M3 2024 Solutions - A Tale of Two Crises: The Housing Shortage and Homelessness

Executive Summary

Dear Minister of State for Housing and Planning,

Homelessness has plagued the world for centuries, and with inflation only exacerbating the situation, the problem has remained unsolved until this very day. Today, we bring you the master key: our solution to open the door to ending the homelessness crisis.

We first calculated the growth of various types of houses, and summed up our numbers to get a total value, utilizing linear regression. By 2034, we will have 138,773 homes in Brighton & Hove and 257,875 houses in Manchester. By 2044, we will have 144,585 homes in Brighton & Hove and 275,511 houses in Manchester. By 2074, we will have 162,019 homes in Brighton & Hove and 328,419 houses in Manchester. We are moderately confident that our predictions are accurate, as our R^2 value is high.

We utilized normal distributions and linear regression in order to predict the change in homelessness over the course of 10, 20, and 50 years. We took in to account the changing housing prices, income, monetary value, and population in order to calculate our final results. We found that by 2034 we will have about 2,623 homeless people in Brighton and Hove and 7,905 in Manchester. We also found that there will be approximately 5582 homeless people in Brighton and Hove in 2,044 and 16,111 homeless people in Manchester in 2044. Finally, we found that there will be approximately 26,638 homeless people in Brighton and Hove in 2074 and 76,145 homeless people in Manchester in 2074, with population and relative population of homeless people increasing in both regions

We utilized a Monte Carlo simulation to calculate a solution to resolve the issue of homelessness within Manchester. We accounted for migration, floods, and economic recessions. We ultimately found that in 10 years, with no government intervention, the homelessness percentages fall on average to 1.403% and the number of houses will be 255,131. In 25 years, the homelessness percentages fall on average to 3.988% and the number of houses will be 279,386. Finally, in 50 years, the homelessness percentages fall on average to 11.779% and the number of houses will be 319,310.

Contents

0 Global Assumptions

1. *The world will not go through any apocalyptic events within the years that we are calculating.*
   For example, unrealistic and fantastical events like the Purge, alien invasions, and the Sun exploding will not occur. However, some realistic occurrences may still happen, like natural disasters and economic crises.

1 Part 1: It Was the Best of Times

1.1 Defining the Problem

We are asked to calculate the changes in housing supply in either two provided UK regions or two provided US regions in the next 10, 20, or 50 years.

1.2 Local Assumptions

1. *The amount of mobile houses, annexes, caravans, and houseboats are negligible when calculating the change in housing supply.*
   The number of these houses is minuscule compared to the total number of houses. For example, in 2021, there were 170 combined mobile houses, annexes, caravans, and houseboats within Brighton and Hove, compared to around 131,240 total houses. Using this data, we find that these types of houses only come out to be roughly 0.1% of total housing units [1]. This means that these types of houses can be treated as negligible, as they do not significantly impact the data used to solve this problem.

2. *Only data post-2009 is relevant.*
   In 2008 and 2009, the housing crisis significantly skewed housing data globally, leading to extreme changes in these values [4]. Thus, we must use data after the housing crisis to create an accurate model to predict the future housing supply.

1.3 Defining Variables

- Y : Year
- T : Total Houses
- B : Bungalow
- F : Flat/Maisonette
- R : Terraced House
- S : Semi-Detached House
- D : Detached House

- U : Unknown Type House
- N_h : Number of Houses of Type h

1.4 Procedure

To find the number of total houses at a certain time, we used the following model:

$$T = \sum_{h \in \{B,F,R,S,D,U\}} N_h \tag{1}$$

This equation sums the amount of different houses to find the total amount of houses in a certain city at a certain time. To find the number of houses for each type of house, we conduct linear extrapolations based on house quantities data post-2009 [1]. Then, we sum these extrapolations to find the total housing in the region at a certain year.

1.5 Applying the Model

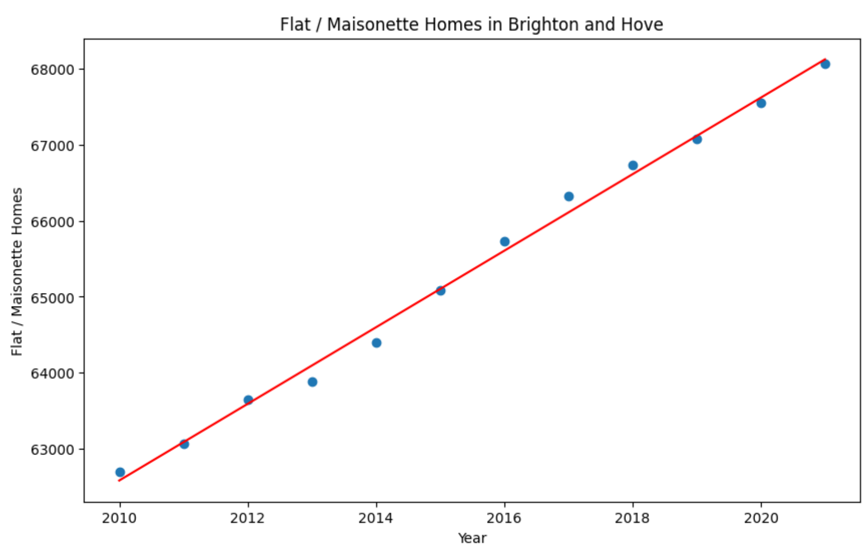

Figure 1: Flat/Maisonette Homes in Brighton and Hove over time

This graph represents the number of flat/maisonette homes in Brighton and Hove over time. We used the same linear regression model for all types of homes in Brighton and Hove and Manchester to extrapolate the numbers of total houses 10, 20, and 50 years later. The slope (m), y-intercept (b), and R^2 value for each linear regression are shown below in Table 1.

Linear Regression Results for Housing in the UK			
Type of Housing		Brighton and Hove	Manchester
Bungalow	m	-3.2867	14.371
	b	$15,451$	-26165
	R^2	0.9269	0.8513
Flat/Maisonette Homes	m	503.74	$1,104.1$
	b	$-949,936$	$-2.143 * 10^6$
	R^2	0.9269	0.8673
Terraced House	m	50.035	210.42
	b	-73328	$-344,849$
	r^2	0.9707	0.9476
Semi-Detached House	m	21.084	314.16
	b	-24905	-578742
	R^2	0.9918	0.9863
Detached House	m	20.874	83.252
	b	$-34,783$	$-161,933$
	R^2	0.9939	0.995
Unknown Type House	m	601.92	37.273
	b	$-1.086 * 10^6$	$-74,708$
	R^2	0.9946	0.7372

Table 1: The slope, m, the y-intercept b, and the R^2 value for each type of housing's linear regression by city

Linear Extrapolation Results of Total Number of Houses by City		
Year	Brighton and Hove	Manchester
2034	138773.846	257875.793
2044	144585.385	275511.807
2074	162020.000	328419.848

Table 2: The predicted number of houses in 2034, 2044, and 2074 in the cities of Brighton & Hove and Manchester

1.6 Strengths and Weaknesses

Strengths: The majority of our R^2 values are well over 0.9 which tells us that there is a strong linear relationship in our data. Moreover, this model is easy to implement while also being able to be scaled up to add more types of houses to be more precise.

Weaknesses: We failed to account for other extreme events such as another housing crisis, such as the Great Recession in 2008. However, it is close to impossible to account for these as they are extremely unpredictable and sporadic. Another weakness of our model is that it assumes that the city will remain in an inflationary gap and never enter a recessionary gap. This means our model is at its most accurate during times of prosperity, but is weaker during recessions.

1.7 Conclusion

After using our linear regression model, we found the future housing supply in our two UK regions. For the number of housing units in Brighton and Hove, we estimate a result of **138,774 for 2034, 144,585 for 2044, and 162,020 for 2074**. For the number of housing units in Manchester, we estimate a result of **257,876 for 2034, 275,512 for 2044, and 328,420 for 2074**.
From the strengths and weaknesses of our model, we conclude that we are **moderately confident** in our model.

2 Part 2: It Was The Worst of Times

2.1 Defining the Problem

We are asked to predict the changes within the homeless population in the next 10, 20, or 50 years, utilizing the same regions that we used in question one.

2.2 Local Assumptions

1. *Homeless people represent the demographic with the lowest income.*
   Many homeless people become homeless as a result of individual vulnerability and structural causes, including insufficient affordable housing, unemployment, and cuts in welfare expenditure [8]. Those most vulnerable to these structural changes are a part of the lowest income brackets, as they are those with the least financial security.

2. *Those who are outliers in the distribution of incomes will not become homeless.*
   We assume that extremely right-skewed outliers, which are those with the largest incomes, will not be homeless. Being an outlier indicates wealthiness, which also indicates a low chance of being homeless.

3. *People are going to try not to be homeless*
   We assume that everyone will try to maintain a house or place to stay, even accepting various inconveniences like moving around or reducing other expenses.

4. *Rent is proportional to the price of a house*
   We assume that the price of a house is proportional to rent because both the housing market and rental market are driven by similar factors, including location, demand, and economic conditions. [7]

5. *Income distribution will remain relatively constant*
   Factors relating to income distribution, such as mean household disposable income, have remained relatively constant post-2009, which implies that income distribution has also remained relatively constant [2].

2.3 Defining Variables

- x : The income for one randomly selected, non-outlier person in a region for a certain year.
- y : The cost of rent for a person for some year.
- x_{low} : The lowest income required to not be homeless.
- y_{low} : The lowest available cost of rent for some year.
- p : The median cost of a home for some year.
- $P(A)$: Probability of an event A, as standardized in statistics
- r : Proportion of people who are homeless.
- l : Value of a pound now compared to a future year to account for inflation.
- k : The maximum proportion $y : x$ to avoid homelessness. A constant.
- m : A constant to describe the ratio of rent to house cost.

2.4 Proof of Normality

Looking at the distribution of x with data from a random sample of people living in the UK [13], we can plot a histogram of the income of these people excluding outliers of the very rich to get the following graph:

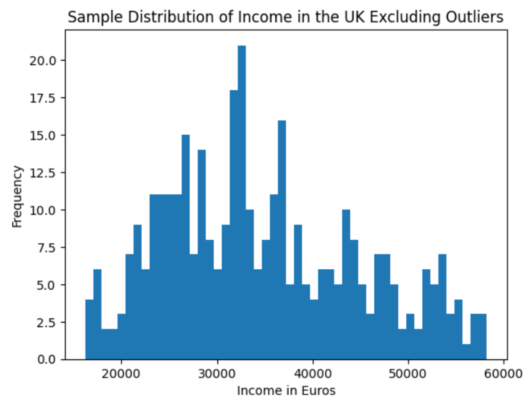

Figure 2: UK Income Distribution Excluding Outliers

To prove normality, we used a Quantile-Quantile (Q-Q) Plot, which compares two probability distributions by plotting their quantiles against each other. In this case, the distributions were income distribution and a normal distribution. In a Q-Q plot, forming an approximate line indicates the distributions are similar, and other patterns/no patterns indicate the distributions are not similar.

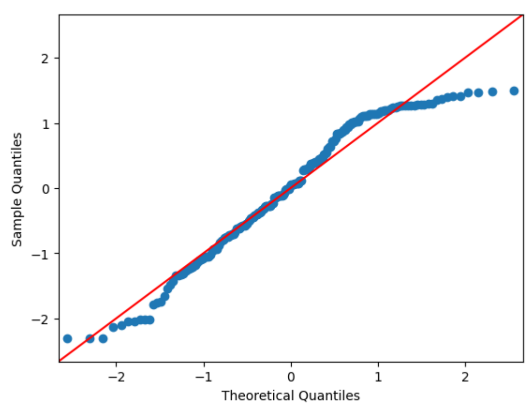

Figure 3: Income Quantiles vs. Normal Quantiles

This Q-Q plot is approximately linear, so $\mathbf{x} \sim \textbf{Normal}$.

2.5 "Homelessness Constant"

For a person with an income of x and who has to pay a rent of y, there will be a proportion of $\frac{y}{x}$ for which a person's income is no longer enough to pay their rent and they will become homeless. We can set this constant as k. If a person's rent is too high, we assume they will try to move to a place with lower rent. However, there is a point where they will no longer be able to find a place with lower rent since they are already staying at the place with the lowest possible rent, y_{low}. As we assumed the distribution of housing prices is constant, y_{low} is proportional to p and therefore is constant for a certain year. Then, $k < \frac{y_{low}}{x}$ for people who are not homeless. Therefore, whether a person is homeless is determined by their income, x. So, we can find k by finding the minimum value, x_{low} for which a person is borderline to being homeless, This means, we have

$$k = \frac{y_{low}}{x_{low}}$$

Note that k is constant even throughout different years.

Next, following our assumption that y is proportional to the median price for housing, we set $y = mp$, where m is a constant, even in different years . Therefore, we can plug this into our equation for $k = \frac{y_{low}}{x_{low}}$ to get $\frac{k}{m} = \frac{p}{x_{low}}$ or

$$\frac{x_{low}}{p} = \frac{m}{k}$$

We can call $\frac{m}{k}$ the "Homelessness Constant" (for lack of a better word) since we can use it to represent whether a randomly selected person in a certain year is homeless or not. If $\frac{x}{p} > \frac{m}{k}$, then a person's income is sufficient for the cost of their rent. Otherwise, it is not and they are homeless.

2.6 Dividing the Distribution

Based on the normality proof above, we know that x for some year is normally distributed. This also means that we can approximate μ_x with the median values of x since the mean of a normal distribution is equal to its median.

Based on normal properties, then we know that we can divide x by some constant, and the distribution of $\frac{x}{p}$ will remain normal. Therefore, $\frac{x}{p}$ for any year is normal as well. A new region is also drawn to represent the homeless population, bounded by $\frac{x_{low}}{p}$.

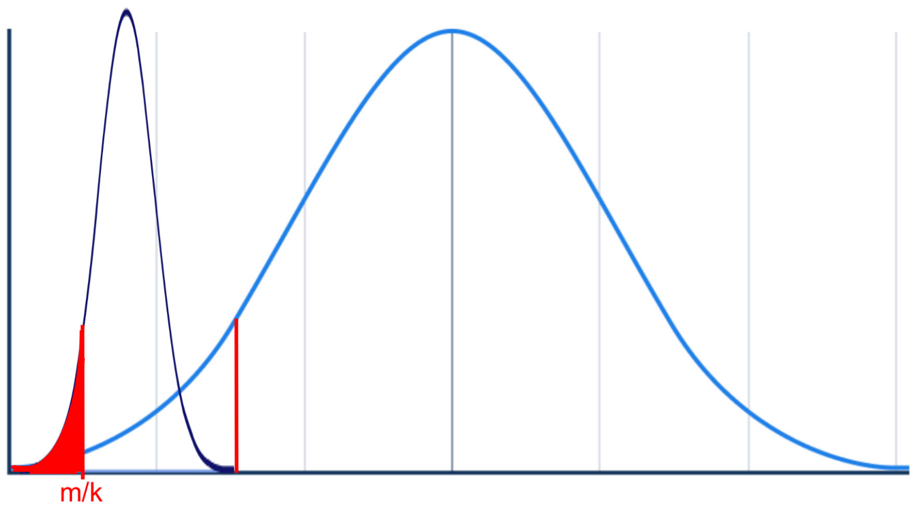

Figure 4: Comparing distribution of x (wide) with that of $\frac{x}{p}$ (thin)

However, since we know that $\frac{x_{low}}{p} = \frac{m}{k}$, the probability that a randomly selected person is homeless, or the proportion of people who are homeless, can be found with

$$P(\frac{x}{p} < \frac{m}{k})$$

So, since $\frac{m}{k}$ is constant over time, we calculate it with our values $\frac{m}{k} = \frac{x_{low}}{p}$

2.7 Calculating the "Homelessness Constant"

We find $\frac{x_{low}}{p}$ value for different years by taking the proportion of people who are homeless, r, which we can find by taking the total number of homeless people divided by the total population. Then, since we are assuming homeless people also have the lowest income, we want to find

$$P(\frac{x}{p} < \frac{x_{low}}{p}) = r$$

We can get $\frac{x_{low}}{p}$ by plugging this into an inverse normal function. We can get the standard deviation of x with the formula

$$\sigma_x = \frac{x_{40\%} - x_{20\%}}{z_{40\%} - z_{20\%}}$$

Where n is the percentile, x_n is the nth percentile for n, z_n is the z-score of the percentile. (We chose these values since we want more accurate data for those with low income.) This gets us average standard deviations of 10532.05366 pounds for Brighton and Hove and 8203.128984 pounds for Manchester.

To get an accurate $\frac{x_{low}}{p}$, we can solve $P(\frac{x}{p} < \frac{x_{low}}{p}) = r$ for $\frac{x_{low}}{p}$ (But this time using a standard deviation of $\frac{\sigma_x}{p}$) for every single year and solve for an average. This gets us

$$\frac{x_{low}}{p} = \frac{m}{k} = 0.01610675325$$

for Brighton and Hove, and

$$\frac{x_{low}}{p} = \frac{m}{k} = 0.03830304155$$

for Manchester.

2.8 Linearizing

In the below plots, we took the data given and performed a linear regression on them. From these plots, we got multiple values for the respective quantity for each year. For the median housing cost of Brighton and Hove, we got values of £580,946 for 2033, £740,636 for 2044, and £1,219,706 for 2074. For the median income of Brighton and Hove, we got values of £40,054 for 2034, £45,364 for 2044, and £61,294 for 2074. For the median housing cost of Manchester, we got values of £311,374 for 2034, £396,484 for 2044, and £651,814 for 2074. For the median income of Manchester, we got values of £34,462 for 2034, £39,892 for 2044, and £56,182 for 2074. For the population of Brighton and Hove, we got values of 295,676 for 2034, 303,816 for 2044, and 328,236 for 2074. For the population of Manchester, we got values of 619,574 for 2034, 677,684 for 2044, and 852,014 for 2074. These graphs had a strong correlation coefficient, which tells us that linear regression was an effective method for extrapolating the data.[13]

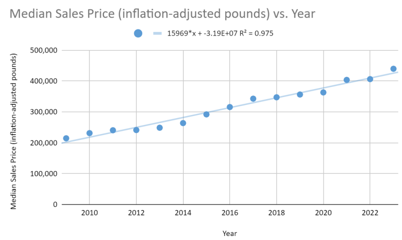

Figure 5: Median Housing Cost(Brighton and Hove) vs. Year

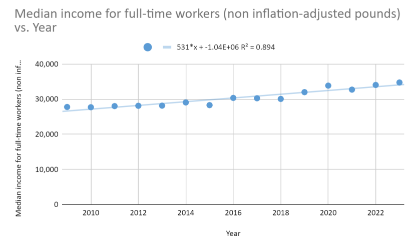

Figure 6: Median Income(Brighton and Hove) vs. Year

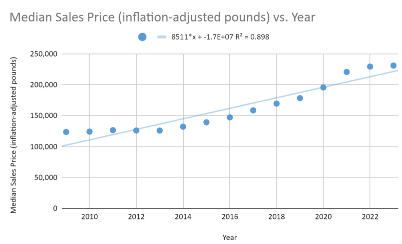

Figure 7: Median Housing Cost(Manchester) vs. Year

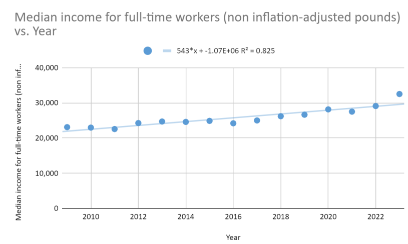

Figure 8: Median Income(Manchester) vs. Year

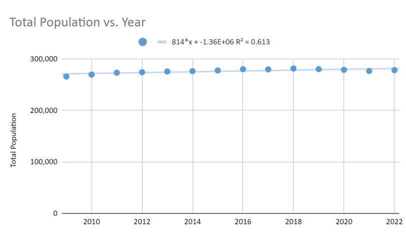

Figure 9: Population (Brighton and Hove) vs. Year

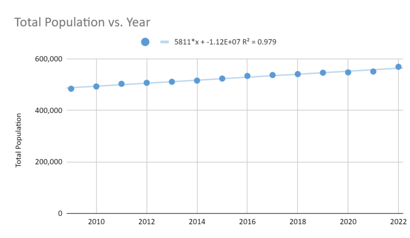

Figure 10: Population (Manchester) vs. Year

2.9 Putting It All Together

Now that we have all of the data for the calculations, we can perform them. Using Google Sheets to solve the proportion below $\frac{m}{k}$ a normal probability distribution. We calculate:

$$\text{NORMDIST}(\frac{m}{k}, \frac{\mu_x}{p}, \frac{\sigma_x * l}{p}, \text{true})$$

. Where $\frac{m}{k}$ is the maximum value of our distribution (and we measure area to the left), $\frac{\mu_x}{p}$ is the true mean of the distribution $\frac{x}{p}$, and $\frac{\sigma_x * l}{p}$ is the standard deviation of $\frac{x}{p}$. We include l since the standard deviation of x will increase as inflation does. As Pounds are worth less, then people will have more of them. As proportion will remain the same, ("true" is just part of Google Sheets). Doing these calculations to find the area for different years gets us,

Normal Distribution Results for Brighton and Hove			
Year	2034	2044	2074
Home Price (p)	£580946	£740636	£1219706
Mean Income (μ_x)	£40054	£45364	£61294
Inflated Value of Pound l [5]	1.23	1.52	2.83
Population	294676	303816	328236
Proportion of Homeless	0.008903598285	0.0183748475	0.08115677042
Homeless People	2623	5582	26638

Table 3: Inputted Values and Corresponding Results

We repeat these calculations for

2.10 Strengths and Weaknesses

Strengths:
One strength of our model is our use of linear regression. In our model there was a strong r^2 value which allowed for us to use linear regression. Linear regression is a strong yet simple tool which makes it a perfect tool to use when it is applicable. Another strength is our use of a normal distribution. The normal distribution is a simple, yet highly accurate tool and allows for us to make calculations which would be complex significantly simpler. This allows for our model to be more effective because of it's simplicity.

Weaknesses:
One weakness of our model is that our model is not completely a normal distribution, but rather is an estimated normal distribution. This will cause some small error within our calculations. Another weakness is that we are assuming that there is a constant y_{low}. The y_{low} value will generally change based on the economic state of the city at a a given time and will not usually remain constant for long periods of time. Another weakness of our model is that it is possible that $\frac{m}{k}$ will not be constant. In our model we held $\frac{m}{k}$ as a constant, however it is possible that it does not remain constant. Finally the inflation rate could suddenly spike or suddenly drop down which is another weakness of our model.

2.11 Conclusion

After using our Normal Distribution model, we found the number of homeless in our two UK regions. For Brighton and Hove, we estimate about **2623** homeless people for 2034 or about **0.89%** of the total population, **5582** homeless people for 2044 (**1.84%**), and **26638** homeless people for 2074 (**8.12%**). For Manchester, we estimate a result of **7905** homeless people for 2034 (**1.28%**), **16111** homeless people for 2044 (**2.38%**), and **76145** homeless people for 2074 (**8.94%**). Both cities will see an increase of the homeless population and relative proportion to the total population.

3 Part 3: Rising from This Abyss

3.1 Defining the Problem

We are asked to create a model that would help a city determine a long-term plan to combat homelessness. We are also asked to analyze how adaptive our model is to unforeseen circumstances, like natural disasters and economic crises.

3.2 Local Assumptions

1. *No legislation will dramatically alter the number of homeless people in the given regions.*
   We assume that there will be no panacea for homelessness and that the problem of solving homelessness will not be a quick and easy fix [10]. We assume that the issue of homelessness will steadily be tackled and dealt with over time, via our long-term plan.

2. *Homeless individuals will not stay homeless if they are provided the means to find a home.*
   We assume that all currently homeless individuals want to escape homelessness and if given the chance to, will [11].

3. *Certain natural disasters, like hurricanes, earthquakes, and tornadoes are negligible for our calculations.* Based on historical trends, natural disasters such as earthquakes are rare in the UK. For example, it is estimated that the largest strength earthquake that could occur in the UK is 6.5 on the Richter magnitude scale, which is only moderate [14]. This is similar to other natural disasters such as tornadoes and hurricanes which could possibly cause severe damage. Therefore, we omit these disasters from our simulation.

4. *We assume that all small economic recessions have a negligible effect on our calculations.* In general, a less severe recession, where a nation's GDP declines by 2% or less, last for roughly 11 months [3] [9]. These recessions have minimal effect on a nation's economy because they often reflect fluctuations in the global economy.

3.3 Defining Variables

- $thouses$: total number of houses
- $thomeless$: total number of homeless people
- $phomeless$: proportion of homeless people
- $tpop$: total number of population
- $mincome$: median income
- $mhouseprice$: median home price
- hc : rate of house growth per year
- hsc : rate of homeless people growth per year
- pc : rate of population growth per year
- ic : rate of income change
- hpc : rate of price of homes change
- fp : flood chance in Manchester
- fh : houses flooded beyond repair
- rp : likelihood of economic recession
- rm : percent income and house price decrease due to recession
- mp : probability of migration percent
- $mpop$: percentage of population effected by migration

3.4 Procedure

First, we used linear regression models using our data to calculate pc ($pc = 5652.1$), ic ($ic = 581.02$), and hpc ($hpc = 9216.1$)[1]. We analyzed these linear regression models and used their m (slope) value, setting these equal to our variables since these variables are the rate of increase, which is exactly what the slope measures. Then, we calculated fp ($fp = 0.0333$) and fh

($fh = 0.010477$), representing our extreme variable related to floods in our calculations, which we gathered from our data.[12] We also calculated rp ($rp = 0.07$) and rm ($rm = 0.048449$), representing our extreme variable related to recessions in our calculations, which we gathered from our data.[5] [9] Finally, we calculated mp ($mp = 0.01$) and $mpop$ ($mpop = 0.00313067749$, representing our extreme variable related to migrations in our calculations, which we gathered from our data. [6]

Next, we created a Monte-Carlo Simulation that ran 2000 times, each time iterating one-by-one over 50 years storing *thomeless*, *phomeless*, *tpop*, *mincome*, and *mhouseprice* and adjusting these values based on general trends and the extreme events of recessions, floods, and major migrations.

To begin, we store $thomeless = 2796, thouses = 238800, phomeless = \frac{2796}{238800}, tpop = 550630, mincome = 27500, mhouseprice = 220625$ based on the given initial data for 2021. One can change these values to the year of use.

Then, we iterate over 50 years, doing the following in each year. Initially, we do $thouses = thouses + hc$, $tpop = tpop + pc$, $mincome = mincome + ic$, and $mhouseprice = mhouseprice + hpc$, assuming that on a normal year with no major events, these values will increase according to its average rate. Next, we account for different extreme events. Since the chance of a flooding is 0.333, for each year, we take a random number from 0 to 1, and if the number is less than or equal to 0.333, this means there is a major flood. When there is a flood, we do $thouses = thouses - hc$ since the default increase of houses doesn't apply. Then, we say $thouses = thouses * (1 - fh)$. This sets the number of houses equal to the previous year's number of houses while removing the proportion of houses destroyed by the flood. Finally, we do $mhouseprice = pthouses * mhouseprice/thouses$. This takes the total amount of money spent on houses from previous years and divides this across the new amount of houses, increasing the average house price. To see whether there is a recession, we use the same independent random number method, meaning a recession and flood could happen in the same year. In the case of a recession, we do $mincome = mincome - ic$ and $mhouseprice = mhouseprice - hpc$ since the normal changes in income and house price are not accountable in this major event. Then, we do $mhouseprice = mhouseprice * (1 - rm)$ and $mincome = mincome * (1 - rm)$. This sets the house price and income price lower based on the average proportion that income and total money decrease during a recession period. Then, we do the same thing for migration percentages with random numbers to account when when there is a major immigration. During this major event, we do $tpop = tpop * (1 + mpop)$, which sets the new population equal to the previous population added with the proportion of new people who migrated to Manchester. Then, we do $mhouseprice = (tpop/prevpop) * mhouseprice$, which varies the median home price directly to the population. Finally, no matter the major events, we use the model in part two and input *mincome* and *mhouseprice* to find *phomeless* and finally do $thomeless = phomeless * tpop$.

To conclude our findings, we iterate this process over 50 years and run this 2000 times. Then, we take the median percent homeless and median income and average it over the 2000 trials to get our results.

This will drastically help policy-makers since they can change the model to edit the number of homeless people, income and other factors based on the suggested policy. Taking the results from the simulation, policymakers can make the best decision based on homeless percentage and number of houses in the future.

3.5 Strengths and Weaknesses

Strengths:
We used a Monte Carlo simulation, which is very good at accounting for confounding variables with respect to a event. We are reasonably confident that our model successfully accounts for the confounding variables related to homeless individuals.

Weaknesses:
Some weaknesses of a Monte Carlo simulation rely on excessive reliance of input data, and its sensitivity to our assumptions. As such, there may be some degree of variability to our actual solutions to remedy homelessness from the correct solution.

3.6 Conclusion

We find that in 10 years, with no government intervention, the homelessness percentages fall on average to 1.403% and the number of houses will be 255,131.476. We find that in 25 years, the homelessness percentages fall on average to 3.988% and the number of houses will be 279,386.075. We find that in 50 years, the homelessness percentages fall on average to 11.779% and the number of houses will be 319,310.360.

References

[1] *A Tale of Two Crises, MathWorks Math Modeling Challenge 2024, curated data*. URL: https://m3challenge.siam.org/kdfrldh/. Accessed March 3rd, 2024 (cit. on pp. 4, 5, 18).

[2] Paula Emily Andrews. *The effects of taxes and benefits on household income, disposable income estimate*. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/datasets/householddisposableincomeandinequality. Accessed March 3rd, 2024 (cit. on p. 8).

[3] *Business Cycle Dating*. URL: https://www.nber.org/research/business-cycle-dating. Accessed March 3rd, 2024 (cit. on p. 17).

[4] *Chart Book: Tracking the Post-Great Recession Economy*. URL: https://www.cbpp.org/research/economy/tracking-the-post-great-recession-economy. Accessed March 3rd, 2024 (cit. on p. 4).

[5] *CPI Inflation Calculator*. URL: https://www.officialdata.org/uk/inflation/2024?endYear=2074&amount=1&future_pct=0.021. Accessed March 3rd, 2024 (cit. on pp. 15, 19).

[6] Statista Research Department. *Estimated number of refugees from Ukraine recorded in Europe and Asia since February 2022 as of September 12, 2023, by selected countrye*. URL: https://www.statista.com/statistics/1312584/ukrainian-refugees-by-country/. Accessed March 3rd, 2024 (cit. on p. 19).

[7] Joshua Gallin. *The Long-Run Relationship Between House Prices and Rents*. URL: https://www.federalreserve.gov/pubs/feds/2004/200450/200450pap.pdf. Accessed March 3rd, 2024 (cit. on p. 8).

[8] Anwen Jones and Nicholas Pleace. *A Review of Single Homelessness in the UK 2000-2010*. 2010 (cit. on p. 8).

[9] Robert C. Kelly. *Recession: Definition, Causes, Examples and FAQs*. URL: https://www.investopedia.com/terms/r/recession.asp. Accessed March 3rd, 2024 (cit. on pp. 17, 19).

[10] Juhohn Lee. *With more than half a million Americans unhoused, the U.S. is still struggling to solve the homelessness crisis*. URL: https://www.cnbc.com/2022/01/07/why-the-us-cant-solve-the-homelessness-crisis.html. Accessed March 3rd, 2024 (cit. on p. 17).

[11] *Myths and Facts*. URL: https://www.coalitionforthehomeless.org/myths-and-facts/. Accessed March 3rd, 2024 (cit. on p. 17).

[12] Ralph Rayner. *Managing climate risks to well-being and the economy.* URL: https://www.theccc.org.uk/wp-content/uploads/2014/07/Final_ASC-2014_web-version-4.pdf. Accessed March 3rd, 2024 (cit. on p. 19).

[13] Daniel Smiech. *Average Salary in the UK (2023)*. URL: https://www.kaggle.com/datasets/danielsmiech/average-salary-in-the-uk-2023. Accessed March 3rd, 2024 (cit. on pp. 9, 13).

[14] *Where do earthquakes occur?* URL: https://www.bgs.ac.uk/discovering-geology/earth-hazards/earthquakes/where-do-earthquakes-occur/#:~:text=A%20magnitude%204%20earthquake%20happens,earthquake%20hazards%20for%20engineering%20projects.. Accessed March 3rd, 2024 (cit. on p. 17).

Appendix

```
import pandas as pd
import numpy as np

df_bh_housing = pd.read_csv("brightonhovehousing.csv")
df_bh_housing = df_bh_housing.iloc[17:].copy()
df_bh_housing['Flat / Maisonette'] = df_bh_housing['Flat / Maisonette'].str.replac

df_man_housing = pd.read_csv('manhousing.csv')
df_man_housing = df_man_housing.iloc[17:].copy()

df_man_housing['Flat / Maisonette'] = df_man_housing['Flat / Maisonette'].str.repl

df_man_housing

import matplotlib.pyplot as plt
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import classification_report, confusion_matrix

#x = df[['Year']].copy()##.to_numpy()
#y = df['Bungalow'].copy()##.to_numpy()


# Step 3: Create a model and train it
#model = LogisticRegression(solver='liblinear', C=10.0, random_state=0)
#model.fit(x, y)

# Step 4: Evaluate the model
#p_pred = model.predict_proba(x)
#y_pred = model.predict(x)
#score_ = model.score(x, y)
#conf_m = confusion_matrix(y, y_pred)
#report = classification_report(y, y_pred)

from sklearn.linear_model import LinearRegression

def houseNumReg(df, year):
```

```
  data = {'Year': [2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 202
  X_train = pd.DataFrame(data)

  Y_bung = df['Bungalow'].str.replace(',', '').astype(float)
  modelbung = LinearRegression(fit_intercept=True) ## Create a model
  modelbung.fit(X_train, Y_bung) ## Fit the model to the data
  bunga = modelbung.predict([[year]])[0]

  Y_flat = df['Flat / Maisonette'].str.replace(',', '').astype(float)
  modelflat = LinearRegression(fit_intercept=True) ## Create a model
  modelflat.fit(X_train, Y_flat) ## Fit the model to the data
  flata = modelflat.predict([[year]])[0]

  Y_terr = df['Terraced House'].str.replace(',', '').astype(float)
  modelterr = LinearRegression(fit_intercept=True) ## Create a model
  modelterr.fit(X_train, Y_terr) ## Fit the model to the data
  terra = modelterr.predict([[year]])[0]

  Y_semi = df['Semi-Detached House'].str.replace(',', '').astype(float)
  modelsemi = LinearRegression(fit_intercept=True) ## Create a model
  modelsemi.fit(X_train, Y_semi) ## Fit the model to the data
  semia = modelsemi.predict([[year]])[0]

  Y_detach = df['Detached House'].str.replace(',', '').astype(float)
  modeldetach = LinearRegression(fit_intercept=True) ## Create a model
  modeldetach.fit(X_train, Y_detach) ## Fit the model to the data
  detacha = modeldetach.predict([[year]])[0]

  Y_unkown = df['Unknown'].str.replace(',', '').astype(float)
  modelunkown = LinearRegression(fit_intercept=True) ## Create a model
  modelunkown.fit(X_train, Y_unkown) ## Fit the model to the data
  unkowna = modelunkown.predict([[year]])[0]

  return(bunga + flata + terra + semia + detacha + unkowna)

print(houseNumReg(df_bh_housing, 2034))
print(houseNumReg(df_bh_housing, 2044))
print(houseNumReg(df_bh_housing, 2074))
```

```
print(houseNumReg(df_man_housing, 2034))
print(houseNumReg(df_man_housing, 2044))
print(houseNumReg(df_man_housing, 2074))

import pandas as pd
import matplotlib.pyplot as plt

# Assuming df is your DataFrame
# Assuming m and b are the slope and intercept for the regression line

# Plot scatter plot
plt.figure(figsize=(10, 6))
plt.scatter(df_bh_housing['Year'], df_bh_housing['Flat / Maisonette'], label='Data

# Plot regression line
plt.plot(df_bh_housing['Year'], 503.74 * df_bh_housing['Year'] - 949936, color='re

# Add labels and title
plt.xlabel('Year')
plt.ylabel('Flat / Maisonette Homes')
plt.title('Flat / Maisonette Homes in Brighton and Hove')

# Show plot
plt.show()

import pandas as pd
import matplotlib.pyplot as plt

# Assuming df is your DataFrame
# Assuming m and b are the slope and intercept for the regression line

# Plot scatter plot
plt.figure(figsize=(10, 6))
plt.scatter(df_man_housing['Year'], df_man_housing['Flat / Maisonette'], label='Da

# Plot regression line
plt.plot(df_man_housing['Year'], 1104.1 * df_man_housing['Year'] - 2143000, color=
```

```
# Add labels and title
plt.xlabel('Year')
plt.ylabel('Flat / Maisonette Homes')
plt.title('Flat / Maisonette Homes in Manchester')

# Show plot
plt.show()

df_income = pd.read_csv('Average_Salary.csv')

df_income['Salary'] = df_income['Salary'].str.replace(',', '').str.replace('£', '
df_income.sort_values(by='Salary', inplace=True)

#df_income = df_income.iloc[:199]
df_income = df_income[ df_income['Salary'] <= 60000 ]

# Calculate the first quartile (Q1) and third quartile (Q3)
Q1 = df_income['Salary'].quantile(0.25)
Q3 = df_income['Salary'].quantile(0.75)

# Calculate the interquartile range (IQR)
IQR = Q3 - Q1

# Define the lower and upper bounds for outliers
lower_bound = Q1 - 1.5 * IQR
upper_bound = Q3 + 1.5 * IQR

# Remove outliers
#df_income = df_income[(df_income['Salary'] >= lower_bound) & (df_income['Salary']

df_income

import matplotlib.pyplot as plt

# Assuming df_income is your DataFrame
value_hist = df_income["Salary"].hist(bins=50)  # Adjust the number of bins as nee
plt.title("Sample Distribution of Income in the UK Excluding Outliers")
```

```
plt.xlabel("Income in Euros")
plt.ylabel("Frequency")
plt.grid(False)
plt.show()

import statsmodels.api as sm
import pylab as py
from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()
df_income['Normalized_Salary'] = scaler.fit_transform(df_income[['Salary']])

sm.qqplot(df_income['Normalized_Salary'].to_numpy(), line ='45')
py.show()

import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm

# Generate data for the original normal distribution
mu = 0  # Mean
sigma = 1  # Standard deviation
x = np.linspace(mu - 3*sigma, mu + 3*sigma, 1000)  # Range of x values
y = norm.pdf(x, mu, sigma)  # Probability density function of normal distribution

# Calculate x-value corresponding to 0.07 area from the left
x_value_at_007_area = norm.ppf(0.07, mu, sigma)

# Plot the original normal distribution
plt.plot(x, y, color='blue', label='Original Normal Distribution')

# Generate data for the second normal distribution
# Generate data for the original normal distribution
mu2 = -2.2  # Mean
sigma2 = 0.2  # Standard deviation
x2 = np.linspace(mu - 3*sigma, mu + 3*sigma, 1000)  # Range of x values
y2 = norm.pdf(x, mu, sigma)  # Probability density function of normal distributio
```

```
# Plot the second normal distribution
plt.plot(x2, y2, color='green', label='Second Normal Distribution')

# Add a vertical line at x_value_at_007_area
plt.axvline(x=x_value_at_007_area, color='red', linestyle='--', label='0.07 Area')
plt.legend()

plt.title('Comparison of Normal Distributions')
plt.xlabel('X')
plt.ylabel('Probability Density')
plt.grid(True)
plt.show()

data = {
    "population": [550630],
    "homeless": [2796],
    "percent homeless": [2796/550630],
    "houses": [238800],
    "house price": [220625],
    "income": [27500]
}


df_sim = pd.DataFrame(data)

df_sim

df_inflation = pd.read_csv('inflation_data.csv')

df_inflation.loc[0]['amount']

from scipy.stats import norm

def prop_homeless(x, p, year):
  if year < 2024:
    return(norm.cdf((0.03830304155 - x/p)/(( 8203.128984* df_inflation.loc[2024 - 
  return(norm.cdf((0.03830304155 - x/p)/(( 8203.128984* df_inflation.loc[year - 20
```

```
  import warnings
warnings.filterwarnings('ignore')

df_monte1 = pd.DataFrame({'percent homeless': [1], 'houses': [1]})
df_monte2 = pd.DataFrame({'percent homeless': [1], 'houses': [1]})
df_monte3 = pd.DataFrame({'percent homeless': [1], 'houses': [1]})

for j in range(2000):
  data = {
      "population": [550630],
      "homeless": [2796],
      "percent homeless": [2796/550630],
      "houses": [238800],
      "house price": [220625],
      "income": [27500]
  }


  df_sim1 = pd.DataFrame(data)

  thouses = 238800
  thomeless = 2796
  phomeless = 2796/238800
  tpop = 550630
  mincome = 27500
  mhomeprice = 220625

  houses_change = houseNumReg(df_man_housing, 2025) - houseNumReg(df_man_housing,
  homeless_change = -0.01
  pop_change = 5652.1
  income_change = 581.02
  homeprice_change = 9216.1

  flooding_percent = 0.0333
  flooding_houses = 0.010477

  recession_percent = 0.07
  recession_money = 0.048449
```

```
migration_percent = 0.01 #random
migration_pop = 0.00313067749

for i in range(50):
  thouses += houses_change
  mhomeprice += homeprice_change
  mincome += income_change
  prevpop = tpop #for migration
  tpop += pop_change

  if (np.random.random() <= flooding_percent):
    #print('flooded')
    thouses -= houses_change

    prev_houses = thouses
    thouses = thouses * (1 - flooding_houses)

    house_total_money = prev_houses * mhomeprice
    mhomeprice = house_total_money / thouses

  if (np.random.random() <= recession_percent):
    #print('recession')
    mincome -= income_change
    mhomeprice -= homeprice_change
    mhomeprice = mhomeprice*(1 - recession_money)
    mincome = mincome*(1 - recession_money)

  if (np.random.random() <= migration_percent):
    #print('migration')
    tpop = tpop * (1 + migration_pop)
    mhomeprice = (tpop/prevpop) * mhomeprice

  phomeless = prop_homeless(mincome, mhomeprice, i + 2021)
  thomeless = tpop*phomeless
  df_sim1 = df_sim1.append({'population': tpop, 'homeless': thomeless, 'percent

df_monte1 = df_monte1.append({'percent homeless': df_sim1.loc[10]['percent homel
```

```
  df_monte2 = df_monte2.append({'percent homeless': df_sim1.loc[25]['percent homel
  df_monte3 = df_monte3.append({'percent homeless': df_sim1.loc[50]['percent homel
  if (j%100 == 0):
    print(j)

print(df_monte1)
print(df_monte2)
print(df_monte3)
```