Modeling the Population of California

Brady Biehn

April 2024

1 Abstract

We analyze two models of population dynamics in California and the rest of the United States. We assume that both regions had a constant death and birth rate, and each received an influx or efflux of migrants for each 5-year time step. We calculate the birth, death, and either the net migration or transition rates (both forms of migration rates) and use the summation (we multiple the death rate by -1) of these rates to model the population growth rate for each 5-year time step. We use Linear Algebra to make long-term predictions of the populations of each region. For the model in which we used the net migration rate, we found that in the long term, the population of California would grow by 17.077%, and 100% of the total population would end up in California. For the model in which we used the transition rate, we found that in the long term, the total population would grow by 8.03% and that 17.827% of the total population would end up in California.

2 Introduction

In this paper, we focus on a comprehensive analysis of population dynamics between California and the rest of the United States. California is known for having one of the biggest economies in the world, starting the most new businesses, and being the top state for getting money for new businesses (California, State of). Moreover, California's appealing Mediterranean-like climate and breathtaking beaches are attractions for visitors and residents alike, as noted by Zack & Allen. With its highly diverse population much of California is in several prominent urban centers. California is a focal point of demographic shifts, as highlighted by the World Population Review. Understanding these movements is necessary because it helps us to understand the trend of population changes over time.

We examine factors influencing population changes, including migration, births, and deaths. By analyzing the number of people moving in and out, as well as birth and mortality rates, we aim to understand long-term population trends. We use the summation of the birth, death (we multiple the death rate by -1), and migration rates to model the population growth for every five-year time step. Our model is designed to shed light on population dynamics and their broader societal implications, guiding policymakers and planners in address population shifts and changing migration rates in their future decisions.

3 Model Description

We use data compiled from Mooney & Swift, the United States Census Bureau and the website Population Pyramids to find the actual population of California and the rest of the United States. For Mooney & Swift we use their population data for the years 1955 and 1960. The authors found that in 1955 the population were 12988 and 152082 (per 1000 people) in California and the rest of the United States, respectively. We use these values as our initial populations for each region. For 1960, Mooney & Swift found that the population of California was 15206 (per 1000 people), and the population of the rest of the U.S. was 163040 (per 1000 people). We will use time steps of every five years, so the year 1960 will serve as our first time step. The U.S. Census takes place every ten years and it is mandated by the U.S. Constitution to count all residents within the U.S. It is considered to be the best estimation of the United States' population. However, one problem with using the Census is due to its nature of only counting people every ten years. Since we use five-year time steps, the Census would not aid us half the time. To combat this issue we also used data from the website Population Pyramids, which is a website dedicated to displaying and visualizing populations. We used the Census for every year that is a multiple of 10 (the year the Census takes place) and Population Pyramids for the other years within our time steps (outside of 1955 and 1960). We stop compiling population data at the year 2020.

For both models, we assume that individuals (per 1000 people) die where they reside. This concept means that if someone lives in California, they can not drive to Arizona and die. If they live California, they die in California. If they live in the rest of the United States, they die in the rest of the United States. Our second general assumption is that if someone is born in the region where their parents live. For example, if someone's parents live in California, they are born in California. That individuals parents can not travel to Oregon, have the baby and go back to California. This assumption is similar to the previous one but for birth rates. The final general assumption for both models is that we assume constant birth and death rates for each time-step. After we calculate these rates, they will never change. It should also be noted that all population data is in units of per 1000 people.

3.1 Net-Migration Model

To construct the net-migration model, we use a recursive equation to find the population of two different regions based up upon previous populations. There are two different regions within our model which are California and the rest of the United States. We calculate the birth, death, and net-migration rates for each region. Note that the death rate is subtracted. We assume that the rates remain constant for all time step. Each of these rates is derived from using our initial populations of 12,988 and 152,082 units of California and the rest of the U.S respectively. The birth, death, and net-migration rates are inserted into their own distinct 2x2 matrix. We calculated the birth rate by using dividing the total number of births between 1955-60 within that particular region by the total population of that region in 1955. There were 1708 units born between 1955 through 1960 in California, and 19499 units born in the rest of U.S. during that same time period. To calculate the death rate we divided the total number of deaths in the region between 1955-60 by the region's total population in 1955. In California, there were 614 people units who died between 1955 and 1960. In the rest of the U.S., there 7417 units who died between 1955 and 1960. Finally, to calculate the net-migration rate, we divided the net migrates (the amount of immigrants subtracted by the emigrants) divided by that region's initial population. Between 1955-60, a net of 1124 units migrated to California, and -1124 units emigrated from the rest of the U.S. Using these formulas we calculated the rates shown in

Table 1:

Table 1

Region	California	Rest of U.S.
Birth Rate	0.1315060	0.128213727
Death Rate	0.0472744407	0.04876974
Migration Rate	0.086541423	-0.00739074973

Rates 1955-1960 per 1000 people

For matrix B, which represents the birth rate, we get the matrix

$$B = \begin{pmatrix} 0.1315060 & 0 \\ 0 & 0.128213727 \end{pmatrix}. \tag{1}$$

The value at B_{11} represents the additional 13.15060% units that are born for each time step. For example in California, if birth rate was the only factor being applied to the initial population, in 1960 (the first time step), the population would now be 14696 units. There is additional 1708 units, or in other terms of rates an additional 13.15060% of newborns added to the initial population. The rate at B_{22} represents the same thing as B_{11} but for the rest of the United States. The rates for B_{12} and B_{21} are both 0, because you can not be born in California while being in rest of the United States and vice versa. For matrix D, which represents the death rate, we get the matrix

$$D = \begin{pmatrix} 0.0472744407 & 0 \\ 0 & 0.04876974 \end{pmatrix}. \tag{2}$$

The rate at D_{11} is the amount of people (per 1000) that die for each time step in California. For example, if death rate was the only affecting our initial population in California, in our first time step of 1960 the new population of California would be 12374 people (per 1000). This value is a 4.72744407% decrease to the initial population of 12988 (per 1000) in California. The rate for D_{22} represents the same thing as D_{11} but for the rest of the United States. The rates at D_{12} and D_{21} are zero, because it is illogical for someone to die in California while being somewhere in the rest of the U.S. The final matrix is matrix M, which shows the net-migration rate for each region. These migration rates were placed into the matrix M where

$$M = \begin{pmatrix} 0.086541423 & 0 \\ 0 & -0.00739075 \end{pmatrix}. \tag{3}$$

The rate at M_{11}, which is equal to 0.086541423, means that there is an 8.6541423% increase in the number of people that migrated to California. The rate at M_{22} means a 0.739075% decrease to the region's total population. If a net-migration rate is positive that means that more people immigrated than emigrated, which, in other words, means that more people moved into the region than moved out. If a net-migration rate is negative that means more people moved out of the region than moved into it. The two rates at M_{12} and M_{21} represent the rate at which the regions intermingle with each other. For the net-migration Model, this phenomenon does not occur so the proportion is 0.

These matrices were used to create our final matrix G, or what we will refer to as our "growth" matrix because it represents the population growth rate for each region. To obtain the growth matrix we used the formula

$$G = I + B - D + M \tag{4}$$

where I is the 2x2 identity matrix. We use the identity matrix because it represent the initial population percentage of 100%. We add B because when someone (per 1000 people) are born there are added to the population. We subtract D because, as seen in the description of D, when someone dies (per 1000 people) the population decreases. Finally, we add the net-migration rate. As shown above, one of the migration rates are positive and the other is negative. Due to the nature of addition the sign is kept, so when M_{11} is added to G_{11} the population growth rate will increase, which reflects the fact that more people are immigrating to California than are emigrating for it. When M_{22} is added to G_{22} the population rate decreases, because more people are leaving the rest of the U.S. than are moving into it. When using equation 4 the resulting matrix is

$$G = \begin{pmatrix} 1.17077 & 0 \\ 0 & 1.07205 \end{pmatrix}. \tag{5}$$

The rate at G_{11} shows the rate that population of California increases for each time step. The rate for G_{22} represents the rate that the rest of the United States' population increase per time step. We use these rates in equation 6.

$$X(t+1) = GX(t) \tag{6}$$

t is our time step, which is every five years, and X(t) is our 2x1 population vector with the components $x_{California}$ and $x_{RestofU.S.}$. The component $x_{California}$ represents the population of California, and the component

$x_{rest of the U.S.}$ represents the population of the rest of the United States. We will show an example of equation 6 by calculating the population of 1960, or when t = 1 (remember 5 year time-steps). We input the matrix of our initial populations (1955) for X(t), that matrix is $\begin{pmatrix} 12988 \\ 15082 \end{pmatrix}$. When then multiply X(t) from the left by our growth matrix. The resulting matrix is $\begin{pmatrix} 15206 \\ 163040 \end{pmatrix}$. For our equation the $X(t+1)_{11}$ is the population of California and $X(t+1)_{21}$ is the population of the rest of the United States. For our example the resulting matrix the population of California is 15206 (per 1000 people) and the population of the rest of the U.S. is 163040 (per 1000 people). These values match exactly with those found by Mooney & Swift.

3.2 Transition Model

In the transition model, we use the same matrices for birth and death rate. However, we made a significant change in how we are modeling the migration rate. Instead of using the net-migration rate, we use the actual transition values, which allows us to calculate the probability that a group of 1000 people will migrate between the regions. We show the actual transition results in Table 2:

Table 2

	To California	To Rest of U.S.
California	12174	814
Rest of U.S.	1938	150144

Actual Number of Transitions Between 1955-1960 in Thousands

The 12174 values in the first row and column in Table 2 tells us that 12174 (per 1000 people) who lived in California still live there in the next time step. The 150144 (per 1000 people) tells us the same thing but for the rest of the United States instead of California. The 814 number is the first row and second column shows that 814 (per 1000 people) who lived in California move to the rest of the U.S. in the following time step. The 1938 value in the second row and first column represents that 1938 (per 1000 people) who lived in the rest of the U.S. move to California in the following time step. We use these values to calculate the transition rate to do this we take the for actual for each type of transition, and divide by the initial population of that region. For example, we take the 814 (per 1000 people) value and divide by California's 1955 population of 12988 (per 1000 people). When we do this we get the actual transition rate of 0.06267328. We calculate this rate for each type of transition within the system. The resulting matrix T shows the transition rate per time step for each type of transition:

$$T = \begin{pmatrix} 0.937326732 & 0.06267328 \\ 0.012743125 & 0.987256875 \end{pmatrix}. \tag{7}$$

The rates at T_{11} and T_{22} are the probability that a group of 1000 people is going to remain where they are, which means that if they lived in California in 1955, then there is 93.732 percent chance that they would remain there in 1960. The transition rate for T_{22} shows the same phenomenon if the group would remain within the bounds outside of California. The rates T_{12} and T_{21} represent the probability that someone will migrate between regions. The transition rate at T_{12} represents the chance someone (per thousands) who lived in California during the previous time step moves somewhere within the rest of the U.S. The rate T_{21} shows the opposite. We used the transpose of T to construct a new growth matrix. The transition matrix is a non-absorbing Markov chain due to the ability of the population to fluctuate between the two regions without ever getting stuck in either. It should also be noted that rows for the transition matrix sum to 1. We used the formula $G^* = B - D + T^T$, the resulting matrix was

$$G^* = \begin{pmatrix} 1.02156 & 0.0127431 \\ 0.0626733 & 1.0667 \end{pmatrix}. \tag{8}$$

We use the transpose of T because we want the column to match the other matrices. For B and D, column 1 represents California, and column 2 represents the rest of the U.S. For T, the first row represents California, and the second row represents the rest of the U.S. To combat this issue when we added them together we used the transpose of T. For G^*, G^*_{11} tells us the rate at which the population of California grows for every 5-year time step, and G^*_{22} tells us the rate that the rest of the U.S grow per time step. The rate for G^*_{12} tells us the chance that someone who lived in the rest of the U.S. moves to California in the next time step, and the rate for G^*_{21} tells us the opposite relationship (California to the rest of the U.S.). Due to the new nature in which we compute the migration rate, the identity matrix is no longer required when computing G^* because it already factors in the previous time step's population. We then calculate the population

$$X(t+1) = G^* X(t) \tag{9}$$

using equation 9. Equation 9 is implemented the same way as equation 6 but it uses our new population growth rates.

4 Model Analysis

To analyze the Net-Migration and Transition Models, we compute the dominant eigensystem. This system consists of the largest eigenvalue and that value's associated eigenvector. The dominant eigenvalue tells us the rate at which the population grows in the long term, and the dominant eigenvector tells us the long-term proportions of the population. For example, let's say, hypothetically, that the dominant eigenvalue was 1.1 and its associated eigenvector was equal to $\begin{pmatrix} .5 \\ .5 \end{pmatrix}$. These values would tell us that the total long term population would grow by 10% and that exactly 50% of the total population would end up in each region in the long-term.

For both models, we simulated the population of each region starting at 1960 through 2020 using 5 year time steps. To simulate the population predicted by the net migration model we use Equation 6. We use Equation 9 to calculate the population predicted by the transition model. For both simulations, we use our initial population of 12988 and 152082 (per 1000 people) as our initial inputs. We compare the results for each simulation to the actual population at that given year.

4.1 Net-Migration Model

To calculate the dominant eigensytem for the Net-Migration Model we use the the growth matrix G. The results of this calculation are

$$\text{Dominant Eigenvector} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{10}$$

$$\text{Dominant Eigenvalue } (\lambda) = \begin{pmatrix} 1.17077 \end{pmatrix} \tag{11}$$

Thus the dominant eigenvalue is 1.177077. This tells us that the population in California is growing at a rate of 17.077% in the long term. Furthermore, the associated dominant eigenvector provides insights into to a point in the future where the population distribution stabilizes. This tells us that eventually 100% of the population will be in California and none in the rest of the U.S. The reason our net migration model yielded 100% population for California and 0% for the rest of the U.S. lies in the nature of the data and assumptions used in the model. Net migration models typically focus on migration patterns between regions, considering factors such as immigration, emigration, and internal migration. In this case, if the net migration into California significantly outweighs migration out of California, it can lead to a scenario where California's population grows while the population in the rest of the U.S. remains stable or declines which causes the long-term population in California to increase to include 100% of the total population.

Conversely, if the net migration out of the rest of the U.S. is substantial or if the population growth rate in other regions is relatively low, it could result in a situation where the population in the rest of the U.S. remains stagnant or declines over time.

Overall, the 100% population in California and 0% in the rest of the U.S. in the net migration model may reflect the model's assumptions about migration patterns and population dynamics rather than an accurate representation of real-world population distribution.

4.2 Transition Model

The transition matrix T captures the probabilities of transitioning between different states, reflecting movements between California and the rest of the U.S. We used the matrix T and calculated its eigenvector and associated dominant eigenvalues, and we determine the stable populations for people in California and rest of the U.S.

$$\text{Dominant Eigenvector} = \begin{pmatrix} 0.16897 \\ 0.83103 \end{pmatrix} \tag{12}$$

$$\text{Dominant Eigenvalue } (\lambda) = \begin{pmatrix} 1 \end{pmatrix} \tag{13}$$

The eigenvector indicates the steady-state proportions of the population, suggesting that approximately 16.9% will reside in California, while the remaining 83.1% will be outside of California.

With an eigenvalue of 1, the system is stable, signifying that the population distribution will reach an equilibrium over time. This means that the proportions of residents in California and the rest of the U.S., as described by the eigenvector, will remain constant in the future, reflecting a balanced distribution between the two regions.

To calculate the dominant eigensystem for the Transition Model we use the the growth matrix G^*. The results of this calculation are

$$\text{Dominant Eigenvector} = \begin{pmatrix} 0.17827 \\ 0.82173 \end{pmatrix} \tag{14}$$

$$\text{Dominant Eigenvalue } (\lambda) = \begin{pmatrix} 1.0803 \end{pmatrix} \tag{15}$$

In the transition model, the eigenvectors reveal the stable proportions of the population between California and the rest of the U.S., once the population dynamics stabilize. Specifically, the eigenvector values of 0.17827 for California and 0.82173 for the rest of the U.S. suggest that approximately 17.827% of the population will be in California, while about 82.173% will be outside California.

We compare this with the net migration model, where the dominant eigenvector indicated that eventually 100% of the population would be in California, we see a notable difference. In the transition model, the proportions show a more balanced distribution between California and the rest of the U.S. Additionally, the eigenvalue of 1.0803 in the transition model suggests a population growth rate of approximately 8.03%, which is lower than the 17% growth rate observed in the net migration model. These differences highlight the impact of considering additional factors such as births and deaths in the transition model, providing a more nuanced understanding of population dynamics.

The lower growth rate observed in this part suggests that demographic factors significantly influence population trends. Including births and deaths in the model provides a more realistic representation of population dynamics. Our findings indicate that the comprehensive model offers a more nuanced depiction of population growth and distribution, as it considers a broader range of factors influencing population change over time. Ultimately, by considering a wider array of factors, the transition model enhances the accuracy and reliability of population projections, providing valuable insights for demographic analysis and informed decision-making.

4.3 Model Comparison

Figure 1 shows the results of our simulation from 1960-2020 for each model compared to the real population.

Figure1.png

Figure 1. Population Trends in California and Rest of U.S. (1955-2020) - This line graph illustrates the population dynamics over time, spanning from 1955 to 2020. The graph depicts both the actual population data and the predicted population trends for California and the rest of the U.S.

The population predictions from 1955 to 2020 using the net migration model and the transition model for California and the rest of the U.S. reveal notable disparities when compared with actual population data. We compare the simulated predictions with the actual population data, we observe notable disparities. Specifically, the actual population growth in California appears to be slowing down compared to the population predictions generated by both models. Interestingly, the net migration model predicts a more rapid increase in California's population compared to the transition model. Similarly, for the rest of the U.S., both the net migration and transition models project population growth, but the net migration model forecasts a higher rate of growth. Despite these differences, neither model closely aligns with the actual population data for California and the rest of the U.S.

5 Discussion/Limitations

Although our study offers valuable insights into population dynamics between the regions, it is important to acknowledge certain limitations within our models. Our models simplify complex processes such as migrations, births, and deaths, which may not fully capture real-world data. For instance, while our matrix models consider migration, births, and deaths separately, these processes are interrelated and influenced by various social, economic, and environmental factors. Additionally, our analysis assumes linear growth rates over time, which may not hold in the long term due to factors like changing fertility rates, economic fluctuations, and policy interventions.

Furthermore, our models do not account for potential future events and changes in social and economic conditions that could significantly impact population dynamics. Events such as natural disasters, pandemics, and alterations in governmental policies can have profound effects on migration patterns. Despite these limitations, our study provides a useful framework for understanding population trends. Future research could refine models, improve data quality, and consider additional factors for more accurate prediction. For example, future research could consider additional factors such as socio-economic indicators, urbanization trends, and cultural shifts to improve the accuracy of population predictions. Additionally, incorporating advanced statistical techniques and machine learning algorithms could enhance the predictive power of models. Moreover, studying historical migration patterns and their drivers in more detail could provide valuable insights into future population dynamics. Furthermore, we believe continual validation against empirical data will be crucial for enhancing the reliability of population projections in the future.

6 Works Cited

References

[1] Bureau, U.S. Census. “Historical Population Change Data (1910-2020).” Census.Gov, 6 Aug. 2022, www.census.gov/data/tables/time-series/dec/popchange-data-text.html.

[2] “California Population 2024.” California Population 2024 (Demographics, Maps, Graphs), worldpopulationreview.com/states/california-population. Accessed 7 Apr. 2024.

[3] California, State of. “California Governor’s Office of Business and Economic Development (Go-Biz).” California Governor’s Office of Business and Economic Development, business.ca.gov/. Accessed 7 Apr. 2024.

[4] Mooney, Douglas D., and Randall J. Swift. *A Course in Mathematical Modeling.* Mathematical Association of America, 1999.

[5] “Population Pyramids of the World from 1950 to 2100.” PopulationPyramid.Net, www.populationpyramid.net/united-states-of-america/1965/. Accessed 1 Apr. 2024.

[6] Zack, Liz. “Is Moving to California Worth It?” Forbes, Forbes Magazine, 15 Mar. 2024, www.forbes.com/home-improvement/moving-services/moving-to-california/.