b) of the 600 simulations, colored according to 3-means clustering result shown in Figure 2.18. The empirical trends are included in the background, but are barely visible because they are covered by the simulation lines.	38

2.20	Comparison of average perceived amenity (a) and the initial number of low-income agents (b) in the inner city between the 3 clusters. The most successful gentrification cluster is in red, the least successful gentrification cluster is in blue. Boxes represent the central 50% of simulation values for that cluster, horizontal black line stands for the average value, whiskers cover the full range of the distribution apart from the few outliers displayed as individual dots.	40
2.21	Boxplots of all the evaluated metrics.	41
2.22	City state at the end of the dynamics after the public housing intervention. Capitalized rent as a background. Public houses are displayed in white. They are randomly selected in (a) and selected by rent gap in (b) . (Public housing fraction set to 0.05).	44
3.1	Trends during the gentrification dynamics in the three defined scenarios: no public housing intervention, public housing selected randomly, and public housing selected based on rent gap. (a) Inner-city median income: In the two scenarios involving public housing, a few steps after the intervention, this indicator stops increasing. (b) Number of agents inside inner-city: In the free-market scenario, the number of high-income agents in the inner city continues to grow during the process, while the number of low-income agents decreases. Once public housing is introduced, both trends quickly stabilize. (Public housing fraction set to 0.05).	46
3.2	Number of developer's investments in the inner-city throughout the dynamics, for a single simulation. Randomly selected public housing units (a) and selected by rent gap (b) . In both plots the free dynamics trend is shown in the background in blue and two vertical, dashed lines indicate the start of the dynamics and the policy intervention. In both public housing scenarios the number of investments drops after the intervention. (Public housing fraction set to 0.05).	47
3.3	Number of developer's investments in the inner-city throughout the dynamics, averaged across 15 simulations. Randomly selected public housing units (a) and selected by rent gap (b) . (Public housing fraction set to 0.05).	48
3.4	(a), (b) Trends of inner-city median income, aggregated over 15 simulations, plotted for different fractions of public housing units. Panels (a) and (b) correspond to public housing units selected randomly and based on the rent gap, respectively. (c) Plot of the final values of income for all the fractions simulated.	49

3.5	(a), (b) Trends of neighbor segregation index, aggregated over 15 simulations, plotted for different fractions of public housing units. Panels (a) and (b) correspond to public housing units selected randomly and based on the rent gap, respectively. (c) Plot of the final values of the index for all the fractions simulated.	51
3.6	(a), (b) Trends of dissimilarity index, aggregated over 15 simulations, plotted for different fractions of public housing units. Panels (a) and (b) correspond to public housing units selected randomly and based on the rent gap, respectively. (c) Plot of the final values of the index for all the fractions simulated.	53
3.7	Evolution of the pairwise dissimilarity index considering three different income groups, averaged across 15 simulation runs. No public housing intervention scenario.	55
3.8	Evolution of the pairwise dissimilarity index considering three different income groups, averaged across 15 simulation runs. Public housing fraction set to 0.05. (a) Random selection of public housing units. (b) Selection by rent-gap.	56
3.9	Evolution of the different contributions to the Theil segregation index, averaged across 15 simulation runs. No public housing intervention scenario.	59
3.10	Evolution of the different contributions to the Theil segregation index, averaged across 15 simulation runs. Public housing fraction set to 0.05. (a) Random selection of public housing units. (b) Selection by rent-gap.	60

Introduction

In recent years, for the first time in human history, the world’s urban population surpassed 50%, and, according to UN projections, it is expected to reach 68% by 2050 [1]. As a consequence of this dramatic growth, cities worldwide are undergoing profound transformations, and numerous efforts are being made across different disciplines to study and predict their evolution. Given their strongly interactive nature and the abundance of available data, cities offer a unique opportunity to apply tools from complexity science. Indeed, this approach has been increasingly adopted over the past few decades to investigate various aspects of urban systems, including mobility, infrastructure networks, urban scaling and sprawl, and patterns of socio-economic segregation, among others.

Gentrification is one of the main drivers of urban change, transforming degraded areas historically inhabited by low-income residents into trendy and expensive neighborhoods. Because of its complex and controversial nature, it has been widely studied and intensely debated by sociologists and economists. However, relatively few attempts have been made to tackle it from a quantitative, model-based perspective. Among the most frequently discussed topics surrounding gentrification are the impact this process has on urban segregation, and the role that public policies can play in mitigating its effects.

In this thesis, we present an agent-based modeling approach to study how segregation between different income groups evolves during gentrification dynamics, and whether public housing policies can effectively mitigate the process. Thinking of gentrification as the macroscopic outcome of many individual decisions — made by investors deciding which areas to redevelop, residents choosing whether to remain in their neighborhood or relocate, and local governments regulating the housing market — an agent-based approach proved to be a natural modeling choice.

The thesis is structured as follows:

1. **Chapter 1:** We briefly review the literature on gentrification and present the economic theory adopted as the basis for our model. We also introduce agent-based models, discussing in detail the first example of an ABM: Schelling’s model. Finally, we present a few existing models of gentrification found in the literature.

2. **Chapter 2:** We begin the chapter by introducing a simple, agent-based toy-model of gentrification. We then present the central ABM, first describing the environment and the agents, then the logic of the dynamics. We validate the model by comparing its outcomes with real-world data, and conclude by explaining the mechanism implemented to simulate public housing policies.
3. **Chapter 3:** We analyze the behavior of our model. First, we compare the gentrification outcomes produced under different policy scenarios. Then, we define three segregation indices and study how urban segregation evolves during the dynamics from these three different perspectives.

The model was implemented in Julia using the Agents.jl library. Part of the code — including the core functions needed to run the simulation — is available on GitHub at <https://github.com/FlavioBrandoli/ABM-Rent-Gap-Gentrification>.

Chapter 1

Gentrification and Agent Based Modelling

In this chapter, we present a literature review on gentrification, stressing the controversies and the debate on this topic. We also introduce Schelling-like agent-based models, and justify why we chose these types of model to tackle the process of gentrification.

1.1 Review of the literature debate on gentrification

In the literature, gentrification is often defined as the process by which higher-income households move to a neighborhood traditionally inhabited by lower-income households, changing the essential character and costs of that area, often causing the displacement of long-term residents [2, 3, 4]. From this general definition, it is already evident that the consequences of gentrification are intrinsically controversial: on the one hand the physical changes of a gentrifying area are generally positive, such as revitalization of parks and streets, emergence of new commercial activity and infrastructure, renovation of the housing stock [2, 4, 5]; on the other hand, these improvements lead to a rise in the cost of living, that may force the displacement of long-term low-income residents [3, 6].

The role of displacement in gentrification is actually debated within the literature: Freeman and Braconi [7] compare mobility data from New York City between gentrifying neighborhoods and similar neighborhoods that did not undergo the process and conclude that gentrification is not associated with a higher level of displacement, in contrast to anecdotal evidence. However other studies challenged these results, such as Johnson's [5], who argued that Freeman and Braconi did not

observe significant displacement, just because they considered the initial phase of gentrification only, ignoring long-term effects; or Newman and Wyly who, analysing the same dataset used by Freeman and Braconi, drew opposite conclusions [8].

Displacement is related to another debated issue around gentrification: the effect the process has on integration (or segregation) between different social and economic classes. Indeed, one of the arguments often mentioned in favor of gentrification is that it leads to greater social mix, thereby improving integration between classes (whether a more socially mixed city actually results in better living conditions is not discussed here, but a critical perspective on this issue can be found in [9]). However, the effectiveness of this mixing process naturally depends on how significant displacement is: social mix will not be achieved if the arrival of high-income residents in a neighborhood forces the departure of low-income ones. Slater [6] takes a skeptical position on the idea of gentrification as a way to integrate classes, he expresses his concern that the actual outcome, due to displacement, may be a more segregated city; Johnson [5], on the other hand, argues that gentrification can lead either to better integration or to further segregation, and that it is up to policy and planning to drive the process in one direction or the other.

The argument that gentrification consequences are not universal, but strongly related to the laws and policies established in cities is largely shared within the literature. Kennedy and Leonard [2] expose nine steps to follow to optimize the end result of gentrification. These strategies involve prevention, community organization and housing policies. Again, Marcuse [3] provides a detailed, zone by zone guide to housing policies that should be adopted in New York City to ensure residents of gentrifying neighborhoods can benefit from the changes, without being threatened by displacement.

The mentioned studies all agree on the fact that housing policies are one of the most effective tools governments have to contrast and mitigate the negative consequences of gentrification. Long-term residents can be protected from displacement by different strategies: such as increasing the number of affordable housing units in gentrifying neighborhoods, or prohibiting new development and construction that would increase rents and prices in the most vulnerable areas. A relevant example of the effectiveness of these types of policies is the city of Vienna where approximately half of the population live in municipal housing estates (houses owned and managed by the City of Vienna) or subsidized dwellings (houses built with economic contributions from the Federal Province of Vienna). This massive presence of non-private houses makes Vienna the biggest public owner of social housing estates in Europe, and it allows the city to keep prices and rent down and to contrast social segregation. Franz [10] studied how this unique scenario offered by Vienna affects gentrification: she compared Vienna, Berlin and New York, confirming that the significant stock of public housing in the Austrian capital acts as a crucial mitigating factor, changing significantly the dynamics of the

process. In New York and Berlin the private rental market has stronger influence and gentrification occurs faster and more intensively with respect to Vienna.

Gentrification is responsible for deep transformations in cities all over the world, so it is not surprising that the discussion around this topic has been so animated over the last decades. Studying the literature it emerged as an interesting and debated topic whether gentrification brings to further segregation or to a more integrated city. It also emerged the significant role played by public housing policies in affecting its dynamics. We will develop an agent-based model to explore these two aspects of gentrification. The rest of this chapter is dedicated at providing the knowledge and context necessary to understand the model we present in Chapter 2.

1.1.1 Smith's rent-gap theory

The factors that can cause gentrification are several, and change from one country to another. In this section, we discuss the theory used as the basis for the model we present in Chapter 2. This theory refers to cities in the United States, and focuses on housing age as the main trigger of the process. Information on different gentrification causes can be found in [2].

This theory is the Rent-Gap Theory of Gentrification, developed by Geographer Neil Smith in 1979 [11]. According to Smith, the well-known division in US cities, between higher-income residents living in the suburbs and lower-income ones living in the inner-city, is largely due to the age and deterioration of buildings: as a city grows, new housing is built at the edges, attracting wealthier residents, who can afford newer and more expensive houses. Meanwhile, the inner-city housing stock deteriorates, due to lack of investment. This creates what Smith calls the rent-gap: a gap between

- the potential rent: the ground rent inner-city could generate in perfect conditions,
- the capitalized rent: the actual rent being capitalized in the current degraded state.

If the city continues to expand, the potential attractiveness of the inner-city raises. Indeed, assuming all the residents need to reach the center (as in the Alonso-Muth-Mills model [12, 13, 14]), the larger the city grows, the longer it takes to commute from the suburbs to the center. So, in a large enough city, it may become preferable to some residents to live in the inner-city rather than in the suburbs. This results in a raise in the potential rent of the inner-city. On the other hand, the capitalized rents decrease as long as the housing stock continues to

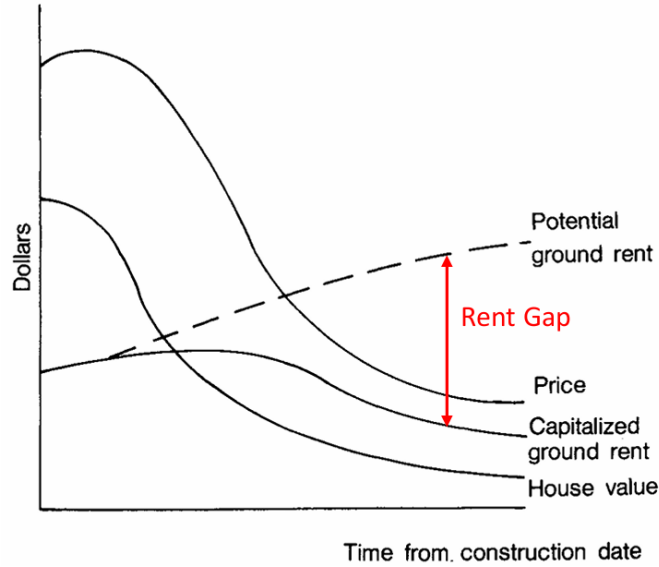


Figure 1.1: The depreciation cycle of inner-city neighborhoods, from [11].

deteriorate. This means that the rent-gap widens over time, as shown in Figure 1.1. A higher rent gap is associated with a greater potential profit in case of investment, since generally a low capitalized rent corresponds to a low purchase price, while a high potential rent implies a high resale price or rental income after renovation. So, when the gap is wide enough, it triggers reinvestment: some developers start buying and renovating properties in the inner-city, because the returns become better than building new houses in the suburbs. As a result, part of the inner-city housing stock gets redeveloped and the area starts attracting higher-income residents. Therefore, this reinvestment process marks the beginning of gentrification (understood as a flow of higher-income residents from the suburbs to the inner-city), that is, according to Smith, a natural consequence of the housing market dynamics.

1.2 Schelling’s model of segregation

Schelling’s model of segregation (developed by economist Thomas Schelling in 1971 [15], [16]) is considered one of the first examples of agent-based models (ABM). It effectively shows how individual behaviors impact the macroscopic scale of a system. In particular Schelling found that a mild preference by individuals for neighbors of their same type is enough to result in well-defined segregation patterns. Despite its simplicity, the model embodies the core concepts of agent-based modeling, illustrating how local rules governing individual agent behavior and interactions can

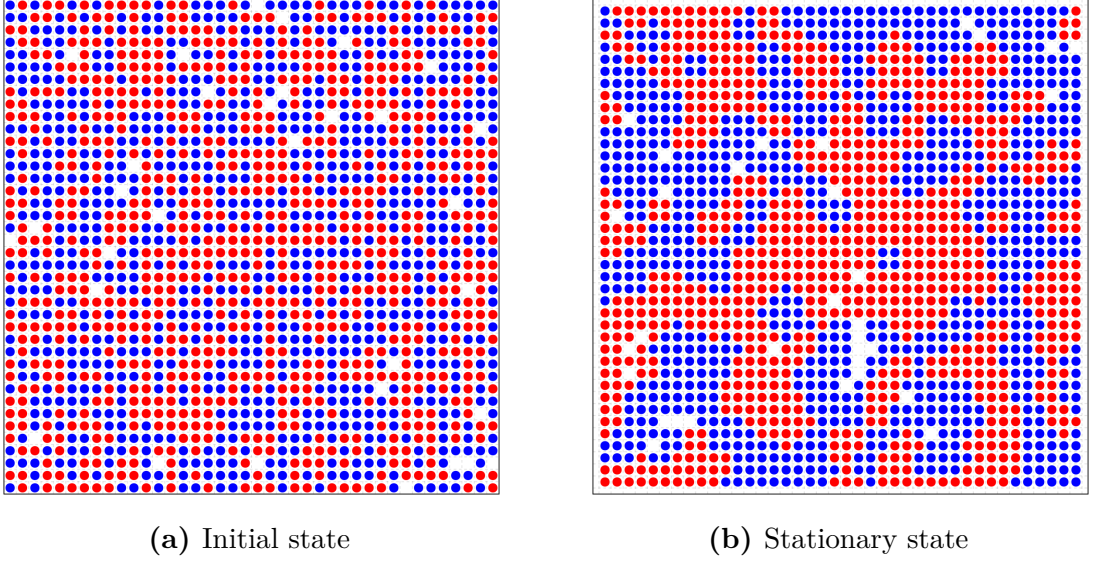


Figure 1.2: Comparison between the initial state of Schelling's model, where agents are randomly distributed on the grid **(a)**, and the stationary, visibly segregated state reached after few steps, where all the agents are happy **(b)**. (Vacancy Density: $\rho_0 = 0.02$, Tolerance Threshold: $f^* = 0.65$)

give rise to emergent, aggregate outcomes. This foundational model has inspired numerous later studies on segregation (see, for example, [17, 18, 19, 20, 21, 22]). Indeed, its adaptability makes it particularly well suited to serve as a starting point for building more complex models, which is what we did in this thesis too.

The setting of the model is very simple: two, equally sized groups of agents are randomly distributed on a grid (the two sets can represent ethnic groups, economic status, or any other form of social classification); agents don't fill all the patches, so there are also some vacancies; the number of vacancies is one of the parameters of the model (vacancy density ρ_0 [23]). An agent can be in two possible states: happy or unhappy. Happy agents are satisfied with their positions and stay still. Unhappy ones, instead, look for vacancies where they would be happy and, if they can find one, they move there. Happiness is described by a binary utility function that depends on the composition of the agent's neighborhood (Moore). Given an agent, f is the fraction of her neighbors belonging to the opposite group (here only the agents are counted as neighbors, but it exists also a version of the model in which vacancies are counted too), f^* is the tolerance threshold (second parameter of the model [24]) and the utility function is simply defined as:

$$U(f) = \begin{cases} 1 & \text{if } f \leq f^* \quad (\text{happy}) \\ 0 & \text{if } f > f^* \quad (\text{unhappy}) \end{cases}$$

which means that an agent is happy as long as the fraction of other color neighbors don't exceed a certain threshold. Once defined these simple rules, different scenarios can be tested by changing the two main parameters of the model: the tolerance threshold (f^*) and the vacancy density (ρ_0). Other tests may be done by changing the shape of the grid, the proportion of agents in the two groups or the definition of neighborhood.

Figure 1.2 shows an example of Schelling's model simulation. A squared, 40×40 grid is populated with the same number of red and blue dots. The 2% of the grid is unoccupied (vacancy density: $\rho_0 = 0.02$), and the tolerance threshold is set at $f^* = 0.65$. The grid on the left is the initial random state, the one on the right is the stable, stationary state the model reaches in few steps: all the agents are happy in this state, so they remain still (this can be interpreted as a Nash equilibrium within a game-theoretic framework). The segregation patterns in the stationary state are clearly visible, as red and blue dots have moved forming large clusters of the same color. A quantitative proof of segregation is given by the neighbor segregation index. Considering the grid as a lattice, each patch is connected by a link to the eight patches in its Moore neighborhood. The neighbor segregation index is the fraction of neighbor links between dots of the same color over the total number of neighbor links (for neighbor link we mean a link that connect two occupied patches):

$$N = \frac{|E_{bb}| + |E_{rr}|}{|E|} \quad (1.1)$$

where $|E_{bb}|$ and $|E_{rr}|$ are the number of edges between two blue and two red dots respectively, and $|E|$ is the total number of neighbor links in the grid. It is a simple and common measure of the exposure between the two groups [15]. Figure 1.3 displays the trend of this index during the dynamics of the model from the initial random state to the final stable configuration. The index is averaged over 20 simulations that share the same values of ρ_0 and f^* , but differ in the initial random setting. This analysis confirms the visual intuition of segregation provided by Figure 1.2: the fraction of connections between matching colors increased from 50% (that is expected for the initial random state) to over 75%, meaning that the exposure between the two groups approximately halved. A similar behavior is common to all the simulations, independently of the initial setting. It is interesting to notice how significant segregation emerges despite the relatively high individual tolerance: indeed, $f^* = 0.65$ means that each agent is comfortable living in a neighborhood where up to 65% of neighbors are of the opposite color. This effect is somehow unexpected and it is the most significant result of the Schelling's model.

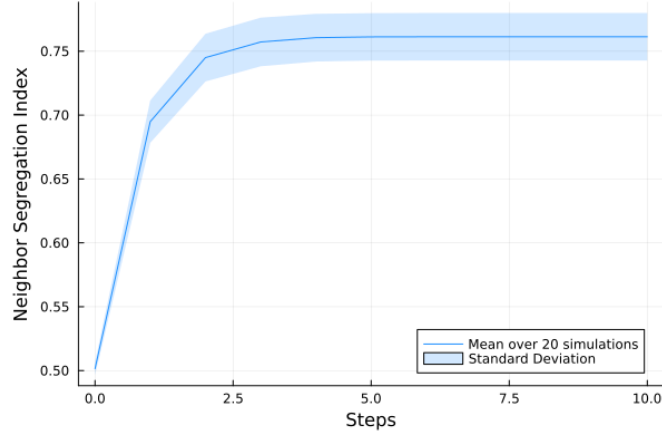


Figure 1.3: Trend of the neighbor segregation index (the fraction of neighbor links connecting two agents of the same color) averaged over 20 simulations that differ in the initial random configuration. The index significantly increases over the dynamics, meaning that the system evolves towards a more segregated state. (Vacancy Density: $\rho_0 = 0.02$, Tolerance Threshold: $f^* = 0.65$)

1.3 Models of gentrification

In recent years, as the debate on socio-economic aspects of gentrification intensified (see Section 1.1), a few attempts have been made to provide some models of the process, trying to tackle it from different perspectives, both analytically and computationally.

Brueckner and Rosenthal [25], starting from the famous Alonso-Muth-Mills monocentric city model [12, 13, 14] and inspired by Smith’s theory [11] (discussed in Section 1.1.1), developed an analytical model of gentrification in US cities. They considered, as Smith, the age of the housing stock as the crucial driver of gentrification, assuming that affluent inhabitants have both the preference and the purchasing power to live in newer dwellings. Indeed, in their model they explain the gentrifying flows of high-income residents as a consequence of the periodic cycle of the housing stock age: initially, assuming the city expanded over time from a single core, housing age decreases linearly with distance from the center, so wealthier residents are found in the suburbs and lower-income ones in the inner-city. However as the city continues to grow, in addition to the construction of new buildings at the periphery, older houses are also renewed. This process transforms the housing age function from a linear pattern into a sawtooth wave (as shown in Figure 1.4), and the income distribution will have a similar, complementary behavior. Therefore, according to this model, gentrification is a natural consequence

of the housing market cyclical dynamics. The city, as it expands, will result in a series of concentric gentrified zones.

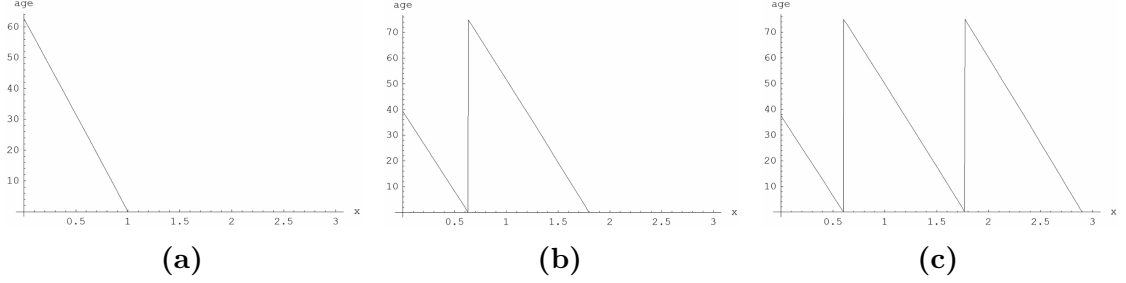


Figure 1.4: Housing age as a function of the distance from the center (x), from [25]. Comparison between three phases of the city expansion: **(a)** first phase, redevelopment has not occurred yet; **(b)** second phase, first round of redevelopment started; **(c)** third phase, second round of redevelopment started.

An interesting epidemic approach to gentrification was adopted by Quan [26], who treated old buildings as susceptible states and new (or renovated) buildings as infected states in a SI epidemic model. He used his model to study how rapidly the "gentrification virus" spreads in different areas of Seoul, concluding that the more land use is allowed in an area, the faster gentrification spreads.

Rossi Mori, Loreto, and Di Clemente recently published a study [27] introducing an interesting and innovative methodology to analyze the spatio-temporal dynamics of income segregation in the city of Milan, using mobile phone data. Although their work focuses on a specific ten-month period, their approach could be extended to longer timeframes, potentially revealing patterns of gentrification. Apart from these models and a few other examples [28, 29], the most common choice to model gentrification in the literature is by far through Schelling-like agent-based models [30, 31, 32, 33, 34, 35].

Ortega, Rodriguez-Laguna and Korutcheva in [30] tried to tackle the process from a statistical physics perspective: they developed a Schelling-like model, in which the dynamics is driven by minimizing a particular version of the well-known Blume–Emery–Griffith Hamiltonian [36]. The three possible spins values can represent: a vacancy ($s_i = 0$) or an agent belonging two groups ($s_i = \pm 1$). Between the two groups there is an economic gap. The model shows how initial economic inequalities between two classes may lead to avalanches of displacement of the lower-income population, or to the formation of segregated ghettos.

Another, original agent-based model of gentrification was recently built by Mauro and Pappalardo [31]. They sketched the city as a 7×7 grid, each patch represents a neighborhood, and can host up to a maximum number of agents. The grid is populated with 2^{12} agents who are associated with fixed income, sampled

from real-world data. The population is split into three classes: L, M and H (low-, medium- and high-income), each of them covers a different share of the total population. The agents are initially spread on the grid such that the spatial distribution of incomes displays a radial gradient, as shown in Figure 1.5a. Notice that in this case the division is opposite with respect to the US cities treated by both Smith and Brueckner [11, 25]. Indeed, here the European city prototype is considered, in which traditionally wealthier residents occupy the center, while lower-income ones the periphery. At each step agents have a certain probability to change neighborhood. The innovative part of the model is that this probability is simply a function of agent's relative income percentile within her cell. The function differs from one class to another. L agents have a probability density function (PDF) for moving out of a cell that is monotonically decreasing with their relative percentile (Figure 1.5b). Conversely, the PDF of moving into a new cell is monotonically increasing with the relative percentile they would occupy in that cell. In other words, low-income agents are assumed to prefer being in the higher income tier within their neighborhood. M agents' PDFs instead are symmetric, centered at $1/2$. The one for moving is convex (Figure 1.5c), while that for moving in is concave. This means that, in this model, medium-income residents prefer to occupy a central position in the income distribution of their neighborhood. Things change for H agents: they have a fixed and constant moving out probability, but they select the new neighborhood looking at how the median income changed in that cell over the last steps. The faster the median increased in a cell the higher is the probability that H agents will move there, further increasing the income of that neighborhood. The main result of this model is that a small fraction (5%) of high income residents, who move according to the logic just explained, are able to induce significant gentrification, even if their moving out probability is set at a low value (0.01).

Bagheri-Jebelli, Crooks and Kennedy [32] developed an agent-based model of gentrification in which the process is not triggered by the residents themselves (as in the two previous cases). Instead, there is another type of agent, not physically on the grid, who looks for old houses and renew them in order to make profit. This idea is borrowed from their study and will also be used in the model presented in Chapter 2.

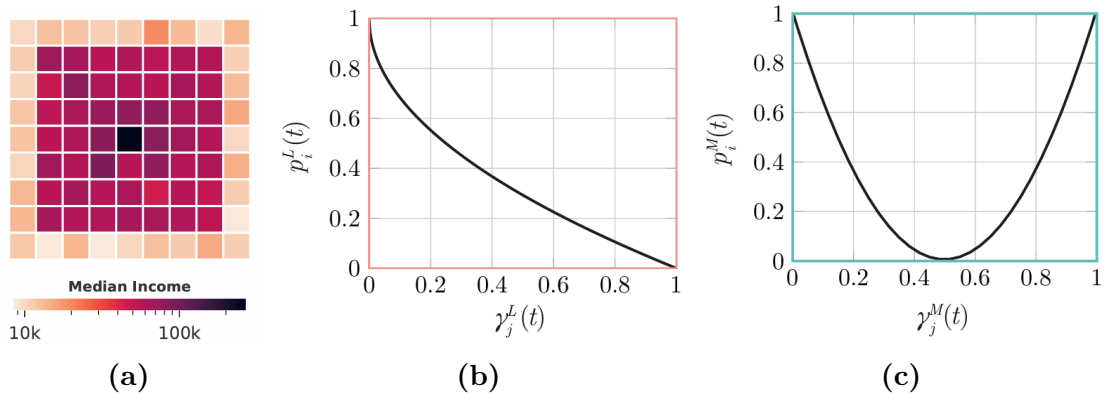


Figure 1.5: Plots from Mauro and Pappalardo's model [31]. **(a)** Initial spatial distribution of agents in Mauro and Pappalardo agent-based model. High-income agents predominantly occupy the city center, medium-income the inner areas, and low-income the periphery. **(b,c)** Moving out probability as a function of the agent's relative income percentile, respectively of low and medium-income classes.

Chapter 2

Our Modeling Scheme

In this chapter we present the agent-based model we developed to explore the previously discussed issues and controversies around gentrification (see the final part of Section 1.1). The model provides a simplified representation of a city. We deliberately omitted several real-world aspects, such as urban expansion, migration flows, and the housing sales market. We focus exclusively on the rental market. Including both rental and sales dynamics would have been too complex, and we consider the rental sector to be more suitable for representation through a dynamic agent-based model.

Furthermore, while we based income and rent values on real-world data, we did not for housing ages. We defined them artificially, with the sole purpose of producing rent values that are consistent with realistic expectations, as we will explain in the next paragraphs.

Before focusing on the main model, we briefly discuss a preliminary and simplified version developed during the early stages of this project. This initial model served as a preparatory step toward the construction of the more sophisticated model that we will discuss later in this chapter.

2.1 A first toy agent-based model of gentrification

The city is initialized as an exaggerated version of the European prototype (also adopted by Mauro and Pappalardo [31]). High-income residents are concentrated in the city center, while low-income residents are located in the peripheral areas. The city is composed of five square quadrants arranged in a cross shape, as shown in Figure 2.1. Each quadrant is a 10×10 grid, and agents can occupy any spot on the grid. The city center consists solely of the central quadrant, while the remaining four quadrants make up the periphery. Initially, the four corner quadrants were

also included as part of the periphery and filled with low-income agents, forming a square-shaped city. However, in that configuration, the number of low-income agents was approximately eight times greater than that of high-income agents, which was not optimal for the dynamics we aimed to study. We could simply decrease the density of low-income agents in the periphery, but that would have changed the dynamics, since agents' utility function depends also on their neighborhood density. For this reason, we switched to the cross layout. However, the shape of the city is not particularly important for the process we aim to simulate.

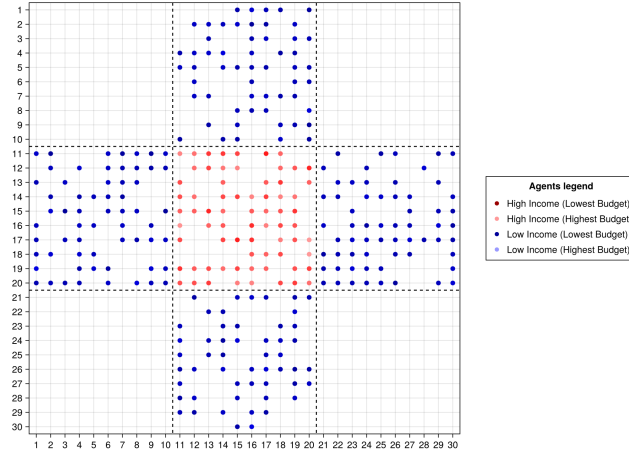


Figure 2.1: Initial state of the city. High-income agents (red dots) are spread across the central quadrant, low-income ones (blue dots) are distributed in the four suburbs. The four corner quadrants are not part of the city. Imagine them as mountains.

Agents in the model are divided into two groups—low- and high-income—and each is assigned a fixed income value, drawn from a Gaussian distribution corresponding to their group, for simplicity. A more accurate approach would have been to draw incomes from a single Pareto distribution instead of two separate Gaussians; however, in this first model, we did not aim to provide a realistic representation of wages. In the more complex model we will introduce later, we sample incomes directly from an empirical distribution. High-income agents are initially distributed randomly within the central quadrant. The initial density ρ (defined as the number of agents over the number of spots) of this quadrant is a fixed parameter provided as input to the model. In the four peripheral quadrants, by contrast, the average density is fixed, but the actual number of agents is drawn from a Gaussian distribution centered around this average. Here too, once the initial density is defined, agents are uniformly distributed across the quadrant. High- and low-income agents are represented by red and blue dots, respectively. The

color shade of each dot corresponds to the agent's income: lighter shades represent individuals in the upper tail of their group's distribution. Figure 2.1 shows how agents are initially distributed across the grid in a simulation, while the right plot in Figure 2.2 displays the resulting densities in the occupied quadrants in the same simulation. The corner quadrants appear black in the plot and are considered prohibited areas. In addition to agent density, each quadrant is also associated with two further attributes: beauty b and cost of living c . The initial beauty level of the central quadrant is set to the maximum value, $b = 1$. In the peripheral areas, by contrast, the beauty level is drawn from a Gaussian distribution centered at 0.4. The cost of living depends on the presence of high-income agents. Initially, it is set to the maximum value in the center ($c = 0.62$) and to the minimum value in all peripheral quadrants ($c = 0.3$). The initial values of beauty and cost in each quadrant are shown in the left and central plots of Figure 2.2 for the considered simulation.

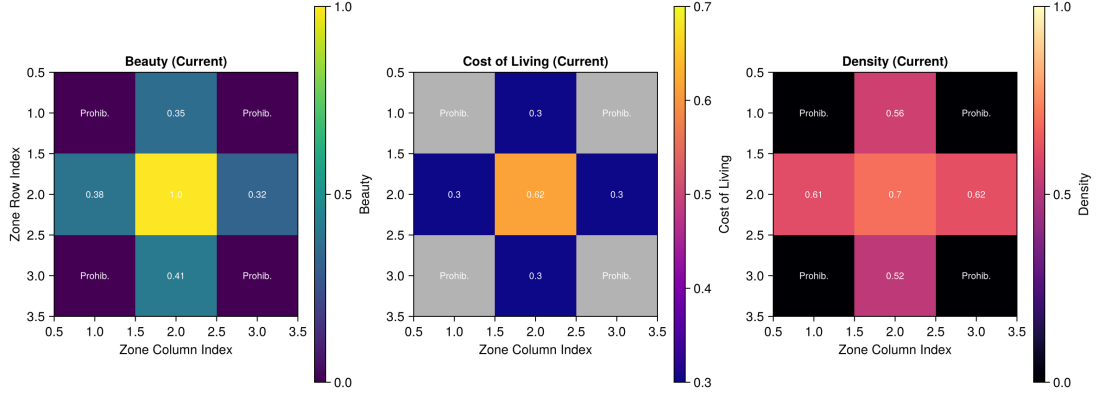


Figure 2.2: Initial quadrant attributes in a specific simulation. **Beauty:** It is initially set to be high in the center and low in the suburbs; **Cost of living:** It is determined by the fraction of high income agents present; **Density:** It is initially fixed in the center and sampled from a Gaussian distribution in the suburbs.

After the model is initialized, the dynamical part can start. Each agent is associated with a utility function that depends on the quadrant they are occupying and on their income. The function is a linear combination of three terms:

$$U = \begin{cases} \alpha \cdot U_{\text{beauty}} + \beta \cdot U_{\text{density}} + \gamma \cdot U_{\text{income}} & \text{if } U_{\text{income}} > 0 \\ 0 & \text{if } U_{\text{income}} \leq 0 \end{cases} \quad (2.1)$$

U_{beauty} is simply the beauty level b of the quadrant the agent is occupying.

U_{density} is defined as a quadratic function of quadrant density ρ :

$$U_{\text{density}} = 4\rho(1 - \rho)$$

centered at $\rho_c = 0.5$. The idea is that agents prefer to live in a neighborhood that is half-filled—neither too crowded nor too empty. U_{income} is a function of the agent's income i and the quadrant cost c :

$$U_{\text{income}} = \begin{cases} \sqrt{i - c} & \text{if } i > c \\ 0 & \text{if } i \leq c \end{cases}$$

Both U_{income} and the total utility U (Eq. 2.1) are set to 0 if the agent cannot afford the quadrant they are in. All three contributions to the utility function take values between 0 and 1, and the weights make the three terms comparable. During the dynamics, at each model step, high-income agents evaluate their satisfaction (the value of the utility function) in the quadrant they are currently occupying, and the potential satisfaction they would have in the other areas. The potential satisfaction is computed as in Eq. 2.1, minus a constant displacement cost. Then, they sample the quadrant to move to from a Boltzmann-like distribution. The probability of selecting quadrant j is given by:

$$P_j = \frac{e^{\frac{U_j}{T}}}{\sum_i e^{\frac{U_i}{T}}} \quad (2.2)$$

where U_j is the value of the potential utility function in quadrant j , the sum in the denominator is over all the quadrants in the city, and T is a temperature-like noise parameter. In the limit $T \rightarrow 0$, the process becomes deterministic and the quadrant with the highest utility is selected with probability 1. In our case, T was set to 0.12. In this way, agents have a good chance of selecting the best quadrant according to their utility function (which may be the one they are already in), but they are not completely certain. Low-income agents follow the same procedure, but only if their satisfaction is below a certain threshold. In that case, they select a new area to move to using the Boltzmann distribution restricted to the quadrants they can afford.

At the beginning of the dynamics, we ran the model for a few steps in which agents are free to move according to the logic described above, and quadrant densities and costs are updated at each step. The cost of living in quadrant i is updated according to:

$$c_i = c_{\text{max}} + (1 - \rho_i^h)(c_{\text{min}} - c_{\text{max}}) \quad (2.3)$$

where c_{max} and c_{min} are constant upper and lower bounds, and ρ_i^h is the fraction of spots in quadrant i occupied by high-income agents. In this model, it is assumed

that the presence of high-income residents is the main driver of the cost of living in an area. During this phase (which we refer to as "thermalization", borrowing the term from physics), agents can move according to their preferences, and a stable configuration is reached after a few steps. This stable state, displayed in Figure 2.3a, is very similar to the initialized one (Figure 2.1). The only difference is that a few high-income agents have moved out of the center. In contrast, no low-income residents have moved into the center, since they usually cannot afford it.

After thermalization, the gentrification dynamics begin. As in [32], a third type of agent is introduced to trigger gentrification: the developer. This agent represents private investors who redevelop peripheral areas to attract high-income residents and generate profit. In this model, redevelopment is simply represented by an increase in the beauty level. In the model we present later in this chapter we will simulate this process in a more realistic way. The developer selects one peripheral quadrant in which to invest. The chosen quadrant k is the one maximizing the score function:

$$k = \arg \max_{i \in \text{periphery}} \left\{ b_i - \frac{\rho_i}{2} \right\} \quad (2.4)$$

where b_i is the beauty level of quadrant i and ρ_i its density. This means that the developer considers most profitable those quadrants that combine low density with high beauty, since these are more likely to attract high-income agents. Once selected, the developer increases the beauty of the chosen quadrant by a fixed amount. Then, as in the thermalization phase, all agents evaluate whether to move

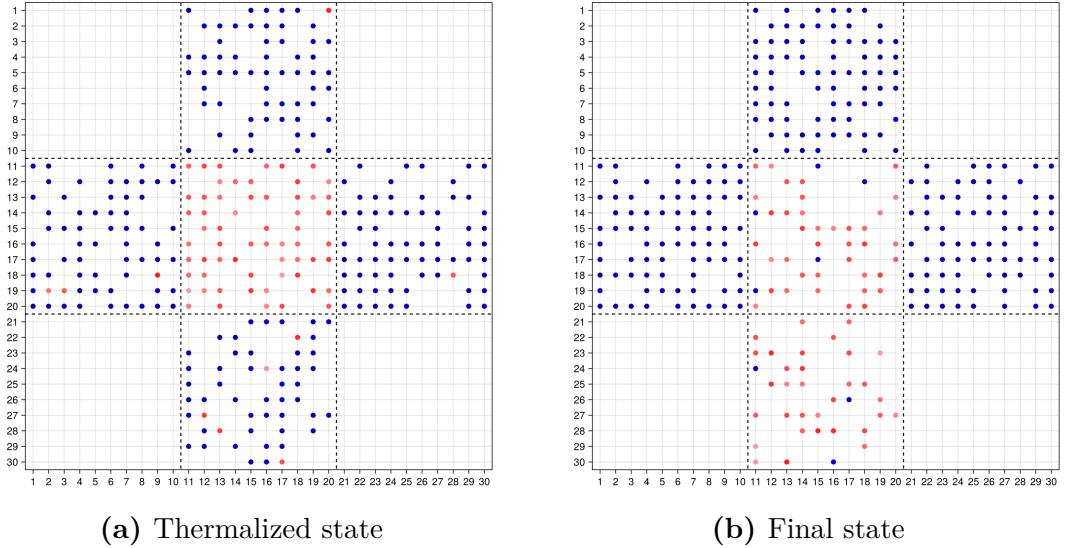


Figure 2.3: Comparison between the thermalized state (a) and the final state (b) reached after the lower quadrant has undergone gentrification.

or stay, following the same utility-based logic. Simultaneously, the cost of living in each quadrant is updated at every step according to Eq. 2.3. This process is repeated over a few model steps. Since the developer increased the beauty of the selected quadrant, its score (Eq. 2.4) is likely to increase too, leading to repeated selection of the same area. The developer's activity ends when the beauty of the gentrified quadrant reaches the maximum value, $b = 1$. As the attractiveness of the area grows, some high-income agents are drawn to it, raising its cost of living. This, in turn, forces the displacement of some low-income agents who previously lived there. At the same time, the cost of living in the central area decreases as high-income agents leave.

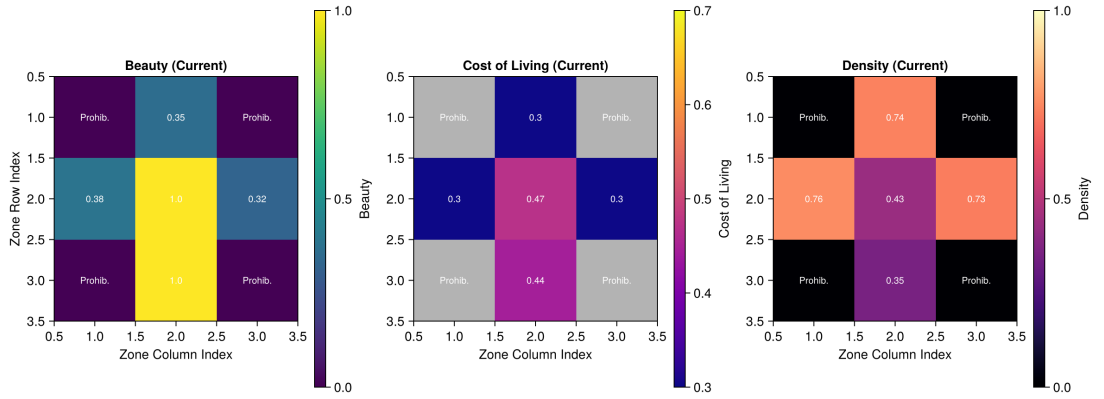


Figure 2.4: Final quadrant properties. The central and the peripheral quadrants are now quite balanced in each property, while in the other peripheral areas the density increased.

The stable configuration reached after 10 model steps is shown in Figure 2.3b, while the resulting quadrant attributes are displayed in Figure 2.4. It is evident that the simulated gentrification dynamics result in the segregation of low-income agents into just three quadrants (initially, they were spread across four), forcing them to live in denser areas. In the final state (Figure 2.3b), very few blue dots are found in the central and gentrified quadrants. High-income agents, instead, are almost equally split between the two neighborhoods. In this way, they benefit from a lower cost of living compared to what they initially had in the central quadrant, while still residing in areas with the highest possible beauty level.

According to this model, high-income residents appear to be the only ones who benefit from the process of gentrification, as they have the opportunity to choose between two attractive and affordable areas. Low-income agents, on the

other hand, cannot benefit from the improvements made in their neighborhoods, as the increase in the cost of living forces their displacement. The simulation was repeated using different input random seeds. The seed is the number that determines all the random processes simulated during a model run. Figure 2.5 shows the aggregate trends, over 20 simulations, of the average satisfaction for both groups. These results confirm the earlier observation: in this model, gentrification dynamics improve the conditions of high-income agents while worsening those of low-income agents.

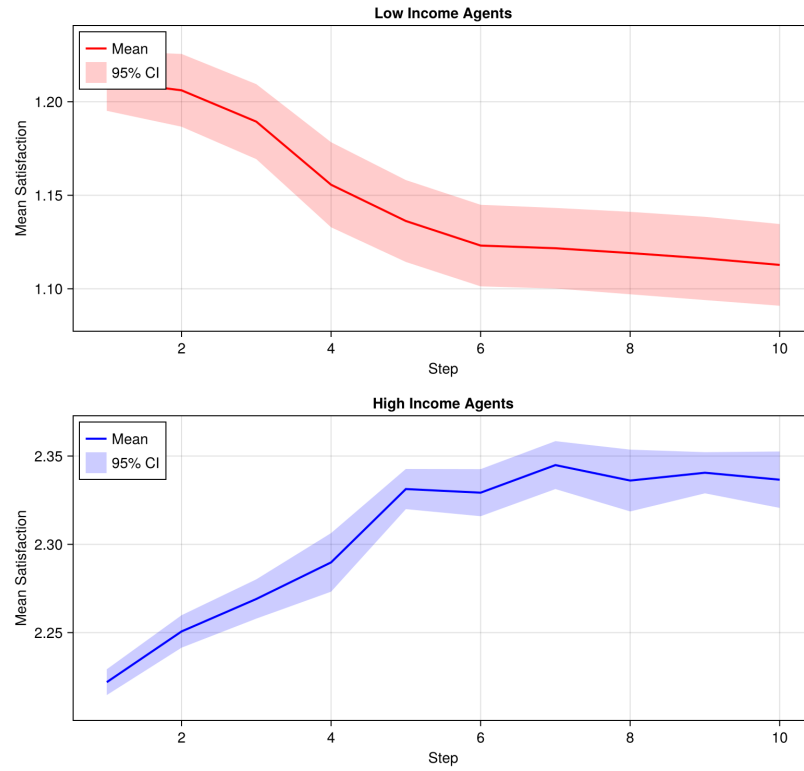


Figure 2.5: Trends of the average satisfaction of the two groups during the gentrification dynamics. The trend is aggregate over 20 different simulations that differ in the initial random setting.

In the rest of this chapter, we discuss the main ABM. As it is significantly more complex and detailed than the previous one, we divide the discussion into several sections and subsections.

2.2 City Environment

The environment of the model is a squared, 40×40 grid, as in Schelling's (Section 1.2). However, since the grid is intended to represent a real city, the patches are not all identical: they can be of three different types. The way some of these patches are distributed across the grid is subject to pseudo-random processes, which are all determined by the seed provided as input to the model.

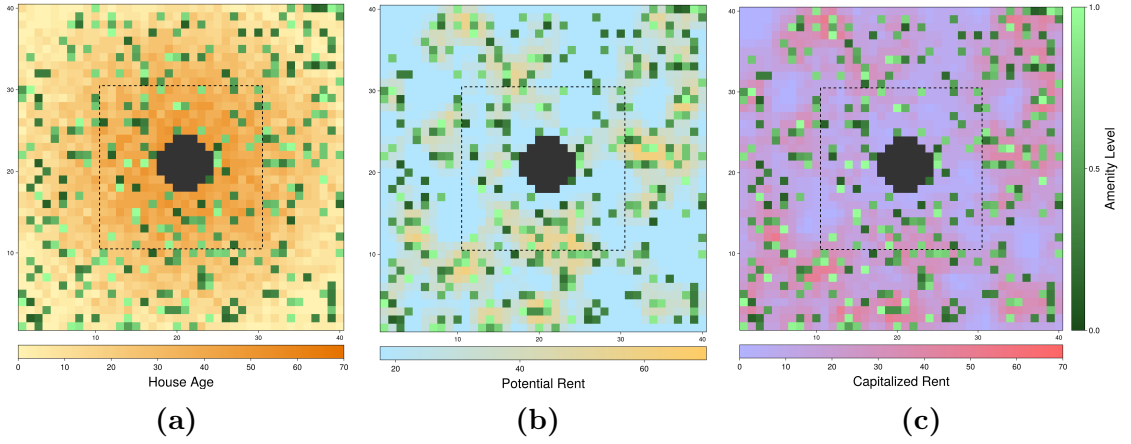


Figure 2.6: Plot of the model's environment. The three grids represent the same sketched city. The green patches (representing the amenities) and the black ones (representing the CBD) are the same in the three plots, what changes is the property of the housing patches displayed. **(a) Age:** The shade of orange represent the age of the dwellings, which, on average, decreases with the distance from the center. **(b) Potential rent:** The potential rent is higher for houses that are surrounded by a greater number of amenities. **(c) Capitalized rent:** The capitalized rent is a combination of the potential rent and the age of a house. This environment results in a city with cheaper houses in the inner-city, and some expensive hotspots in the suburbs.

At the center of the grid there is a region of special patches that represent the Central Business District (CBD) of the city, they are colored black in the plots (Figure 2.6). Agents cannot occupy those patches, which means that residents don't live there. They represent a part of the city every agent needs to reach (for example the working place, as assumed in the Alonso-Muth-Mills model [12, 13, 14]), and so the distance from this region will be considered in the agents' utility function. A part from the CBD, only two other macro-areas of the city are defined: the inner-city and the suburbs. They are divided by simply separating the central quarter of the grid from the rest (dashed line in Figure 2.6). The second type of

patches are the amenities, colored green. They represent all non-residential areas of a city that can contribute to making a neighborhood more or less attractive, such as parks, squares, monuments, and public spaces in general. We assigned to each of these patches an amenity level, represented by the shade of green, which is just a value between 0 and 1 and determines how attractive an amenity is for the agents (light shade of green means the value is close to 1 and so the attractiveness is high). While the CBD patches are precisely selected by taking the central part of the grid and remain the same across different simulations, the amenities are distributed randomly: the fraction of patches that must be amenities is fixed (with different values for inner-city and suburban areas), but their specific distribution on the grid depends on the random seed used for each simulation. The random seed is simply a number that determines and makes reproducible all the random draws performed by the program. Besides the spatial distribution, the single amenity level of each green patch is also random, and so it is determined by the seed. Thus, changing the input seed we can simulate the model on different initial environments.

All the remaining patches are houses. Houses have three different properties (shown in the three plots of the same simulation in Figure 2.6) that are necessary to model the rent-gap mechanism explained in Section 1.1.1. The first property is the dwelling age. It is common to assume that a city expands over time from a single central core, and that buildings near the center are generally older than those in the outer regions. To replicate this scenario in the model, the average age of housing patches was defined as a decreasing, linear function of the distance from the center. In this way, houses next to the CBD have, on average, the maximum age (a parameter of the model), while those at the edge have an average age of zero. Then, to add realistic fluctuations, the effective age of each patch was extracted from a Gaussian distribution, centered on its corresponding average value (a clamp function ensures that the final values do not fall below 0, or above the parameter maximum age $a_{\max} = 50$). The result of this process is displayed in Figure 2.6a: the age value is represented by the shade of orange (darker means older), and it is visible that, with some fluctuations, age decreases moving away from the CBD. The second property of housing patches is the potential rent, which represents the rent a house would have if it were new (see Section 1.1.1). Two features contribute to this value: the proximity to the center (once again), and especially the level of amenity around the patch. The idea is that a house is more valuable if it is surrounded by attractive amenities and if it is closer to the center, since, as already mentioned, agents are supposed to commute to the CBD. The potential rent is defined as follows: we determined a base value for each house, similarly to the average age, defining a decreasing function of the distance from the center. The steepness of this function is low, which is why this contribution is only slightly visible in Figure 2.6b. Then, for a specific housing patch, a contribution is added to the base value. This contribution is based on the sum of the amenity levels of

the green patches present around the house - we will refer to it as the amenity score of a housing patch. The potential rent of house i is so defined as:

$$PR_i = PR_i^{\text{base}} + \alpha \left(\sum_{j \in \text{Moore}_1(i)} A_j + \sum_{k \in \text{Moore}_2(i)} \frac{A_k}{2} \right) \quad (2.5)$$

where PR_i^{base} is the base value, α is a constant factor multiplied to the amenity score, A_j is the amenity level of the green patch j , and $\text{Moore}_r(i)$ is the set of patches in i 's Moore neighborhood of radius r . If j is a second neighbor of i , its amenity is divided by 2 to account for the fact that it contributes less to the value of the house, since it is farther away. From Figure 2.6b it is evident that, according to this definition, the houses with the highest potential rent are those surrounded by more, and better amenities. The last property of housing patches is a combination of the previous two: the capitalized rent. According to Smith's theory, the capitalized rent is the effective rent associated with a dwelling, given its state of deterioration. In this model it is defined as the potential rent decreased proportionally to the house age, plus a small zero-mean Gaussian noise term ($\sigma = 0.05$) representing the maintenance effect for each specific house. This noise can slightly improve or worsen the deterioration caused by aging:

$$CR_i = PR_i \left(1 - \beta \frac{a_i}{a_{\max}} + \eta_i \right) \quad (2.6)$$

where CR_i is the capitalized rent of house i , $\beta = 2/3$ is a constant factor, a_i is the age of the house and η_i is the Gaussian noise term associated with maintenance for house i . Here too, a clamp function is used to keep the capitalized rent value within the bounds of a positive threshold and the potential rent. The constant parameters present in these functions have been defined to make the resulting capitalized rent distribution qualitatively consistent with an empirical one, as explained in Section 2.4.1 and Figure 2.12.

The rent-gap of a housing patch is simply defined as:

$$RG_i = PR_i - CR_i = PR_i \left(\beta \frac{a_i}{a_{\max}} - \eta_i \right) \quad (2.7)$$

Figure 2.7 shows how the capitalized rent and the rent-gap change as functions of house age. The trends for two different values of potential rent are displayed. The confidence bands represent the expected variability introduced by the Gaussian maintenance noise η , which causes slight shifts in the curve. Note that, due to the definition of capitalized rent (Eq. 2.6), dwellings with higher potential rent are associated with a larger rent gap at the same age.

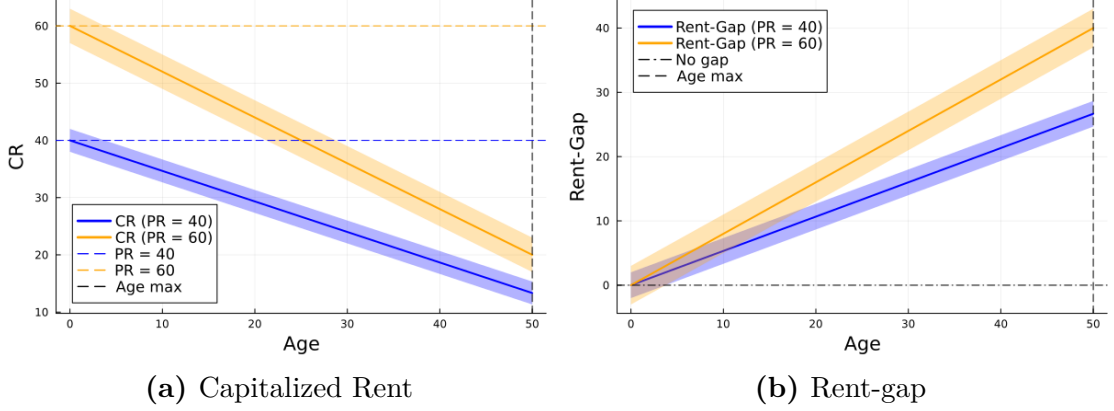


Figure 2.7: (a) Capitalized rent as a function of the age for two values of the potential rent. (b) Rent-gap as a function of the age. At the same age, a higher potential rent is associated with a higher rent-gap. In both cases the confidence bands represent the variability due to the Gaussian noise.

2.3 Agents

The grid is populated by agents representing households. Each agent is associated with a fixed annual income. Incomes are sampled from an empirical distribution based on national U.S. data from the 2013 American Community Survey (ACS), a large-scale, annual survey conducted by the U.S. Census Bureau. The distribution is shown in Figure 2.8. Note that the income brackets on the x-axis, although displayed with equal width, actually represent intervals of varying sizes. This reflects the structure of the original data source. The agents are divided into two groups — low-income and high-income — by splitting the income distribution at the 60th percentile (Figure 2.9). Therefore, 60% of the agents on the grid belong to the low-income class. Initially, agents are randomly placed on the grid, as shown in Figure 2.10a. Again, the process is pseudo-random and determined by the seed. The total number of agents is set such that the unoccupied fraction of houses is around 7%, again coherently with the rental vacancy rate from the 2013 ACS. We represented on the grid low-income residents as blue dots and high-income as red dots (Figure 2.10).

Each agent is also associated with a satisfaction level, determined by a utility function, which is given by a weighted sum of four components:

$$U = (a \cdot U_{\text{rent}} + b \cdot U_{\text{amenities}} + c \cdot U_{\text{age}} + d \cdot U_{\text{center}}) \cdot H(U_{\text{rent}}) \quad (2.8)$$

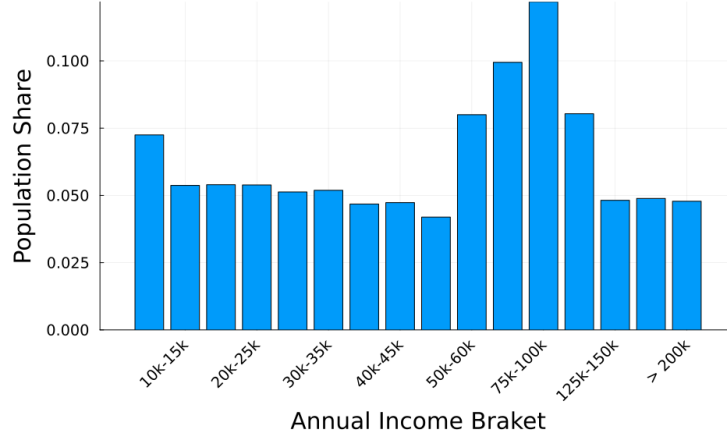


Figure 2.8: Empirical distribution of annual household incomes, based on data from the 2013 American Community Survey (ACS). The distribution was used to assign income levels to agents in the model. Note that the income brackets in the source data have non-uniform sizes.

where $H(x)$ is the Heaviside function, defined as:

$$H(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

The weights a, b, c, d differ between low- and high-income agents, but in both cases they sum to 1. U_{rent} is the normalized difference between the annual income and the annual rent of the occupied patch. If the income is smaller than the rent, U_{rent} is set to 0, and the Heaviside function drops the total utility function to 0 (as in the model we presented in Section 2.1). This means if an agent cannot afford a dwelling, then she is completely unhappy. $U_{\text{amenities}}$ accounts for the level of amenities surrounding the patch; it corresponds to the normalized version of the amenity score added to the base potential rent (see Equation 2.5). $U_{\text{age}} = \frac{a_{\text{max}} - a}{a_{\text{max}}}$ reflects the agents' preference for newer houses. Finally, $U_{\text{center}} = \frac{d_{\text{max}} - d}{d_{\text{max}}}$ captures the preference for proximity to the CBD (d is the Euclidean distance between the patch and the center of the grid). The total utility function is therefore a convex combination of four normalized utility functions, and will thus take a value between 0 and 1.

2.4 Dynamics

Similarly to the toy model we discussed in Section 2.1, also here the model dynamics consists of two parts: a first thermalization phase, in which the agents move to

reach a stable configuration, and a second phase, during which the gentrification process takes place.

2.4.1 Thermalization phase

After the agents are randomly spread on the grid, they start moving to improve their satisfaction level. The individual decision-making rule is inspired by Schelling's model (Section 1.2): the two groups are assigned a "happiness threshold" (with different values for low- and high-income agents). If an agent's utility function, computed for the patch she occupies, returns a value above that threshold, she is considered happy and does not move. Otherwise, if her satisfaction is below the threshold, she looks for a better housing patch: she randomly selects a fixed number of vacant houses and moves to the one with the highest utility function value among them, provided that the new satisfaction is strictly larger than her current one. In this phase, a single step of the model is defined as the process by which all the agents check their satisfaction and, if necessary, move. After few model steps the system reaches a stable configuration, similarly to what happened in the dynamics of the Schelling's model displayed in Figure 1.2. However, while in that example the dynamics stopped because all the agents had switched to the happy state, in this model that condition is never reached. Still, the dynamics stops because unhappy agents (those whose utility function is below the threshold) cannot find any better patch to move to.

In Figure 2.10 it is shown the comparison between the initial random state

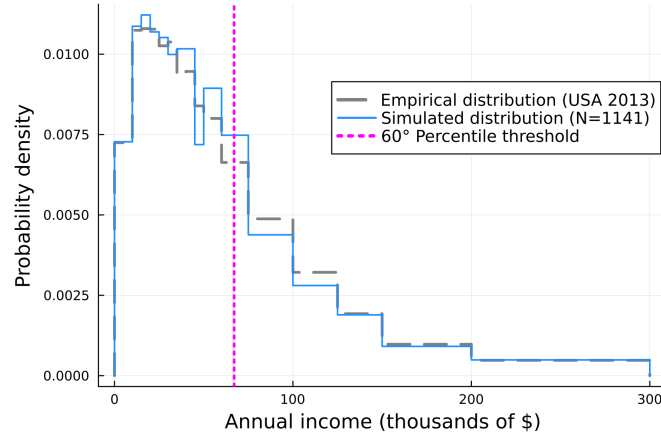


Figure 2.9: Income distributions. The gray dashed line shows the empirical data, while the blue line shows the sampled distribution from a single simulation run. The pink, dotted line represents the percentile threshold dividing the two agent classes.

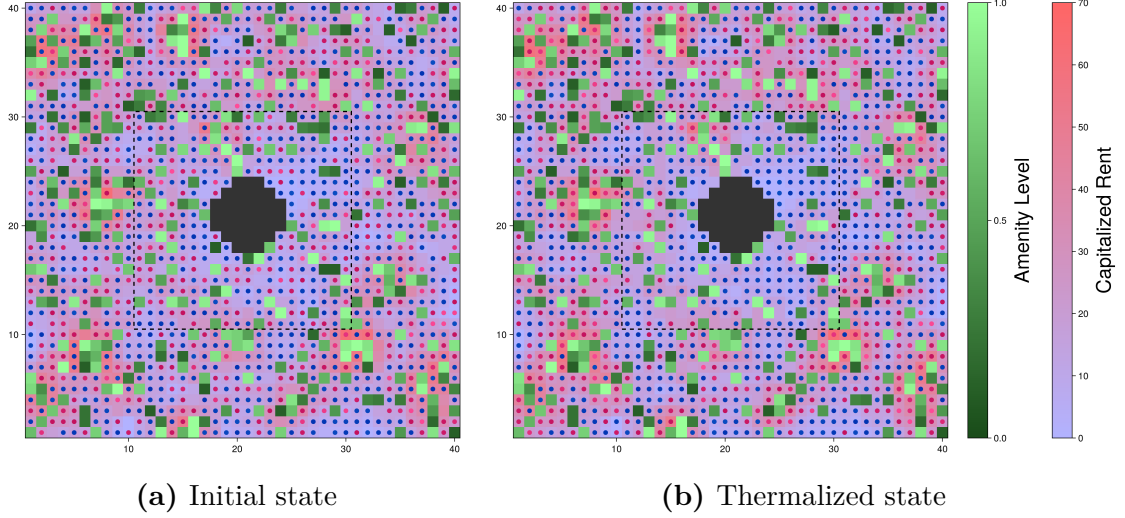


Figure 2.10: Comparison between the initial state in which agents are randomly distributed on the grid **(a)**, and the stable state reached after few steps of thermalization **(b)**. Low-income agents are represented by the blue dots and high-income by the red dots. In the grid background it is displayed the capitalized rent of the housing patches.

and the configuration reached after thermalization. Although this is just a single example, the way the city changes during this phase is qualitatively similar in all the simulations. In the thermalized state it is evident the correlation between the capitalized rent of the housing patches and the agents' group, with low-income agents occupying the cheaper dwellings (purple end of the colorbar) and high-income the more expensive (red end of the colorbar). The inner-city is mostly populated by low-income agents (blue dots), while the areas richer in amenities in the suburbs are filled with high-income residents (red dots). Thus, the defined environment and utility function are effective in making the system thermalize towards a state similar to the prototype of U.S. city introduced in Chapter 1, in which wealthier residents occupy the newer suburban areas of the city, while lower-income inhabitants live in the degraded inner-city. Here, since the low-income group makes up 60% of the total population, and the suburbs are approximately three times larger than the inner city, a significant fraction of blue dots is found in the suburbs as well. To have a more quantitative view on how the contact between the two classes changes during the thermalization phase, we can look at the neighbor segregation index, as we did in Section 1.2. This index represents the fraction of neighbor links connecting agents of the same type (or color) out of the total number of neighbor links (see Eq. 1.1). Its evolution during Schelling's dynamics is shown in Figure 1.3, while Figure 2.11 displays the trend of this index, averaged over 20 different

model runs, during the thermalization phase of the model. It takes only a few model steps for the index to stabilize; and here, as in Schelling’s model, it increases, indicating that exposure between the two groups decreases during thermalization.

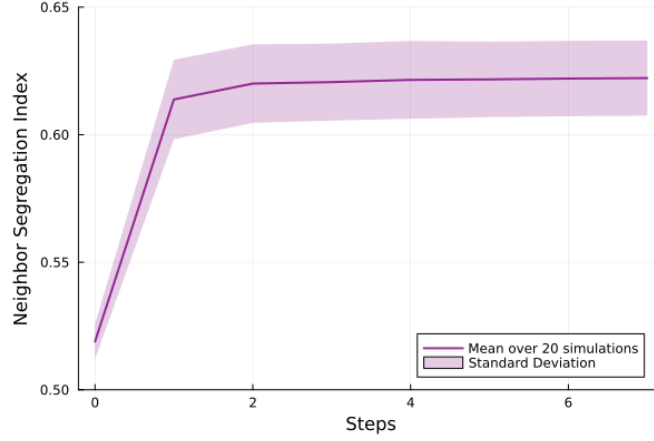


Figure 2.11: Trend of the neighbor segregation index averaged over 20 simulations of the thermalization phase. The index increases over the dynamics, meaning that the system evolves towards a more segregated state.

While income values were sampled from an empirical distribution, capitalized rents are determined by the mechanisms described in Section 2.2, and therefore cannot be directly sampled from real-world data. However, as shown in Equations 2.5 and 2.6, several free parameters influence the final rent values. These parameters can be tuned to make the distribution as close as possible to an empirical one, which is precisely what has been done here. The data source for the empirical distribution was again the 2013 ACS. The available data did not report absolute gross rents, but rather rents expressed as a percentage of agent incomes. For this reason, they were compared with the results of model simulations after the thermalization phase, as this configuration is considered a plausible representation of real cities. Figure 2.12 shows the comparison between the empirical distribution and the one obtained by aggregating data from 20 different simulations of the model. The two distributions are relatively close.

It is important to stress that the analogy with Schelling’s model exists only at the implementation level: in both cases, a threshold defines the boundary between happy and unhappy states, and only the unhappy agents are allowed to move. What determines the utility function value, however, is completely different in the two models: here, the agents do not take into account the composition of the neighborhood in terms of resident groups. In this thesis, we do not study how segregation emerges from social preferences, but rather how it is affected by

economic inequalities within a city. Furthermore, although the dynamics of the thermalization phase is somewhat similar to Schelling’s — in both cases, the system evolves toward a more segregated configuration — the main focus of this thesis is on how segregation changes during the gentrification process. In this light, the thermalized, segregated system is treated as the starting point, rather than the final outcome.

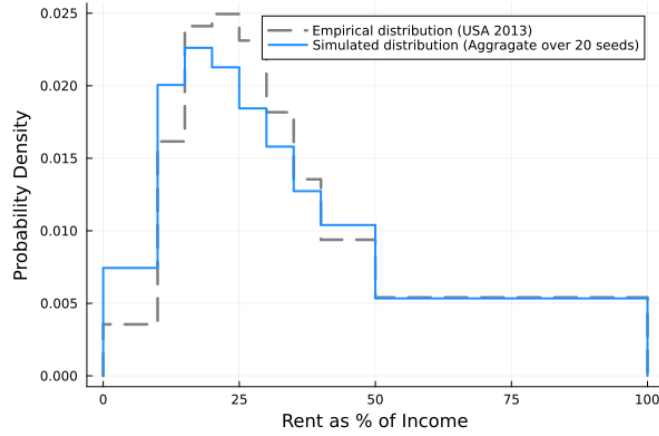


Figure 2.12: Distributions of rents as a percentage of resident income. The gray dashed line shows the empirical data, while the blue line represents the aggregate distribution obtained by combining data from 20 simulations after the thermalization phase.

2.4.2 Gentrification dynamics

The state reached after the thermalization phase serves as the initial setting for the gentrification dynamics. Indeed, the thermalized grid (as previously mentioned) is intended to represent the scenario described by Smith in his theory (see Section 1.1.1), where the city’s residents are strongly divided between the inner-city (low-income) and the suburbs (high-income). According to the theory, as the city grows, redevelopment interventions in the inner-city are triggered. These investments are driven by the possibility of high profit offered by high rent-gap, empty houses. As anticipated at the end of Section 1.3, the responsible for the renovation interventions in this model (as well as in [32]) is third type of agent, who is not represented on the grid. This agent, referred to as the developer, represents professional landlords who purchase property, renew it, and rent it to high-income tenants in order to generate long-term profits.

It is assumed in the model that the city has already expanded enough to

prompt the developer to start investing in the inner-city. Once this new phase is activated, at each model step the developer invests in a certain number of vacant housing patches. She selects randomly a fixed number of unoccupied dwellings, and among them she picks those with the highest rent-gap and renews them. The first effect of the investment is to decrease the age of the renovated house, making it more attractive to agents. This is implemented by sampling the new age from an exponential distribution: $p(x) = \lambda e^{-\lambda x}$ (where λ is set to $1/7$). We chose to draw the new age from an exponential distribution to introduce some stochasticity into the renovation outcome, and to reflect the fact that, although renovation is expected to modernize the dwelling, it does not make it equivalent to a newly built (zero-age) house. The exponential distribution we use tends to return reasonably low values with high probability. As a consequence, the capitalized rent increases according to Eq. 2.6. The second effect is a spillover on the potential rent of the surrounding houses - those present in the Moore neighborhood of radius 1 of the redeveloped patches. This is simply implemented by adding a constant to their potential rent value (Eq. 2.5). The spillover accounts for two realistic consequences of the investment. First, the fact that if a house is redeveloped and its rent raises, the rents of the dwellings around it are likely to increase too [37]. Second, the houses in the neighborhood of a renewed dwelling will become more attractive to further investments, since it is easier to attract a high-income resident in an area already partially redeveloped. Adding a constant to the potential rent is effective in obtaining both these two consequences: indeed, as shown in Figure 2.7, a higher potential rent results both in a higher capitalized rent and a higher rent-gap, making the dwelling more expensive and more likely to be selected and renewed by the investor. After the developer's intervention, the model step proceeds as in the thermalization phase: each agent evaluates her happiness and, potentially, moves. Finally, at the end of each step, the age of all housing patches is increased by a constant, and the capitalized rent is recomputed. This part was absent during the thermalization phase, which was only meant to bring the system into a realistic configuration, not to replicate the city's evolution.

In the first step of this dynamics, the number of investments is fixed, and so is the same in every simulation. Starting from the second step, however, the developer decides the number of investments according to a feedback loop, and the process changes from one simulation to another. She evaluates how successful her previous interventions have been by checking whether the renewed houses have been occupied. In principle, they could be occupied by either high-income or low-income agents. However, because of the new high capitalized rent, low-income agents usually cannot afford redeveloped dwellings. The feedback mechanism works as follows: the developer is associated with a satisfaction value, which increases or decreases at each step depending on how many renewed houses are still empty after three steps since their renovation. The number of investments during a step

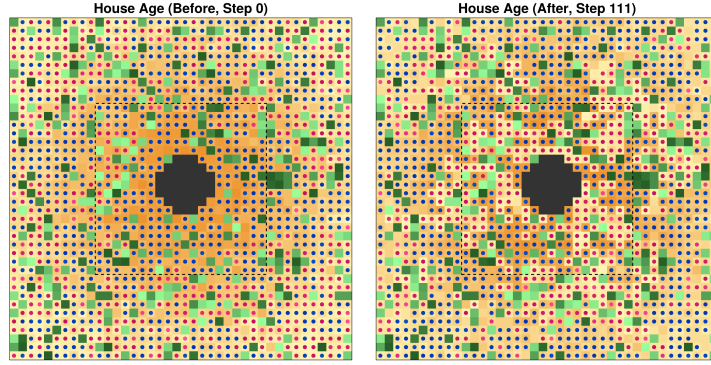
is proportional to this satisfaction value. If the satisfaction falls below a certain threshold, the developer stops investing. Then, once she has stopped, she may start investing again if the fraction of occupied houses among the redeveloped ones rises above another threshold. Note that even if no agents move to a patch during a single model step, this does not mean they will never move there, since at each step unhappy agents randomly select which houses to evaluate.

Figure 2.13 shows how the three properties of housing patches change after a few steps of gentrification dynamics, in a single simulation. The grids on the left display the thermalized state, while those on the right the configuration reached after 111 dynamical steps. To see exactly which houses have been renewed it is more useful to focus on the age plot (Figure 2.13a), since the age is the only property that changes only for redeveloped patches. In this example several areas of the inner-city and few parts of the suburbs have been gentrified, and the majority of redeveloped dwellings have been occupied by a high-income resident. In the second plot (Figure 2.13b) it can be observed how the spillover effects rise the potential rents in gentrifying areas. As previously explained, the potential rent is not increased directly in the renewed house, but in its neighbors. Comparing the plots of these two properties, it becomes evident that the houses affected by the spillover are often the next to be renewed. Indeed, many of the patches whose potential rent increased also show a decrease in age. This is consistent with the point discussed above: it is more realistic for investments to occur in nearby areas rather than sparsely. The last plot (Figure 2.13c) shows how the capitalized rent changes during the dynamics. In this case, both the effects of direct redevelopment and spillover are present. From the plots of the two rents, it may seem that the most significant gentrification has occurred in the suburbs rather than in the inner-city, due to the more intense colors. However, this is only because the initial values of CR and PR were already quite high in those areas. To study gentrification more effectively, it is more meaningful to focus on how the color changes in an area, rather than on the final intensity it reaches.

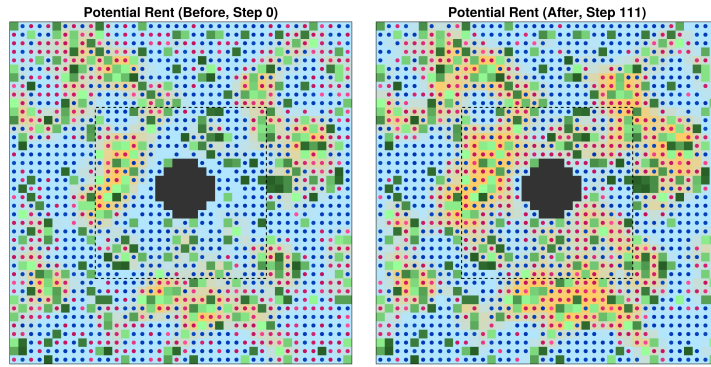
The plots in Figure 2.13 clearly show that the simulated process leads to a city where the inner city is occupied by a greater number of high-income residents and offers higher rents compared to the thermalized state. However, this represents only a single simulation. Although the initial number of investments is the same in every run, the different random seeds used to initialize the model result in slightly different thermalized starting grids. These differences in the initial random setup have a significant impact on the gentrification dynamics.

The quantitative indicators chosen to study the process of gentrification in the inner-city are the medians of capitalized rents and incomes in that area. The median is one of the most commonly used indicators in urban studies. Observing

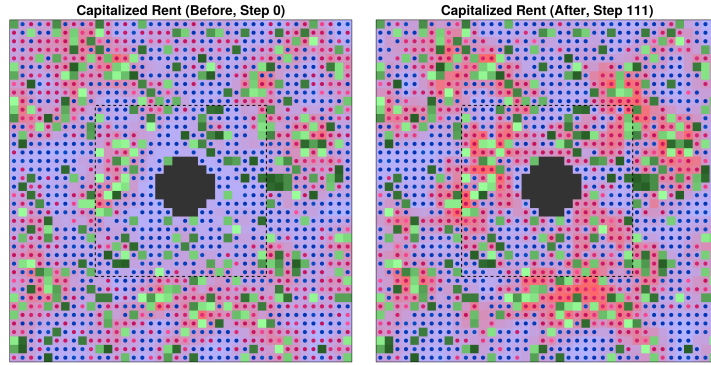
capitalized rent allows us to see changes in the housing stock, while the median income reflects changes in the population. Note that agents' incomes are fixed, so the only way the median of the distribution can change is through the arrival or displacement of residents. Figure 2.14 shows the behavior of these two indicators during the dynamics for two different simulations. The green plot corresponds to the same simulation shown in Figure 2.13. In this case, both rent and income increase significantly, and it can be considered an example of successful gentrification. However, the red plot shows a very different scenario: both indicators rise at the beginning of the process but then stop after a few steps. The median income remains constant, while the median rent gradually decreases. This latter effect is simply due to the aging of the housing stock at each step, which results in a slow decline in capitalized rent. To better understand what is happening in the red simulation, we can explicitly examine the grid before and after the dynamics, shown for housing age in Figure 2.15. The plot confirms that the developer did not invest significantly in the inner-city housing stock. However, she did make a few investments in some areas of the suburbs near the inner-city border. Apparently, those patches—despite being newer—offered a higher rent-gap. In fact, the rent-gap is not determined by age alone, but also by the potential rent (see Eq. 2.7), which depends on surrounding amenities.



(a) Age



(b) Potential rent



(c) Capitalized rent

Figure 2.13: Comparison of the three housing patch properties before and after the gentrification dynamics. **(a) Age:** The age of redeveloped houses decreases, lightening them from orange to yellow. **(b) Potential rent:** Color yellow is associated with high potential rent, it increases because of spillover effects. **(c) Capitalized rent:** Rents of renewed houses and of their neighbors increase, changing color from purple to red.

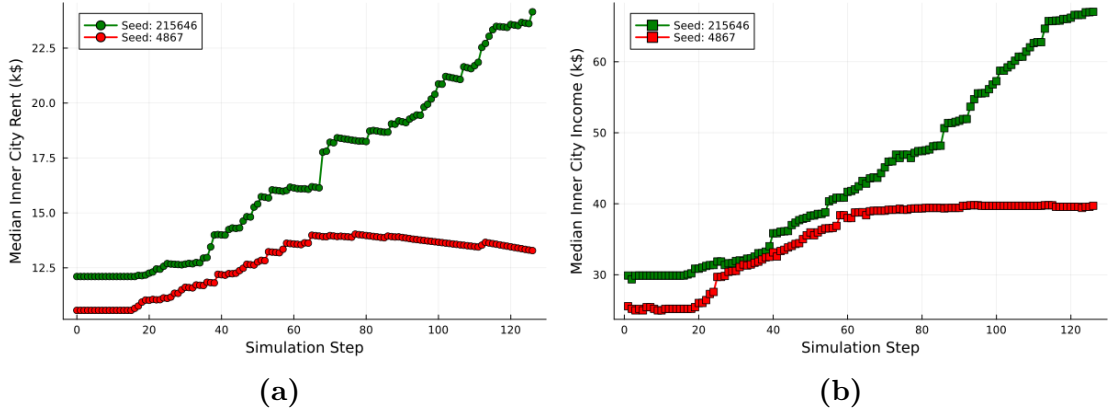


Figure 2.14: Trends of the median of the rents and the median of the incomes in the inner-city for two different simulation. Seed 215646 (green) represents a successful gentrification case: both the indicators grow significantly; seed 4867 (red) is instead an unsuccessful example.

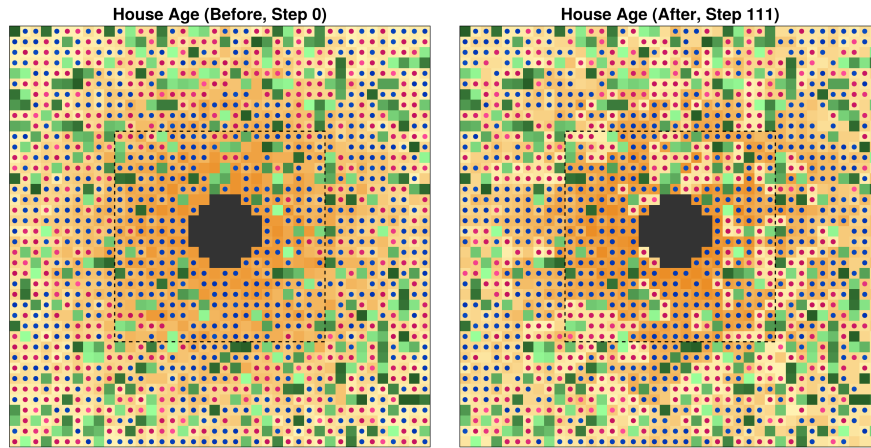


Figure 2.15: An example of unsuccessful gentrification: few houses are renewed in the inner-city, while the investments in the suburbs are significant.

Therefore, the model is able to simulate different scenarios: some in which the inner city undergoes strong gentrification, and others in which the process halts early. The factors driving these different outcomes are related to the initial random setting and are examined in the next section.

2.5 Model validation

In this section, we compare the simulated growth of inner-city rent and income with real-world data from several U.S. gentrifying neighborhoods. Based on this comparison, the model simulations have been clustered in different groups, and we selected a set of successful gentrification seeds. After studying what causes the different behaviors in the simulations, we will focus the analysis on the selected gentrifying seeds.

2.5.1 Real-world examples of gentrification

Since the prototype of city used as a reference to build the model is that of the United States - that is the opposite in terms of income spatial distribution with respect to European cities, as discussed in Section 1.3 - we took also the data to validate the model from the U.S. . A list of neighborhoods that have undergone a strong gentrification in the last years was taken from a 2018 CNBC article [38]. All these areas are parts of the inner-city in their respective cities. The data for the median gross rents and incomes in these neighborhoods were, once again, taken from the American Community Survey, available on the U.S. Census Bureau website. While the article refers to gentrification since the year 2000, the oldest data available from the source are from 2011. Therefore, the time interval considered for this analysis is from 2011 to 2020. The data used are inflation-adjusted and reported as 5-year estimates, meaning that the value associated with a specific year is the average of that variable over the previous five years. For example, the median rent value referred to as 2011 is actually the average of the values from 2007 to 2011. Because the considered neighborhoods are in different cities, the magnitude of the data varies significantly among them. To make the data comparable, they were indexed to the year 2011 to represent the percentage growth. Given the original set $X = \{x_1, x_2, \dots, x_n\}$ of median rents or incomes from a specific neighborhood, it was transformed into the set $Y = \{y_1, y_2, \dots, y_n\}$ where $y_i = \frac{x_i}{x_1} \cdot 100$. This process results in data where the y-value represents the percentage of the 2011 baseline. The trends of these indexed data are shown in Figure 2.16, the neighborhood are identified through the name of the city and their ZIP code. The continuous lines represent the gentrifying neighborhoods mentioned in the CNBC article, while the dashed lines show two control neighborhoods. These controls are not in the list of gentrification examples and are included to demonstrate that the observed growth in rents and incomes is not due to some general factor affecting all areas. Indeed, while all the gentrifying examples show a final percentage growth larger than 50% for the rents, and than 60% for the incomes (meaning that the last point of the curves is respectively above 150 and 160) the controls show little or no growth.

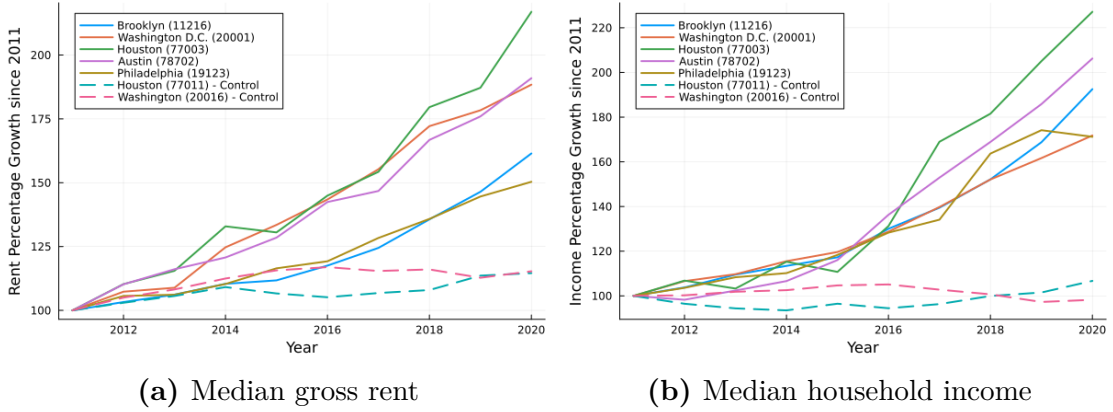


Figure 2.16: Trends of median rents (a) and incomes (b) in real U.S. neighborhoods from 2011 to 2020. The gentrifying areas (continuous lines) display a significant growth in both indicators, while the control neighborhoods (dashed lines) show little or no growth. The numbers in the legend represent the ZIP code identifying that area.

2.5.2 Simulations clustering

The plots of inner-city median rents and incomes from the model simulations, shown in Figure 2.14, illustrate how these indicators evolve step by step. In contrast, the real-world data in Figure 2.16 are based on 5-year estimates, meaning that only one value is reported per year, each averaged over a relatively long time interval. In order to compare the simulation results with the empirical data, it is necessary to define the temporal duration of a model step and to replicate the 5-year averaging process. In this model, one year corresponds to 8 steps, meaning that each step represents approximately one and a half months. We ran the simulations run for 112 steps, corresponding to the 14 years encompassed by the real data. Note that although the data source ranges from 2011 to 2020, the 2011 value is actually an average based on data starting from 2007 — hence, the total time span is 14 years. To make the simulation output comparable with the 5-year estimates in the empirical data, we applied a rolling 5-year moving average to the 112-step simulation results, using overlapping windows of 40 steps (equivalent to 5 years each). An example of the result of this process is shown in Figure 2.17. The simulation reported there is the same one shown in green in Figure 2.14; the number of model steps is identical — what differs is the way the data are represented. The gray curves in the background represent the empirical data from the gentrifying neighborhoods shown in Figure 2.16. This specific simulation exhibits rent and income growth that is comparable to the real data — at least in terms of growth magnitude (though the shape is more linear than that of the empirical curves).

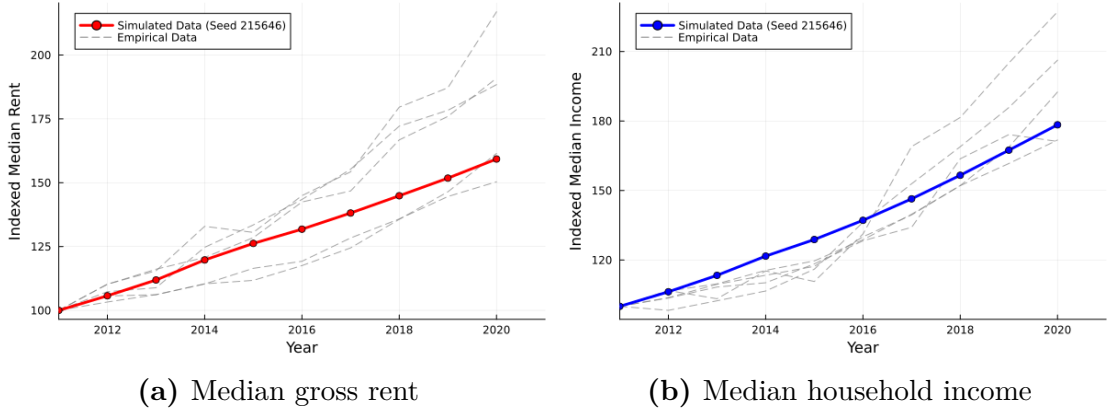


Figure 2.17: Trend of median inner-city rent and income from a single simulation. The 10 values shown are obtained by applying a rolling moving average over overlapping intervals of 40 steps each on the 112 simulated steps. All values are indexed to the first one. The dashed gray curves in the background represent the empirical data.

However, as discussed at the end of Section 2.4.2, model simulations exhibit very different behaviors.

A K-means clustering was performed to group the model simulations based on their agreement with empirical data. The K-means algorithm clusters a set of data into K groups according to their Euclidean distance in the data space: it starts by defining K random centroids and forming the K groups by assigning each point in the set to the closest centroid; then, the centroids are recomputed as the average position of the points in each cluster, and the groups are redefined accordingly. This process is iterated until the centroids stop changing significantly. In our case, for each seed, four parameters were defined, and a 3-means clustering was then performed in the resulting four-dimensional data space. The four values were computed from the 20 indexed rent and income data points of each simulation — such as those shown in Figure 2.17. The parameters are defined as follow:

- **Combined Root Mean Square Error (RMSE):** For the rent and income values of each year, the average distance from the corresponding values in the empirical data is computed. These distances are squared, and a single average value is then calculated — averaged over both indicators and all ten years. Finally, the square root is taken. This parameter represents the overall distance between the simulation outputs and the real data: the lower is the value the better is the agreement.
- **Combined Shape Distance:** The procedure is identical to the RMSE, but instead of using the original rent and income values, it is applied to the

differences between consecutive years. For example, the rent value of 2011 is subtracted from that of 2012, and so on. At the end, the RMSE is computed on the resulting 18 difference values (9 for rents and 9 for incomes), measuring their distance to the differences in the empirical data. This parameter captures the similarity in shape between the simulated and empirical trends. Again, the lower is the value the better is the agreement.

- **Final rent index:** The final indexed rent value for the year 2020.
- **Final income index:** The final indexed income value for the year 2020. These last two parameters were included to give weight to the final percentage growth.

The model was run for 600 different simulations, each using a randomly selected seed. For each simulation, the four parameters were computed, and clustering was performed on the resulting dataset. To make the distances between these different parameters comparable, the data were normalized before clustering using the formula:

$$y_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}$$

where x_i is the value of any of the four parameters for a specific seed, and $x_{\min}$, $x_{\max}$ are the minimum and maximum values of that parameter across all simulations. The result of the clustering process in the original parameter space (non-normalized) is shown in Figure 2.18.

The four parameters are clearly correlated: in each of the six plots in Figure 2.18, the data exhibit an approximately linear relationship. The first plot is particularly informative about the agreement between the simulations and the empirical data, as it displays the combined RMSE and the combined shape distance. The simulations that best match the real-world data belong to the red cluster, which shows the lowest values for both of these parameters. These seeds are also, on average, those with the highest final values of indexed rents and incomes. Figure 2.19 shows the actual trends of the clustered simulations, colored according to the 3-means clustering results shown in Figure 2.18. It is clear that the majority of the simulated dynamics exhibit a slower growth in inner-city rents and incomes compared to the empirical data. This is actually a realistic property of the model. Indeed, the real-world data were taken from some of the most gentrified neighborhoods in the United States, so it is plausible that they are a sort of extreme case for the model simulations. The real U.S. neighborhood trends are faintly visible in the background of Figure 2.19, but are the same as those shown in Figures 2.16 and 2.17. They show a superlinear growth of the indexed indicators, while our simulations tend to grow sublinearly. However, the model did not aim to exactly reproduce real gentrification indicators trends. As mentioned at the beginning of this chapter,

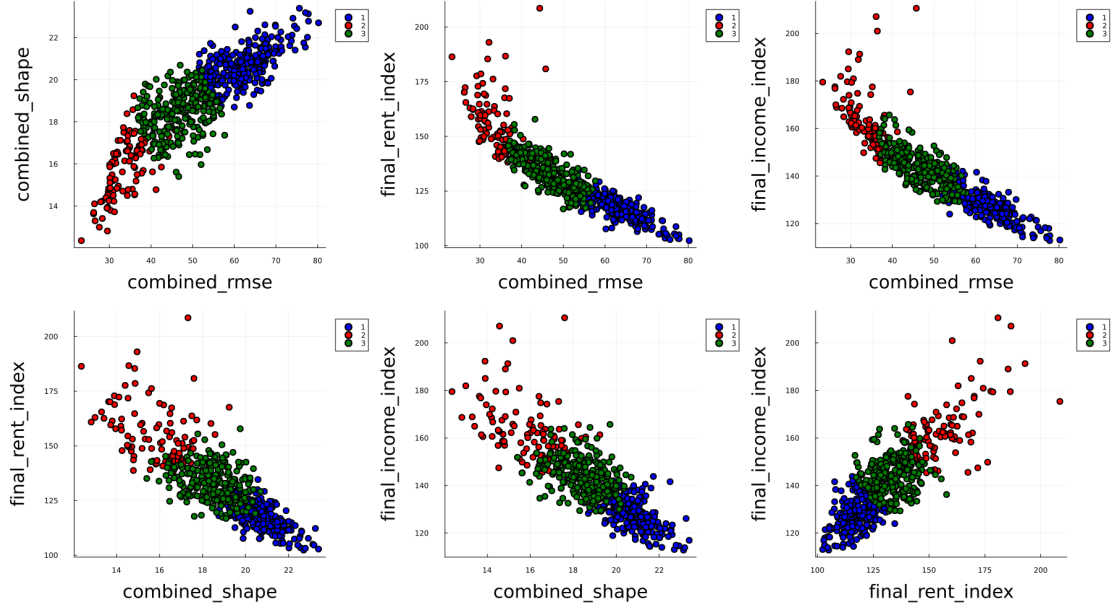


Figure 2.18: Clustering of the 600 simulations in the 4-parameter space, shown in all six possible pairwise 2D projections.

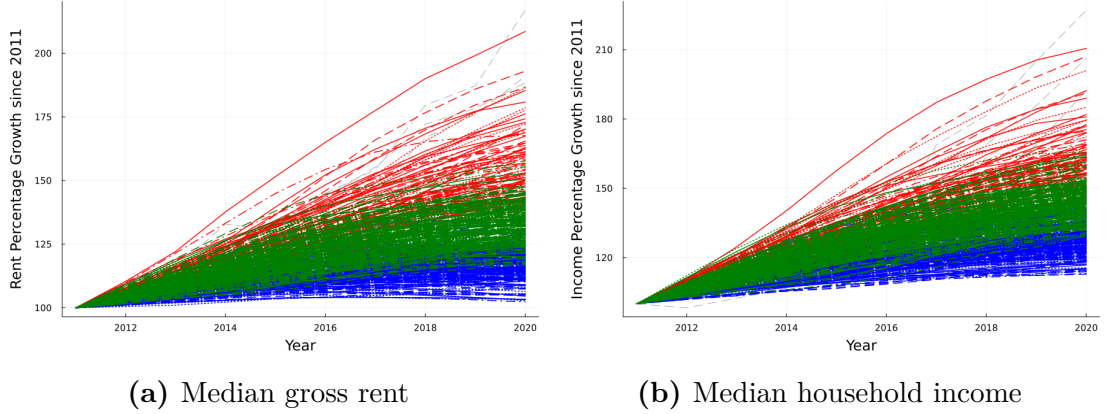


Figure 2.19: Trends of median rents **(a)** and incomes **(b)** of the 600 simulations, colored according to 3-means clustering result shown in Figure 2.18. The empirical trends are included in the background, but are barely visible because they are covered by the simulation lines.

several real-world aspects are neglected, making it unrealistic to attempt a precise quantitative replication of the process. Still, the simulations shown in red display final indexed values comparable to the empirical ones, indicating that in those cases the final effect of the gentrification process is similar to the real examples. These

simulations are considered the most successful examples of gentrification and, since one of the purposes of this thesis is to investigate how gentrification affects urban segregation, in the next chapter we will focus exclusively on this set of seeds, as they are expected to exhibit more significant effects.

Before concentrating exclusively on the successful seeds, we aim to investigate what determines the different behaviors observed across simulations, which are apparently very similar in their initial settings (see Sections 2.2, 2.3). It is difficult to isolate the effects of the random processes that occur during the gentrification dynamics, such as the specific housing patches selected by agents before moving, or those evaluated by the developer. However, these sources of randomness are unlikely to significantly affect the overall outcome, as the large number of agents and time steps should compensate for differences among simulations. What should cause the difference evolutions of the dynamics is the initial random setting of the grid - before the thermalization phase. One of the aspects that can vary the most from one seed to another is the distribution of amenities, both in terms of spatial location and attractiveness levels. Another is the initial spatial distribution of high- and low-income agents across the grid. A third source of variation is the age of housing patches. However, while the number of amenities is small and their attractiveness levels can vary significantly, the number of housing patches is large, and their age tends to fluctuate only slightly around an average value determined by their distance from the center. For this reason, the effect of housing age variability is also expected to be smoothed out when considering the grid as a whole. Several metrics describing different aspects of the initial random configuration were defined and compared across the three simulation clusters. Among these, the two that exhibited the most significant differences are: the average perceived amenity in the inner-city, and the initial random number of low-income residents located in the inner-city before thermalization occurs. The average perceived amenity is calculated as the mean amenity score of all housing patches in the inner-city. This score corresponds to the second component of the potential rent, as defined in Eq. 2.5. Importantly, amenities outside the inner-city can still influence this score—for instance, a green patch located in the suburbs may contribute to the amenity of a nearby housing unit situated at the edge of the inner-city.

The comparison of these two metrics across clusters is presented in Figure 2.20 using boxplots. The color scheme is consistent with Figures 2.18 and 2.19. Each box represents the interquartile range (i.e., the central 50% of simulation values for a given cluster), the horizontal black segment stands for the average value, while the whiskers extend to cover approximately the full range of the distribution, excluding outliers, which are shown as individual dots. From this figure, we can observe that simulations belonging to the red cluster—those exhibiting the most significant gentrification—are, on average, associated with higher levels of perceived amenity

and lower initial numbers of low-income agents in the inner-city. Not only is the red box the top one in terms of perceived amenity and the bottom one for initial low-income presence, but the remaining two boxes also follow a consistent pattern with the trends shown in Figure 2.19. In both cases, the green box lies between the red and the blue, just like the green trends, which fall between the other two groups. This observation supports the idea that the initial values of these two metrics—perceived amenity and presence of low-income agents—help determine, at least in part, the degree of gentrification that will occur in the inner-city over the course of the simulation.

These results are reasonable: the degree of gentrification depends on how many high-income agents the developer’s investments are able to attract into the inner-city. If the random initial configuration of a simulation features a high level of perceived amenity in the inner-city, it implies that the amenity values of green patches in or near that area are particularly high. This, in turn, makes it easier to attract high-income residents, since the amenity score is a key component of their utility function (Eq. 2.8). Conversely, if the initial random distribution of agents places a high number of low-income residents in the inner-city, this tends to result

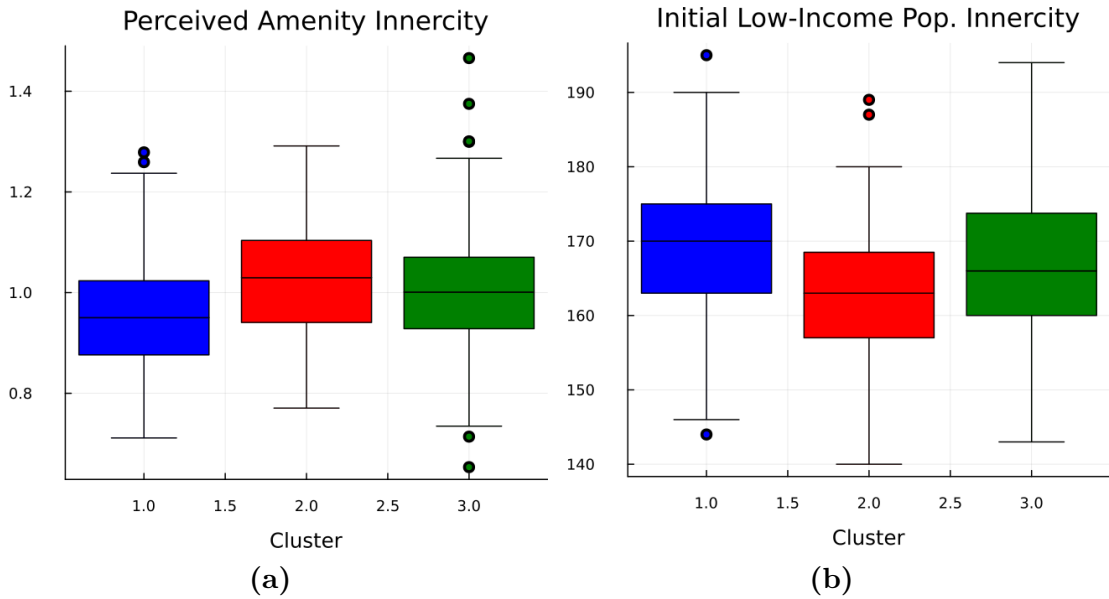


Figure 2.20: Comparison of average perceived amenity **(a)** and the initial number of low-income agents **(b)** in the inner city between the 3 clusters. The most successful gentrification cluster is in red, the least successful gentrification cluster is in blue. Boxes represent the central 50% of simulation values for that cluster, horizontal black line stands for the average value, whiskers cover the full range of the distribution apart from the few outliers displayed as individual dots.

in a denser presence of low-income agents even after thermalization. Low-income agents are generally satisfied with living in the inner-city due to the availability of affordable housing; therefore, a higher initial presence leads to a greater number of "happy" agents who will not move away during thermalization. As a consequence, a densely occupied inner-city makes it more difficult for the developer to attract many high-income agents later on.

Similar boxplots were produced for several other metrics, such as the initial number of high-income agents in the inner-city, the average house age, and the fragmentation of amenities. However, in all these cases, the boxes were largely aligned across clusters, suggesting that these aspects are less relevant to the evolution of gentrification. All of these plots—including the two previously discussed—are shown in Figure 2.21. Apart from the two metrics already analyzed, only the average potential rent in the inner-city shows a notable divergence between clusters. This is not surprising, as potential rent is directly influenced by the perceived amenity.

The fact that the initial random number of low-income agents influences the gentrification process is, to some extent, an unintended effect. The purpose of the thermalization phase was to guide the system toward a realistic configuration and to eliminate the influence of the initial random setup. However, due to the way the relocation rules are defined, it is impossible to fully remove this initial bias. By

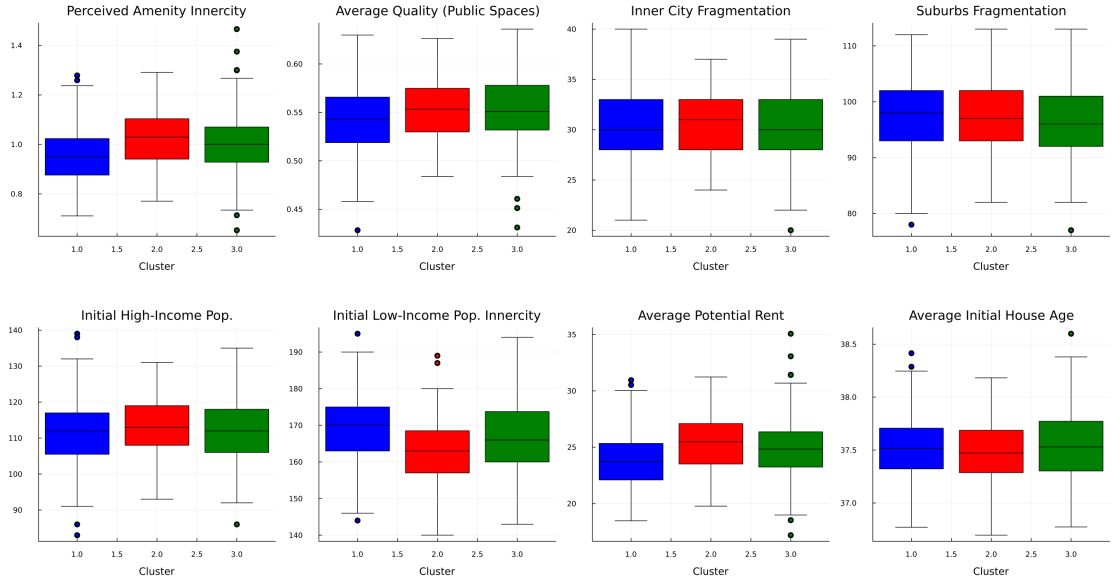


Figure 2.21: Boxplots of all the evaluated metrics.

contrast, the influence of inner-city amenities as a driving factor for gentrification is both expected and realistic. This aspect could also be leveraged in future work to make the model predictive: by initializing the model with the spatial distribution of amenities from a real city, it would be possible to compare the gentrification trends predicted by the model with those observed empirically.

2.6 Introduction of public housing

The model described so far was designed to simulate gentrification dynamics in a real estate market left entirely to private forces, without any form of public regulation or intervention. In this section, we present a modified version of the model that includes the introduction of public housing. As discussed in Section 1.1, there is broad agreement in the literature that public policies can play a crucial role in controlling and mitigating the consequences of gentrification — and public housing is often considered one of the most effective measures. Several cities, such as Vienna, have already moved in this direction.

The model initialization, thermalization, and the initial phase of the gentrification dynamics remain exactly the same as in the previously described version of the model. The key difference in this new version is that rent growth in the inner-city is monitored at each step. When the median rent reaches a 50% increase compared to its initial value, a public housing policy is triggered. This mechanism represents a form of government supervision that continuously monitors the effects of gentrification and intervenes when rents rise excessively. Once the policy is activated, a number of housing patches are selected and converted from private to public. To be eligible for selection, a dwelling must either be vacant or occupied by a low-income resident, and must not have been renewed by the developer. A public house is associated with a fixed low rent, is accessible only to low-income agents, cannot be selected by the developer, and is not affected by gentrification spillover effects: even if one of its neighboring patches is renewed, its potential rent does not increase. On the contrary, public houses generate a negative spillover effect: the potential rents of the surrounding dwellings decrease by a fixed amount when a nearby house is converted to public. As explained in Section 2.4.2, changes in potential rent affect both the capitalized rent and the rent gap of a patch. Therefore, the negative spillover makes public housing a kind of counterbalance to private investment: while the redevelopment of a property raises the rents of adjacent dwellings, the introduction of a public house — with its low, fixed rent — lowers them. In addition, the negative effect on the rent gap reflects the idea that, whereas partially redeveloped areas tend to attract further investment, the surroundings of a public house become less attractive to developers — also because

public houses cannot themselves be gentrified. Finally, once a dwelling becomes public, its age stops increasing with time steps. If its age exceeds a certain threshold at the moment of intervention, it is reset to that threshold — the idea being that, once under public ownership, the house is kept in decent condition and regularly maintained by the government.

The fraction of patches to be converted into public housing is provided as an input to the model, and they are all transformed at once as soon as the policy is activated. After the intervention, the simulation dynamics proceed as before: unsatisfied agents relocate, and the developer continues to invest. The only difference is that low-income agents who are dissatisfied and searching for a new dwelling now consider public housing options first. We simulated two different public housing selection strategies:

- **Random:** Among the eligible houses, those to be converted from private to public are selected randomly (see Figure 2.22a).
- **By rent gap:** The eligible houses with the highest rent gaps are converted into public housing (see Figure 2.22b).

The random strategy represents a sort of blind intervention: the government acts uniformly across the city, without considering which areas are experiencing the most intense gentrification effects. The rent gap-based strategy, on the other hand, is targeted: dwellings with the highest rent gaps are likely to be located in the inner city and are the most at risk of imminent gentrification, since the developer selects investments based on rent gap maximization.

Figure 2.22 shows the state of the grid at the end of the simulation, after the public housing intervention. The public housing fraction was set to 5% in this case. Public houses are shown in white. In the left panel, they are selected randomly, while in the right panel they are selected based on rent gap. In this case, as expected, they are primarily concentrated in the inner city. The simulation shown corresponds to the same seed presented in Figure 2.13 for the previous version of the model, meaning that the distribution of amenities and the dynamics prior to the intervention are exactly the same.

The way public housing policy is simulated in this model is highly simplified and does not aim to be realistic from an urban economics perspective. Nonetheless, the way the introduction of public housing can influence gentrification dynamics remains of interest and will be examined in detail in the next chapter.

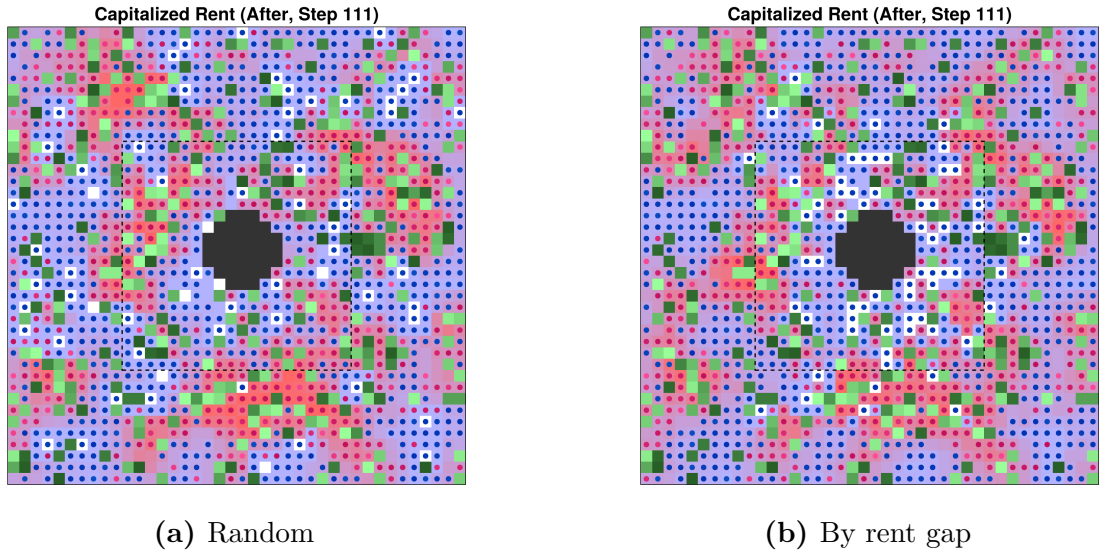


Figure 2.22: City state at the end of the dynamics after the public housing intervention. Capitalized rent as a background. Public houses are displayed in white. They are randomly selected in **(a)** and selected by rent gap in **(b)**. (Public housing fraction set to 0.05).

Chapter 3

Results

In this chapter, we first analyze the effects of public housing policy on simulated gentrification dynamics, then we define a set of segregation measures and examine their evolution throughout the process. All simulation seeds used in this chapter are drawn from the successful gentrification cluster identified in Section 2.5.2.

3.1 Mitigation effects of public housing

Figure 2.22 shows the state of the city after the public housing intervention, but it provides limited insight into how the gentrification process changes following the transformation of 5% of housing patches from private to public. To investigate this, we compare the trends of inner-city median income across the three scenarios: fully private market, randomly assigned public housing, and rent gap–selected public housing. Figure 3.1a shows the trend of this indicator in the same simulation shown in Figure 2.22, which is also based on the same seed used in Figures 2.13, 2.14, and 2.17. In the figure, the start of the developer dynamics is indicated by the vertical dashed line; the changes before that point are due to thermalization. The three curves begin to diverge only after the public housing intervention—before that, the dynamics are exactly the same across the three scenarios. A few steps after the point of divergence, the orange and green trends—representing the public housing scenarios—stop growing. The curve of the free-market scenario, instead, continues to increase, since in that case nothing changes at that point. Therefore, in this simulation, the introduction of 5% public housing appears to be quite effective in halting the gentrification process: the stabilization of median income reflects the fact that the displacement of low-income agents from the inner city, as well as the arrival of high-income residents, must have stopped. This is confirmed by Figure 3.1b, which displays the changes in the number of low-income and high-income agents in the inner city during the process. In the free gentrification scenario, the

number of low-income agents decreases and the number of high-income agents increases throughout the dynamics, eventually reaching the same value at the end of the run. In the two public housing scenarios, instead, both populations stop varying significantly. Note that in the few steps immediately following the intervention, the number of low-income agents in the inner city under the randomly selected public housing policy is actually lower than in the free-market scenario, as if the policy increased displacement instead of reducing it. This is, however, expected: if public housing units are selected randomly, some of them will be located in the suburbs and may attract a few low-income agents who previously lived in the inner city. However, after few steps in this scenario the number of low-income agents stabilize, while it continues to decrease in the free dynamics.

These plots suggest that the introduction of public housing made the developer either less able to displace low-income agents and attract high-income residents to the inner city, or simply less interested in investing in an area that is now partially controlled by public institutions. Figure 3.2 confirms this idea, showing the number of investments in the inner city at each step for the considered simulation. The blue background plot represents the scenario without intervention, while the red in Figure 3.2a represents the randomly selected public housing case, and the green in Figure 3.2b the rent gap-based case. It is evident that after the policy intervention,

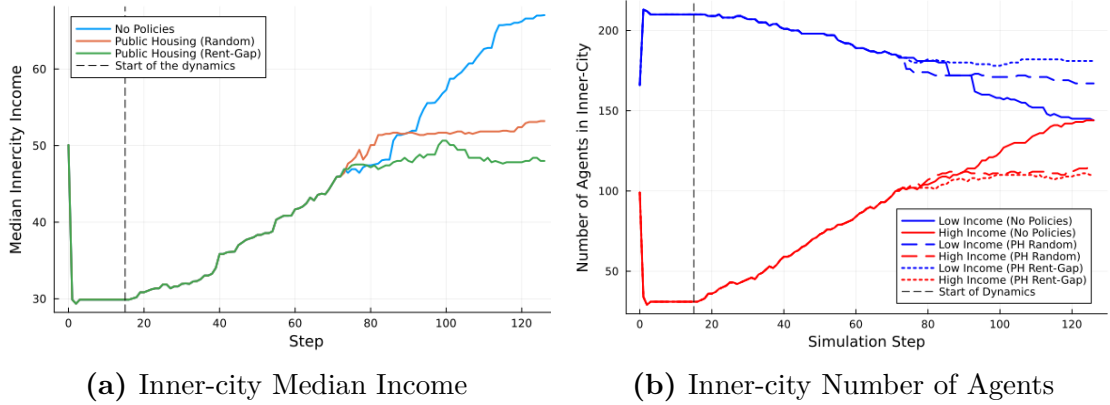


Figure 3.1: Trends during the gentrification dynamics in the three defined scenarios: no public housing intervention, public housing selected randomly, and public housing selected based on rent gap. **(a) Inner-city median income:** In the two scenarios involving public housing, a few steps after the intervention, this indicator stops increasing. **(b) Number of agents inside inner-city:** In the free-market scenario, the number of high-income agents in the inner city continues to grow during the process, while the number of low-income agents decreases. Once public housing is introduced, both trends quickly stabilize. (Public housing fraction set to 0.05).

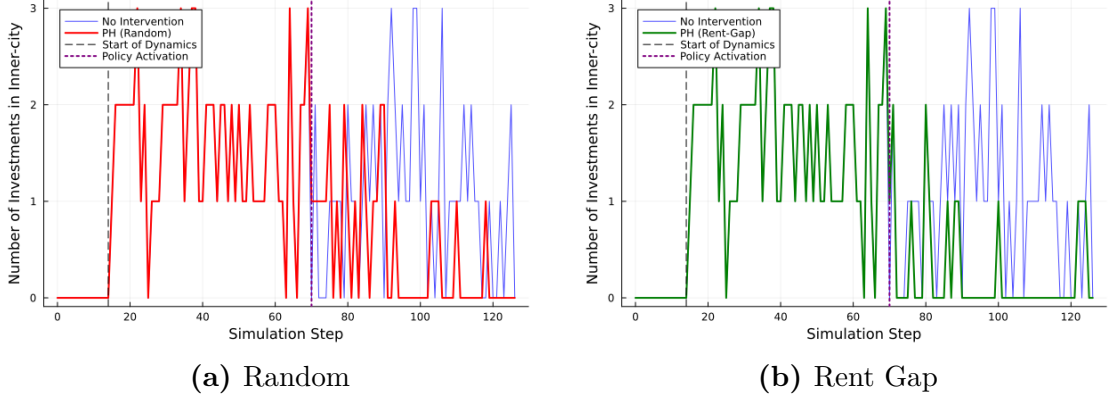


Figure 3.2: Number of developer’s investments in the inner-city throughout the dynamics, for a single simulation. Randomly selected public housing units **(a)** and selected by rent gap **(b)**. In both plots the free dynamics trend is shown in the background in blue and two vertical, dashed lines indicate the start of the dynamics and the policy intervention. In both public housing scenarios the number of investments drops after the intervention. (Public housing fraction set to 0.05).

in both cases, the number of investments in the inner city decreases with respect to the free dynamics, with a stronger effect in the rent-gap-based scenario.

For clarity and robustness, we also plotted the number of investments in the inner city, averaged over 15 different simulations, keeping the public housing fraction fixed at 5%. The seeds were selected from the successful gentrification cluster identified in Section 2.5.2. The plot is shown in Figure 3.3. It is clear from this figure that the average number of investments in the inner city decreases during the dynamics in each scenario, including the one with no policy intervention. For this fraction of public units, the average trend in the random scenario is not very different from the free case, whereas in the rent-gap-based scenario the number of investments drops significantly after the policy is activated. This is understandable: in this scenario, the top 5% of available housing patches with the highest rent gap are converted to public housing. Many of these are likely located in the inner city and represent the most profitable opportunities for the developer. Furthermore, due to the negative spillover effect on potential rent, the rent gap of the neighboring housing patches (most of which are also in the inner city) decreases, making them less attractive to the developer. As a result, the developer loses interest in this area and may instead continue investing in parts of the suburbs.

To better understand the effect of public housing policy on gentrification dynamics, we also aggregated the trends of inner-city median income over the same 15 simulations, and plotted them for different fractions of public housing units.

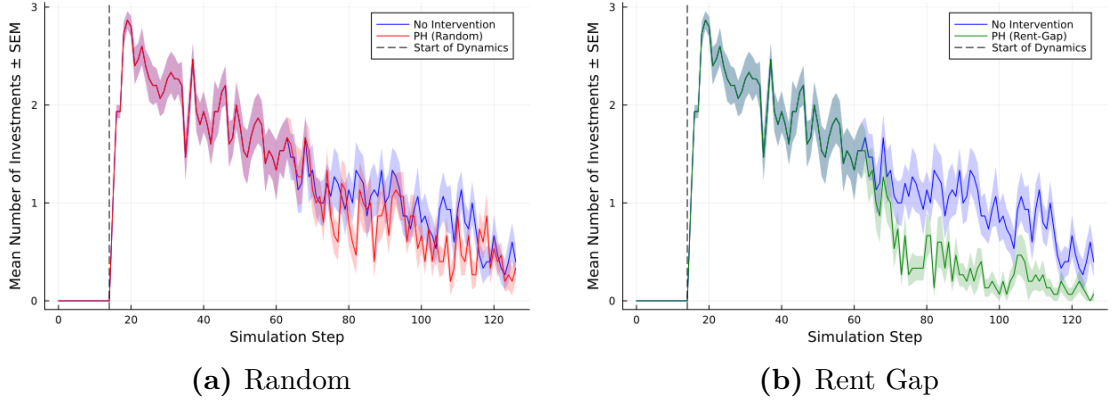


Figure 3.3: Number of developer’s investments in the inner-city throughout the dynamics, averaged across 15 simulations. Randomly selected public housing units (a) and selected by rent gap (b). (Public housing fraction set to 0.05).

The results of this analysis, for both random and rent gap–based selection methods, are shown in Figures 3.4a, 3.4b. This time, the model was run for a larger number of steps, with respects to the simulations displayed so far, to allow the system to reach a stationary state. By doing so, we observe that even in the scenario without public housing (top blue trend in the figure), the median income eventually stops increasing after a sufficient number of steps, suggesting that the inner city becomes saturated. Increasing the fraction of public units, the main effect is that the income growth slows down, and the plateau reached in the final steps shifts downward. To better illustrate the difference between the two policy scenarios, Figure 3.4c shows how the final value of the average median income varies with the fraction of public housing. It is clear from this plot that the rent gap case is faster and more effective in stopping the process: a fraction of 5% public housing is already enough to keep the final income value around 50, which is approximately the median income at the time of the policy intervention. Meaning that gentrification of the inner-city stopped very fast after the policy activation. Increasing the fraction over 0.05 does not result in a lower value of the median income. Indeed, since the privately renewed dwellings, or in general those occupied by high-income agents, cannot be made public, the median income will not fall below the value it was at the step of the intervention. Instead, interestingly, for the highest considered values of the public housing fraction, the final value of the inner-city median income begins to increase again. This is due to the fact that, as the fraction increases, some of the selected houses with the highest rent gaps start to be located in the suburbs, and, once made public, they may attract low-income agents from the inner-city. In the random case instead the final income decreases more slowly with the public

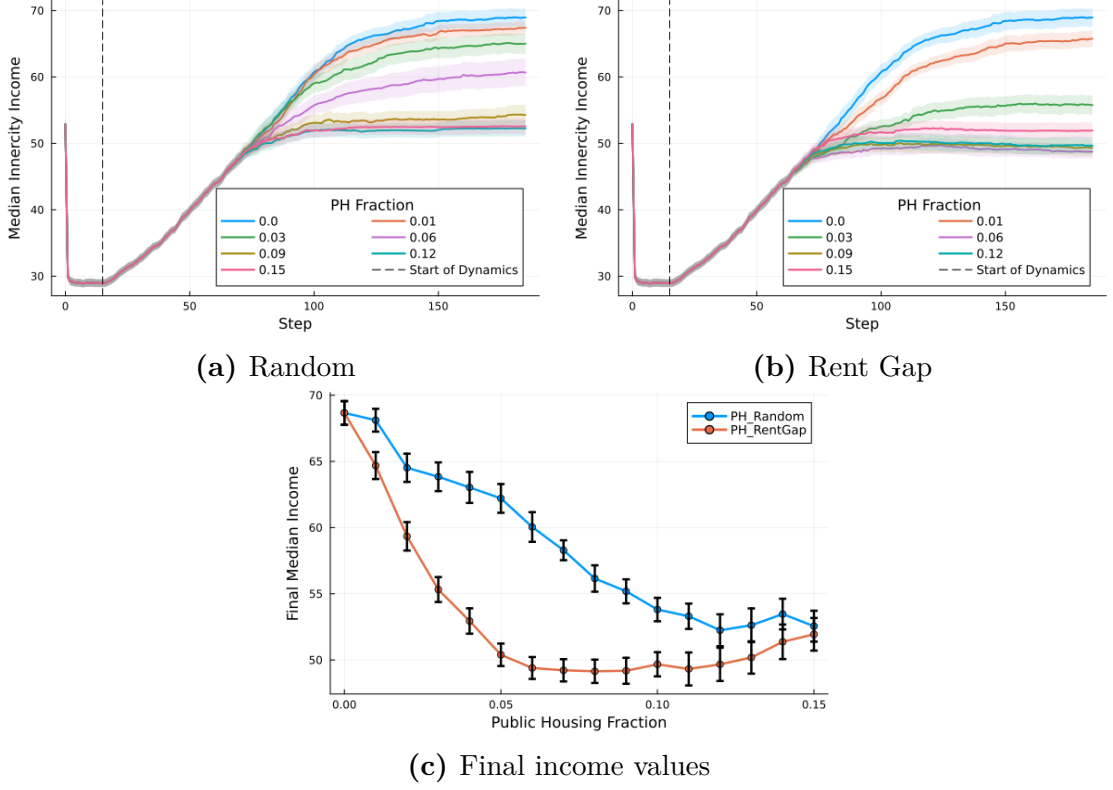


Figure 3.4: (a), (b) Trends of inner-city median income, aggregated over 15 simulations, plotted for different fractions of public housing units. Panels (a) and (b) correspond to public housing units selected randomly and based on the rent gap, respectively. (c) Plot of the final values of income for all the fractions simulated.

housing fraction and seem to stabilize only for the last considered values. In this case, the final median at the fraction 0.05 is closer to the value associated to the free dynamics rather than the corresponding value in the rent gap case. This is somewhat in agreement with the trends of the number of investments shown in Figure 3.3 for the 5% of public housing units.

3.2 Segregation measures

In this section, we introduce several measures of urban segregation and analyze how they evolve during the gentrification process, in order to understand whether this dynamics leads the city toward a more or less segregated state.

3.2.1 Neighbor segregation index

The first index we consider is the neighbor segregation index, which was introduced in the context of Schelling’s model (Section 1.2, Figure 1.3). It is defined as the fraction of neighbor links connecting agents belonging to the same group. Considering the grid as a lattice, each patch is connected by a link to the eight patches in its Moore neighborhood. A neighbor link refers to a connection between two occupied housing patches. The index can thus be expressed as:

$$N = \frac{|E_{ll}| + |E_{hh}|}{|E|} \quad (3.1)$$

where $|E_{ll}|$ and $|E_{hh}|$ denote the number of links between two low-income and two high-income agents, respectively, and $|E|$ is the total number of neighbor links in the grid. We previously analyzed the behavior of this index during the thermalization phase (Section 2.4.1), showing that the thermalized state from which the gentrification dynamics begins is already quite segregated. Indeed, Figure 2.11 shows that, on average, about 62% of the links connect residents from the same group. Examining the evolution of the neighbor segregation index during the gentrification dynamics, we observe that it increases further.

Figures 3.5a and 3.5b show the trend of the index, averaged over the usual 15 simulations, for the free dynamics scenario as well as for the two policy scenarios with 5% and 10% of public housing. In the absence of any policy intervention, the index increases significantly throughout the simulation, exceeding 70% in the final steps. Although the growth is slower when a policy is implemented, it remains significant. Unlike the trends in inner-city median income shown in Figure 3.4, here the difference between the two public housing policies is minimal. This is confirmed by Figure 3.5c, which shows the values of the index at the end of the dynamics for all simulated fractions of public housing. The plot reveals that, in both policy scenarios, final segregation follows the same linear decreasing trend with respect to the fraction of public units.

This coherence in terms of resulting segregation between the two methods of selecting public housing units may seem surprising. One might expect random selection to be more effective in mixing the two groups, compared to the rent-gap-based method, since patches selected based on rent gap are more likely to be already linked or clustered. However, it is important to remember that, even in the random case, the patches available for selection are only those that are either empty or already occupied by low-income agents. This makes it unlikely for a selected patch to be located in an area predominantly inhabited by high-income residents, and thus to have a strong mixing effect. As shown in Figure 2.22a, for example, the majority of randomly selected houses are located in cheaper neighborhoods and are surrounded by housing patches occupied by low-income agents. However,

when increasing the fraction of public units, a direct mixing effect emerges, even if limited. Indeed, looking at the green plot in the figure — which corresponds to the scenario with the highest fraction of public housing — it is evident that, in the steps immediately following the policy activation, the segregation index slightly decreases. These steps likely correspond to the period in which the newly introduced unoccupied public units get filled. The fact that the index decreases during this phase indicates that public units do help in mixing the two income groups. Still, the effect is relatively small, and after a few steps, the index starts increasing again, although more slowly compared to the dynamics without policy intervention.

Another interesting difference, when compared to the behavior of inner-city median income shown in Figure 3.4, is that while income growth stops after a few steps in all scenarios, the neighbor segregation index continues to rise throughout the entire simulation. Take, for example, the case with 5% rent-gap-selected public

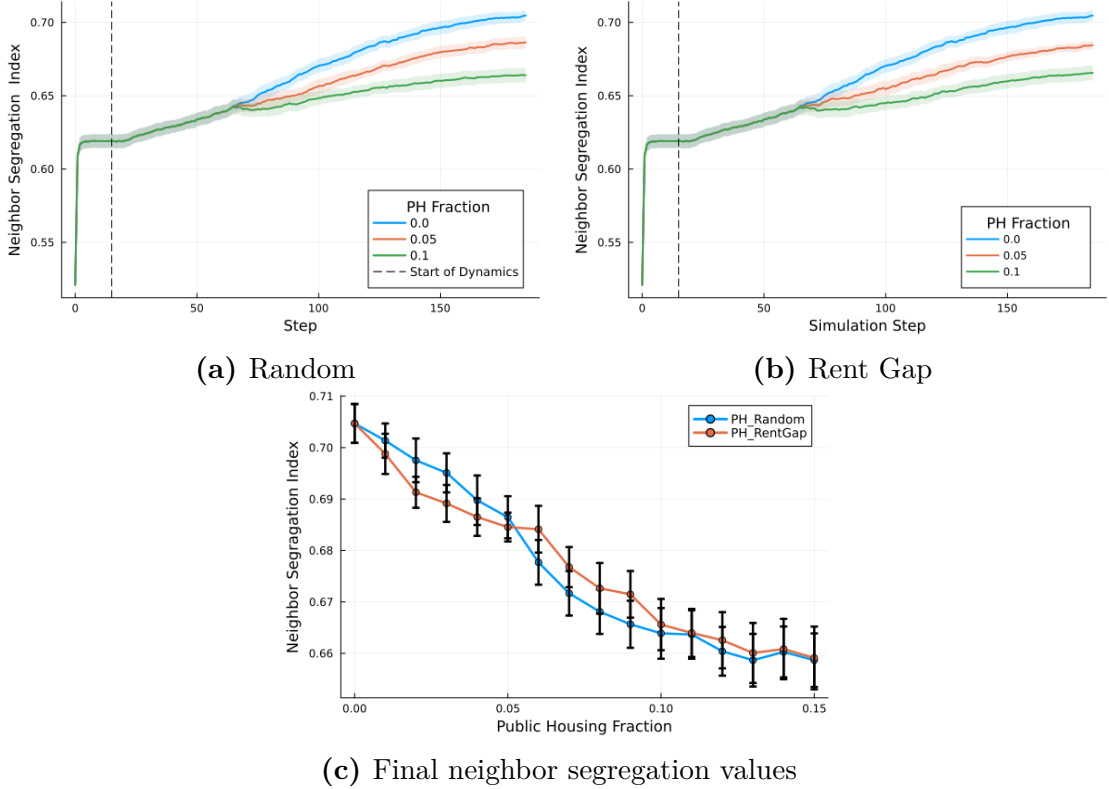


Figure 3.5: (a), (b) Trends of neighbor segregation index, aggregated over 15 simulations, plotted for different fractions of public housing units. Panels (a) and (b) correspond to public housing units selected randomly and based on the rent gap, respectively. (c) Plot of the final values of the index for all the fractions simulated.

housing: the median income in the inner city stops increasing shortly after the intervention, whereas segregation continues to grow significantly (orange plot in Figure 3.5b). If we interpret the stabilization of inner-city income as a sign that gentrification in that area has halted — along with most of the in- and outflows of high- and low-income agents — then the continued rise of the segregation index could suggest that the developer is continuing to invest, probably in parts of the suburbs, since gentrification seems to be associated with an increase in neighborhood segregation. This would also explain why segregation grows more slowly when the policy is in place: by mitigating gentrification, the policy also slows down the increase in segregation.

3.2.2 Dissimilarity index

The second index we study is the dissimilarity, or displacement, index introduced in [39]. It is a very common and effective choice to measure the segregation between two groups of individuals [40]. To define this index, different neighborhoods in a city must be identified. In our case, we divided the 40×40 grid into 16, 10×10 quadrants. The dissimilarity index is defined as:

$$D = \frac{1}{2} \sum_{i=1}^{16} \left| \frac{l_i}{L} - \frac{h_i}{H} \right| \quad (3.2)$$

where l_i and h_i are the numbers of low- and high-income agents present in neighborhood i , while L and H are the total numbers of agents in the two groups. Thus, this index measures the distance in terms of the two groups' shares in each quadrant, and combines them into a single value. It returns a number between 0 and 1: 0 means complete integration and 1 complete segregation. Indeed, in the case where each neighborhood contains the same share of the two groups:

$$\frac{l_i}{L} = \frac{h_i}{H} \quad \forall i$$

Eq. 3.2 returns 0; while in the opposite case, in which the two groups do not coexist in any of the quadrants, it returns 1. The dissimilarity index can be interpreted as the fraction of one of the two population groups that would need to move to a different neighborhood in order to achieve perfect integration. This index provides a broader, more large-scale perspective on segregation compared to the neighbor segregation index discussed in Section 3.2.1. While the latter focuses on short-range contact between individuals of different types (i.e., whether they live next to each other), the dissimilarity index considers whether the two groups are present in comparable proportions within the same neighborhood — regardless of how mixed or clustered they are within it. Clearly it is influenced by how the neighborhoods are defined.

We produced, for the dissimilarity index, the same types of plots previously discussed for the neighbor segregation index (Figure 3.5). These are shown in Figure 3.6, and they display a very different behavior compared to that of the first index. In fact, the dissimilarity index decreases during the initial phase of the process (before the policy intervention), indicating that gentrification is initially making the neighborhoods more balanced. However, in the free dynamics scenario, the index rises again in the second phase, returning almost to its initial value. This behavior can be explained as follows. In the initial thermalized state, the inner-city is highly segregated: it is mainly inhabited by low-income agents (see Figure 2.10b). When the gentrification process begins, high-income residents start moving into the inner city, making its composition more balanced. From the perspective of the dissimilarity index, this corresponds to a decrease in segregation. As the process continues, however, the most attractive areas of the inner-city become heavily gentrified. This leads to a reversed segregation pattern: some parts of the inner-city

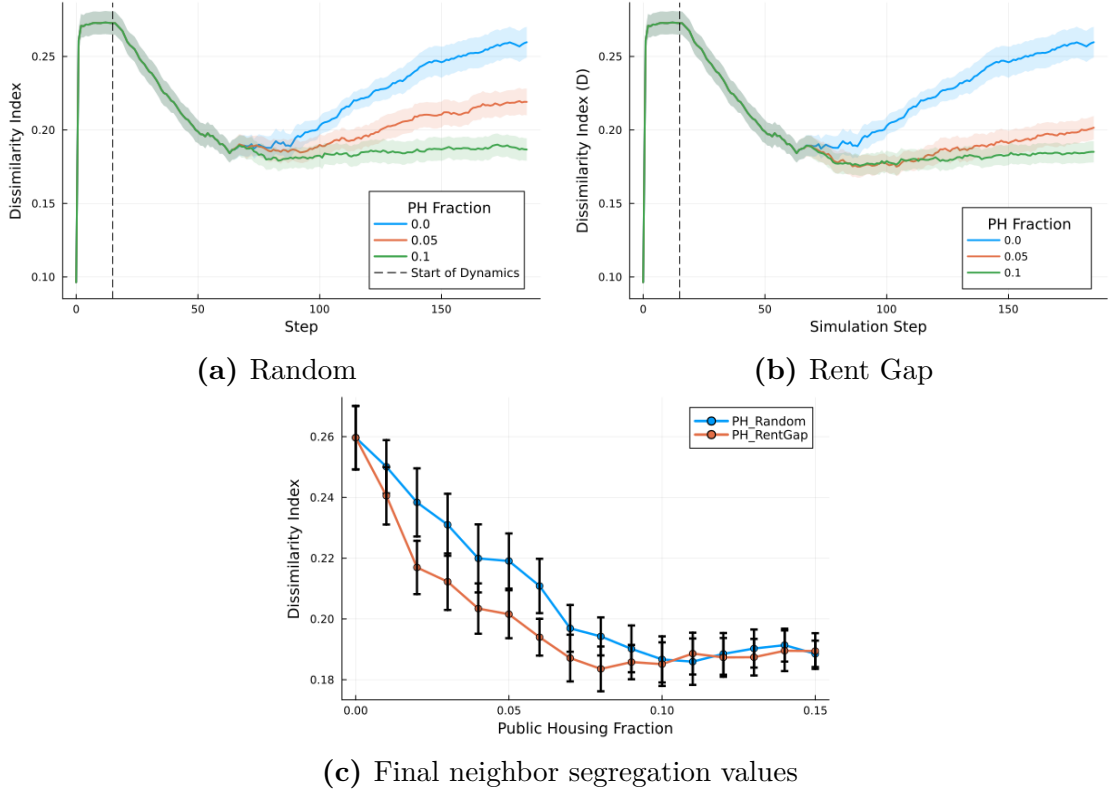


Figure 3.6: (a), (b) Trends of dissimilarity index, aggregated over 15 simulations, plotted for different fractions of public housing units. Panels (a) and (b) correspond to public housing units selected randomly and based on the rent gap, respectively. (c) Plot of the final values of the index for all the fractions simulated.

are now predominantly occupied by high-income agents, while low-income residents are displaced to cheaper suburban areas (see Figure 2.13).

When the public housing policy is activated, Figures 3.6a and 3.6b show that, for a sufficiently large fraction, the dissimilarity index can be effectively stabilized at a low value for both selection methods. In particular, the green curves, corresponding to a 10% share of public units, remain nearly constant after the intervention, both in the random and rent-gap cases. For smaller fractions (e.g., the orange curve for 5%), the rent-gap strategy appears more effective in keeping segregation low. This may seem counterintuitive, as randomly distributing public units might be expected to better counteract segregation from the perspective of the dissimilarity index. However, as previously discussed, the inner city plays a crucial role in the dynamics, and in the rent-gap case, it is specifically targeted with public units. This likely prevents the reversed segregation process that would otherwise cause the index to rise again. That said, according to the final dissimilarity index values shown in Figure 3.6c, the difference between the two policy scenarios is not particularly significant.

Up to now, we have considered only the division between low and high-income agents, since the entire dynamics of the model is based on this binary division, and we measured exclusively the segregation between these two groups. However, as explained in Section 2.3, we sampled agents' incomes from an empirical income distribution, and we defined the two groups by splitting the sample at the 60th percentile (see Figure 2.9). This implies that, within each group, there may be agents whose incomes differ significantly, particularly within the high-income group. Since income makes an important contribution to the agents' utility function (see Eq. 2.8), and thus significantly affects their decision-making process, it is likely that, even within a single group, there are hidden behavioral differences. For this reason, in the final part of the analysis, we decided to further divide the high-income population — which represents 40% of the agents — into two subgroups, by splitting the distribution again at the 80th percentile. It is important to note that the dynamics remain unchanged; the only difference is that, whereas previously we collected data for the high-income group as a whole, we now observe separately the behavior of the 60th–80th percentile subgroup (from now on referred to as the middle-income group) and the top 20% income group (from now on referred to as the top-income group).

Since the dissimilarity index is used to measure segregation between two groups, we can also apply it here to study the evolution of pairwise segregation among the three groups. The interpretation remains the same: it quantifies the fraction of agents from one group that would need to change quadrant to achieve perfect integration with another group. Previously, with only two groups, the index also captured overall (global) segregation. However, in the current case with three

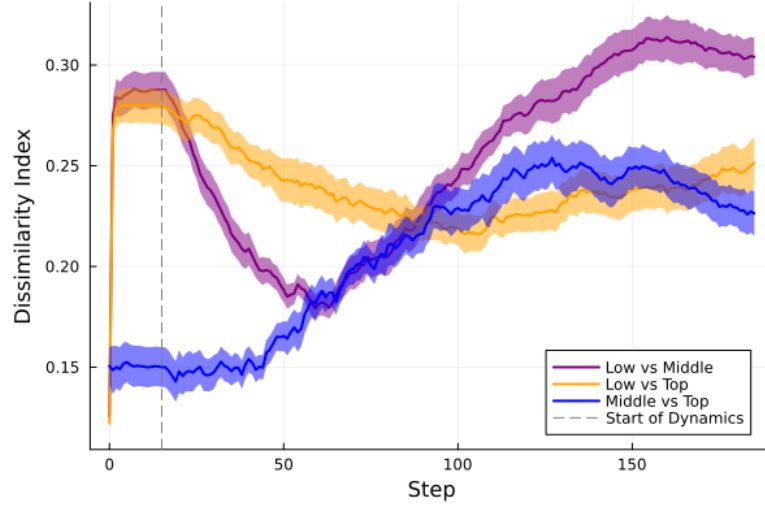


Figure 3.7: Evolution of the pairwise dissimilarity index considering three different income groups, averaged across 15 simulation runs. No public housing intervention scenario.

groups, this is no longer true. Based on the definition in Eq. 3.2, we now compute a separate index for each pair of groups:

$$\begin{cases} D_{LM} = \frac{1}{2} \sum_{i=1}^{16} \left| \frac{l_i}{L} - \frac{m_i}{M} \right| \\ D_{LT} = \frac{1}{2} \sum_{i=1}^{16} \left| \frac{l_i}{L} - \frac{t_i}{T} \right| \\ D_{MT} = \frac{1}{2} \sum_{i=1}^{16} \left| \frac{m_i}{M} - \frac{t_i}{T} \right| \end{cases}$$

where the letters m, M stand for "middle" and t, T for "top" - the difference between small and capital letters is the same as before.

The evolution of these measures in the scenario with no policy intervention is shown in Figure 3.7. As before, we display the average value of the index across the usual set of 15 simulations. From the figure, it appears that among the previously defined high-income groups, the middle-income agents are the ones most involved in the gentrification process. In the first steps, during the thermalization phase, the segregation between top- and low-income agents is comparable to that between middle- and low-income groups, while the segregation between middle and top is very low. This aligns with the initial setup previously discussed: the inner city is mostly populated by low-income agents—resulting in high segregation between them and the other two groups; while middle- and top-income agents are more evenly distributed and well mixed across the suburbs, leading to low segregation. However, at the beginning of the dynamic phase, the purple curve (Low vs Middle) drops rapidly. This suggests that it is mainly the middle-income

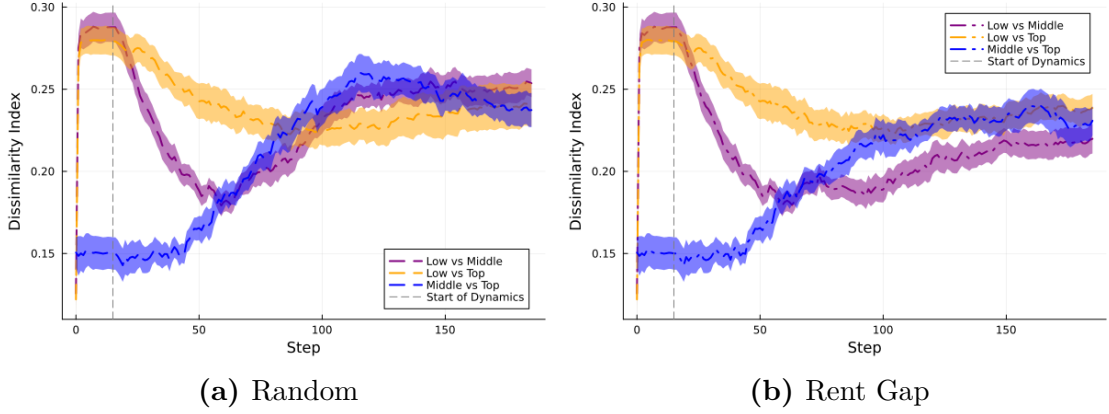


Figure 3.8: Evolution of the pairwise dissimilarity index considering three different income groups, averaged across 15 simulation runs. Public housing fraction set to 0.05. **(a)** Random selection of public housing units. **(b)** Selection by rent-gap.

group that relocates to the inner city during this phase. Subsequently, the purple curve rises again after a few step, reaching a value even higher than its initial level. This indicates that, as low-income agents begin to be displaced from the inner city, they become again segregated from the middle group, which has partially replaced them in the inner-city. Although displaced low-income agents will necessarily move in the suburbs, where we expect to find the majority of top-income agents, not that the Low-Top segregation curve (orange) does not decrease significantly during the dynamics, suggesting that these two groups remain quite segregated also within the suburbs. The increasing segregation between middle and top groups (blue) confirms that these two groups get more separated during the process, as many middle-income agents likely relocate to the inner city, while most top-income agents remain in the suburbs.

Figure 3.8 displays the same plots in the case where 5% of housing patches are converted to public housing, both under random selection and rent-gap-based selection. The main effect is a downward shift of the purple trend (Low–Middle) after the intervention, confirming that these two groups are the most affected by the process. Once again, the effect is stronger in the rent-gap scenario. In that case, even the blue curve (Middle–Top) shows a slight decrease.

3.2.3 Theil index

The last index we discuss is the Theil H index of segregation (also known as multigroup entropy index or information theory index). In [41] many multigroup segregation indices are compared, and the Theil index results the more robust and complete among them. Thus we decided to adopt it since, having now three income

groups, we are dealing with a multigroup scenario. One of the most interesting properties of the Theil index is its additive organizational decomposability. This means that, if a city is divided into multiple regions, the index can be broken down into contributions from segregation between regions and within each individual region. This kind of decomposition is not possible, for example, with the dissimilarity index. Again, we divide the city in 16 quadrants, four belonging to the inner-city and 12 belonging to the suburbs, and we exploit index H decomposability to study how segregation between inner-city and suburbs and within them change during the dynamics.

To define the Theil index H we need the definition of the entropy first. Considering a city populated by K different groups of individuals, we define the total entropy of the city as¹:

$$E = \sum_{r=1}^K \pi_r \ln \left(\frac{1}{\pi_r} \right) \quad (3.3)$$

where π_r is the fraction of agents belonging to the group r . We can define also the entropy in a smaller region j of the city:

$$E_j = \sum_{r=1}^K \pi_{rj} \ln \left(\frac{1}{\pi_{rj}} \right)$$

the only difference is that here the fractions of the groups π_{rj} are restricted to the area j . The entropy measures the "diversity" of an area: it is maximum when all the groups are present in the same proportion, as in that case the region is as diverse as possible. The Theil index estimates segregation by comparing the total entropy of the city to the entropies of its smaller parts. Given a city divided in N regions (for us they are the quadrants, so $N = 16$), H is defined as:

$$H = \sum_{i=1}^N \frac{t_i (E - E_i)}{TE} = \frac{E - \sum_{i=1}^N \frac{t_i}{T} E_i}{E} = 1 - \frac{\sum_{i=1}^N t_i E_i}{TE} \quad (3.4)$$

where t_i is the number of agents in region i and T is the total number of agent in the city. The index returns again a value between 0 and 1. Indeed entropies are ≥ 0 and it is always true that:

$$E \geq \sum_{i=1}^N \frac{t_i}{T} E_i$$

The proof of this inequality is omitted in [41, 42], therefore we derive it explicitly in Appendix A. The index H is equal to 0 when each area has the same entropy of

¹The notation used in this section follows that of [41, 42], in order to maintain consistency and facilitate comparison.

the entire city, so the diversity is the same across the whole system, in this scenario the segregation is minimum; 1 is instead returned if different groups do not coexist in any part of the city.

If the N regions are grouped in $M < N$ clusters (in our case we consider as clusters of quadrants the inner-city and the suburbs, so $M = 2$), the additive organizational decomposability allow us to decompose H in two term, one for the segregation between the clusters and one for the segregation within them:

$$H = H_{\text{between}} + H_{\text{within}} \quad (3.5)$$

where H_{between} and H_{within} are defined as:

$$H_{\text{between}} = \sum_{\alpha=1}^M \frac{t_{\alpha} (E - E_{\alpha})}{TE}, \quad H_{\text{within}} = \sum_{\alpha=1}^M \frac{t_{\alpha} E_{\alpha}}{TE} H_{\alpha} \quad (3.6)$$

where the index α indicates the cluster of quadrants, t_{α} is the total number of agent in the cluster and E_{α} is the entropy of the entire region that makes the cluster. The definition of H_{between} is formally identical to that of H in Eq. 3.4, but the clusters are considered instead of the quadrants. Indeed, the segregation between larger areas is given by the index H computed as if the city were divided into clusters only, ignoring the smaller division into quadrants. Instead, the H_{α} in the definition of H_{within} must be interpreted as the segregation within a single cluster, obtained by computing the index H restricted on the quadrants of the cluster, as if that were the entire city. H_{within} is then given by a weighted average of those values. In our case $\{H_{\alpha}\}$ is made of:

$$H_{\text{suburbs}} = \sum_{j \in Q_s} \frac{t_j (E_s - E_j)}{t_s E_s}, \quad H_{\text{inner-city}} = \sum_{k \in Q_{ic}} \frac{t_k (E_{ic} - E_k)}{t_{ic} E_{ic}} \quad (3.7)$$

where Q_s and Q_{ic} are the sets of quadrants respectively inside the suburbs and inside the inner-city, and all the elements indexed by s or ic refer to the property of the entire clusters.

Substituting the definitions given by Eq. 3.7 in Eq. 3.5 it can be shown that this version of H is the same as the one expressed in Eq. 3.4. Since we did not find the proof in the literature we consulted, we derive it step by step in Appendix B.

Figure 3.9 shows the average trend of the different components of the Theil index defined above, in the scenario without intervention. As in previous cases, the average is computed over 15 random seeds. The black curve represents the behavior of the global index defined in Eq. 3.4, which captures segregation among the three income groups across the entire city. The shape of this curve is qualitatively similar to that of the global dissimilarity index in the no-intervention scenario, shown in

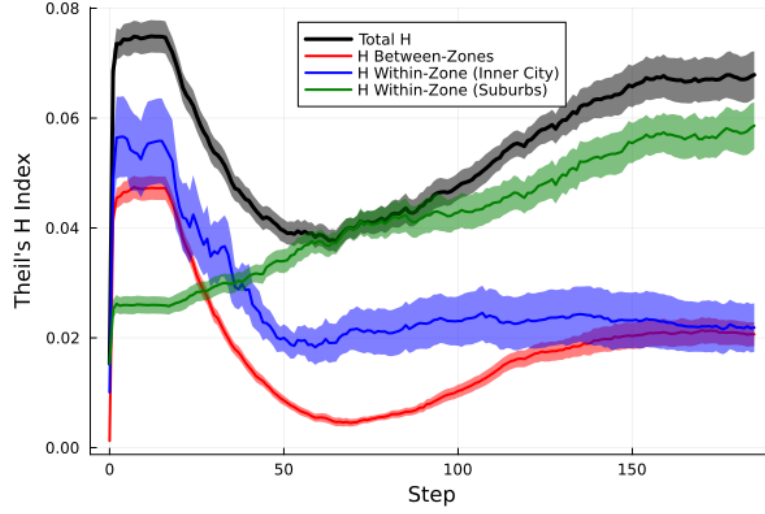


Figure 3.9: Evolution of the different contributions to the Theil segregation index, averaged across 15 simulation runs. No public housing intervention scenario.

blue in Figure 3.6. Although the two indices differ both in their definitions and in the number of income groups considered, they both describe the evolution of a global measure of segregation during the simulation. In both cases, the spatial unit used to compute the index is the 10×10 quadrant, which explains why the two curves exhibit a similar behavior over time. The red curve represents the level of segregation between the suburbs and the inner-city, as defined by the first term in Eq. 3.7. It measures how the populations of these two macro-areas differ from the overall city population in terms of income group distribution. Once again, the trend shows a rapid decrease during the initial phase of the dynamics, followed by a slight increase and eventual stabilization at a value significantly lower than the starting point. This confirms that the most favorable configuration in terms of segregation is reached in the early stages of the gentrification process. This curve corresponds to the lowest levels of segregation, as it reflects the most coarse-grained perspective: it considers only two large areas. Therefore, it is expected not to capture high levels of segregation. The blue and green curves represent segregation within the suburbs and the inner-city, respectively, treating each as an independent city. In the inner-city, segregation again decreases during the initial phase and then stabilizes. Since this index compares the entropy of the entire inner-city with that of its individual quadrants, this behavior can be interpreted as follows: in the early stages of the gentrification process, only a few parts of the inner-city undergo change, leading to heterogeneity among quadrants in terms of income composition. However, after a few steps, gentrification likely extends to all four inner-city quadrants, making them more balanced and maintaining

this balance over time. Segregation within the suburbs stands out from the other curves, displaying a consistently increasing trend from the very beginning. This suggests that, interestingly, although gentrification occurs mainly in the inner-city, it contributes to a growing level of segregation within the suburbs. At the end of the dynamics the suburbs are significantly more segregated than the inner-city, while at the beginning it was the opposite. This increasing trend may be explained by the fact that both the departure of middle-income agents and the arrival of low-income agents are not evenly distributed across the suburbs, but rather concentrated in specific areas. As a result, certain suburban quadrants experience sharper changes in their income composition, becoming progressively more unbalanced over time compared to the entire suburban area. Figure 3.10 displays the same curves in the presence of 5% public housing units, for both patch selection methods. In this case, the plots are not very different from the free dynamics scenario: both display the expected decrease in the total segregation curve (black). Moreover, in the rent-gap case, the final rise in segregation between suburbs and inner city (red curve) does not occur, confirming the effectiveness of this policy in halting the agents flow between the two macro-areas of the city.

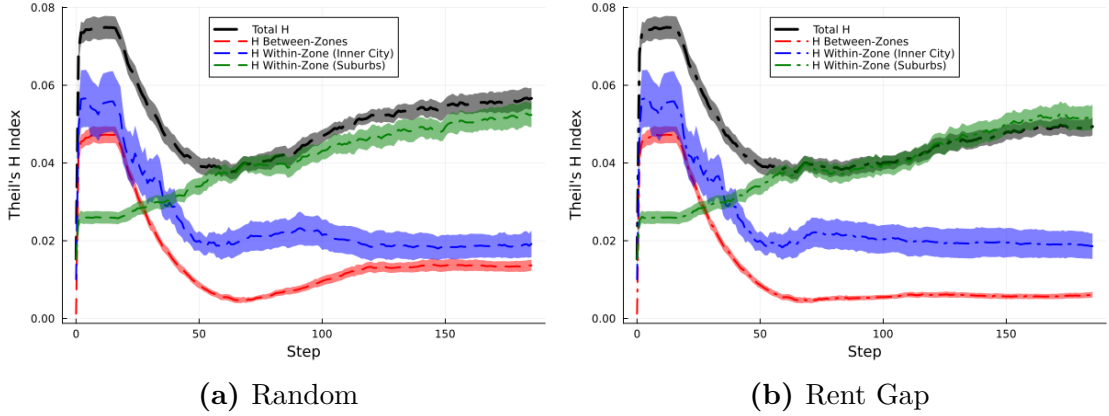


Figure 3.10: Evolution of the different contributions to the Theil segregation index, averaged across 15 simulation runs. Public housing fraction set to 0.05. **(a)** Random selection of public housing units. **(b)** Selection by rent-gap.

Conclusions

In this thesis, we presented the agent-based model we developed to simulate gentrification dynamics inspired by the rent-gap economic theory, typically observed in U.S. cities. The model is designed as a simplified prototype of an American urban environment, intentionally neglecting various real-world aspects such as urban expansion and the housing sales market. The stochasticity introduced in the model setup allowed us to simulate the process across different environments, resulting in a variety of gentrification outcomes. The most representative runs exhibited a growth of common gentrification indicators, such as median rents and incomes, that is comparable to that of real gentrifying neighborhoods. Future work could involve calibrating the model on the actual spatial and economic configuration of a specific city, replicating its distribution of amenities, to explore whether the model can also serve a predictive purpose.

One of the objectives of our model was to investigate the effectiveness of public housing policies in mitigating the effects of gentrification. We simulated two possible strategies for selecting housing units to be made public: randomly, or based on the rent-gap. According to our model, in the rent-gap selection scenario, a small fraction of public units (5%) is sufficient to halt the increase in inner-city median income—a common indicator of gentrification—just a few steps after the intervention. Increasing this fraction further does not produce additional effects in this scenario. The random selection case slows down the process more gradually but remains effective in reducing income growth.

The other main goal of this work was to investigate the impact of gentrification on urban segregation. To do so, we analyzed the evolution of three different segregation indices throughout the gentrification dynamics. The most intuitive among them is the neighbor segregation index, which measures the exposure of agents to others belonging to the same income group in terms of local neighbor links—complementary to the contact with agents from different groups. The analysis of this index revealed that, according to our model, gentrification is associated with a reduction in short-range contact between different income groups, leading to increased segregation. In fact, this index displayed a steady increase over time

across all policy scenarios. Public housing policies, by halting the gentrification of the inner-city, also proved effective in mitigating the rise of this segregation measure.

The second measure we considered is the dissimilarity index, which captures segregation at the neighborhood scale. It represents the proportion of residents from one group who would need to relocate to different neighborhoods in order to achieve an even distribution of the two groups in each part of the city. This analysis revealed that, from this perspective, gentrification is initially effective in reducing segregation, improving the social mix in the inner-city compared to the initial state. However, in the absence of policy intervention, the dissimilarity index rises again in the later stages of the process, eventually returning near to its original value. This suggests that uncontrolled gentrification does not lead to lasting integration between groups, but instead produces a reversed pattern of segregation: low-income residents, initially concentrated in the inner-city, are gradually pushed out and end up segregated elsewhere. By splitting the high-income group into two subgroups — middle- and top-income — and analyzing the pairwise dissimilarity indices among the three resulting groups, we also found that the group most involved in the gentrification process is the middle-income one, corresponding approximately to the 60th–80th percentiles of the income distribution.

Finally, we measured the Theil H index of segregation which, thanks to its decomposability properties, allowed us to separately examine the contributions of within-zone and between-zone segregation. The global version of the index exhibited a trend consistent with that of the dissimilarity index, confirming that uncontrolled gentrification leads to a temporary reduction in segregation, followed by a return to high levels. However, when decomposing the index, we found that, although the inner-city is the area most directly affected by gentrification in terms of housing stock changes, the most severe consequences in terms of segregation occur in the suburbs.

Overall, the model partially challenges the notion that gentrification can serve as a tool to reduce income segregation in our cities—at least when left to evolve without regulation. At the same time, it confirms the potential effectiveness of public housing as a policy instrument to slow down or even halt the effects of gentrification, and to preserve more balanced levels of social mix.

Appendix A

In this appendix, we provide a brief justification of the inequality $E \geq \sum_{i=1}^N \frac{t_i}{T} E_i$, mentioned in Section 3.2.3. This result ensures that the value of the Theil index H remains between 0 and 1. In what follows we will use index r for agents groups and index i for city quadrants, without specifying every time the extremes of the sums.

Recall the definitions of global and quadrant entropy:

$$E = - \sum_r \pi_r \ln(\pi_r), \quad E_i = - \sum_r \pi_{ri} \ln(\pi_{ri})$$

where π_r is the global proportion of group r , and π_{ri} is the proportion within quadrant i .

Defining n_r, n_{ri} respectively as the number of agents belonging to the group r present in the entire city and in the region i , then π_r can be expressed as:

$$\pi_r = \frac{n_r}{T} = \sum_i \frac{n_{ri}}{T} = \sum_i \frac{t_i}{T} \frac{n_{ri}}{t_i} = \sum_i \frac{t_i}{T} \pi_{ri}$$

where t_i is the number of agents in quadrant i and T is the total number of agents in the city.

Substituting the last expression in the definition of E , and naming $a_i = \frac{t_i}{T}$ and $f(x) = -x \ln(x)$ we get:

$$\begin{aligned} E &= - \sum_r \pi_r \ln(\pi_r) = - \sum_r \left(\sum_i \frac{t_i}{T} \pi_{ri} \right) \ln \left(\sum_i \frac{t_i}{T} \pi_{ri} \right) = \sum_r \left[f \left(\sum_i a_i \pi_{ri} \right) \right] \geq \\ &\geq \sum_r \left[\sum_i a_i f(\pi_{ri}) \right] = - \sum_r \sum_i \frac{t_i}{T} \pi_{ri} \ln(\pi_{ri}) = - \sum_i \frac{t_i}{T} \sum_r \pi_{ri} \ln(\pi_{ri}) = \sum_i \frac{t_i}{T} E_i \end{aligned}$$

where we applied the Jensen inequality since the function f is concave and $\sum_i a_i = 1$:

$$f \left(\sum_i a_i x_i \right) \geq \sum_i a_i f(x_i)$$

Appendix B

In this appendix we demonstrate the equivalence between the two versions of the Theil index H provided in Section 3.2.3.

The standard definition of the index is:

$$H_1 = \sum_{i=1}^N \frac{t_i (E - E_i)}{TE} = 1 - \frac{\sum_{i=1}^N t_i E_i}{TE}$$

where the index i runs over all the N quadrants of the city, t_i and E_i are respectively the number of agents and the entropy of quadrant i , E and T are instead the total entropy and the total number of agents.

The second version is given by the decomposition in H_{between} and H_{within} :

$$H_2 = H_{\text{between}} + H_{\text{within}} = \sum_{\alpha=1}^M \frac{t_{\alpha} (E - E_{\alpha})}{TE} + \sum_{\alpha=1}^M \frac{t_{\alpha} E_{\alpha}}{TE} H_{\alpha}$$

where index α runs over the $M < N$ clusters. All the terms indexed by α then refer to the entire region inside the cluster, that will be made by the union of more quadrants. H_{α} is Theil index restricted only on the cluster α , as if it were an independent city:

$$H_{\alpha} = \sum_{j \in Q_{\alpha}} \frac{t_j (E_{\alpha} - E_j)}{t_{\alpha} E_{\alpha}} = 1 - \frac{\sum_{j \in Q_{\alpha}} t_j E_j}{t_{\alpha} E_{\alpha}}$$

The second version H_2 can be expressed as:

$$H_2 = \frac{1}{TE} \sum_{\alpha=1}^M t_{\alpha} [E - E_{\alpha} + E_{\alpha} H_{\alpha}] = \frac{1}{TE} \sum_{\alpha=1}^M t_{\alpha} [E - E_{\alpha} (1 - H_{\alpha})]$$

then substituting the definition of H_{α} we get:

$$H_2 = \frac{1}{TE} \sum_{\alpha=1}^M t_{\alpha} \left[E - E_{\alpha} \left(\frac{\sum_{j \in Q_{\alpha}} t_j E_j}{t_{\alpha} E_{\alpha}} \right) \right]$$

that can be finally reexpressed in H_1 :

$$\begin{aligned}
H_2 &= 1 - \frac{1}{TE} \sum_{\alpha=1}^M t_{\alpha} E_{\alpha} \left(\frac{\sum_{j \in Q_{\alpha}} t_j E_j}{t_{\alpha} E_{\alpha}} \right) = \\
&= 1 - \frac{1}{TE} \sum_{\alpha=1}^M \sum_{j \in Q_{\alpha}} t_j E_j = 1 - \frac{\sum_{i=1}^N t_i E_i}{TE} = H_1
\end{aligned}$$

Bibliography

- [1] United Nations, Department of Economic and Social Affairs, Population Division. *World Urbanization Prospects: The 2018 Revision, Online Edition*. <https://population.un.org/wup>. Accessed: 2025-10-01. 2018 (cit. on p. 1).
- [2] Maureen Kennedy and Paul Leonard. «Gentrification: Practice and politics». In: *Washington, DC: Local Initiatives Support Corporation Center for Homeownership and Knowledge Sharing Initiative* (2001) (cit. on pp. 3–5).
- [3] Peter Marcuse. «Gentrification, abandonment, and displacement: Connections, causes, and policy responses in New York City». In: *Wash. UJ Urb. & Contemp. L.* 28 (1985), p. 195 (cit. on pp. 3, 4).
- [4] David Christafore and Susane Leguizamon. «Neighbourhood inequality spillover effects of gentrification». In: *Papers in Regional Science* 98.3 (2019), pp. 1469–1485 (cit. on p. 3).
- [5] Olatunde CA Johnson. «Unjust Cities: Gentrification, Integration, and the Fair Housing Act». In: *U. Rich. L. Rev.* 53 (2018), p. 835 (cit. on pp. 3, 4).
- [6] Tom Slater. «The eviction of critical perspectives from gentrification research». In: *International journal of urban and regional research* 30.4 (2006), pp. 737–757 (cit. on pp. 3, 4).
- [7] Lance Freeman and Frank Braconi. «Gentrification and displacement New York City in the 1990s». In: *Journal of the American planning association* 70.1 (2004), pp. 39–52 (cit. on p. 3).
- [8] Kathe Newman and Elvin Wyly. «The Right to Stay Put, Revisited: Gentrification and Resistance to Displacement in New York City». In: *Urban Studies* 43 (Jan. 2006), pp. 23–57. DOI: 10.1080/00420980500388710 (cit. on p. 4).
- [9] Guillermo Prieto-Viertel, Mikhail Sirenko, and Camilo Benitez-Avila. *The Ambiguity of 'Balanced Neighbourhoods': How Rotterdam's Housing Policy Undermines Urban Social Resilience*. Oct. 2024. DOI: 10.21203/rs.3.rs-5337680/v1 (cit. on p. 4).
- [10] Yvonne Franz. *Gentrification in neighbourhood development: case studies from New York City, Berlin and Vienna*. V&R Unipress, 2015 (cit. on p. 4).

- [11] Neil Smith. «Toward a theory of gentrification a back to the city movement by capital, not people». In: *Journal of the American planning association* 45.4 (1979), pp. 538–548 (cit. on pp. 5, 6, 9, 11).
- [12] William Alonso. *Location and land use: Toward a general theory of land rent*. Harvard university press, 1964 (cit. on pp. 5, 9, 20).
- [13] R.F. Muth. *Cities and Housing*. University of Chicago Press, 1969. URL: <https://books.google.it/books?id=zrUnzQEACAAJ> (cit. on pp. 5, 9, 20).
- [14] Edwin S Mills. «An aggregative model of resource allocation in a metropolitan area». In: *The American Economic Review* 57.2 (1967), pp. 197–210 (cit. on pp. 5, 9, 20).
- [15] Thomas C Schelling. «Dynamic models of segregation». In: *Journal of mathematical sociology* 1.2 (1971), pp. 143–186 (cit. on pp. 6, 8).
- [16] Thomas C Schelling. «Sorting and mixing». In: *Micromotives and macrobehavior* (1978) (cit. on p. 6).
- [17] Elizabeth E Bruch and Robert D Mare. «Neighborhood choice and neighborhood change». In: *American Journal of sociology* 112.3 (2006), pp. 667–709 (cit. on p. 7).
- [18] Vasco Cortez and Sergio Rica. «Dynamics of the Schelling Social Segregation Model in Networks». In: *Procedia Computer Science* 61 (2015). Complex Adaptive Systems San Jose, CA November 2-4, 2015, pp. 60–65. ISSN: 1877-0509. DOI: <https://doi.org/10.1016/j.procs.2015.09.148>. URL: <https://www.sciencedirect.com/science/article/pii/S1877050915029786> (cit. on p. 7).
- [19] Mark Fossett and Warren Waren. «Overlooked implications of ethnic preferences for residential segregation in agent-based models». In: *Urban Studies* 42.11 (2005), pp. 1893–1917 (cit. on p. 7).
- [20] Victor Stoica and Andreas Flache. «From Schelling to Schools». In: *Journal of Artificial Societies and Social Simulation* 17.1 (), p. 5 (cit. on p. 7).
- [21] Yu Xie and Xiang Zhou. «Modeling individual-level heterogeneity in racial residential segregation». In: *Proceedings of the National Academy of Sciences* 109.29 (2012), pp. 11646–11651 (cit. on p. 7).
- [22] Junfu Zhang. «Residential segregation in an all-integrationist world». In: *Journal of Economic Behavior & Organization* 54.4 (2004), pp. 533–550 (cit. on p. 7).
- [23] Marc Barthélemy. «The structure and dynamics of cities». In: Cambridge University Press, 2016. Chap. 7.3 (cit. on p. 7).

- [24] Luca Dall'Asta, Claudio Castellano, and Matteo Marsili. «Statistical physics of the Schelling model of segregation». In: *Journal of Statistical Mechanics: Theory and Experiment* 2008.07 (2008), p. L07002 (cit. on p. 7).
- [25] Jan K Brueckner and Stuart S Rosenthal. «Gentrification and neighborhood housing cycles: will America's future downtowns be rich?». In: *The Review of Economics and Statistics* 91.4 (2009), pp. 725–743 (cit. on pp. 9–11).
- [26] Steven Jige Quan. «Modeling Gentrification with Theoretical Physics of Complex Systems: Epidemic Spreading». PhD thesis. 2018 (cit. on p. 10).
- [27] Lavinia Rossi Mori, Vittorio Loreto, and Riccardo Di Clemente. «Time-space dynamics of income segregation in the city of Milan». In: *PNAS Nexus* (Sept. 2025), pgaf283. ISSN: 2752-6542. DOI: 10.1093/pnasnexus/pgaf283. eprint: <https://academic.oup.com/pnasnexus/advance-article-pdf/doi/10.1093/pnasnexus/pgaf283/64215629/pgaf283.pdf>. URL: <https://doi.org/10.1093/pnasnexus/pgaf283> (cit. on p. 10).
- [28] Grace KM Wong. «A conceptual model of the household's housing Decision-Making process: The economic perspective». In: *Review of Urban & Regional Development Studies* 14.3 (2002), pp. 217–234 (cit. on p. 10).
- [29] David O'Sullivan. «Toward micro-scale spatial modeling of gentrification». In: *Journal of Geographical Systems* 4.3 (2002), pp. 251–274 (cit. on p. 10).
- [30] Diego Ortega, Javier Rodriguez-Laguna, and Elka Korutcheva. «Avalanches in an extended Schelling model: An explanation of urban gentrification». In: *Physica A: Statistical Mechanics and its Applications* 573 (2021), p. 125943 (cit. on p. 10).
- [31] Giovanni Mauro, Nicola Pedreschi, Renaud Lambiotte, and Luca Pappalardo. «Dynamic models of gentrification». In: *arXiv preprint arXiv:2410.18004* (2024) (cit. on pp. 10, 12, 13).
- [32] Niloofar Bagheri-Jebelli, Andrew Crooks, and William G Kennedy. «Capturing the effects of gentrification on property values: An agent-based modeling approach». In: *Conference of the Computational Social Science Society of the Americas*. Springer. 2020, pp. 245–264 (cit. on pp. 10, 11, 17, 28).
- [33] Avishai Halev. «Modeling Gentrification: An Agent-Based, Amenities-Driven Approach». In: (2018) (cit. on p. 10).
- [34] Jonathan D Shaw, Juan G Restrepo, and Nancy Rodriguez. «Dynamical system model of gentrification: Exploring a simple rent control strategy». In: *Physical Review E* 110.5 (2024), p. 054305 (cit. on p. 10).
- [35] Shipeng Sun and Steven M Manson. «Simple agents, complex emergent city: Agent-based modeling of intraurban migration». In: *Computational approaches for urban environments*. Springer, 2014, pp. 123–147 (cit. on p. 10).

- [36] Martin Blume, Victor J Emery, and Robert B Griffiths. «Ising model for the λ transition and phase separation in He 3-He 4 mixtures». In: *Physical review A* 4.3 (1971), p. 1071 (cit. on p. 10).
- [37] Mats Wilhelmsson, Mohammad Ismail, and Abukar Warsame. «Gentrification effects on housing prices in neighbouring areas». In: *International Journal of Housing Markets and Analysis* 15.4 (2022), pp. 910–929 (cit. on p. 29).
- [38] Shawn M. Carter. «This LA Neighborhood Is the Most Gentrified in the US—and These Other Zip Codes Made the Top 10». In: *CNBC* (Mar. 2018). URL: <https://www.cnbc.com/2018/03/06/la-and-the-other-top-10-most-gentrified-zip-codes-in-the-us.html> (cit. on p. 34).
- [39] Julius Jahn, Calvin F. Schmid, and Clarence Schrag. «The Measurement of Ecological Segregation». In: *American Sociological Review* 12.3 (1947), pp. 293–303. ISSN: 00031224, 19398271. URL: <http://www.jstor.org/stable/2086519> (visited on 09/12/2025) (cit. on p. 52).
- [40] Otis Dudley Duncan and Beverly Duncan. «A methodological analysis of segregation indexes». In: *American sociological review* 20.2 (1955), pp. 210–217 (cit. on p. 52).
- [41] Sean Reardon. «Measures of Multi-Group Segregation». In: *Sociological Methodology* 32 (Dec. 2002), pp. 33–67. DOI: 10.1111/1467-9531.00110 (cit. on pp. 56, 57).
- [42] John Iceland. «The Multigroup Entropy Index (Also Known as Theil’s H or the Information Theory Index)». In: *Retrieved July 31* (Jan. 2004) (cit. on p. 57).

Acknowledgements

I would like to sincerely thank Prof. Vito D.P. Servedio for supervising my thesis project at the Complexity Science Hub Vienna (CSH). His guidance, availability, and thoughtful advice played a key role throughout the development of this work, and I truly appreciated the opportunity to collaborate with him.

I also thank the staff at the CSH for welcoming me into their research environment and for allowing me to take part in a stimulating and valuable experience during my stay in Vienna.

I am grateful to all the professors at Politecnico who have contributed to my academic path over the past two years. In particular, I would like to thank Prof. Luca Dall'Asta, my internal supervisor, for his support and availability during the thesis project.