MathWorks Math Modeling Challenge 2024

Canyon Crest Academy

Team #17895, San Diego, California

Coach: Brian Shay

Students: Michelle Liang, Evan Luo, Larry Wu, Jonathan Xue, Stephen Zh

M3 Challenge FINALIST—$5,000 Team Prize

JUDGE COMMENTS

Specifically for Team #17895—Submitted at the close of triage judging

COMMENT 1: The executive summary was very nice, and overall, the paper was well organized. Including the assumptions and sensitivity analysis strengthened the paper and demonstrated the thought that went into each model. Include some findings from the rejected models to provide more evidence to support the chosen model. The inclusion of units within the tables was wonderful. Very nice job on Question 3.

COMMENT 2: Good discussion and testing of model in Q1. Nice research for Q3 and inclusion of different classifications of homelessness, including PSH.

COMMENT 3: The team gave clear reasons for picking the models. The Data wrangling employed by the team helps to create a better model. The model used in addressing the last question lacks an equation, which would have provided a better understanding of the solution.

COMMENT 4: Excellent job with the executive summary. The long-term policy recommendation could be described more clearly, without abbreviations. Good work simulating a model that builds on a program Seattle implements, the homelessness-alleviation program PSH. You found that PSH offers long-term access to resources for homeless individuals and can be very impactful as well as cost-effective. In setting up the models, values of constants and parameters were clearly stated and justified.

Good job listing all assumptions with their justifications and providing definitions for each variable used in your model. Sensitivity analysis for the models is given for each part of the challenge. However, Sensitivity analysis was based on the percent error.

The long-term plan did not account well for unforeseen circumstances. Suggested improvements of their models for the cities were brief, and the generalization of their model is not discussed in parts of the challenge. The team did not integrate other sources of data, which can be an avenue to further validate their results.

It should be remarked that the team successfully integrated results from previous affordable housing projects around the States, informing their long-term planning for addressing homelessness in Seattle and potentially different cities. They had fresh ideas, demonstrating creative problem-solving skills.

COMMENT 5: Good ideas and well explained, very creative overall. In Q2, showing the data, parameters, and models contributing to the multivariate regression would greatly enhance the discussion.

A Tale of Two Crises: The Housing Shortage and Homelessness

1 Executive Summary

To the Secretary of the U.S. Department of Housing and Urban Development and the Minister of State for Housing and Planning,

Homelessness is a prevalent and unsolved issue in the USA. In the United States alone, we have over half a million[1] American adults experiencing homelessness today. To help remedy this, one of the best ways to solve homelessness is to give homeless people external aid to help them get back on their feet. To better understand how to aid homeless people the best, we must understand and be able to predict the most important underlying causes of homelessness. To investigate further, we focused on two regions: Seattle, Washington, and Albuquerque, New Mexico.

For the first problem, we correlated our housing supply growth to the growth of the US Population. We then created a linear regression model on housing supply since US population growth has been roughly linear for the past 50 years. We then extrapolated the data for 10, 20, and 50 years into the future. We predict that Seattle will have 447,823, 512,475, and 706,429 housing units and Albuquerque will have 268,780, 282,792, and 324,827 housing units for 10, 20, and 50 years from now respectively.

We then created a combination of multiple linear regression models as well as a singular multivariate regression model to predict the homeless population for the two regions. By looking at the correlation between the homeless population and various other related factors of homelessness, we narrowed our focus to four factors: Real GDP, housing prices, total population of the region, and poverty rates. We used linear regression to predict the values of each factor for any future year. From there, our multivariate regression model considers each of the predicted factors to predict the homeless population. We predicted that the predicted homeless population of Albuquerque in 10, 20, and 50 years will be 1173, 1019, and 531 people, respectively, and we predicted that Seattle's homeless population in 10, 20, and 50 years will be 18169, 22837, 37916 people, respectively.

Finally, we went to the heart of the problem: Finding a way to use models to help identify the best course of action in tackling homelessness. We simulate a model in which the city of Seattle implements the homelessness-alleviation program PSH on a wider scale, and find that PSH can be very impactful as well as cost-effective.

We hope our findings will assist the country in helping policymakers make the best decisions in helping the homelessness crisis.

Contents

2 Global Assumptions

1. **The US economy will follow a business cycle of around 6 years**
Justification: The average business cycle since 1945 for the US is around 6.25 years but has been increasing in recent years[3]. This means that the US economy will predictably go into recession and also have economic growth. This also assumes that there are no major economic recessions such as the 2008 financial crisis or the Great Depression.

3 Part I: It Was the Best of Times

3.1 Restating the Problem

With housing prices increasing at a faster rate than the average income, housing is becoming a larger problem in the United States. In this section, we aim to create a model that will predict the housing supply in two United States cities—Seattle, Washington, and Albuquerque, New Mexico—in 10, 20, and 50 years.

3.2 Assumptions

1. **There are no significant policy changes, natural disasters, or unforeseen influences related to housing growth in either Seattle or Albuquerque.**
Justification: It isn't feasible to numerically quantify the impact of these influences on housing units, nor is it feasible to find a predictable trend.

2. **Housing supply increases proportionally to population growth.**
Justification: As population growth increases, there will be more demand for houses, causing the housing supply growth rate to increase. As population growth decreases, there will be less demand, thus lowering the housing supply growth rate. Thus, it is reasonable to assume that housing and population exhibit a correlative relationship.

3. **Population growth in all US cities are roughly linear.**
Justification: Judging from the past 70+ years of U.S. population data[4], population growth is roughly linear. The population trend line and residuals are shown in FIG. 1. and FIG. 2. Because it is difficult to extrapolate the growth rate of specific cities 50 years from now, we will assume them to grow according to the US growth rate as a whole, which is linear.

4. **The housing supply data is homoskedastic**
Justification: Assuming the housing supply data is heteroskedastic, that would mean that some of the years exhibit little to negative growth. This is an unreasonable assumption because it is costly to destroy that amount of houses and for that amount of people to migrate from city to city; therefore, we can safely assume homoscedasticity.

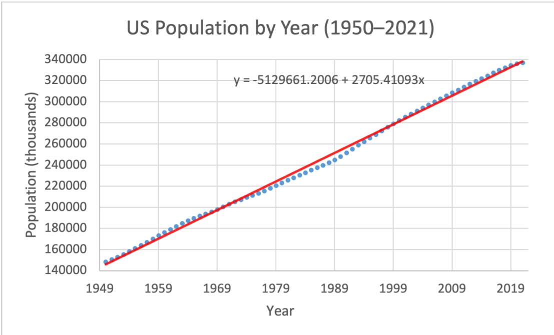

FIG. 1. Plot of US population in thousands by year. The red line is the trend line, and the scatterplot represents US population in thousands each year from 1950–2021.

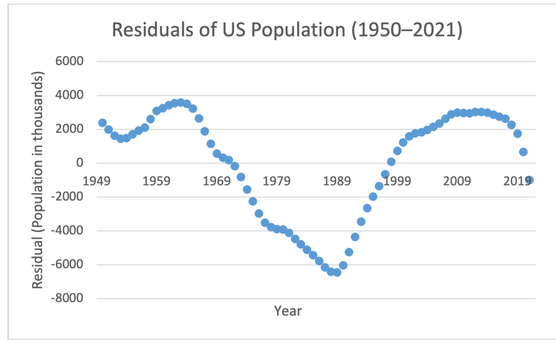

FIG. 2. Plot of US population residual in thousands by year. The residuals are extremely small; therefore, we can assume that the US population grows linearly.

5. **Housing supply growth will be roughly linear.**
   Justification: Given assumptions 2 and 3, we can then assume housing supply to grow linearly.

3.3 Variables

Symbol	Definition
x_s	Seattle Housing Supply
x_a	Albuquerque Housing Supply
y	Year

3.4 Model

3.4.1 Model Development

We chose a univariate linear regression model of y to x to predict housing supply in future years because we saw from FIG. 1. and FIG. 2. that population growth is roughly linear. Consequentially, housing supply growth is also roughly linear. The formula for a univariate linear regression is below: independent variable denoted by x, coefficients denoted by A and B, and dependent variable denoted by y.

$$y = Ax + B$$

The formula for univariate linear regression

Because we are extrapolating so far into the future, it is difficult to predict major events that cause huge temporary increases/decreases in housing supply, so it is best to predict the average housing supply. Furthermore, through Assumption 4, we now know that the data is homoskedastic and thus increases the validity of using a linear regression model. Therefore, a linear regression model, which aims to minimize the variance in our data by fixing a best-fit line, serves us best.

We thought about using other regression models, such as polynomial or exponential models, but we don't see any exponential nature exhibited in our data, whether it's population growth or housing unit growth. Using any other more complex models would overcomplicate our assumed linear relationship between year and housing units.

3.4.2 Executing the Model

We took the housing supply data of Seattle and Albuquerque in the years 2010–2022 given by the Mathworks Math Modeling Challenge[2] as shown below. We put this data into an excel sheet and found the equation of the line of best fit through linear regression and extrapolated it for 10, 20, and 50 years from now.

Year	**Housing Units**
2010	302465
2011	304164
2012	306694
2013	309205
2014	311286
2015	315950
2016	322795
2017	334739
2018	344503
2019	354475
2020	367337
2021	362809
2022	372436

3.5 Results

FIG. 3. and FIG. 4. are the graphs of our proposed linear regression for the two US cities, Seattle and Albuquerque.

3.6 Discussion

From our graphs, we find that from 10, 20, and 50 years from now, Seattle will have 447 823, 512 475, and 706 429 housing units, respectively. We also find that 10, 20, and 50 years from now, Albuquerque will have 268 780, 282 792, and 324 827 housing units, respectively. Given that the growth in housing units is increasing linearly, the growth rate is decreasing, which is expected since the US is experiencing a population growth rate decline.

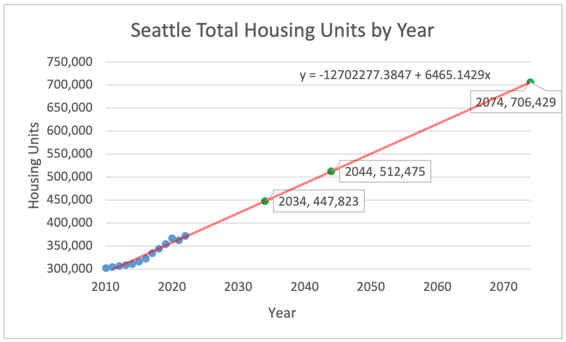

FIG. 3. Plot of Seattle Housing Units by Year. Blue dots represent original data while green dots represent our predicted amount of housing units in the future. (Data)

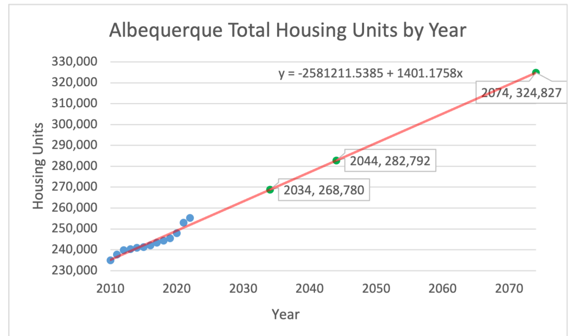

FIG. 4. Plot of Albuquerque Housing Units by Year. Blue dots represent original data while green dots represent our predicted amount of housing units in the future. (Data)

3.7 Sensitivity Analysis

We took out the 2022 data, re-linearized our model and tested our predicted 2022 housing units to the actual data. We calculated our error using the formula:

$$\text{Error} = \frac{|\text{predicted} - \text{real}|}{\text{real}} \cdot 100\%$$

And our data is as follows:

City	Real	Predicted	Error
Seattle	369410.195	362809	1.82%
Albuquerque	250749.252	252924	0.86%

3.8 Strengths and Weaknesses

3.8.1 Strengths

1. By using a linear regression model, we can predict the average housing supply in the future, which is superior because it is difficult to predict major events that cause huge temporary increases/decreases in housing supply.

2. By having a polynomial degree of 1, we are less sensitive to outliers when compared to multi-degree polynomials such as quadratics.

3. Due to its simplicity, our model is easy to understand and interpret, allowing for an easier time to notice errors or abnormalities in data.

3.8.2 Weaknesses

1. Using a linear regression model assumes that the housing supply growth is constant, which could be not true. The housing unit growth rate could be on a downward or upwards trend, but the small dataset prevents us from seeing that trend.

2. Our model is not sensitive to more complex relationships between the housing supply and growth. Many other factors not shared between population and housing could affect the housing supply growth rate, and using a linear regression model will not catch that.

3. The population growth rate might not be proportional to the housing supply growth rate, which we are assuming to be true. Whether it's because housing companies are incentivized to build more houses than people or each person buys multiple houses and leaves them vacant, the housing unit growth rate could be higher than the population growth rate.

4. COVID might have negatively affected the housing supply growth rate, causing $\frac{3}{13}$ of our data (2020–2022) to be lower than expected. This could cause our projected growth rate for housing supply to be lower than expected.

5. More data can ensure that our model is accurate and our assumed relationships are true.

4 Part II: It Was the Worst of Times

4.1 Restating the Problem

In the second problem, we are tasked to predict the changes in the homeless population over the next 50 years, which we interpret as predicting the homeless population over the next 50 years. We chose the following 4 factors (housing price, poverty rates, real GDP, population) to predict the changes in the homeless population in these cities.

4.2 Assumptions

1. **There are no significant policy changes, natural disasters, or unforeseen influences in either Seattle or Albuquerque.**
   Justification: It isn't feasible to numerically quantify the impact of these influences on the homeless population, nor is it feasible to find a predictable trend.

2. **The amount of intervention of non-profits and volunteer groups in supporting the homeless will not be considered.**
   Justification: The amount of homelessness is influenced based on the support received from non-profits, so it can't be ignored entirely. However, their level of invention can be fairly erratic due to losses in donations, the number of volunteers increasing/decreasing, etc. As such, it will not be a factor in our model.

3. **The housing price is related to the homeless population**
   Justification: For housing price, if the housing price goes up, the houses are less affordable, thus fewer people can obtain houses to live in, so more people become homeless.

4. **The real GDP is related to the homeless population**
Justification: GDP is correlated with economic strength, which is correlated with interest rates. When interest rates are low, people are more willing to take loans to buy houses, which decreases the homeless population,

5. **The poverty rate is related to the homeless population**
Justification: When the poverty rate is higher, fewer people can purchase houses, which increases the homeless population.

6. **The whole population is related to the homeless population**
Justification: We assume that the distribution of income ranges for the whole population is about constant for the simplicity of the model. When the population increases, we assume the more impoverished are incapable of buying houses, which increases the homeless population.

7. **The only related factors that influence homelessness are housing prices, poverty rates, Gross Domestic Population, and the regions' population**
Justification: From assumptions 3 to 6, the factors are correlated with homelessness. By choosing various factors, plotting the factors against homelessness, and comparing the variances, we determined that the aforementioned factors have the highest correlation and will be the factors we focus on for our model.

8. **Each of the listed four factors has an approximate constant weight of influence on the homeless population of the regions.**
Justification: Since we ignore all other external factors of homelessness, including the year, the impact of each factor on the homeless population does not change. For example, suppose the poverty rate is twice as significant in predicting the homeless population as real GDP in the year 2019. In that case, this relationship will not change in the year 2020 if no other factors are considered.

4.3 Variables

Symbol	**Definition**	**Units**
H_y	The homeless population at year y	People
P_y	The average housing price at year y	2015 US Dollar
R_y	The poverty rates at year y	Percentage
G_y	The real GDP at year y	Billions of 2017 US Dollar
N_y	The population of the region at year y	People

4.4 Model

4.4.1 Model Development

We split the model into two parts. One part predicts the homeless population for any year in the future based on the four factors stated earlier, and the other part predicts the value of each factor for any year in the future. The individual regression models for each factor is necessary because to predict the homeless population 50 years in the future, the multivariate requires a predicted value for each factor as input. Since the number of data points we have is quite small, we have to be careful as to how we predict and input the stated factors, especially when extrapolating 50 years into the future while having only 13 data points.

To increase the accuracy of the model, we removed one year of data in Seattle (2021) because the homeless data was incomplete[5], and one year of data in Albuquerque (2022) because the homeless numbers spiked up in 2023[7] after dipping in 2022, and while the ideal solution would be to include 2023 data, since 2023 is recent, there is no reliable public source of the Albuquerque GDP in 2023, so omitting both 2022 and 2023 from the Albuquerque data is the solution that we came up with to maintain the integrity of the model.

To predict the homeless population for each region, we use a multivariate linear regression based on the four factors above: housing price, poverty rates, real GDP, and population. This integrates all of the four factors in the optimal proportion to create a decent linear regression. This linear regression will allow us to predict the homeless population based on the other factors previously mentioned. The linear regression will be used instead of other forms of regression because for values 50 years into the future, it will be quite difficult to accurately predict the data and a linear regression minimizes the noise from the data, as well as not growing too fast/slow so the numbers after 50 years will be reasonable. The formula for multivariate linear regression is below, where independent variables are denoted by x, coefficients are denoted by A and B, and the dependent variable is denoted by y.

$$y = \sum_{i}^{n} A_i x_i + B$$

The formula for multivariate linear regression

For predicting the independent variables themselves, the housing price[2], real GDP[8][9][10], population[2], and poverty percent[2] will all be modeled with linear regressions, for the same reason. However, as mentioned in the next subsection, there will be slight modifications to the data to make them more linear to allow a linear regression. While the residuals for the linear regression may look polynomial, making the regression polynomial would scale too fast and be unrealistic for predicting actual homeless population data. By keeping the model linear, we preserve the realistic slow but consistent change in these values, while still being able to relatively accurately predict the homeless population. The formula for linear regression is below. The independent variable is denoted by x. Coefficients are denoted by A and B. The dependent variable is denoted by y, and n is the number of independent variables.

$$y = Ax + B$$

The formula for univariate linear regression

4.4.2 Executing the Model

We adjust our housing prices to not include inflation. Otherwise, the exponentially compounding effects of inflation would render the linear regression ineffective. Thus, we adjusted each year's housing cost to 2015 US Dollars, mainly out of convenience, because the deflator data[11] was relative to the 2015 US Dollar value.

When extrapolating a linear regression model of poverty rate vs time, the poverty rate in 50 years will be negative, which doesn't make sense. Therefore, we did a linear regression between the year and the natural log of the poverty rate to capture the aspect that poverty will decrease at a decreasing rate so that the poverty rate can reach close to 0 but never 0, much less be negative. For the other variables, GDP was already adjusted for inflation, since it was real GDP, and the population is already linear.

To compute the logarithm, we used the numpy.log function in Python, and to do the linear regression, we used the sklearn. We then use the LinearRegression class to process the data and spit out a linear model. The library itself implements the formulae below to compute the coefficients of the linear regressions, after adjusting the data and taking the logarithm of the poverty percentage. Here, x denotes the independent variable or the year, and y denotes the dependent variable, after being adjusted to be more linear. $\overline{x}$ and $\overline{y}$ are used to denote the average of these variables.

$$A = \frac{\sum (y-\overline{y})(x-\overline{x})}{\sum (x-\overline{x})^2}$$
$$B = \overline{y} - A\overline{x}$$

The formulae for coefficients of univariate linear regression

To compute the multivariate linear regression, we can compute the following, where A is the column vector of coefficients, X is the $n \times k$ matrix of independent variables, and y is the column vector of the dependent variables, and B the y-intercept. Let $\overline{y}$ denote the mean of the dependent variable:

$$A = (X^T X)^{-1} X^T y$$
$$B = \sum_{i}^{k} y_i - \overline{y}$$

The formulae for coefficients of multivariate linear regression

Before computing any of the above regressions, we first used min-max normalization to normalize the data to make the coefficients reasonably sized, and minimize the precision issue caused by programming languages. After normalizing the data and obtaining the projected values for the independent variables to the multivariate linear regression, we solve for the multivariate linear regression, obtain the projected homeless population, and undo the normalization to find the actual projected values.

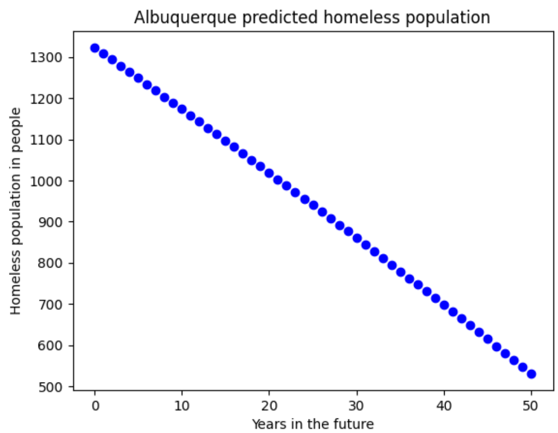

FIG. 5. Graph of Albuquerque homeless population over the next 50 years

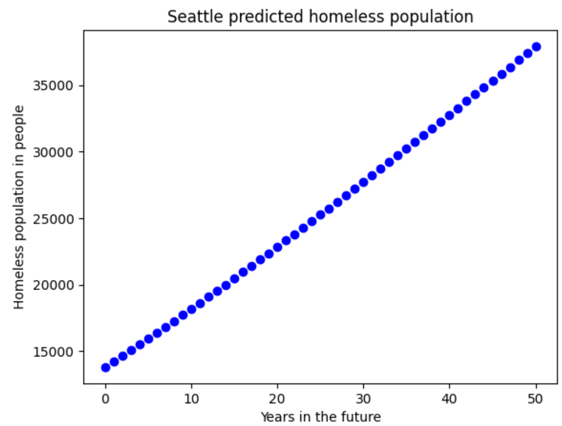

FIG. 6. Graph of Seattle homeless population over the next 50 years

y	H_y	G_y	R_y	N_y	P_y
2034	1173	48.11	14.79	597155	262388
2044	1019	51.64	13.39	621297	303909
2074	531	62.23	9.95	693724	428472

TABLE 1. Albuquerque in year y

4.5 Results

In FIG 5, FIG 6, TABLE 1, and TABLE 2 are the Albuquerque and Seattle homeless populations that are predicted over the next 50 years. We found that in 10 years, the predicted homeless population of Albuquerque will be 1,173 people, in 20 years it will be 1,019 people, and in 50 years it will be 531 people. In Seattle, in 10 years it will be 18,169 people, in 20 years it will be 22,837 people, and in 50 years, it will be 37,916 people.

4.6 Discussion

Our linear regression models show that in the next 50 years, we can expect an increase in the homeless population of Seattle, while we expect a decrease in the homeless population of Albuquerque. Although the graphs look linear, they are not because of the logarithmic component we used to compute the regression of the poverty rates.

y	H_y	G_y	R_y	N_y	P_y
2034	18169	688.36	6.84	920496	790572
2044	22837	879.82	4.91	1058641	1020246
2074	37916	1454.22	1.81	1473075	1709269

TABLE 2. Seattle in year y

4.7 Sensitivity Analysis

To test the sensitivity of our model, we will remove the last year in our data and retrain the model, then compare the predicted homeless population and the actual homeless population. Our model predicted that the homeless population in Albuquerque in 2021 would be 1296, compared to the actual value of 1567. Our model also predicted that the homeless population in Seattle in 2022 would be 12772, compared to the actual value of 13368.

$$\text{Percent Error} = \frac{|\text{actual} - \text{predicted}|}{\text{actual}} \times 100\%$$

Plugging into the formula, we get a 17.31% error for Albuquerque, and a 4.46% error for Seattle. Given the relatively high percent errors, it is reasonable to assume that the model roughly predicts the magnitude of the homeless population, but the accuracy leaves much to be desired.

4.8 Strengths and Weaknesses

4.8.1 Strengths

1. A multivariate regression can account for all four factors when predicting the homeless population. Because of the simplicity of how the model predicts, it can quickly predict the homeless population many years ahead.

2. Because the predictions are decades into the future, the linear regression model in predicting each factor is less prone to being influenced by long-term predictions compared to more complex models. For example, a quadratic regression of the real GDP could have a much more extreme prediction in 50 years compared to our linear regression model.

4.8.2 Weaknesses

1. The model only considers the real GDP, the population, poverty rate, and housing prices. Therefore, other factors that are related to homelessness, such as the percentage of drug abuse and employment rate, could have better refined our homeless population prediction. In reality, this would also break the assumption that each factor has a constant percentage of significance in influencing the homeless population. It is also possible that our chosen factors are not the best for predicting the homeless population.

2. The model assumes, for real GDP, population, and housing prices, a linear relationship between each factor and the homeless population, but this may not be the most accurate modeling of their relationship. The linear regression between the natural log of the poverty rate and time may also not be accurate. More complex models, such as a polynomial or Long-Short Term Memory (LSTM) model, could be more accurate in predicting the values of the factors.

3. The data for the regression models does not reflect the assumptions made. For example, in 2016, the government implemented higher public spending and reduced taxes, which

is an external, uncontrollable factor in the data. However, both the linear regression and multivariate regression don't account for this factor in the data, so the model is not completely accurate in predicting the homeless population based on real GDP, population, poverty rates, and housing prices.

4. Because the linear regression models for each factor are not the most accurate, the multivariate regression model, which takes in the predicted values of each linear regression model, becomes even less accurate.

5 Part III: Rising From the Abyss

5.1 Restating the Problem

In this final section, we are tasked with developing a model that could potentially assist a government in developing a long-term program to address homelessness. It should be responsive to external factors, such as an influx of migrants, a natural disaster, or an economic recession.

5.2 Assumptions

1. **When developing plans to address homelessness, the primary limiting factor is cost.**
Justification: Cities have limited monetary resources to tackle this problem. Therefore, we will develop our model under the presumption that cities would like to minimize the cost of reducing homelessness.

2. **The homeless population can be classified into two main groups: the chronically homeless and those who are temporarily homeless**.
Justification: Some individuals, such as victims of natural disasters or eviction, may be briefly homeless while looking for a new place to live, but typically require help only on a short term – these we term *temporarily* homeless. Other individuals, such as those facing mental illness or substance abuse, may continually require intensive assistance and costly resources – these we term *chronically* homeless. Most homeless individuals can be approximately categorized as one of the two types, and as their situations are very different, we will separately analyze these two groups.

3. **Temporarily homeless individuals and chronically homeless individuals cost the same to house per month**
Justification: Homeless support programs typically provide the same resources to all of the individuals they serve, so there is no expense distinction between the two types of homelessness per unit of time. Thus, we can assume that housing someone temporarily has a roughly equivalent monthly cost to housing someone more permanently.

4. **The unit cost of housing a homeless person remains roughly constant across the years.**
Justification: After the threshold of their invention, the same tools, resources, and

technologies should be available across the homeless population, no matter which year it is. Therefore, we assume the unit cost of housing a homeless person is stable across different years.

5. **A temporarily homeless individual will not need assistance for more than two months**
Justification: Temporarily homeless individuals are those who come into homelessness by a ruinous but rare event such as a layoff or a natural disaster. Thus, these individuals often have a support system (insurance, family, technical skills, etc) around them that can help them exit homelessness quickly. Based on this, we can assume they will not be homeless for more than two months. In fact, on average, stays in a homeless shelter last 77 days[17]

6. **A widespread permanent supportive housing program is enacted to improve the chronic homelessness situation**
Justification: Our model is designed to simulate the costs and impact of a permanent supportive housing program in addressing homelessness, so that community leaders can evaluate whether this is a feasible solution. Hence, we will be operating under the condition that this program has been implemented.

7. **The population of Seattle can be classified into four groups: the *not homeless*, *chronically homeless*, *temporarily homeless*, and *in permanent supportive housing* (PSH).**
Justification: Outside of the typical "not homeless" group and the two homeless groups, we also separately classify those whose chronic homelessness is being actively resolved through the permanent supportive housing program. As stated before, we are simulating the solution of PSH and thus assume that Seattle creates and encourages chronically homeless people to use this opportunity.

8. **Individuals currently in PSH will either remain in PSH or become not homeless; in particular, they will not revert to being chronically homeless.**
Justification: Studies have shown that Permanent Supportive Housing programs have high retention rates and are very effective in rehabilitating chronically homeless individuals[16]

9. **Migrants to Seattle are initially classified as *not homeless.***
Justification: Homelessness is caused by a variety of factors. It is not well-defined how many people who are born or have migrated to Seattle are homeless; therefore, for the sake of simplicity, we will assume that additions to the population in Seattle originally have housing.

5.3 Variables

Symbol	Definition	Units
n_t	Number of Temporarily Homeless Individuals	People
n_c	Number of Chronically Homeless Individuals	People
n_h	Number of Not Homeless Individuals	People
n_p	Number of Individuals in PSH	People
p	Percent of Chronically Homeless Individuals that take part in PSH	Percentage
C	Cost Per Month of Housing One Individual in PSH	US Dollars
C_h	Cost Per Month of a Homeless Individual Living on the Streets	US Dollars
N	Number of Years	Years

5.4 Constants

Symbol	Definition	Value
p_c	Percent of Newly Homeless Individuals who become Chronically Homeless	10%[12]
p_h	Percent of Not Homeless Individuals who become Homeless Per Month	0.15%[12]
K	Population Growth of Seattle per Month	1151[2]
Y	Average Length of Time Spent in PSH	5 Years [15]

5.5 Model

5.5.1 Model Development

City authorities must weigh various factors in designing and implementing plans to alleviate homelessness, including administration costs, infrastructure expenses, and program efficacy.

One particularly promising approach is Permanent Supportive Housing (PSH), which offers long-term access to resources to homeless individuals. Existing ways of handling the problem of chronic homelessness are often very expensive, costing anywhere from 35,000 to 150,000 dollars per year in medical fees and emergency services. In fact, it is often more cost-effective to simply provide housing to the homeless individual – which is precisely what the PSH program does.

To facilitate city authorities in deciding whether PSH is a feasible option, we create the following model, which simulates the changing demographic of Seattle with the implementation of a PSH program.

Our model aims to estimate the persistent and potential temporary costs of homelessness in a city. To do so, we split the problem into two parts; first, we used a modified Markov chain to find the number of individuals who are temporarily homeless, chronically homeless, or in permanent supportive housing given a particular starting month configuration by iteratively updating the state of Seattle's homeless population. We chose to model the situation with a Markov Chain as there are various separate groups of people that interact and influence the sizes of each other, making a matrix a neat way to organize all these states and interactions. However, Markov Chains typically require all changes to be a linear combination of the group sizes, and, as established in Part 1, the population growth of Seattle is a scalar constant. Hence, we modify the model to account for linear population growth.

5.5.2 Executing the Model

First, we create our matrix to determine the number of people who will enter or exit homelessness; we use the matrix to calculate the change in the homeless population and modify the resulting vector to account for population growth.
We note that in Seattle, the percentage of people who fall into homelessness is hard to determine; however, therefore, based on the current homeless population of Seattle, we estimate that roughly 1.8% of individuals in the "not homeless" population fall into homelessness per year, translating to $p_h = 0.15\%$ per month. To determine whether they are chronically or temporarily homeless, we note that, generally, $p_c = 10\%$ of individuals are chronically homeless [citation needed]. Therefore, we will split this 0.15% into $p_c \cdot p_h = 0.015\%$ falling into chronic homelessness and $0.15 - 0.015 = 0.135\%$ falling into temporary homelessness. Now, how many chronically homeless individuals take part in PSH is dependent on how thoroughly the city promotes the program, so this is a variable p; we will demonstrate our model using $p = 20\%$ for now, but the precise value can be varied to observe the concomitant differences in homelessness. Thus, in the next month, 20% of those in chronic homelessness will be placed in the PSH population, while around half of those in temporary homelessness will find a new home. Additionally, we find that the average length of time spent in PSH is $Y = 5$ years[15], hence every month around $\frac{1}{5 \cdot 12} = \frac{1}{60}$ of PSH members become not homeless. At the end of each month, we will add $K = 1151$ people to the *not homeless* population because, based on our model in Phase II, we divide Seattle's yearly population growth by 12 to find that the monthly population increase would be around 1151 individuals.

Second, we perform cost analysis for each of the groups. We note that at present, it costs around $C = \$1153$ per person in Permanent Supportive Housing (PSH) [citation needed]. Optimally, everyone who is chronically homeless is placed in PSH within a month, since PSH is cheaper [citation needed]. Before that, however, someone who is chronically homeless and unsheltered costs the city $C_h = \$2897$[13]. Since people who are temporarily homeless require roughly the same amount of resources, we treat those who are temporarily and chronically homeless similarly in terms of cost to the city.
Finally, we set our starting values for the number of not homeless individuals n_h, temporarily homeless individuals n_t, chronically homeless individuals n_c, and the number of individuals in PSH n_p. We also set the number of years N we want to run the model for. For now, we let the $n_h = 715700$, $n_t = 9224$, $n_c = 4144$, $n_p = 5000$ to match 2022 Seattle data.

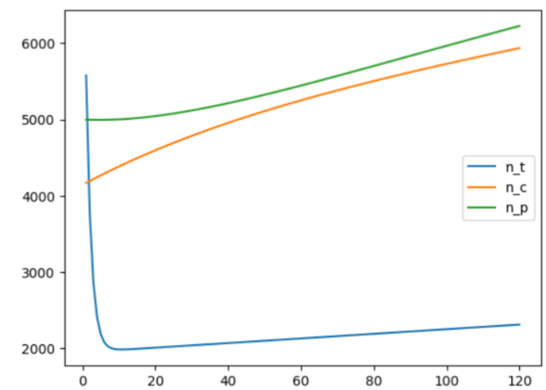

FIG. 7. Graph of Seattle Populations with No Policy Change. Total Cost: 3.3 billion

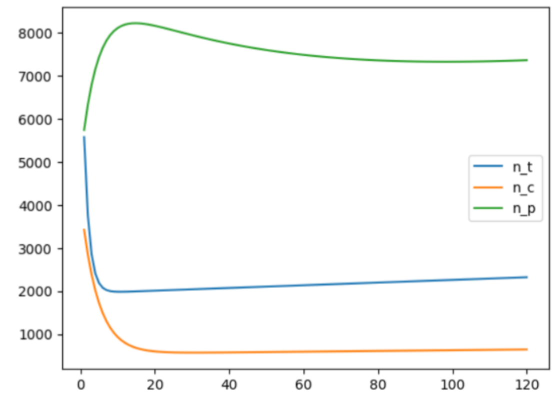

FIG. 8. Graph of Seattle Populations with Major Policy Change. Total Cost: 2.5 billion

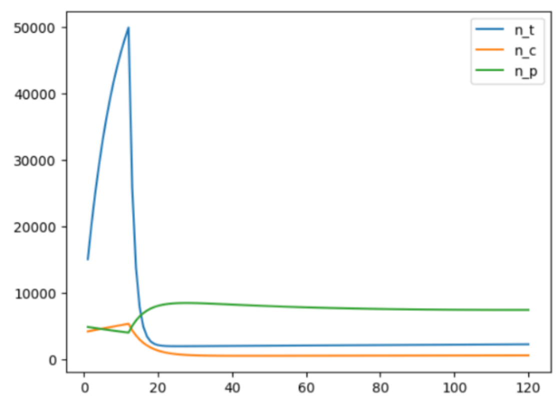

FIG. 9. Graph of Seattle Populations with Major Recession and Major Policy Change. Total Cost: 3.5 billion, with 1.47 billion first year cost

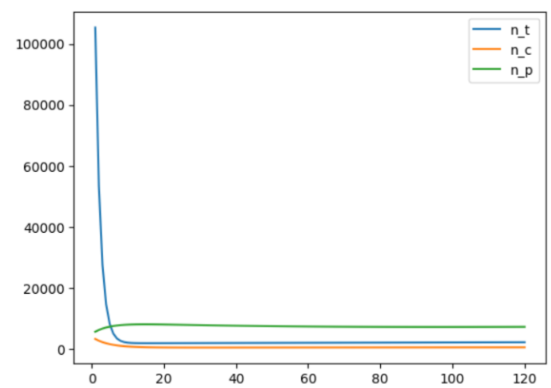

FIG. 10. Graph of Seattle Populations with Natural Disaster. Total Cost: 2.6 billion, 821 million in first year

5.6 Results

Our results are demonstrated in Figures 7 through 10. We can see the effects of various events on the Seattle homeless population. Population data generated by the model is graphed over a span of 10 years.

5.7 Discussion

We see that the implementation of PSH causes an initial spike in the number of homeless people using it, but as time progresses the chronically homeless population and also number of PSH users steadily decrease. The temporary homeless population remains fairly constant, but temporary unemployment is a natural consequence of a changing society and hence not too major a concern. This demonstrates that PSH is effective in addressing the homelessness problem.

On the contrary, in a model where PSH is not promoted, we see that chronic homelessness steadily increases, incurring hundreds of thousands of dollars of cost.

We see that the projected cost of PSH across 10 years is \$2.5 billion, which is cheaper than the projected cost with current levels of PSH, which is \$3.3 billion.

5.8 Sensitivity Analysis

We see that the benefits of PSH persist despite various unexpected events causing changes in the configuration. For example, when we introduce a natural disaster event to Seattle, we account for this in the model by increasing n_t by $200,000$ and decreasing n_h by $200,000$ people – that is, making $200,000$ non-homeless individuals temporarily homeless instead, as many individuals are displaced from their houses. This manifests in a huge spike of temporary homelessness in the short run, but stabilizes to a similar pattern as before when PSH was implemented, costing \$2.6 billion over 10 years.

Similarly, when we simulate an economic recession, we assume that a greater proportion of individuals become homeless, increasing p_h, and that less taxpayer dollars are available to fund PSH, decreasing p. This results in an abrupt change the first year the recession hits, but also stabilizes over time, costing \$3.5 billion over 10 years.

Overall, we see that our model is relatively robust, and handles extreme circumstances relatively well.

5.9 Strengths and Weaknesses

5.9.1 Strengths

1. Our model accounts for many factors that influence the population of homeless individuals in Seattle and is thus better able to capture the complexity of how these populations behave relative to each other.

2. Our model is versatile in allowing for the customization of various parameters of the situation, thus allowing cities to plug in data that reflects their own situation and experiment with varying degrees of PSH implementation.

5.9.2 Weaknesses

1. Many general homelessness statistics are not well-documented, so gathering accurate, consistent data about homelessness is difficult. Hence, the parameters we use in our model may be somewhat inaccurate.

2. Our model fails to account for various other barriers to implementing PSH on a wide scale such as infrastructure costs, promotional costs, public resistance, and political

barriers. These obstacles make the implementation of PSH possibly more costly and unfeasible.

3. There are many more factors that affect homelessness such as affordable healthcare, inflation, and economy that we do not specifically account for in our model.

6 Conclusion

6.1 Further Studies

The COVID pandemic in 2020 affected people's ability to travel from city to city. With more time, we could develop two models, one before COVID and one after COVID, to remove potential outliers in our data. Additionally, more research could be done to determine the movement of homeless populations throughout Albuquerque and Seattle.

6.2 Conclusion

In the first question, we used linear regression to model the growth of the housing supply in two U.S. cities, Albuquerque and Seattle. This model predicted that in 10, 20, and 50 years, the housing supply increased to 268780, 282792, 324827, 448823, 512475, 706429 units in Albuquerque and Seattle, respectively.

In the second question, we used multivariate linear regression supported with linear regression to predict the homeless population. The model predicted a decreasing homeless population in Albuquerque and an increasing homeless population in Seattle.

In the third question, we built a model to estimate the effects of various policy changes and events on Seattle's homeless population. We also determined the cost of implementing PSH, a potential solution to the homelessness problem.

In conclusion, our findings have concluded that in the midst of cities with predicted increasing and decreasing homeless population, we have found a viable solution to decrease the cost of homelessness.

7 References

1. https://usafacts.org/articles/how-many-homeless-people-are-in-the-us-what-does-the-data-miss/
2. https://m3challenge.siam.org/kdfrldh/
3. https://www.nber.org/research/data/us-business-cycle-expansions-and-contractions
4. https://population.un.org/wpp/Download/Standard/MostUsed/
5. https://dchsblog.com/2020/12/17/2021-point-in-time-street-count-canceled/
6. https://data.census.gov/table/ACSDP1Y2022.DP04?q=albuquerque%20average%20household%20si
7. https://www.abqjournal.com/this-years-point-in-time-count-identified-2-394-people-experiencing-homelessness-in-albuquerque-organizers/article_072d451c630d11ee8b2917db0ad4a084.html
8. https://fred.stlouisfed.org/series/RGMP10740
9. https://tradingeconomics.com/united-states/total-gross-domestic-product-for-albuquerque-nm-msa-fed-data.html
10. https://www.statista.com/statistics/183863/gdp-of-the-seattle-metro-area/
11. https://data.worldbank.org/indicator/NY.GDP.DEFL.ZS?end=2022&locations=US&start=1960&v
12. https://www.theurbanist.org/2019/04/01/seattle-has-the-solution-to-chronic-homelessness-we-just-need-to-bring-it-to-scale/
13. https://www.politifact.com/factchecks/2012/mar/12/shaun-donovan/hud-secretary-says-homeless-person-costs-taxpayers/
14. https://repository.upenn.edu/spp$_{p}apers$/148/
15. https://www.security.org/resources/homeless-statistics/
16. https://www.desc.org/what-we-do/housing/
17. https://www.affordablehousingpipeline.com/blogs/california-affordable-housing/what-is-permanent-supportive-housing
18. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8833310/