Exploring Residential Distribution of Income Groups through Agent-based Modeling

Xiaofan Liang
Minerva Schools at KGI

Table of Contents

Abstract

As cities continue to expand in the 21st century, there is a growing need to develop a deeper understanding of the spatial distribution of a city's population. Policy makers are constantly seeking for better models to predict residential patterns of different income groups to identify areas with high social risks, design future infrastructure, and plan affordable housing complex. Agent-based modeling is a powerful tool for revealing system dynamics and understanding how simple individual behaviors following prescribed rules can give rise to emergent properties. This paper presents an agent-based model that focuses on examining the residential dynamics of different income groups. Agents are endowed with heterogeneous income and social capital to choose their residences under a housing market mechanism that responds to income inequality. The model builds upon the traditional urban economic theories and simulates how individuals make residential choices given the tradeoff of between job opportunities and housing price. We found that residential patterns are qualitatively different based on income and these patterns are sensitive to the changes in the total population, housing market responsiveness, and income group ratio. In particular, our model replicates how urban processes, such as gentrification and poverty concentration in the inner-city, emerge when the housing market responds to residents' demands. The model outcomes also reveal power dynamics between income groups such that those with higher income can have more power to impact the locations of others, and sometimes even force them to settle in suboptimal choices. We suggested this model to help urban planners and policymakers to better understand the complexity of residential dynamics, but further extensions are required to make posterior predictions that can be integrated into the planning practices.

Introduction

Why is it important to study where people live?

According to the UN World Urbanization Prospects Report, 68% of the global population will live in cities by 2050. Preparing cities to face such a challenge requires upfront resource planning from a wide range of city stakeholders in both public and private domain, and yet the quantities they rely on to make such decisions —— residential patterns of income groups —— are difficult to predict. Without knowing where different income groups will live, government officials may design public infrastructures that fall short of the need of local residents. A recent trend of the inverse population flow in London, as in the rich heading to city centers and the poor to the suburbs, exemplifies the consequences of such mismatch, causing a shortage of public transit in the suburbs since the infrastructures were originally designed for small and wealthier families but not the poor (O'Sullivan, 2015). Real estate developers and companies also look at income as a crucial indicator for local consumption for various housing, service, and life products. In addition, residential distribution of people with different state of wealth speaks to a city's social sustainability such as integration, inequality, neighborhood resilience, and to what extent space contributes to the exacerbations or improvements of these characteristics. As early as the 1890s, London businessman Charles Booth, who realized the importance of a spatial understanding of population, orchestrated decade-long efforts to construct the first poverty map that colored every street in London according to residents' employment and wealth (Vaughan, 2018). This map enables London government in the 1930s to initiate a series of large-scale slum clearance projects to improve the living conditions of urban poor and expand the public understanding of the social and spatial structures that contribute to poverty.

Why is residential distribution hard to predict?

Despite the interests of all parties in cities, understanding and predicting the dynamics of residential distribution is not trivial. Where people choose to live is more than a personal preference, but also a response underpins the complex interactions of social, cultural, and economic factors embedding in the particular built environment of a city. Researchers from the Chicago school of Sociology in the early 20th century were the first to theorize about residential areas for different social class as part of the land use patterns of cities. One of the paradigmatic land use patterns is the concentric model, proposed by Burgess (1925) to describe layers of symmetric functional zones around the center of business development (CBD). Burgess (1925) relates the formation of "working-men" zone to the social and economic organization of population in cities as waves of new immigrants pushing to integrate into a new urban environment. Building on Burgess's monocentric assumption of cities, economists came up with the bid-rent curve that mathematized the declining gradient of land prices, rents, and density along the distance away from the city center. This distribution emerges from the bids that different people and businesses are willing to pay for the land. It suggests an analytical framework to predict various land uses, including residential locations, and people commonly referred it as the AMM model (Alonso, 1964; Muth 1969; Mills 1972). This analytical framework has proven to be robust and predictive to the changing landscapes of density, land price, and housing price from cities around the world (Bertaud, 2018). Furthermore, Smith (1979) introduced the rent-gap theory that captures the role of negative externalities (e.g., polluted environment, industrial parks, degrading building values over time) at creating concaves in the bid-rent curve, which accounts for one type of deviation that the AMM model encounters in reality. In summary, we can expect that in a monocentric city, when people move further away from city centers (where most of the jobs are), density and housing price per unit will decrease, but transportation cost, land consumption, and floor space per person will increase. People's residential locations can be seen as their personal choices constrained by their income to balance these tradeoffs (Bertaud, 2018).

The exogenous factors that determine the land use patterns and housing markets in a city can also change when a city is experiencing different phrases of urbanization. A key criticism of Burgess's concentric model is the lack of consideration of commute. Hoyt's fan-shaped model of cities (1939) complemented Burgess's model by emphasizing the distribution of different income groups along transportation lines. Multi-nuclei model of cities became more popular as governments increasingly involved in regulating economic activities, establishing top-down urban planning practices, and encouraging industrial parks and their related worker communities to move to the city periphery (Harris & Ullman, 1945). With the expansion of the realm of job opportunities and transportation conveniences in the post-industrialization era, scholars observed wealthier class moving to car-dependent suburbs, resulting in a massive scale of urban sprawl (Bruegmann, 2006).

This income location pattern can also vary dramatically given the historical context a city developed. Most European cities (e.g., Paris) tend to have rich centers and impoverished peripheries, which is different from most American cities (e.g., Detroit) that have poor centers and rich peripheries. The latter is consistent with the standard predictions from the AMM model. Brueckner, Thisse and Zenou (1999) explained this contrast as a result of the distribution of amenities in cities because places like Paris has accumulated a significant amount of cultural facilities in the city center before it has to expand to encompass more population, while most of the American cities were born with the population influx during industrialization.

The majority of the sociological and economic models above have treated the spatial forms of cities as a static phenomenon, while where people choose to locate is mainly a dynamic decision-making process. Sociological theories capture and explain the spatial and temporal changes of income groups based on observations, but lack an analytical framework to develop accurate predictions. The economic models provide an analytical framework, but are criticized heavily as people increasingly realize the limitations of perfectly rational macroeconomic models that do not reflect the reality derived from individual actions. Besides, economic models tend to assume homogeneity in agents to reduce complexity in equations, which is the opposite of what I want to investigate in this project. Agent-based modeling (ABM) provides an ideal playground to study a bottom-up stochastic process of individuals choosing residences and experiment with various parameter combinations.

Agent-based Modeling Approach to Residential Dynamics

Various agent-based models (ABMs) have been used in urban economics to simulate the interactions of agents and space. They can range from the purely theoretical and stylized models to practical ones. The most famous theoretical models include Schelling's model of segregation (1971), the first attempt to illustrate how complex segregated neighborhoods can emerge from simple rules of choosing a location where one-third of the neighbors are of similar race. Benenson (1998)'s model further incorporated the dynamics of cultural dissonance and the cost of moving on top of what Schelling proposed. More recent ABM research work on residential dynamics tend to serve two purposes: one purpose is to explore different conceptualizations of this process in ABM models that help build theories and explanations, such as how may multi-workers households contribute to polycentric city structure (Lemoy, Raux & Jensen, 2017), while the other type is to situate the ABM into a particular residential phenomenon such as slums (Patel, Crooks & Koizumi, 2012) or to calibrate the model to include vectors of metrics for education, transportation, infrastructure etc., and make posterior predictions validated by real-life data (very often GIS data) (Liu, 2010). As Straszhem (1987) pointed out, the classical urban economic models did not consider multiple income classes of bidders because it is difficult to find particular income distribution functions with realistic specification and at the same time, generate tractable results. Agent-based modeling can provide an approximated solution that has proven to converge to theoretical predictions using local search optimization algorithms (Lemoy, Raux & Jensen, 2017). Moreover, ABM also has the advantage of rich extensibility. Any changes to the agent behaviors or spatial settings (e.g., changing a city from monocentric to polycentric) can be rendered in a few lines of code but may be incredibly difficult to express in analytical functions.

Though the more recent work in agent-based modeling tends to include differentiation of agents by income, either as a category or a variable in vectors representing agents' preferences or characteristics, very few focus on spatial distribution of income groups as the outcome, which is the point of this study. Huang, Parker, Filatova, and Sun (2014) reviewed fifty-one ABMs on urban residential choice modeling and summarized three fundamental features introduced by ABMs. This model explores extensions in the first feature, agent heterogeneity, and the second feature, representation of the land-market processes. Since the research question concerns the spatial distribution of income groups, I separate agents into three different income levels. Their behaviors also diverge in terms of the range of neighbors they can search to make residential decisions: richer agents have larger search range than the poor, implying that the rich have more social capital, such as life flexibility, technology, local knowledge, and connections, to navigate the search. The implementation to separate agents in this model is not the most sophisticated type among all the existing ABMs, but is sufficient for this research. Different from other ABMs that either combine land price and housing price into one variable or assign fixed values to both (Huang, Parker, Filatova & Sun, 2014), I set apart land price and housing price so that the former is imposed by the monocentric assumption, while the latter is correlated with the land price while being dynamically update based on the demand and the types of agents reside on it. In another word, I used housing price as a single container to capture the influence of density, income groups (which manifested as their preferences for floor space), and market responsiveness (how quickly will the price raise given the demand) in residential choices.

The reason for such design rests on the goal of this agent-based model, which is to simulate the basic “push” and “pull” economic forces — job opportunities and housing price —in cities that motivate people to move. From classical urban economic theories, we can expect that people will be pushed toward the city center to secure higher job probability and pulled away from the city center due to the housing price. Therefore, the distribution of income groups may seem to be straightforward: rich people will take over the city center because it has the highest job probability and only they can afford the price, while the poor go to the periphery. However, in reality, we observe a much more mixed-income population in the city center. Besides, we do not know how the total population in the city or the proportions of different income groups will change these dynamics. Therefore, the model used in the present work will simplify city spatial structure as a monocentric form and explore the dynamics of the uncertainties above.

Methods

Model Introduction

The model in this paper can be conceptualized as a spatial agent-based model of a city in a confined space (no immigration) with agents (people) and patches interacting through space and time. Every agent has only one functional attribute, which is income. Each patch has four attributes: 1) land price, 2) job probability, 3) density and 4) housing price. Similar to income and housing price, land price is also a categorical variable with same value range and is assigned when the model is initialized to follow a concentric distribution (high in the center and low in the periphery). Job probability is static at each patch with a fixed continuous value sampled from probability distributions with varying mean and variance. The hierarchical model that determines the values of such means and variance is an input to the model when it is first initialized. Density is updated every step to calculated the number of agents on top of the patch. Housing price at patch i is a composite variable recalculated at every step as follows:

$$H_i = K * L_i * \sum_{j=1}^{N_i} I_j$$

H_i : Housing price at patch *i*. $0 <= H_i <= 3$
K : Market responsiveness constant. $0 <= K <= \frac{1}{3}$
L_i : Land price at patch *i*. $L_i \in [1, 3]$
I_j : Income of agent *j* at patch *i*. $I_j \in [1, 3]$
N_i : Density (total number of agents) of patch *i*.

In this equation, housing price is a continuous variable ranging from zero to three. When no one resides on a patch, housing price will be zero, and when the demand is high, the price will be added to maximum. Market responsiveness constant (K) is also continuous ranging from zero to one-third. When K = 0, the housing price is equivalent to nothing, but K cannot exceed one-third because if K is larger than one-third, not a single rich people can move into the most expensive land. Land price and income are both discrete values between one and three. Density is also an discrete integer that can be as small as zero and as large as the total population.

In short, both endogenous factors such as the land price, density, and resident compositions on that patch, and exogenous factors such as housing market responsiveness determine the housing price of a particular patch. If the demand of a patch is high, the resulting housing price can exceed the land price of the patch and vice versa. The housing price formula captures a few salient dynamics of the housing market in reality:

1) Housing price is partially driven by the density of people on the patch as embedded in the sum of agents' income. Higher density (as reflected by a larger sum of income) signals higher demand, which generally leads to the raising housing price (see Figure 1). This correlation may sound counterintuitive because, with a fixed piece of land, a denser population should result in a lower land price per person. However, we know that the city center is often compact but also expensive because of the high demands. Therefore, this equation considers density as an indicator for demand, but also allows different income groups to have separate effects on housing price (see next point).
2) Housing price is also driven by the composition of residents (as reflected by sum of income) living on the patch. Increasing the density of rich people rather than poor people should have a differentiated effect on the housing price. Another way to look at the sum of agents' income is to think about agents' income as equivalent to agents' floor space consumption. Therefore, rich agents with higher income will translate into larger floor space consumption and thus will have greater impacts at raising the housing price when they join the patch. This mechanism also simulates the gentrification process as housing price that used to be affordable to the poor agents will be raised above the affordability limit as richer agents move in.
3) Housing price also mediates how land price is translated into the actual housing market, and thus the affordability of a patch to different income groups. For example, in the equation, poor agents can still reside on patches that have the high land price (which often locate at the city center) as long as the demand for that land is low (may be due to negative

externalities). Housing price also raises quicker for more expensive lands when people start settling in (see Figure 2).

4) Market responsiveness constant represents how sensitive the housing price is to the density and composition of the residents. The larger the K, the faster the housing price changes with an additional person added to the patch. (see Figure 2).

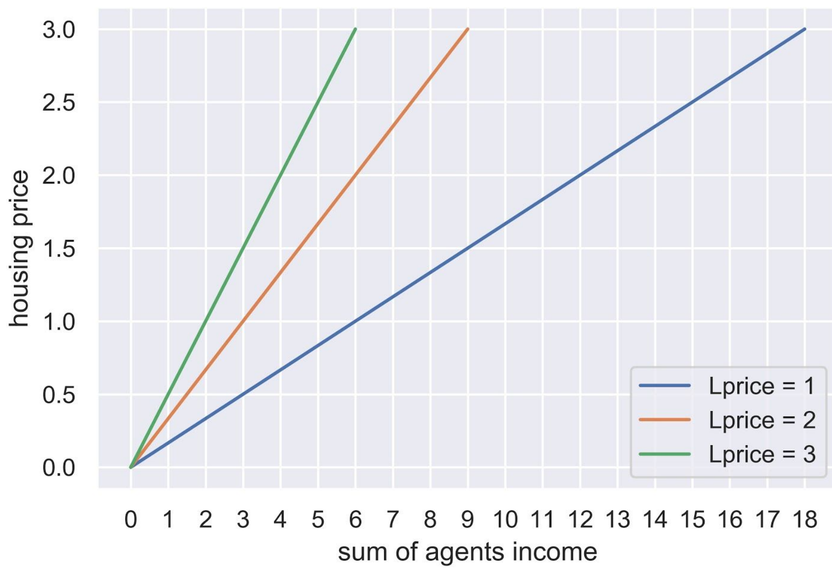

Figure 1: Housing price versus sum of agents' income at a different land price with fixed market responsiveness constant (K = ⅙) according to the housing price equation.

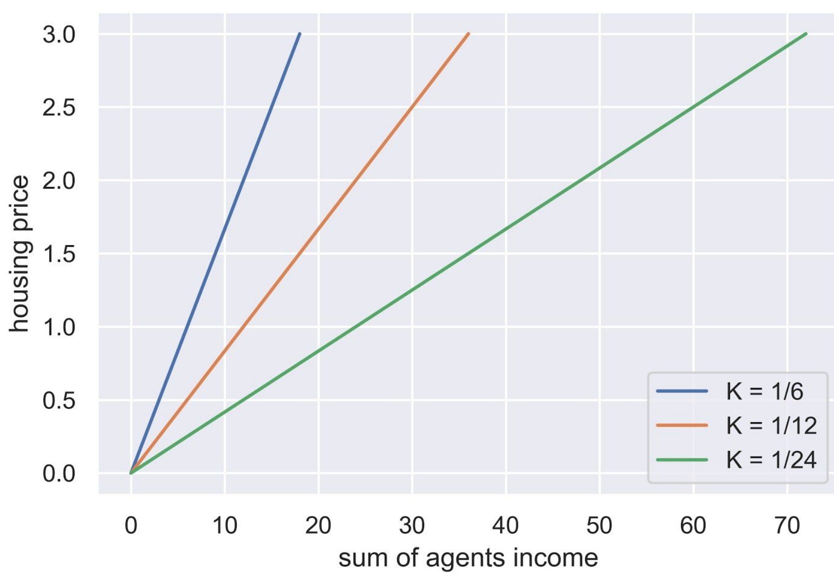

Figure 2: Housing price versus sum of agents' income with a different value of market responsiveness constant (K) with a fixed land price (Lprice = 1) according to the housing price equation.

Model Assumptions

A fundamental assumption of this model is the monocentric structure that I imposed to influence the distribution of land price and job probability. The gradient of land price and job probability follows a linear descent that roughly aligns with the bid-rent curve from the classic AMM model (see Figure 3 and 4). The parameters of functions that distribute land price and job probability are fixed in the current analysis. The sampling probabilities for land price from one to three at each distance is (0, 0.2, 0.8), (0, 0.4, 0.6), (0.2, 0.3, 0.5), (0.6, 0.2, 0.2), and (0.8, 0.2, 0), which essentially assigns higher probability of high land price when the distance is close to CBD and decreases such probability when the patch is further away. Job probability is sampled from normal distributions with varying mean but the same standard deviation based on the distance. From center to periphery, the mean and variance of the distributions drops linearly from (1.15, 0.1), (0.9, 0.1), (0.7, 0.1), (0.5, 0.1), to (0.3, 0.1). Noticed that at distance zero, the job probability is higher than one, which is used to guarantee almost 100% job opportunities in the city center and thus extend the chances that people can stay for longer. Such a design removes the stochastic moving at the center and allows the model to converge faster. There could exist better probability functions for the land price and job probability to create more realistic distributions (see #distribution for more details), but this model is sufficient to match the concentric assumptions adopted from the classic AMM model.

Moreover, housing is assumed to be the only good in the market that agents will spend all their income on. Under this simple framework, values of income and housing price are normalized so that they can be easily compared to figure out affordability. However, this housing equation inevitably omits commute cost, which is widely cited as an impact factor to housing price in residential choice models (Fujita, 1989). Accounting for commute cost requires expanding income (or budget) into a composite variable, which calls for a more nuanced utility function with more parameters to accommodate preferences and a different people-job interaction than what is encoded in this model. Since the focus of the model is to explore residential dynamics of different income groups, tradeoffs are made favoring simplistic representations.

Figure 3: Average distribution of land price along distance to CBD over 100 simulation runs, with 95% confidence intervals.

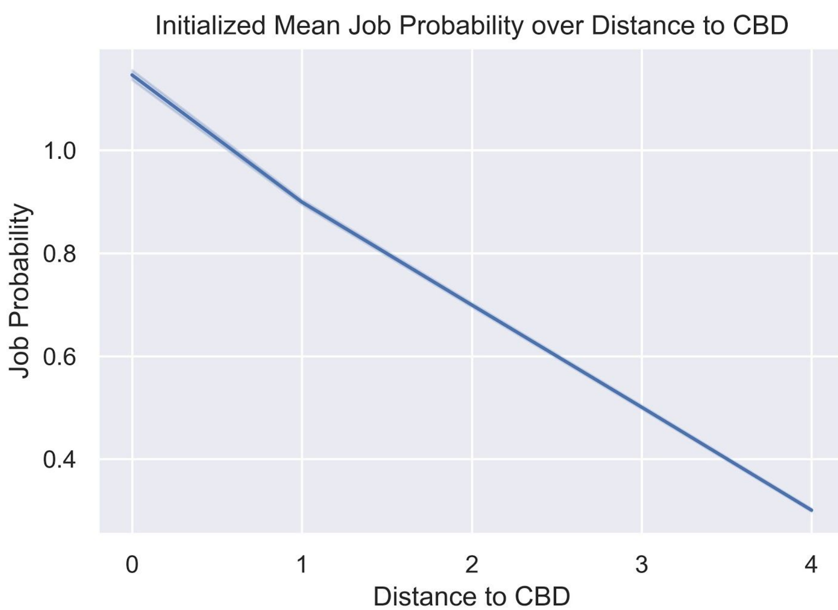

Figure 4: Average distribution of job probability along distance to CBD over 100 random initializations, with 95% confidence interval.

Model Interface

The model interface is coded in Mesa, an open-sourced Python library that is built for agent-based models (Masad & Kazil, 2015). The biggest advantages of Mesa compared with other agent-based modeling software (e.g., NetLogo) is that it is based in Python and it automatically generates a browser interface for simulation. It is also easy to debug in the console, extend the source files to customize the outputs, and output data collected during the model runs as pandas data frames.

When the model is initialized, agents in different income groups are created according to the proportion defined by users and placed in a spatial grid randomly. Each patch of land is assigned a land price and a job probability. Figure 5 shows four canvases that update at the same time and display different information in the background. From left to right, the canvases represent patches' land price, job probability, density, and percentage of income groups. Circles represent agents, while squares represent patches. The red color represents the poor, the yellow represents the middle-income group, while the green represents the rich, with the dot size decreasing so that multiple populations can be seen on a single patch. Therefore, if different types of agents overlap, they will be shown as layered circles. Lighter grayscale colors refer to higher values, which can be higher density, job probability, or land price. Land price layout only has three color gradient, corresponding to the categorical nature of the variable. Job probability is continuous, and thus the layout shows more varieties in color.

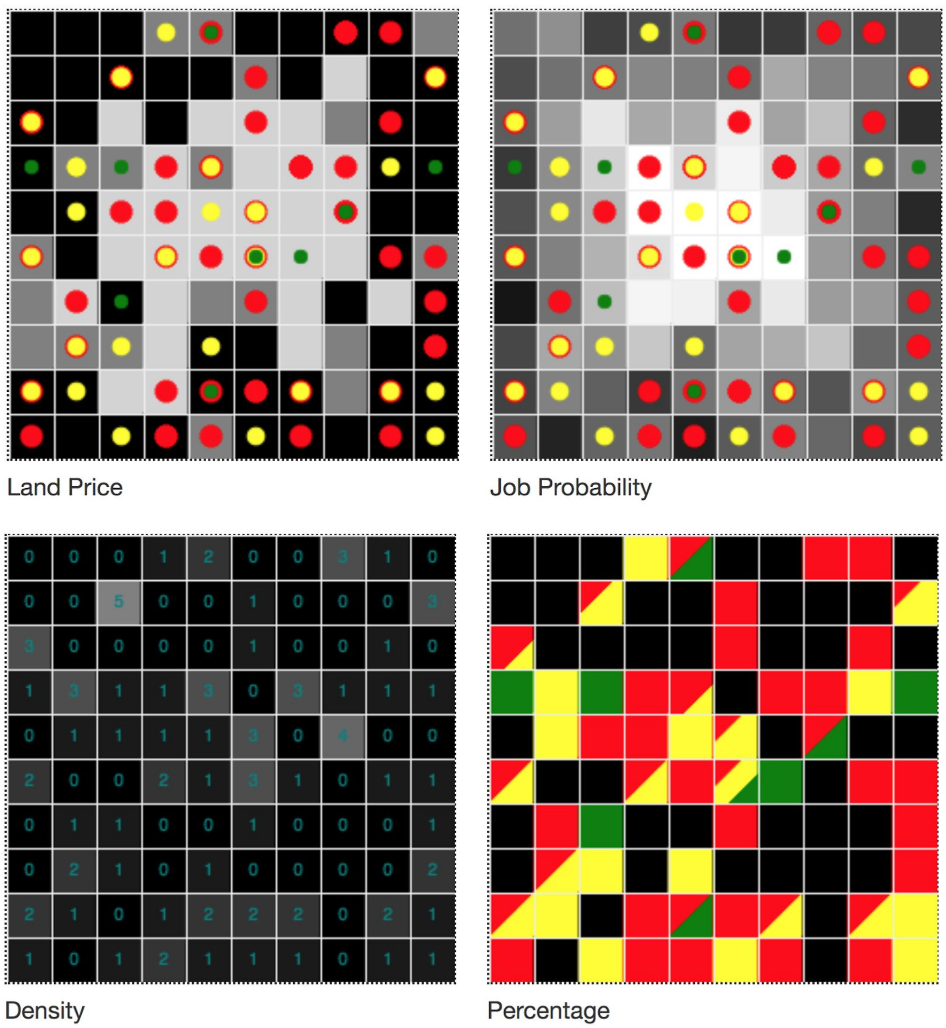

Figure 5: The simulation interface when the model is initialized.

Model Logic and Behavioral Rules

At each time step, each agent will be activated in a random order and follows the decision tree (see Figure 7) to decide on its action based on information of patches from the last time step (which essentially creates simultaneous update among all agents). The model will first evaluate whether the agent is "happy", which means the agent's income is greater or equal to current patch's housing price, and it can secure a job according to the current patch's job probability. If all these evaluations are satisfied, the agent can stay at the current spot and wait for the next step. If not, then the agent will pick a random neighbor within its searching radius based on their income levels. The target neighbor will be available to move in if 1) the target patch's (projected) housing price is still affordable after the agent has joined, and 2) the agent will win a job by rolling a dice according to the job probability at the target patch. If this random neighbor failed, then the agents will choose another one until all neighbors ran out. In that case, if the agent still cannot find an option within three steps, it will be removed from the model forever. If the agent finds a satisfactory patch, it will move to the target patch and wait for all other agents to finish. After all the agents have progressed, the patches will update their density and housing price.

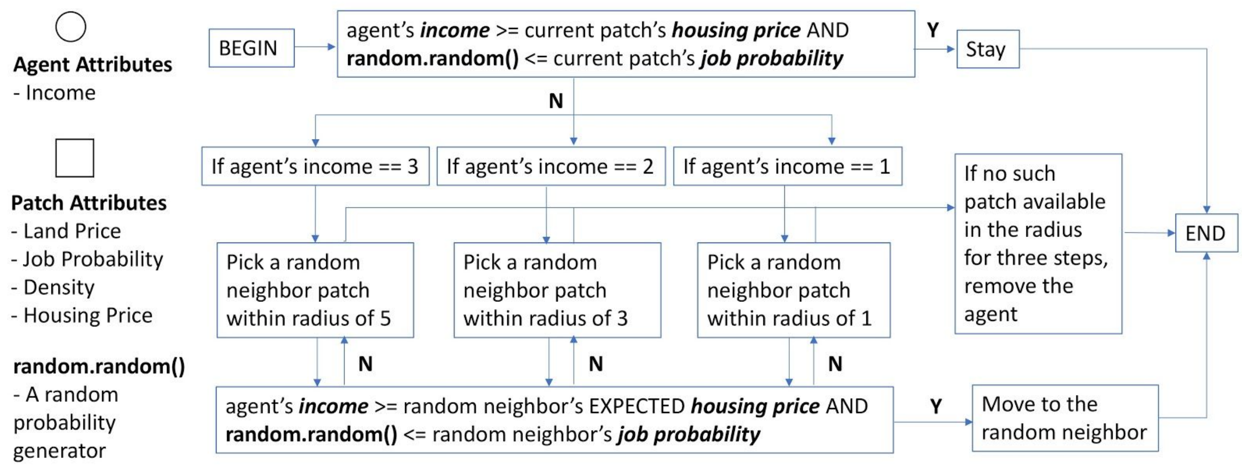

Figure 6: Decision tree of each agent's behavioral rules for one time-step.

Model Outcomes and Metrics

Only one type of outcome is studied, which is the spatial distribution of different income groups. We applied two different metrics to analyze the spatial dynamics of income groups under different model parameters. Since the model is concentric, patches tend to have similar statistics (e.g., land price, job probability) at the same unit of distance with respective to CBD. Both metrics simplify 2D spatial representation (e.g. Figure 5) into 1D by flattening spatial coordinates into units of distance away from CBD. The first metric is the normalized median percentage of people within their own income groups over distance away from CBD. It shows where most of the people within a certain income group locate and whether people from different income groups overlap on their locations or not. The second metric is median density per patch by distance, which can be further broken down into absolute numbers of people in each income group at each distance. It shows how the overall population distribution change and what are the income compositions look like at each distance. We examined these two metrics under various model parameters, including the number of agents, the market responsiveness constant, and proportion of different income groups. These parameters are chosen because they have direct impacts on housing price, and thus suggest a good first step to examine the model dynamics set up by the equation. Future extensions can be made to vary the initial distribution of jobs or land price, which will be harder to track analytically from the equation.

Results

Each simulation was run for 100 steps to average the temporal outputs. For each set of parameter values, 100 simulations were run so that the data could be independent of the initial randomization and the stochastic agent moving processes. Based on how the model is set up, we hypothesized that if the housing market works correctly according to the equation, qualitative difference in income should translate into a differentiated spatial distribution of income groups based on affordability. The results confirm such prediction, and it is robust under a variety of parameter settings: rich people do tend to live at the city center with higher land price, followed by middle-class people, and lastly the poor at the periphery. Density also persists in concentrating in the center and dispersing in the peripheral.[1] However, having most people from one income group at a particular distance does not necessarily makes them the dominant population at that distance. The absolute compositions of residents have to do with both their normalized median percentage within their income group and the percentage this group takes up in the total population. In the model setting where rich people are only 10% of the total population, even if they concentrate at the city center with the highest in-group percentage, their absolute number is still overwhelmed by the middle-class residents, which projects a very audience group to urban planners. Moreover, population overload, sensitive housing market, or high ratio of rich people, can amplify the economic and power inequality between different income groups, creating market failure and causing the low-income or even middle-class population to locate at places deviated from their optimal choices.

How Population Changes Spatial Distribution of Income

When population increases in a confined space, people from all income groups move slowly to the city periphery but in a different order (see Figure 7). The low-income group moves first, then follows the middle income group, and lastly the high-income group. The housing price dynamics can explain this observation as we expect that when the population increases, demand for housing in the market is higher and thus the actual housing price will rise quickly. Therefore, different income groups start to have differentiated residential locations, corresponding to their ability to afford the land price.

However, such differentiation does not translate into complete segregation. In fact, when the total population rises, the city center becomes increasingly polarized with both high and low-income groups. This counterintuitive result has to do with both the housing price dynamics and the searching radius rendered from income. Considered the model setting in Figure 8 as an example, ever since the total population grows beyond 360, rich people stop moving into the city center, and poor people start to accumulate (see Figure 8). This threshold number is partially an artifact of the housing price equation as we expect a maximum of five rich people at the city

---

[1] Density per patch at the particular distance is normalized by the number of patches at that distance. The number of patches for each distance away from CBD in a 10 * 10 grid is 4, 12, 32, 60, 96.

center with a market constant of fifteen, which is evident in the subfigure[2] with N = 360. However, the housing price should be unaffordable for the poor at the city center at this point, so how do they manage to stay? The answer is that they are trapped. Most of the poor people at the center were assigned to be there as the model initialized, but since the population density is so high that the housing price is raised to the maximum immediately, they cannot move to cheaper land further away as they only look for neighbors within one unit of the searching radius. Though they stay unpleasantly at the city center and will eventually be removed by the model, poor people still occupy a portion of the limited floor space when they are still present. As a result, some rich people are forcefully pushed away from the city center despite their ability to afford the land and their motivations for the higher job probability. The tension between the rich and the poor at the city center also has a spillover effect on the middle income group. Different from the poor, the middle income group can escape from the center when the price became unaffordable, but as the rich people start to flood into medium distance areas and lift the housing price, the middle income group is further pushed into the more affordable patches at the city periphery.

Looking at the absolute resident composition by distance also generates a different narrative. Despite most of the rich people occupy the city center while population changes (see Figure 7), they do not always have the dominant number. In the particular model setting used in Figure 8, it is only when the total population increases beyond 270 that the rich replace the middle-class and become the main demographic group at the city center. This is because rich people only share a small percentage of the total population as defined by the model parameters, and thus their only way to be the majority at a particular distance is to raise the housing price enough to be exclusive. To achieve this goal, the population of the rich needs to pass a certain threshold which can be varied by the market responsiveness and the initial ratio. This threshold can be deduced analytically from the housing equation given the values of all other variables.

[2] This is how we deduce the maximum amount of people at a particular patch: land price at city center is roughly 3, K is 15, and thus the maximum sum of income is 15, which means it can accommodate at maximum five people with income equal to 3 before the land price goes over 3.

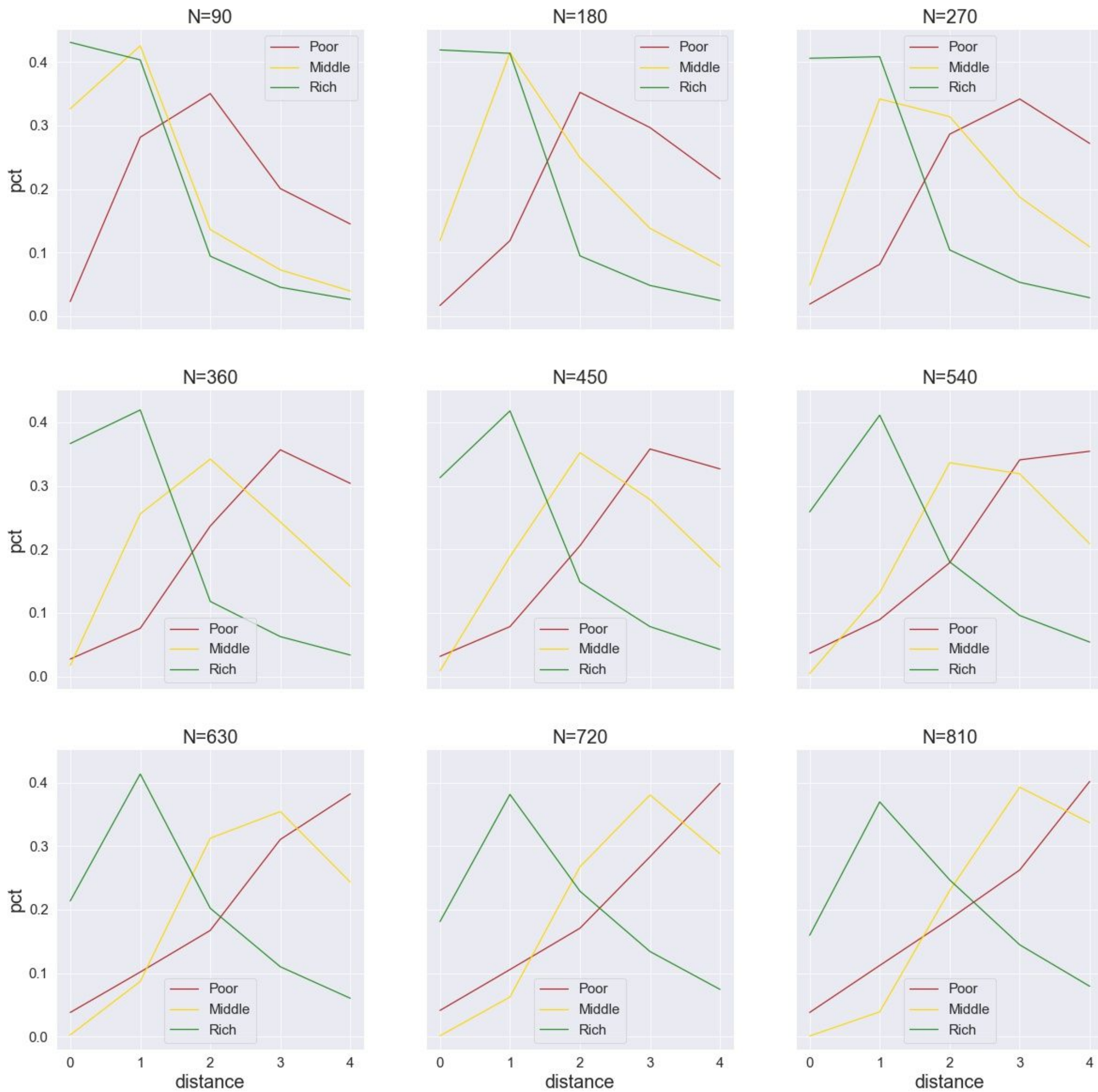

Figure 7: Normalized median percentage of population within each income group versus distance away from CBD, under different values of N (number of agents), with K= $\frac{1}{15}$ and ratio of poor, mid, and rich = 5:3:1.

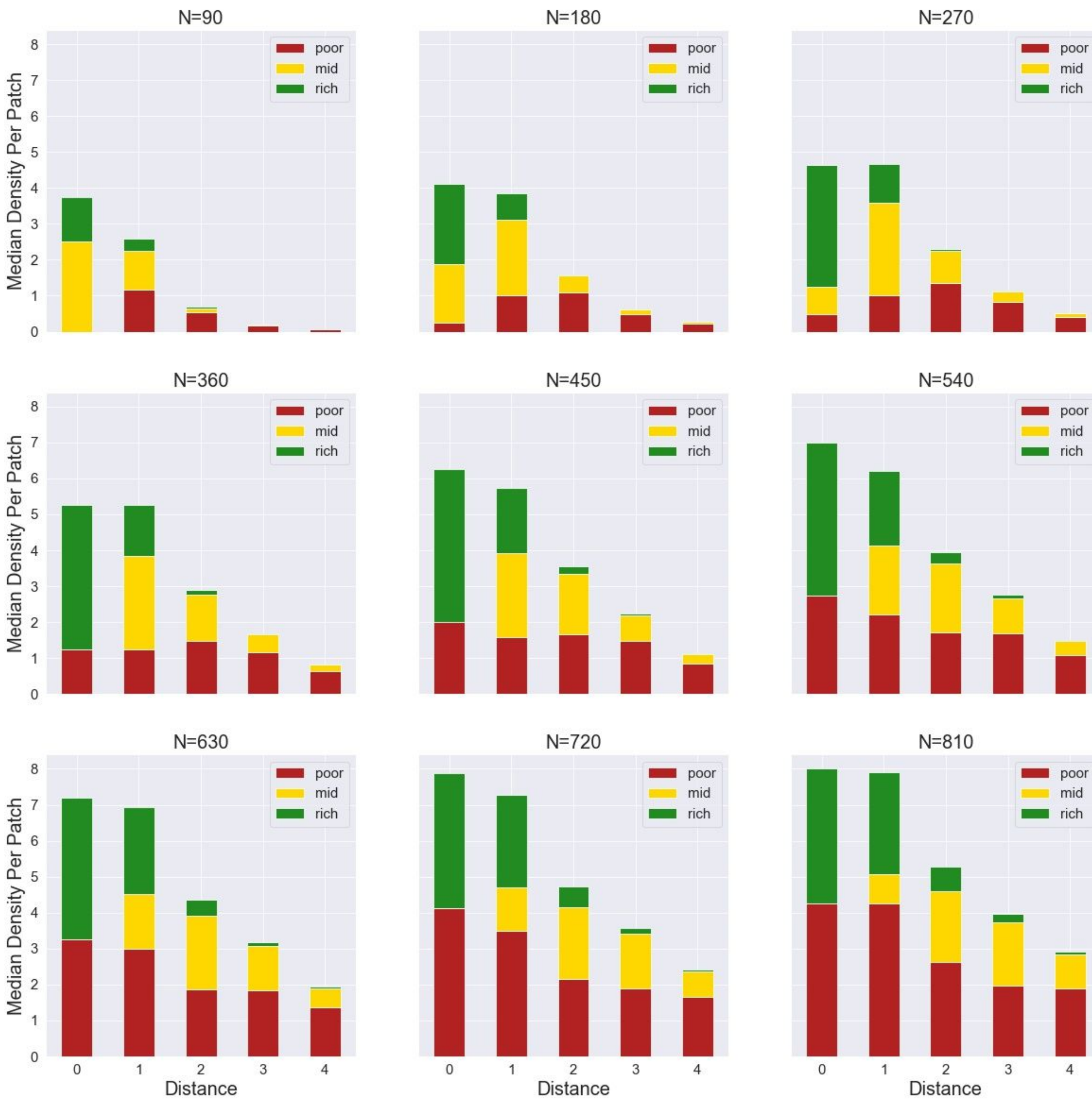

Figure 8: Median density per patch broken down in each income group versus distance away from CBD, under different values of N (number of agents), with K = $\frac{1}{15}$ and ratio of poor, mid, and rich = 5:3:1.

How Market Responsiveness Constant Changes Spatial Distribution of Income

From the housing price equation, we can predict that when the market responsiveness constant is small or even close to zero, all agents will want to move to the city center as it effectively removes the housing price constraint. On the contrary, when the market responsiveness is large, the housing price of the patch will converge to the maximum quickly, and thus agents will move to affordable patches that align with their income levels. The subfigures with K in a decreasing order in Figure 9 confirmed this prediction, as we observe clear differentiation of location choices with larger K, and clustering at the city center with smaller K. In the process of converging to the center, the advantage of social capital, rendered by the radius of search range, allows the majority of the rich people to first move into the center, then followed the middle income group, and lastly the poor. It means that even if the housing market becomes completely affordable for everyone, the inherent power dynamics still dictate who will have the cake first.

Increasing market responsiveness constant also has a similar effect on housing price as the population influx. Comparing subfigures from $K = \frac{1}{5}$ to $K = \frac{1}{20}$ in Figure 10, we can see that poor people can locate in the city center when $K = \frac{1}{5}$ and $K = \frac{1}{20}$ and beyond but disappeared when the value of K is in the middle of the range. The reason is that there are two different mechanisms at play determining whether the poor have a place to stay. When $K = \frac{1}{5}$, the market is so responsive that the housing price escalates within a few steps, resulting in the same "poverty trap" discussed in the case of population overflow. When $K = \frac{1}{20}$, market is so not responsive that the housing price is still affordable for the poor despite the presence of a large amount of middle and high-income people.

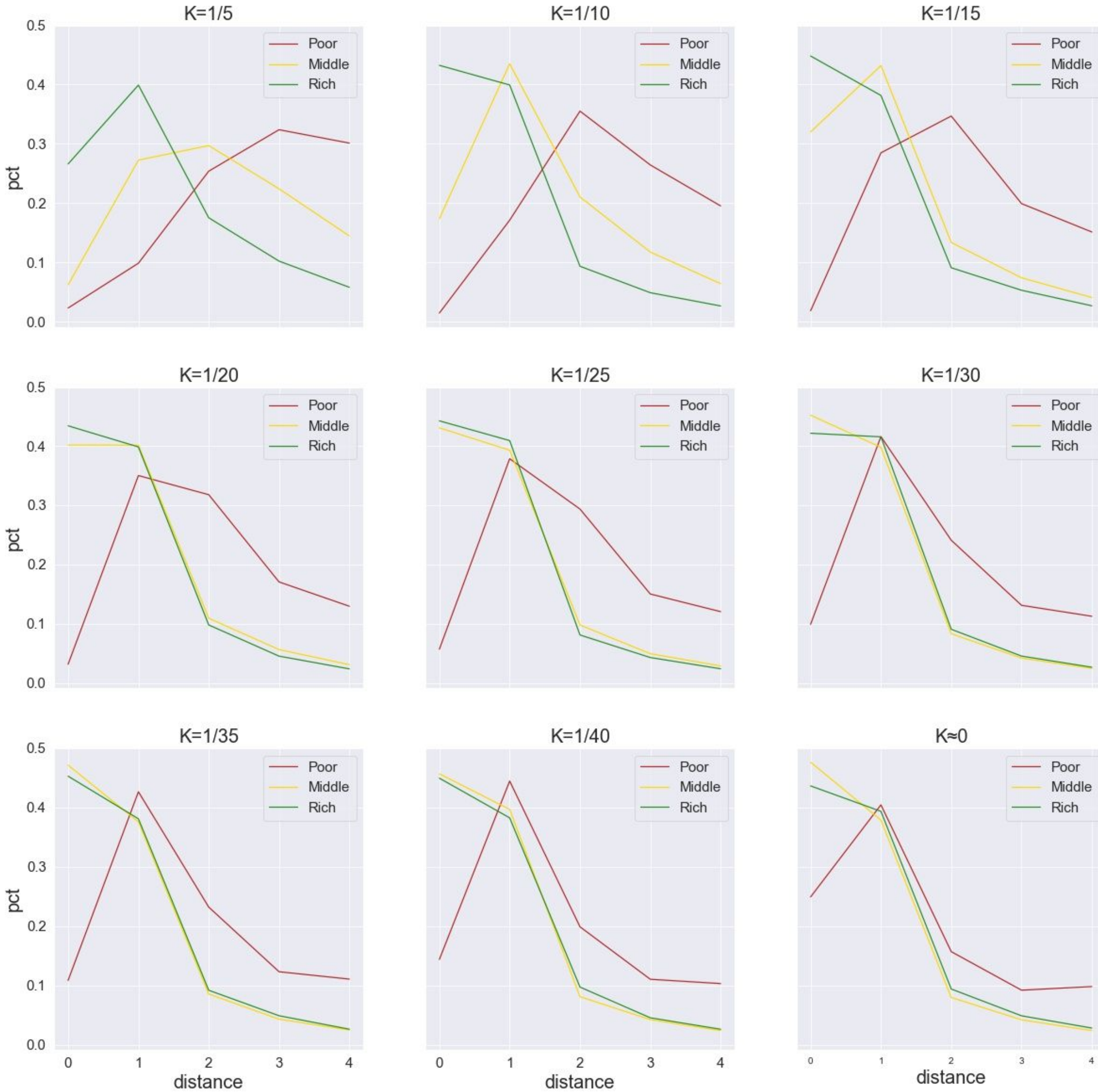

Figure 9: Normalized median percentage of population within each income group versus distance away from CBD, under different values of K (market responsiveness constant), with N = 90 and ratio of poor, mid, and rich = 5:3:1.

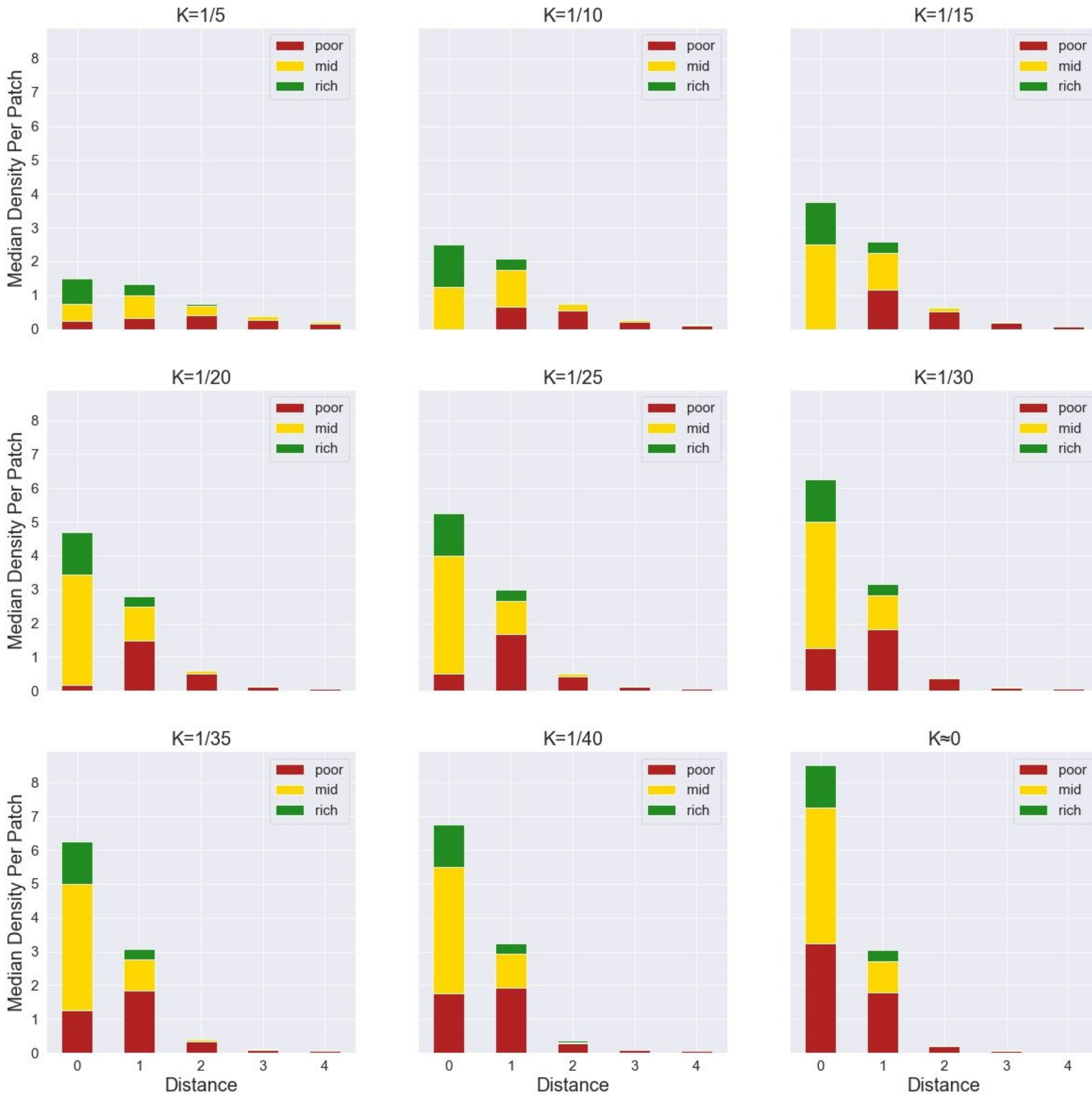

Figure 10: Median density per patch broken down in each income group versus distance away from CBD, under different values of K (market responsiveness constant), with N = 90 and ratio of poor, mid, and rich = 5:3:1.

How Income Group Ratio Changes Spatial Distribution of Income

Income group ratio impacts the average composition of residents on patches and the sum of their income levels. Given the housing price equation, a higher proportion of wealthy people means that more of them will come to the city center as they can afford it and make it unavailable for other population. Figure 12 shows that the city center is completely segregated by the high-income community when they have the dominant ratio of the entire population. The impacts of their presence also spill over to areas at one unit of distance where the housing price becomes much less affordable for the other two income groups. In particular, it affects middle income population more than the lower income one (see Figure 11 with ratio = 1:1:7 and ratio = 1:3:5).

Increasing the percentage of middle income groups does not seem to impact the distribution of both rich and poor population compared to the baseline ratio (see Figure 11 with ratio = 1:7:1 and ratio = 3:3:3). Higher ratio of the low-income group (see Figure 11 with ratio = 7:1:1 and ratio = 5:3:1) seems to create win-win situations for all income groups as in the housing price is quite affordable at all distance, allowing poor people to share some land in the city center as well as the nearby areas. Nonetheless, no matter how the ratio changes, there is still a qualitative differentiation of where the majority of people in each income group will live (e.g., rich in the center, middle income people in the middle, and the poor at the periphery), which roughly follows the land price distribution along the distance.

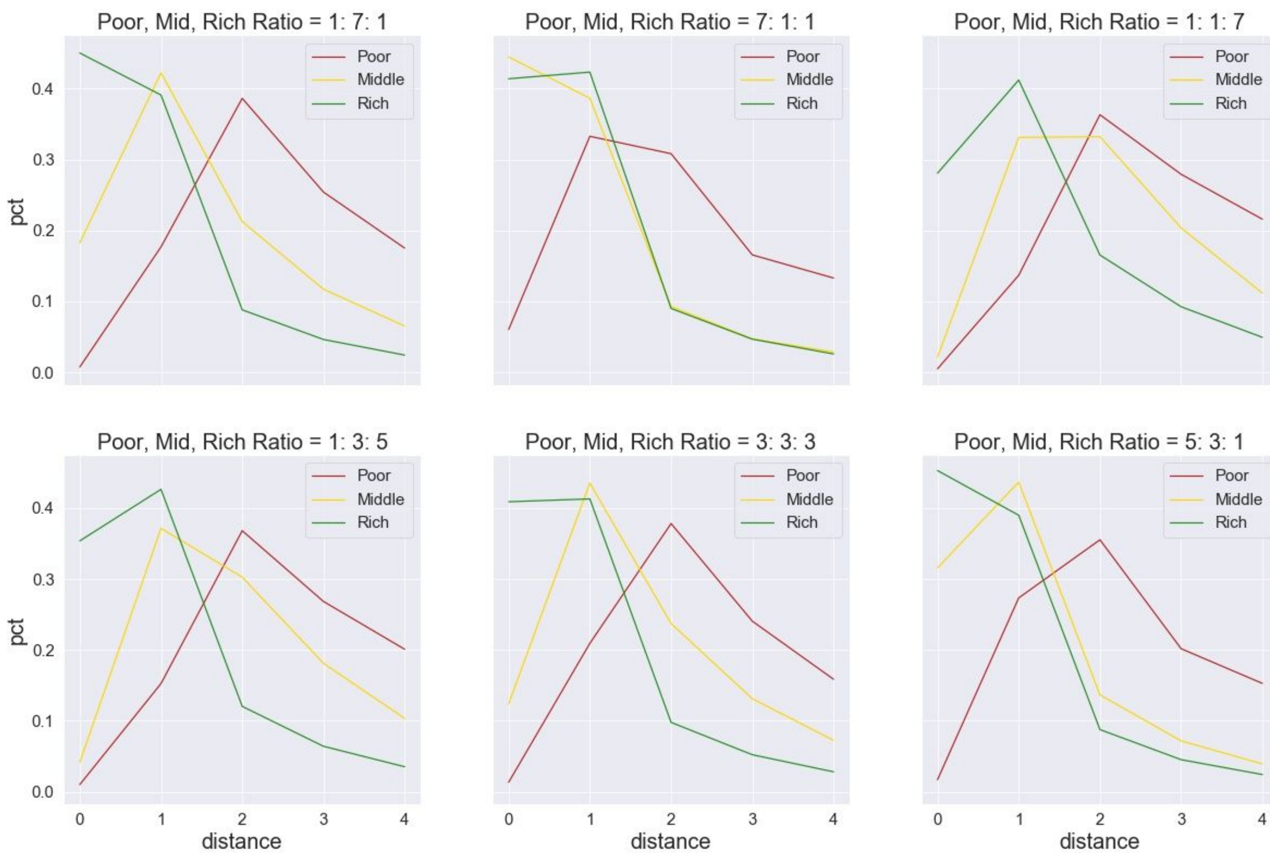

Figure 11: Normalized median percentage of population within each income group versus distance away from CBD, under different values of income group ratio, with N = 90 and K = 1/15.

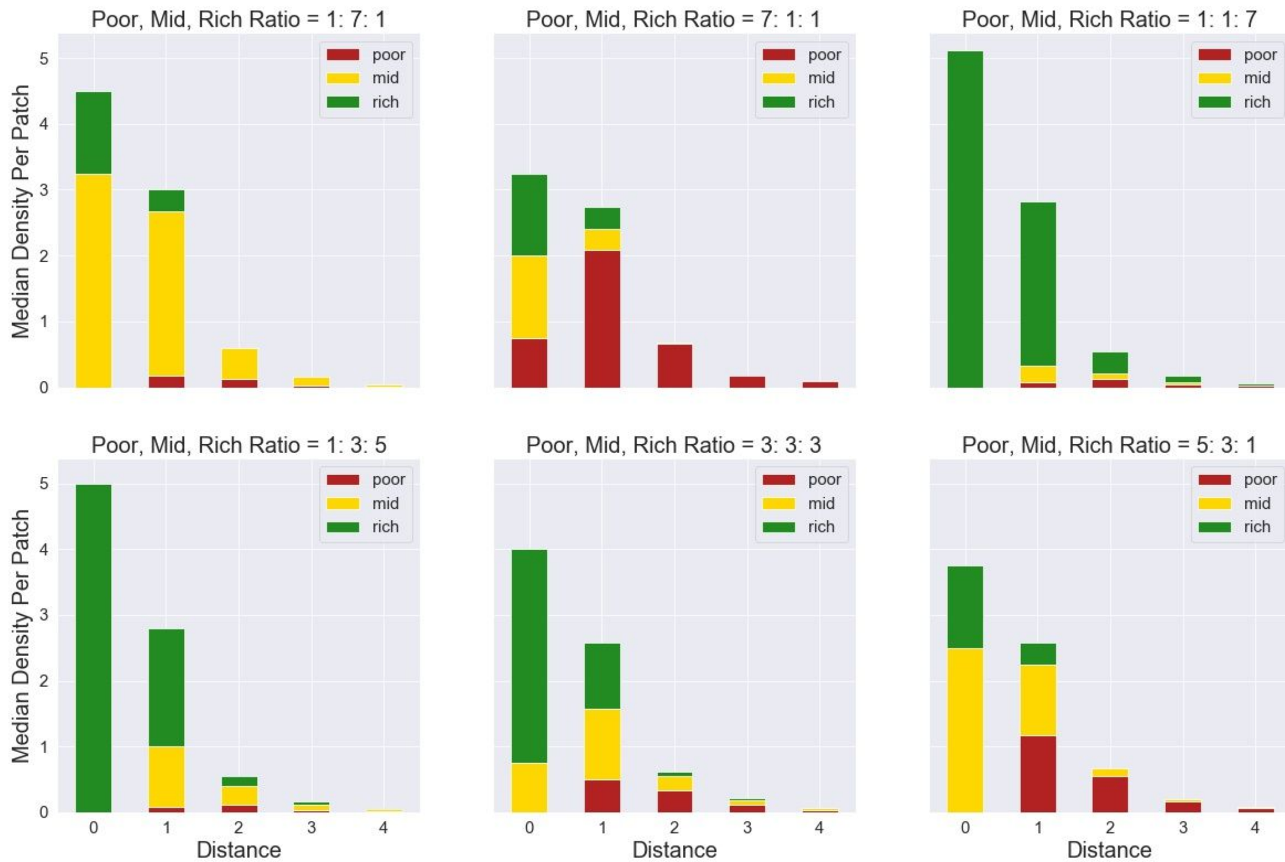

Figure 12: Median density per patch broken down in each income group versus distance from CBD, under different values of income group ratio, with N = 90 and K= $\frac{1}{15}$.

Discussion

Despite the simplistic nature of the model and its focus on the economic aspect of the residential dynamics, this model reproduced some of the outcomes that strongly resonate with the demographic changes in the real urban environments for the past century. For example, similar to how the rich and poor polarized in the city center as population increases in the model, cities worldwide that experienced industrialization in the early 20th century faced the same challenge. The agglomerated nature of the industrial work attracted a large population of rural and foreign immigrants to cluster at the city center. As a result, cities became more crowded and slowly gentrified the centers. On top of the economic narrative is the fact that new immigrants, ignorant of the new urban environment, often have strong motivations to locate wherever the jobs and their social ties are present in the city. The social capital constraint in the model captures this dynamics as the “poverty traps”, which refers to the spatial concentration of poverty at the city center. When the high demand for affordable housing clashes with the raising economic values of the land, the urban poor often resort to informal settlement as the solution. What this model does not capture is the story of transportation, which not only enables the separation of work and life, but also diverts the social tensions in the city center by dispersing the population into cheaper suburbs. In particular, middle income and high-income population take the lead for the spread as they can afford cars, larger floor spaces, and the cost to maintain their social ties while moving around in the cities. Even without transportation, our model shows how middle income group is forced to city periphery when population increases, a trend of which will be even more encouraged if transportation resolves the physical proximity constraint of work.

The model also shows that residential location distribution, even though based on economic factors, is still inherently social. The economic capital from each income group is not only translated into affordability for housing, but also the relative power for it to achieve optimal choices and impact the location status of other groups. The spatial distribution of income groups is essentially a product of power dynamics that pray on inequality. For example, no matter how the parameters change, the urban poor is always the population that has the highest representations in city periphery (the least favorite spots in the cities), which means they have to move frequently to find jobs. They are also particularly vulnerable, because their locations, especially those at the city center, are contingent on changes from various parameters, including total population, how fast the market responds to demand, and the ratio of high or middle income groups. Their limited social capital combined with the rapidly changing housing price in the city center can create a “poverty trap” where no affordable housing exists within their reach of information and network, forcing them to leave the cities after some time. As a contrast, rich people can create complete income segregated neighborhood in the city center when their population grows beyond a certain threshold. The distribution of their locations is also less affected by changing ratio in the other two groups (always stays close to the city center) and can reach the maximum density prescribed by the housing price equation. When the housing market is quickly saturated (with high market responsiveness), they are the last to move away from the city center, but when the market is full of mobility (with low market responsiveness), they are the first to move into the city center given their advantages at monitoring housing price at a larger scale. Wherever they go, they raise the housing price and thus squeeze the room for other income groups to reside, which can be interpreted as the process of gentrification. The middle-class people sit between the rich and the poor. They tend to locate between the city center and the periphery and follow the movement of the wealthy ones. On one hand, the majority of this population can move closer to the CBD when the market is not very responsive, the total population is small, or the ratio of the rich is low. They are also flexible enough to escape the traps for unaffordable housing when things turn south. On the other hand, they have little leverages to affect the high-income class and can be forced into suboptimal locations due to the tension between the rich and the poor at the city center when the population growth reaches a certain point. Therefore, this model can not only account for the analytical outcomes from the equation, but also consider how the interactions of social and economic capital mediate the operation of the market.

Though the model needs to be further developed to answer policy questions, it can help urban planners and policymakers to rethink residential distribution in cities in two ways. First of all, the same pattern of residential distribution can have multiple causes. Increasing the ratio of rich people, total population, or market responsiveness can all have the same effects of expanding the presences of the upper class in city center while shrinking the space for the lower-middle class. Residential shifts also do not follow parameter changes linearly. For example, the density of low-income group starts to go down when the total population goes up, but once the total population passes a certain point, social capital begins to mediate the mobility of agents significantly, which breaks the predictions from the housing equation. Therefore, when predicting the future trend of income distribution, urban planners and policymakers should use computer simulations to explore the whole landscape of interactions to complement their intuitions and analysis. Secondly, the difference between the absolute compositions of residents at each distance, as opposed to the in-group percentages, challenges the concept of "planning for the majority". Typically, city stakeholders will look at the local income demography and make decisions based on what to invest so that the majority in the region can be served, which refers to the absolute density. However, the population by income in cities is inherently disproportional, and thus certain income group may never be the majority even though most of the people within this income group have concentrated at a specific area. For example, the model shows that in an extremely dense city, most of the middle income people (in terms of in-group percentages) have moved to city peripheries, but their absolute density is still lower than the poor (see Figure 8). This happens because the low-income group has a much higher total density and can still outweigh the presences of others with a smaller percentage. Consequently, a more sophisticated framework is required to ensure equal prioritizations of public resources so that they can be tailored to different income groups.

For future extensions, including transportation can break the tie between people's residential locations and people's job locations, suggesting a new modeling relationship. For the model to be nontrivial, more deliberation is needed to address how people may choose jobs and whether the differentiation of jobs is appropriate. A decentralized labor market may be another interesting direction to experiment with. Thanks to the speeding communication technology, companies in the 21st century have begun to disintegrate into different functional parts that can locate further away from each other in order to maximize efficiency. Most of the metropolitan cities today are polycentric with multiple job nucleus. Therefore, instead of having one set of economic forces that "push" and "pull" the population around, having multiple attractors can be not only analytically interesting, but also empirically worthy to help predict the population trend in emerging megaregions.

In summary, the agent-based model in this paper expands the schematic economic model in classic economic theories with heterogeneous agent income, the endowment of social capital, and a more sophisticated housing market dynamic that accounts for various factors. The model outcomes may still be too coarse to be extrapolated for policy recommendations, but they are one step closer to understand and replicate the mechanisms that underlie the spatial changes of demographic groups. In particular, attention should be drawn to income distribution under the market failure and the "poverty trap" created by the interaction between the population expansion and housing market's immediate response to close the pricing gap. This instance exemplifies the role of power as a second hand to reallocate residential resources beyond affordability as well as the complexity of residential dynamics that may not be predicted by an analytical equation.

References

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

Benenson, I. (1998). Multi-agent simulations of residential dynamics in the city. *Computers,* *Environment and Urban Systems*, 22(1), 25-42.

Bertaud, A. (2018). *Order without Design: How Markets Shape Cities*. Cambridge: MIT Press.

Brueckner, J. K., Thisse, J.-F., & Zenou, Y. (1999). Why is central Paris rich and downtown Detroit poor? *European Economic Review*, 43(1), 91–107. doi:10.1016/s0014-2921(98)00019-1

Bruegmann, R. (2006). *Sprawl: A compact history*. University of Chicago press.

Burgess, E. (1925). The Growth of the City: An Introduction to a Research Project. *The city*. 38.

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

Jaffe, E. (2015, November 17). An Intriguing Explanation for the 'Roots of Gentrification': A Hatred of Long Commutes. *CityLab*. Retrieved from https://www.citylab.com/transportation/2015/11/why-the-wealthy-have-been-returning-to-the-city-center/416397/

Harris, C. D., & Ullman, E. L. (1945). The nature of cities. *The Annals of the American Academy of Political and Social Science*, 242(1), 7-17.

Hochstenbach, C., & Musterd, S. (2018). Gentrification and the suburbanization of poverty: Changing urban geographies through boom and bust periods. *Urban Geography*, *39*(1), 26-53.

Hoyt, H. (1939). *The structure and growth of residential neighborhoods in American cities.* Washington, DC: Federal Housing Administration.

Huang, Q., Parker, D. C., Filatova, T., & Sun, S. (2014). A review of urban residential choice models using agent-based modeling. *Environment and Planning B: Planning and Design*, *41*(4), 661-689.

Liu, X., Xia, L., Chen, Y., Liu, T., & Li, S. (2010). 基于多智能体的居住区位空间选择模型 (Residential Spatial Distribution Model Based on Multi-agent Systems). *地理学报 (Geography Report)*, 65(6). Retrieved from https://wenku.baidu.com/view/f0e9de2aed630b1c59eeb50f.html

Masad, D., & Kazil, J. (2015). MESA: an agent-based modeling framework. *In Proceedings of the 14th Python in Science Conference (SCIPY 2015)*, 53-60.

Mills, E. S. (1972). Welfare aspects of national policy toward city sizes. *Urban Studies*, 9(1), 117-124.

Muth, R. (1969). *Cities and housing: The spatial patterns of urban residential land use*. University of Chicago, Chicago, 4, 114-123.

O'Sullivan, F. (2015, December 15). London's Wealthy Are Returning to the City as Its Poor Are Heading to the Suburbs. *CityLab*. Retrieved from https://www.citylab.com/equity/2015/12/london-city-suburb-wealthy-poor-centre-report/420504/

Patel, A., Crooks, A., & Koizumi, N. (2012). Slumulation: an agent-based modeling approach to slum formations. *Journal of Artificial Societies and Social Simulation*, *15*(4), 2.

Lemoy, R., Raux, C., & Jensen, P. (2016). Where in cities do" rich" and" poor" people live? The urban economics model revisited. In *ERSA conference papers* (No. ersa16p524). European Regional Science Association. Retrieved from: http://www-sre.wu.ac.at/ersa/ersaconfs/ersa16/Paper524_RemiLemoy.pdf

Lemoy, R., Raux, C., & Jensen, P. (2017). Exploring the polycentric city with multi-worker households: an agent-based microeconomic model. *Computers, Environment and Urban*

Systems, *62*, 64-73.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of mathematical sociology*, 1(2), 143-186.

Smith, N. (1979). Toward a theory of gentrification: a back to the city movement by capital, not by people. *J. Am. Plan*. Assoc, 45, 538-548.

Straszhem, M. (1987). The theory of urban residential location. In *Handbook of regional and urban economics* (Vol. 2, pp. 717-757). Elsevier.

Vaughan, L. (2018). *Mapping Society: The Spatial Dimensions of Social Cartography*. UCL Press.

Wu, F. (2002). Complexity and Urban Simulation: Towards a Computational Laboratory. *Geography Research Forum*, 22. Retrieved from http://raphael.geography.ad.bgu.ac.il/ojs/index.php/GRF/article/view/242/234

Appendix A - LOs and HCs Applications

Index

Capstone LOs:

- #cp194.101-accountability
- #cp194.101-planningarchitecture
- #cp194.101-qualitydeliverables

CS Major LOs:

- #cp194.101-ABMimplementation
- #cp194.101-ABMmetric
- #cp194.101-ABMviz
- #cp194.101-cs146_rightdistribution
- #cp194.101-IL181_urbanmodeling
- #cp194.101-IL181_urbantheory

Foregrounded HCs (8):

- #variables
- #multipleagents
- #simulation
- #responsibility
- #professionalism
- #organization
- #evidencebased
- #sourcequality

Additional HCs (15):

- #selfawareness
- #algorithm
- #descriptivestats
- #dataviz
- #distribution
- #modeling
- #systemdynamics
- #emergentproperties
- #multiplecauses
- #hypothesisdevelopment
- #designthinking
- #critique
- #interventionstudy
- #decisionselection
- #powerdynamics

LOs

***#cp194.101-accountability* (LO1)**: *Demonstrate ownership and stewardship of the Capstone project (e.g., keep commitments, present self and work products in a professional manner, take responsibility for making forward progress and demonstrate resilience in the face of challenges, mitigate behaviors that impair effective performance).*
*+ **#selfawareness** (**HC1**): Identify and monitor your strengths and weaknesses; mitigate behaviors and habits that impair effective performance. (H) CS*
*+ **#responsibility** (**HC2**): Follow through on commitments, be proactive, and take responsibility. (H) CS*

Demonstrating ownership of a project to me means keeping myself motivated, taking control of the pace of work based on my strengths and weaknesses, being proactive about countering challenges. Over the past year, I have to overcome many challenges, one of which is to juggle time between career development (e.g., grad school applications, grad school interviews, and grad school visits), academics (e.g. schoolwork), life (e.g., relationship and friendship broke up), and this big capstone project. The other big challenge is dealing with an unsatisfactory proposal at the beginning and the push-back from my advisor to not change topics. At first, I was very stressed out and not very motivated to pursue the current topic as it deviated from my original interests. On top of that were my advisor's expectations of having a working model as soon as possible.

I was able to overcome these challenges because I adopted strategies based on my strength and weakness and took responsibility for my work. I understand that my weakness is caring too much about grades, which not only adds to my stress but also makes me follow blindly along what my advisor suggested. Ironically, this is precisely the opposite of what my advisor wanted me to do — taking ownership of the project. I also understand that once I pick up my motivations, which means I have a clear idea of why I am working on a topic and truly found interests and meanings in doing so, I will perform better and will be able to produce progress steadily. Therefore, I decided to spend lots of time reading relevant literature, and even books to understand the problem landscape, even though I was under tremendous pressures to produce a model as soon as possible. I also sorted out help from my Minerva peers and selected Mesa (based in Python) instead of NetLogo as the programming framework because I can discuss codes and model structures with my friends. As a result, I was able to re-establish my motivations and built up a complete model in time. I am also aware that it is important to maintain my mental health and make the research experience enjoyable, especially if I will enter grad school doing research for the next five years of life. Therefore, I took time and worked on the project on my own pace, even though that sometimes means I had to use extensions and did not meet my advisor's expectations squarely. It is incredibly challenging because my grades have not been well for the capstone classes in the senior year (the average scores translate into B). I suppressed my urges to argue about grades because I know arguing will harm my motivations and focused on producing work according to my own standard instead.

#cp194.101-planningarchitecture *(**LO2**): Create systems (e.g., rubrics, timelines, intermediate deadlines, work groups, accountability groups, feedback mechanisms) that facilitate achievement of project goals and deadlines.*

I adopted multiple mechanisms to keep my work on progress and receive feedback.

1. I have a capstone work accountability group with Jonathan and Parthiv since London semester, and we worked together on Friday every week. I will discuss my codes with Parthiv, who is a master at programming, and refactor them consistently.
2. I kept track of the points raised in the check-in meeting and ran a to-do list on top of every check-in submissions to make sure items with priorities were addressed first.
3. I actively seek feedback from urban practitioners and professors to ground my research in the real world context and reveal the meanings that can be fueled to support my motivations. I presented my half-done capstone at Foster + Partner, a famous architecture and urban planning firm at London, and joined a book club at San Francisco with like-minded people to read about the book *Order Without Design*, which supply terminologies for me to discuss about my work and an overview of the space my work should fit in. I also bounced ideas back and forth with my mentor at Santa Fe Institute and followed up with their suggestions.

4. I have strategically leveraged my assignments from multiple classes in London semester (e.g., Poverty Maps —— The History of Mapping Urban Poor in London) to explore the question of income distribution, which helps me situate my work in the field and justify its significance.

#cp194.101-qualitydeliverables *(**LO3**): Deliver high-quality, substantial work products*

For the coding part, I have a quality simulation with deliberated logic and generates non-trivial results. I have iterated through the behavioral rules multiple times over the past year, testing out how each variable changes the model dynamics and improving on including new intricacies that strike the right balance between complexity and tractability. Both the interface, with four different canvases, and the graphs on average statistics (e.g. Figure 3 and 4) confirms that my model runs according to the assumptions and rules and it updates the data correctly. The codes are refactored multiple times for better performances, reorganized for convenient replication, uploaded to Github for quick access, and commented for easy understanding. I also carefully examined the sensitivity of model outputs and how they may change with each parameter. You can see more justifications in #ABMimplementation, #ABMmetrics, #ABMviz, #modeling, #simulation, #designthinking, and #variables.

For the paper part, I have written 22 pages and roughly 8000 words (including LOs and HCs Appendix 12700 words) with a substantial amount of codes for both simulation and analysis. The paper follows a standard publication format, including an abstract, introduction, method, result, and discussion section. You can see more justifications in #dataviz and #professionalism.

#cp194.101-ABMimplementation *(**LO4**): Implement the ABM model with my choice of a programming language that runs at a reasonable speed, with proper comments, clean codes, salient graphic styles.*
+ ***#algorithm*** *(**HC3**): Apply algorithmic thinking strategies to solve problems and effectively implement working code. (C) FA*

I justified why I chose Mesa library in Python language to implement the ABM model for both technical (e.g., easier to debug, based in Python, etc.) and personal reasons (e.g., easier to discuss with peers). I refactored the codes multiple times to improve the performance of the model. It used to take five seconds to update one frame in the browser, and after refactoring, now it can update three frames within one second. I also organized the codes nicely in both Github and Jupyter Notebook format, commented each section of the codes, and set a fixed context for the seaborn library for visualizing the metrics.

To achieve the optimal implementation, organization, and visualization, I thoroughly explored the functionalities of Mesa library and its advanced features, such as inserting Javascript codes to render the browser interface so that four canvases, instead of only one in the default, can be displayed side by side. To fully utilize Mesa, I did not rely on the documentation provided but often dug into the source codes to find the corresponding component and figure out what are the constraints of the program. For example, when I wanted to visualize percentage at each cell, the module for visualization only takes in color as a variable or color as a list, which is not friendly for percentages. I creatively used a format like [red * 20 + yellow * 30 + green * 50] to represent percentages for color in a single cell. Moreover, when I realized when the CanvasGrid module is not compatible with my implementation of patches, I overwrote the original function so that my codes can run. You can find both of these examples in server.py.

Here I listed a few examples of how I use algorithmic thinking to improve the performance of the model. I used to implement patches in the same way as agents (human), which is similar to the classic WolfSheep example in NetLogo. However, I soon realized that in my model, patches do not move and do not involve any methods that restricted to agents. It is also costly to calculate the number of people from each income group on each patch, grouped by distance, because the computer has to go through all agents and filter which agent is a patch and which agent is a human. Therefore, I refactored patches as dictionary items, and the key is their coordinates. In this way, I can retrieve patch attributes through their locations in one step and reduce the calculation time from O(N) to O(1). Another significant improvement is how I implement the neighbor search. I used first to filter out all the neighbors that satisfy the requirements and then randomly choose one of them as the target. Soon I realized it would be much faster if I implement a while loop and then randomly search through the neighbors. As soon as the agent found a satisfactory match, then the while loop stop. In this way, agents on average will find a target within a few samplings instead of going through a full list of twenty-four neighbors. It is a tremendous improvement of model speed because every agent needs to do neighbor-search within the radius for every step
(see agents.py).

***#cp194.101-ABMmetric* (LO5):** *Evaluate and design metrics and improve agent-based model's performance in terms of reliability, validity, and space for insights.*
+ ***#descriptivestats* (HC4):** *Calculate and interpret descriptive statistics appropriately. (H) FA*

My model essentially has three dimensions: two of dimensions are about how attributes change in a 2D space, and the third one is how the model changes over time. To design appropriate ABM metrics, I thought deeply about what will best describe such 3D model, which is very different from dealing with 1D distributions. Since the research question concerns about spatial distribution, I simply used the median to approximate the average spatial dynamics (e.g., either percentages or the number of income groups on patches) for 100 steps. I believe that median may be better than mean in this case because 1) median is more robust with outliers, and 2) median is more robust when the distribution of data is not symmetrical. I plotted a few histograms and observed some non-symmetric data distribution, but when I empirically tested the difference, mean and median does not seem to produce very different outcomes (See analysis.py). Therefore, I selected median as the summary statistics for the dependent variables (e.g., median percentage or median density per patch) over 100 steps for 100 simulations.

To decide what metrics to use to display spatial distribution, I first decided on what are the outcomes that I care to discuss. You can find more justifications in the paper section Model Outcomes and Metrics. The two metrics are complementary at describing the model's spatial dynamics: the bar chart focuses on absolute scale per distance, and the line chart focuses on the relative scale regarding each income group. In particular, I carefully normalized density at each distance according to the number of patches at that distance for the bar chart so that the data is comparable. You can see more details at #ABMviz about the intentions behind the visualization.

#cp194.101-ABMviz *(**LO6**): Visualize the outcome of the agent-based model that is intuitive and user-friendly for explorations, salient for deriving insights, informative and professional in graphic style for the outcome.*
#dataviz *(**HC5**): Interpret, analyze, and create data visualizations. (C) EA*

There are two major data visualizations in this paper. One is the simulation interface by Mesa library, and the other is the visualization for metrics (line chart and bar chart). First of all, throughout the paper, I consistently use one set of color to represent one type of income group and thus reduce the cognitive loads for readers. I labeled every graph and carefully set the context of the seaborn library so that all the figures are exported with high resolutions, large axis labels and tick labels, and large legend size. This removes any difficulties that the readers may have at examining details of the graphs. There are also color gradients on the simulation interface so that people can capture the spatial changes of density and percentages intuitively. The simulation interface is also user-friendly to manipulate.

Secondly, since I will use the same metrics to test model sensitivity against each parameter, I design the visualization carefully so that it is scalable, simplistic, but at the same time, rich in information. I used the line graph for percentage metric because it is straightforward to capture how the percentages change over distance and within its own income group in contrast with other groups. I used the bar chart for density metric because it is easier to interpret how the total density and its subcomponents (density by income) along the distance. It allows the readers to interpret spatial distribution from various angles and generate narratives on both population and income group level.

***#cp194.101-cs146_rightdistribution** (**LO7**): Given a dataset or real-world situation, use standard properties of probability distributions to select appropriate distributions in one or more variables to represent the scenario.*
+ ***#distribution** (**HC6**): Identify different types of distributions and make inferences based on samples from distributions appropriately. (C) FA*

I originally introduced this LO expecting that I may have time to play with the parameters that determine job distribution but did not make it. The applications of distribution in this paper include 1) setting up job distribution and land price parameters so that the job probability and land price sampled from the distributions are more likely to be higher in the city center but lower in the periphery, and 2) selecting appropriate summary statistics (see #ABMmetric). For both job probabilities and land price, I set up a sequence of distributions at each distance scale. For job probability, because I want higher job probability in the center and lower at the periphery, I designated the mean parameters of the normal distributions to decrease linearly (except for city center). I chose the normal distribution because probability is continuous and the sampling values will converge to mean and thus follow the expectation of a linear descent. I set a small variance so that the sampling values will not go below zero. Land price is categorical and thus is sampled from a categorical distribution (multinoulli distribution) instead of a normal distribution. I manually set the probability parameters for each land price value at each distance, making sure the overall probability of having a higher land price decreases when distance increases (see model.py). I also plotted the simulated distribution of mean job probability and land price at each distance to ensure the gradient of values aligns with the expectations (see Figure 3 and 4).

The distributions I chose are not perfect for the scenario and the mechanisms of assigning jobs and land price can be improved. A further extension is to relax the bundling between jobs and patches. In this way, I can use one overarching distribution (or hierarchical distributions) to determine how the jobs will be distributed. For example, Beta distribution is a better option to sample job probabilities from as it is confined in the interval [0, 1] and by changing the beta distribution parameters, we can change the job distribution from concentric (e.g. $\alpha = \beta = 2$) to skewed (e.g. $\alpha = 2$, $\beta = 5$) or polycentric (e.g. $\alpha = \beta = 0.5$). More details regarding the concentric assumptions and the interpretations made based on the sampling values can be found at #variable, #modeling, and #simulation.

***#cp194.101-IL181_urbanmodeling** (**LO8**): Identity and/or create a model and/or explain how a given model is appropriate or inappropriate for an urban context. In my case, it is creating an agent-based model that can explain the spatial distribution of migrants in the city and justify the strength and weakness of the model.*

This entire capstone project is to model the residential distribution of income groups, and the whole paper is explaining how this model answers the research question. In particular, I interpreted the model results in a language that urban planners and policymakers are familiar with (see Discussion). The strengths of the model are discussed in both the Introduction and Method section. The weaknesses of the model are outlined in the Model Assumption section and a paragraph in the Discussion section regarding future extensions.

#cp194.101-IL181_urbantheory *(**LO9**): Apply the primary ideas, concepts, theories, and language involved in computation and urbanism.*

The Introduction and the Method sections of the paper have engaged widely with urban theories in Sociology, Economics, Planning field, as well as the ABM residential modeling literature. I also adopted terminologies from both urban and ABM literature to describe, discuss, and interpret my model results (e.g., gentrification, heterogeneous agents, segregation, etc.). You can see #professionalism, #evidencebased, and #sourcequality for more details.

HCs

#variables *(**HC7**): Identify and classify the relevant variables of a system, problem, or model.*

All the variables in my model were widely discussed and supported in all kinds of urban literature (e.g., density, land price, housing price), some of which I discussed in the Introduction section. Some of the variables can be included as the input to the model (e.g., density) or the estimation of housing price, but I believe that density should not be constrained or be the cause of model outcome because the constraints for people to move freely should be as minimal as possible. This idea is also advocated in the book Order Without Design. I also explained why model parameters are chosen in the section Model Outcomes and Metrics. By setting up the housing price equation, I not only identify the relevant variables to housing price, but also create a way that these variables can be tied together to capture residential dynamics of income groups. In #decisionselection, I further reveal the deliberation behind the variable selection as they constitute the basic logic for people to decide where to locate.

#simulation ***(HC8):*** *Apply and interpret simulation modeling. (C) FA*
+ ***#modeling*** ***(HC9):*** *Recognize how models can be used to explain a set of data and generate new predictions. (C) EA*

This entire paper is about designing a generative agent-based model that simulates the dynamics of residential distribution by income groups. The simulation is based on #multipleagents and captures various #systemdynamics, #emergentproperties, derived from #multiplecauses (see those HCs for details). I introduced the model in the context of literature, justified its significance, outlined the method and programming framework, describe the model results as they change under different parameter values, and discussed the interpretations of the model. I also discussed what predictions the model could make for urban planners and policy makers, what are the constraints, and the possibilities with extensions (see Discussion section)

***#multipleagents* (HC10):** *Analyze and apply decompositions of complex systems into constituent parts. (C) CS*

One of the advantages of my agent-based model over the existing ones is that it not only differentiates agents into different income group but also applies a whole new modeling framework based on agent's income (e.g., housing price equation). In the Introduction section, I justified the reason to decompose a residential system into multiple types of agents and why understanding the residential distribution of income groups is important. The results of my model generate novel perspectives on how agents from different income groups interact with each other and with the housing market, resulting in distinct residential patterns (see Results and Discussion sections).

***#professionalism* (HC11):** *Follow established guidelines to present yourself and your work products professionally. (H) MC*

Professionalism can be broken down in many ways, especially in the context of a research paper. My paper demonstrates that 1) the content is robust and well justified, 2) the writing is clear and the organization is easy to follow (see #organization), 3) the references, graphs, equations, and in-text citations are formatted correctly (APA style for this paper). Graphs are also exported with high-resolution quality. I reverted some of the suggestions my advisor made (e.g., bolded Figure number and deleted Running head) because I think they do not follow the APA style 6th edition.

***#organization* (HC12):** *Effectively organize communications. (H) MC*

This paper is organized after the structure of a publishable paper. I also referenced papers published in the agent-based modeling field to identify sections they use to discuss the models. I tried to give a topic sentence at each paragraph to summarize what the paragraph is about. Moreover, I summarize my findings in the abstract, at the top of the result section, and in the last paragraph with a signifying language, so that readers who want to skim through the paper can quickly locate the main points. I also leveraged the headings of the section (e.g., Why is it important to study where people live) to convey straightforward messages to readers.

***#sourcequality* (HC13):** *Distinguish between categories and types of information to determine source quality. (H) MC*
+ ***#evidencebased* (HC14):** *Identify and appropriately structure the information needed to support an argument effectively. (H) MC*

In the *Introduction* section, I cited literature or research work that ranges from 1889 to 2018 to provide justifications on the importance and the discourse of the research question. Most of the sources are well-cited papers and books that help me clarify model assumptions and model interpretations.

#systemdynamics* (HC15): Recognize the role of attractors and sensitivity to varying conditions in the behavior of complex systems. (C) CS*

In agent-based modeling, attractors can mean states that the model results converge to. In the Results section, I discussed various states that the model converges to after 100 steps and averages over 100 runs, and why these state changes happen after certain thresholds (e.g., poverty traps by social capital). I also show how the model performance is sensitive to the model parameters and the implications behind.

#emergentproperties* (HC16): Identify emergent properties of complex systems and discern their causes. (C) CS*

In the Introduction section, I illustrated why residential distribution itself is an emergent property as it underpins complex interactions between social and economic factors. This entire paper discusses what are the residential distributions given the interactions of these factors (details see #multiplecauses) and explain why phenomenon such as slum or poverty concentration in the inner city can emerge out of the prediction of the analytical equation.

#multiplecauses* (HC17): Identify ways that multiple causes interact to produce complex effects. (C) CS*

In the Introduction section, I outlined how a specific residential distribution can come from multiple causes, some of which are endogenous (e.g., cultural dissonance, affordability), while others are exogenous (population, transportation, urbanization). In the Discussion section, I also have an entire paragraph about how residential patterns have multiple causes and the implications for urban planners and policymakers. The process to construct the housing equation also relies on deep intuitions of how different factors interact to impact housing price.

***#decisionselection* (HC18)**: *Apply decision making frameworks and heuristics to solve problems effectively. (H) EA*

Designing the behavioral rule in an agent-based model is essentially setting up a decision making framework for the agents. In the Introduction section, I discussed various ways in the literature that people have used to set up such structure, including using analytical functions or vectors to include as many factors as possible. Recognized that complicated and rational decision framework may not always reflect the reality (e.g., macroeconomics), I adopted the "less is more" heuristic to design simple behavior rules by focusing on the fundamental forces that motivate people to move their residences in cities, which are housing price and job probability. My results show that a model with simple structure and assumptions can also produce intricate outcomes that resemble urban processes in real-life.

***#interventionstudy* (HC19)**: *Design and interpret experimental studies. (C) EA*

Experiment in agent-based modeling means playing around with model parameters. In this paper, I swipe through three model parameters with a different range of values while keeping others as constant. As a result, I was able to derive predictions about residential distribution based on how these parameters change. I carefully selected the model parameters (see #variables) and the values they should be experimented with to maximize the coverage of different model states. My Results and Discussion sections are dedicated to interpreting the results of these experiments. In the process of developing the model, I also treated the model as an experiment, starting from the design with only one variable to more complicated mechanisms. This helps me understand how each factor contribute to the model (see #designthinking).

***#designthinking* (HC20)**: *Apply iterative design thinking to conceive and refine products or solutions. (H) MC*

In the process of developing a good agent-based model for my research question, I adopted design thinking to iterate my model so that it functions as expected and is capable of replicating more nuanced urban dynamics. I developed the model interface in the browser and the visualization metrics, as the instruments to gain intuition and feedback for further improving the model. For example, if my data collection or calculation process is wrong, I can notice that from the browser interface and correct it quickly. I followed my advisor's suggestion to build up a model to play with along with the literature review, which helps me hone my thoughts. By the end of the London semester, I have iterated the model from having only a single variable to multiple ones and was able to articulate the impact of each on the model performance. Having all these tools in place also allowed me to rethink the entire model structure at week 5 of this semester and make editions to the model logic that produce the current paper.

#critique ***(HC21)***: *Actively and critically engage with texts and other forms of communication. (H) MC*

In my Introduction section, I actively engage with the literature in both social science theories and agent-based modeling fields, showing how my research question fit into the broader discourse. I critique the inadequacy in both the observations from sociology theories and the analytical equations from economic models. By critiquing the inadequacy of the existing literature, I figure out the gap where my work will fit in and provide values to the field.

#hypothesisdevelopment ***(HC22)***: *Evaluate the link between hypothesis-driven research and the theories or observations that motivate it. (C) EA*

Though hypotheses are not addressed explicitly in my paper, they are essentially expectations I derive from models and equations. In the Method section, I discussed in details what are the hypotheses of residential distribution (how the model will behave) according to analytical function and model assumptions (see Figure 3 and 4). Then I used these hypotheses to drive the narratives of the paper and justified why this research question may be more complicated than what people will expect from linear thinking. Furthermore, in the Results section, I reminded the readers what we could expect from the equations and assumptions before contrasting them with the actual model results. Through this contrast, I highlighted how differentiated residential distribution emerges and how social capitals can break the predictions.

#powerdynamics ***(HC23)***: *Recognize how to influence group interactions by exerting different types of power. (H) CS*

In the Results and Discussion sections, I recognized how power could be exerted through income inequality and social capital by restraining people's choices and mobility. When interpreting and discussing the model outcomes, I have a full paragraph explaining how rich people can have more power at choosing where to locate and as a result, influence where other people can locate or even force them into suboptimal options.

Appendix B - Model Implementation (Github Link)

https://github.com/xiaofanliang/ResidentialDistribution