Executive Summary

Our aim is to develop a realistic model simulating the evolution of house prices across a city and the resulting formation of neighbourhoods. We extended the principle that house prices are influenced by householders' economic status and surrounding house prices by incorporating the desire to live near amenities. We also considered the addition of new houses and adjustments to the prices of existing ones. This was achieved using an agent-based model, where householders move based on house prices and proximity to amenities, shaping house prices over time. We validated the algorithm with a base model on a grid, ensuring robustness and convergence, before applying it to model house price dynamics in New York City using real-world data. Our base model findings highlight the significant impact of proximity to amenities through the formation of affluent neighbourhoods around these features. In our NYC model, we observe a notable discrepancy between amenity influences reported in the literature and those aligning with actual price data. The discrepancy suggested that the theoretical importance of amenities may not fully capture their real-world impact on housing markets, emphasizing the need for data- or context-driven adjustments. Adopting a discrete optimization approach, we considered 60 permutations of amenity weightings and selected the one fitting actual price data best, noting that supermarkets were consistently the most valued amenity and parks the least before disruption in 2020. Computational constraints prevented testing many parameter combinations. Further work is needed to implement a continuous optimization method to determine the model's optimal parameters.

1 Introduction

Numerous studies on housing markets and urban migration highlight a strong relationship between residential location choices and the socio-economic characteristics of households [12]. Emerging patterns of neighbourhood segregation in cities often correlate with residents' income levels [7, p. 20]. In the United States, where residential location is tied to social status, moving towards neighbourhoods of similar class position is widely observed [14, pp. 163-164]. The primary aim of this report is to incorporate proximity to amenities to explore its influence on residential decisions and therefore house prices. We investigate the relationship between migration patterns and income by building on studies that focus on desirable locations proximal to educational, recreational and retail facilities [12, p.10-11]. Additionally, we examine the impact of introducing new houses and adjusting the values of existing ones.

2 A Theoretical Base Model

We used an agent-based model [8, p. 3] to investigate how householders' economic status and neighbourhood house values influence property values in a simplified city. Our approach builds on Aguilera and Ugalde's model [2, p. 4]. This base model is nondimensionalised, making units irrelevant. We began by initialising an $n \times m$ grid, with each cell x representing a house with a value $V^t(x)$ at time t, and placing one householder (henceforth referred to as an agent) per house. Each agent's economic status is denoted by an affluence level $A^t(x)$ taking one of three values $j, k, l \in [0, 1]$, where $j < k < l$, reflecting different wealth levels.

In each iteration, we updated the model from time t to $t+1$ and performed the following steps for all houses and agents. First, the house value at x evolves according to

$$V^{t+1}(x) = A^t(x) + \lambda \frac{\sum_{y \in \mathcal{N}(x)} V^t(y)}{\#\mathcal{N}(x)}, \tag{1}$$

where $\mathcal{N}(x)$ is the square neighbourhood of radius 2, $\#\mathcal{N}(x)$ corresponds to the number of houses within it and λ is the inflationary parameter weighting the influence of the neighbourhood's mean house value on updated values [2, p. 5]. After updating house values, each agent evaluates 10 randomly selected houses and computes the economic improvement $\delta(x, y)$ for relocating to each based on Aguilera and Ugalde's formula [2, p. 5]. This formula decreases dissatisfaction from mismatches between income and house value, with $\delta(x, y) > 0$ indicating an improvement in satisfaction. If such a house exists, the agent swaps with the option offering the highest $\delta(x, y)$. Limiting evaluations to 10 reflects realistic search behaviour.

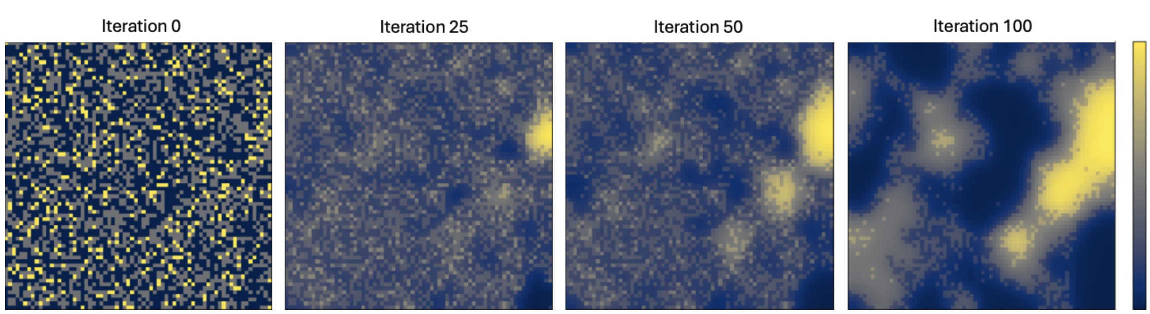

Figure 1: Progression of house values in a 70×70 grid representing a simplified city, with simulation outputs at iterations $t = 0, 25, 50, 100$ and house values indicated by a colour scale (blue = low, yellow = high).

We implemented the model with arbitrarily chosen values: all house values started at $V^0(x) = 5$ and agents were randomly assigned affluence levels of $j = 0.1$, $k = 0.5$, or $l = 1$ in proportions of $0.1, 0.4, 0.5$, respectively. The model is robust for a range of parameter choices; for example, $V^0(x)$ between 4 and 6 yields similar results. For $\lambda = 0.75$, Figure 1 shows the output of our model on a 70×70 grid after 100 iterations, clearly illustrating the segregation of high- and low-value regions. The random initial distribution quickly transitions into a smoother, clustered pattern and converges into a grid with well-defined neighbourhoods.

2.1 Incorporating the Influence of Amenities

We extended the model to account for agents valuing proximity to amenities, assuming that all agents value living near these amenities equally. Accordingly, we modified equation (1) to $V^{t+1}(x) + \alpha W(x)$, where $W(x) \geq 0$ represents the proximity of x to amenities Q. We define $W(x) = \sum_{a \in Q} \max\{(r + 1 - d_a), 0\}/r$, where d_a is the Manhattan distance between house x and amenity a. The multiplier α reflects the importance of nearby amenities in determining the house value, with the term $\alpha W(x)$ increasing the value when proximity to amenities is favourable. As α increases, the effect of amenities on nearby house values intensifies, while a larger radius r increases the number of houses influenced by the amenity. The economic improvement function $\delta(x, y)$ suggests that higher house values near amenities will likely attract wealthier agents, whose income aligns better with the higher updated house values. We excluded amenities from $\delta(x, y)$, as their positive effect on one householder would be offset by the negative effect on the other. By initially leaving some grid spaces empty, we also enabled the introduction of new houses over time and the modification of existing house values, either randomly or at selected locations, to simulate renovations.

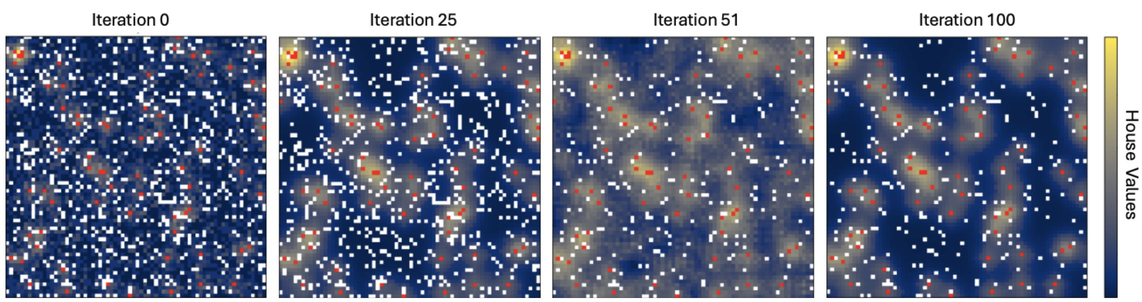

Figure 2: Effect of amenities, new houses and renovations on house values in the grid, with white cells representing unoccupied spaces and red representing amenities. Simulation outputs are shown at iterations $t = 0, 25, 51, 100$.

Using the same parameters as before, we set 15% of the grid to be empty, then randomly added 70 amenities, setting $\alpha = 0.75$ and $r = 4$. At iteration 25, new houses with an initial value of $V^{25} = 8$ were added randomly to 60% of unoccupied spaces. At iteration 50, the values of 1000 randomly selected houses were inflated to $V^{50} = 8$. These parameters are again arbitrary, but have been verified to be robust. Figure 2 illustrates how neighbourhoods of high-value houses form around amenities. Since $W(x)$ is higher for houses closer to amenities, this results in smooth, continuous patterns. The introduction of new houses has minimal effect on neighbourhood formation, as their values quickly align with surrounding houses. Renovations are visible through the increase in high-value (yellow) points across the entire grid between iterations 25 and 51, with similarly little effect on neighbourhood clustering.

We verified this algorithm using a segregation index S_{BO} [2, p. 7], which measures the deviation of a spatial distribution's entropy from that of a random distribution, capturing the degree of spatial clustering in the agent distribution. A lower segregation index indicates greater clustering into affluence-based neighbourhoods, as higher entropy corresponds to randomness. Aguilera and Ugalde [2, p. 13] found that the segregation index decreases with iterations, reflecting increasing segregation. Our model shows a similar trend, validating the algorithm numerically by capturing the emergence of clustering. Running the model with the parameters above with no amenities, new house additions or value increases (where empty spaces were set to the mean house value to allow for continuous patterns), the segregation index drops from approximately -1.85 at iteration 25 to -2.34 by iteration 100. With amenities, the segregation index at iteration 100 is slightly higher at -2.24, suggesting amenities reduce clustering, possibly due to more, smaller clusters forming.

3 Modelling New York City House Prices

Exploiting NYC's borough layout, we used the Python package OSMnx [3] to access real-world data on amenity and house locations, replacing the previous $n \times m$ grid approach with a map-based model. The commercial to residential building ratio is estimated at 1:10 [15]. We adjusted this to 1:5 to account for non-commercial amenities, resulting in 1000 houses and 200 amenities.

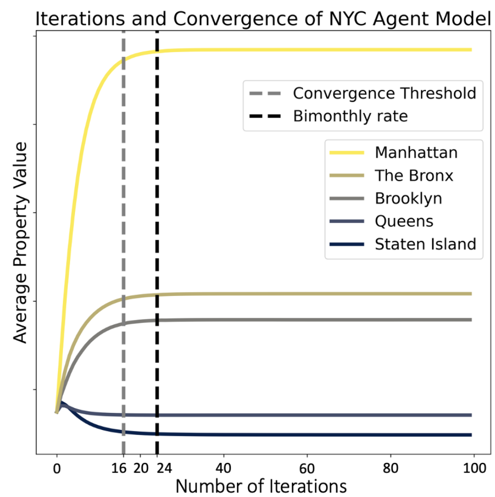

Figure 3: APB for iterations 0 to 100 in 2019 to test for speed of convergence.

The five boroughs of NYC—Staten Island, Queens, Manhattan, Brooklyn and the Bronx—contain 5%, 25%, 25%, 30% and 15% of the city's total houses respectively [4]. We note, however, that some residences in the Bronx are mislabelled as "buildings" instead of houses, leading to an under-representative sample that may distort these proportions and produce skewed patterns. Between 2017 and 2021, the boroughs grew 3%, 6%, 5%, 5% and 5% respectively [10] with assumed continuous growth. Despite this increase, the distribution of houses across boroughs remains relatively constant year to year [10]. To initialize our model, we used the median of yearly average house prices per borough (USD) from the NYC Department of Finance [9]. We used income distribution data [5] to assign 17 income levels (USD) to households in each borough proportionately, providing a more continuous representation of affluence compared to the base model. Exact values for house prices and incomes are provided in [1].

To ensure a consistent representation of average house prices in surrounding areas, we implemented a dynamic circular radius R_1 (km) based on the sparsity of the distribution in the map. If we run the model with fewer agents, R_1 will automatically increase, allowing us to include a sufficient number of houses in the average calculation despite the lower house density. We update R_1 by scaling it by the distance between the two agents farthest from each other divided by $\max\{\frac{\#\text{agents}+\#\text{amenities}}{100}, 5\}$. This ensures that even when running the model with fewer than 100 agents, the map will always be divided into at least five slices.

To represent the attractiveness of different amenity types, we first ordered the influence weights based on empirical findings. Distance to schools significantly influences residential location choice within a 5km radius, while grocery stores are highly attractive within 500m [13, p. 266]. Parks have a small positive impact, with a significant effect only when located within 1km [13, p. 268]. As NYC is a tourist hub, we included attractions as amenities akin to parks, enhancing aesthetics but not providing essential services. In our model, we approximated all amenity radii as $R_2 = R_1/5$ (km). This approach incorporates the dynamic radius approach while maintaining stability—division by 5 prevents over-inflation of house prices, which can occur if R_2 is too large, causing too many houses to share the same amenity sum and amplifying price increases via the neighbourhood averaging factor. This approximation reflects a computational limitation, preventing distinct empirical radii from being used.

Based on the above relative weighting, we initially set the weights to $[0.4, 0.2, 0.1, 0.1]$ for parks, schools, supermarkets and tourist attractions respectively. By fixing $\lambda = 0.8$, a reasonable value which allows smooth patterns to emerge, we focused on evaluating the performance of the amenity influence weights. House prices were then updated according to

$$V^{t+1}(x) = A^t(x) + \lambda \frac{\sum_{y \in \mathcal{N}_{R_1}(x)} V^t(y)}{\#\mathcal{N}_{R_1}(x)} + M \sum_{a \in \mathcal{N}_{R_2}(x)} w(a),$$

where $\mathcal{N}_{R_1}(x)$ and $\mathcal{N}_{R_2}(x)$ select neighbouring houses within R_1 and amenities within R_2 respectively. We sum over $w(a)$, the influence weights of each amenity a, scaled by M (USD), the median of average house prices per borough (which we henceforth refer to as APB). M approximates the general price level, ensuring that the valuation of amenities increases proportionally with general price inflation.

3.1 The Simulation and Results

We next analyse the results of our model, implemented using the real-world data above. Additionally, we investigate the significance and effect of adjusting the model's parameters to identify values that better align the model's outcomes with real-world house price data.

We first examined the model's sensitivity to initial conditions by running 100 simulations with different initial agent distributions. The results showed a maximum APB coefficient of variation of 1.8% in the final distribution, indicating that the model clusters quickly, with the initial distribution influencing only the first few iterations. This suggested that testing the model with different initial agent distributions for each simulation is unnecessary, simplifying computation without compromising prediction accuracy. To further reduce unnecessary computation, we tested when the APB converge in our simulation, using the maximum percentage difference of APB between iterations as a measure. Our convergence threshold is 1%, met after 16 iterations. We simulated bi-monthly interactions with 24 iterations per year. Figure 3 shows that all the APB converge by then.

Year	Permutation	Error (%)
2015	[0.1, 0.1, 0.4, 0.2]	28.7
2016	[0.1, 0.1, 0.4, 0.2]	38.8
2017	[0.1, 0.1, 0.4, 0.2]	38.5
2018	[0.1, 0.1, 0.4, 0.2]	27.3
2019	[0.1, 0.1, 0.4, 0.2]	27.1
2020	[0.1, 0.2, 0.4, 0.1]	13.1
2021	[0.1, 0.1, 0.4, 0.2]	36.6
2022	[0.1, 0.2, 0.4, 0.1]	32.8
2023	[0.1, 0.1, 0.4, 0.2]	75.4

Table 1: Best-performing weights (in the order parks, schools, supermarkets, tourist attractions) and corresponding average percentage error in APB for 2015–2023.

For our simulation, we used two distinct methods. In the first approach, we initialised all house prices to be the 2014 median of APB. The model ran for 24 iterations per year (for 9 years) and at the start of each new year, house prices were reset to the median of the final APB as predicted by the model. This approach led to large compounding errors over time, as it relied entirely on model predictions beyond the initial year. To address this, the second approach initialised house prices each year using the median of the APB from actual data rather than model predictions. By modelling each year independently with 24 iterations, we generated 9 separate models that predict the APB for 2015 to 2023. The 2020 model is shown in Figure 4. This method reduced errors by avoiding unrealistic assumptions of continuity over multiple years, especially given disruptions like the COVID-19 pandemic.

We fitted the model using the empirical weights with $\lambda = 0.8$, observing a significant 9-year average percentage error of 51.3%. This suggested that in reality, agents may not value amenities in the hypothesised manner. To better align the results with real-world price data, we initially attempted to refine the model using continuous optimization methods. Despite our attempts to implement the Genetic algorithm [11, p. 30] and the L-BFGS algorithm [6, pp. 1191-1192], computational limitations hindered their success. We thus opted for a discrete optimization approach, exploring 60 permutations of alternative relative amenity weightings by evaluating all possible permutations of the sets $[0.4, 0.3, 0.2, 0.1]$, $[0.4, 0.4, 0.2, 0.1]$, $[0.4, 0.2, 0.2, 0.1]$ and $[0.4, 0.2, 0.1, 0.1]$. This evaluation identified the best-fitting permutation for each year (as seen in Table 1) and validated the model's robustness, as it showed consistent behaviour across variations in these parameters. In particular, we observed that supermarkets are consistently the most valued amenity, contradicting literature that suggests schools are the most important. Parks are consistently valued the least, aligning with the literature. The order of importance remains stable until 2020, thereafter fluctuating yearly, possibly due to housing market instability caused by COVID-19.

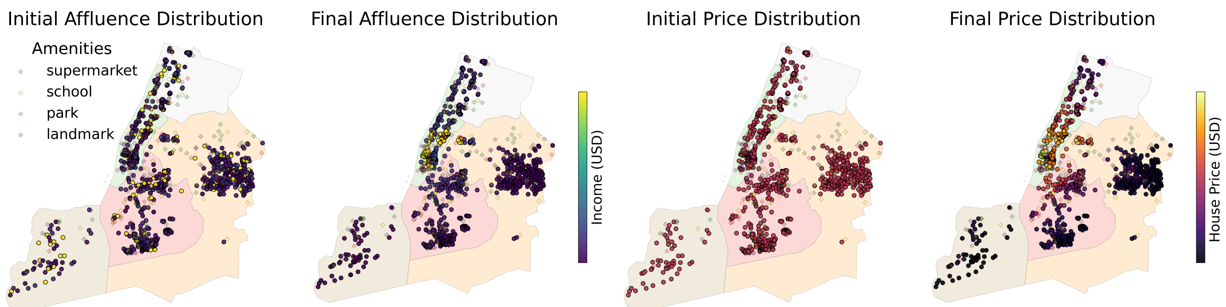

Figure 4: Initial 2019 agent and house price distribution versus the final 2020 predictions using weights from Table 1.

4 Conclusions and Further Work

Implementing our map-based model with real-world data resulted in large percentage errors, prompting tests of different amenity weightings. Supermarkets were consistently the most valued amenity and parks the least, with the ordering stable until the model was disrupted in 2020. We conclude that our model does not accurately predict real house prices when based on literature, but this can be improved by optimising the amenity weights. The map-based model can be adapted for studying house price dynamics in any city worldwide.

Due to time constraints, we have not simulated renovations in NYC, but this can be easily implemented by following the procedure outlined in the basic model. A limitation of this work is the data quality from OSMnx [3], where many residences were mislabelled as "buildings" instead of "houses", leading to an unrepresentative sample of the Bronx. A further limitation was that, due to computational constraints, we used a discrete optimization approach to adapt the model to real-world data. Future work should refine this approach using continuous optimization algorithms and/or greater computational capacity. Another improvement would involve allowing for different amenities to have varying spheres of influence.

References

[1] Industrial Mathematics Group 6. Housing external influences - nyc data, 2024. (Accessed: 1 December 2024).

[2] Antonio Aguilera and Edgardo Ugalde. A spatially extended model for residential segregation. *Discrete Dynamics in Nature and Society*, 2007(1):048589, 2007.

[3] Geoff Boeing. Modeling and analyzing urban networks and amenities with OSMnx. 2024. Working paper. (Accessed: 29 November 2024).

[4] United States Census Bureau. 2014 New York City housing and vacancy survey - series ib: All occupied housing units by tenure. `https://www.census.gov/data/tables/2014/demo/nychvs/s1b-all-other-occupied.html`. Dataset: Borough. (Accessed: 17 November 2024).

[5] United States Census Bureau. 2014 New York City housing and vacancy survey - series ib: All occupied housing units by tenure. `https://www.census.gov/data/tables/2014/demo/nychvs/s1b-all-other-occupied.html`. Dataset: Total Household income. (Accessed: 17 November 2024).

[6] Rh Byrd, Ph Lu, J Nocedal, and Cy Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208, 1995.

[7] William A. V. Clark. Changing residential preferences across income, education, and age: Findings from the multi-city study of urban inequality. *Urban Affairs Review*, 44(3):334–355, 2009.

[8] G. Nigel Gilbert. *Agent-based models*. Quantitative applications in the social sciences ; no. 07-153. SAGE, Los Angeles, [Calif.], 2008.

[9] NYC Department of Finance. Annualized sales update. `https://www.nyc.gov/site/finance/property/property-annualized-sales-update.page#`. (Accessed: 14 November 2024).

[10] NYC Department of Housing Preservation Development. 2021 New York City housing and vacancy survey selected initial findings. `https://www.nyc.gov/assets/hpd/downloads/pdfs/services/2021-nychvs-selected-initial-findings.pdf`. (Accessed: 14 November 2024).

[11] Alain Petrowski and Sana Ben-Hamida. *Evolutionary algorithms*. Metaheuristics set ; volume 9. ISTE, London, 2017.

[12] Patrick M. Schirmer, Michael A.B. van Eggermond, and Kay W. Axhausen. The role of location in residential location choice models: a review of literature. *Journal of Transport and Land Use*, 7(2):3–21, 2014.

[13] Yves van de Vyvere, Harmen Oppewal, and Harry Timmermans. The validity of hierarchical information integration choice experiments to model residential preference and choice. *Geographical Analysis*, 30(3):254–272, 1998.

[14] Myungje Woo and Hazel A. Morrow-Jones. Main factors associated with homeowners' intentions to move. *International Journal of Urban Sciences*, 15(3):161–186, 2011.

[15] Citizens Budget Commission New York. The Potential for Office-to-Residential Conversions. `https://cbcny.org/research/potential-office-residential-conversions`, 2022. (Accessed: 26 November 2024).