Mathematical Modeling of Population Dynamics in California and Ontario: Math 3MB3 Final Project - Group 2

1. Introduction

Population growth and change are critical to planning a sustainable future. Population models are used by governments and organizations to forecast trends and allocate resources accordingly for infrastructure, healthcare systems, education, and other essential services. This is especially true in areas undergoing rapid growth or unusual demographic changes, such as California and Ontario. One case study that is useful in this regard is the data on California's post-war boom from 1955 to 1960, characterized by elevated birth rates, mortality rates, and migration. Just like Quebec, Ontario has also witnessed tremendous industrialization, urbanization, urban migration, and natural resource exploitation, making this region an excellent choice to study population dynamics over time.

Traditional population models are heavily dependent upon simplifying assumptions, like constant growth rates or fixed migration patterns. While these models give insight in certain situations, they are too simplistic to account for the subtleties of human population dynamics; for example, how the rate of internal reintegration might shift based on the state of the economy or specific policy decisions that could be made, not to mention resource constraints that could limit growth. As a result, such models typically generate predictions that do not reflect real-world data, illustrated by the results we obtained, which fell far from real-world population values. The rigidity of these models also limits the sensitivity needed to update and learn over time and makes them less effective as demographic shifts change. Improvement of predictive models can lead to such accurate forecasts that would help policymakers manage population growth and its impact.

The report starts with two base models that assume California and U.S. population trends will hold. The first model includes birth, death, and net migration rates, and the second includes only migration patterns. Then, a third, composite model is built, designed to mitigate the shortcomings of the simple models and deliver a more comprehensive view of population dynamics. The models are then used for application to Ontario and across Canada, to look at how relevant the approaches are in different demographic contexts.

This report uses historical population data from California and the United States from 1955 to 1960 to establish modeling of projected future populations. The research question is: How can the population models based on California and U.S. data from 1955 to 1960 be fitted to make them applicable to predict population growth in Ontario and Canada? To answer it, we make mathematical models that incorporate birth, death, and migration rates, test what they predict against real-world data, and refine them to account for regional differences. Overall, it showed us the strengths and weaknesses of different individual models, especially in the context of improving prediction accuracy and examining long-term trends. We demonstrate in this report that the understanding of population dynamics can be cemented using base models, but a pooling of approaches allows for more flexibility and realism. We illustrate population modeling to aid sustainable development and resource planning by refining these models and applying them to different regions, where the challenge of managing growth combines multiple aspects of such a complex, dynamic world.

2. Model

To predict the future populations based on the information of the population from 1955 to 1960, we need to create a math model which can summarize the patterns of previous population changes and make predictions about future populations.

2.1. Base model

The table below shows all the symbols we use in the base model.

Symbol	**Variable name or equation**	**Unit, Range**
$P(t)$	Populations of California and the US	Individuals (in thousands),

Symbol	Variable name or equation	Unit, Range
	excluding California at time t, respectively.	$P(t) \geq 0$.
A(t)	Populations of California at time t (in thousands).	Individuals (in thousands), $A(t) \geq 0$.
B(t)	Populations of the US excluding California at time t (in thousands).	Individuals (in thousands), $B(t) \geq 0$.
B	Birth matrix (diagonal).	Unit less, Any element in B ≥ 0.
D	Death matrix (diagonal).	Unit less, Any element in D ≥ 0.
M	Migration matrix (diagonal).	Unit less, Any element in $M \in [-1,1]$
I	Identity matrix (diagonal).	Unit less, Only have 1 or 0.
G	Growth matrix (diagonal) $G=I+B-D+M$.	Unit less. Any element in $G \in R$.
T	The migration matrix between the two regions.	Unit less. Any element in [0,1]
G_new	New growth matrix (diagonal) $G\_new=I+B-D$.	Unit less. Any element in $G\_new \in R$.
t	Time starts at 1955 and 1955 is at t=0	Per 5 years, $t \geq 0$.

Model 1

The first model uses births, deaths, and net migration in California and the rest of the U.S from 1955 to 1960 to make predictions about future populations.

This table shows the information of population between 1955 and 1960:

Region	Population 1955	Births 1955-60	Deaths 1955-60	Net Migration 1955-60	Population 1960
California	12,988,000	1,708,000	614,000	1,124,000	15,206,000
Rest of U.S.	152,082,000	19,499,000	7,417,000	-1,124,000	163,040,000
Total	165,070,000	21,207,000	8,031,000	0	178,246,000

Based on the information above, we calculate the birth, death and migration rate.

Birth rate = Births 1955-60/Population 1955

Death rate = Deaths 1955-60/Population 1955

Migration rate = Net Migration 1955-60/Population 1955

After the calculation we get this table below:

Region	Birth Rate	Death Rate	Migration Rate
California	0.1315060	0.0472744	0.0865414
Rest of U.S.	0.1282137	0.0487697	-0.0073908

Let I be the identity matrix and B,D and M be the birth, death, and net migration matrices(diagonal with California's rate in first row first column and rest of U.S's rate in second row second columns), respectively. Then let growth matrix G be I+B-D+M. That makes

$$G = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0.1315060 & 0 \\ 0 & 0.1282137 \end{bmatrix} - \begin{bmatrix} 0.1315060 & 0 \\ 0 & 0.1282137 \end{bmatrix} + \begin{bmatrix} 0.0865414 & 0 \\ 0 & -0.0073908 \end{bmatrix}$$
$$= \begin{bmatrix} 1.170773 & 0 \\ 0 & 1.072053 \end{bmatrix}$$

Then we assume the growth is the same for every 5 years and use the growth matrix G to predict the future population by the equation below.

$$\mathrm{P(t)} = \mathrm{G} \cdot P(t-1)$$

I is needed here to make G the population factor, so that the population in 1955 times the factor G is the population in 1960. Otherwise we wouldn't have proper net growth in our future populations.

For example, if B = D, then without the identity matrix we would have G = M. So for low value entries of -1 <= M <= 1, multiplying M with the previous 5 year's Population vector P(t-1), such that M*P(t-1), would result in a less new population P(t) than the previous population P(t-1), which wouldn't fit a growing population model.

$\mathrm{P(1)} = \mathrm{G} \cdot P(0)$ for t=1 is at 1960, P(1) is the population in 1960 and P(0) is the population in 1955.

Model 2

The second model uses only immigration and emigration between California and the rest of the U.S from 1955 to 1960 to make predictions about future populations.

This model only accounts for net migration between California and the rest of the US, and how initial populations transfer over from one to the other, and does not account for inherent growth or death of populations.

	To California	**To the U.S.**	**Total 1955 population**
California	12,174,000	814,000	12,988,000
Rest of U.S.	1,938,000	150,144,00	152,082,000

Let T be a migration matrix between the two regions.

$$T = \begin{bmatrix} \text{The percentage of California population} & \text{The percentage of Rest of U.S population} \\ \text{that remains in California} & \text{that moves to California} \\ \text{The percentage of California population} & \text{The percentage of Rest of U.S population} \\ \text{that moves to Rest of U.S} & \text{that remains in Rest of U.S} \end{bmatrix}$$

$$T = \begin{bmatrix} 0.937326763 & 0.012743125 \\ 0.062673236 & 0.987256874 \end{bmatrix}$$

So that

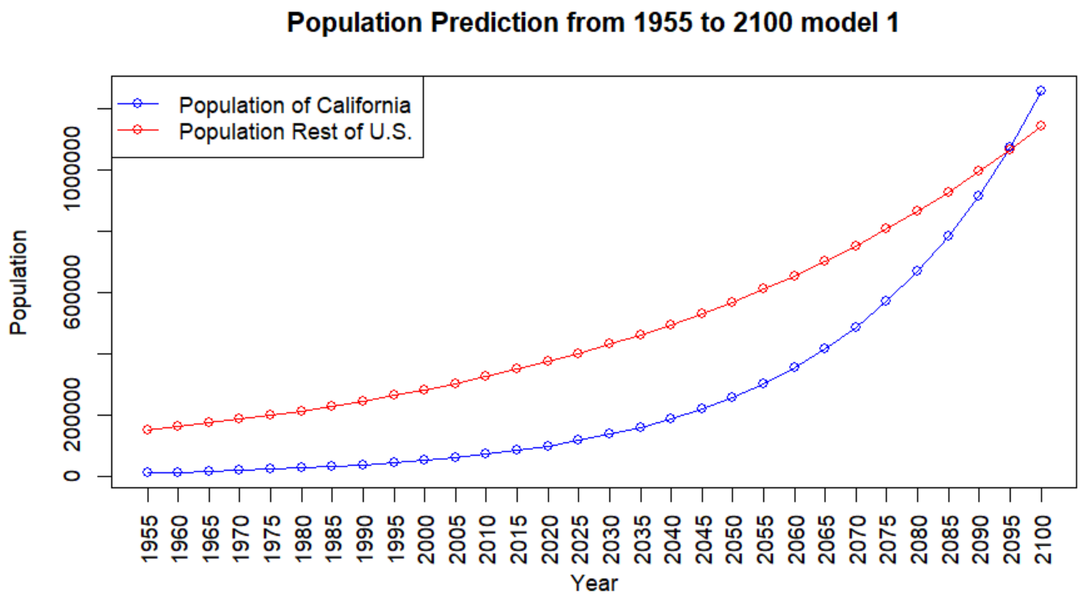
$$P(1) = T \cdot P(0)$$

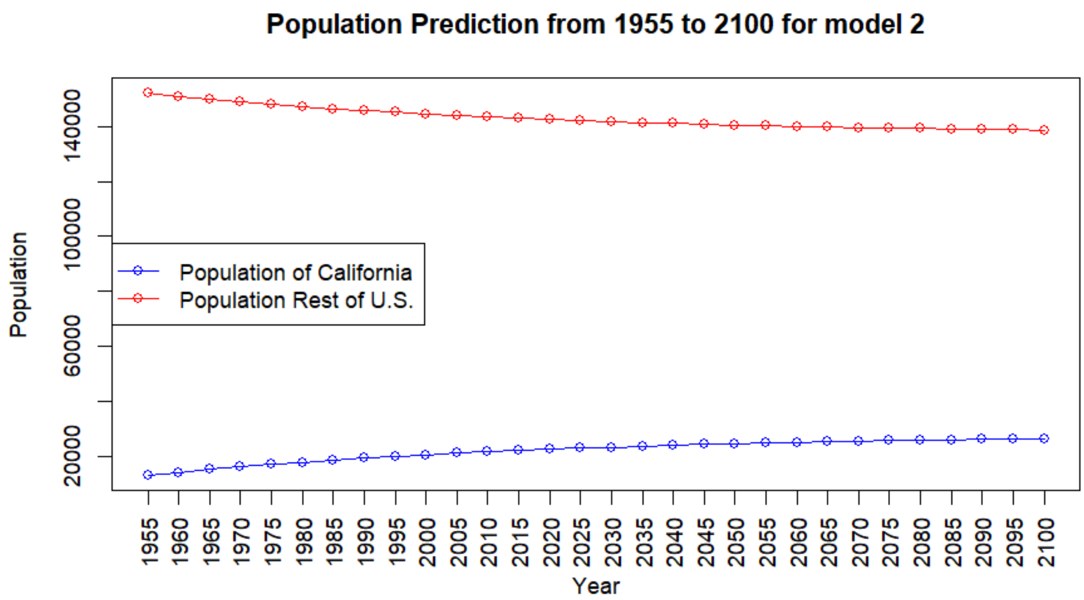
Then we assume the migration rates between the two regions are the same for every 5 years and use the migration matrix between the two regions T to predict the future population by the equation below:

$$P(t) = T \cdot P(t-1)$$

Summarize the results of model 1 and model 2 and compare them with the real data of 2020.

The graph below shows predicted values for the population from 1955 to 2100.





We used the models above to get the predicted value for the population in 2020 and we are given the actual population from [*P1: TOTAL POPULATION - Census Bureau Table*].

The table for population in 2020:

	Population for model 1	**Population for model 2**	**Actual Population**
California	100,852,700	22,514,170	39,538,223
Rest of U.S.	375,742,400	142,555, 830	291,911,058
Total	476,595,100	165,070,000	331,449,281

The table shows that the predicted value for 2020 population for both model 1 and model 2 are very different to the actual population. The reason why this happens is different for each model.

Model 1: The first problem in model 1 is that it only considers net migration rate and does not consider migration between California and the rest of the U.S. The second problem is that we assume that birth, death, and net migration rates will not change but this is not true in reality. In the real world, birth, death, and net migration rates will always change for various reasons. It may be natural or policy reasons. Furthermore, the results of this model are also unreasonable. This model predicts that the population will grow indefinitely because we have a positive growth matrix G and the population in California will be more than the population of the rest of the U.S by about 2095, which is not realistic given the limited natural resources. Also, due to limited natural resources, California and the rest of the U.S. should have a carrying capacity of population. In reality, when we have a carrying capacity for the population, the population growth rate should slow down as it approaches the carrying capacity. However, we ignore that in our model.

Model 2: The biggest problem with the second model is that it does not take into account death and birth rates. This will make the total population of the United States always remain constant. Another problem is similar to the problem in model 1, which is that it assumes that the migration between California and the rest of the U.S. will always be the same. This means that the population of the rest of the U.S. will continue to move to California until California has the entire population of the U.S. ,which does not make any sense.

New combined model

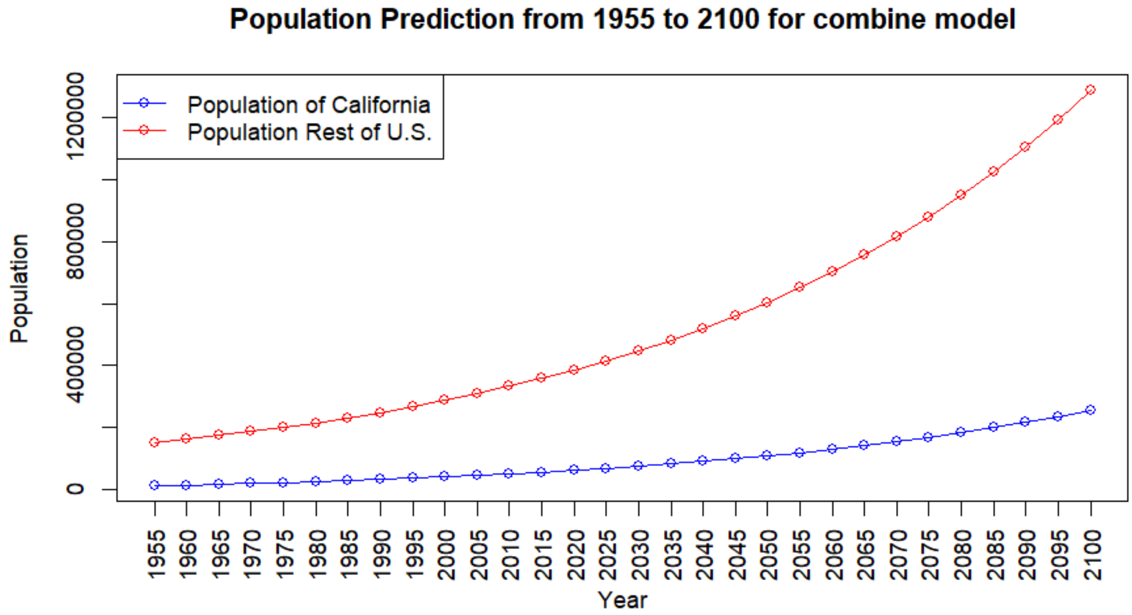
Since the first two models have their own shortcomings, we combined the first two models to create a third model called the new combined model, where we combine model 1 and model 2 separating the migration rates. We replace the migration matrix M with the migration matrix between the two regions T and let I+B-D be G_new. So the G_new will be the new growth rate that only considers the birth and death rate and equation shown below.

$$\mathrm{G}_{new} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0.1315060 & 0 \\ 0 & 0.1282137 \end{bmatrix} - \begin{bmatrix} 0.1315060 & 0 \\ 0 & 0.1282137 \end{bmatrix} = \begin{bmatrix} 1.084232 & 0.000000 \\ 0.000000 & 1.079444 \end{bmatrix}$$

Then we multiply the population at the previous time by the migration between the two regions to get the population after migration and then multiply it by the new growth rate to get the population at the next time.Theoretically this will give us more accurate predictions than using the migration matrix M directly in model 1. Therefore we use G_new times T to replace G in model 1 to get the new combined model which is shown below.

$$P(t) = G_{new} \cdot T \cdot P(t-1)$$

Compare the results of model3 to model1 and model2 and compare with the real data of 2020.



The table for population in 2020:

	Population for model 1	**Population for model 2**	**Population for combined model**	**Actual Population**
California	100,852,700	22,514,170	63,021,940	39,538,223
Rest of U.S.	375,742,400	142,555, 830	385,912,100	291,911,058
Total	476,595,100	165,070,000	448,934,040	331,449,281

The new combined model is closer to the actual data compared to model 1. As shown in the plot, the population of the rest of the U.S. is growing faster than that of California, which is more realistic than having California's population exceed that of the rest of the U.S. However, the model still has the issue of predicting indefinite population growth while assuming that the rates remain constant over time and ignoring carrying capacity, which is not realistic.

2.2. Analysis of Base Model

2.2.1. Model 1

P(t)=G*P(t-1), such that **G = I + B - D + M,** where G is our net growth rate matrix and t >= 1 where P(0) is the initial given population in 1955 and P(1) is predicted population in 1960, etc.

G matrix	**To California**	**To Rest of the US**
California	1.170773	0
Rest of the US	0	1.072053

Since the G matrix is a diagonal matrix, $\lambda_1 = 1.170773$ and $\lambda_2 = 1.072053$. Both eigenvalues are > 1, or since dominant eigenvalue $\lambda_d = \lambda_1 = 1.17077 > 1$, this means both populations will grow without bound. So,

$$\vec{P}_t = G \cdot \vec{P}_{t-1} = G \cdot G \cdot \vec{P}_{t-2} = \cdots = G^t \cdot \vec{P}_{t-t} = G^t \cdot \vec{P}_0$$

such that t >= 1, with initial population $\vec{P}_0$.

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \ B = \begin{bmatrix} 0.1315060 & 0 \\ 0 & 0.1282137 \end{bmatrix}, \ D = \begin{bmatrix} 0.0472744 & 0 \\ 0 & 0.0487697 \end{bmatrix} \ and$$

$$M = \begin{bmatrix} 0.0865414 & 0 \\ 0 & -0.073908 \end{bmatrix}$$

$$\vec{P}_t = G^t \cdot \vec{P}_0 = (I + B - D + M)^t \cdot \vec{P}_0$$

$$\vec{P}_t = \begin{bmatrix} 1 + 0.1315060 - 0.0472744 + 0.0865414 & 0 \\ 0 & 1 + 0.1282137 - 0.0487697 - 0.0073908 \end{bmatrix}^t \cdot \vec{P}_0$$

$$= \begin{bmatrix} 1.170773 & 0 \\ 0 & 1.0720532 \end{bmatrix}^t \cdot \vec{P}_0$$

$$Solve\ det(G - \lambda \cdot I) = 0 = \begin{vmatrix} 1.170773 - \lambda & 0 \\ 0 & 1.0720532 - \lambda \end{vmatrix} = 0$$

Since G is a diagonal Matrix,

$$\boxed{\lambda_1 = 1.170773 \quad and \quad \lambda_2 = 1.0720532}$$

Since the dominant eigenvalue $\lambda_d = \lambda_1 > \lambda_2\ and\ \lambda_1 > 1$, all components of the solution will grow without bound.

In other words, all populations will grow to infinity without bounds as $t \to \infty$.

$$\vec{P}_t = G^t \cdot \vec{P}_0 = \begin{bmatrix} 1.170773^t & 0 \\ 0 & 1.0720532^t \end{bmatrix} \cdot \vec{P}_0 = \begin{bmatrix} \lambda_1{}^t \cdot A_0 \\ \lambda_2{}^t \cdot B_0 \end{bmatrix} = \begin{bmatrix} A_t \\ B_t \end{bmatrix},$$

$where\ \vec{P}_0 = \begin{bmatrix} A_0 \\ B_0 \end{bmatrix}$, A_0 $is\ the\ initial\ California\ population\ (in\ thousands)\ and$

B_0 $is\ the\ initial\ rest\ of\ the\ US\ population\ (in\ thousands),\ such\ that$

$A_0 = 12,988\ and\ B_0 = 152,082,\ and\ A_t\ is\ California's\ population$

$at\ time\ t\ (in\ thousands)\ and\ B_t\ is\ rest\ of\ the\ US\ population\ at\ time\ t\ (in\ thousands)$.

COMPARISON TO REAL 2020 DATA:

Based on https://data.census.gov/table?q=california &
https://data.census.gov/all?q=2020%20usa%20population, California's **true** population in 2020 was **39,538,223**, whereas the total US population was 331,449,281. So, the rest of the US's **true** population in 2020 was 331,449,281 - 39,538,223 = **291,911,058**.

In **2020**, which is t = (2020-1955)/5 = 13, **California's population (in thousands):**
(1.170773^13) * 12,988 = 100,852.6601791608

And the **rest of US population (in thousands)** in **2020**:
(1.0720532^13) * 152,082 = 375,742.3979448416

Our 1st model predicts a population of approximately 100,852,700 for California and approximately 375,742,400 for the rest of the US in 2020, whereas the true population was 39,538,223 and 291,911,058 for California and the rest of US, respectively.

OVERESTIMATION: Predicted Population / True Population
Predicted California Population / True California Population
= 100,852,700 / 39,538,223 = 2.55076461074

Predicted the rest of US Population / True population of the rest of US = 375,742,400 / 291,911,058 = 1.28718111117

So, our model is overestimating the true population by approximately 155% for California and 28.7% for the rest of the US (mostly due to reasons expressed in the base model section, such as not accounting for carrying capacity, assuming constant long-term growth rate, etc.)

LONG-TERM GROWTH RATE OF CALIFORNIA & REST OF US

$$\lim_{t\to\infty} \frac{A_t}{A_{t-1}} = \lim_{t\to\infty} \frac{\lambda_1^{\,t}\cdot A_0}{\lambda_1^{\,t-1}\cdot A_0} = \lim_{t\to\infty} \lambda_1^{\,t-(t-1)} = \lim_{t\to\infty} \lambda_1 = 1.170773$$

$$\lim_{t\to\infty} \frac{B_t}{B_{t-1}} = \lim_{t\to\infty} \frac{\lambda_2^{\,t}\cdot B_0}{\lambda_2^{\,t-1}\cdot B_0} = \lim_{t\to\infty} \lambda_2^{\,t-(t-1)} = \lim_{t\to\infty} \lambda_2 = 1.0720532$$

Thus, in the long-run the growth rate of the population in California is approximately 17.08% every 5 years.

And, the long-run growth rate of the population in the rest of the US is approximately 7.2% every 5 years.

Since the long-term growth rate of California's population is larger than the long-term growth rate of the rest of the US population, at some point the population size will intersect with the population size of the rest of US and eventually California will end up with a much larger population size than the rest of US.

LONG-TERM PROPORTIONATE POPULATION OF CALIFORNIA WITHIN US:

$$\lim_{t \to \infty} \frac{A_t}{A_t + B_t} = \lim_{t \to \infty} \frac{\lambda_1^{\,t} \cdot A_0}{\lambda_1^{\,t} \cdot A_0 + \lambda_2^{\,t} \cdot B_0}$$

And since $\lambda_1 > \lambda_2$, $\lambda_1^{\,t} \cdot A_0$ *grows faster than* $\lambda_2^{\,t} \cdot B_0$

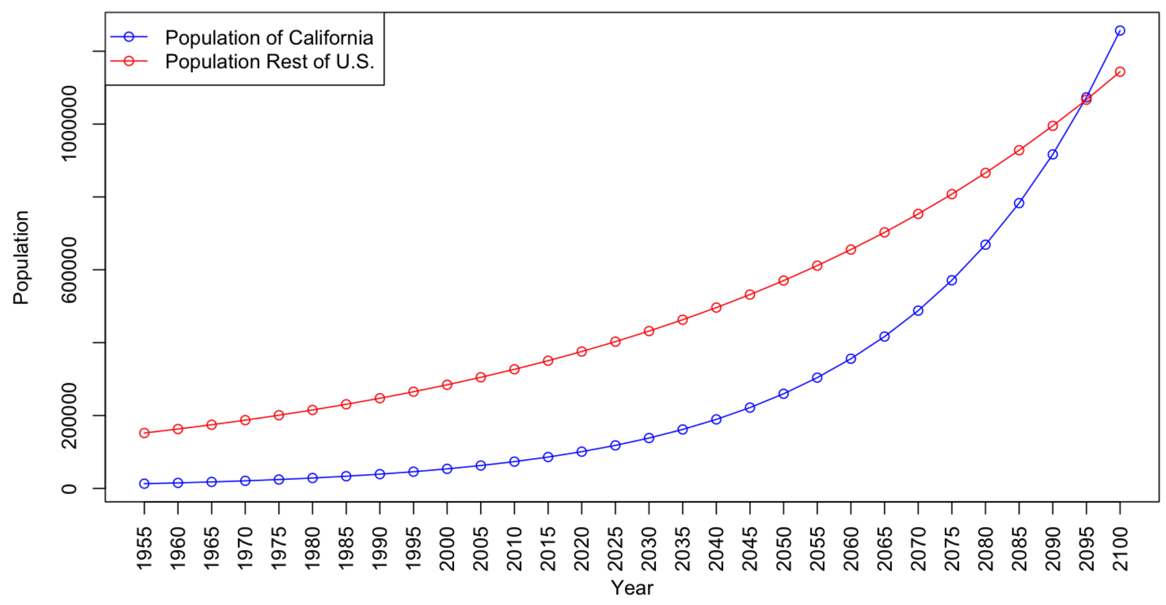
$$\lim_{t \to \infty} \frac{A_t}{A_t + B_t} \approx \lim_{t \to \infty} \frac{\lambda_1^{\,t} \cdot A_0}{\lambda_1^{\,t} \cdot A_0} = 1$$

Thus, in the long-term the proportionate population of California within the US approaches 100%.

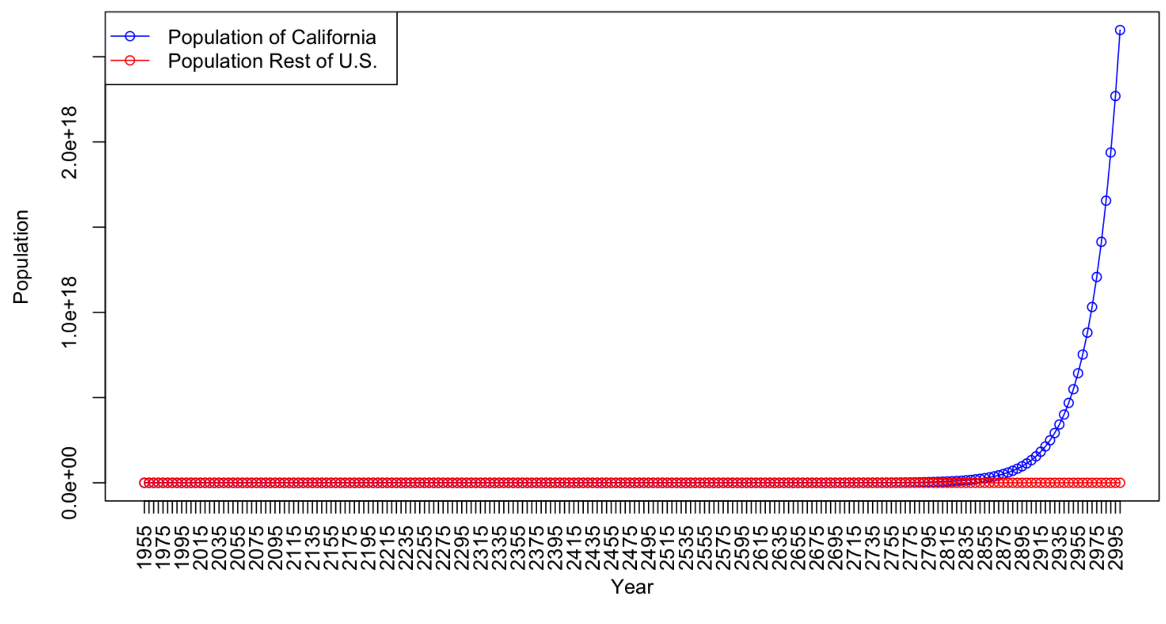
So this means that since the initial population of California is smaller than initial population of the rest of US and the fact that the long-term growth rate of California is larger than the long-term growth rate of the rest of US, the model predicts that at some point the population of California will overreach the population of the rest of US and eventually, in the long-run, approximately the entire US population will be concentrated within California. For obvious reasons such as limited resources in the real-world, this is very unrealistic.

Below we have a couple of graphs of our first model's population prediction for the years 1955 to 2100 compared to a longer term version from 1955 to 3000.

Population Prediction from year 1955 to 2100 for model 1



Population Prediction from year 1955 to 3,000 for model 1



2.2.2. Model 2

P(t)=T*P(t-1), such that T is our transition matrix of people emigrating and immigrating between California and the rest of the US.

$$T = \begin{bmatrix} \frac{12174}{12988} & \frac{1938}{152082} \\ \frac{814}{12988} & \frac{150144}{152082} \end{bmatrix} = \begin{bmatrix} 0.9373267632 & 0.01274312542 \\ 0.06267323683 & 0.9872568746 \end{bmatrix}$$

$$det(T - \lambda \cdot I) = 0 = \begin{vmatrix} \frac{12174}{12988} - \lambda & \frac{1938}{152082} \\ \frac{814}{12988} & \frac{150144}{152082} - \lambda \end{vmatrix}$$

$$\Rightarrow \lambda^2 - \lambda \cdot trace(T) + det(T) = 0$$

$$trace(T) = \frac{12174}{12988} + \frac{150144}{152082} = 1.9245836377$$

$$det(T) = \left(\frac{12174}{12988}\right)\left(\frac{150144}{152082}\right) - \left(\frac{814}{12988}\right)\left(\frac{1938}{152082}\right) = 0.9245836377$$

$$\Delta(T) = tr(T)^2 - 4 \cdot det(T) = 0.005687627703$$

$$\sqrt{\Delta(T)} = 0.0754163623$$

$$\lambda_{1,2} = \frac{tr(T) \pm \sqrt{\Delta(T)}}{2}$$

$$\lambda_1 = \frac{1.9245836377 + 0.0754163623}{2} = \frac{2}{2} = 1$$

and

$$\lambda_2 = \frac{1.9245836377 - 0.0754163623}{2}$$

$$= 0.9245836377$$

Dominant eigenvalue $\boxed{\lambda_d = \lambda_1 = 1}$

Therefore, solutions will approach to a non-zero value given by its associated eigenvector, $\vec{v}_1$.

$$Solve\ \left(T - \lambda_1 \cdot I\right)\vec{v}_1 = \vec{0} \quad and\ \left(T - \lambda_2 \cdot I\right)\vec{v}_2 = \vec{0}$$

$$\left(T - \lambda_1 \cdot I\right) \cdot \vec{v}_1 = \vec{0}$$

$$\begin{bmatrix} \frac{12174}{12988} - 1 & \frac{1938}{152082} \\ \frac{814}{12988} & \frac{150144}{152082} - 1 \end{bmatrix} \cdot \vec{v}_1 = \vec{0}$$

$$\begin{bmatrix} \frac{12174 - 12988}{12988} & \frac{1938}{152082} \\ \frac{814}{12988} & \frac{150144 - 152082}{152082} \end{bmatrix} \cdot \vec{v}_1 = \vec{0}$$

$$\begin{bmatrix} -\frac{814}{12988} & \frac{1938}{152082} & \vdots & 0 \\ \frac{814}{12988} & -\frac{1938}{152082} & \vdots & 0 \end{bmatrix}$$

$$Row\ 2 + Row\ 1 \rightarrow Row\ 2$$

$$\frac{12988}{-814} \cdot Row\ 1 \rightarrow Row\ 1$$

$$\begin{array}{cc} x_1 & x_2 \\ \multicolumn{2}{c}{\left[\begin{array}{cc|c} 1 & \frac{1938}{152082} \cdot \frac{12988}{-814} & 0 \\ 0 & 0 & 0 \end{array}\right]} \end{array}$$

$$1 \cdot x_1 - \frac{1938(12988)}{152082(814)} \cdot x_2 = 0$$

$$x_1 = \frac{1938(12988)}{152082(814)} \cdot x_2$$

$$Let\ x_1 = 1 \Rightarrow x_2 = \frac{152082(814)}{1938(12988)}$$

$$\vec{v}_1 = \begin{bmatrix} 1 \\ \frac{152082(814)}{1938(12988)} \end{bmatrix}$$

And,

$$\left(T - \lambda_2 \cdot I\right) \cdot \vec{v}_2 = \vec{0}$$

$$\left[\begin{array}{cc|c} \frac{12174 - \lambda_2 \cdot 12988}{12988} & \frac{1938}{152082} & 0 \\ \frac{814}{12988} & \frac{150144 - \lambda_2 \cdot 152082}{152082} & 0 \end{array}\right]$$

$$\Rightarrow \left[\begin{array}{cc|c} \frac{1938}{152082} & \frac{1938}{152082} & 0 \\ \frac{814}{12988} & \frac{814}{12988} & 0 \end{array}\right]$$

$$Row\ 1 \cdot \frac{152082}{1938} \longrightarrow Row\ 1 \quad \& \quad Row\ 2 \cdot \frac{12988}{814} \longrightarrow Row\ 2$$

$$= \begin{bmatrix} 1 & 1 & \vdots & 0 \\ 1 & 1 & \vdots & 0 \end{bmatrix}$$

$Row\ 2 - Row\ 1 \longrightarrow Row\ 2$

$$= \begin{bmatrix} 1 & 1 & \vdots & 0 \\ 0 & 0 & \vdots & 0 \end{bmatrix} \qquad \Rightarrow x_1 + x_2 = 0$$

$if\ x_1 = 1,\ then\ x_2 = -1$

$Thus,\ \vec{v}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

$Model\ solution\ using\ \vec{P}_t = \begin{bmatrix} A_t \\ B_t \end{bmatrix} = c_1 \cdot \lambda_1{}^t \cdot \vec{v}_1 + c_2 \cdot \lambda_2{}^t \cdot \vec{v}_2$

Where A_t *is California's Population at time* t*, and* B_t *is the rest of US population at time* t.

and since $|\lambda_d| = |\lambda_1| = 1 > |\lambda_2|$, *as* $t \to \infty$:

$\vec{P}_t \approx c_1 \cdot \lambda_1{}^t \cdot \vec{v}_1$ since λ_1 *dominates the solution long* − *term*.

$$\vec{P}_t = T \cdot \vec{P}_{t-1} = T \cdot T \cdot \vec{P}_{t-2} = \cdots = T^t \cdot \vec{P}_0$$

$$= P \cdot D^t \cdot P^{-1} \cdot \vec{P}_0$$

where $\vec{P}_0 = \begin{bmatrix} A_0 \\ B_0 \end{bmatrix}$ *is the initial total California and rest of US population* (*in thousands*)

at time $t = 0$, *such that* $A_0 = 12,988$ *and* $B_0 = 152,082$.

$$D^t = \begin{bmatrix} \lambda_1{}^t & 0 \\ 0 & \lambda_2{}^t \end{bmatrix} \qquad P = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ \dfrac{152082(814)}{1938(12988)} & -1 \end{bmatrix}$$

$$det(P) = 1(-1) - (1)\left(\frac{152082 \cdot 814}{1938 \cdot 12988}\right) = \frac{-1938(12988) - 152082(814)}{1938(12988)}$$

$$P^{-1} = \begin{bmatrix} -1 & -1 \\ -\dfrac{152082(814)}{1938(12988)} & 1 \end{bmatrix} \cdot \frac{1}{det(P)}$$

$$P^{-1} \cdot \vec{P}_0 = \frac{1}{det(P)} \cdot \begin{bmatrix} -1 & -1 \\ -\dfrac{152082(814)}{1938(12988)} & 1 \end{bmatrix} \cdot \begin{bmatrix} 12988 \\ 152082 \end{bmatrix}$$

$$P^{-1} \cdot \vec{P}_0 = \frac{1}{det(P)} \begin{bmatrix} -12988 - 152082 \\ -\dfrac{152082(814)}{1938} + 152082 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$$

$$\Rightarrow c_1 = \frac{(-12988 - 152082)}{det(P)} = (-12988 - 152082) \cdot \frac{1938(12988)}{-1938(12988) - 152082(814)}$$

$$\Rightarrow c_1 \approx 27,891.9275618544$$

$$\vec{P}_t \approx c_1 \cdot \lambda_1{}^t \cdot \vec{v}_1 = 27,891.9275618544 \cdot (1)^t \cdot \begin{bmatrix} 1 \\ \dfrac{152082(814)}{1938(12988)} \end{bmatrix}$$

$$\vec{P}_t \approx 27,891.9275618544 \cdot \begin{bmatrix} 1 \\ \dfrac{152082(814)}{1938(12988)} \end{bmatrix} = \begin{bmatrix} 27,891.9275618544 \\ 137,178.07243814564 \end{bmatrix}$$

$$as\ t \to \infty\ ,\ \ \vec{P}_\infty = \begin{bmatrix} 27,891.9275618544 \\ 137,178.07243814564 \end{bmatrix} \approx \begin{bmatrix} 27,892 \\ 137,178 \end{bmatrix}$$

So, the long-term behaviour of this model as $t \to \infty$ is that the total populations of California and the rest of the US accounting for migration within the US will approach to approximately 27,892 and 137,178 , respectively.

Note:

Initial total Population = 12,988 + 152,082 = 165,070

And total population in the long-run accounting for only migration = 27,892 + 137,178 = 165,070

Proportion of Californians in US in the long-run (accounting only for migration):

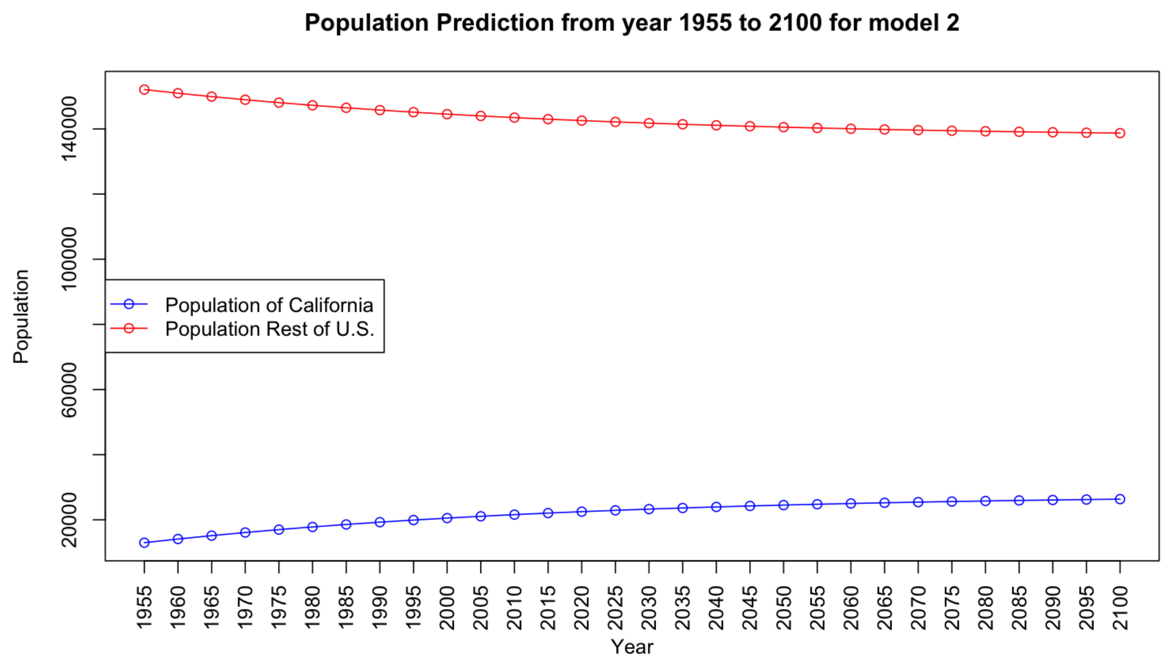
$$\frac{(27,892)}{(27,892 + 137,178)} \approx 0.1689707397$$

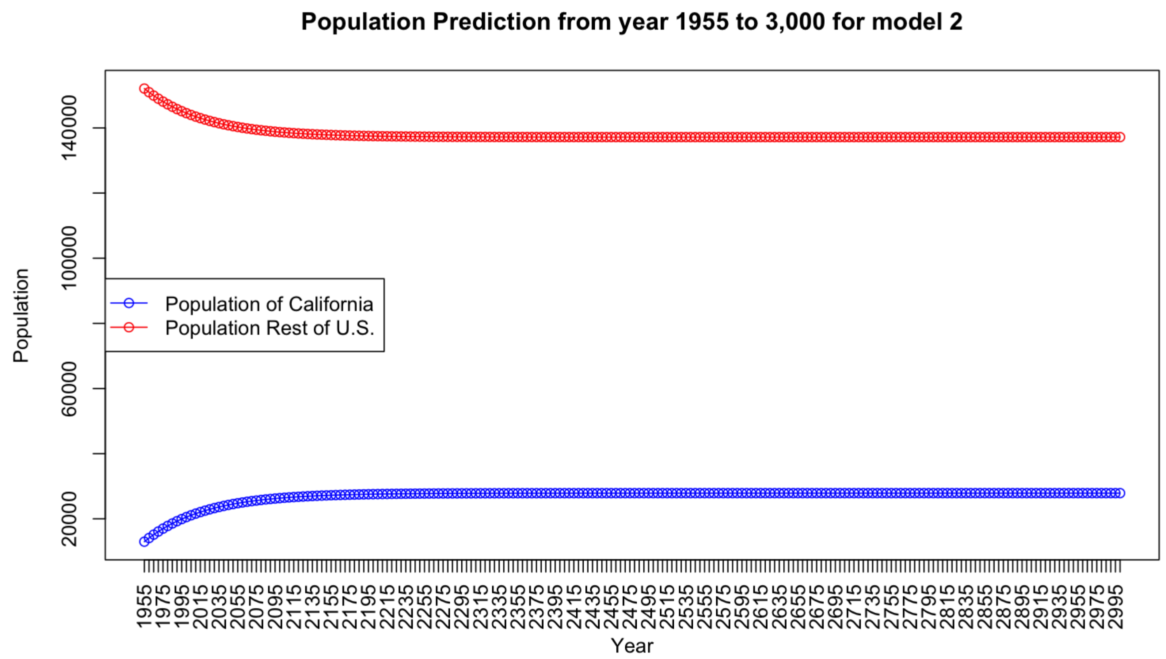
Proportion of the rest of US population in the long-run (accounting only for migration):

$$\frac{(137,178)}{(137,178 + 27,892)} \approx 0.8310292603$$

In other words, compared to the initial population distribution where there were initially 12,988/165,070 $\approx$ 7.87% of the US population within California and 152,082/165,070 $\approx$ 92.13% of the US population within the rest of the USA, the model predicts that long-term population distributions would be approximately 16.897% of US population within California and 83.103% of US population within the rest of US, when only accounting for net migration between the two.

Below we have a couple of graphs of our 2nd model's population prediction accounting only for migration for the years 1955 to 2100 compared to a longer term version from 1955 to 3000.





2.2.3. Model 3 (Combined model)

$$\vec{P}_t = G_{new} \cdot T \cdot \vec{P}_{t-1} = \left(G_{new} \cdot T\right)^t \cdot \vec{P}_0,$$

where t >= 1, such that P(0) is the initial given population in 1955, and P(1) is the predicted population in 1960, etc.

$$G_{new} = I + B - D = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0.1315060 & 0 \\ 0 & 0.1282137 \end{bmatrix} - \begin{bmatrix} 0.0472744 & 0 \\ 0 & 0.0487697 \end{bmatrix}$$

$$G_{new} = \begin{bmatrix} 1.0842316 & 0 \\ 0 & 1.079444 \end{bmatrix}$$

$$T = \begin{bmatrix} \frac{12174}{12988} & \frac{1938}{152082} \\ \frac{814}{12988} & \frac{150144}{152082} \end{bmatrix} = \begin{bmatrix} 0.9373267632 & 0.01274312542 \\ 0.06267323683 & 0.9872568746 \end{bmatrix}$$

$$G_{new} \cdot T = \begin{bmatrix} 1.0842316 & 0 \\ 0 & 1.079444 \end{bmatrix} \cdot \begin{bmatrix} \frac{12174}{12988} & \frac{1938}{152082} \\ \frac{814}{12988} & \frac{150144}{152082} \end{bmatrix}$$

$$= \begin{bmatrix} 1.01627929615 & 0.01381649926 \\ 0.06765224946 & 1.06568850973 \end{bmatrix}$$

$$Solve\ det\left(G_{new} \cdot T - \lambda \cdot I \right) = 0$$

$$= \begin{vmatrix} 1.01627929615 - \lambda & 0.01381649926 \\ 0.06765224946 & 1.06568850973 - \lambda \end{vmatrix} = 0$$

$$\Rightarrow \lambda^2 - \lambda \cdot trace\left(G_{new} \cdot T \right) + det\left(G_{new} \cdot T \right) = 0$$

$$trace\left(G_{new} \cdot T \right) = 1.01627929615 + 1.06568850973 = 2.08196780588$$

$$det\left(G_{new} \cdot T \right) = (1.01627929615)(1.06568850973) - (0.06765224946)(0.01381649926)$$

$$= 1.08210245133$$

$$\Delta\left(G_{new}\cdot T\right) = tr\left(G_{new}\cdot T\right)^2 - 4\cdot det\left(G_{new}\cdot T\right) = 0.006180139401$$

$$\sqrt{\Delta\left(G_{new}\cdot T\right)} = 0.07861386265$$

$$\lambda_{1,2} = \frac{tr\left(G_{new}\cdot T\right) \pm \sqrt{\Delta\left(G_{new}\cdot T\right)}}{2}$$

$\lambda_1 = 1.08029083427$ *and* $\lambda_2 = 1.001676971615$,

where $\lambda_d = \lambda_1 > \lambda_2$ *and* $\lambda_1 > 1$.

So, the solutions of the populations will grow without bound over long-term, as $t \to \infty$ **.**

Solve $\left(G_{new}\cdot T - \lambda_1\cdot I\right)\vec{v}_1 = \vec{0}$ *and* $\left(G_{new}\cdot T - \lambda_2\cdot I\right)\vec{v}_2 = \vec{0}$

$$\left(G_{new}\cdot T - \lambda_1\cdot I\right)\cdot\vec{v}_1 = \vec{0}$$

$$\begin{bmatrix} 1.01627929615 - 1.08029083427 & 0.01381649926 \\ 0.06765224946 & 1.06568850973 - 1.08029083427 \end{bmatrix}\cdot\vec{v}_1 = \vec{0}$$

$$\Rightarrow\begin{bmatrix} -0.06401153812 & 0.01381649926 & \vdots & 0 \\ 0.06765224946 & -0.01460232454 & \vdots & 0 \end{bmatrix}$$

$$\frac{Row\ 1}{-0.06401153812} \to Row\ 1 \quad and \quad \frac{Row\ 2}{-0.01460232454} \to Row\ 2$$

$$\begin{bmatrix} 1 & -0.2158438879 & \vdots & 0 \\ -4.6329780765 & 1 & \vdots & 0 \end{bmatrix}$$

$$4.6329780765\cdot Row\ 1 + Row\ 2 \to Row\ 2$$

$$\Rightarrow\begin{bmatrix} 1 & -0.2158438879 & \vdots & 0 \\ 0 & 0 & \vdots & 0 \end{bmatrix} \Rightarrow x_1 - 0.2158438879\cdot x_2 = 0$$

$$if\ \ x_2 = 1, \quad then\ x_1 = 0.2158438879$$

Thus,

$$\vec{v}_1 = \begin{bmatrix} 1 \\ 4.6329780738 \end{bmatrix} = \begin{bmatrix} 0.2158438879 \\ 1 \end{bmatrix}.$$

And,

$$\left(G_{new} \cdot T - \lambda_2 \cdot I \right) \cdot \vec{v}_2 = \vec{0}$$

$$\begin{bmatrix} 1.01627929615 - 1.001676971615 & 0.01381649926 \\ 0.06765224946 & 1.06568850973 - 1.001676971615 \end{bmatrix} \cdot \vec{v}_2 = \vec{0}$$

$$\Rightarrow \begin{bmatrix} 0.01460232454 & 0.01381649926 & \vdots & 0 \\ 0.06765224946 & 0.06401153812 & \vdots & 0 \end{bmatrix}$$

$$\frac{Row\ 1}{0.01460232454} \rightarrow Row\ 1\ \ and\ \ \frac{Row\ 2}{0.06401153812} \rightarrow Row\ 2$$

$$\Rightarrow \begin{bmatrix} 1 & 0.9461849189 & \vdots & 0 \\ 1.05687586093 & 1 & \vdots & 0 \end{bmatrix}$$

$$-1.05687586093 * Row\ 1 + Row\ 2 \rightarrow Row\ 2$$

$$\Rightarrow \begin{bmatrix} 1 & 0.9461849189 & \vdots & 0 \\ 0 & 0 & \vdots & 0 \end{bmatrix} \Rightarrow x_1 + 0.9461849189 \cdot x_2 = 0$$

$$if\ \ x_2 = 1,\ then\ x_1 = -0.9461849189$$

Thus,

$$\vec{v}_2 = \begin{bmatrix} 1 \\ -1.05687586013 \end{bmatrix} = \begin{bmatrix} -0.9461849189 \\ 1 \end{bmatrix}$$

$$\vec{P}_t = \left(G_{new} \cdot T\right) \cdot \vec{P}_{t-1} = \left(G_{new} \cdot T\right) \cdot \left(G_{new} \cdot T\right) \cdot \vec{P}_{t-2} = \cdots = \left(G_{new} \cdot T\right)^t \cdot \vec{P}_0$$

$$= P \cdot D^t \cdot P^{-1} \cdot \vec{P}_0$$

$$D^t = \begin{bmatrix} \lambda_1{}^t & 0 \\ 0 & \lambda_2{}^t \end{bmatrix}, \quad P = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 \end{bmatrix}$$

$$P = \begin{bmatrix} 0.2158438879 & -0.9461849189 \\ 1 & 1 \end{bmatrix}, \quad P^{-1} = \frac{1}{det(P)} \begin{bmatrix} 1 & 0.9461849189 \\ -1 & 0.2158438879 \end{bmatrix}$$

$$det(P) = (0.2158438879) - (-0.9461849189) = 1.1620288068$$

$$\Rightarrow P^{-1} = \begin{bmatrix} 0.8605638639 & 0.8142525498 \\ -0.8605638639 & 0.1857474502 \end{bmatrix}$$

$$P^{-1} \cdot \vec{P}_0 = \begin{bmatrix} 0.8605638639 & 0.8142525498 \\ -0.8605638639 & 0.1857474502 \end{bmatrix} \cdot \begin{bmatrix} 12988 \\ 152082 \end{bmatrix}$$

$$\Rightarrow = \begin{bmatrix} 135{,}010.1597430168 \\ 17{,}071.8402569832 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$$

$$Model\ solution\ using\ \vec{P}_t = \begin{bmatrix} A_t \\ B_t \end{bmatrix} = c_1 \cdot \lambda_1{}^t \cdot \vec{v}_1 + c_2 \cdot \lambda_2{}^t \cdot \vec{v}_2$$

Where A_t *is California's Population at time t (in thousands),*

and B_t *is the rest of US population at time t (in thousands).*

Therefore,

$$\vec{P}_t = 135,010.1597430168 \cdot (1.08029083427)^t \cdot \begin{bmatrix} 0.2158438879 \\ 1 \end{bmatrix}$$

$$+ 17,071.8402569832 \cdot (1.001676971615)^t \cdot \begin{bmatrix} -0.9461849189 \\ 1 \end{bmatrix}$$

Also, as $t \to \infty$, since

$$\vec{P}_t = c_1 \cdot \lambda_1^t \cdot \vec{v}_1 + c_2 \cdot \lambda_2^t \cdot \vec{v}_2$$

$$and \ |\lambda_d| = |\lambda_1| = 1.08029083427 > |\lambda_2|$$

Then,

$$\vec{P}_t \approx c_1 \cdot \lambda_1^t \cdot \vec{v}_1 \quad \text{since } \lambda_1 \ dominate\ the\ solutions\ long - term.$$

$$\approx (135,010.1597430168) \cdot (1.08029083427)^t \cdot \begin{bmatrix} 0.2158438879 \\ 1 \end{bmatrix}$$

Although the population still grows with no bound in this model, California's population grows at a lower rate than in the first model.

COMPARISON TO REAL 2020 DATA:

In **2020**, which is t = (2020-1955)/5 = 13, **California's population (in thousands):**

135010.1597430168*(1.08029083427^13)*(.2158438879)

+17,071.8402569832*(1.001676971615^13)*(-0.9461849189) = 63,021.9400880644

And the **rest of US population (in thousands)** in **2020**:

135010.1597430168*(1.08029083427^13)*(1)

+17,071.8402569832*(1.001676971615^13)*(1) = 385,912.1040762598

Recall:

California's **true** population in 2020 was **39,538,223**, and the rest of the US's **true** population in 2020 was **291,911,058**.

Our 3rd model predicts that California's and the rest of the USA's population will be approximately 63,021,940 and 385,912,104 , respectively.

OVERESTIMATION: Predicted Population / True Population

Predicted California Population / True California Population

= 63,021,940 / 39,538,223 = 1.59394973315

Predicted the rest of US Population / True population of the rest of US

= 385,912,104 / 291,911,058 = 1.32201947622

So, our new combined model still overestimates the actual population, although this time by approximately 59.4% for California and 32.2% for the rest of the US.

Although it overestimates the true population (most likely for similar reasons expressed in Model 1 in the Base model section, such as lack of accounting for carrying capacity, assuming constant long-term growth rate, etc.) it estimates California's population much better this time than the 1st Model.

LONG-TERM GROWTH RATE:

$$\lim_{t \to \infty} \frac{\left(\vec{P}_t\right)}{\vec{P}_{t-1}} = \lim_{t \to \infty} \frac{c_1 \cdot \lambda_1^{\,t} \cdot \vec{v}_1}{c_1 \cdot \lambda_1^{\,t-1} \cdot \vec{v}_1}$$

$$= \lim_{t \to \infty} \lambda_1^{\,t-(t-1)} = \lim_{t \to \infty} \lambda_1 = \lambda_1 = 1.08029083427$$

Thus, in the long-run, California's population (and the rest of the US population) grows approximately at a rate of 8.03% every 5 years.

LONG-TERM POPULATION DISTRIBUTION:

$$\vec{P}_t = \begin{bmatrix} A_t \\ B_t \end{bmatrix} \approx (135{,}010.1597430168) \cdot (1.08029083427)^{t} \cdot \begin{bmatrix} 0.2158438879 \\ 1 \end{bmatrix}$$

$$Total\ US\ Population\ at\ time\ t = A_t + B_t$$

$$A_t \approx 135{,}010.1597430168 \cdot (1.08029083427)^t \cdot (0.2158438879)$$

$$B_t \approx 135{,}010.1597430168 \cdot (1.08029083427)^t \cdot (1)$$

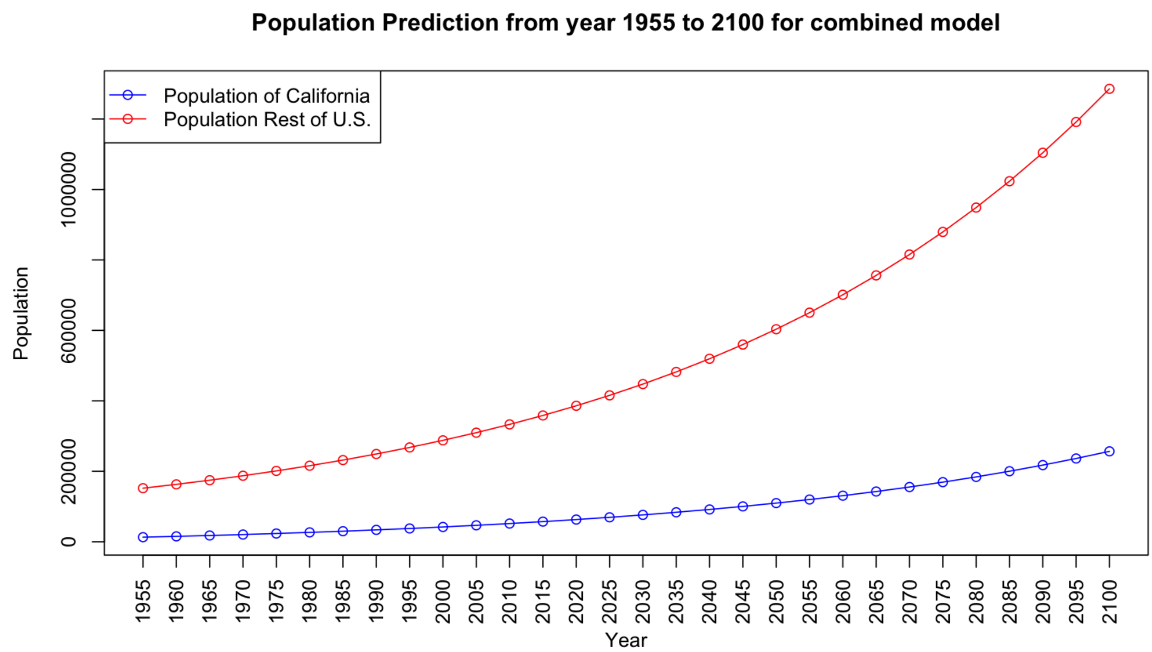
$$\lim_{t \to \infty} \frac{A_t}{A_t + B_t} = \frac{135{,}010.1597430168 \cdot (1.08029083427)^t \cdot (0.2158438879)}{135{,}010.1597430168 \cdot (1.08029083427)^t \cdot (0.2158438879 + 1)}$$

$$\Rightarrow \lim_{t \to \infty} \frac{A_t}{A_t + B_t} = \frac{(0.2158438879)}{(0.2158438879 + 1)} \approx 0.1775259884 \approx 17.75\%$$

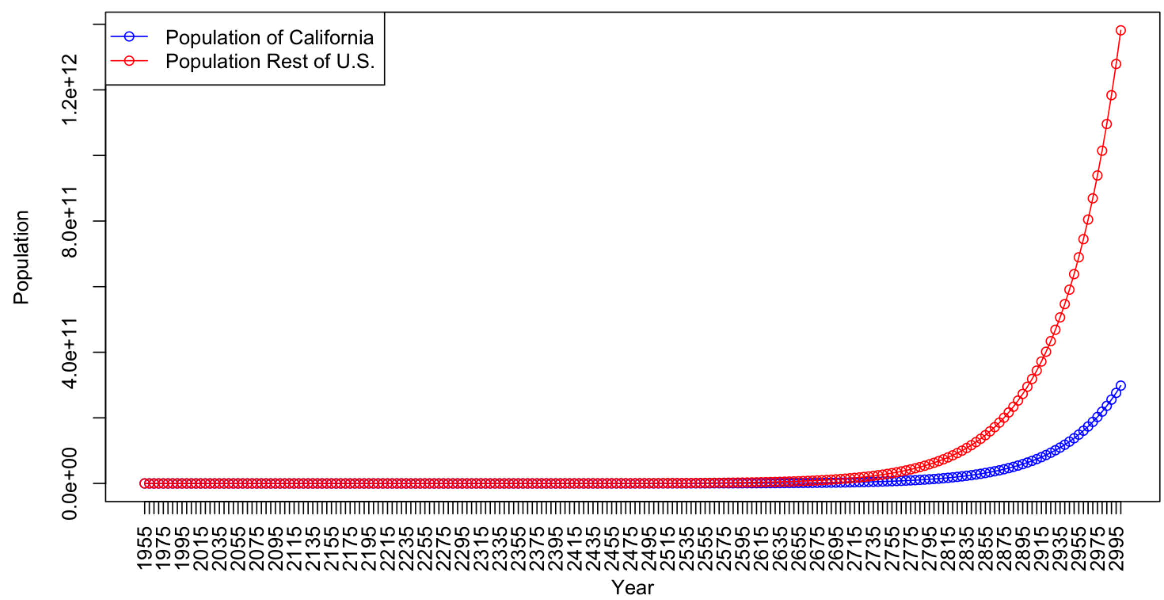
$$and \lim_{t \to \infty} \frac{B_t}{A_t + B_t} \approx 1 - 17.75\% \approx 82.25\%$$

Therefore, the long-term California's population proportion of total US population is approximately 17.75%, and the proportionate population within the rest of US is 82.25% in the long-run.

Below we have a couple of graphs of our 3rd model's population prediction for the years 1955 to 2100 compared to a longer term version from 1955 to 3000.



Population Prediction from year 1955 to 3,000 for combined model



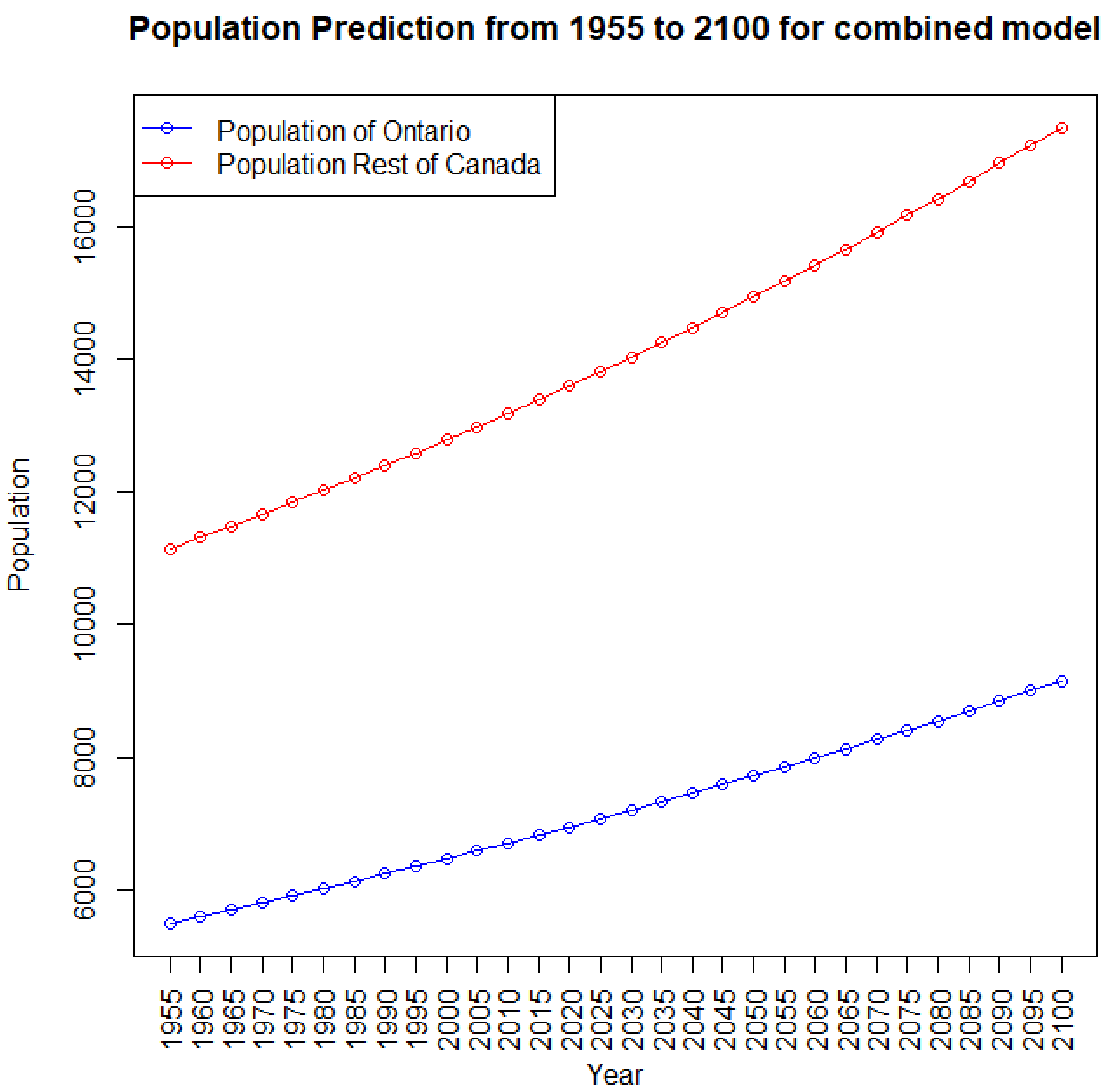
2.3. Extension

We predicted the population growth trends for Ontario and the rest of Canada from 1955 to 2100. The model uses fixed birth rates and death rates (matrices B, D) to calculate the growth matrix G_new and then times the migration rates between the two regions(matrixT) as the same as the 3rd model .

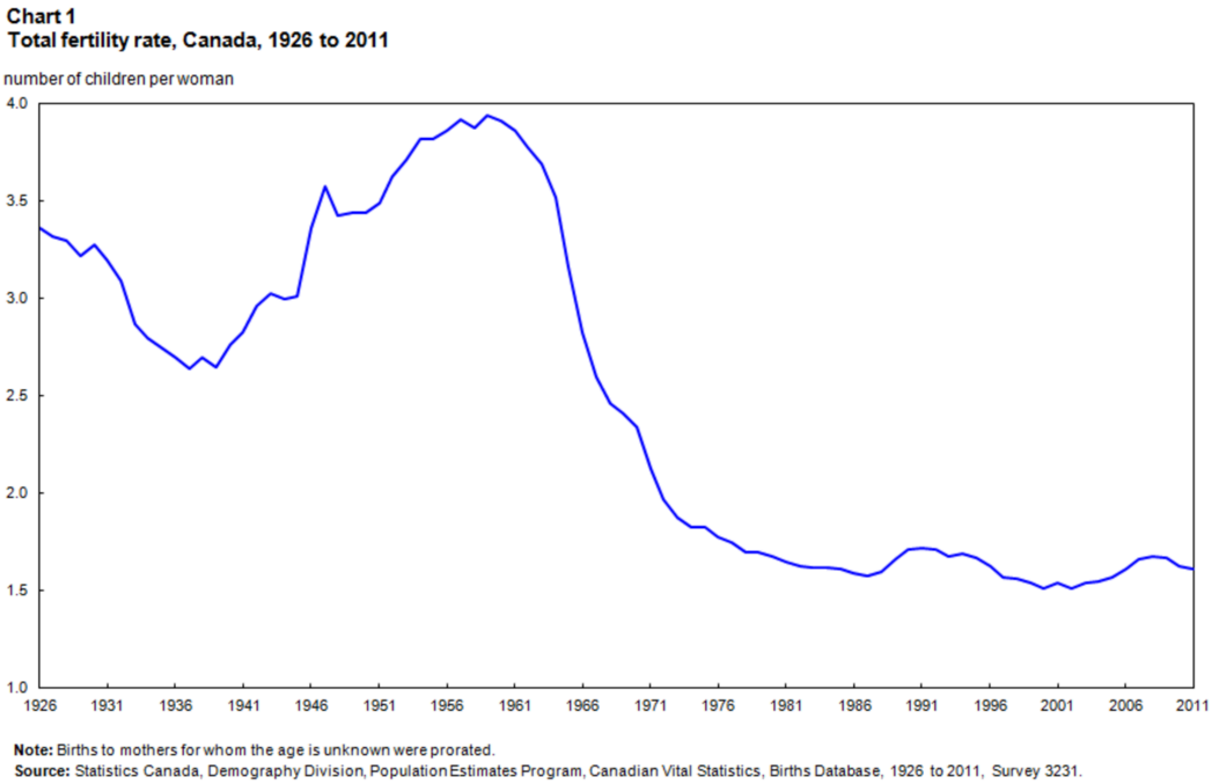
In 1955:

Population	**Birth**	**Death**
16,630,000	414,087	141,355
5,487,900	136,319 (rate: 0.02483992)	46,647 (rate: 0.00849997)
11,142,100	277,768 (rate: 0.024929591)	94,708 (rate: 0.00850001)

	To Ontario	To Rest of Canada	Total 1955 Population
Ontario	5,323,685 (rate: 0.970076896)	164,215 (rate: 0.029923103)	5,487,900
Rest of Canada	178273 (rate: 0.015999946)	10,963,827 (rate: 0.984000053)	11,142,100



Then we get the Estimation here.



Because we observed that the fertility rate dropped sharply between 1960 and 1970, and remained relatively stable after 1970, we decided to split the above 1955-2100 forecast into 1955-1970 and 1970-2100. In this way, the gap between our estimate and the actual value will not be affected by the sudden drop in fertility rate.

Then we search for the data for 1970.

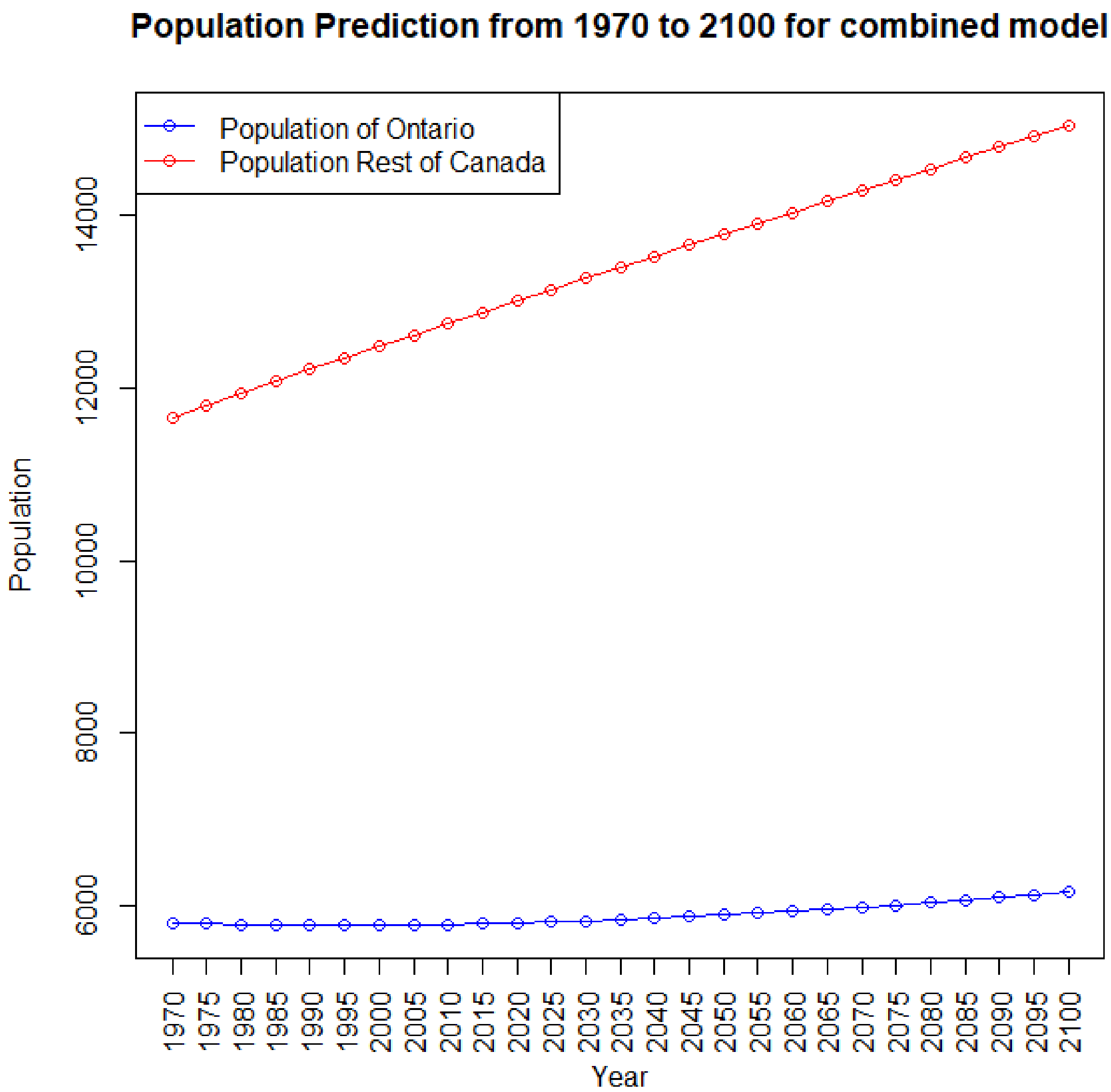
In 1970:

	Population	**Birth**	**Death**
total	21,297,000	315,256	159,533
Ontario	8,483,516	115,914 (rate: 0.013663438)	63,813 (rate: 0.007521999)
Rest of Canada	12,813,484	199342 (rate: 0.015557205)	95,720 (rate: 0.007470255)

	To Ontario	To Rest of Canada	Total 1970 Population
Ontario	8144175 (rate: 0.959999957)	339,341 (rate: 0.040000042)	8,483,516
Rest of Canada	204,936 (rate: 0.015993776)	12,608,548 (rate: 0.984006223)	12,813,484

```
> p3[,4]
[1]  5803.658 11658.023
```

The p_init should be from our first estimation from 1955 until 1970. So in the next code, we will use this p_init as the population initial value



Then we get the final estimation.

3. Result

We created 3 models based on population data for California and the rest of the US from 1955 to 1960 to predict future population trends for the US. Our 3 models are based on the assumption that the future population growth rate and the migration rate will be the same for every 5 years as they were from 1955 to 1960, with Model 1 considering only the population growth rate, Model 2 considering only the migration rate, and Model 3 combining them. Model 3 shows the best results in predicting future population trends for the US, estimating that the population of California and the rest of the US will be around 63,021,940 and 385,912,104 respectively, which is still an overestimate but more realistic compared to the previous ones.

To apply the models to predict population trends for Ontario and the rest of Canada, we use Model 3, which gives us the result that the population of Ontario and the rest of Canada will continue to increase at a similar rate from 1955 to 2100, with Ontario's population being much smaller than that of the rest of Canada. Because of the sharp decline in fertility between 1960 and 1970, we refine the model by splitting the period into 1955 to 1970 and 1970 to 2100 and predict population trends using data from 1970. We find that Ontario's population is still smaller than that of the rest of Canada, but will grow at a much slower rate over the long term than we previously forecast.

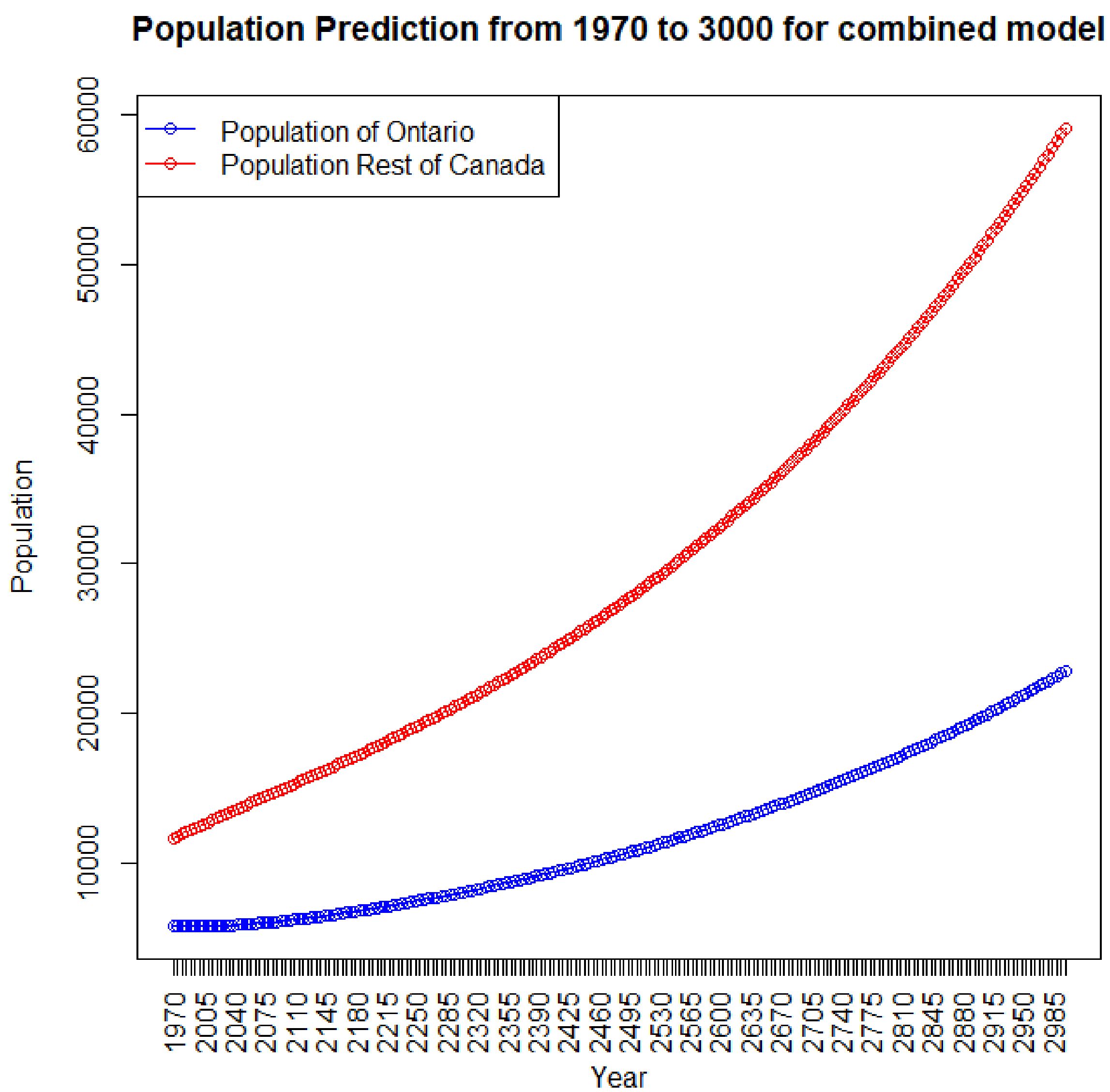
However, there are limitations to the application of the models in reality. Firstly, political factors, global environmental factors (including carrying capacity) and technological development factors affect the growth rate and migration rate in complex ways, and it is impossible for them to remain constant over the years that do not reflect dynamic changes. In addition, different countries and regions within countries have different development processes, which need to be specifically discussed in order to address the prediction of population trends. As a result, our projections are more applicable to scenarios in which external disturbances remain minimal and population trends follow historical patterns. In addition, our models are more applicable to countries and regions that have similar development milestones to California and the rest of the US and Ontario and the rest of Canada.

Our results can be used as a basis for further research on predicting population trends by integrating more recent data and adding different types of variables that affect the birth rate, death rate and migration rate for a particular region, for example, one can use our results as a basis to revise and explore the impact of the development of artificial intelligence on the migration rate of the US population in 2024. Our results can also be used to test the robustness of models in predicting populations in different regions, one can use our models for testing or revise them to get more robust models.

Starting with the 1955 population (p_init), it iteratively computes population changes every 5 years. The results are plotted as a blue curve (Ontario) and a red curve (rest of Canada).

In the long run, the Population of Ontario and the Population Rest of Canada are both increasing.

Since we discovered the birth rates Significantly decreased during 1960-1970, then we decided to use the new data from 1970 to do further estimation.



In the long run, the Population of Ontario and the Population Rest of Canada are both increasing. Because of the significant decline in fertility rates during the 1960s, we predict that Ontario's population will increase much more slowly compared with the first model.

Regional capacity issue:
Although Canada as a whole has not yet faced severe regional capacity issues, continued population growth, especially in densely populated provinces such as Ontario and British Columbia, may increase pressure in certain areas. Therefore, proactive planning and the implementation of sound policies are crucial to achieving a balance between population and resources.

Analysis of Prediction vs. Real Data:
Our prediction model relies on fixed birth rates, death rates, and migration rates, resulting in linear or nearly linear growth over time, but in reality, these factors are dynamically changing. With the improvement of medical, economic, and social conditions, birth and death rates will change, while migration patterns will fluctuate due to policy changes, economic opportunities, and global events. The fixed matrices of B (birth rate), D (death rate), and T (the migration matrix between the two regions) in our model do not reflect long-term trends, such as a decrease in birth rates in Ontario or an increase in growth rates in other provinces, and do not take into account external influences such as changes in immigration policies. In addition, the initial value and migration ratio in 1955 may not be consistent with historical data, ignoring the actual migration flow. Finally, the model lacks consideration for the age structure of the population, which can significantly affect growth as population aging increases mortality rates and reduces the proportion of youth and childbearing age populations, thereby lowering birth rates.

The prediction may be inaccurate because the model assumes fixed parameters for birth, death, and the migration matrix between the two regions(B, D, and T), which can change over time due to factors like economic and cultural shifts or medical advancements. Additionally, global events such as wars, economic crises, and pandemics, as well as the lack of multi-scenario modeling (e.g., high, medium, and low growth), further limit the model's ability to capture future uncertainties.

Compared with the California model, the biggest difference of this model is that we observed the changes in fertility rate and made a segmented model. There are several benefits to dividing the model into multiple parts. It improves clarity by isolating key components such as birth and death rates and migration, allowing for targeted analysis of the impact of each factor. This modular structure increases flexibility and allows for independent adjustments to reflect changes in the real world, such as constantly changing migration patterns or declining birth rates. It also improves accuracy by capturing the dynamics of specific areas and supports scenario testing by modifying various parts to explore different results. Segmenting improves transparency, makes the model easier for stakeholders to interpret, and facilitates error isolation in the event of unexpected results.

4. Discussion

We created 3 models based on population data for California and the rest of the US from 1955 to 1960 to predict future population trends for the US. Our 3 models are based on the assumption that the future population growth rate and the migration rate will be the same for every 5 years as they were from 1955 to 1960, with Model 1 considering only the population growth rate with net migration rate, Model 2 considering only the migration rate between the 2 regions, and Model 3 combining them. Model 3 shows the best results in predicting future population trends for the US, estimating that the population of California and the rest of the US will be around 63,021,940 and 385,912,104 respectively, which is still an overestimate but more realistic compared to the previous ones.

To apply the models to predict population trends for Ontario and the rest of Canada, we use Model 3, which gives us the result that the population of Ontario and the rest of Canada will continue to increase at a similar rate from 1955 to 2100, with Ontario's population being much smaller than that of the rest of Canada. Because of the sharp decline in fertility between 1960 and 1970, we refine the model by splitting the period into 1955 to 1970 and 1970 to 2100 and predict population trends using data from 1970. We find that Ontario's population is still smaller than that of the rest of Canada, but will grow at a much slower rate over the long term than we previously forecast.

However, there are limitations to the application of the models in reality. Firstly, political factors, global environmental factors (including carrying capacity) and technological development factors affect the growth rate and migration rate in complex ways, and it is impossible for them to remain constant over the years that do not reflect dynamic changes. In addition, different countries and regions within countries have different development processes, which need to be specifically discussed in order to address the prediction of population trends. As a result, our projections are more applicable to scenarios in which external disturbances remain minimal and population trends follow historical patterns. In addition, our models are more applicable to countries and regions that have similar development milestones to California and the rest of the US and Ontario and the rest of Canada.

Our results can be used as a basis for further research on predicting population trends by integrating more recent data and adding different types of variables that affect the birth rate, death rate and migration rate for a particular region, for example, one can use our results as a basis to revise and explore the impact of the development of artificial intelligence on the migration rate of the US population in 2024. Our results can also be used to test the robustness of models in predicting populations in different regions, one can use our models for testing or revise them to get more robust models.

5. References

U.S. Census Bureau. (2021). *P1: Total population. 2020 Census Redistricting Data (Public Law 94-171) Summary File*. Retrieved November 27, 2024, from https://data.census.gov

Statistics Canada. (n.d.). *Population Projections for Canada, Provinces and Territories*, from https://www150.statcan.gc.ca/n1/pub/11-630-x/11-630-x2014002-eng.htm

PopulationPyramid.net. (n.d.). *Population Pyramids of the World from 1950 to 2100*, from https://www.populationpyramid.net/

CEIC Data. (n.d.). *Global Economic Data, Indicators, Charts & Forecasts*, from https://www.ceicdata.com/en

6. Appendix: R Code

6.1 R-code of Models for years 1955 to 2100:

```r
#BASE MODEL
#Model 1
I<- matrix(c(1,0,0,1), nrow = 2)
I
B<- matrix(c(0.1315060,0,0,0.1282137), nrow = 2)
B
D<- matrix(c(0.0472744,0,0,0.0487697), nrow = 2)
D
M<- matrix(c(0.0865414,0,0,-0.0073908), nrow = 2)
M
G<-I+B-D+M
G
t_vals <- seq(1955, 2100, by = 5)
p_init <- c(12988, 152082)
p1 <- matrix(0, nrow = 2, ncol = length(t_vals))
p1[,1] <- p_init
for (t in 2:length(t_vals)) {
  p1[,t] <- G %*% p1[,t - 1]
}
plot(t_vals, p1[1 ,], type = "o", col = "blue",
     xlab = "Year", ylab = "Population", ylim = range(p1),
     main = "Population Prediction from 1955 to 2100 model 1",xaxt = "n")
lines(t_vals, p1[2 ,], type = "o", col = "red")
axis(1, at = t_vals, labels = t_vals, las = 2)
legend("topleft", legend = c("Population of California", "Population Rest of U.S."),
        col = c("blue", "red"), lty = 1, pch = c(1, 1))
```

```r
#In 2020
(2020-1955)/5=13  Since 1955 is at t=1 in this code, 1+13=14.
p1[,14]
#Model 2
T <- matrix(c(0.937326763, 0.062673236,0.012743125, 0.987256874), nrow = 2)
T
t_vals <- seq(1955, 2100, by = 5)
p_init <- c(12988, 152082)
p2 <- matrix(0, nrow = 2, ncol = length(t_vals))
p2[,1] <- p_init
for (t in 2:length(t_vals)) {
  p2[,t] <- T %*% p2[,t - 1]
}
plot(t_vals, p2[1 ,], type = "o", col = "blue",
     xlab = "Year", ylab = "Population", ylim = range(p2),
     main = "Population Prediction from 1955 to 2100 for model 2",xaxt = "n")
lines(t_vals, p2[2 ,], type = "o", col = "red")
axis(1, at = t_vals, labels = t_vals, las = 2)
legend("left", legend = c("Population of California", "Population Rest of U.S."),
```

```
       col = c("blue", "red"), lty = 1, pch = c(1, 1))
#In 2020
p2[,14]
#Model 3
G_new <- I+B-D
T <- matrix(c(0.937326763, 0.062673236,0.012743125, 0.987256874), nrow = 2)
t_vals <- seq(1955, 2100, by = 5)
p_init <- c(12988, 152082)
p3 <- matrix(0, nrow = 2, ncol = length(t_vals))
p3[,1] <- p_init
for (t in 2:length(t_vals)) {
  p3[,t] <- G_new%*%T %*% p3[,t - 1]
}
plot(t_vals, p3[1 ,], type = "o", col = "blue",
     xlab = "Year", ylab = "Population", ylim = range(p3),
     main = "Population Prediction from 1955 to 2100 for combine model",xaxt = "n")
lines(t_vals, p3[2 ,], type = "o", col = "red")
axis(1, at = t_vals, labels = t_vals, las = 2)
legend("topleft", legend = c("Population of California", "Population Rest of U.S."),
       col = c("blue", "red"), lty = 1, pch = c(1, 1))
#In 2020
p3[,14]
```

6.2 R-code of Models for year 1955 to 3000:

```
#BASE MODEL
#Model 1
I<- matrix(c(1,0,0,1), nrow = 2)
I
B<- matrix(c(0.1315060,0,0,0.1282137), nrow = 2)
B
D<- matrix(c(0.0472744,0,0,0.0487697), nrow = 2)
D
M<- matrix(c(0.0865414,0,0,-0.0073908), nrow = 2)
M
G<-I+B-D+M
G
t_vals <- seq(1955, 3000, by = 5)
p_init <- c(12988, 152082)
p1 <- matrix(0, nrow = 2, ncol = length(t_vals))
p1[,1] <- p_init
for (t in 2:length(t_vals)) {
  p1[,t] <- G %*% p1[,t - 1]
}
plot(t_vals, p1[1 ,], type = "o", col = "blue",
     xlab = "Year", ylab = "Population",
     main = "Population Prediction from year 1955 to 3,000 for model 1",xaxt = "n")
lines(t_vals, p1[2 ,], type = "o", col = "red", ylim = range(p1) )
axis(1, at = t_vals, labels = t_vals, las = 2)
legend("topleft", legend = c("Population of California", "Population Rest of U.S."),
       col = c("blue", "red"), lty = 1, pch = c(1, 1))
#In 2020
  # (2020-1955)/5 + 1 =14
      # for ex: since (1955-1955) / 5 = 0 / 5 = 0, but R starts at 1
      # so (1955-1955) / 5 + 1  = 1
p1[,14]
```

```
#Model 2
T <- matrix(c(0.937326763, 0.062673236,0.012743125, 0.987256874), nrow = 2)
T
t_vals <- seq(1955, 3000, by = 5)
p_init <- c(12988, 152082)
p2 <- matrix(0, nrow = 2, ncol = length(t_vals))
p2[,1] <- p_init
for (t in 2:length(t_vals)) {
  p2[,t] <- T %*% p2[,t - 1]
}
plot(t_vals, p2[1 ,], type = "o", col = "blue",
     xlab = "Year", ylab = "Population", ylim = range(p2),
     main = "Population Prediction from year 1955 to 3,000 for model 2",xaxt = "n")
lines(t_vals, p2[2 ,], type = "o", col = "red")
axis(1, at = t_vals, labels = t_vals, las = 2)
legend("left", legend = c("Population of California", "Population Rest of U.S."),
       col = c("blue", "red"), lty = 1, pch = c(1, 1))
```

```
#Model 3
G_new <- I+B-D
G_new
T <- matrix(c(0.937326763, 0.062673236,0.012743125, 0.987256874), nrow = 2)
T
t_vals <- seq(1955, 3000, by = 5)
p_init <- c(12988, 152082)
p3 <- matrix(0, nrow = 2, ncol = length(t_vals))
p3[,1] <- p_init
for (t in 2:length(t_vals)) {
  p3[,t] <- G_new %*% T %*% p3[, t - 1]
}
```

```
plot(t_vals, p3[1 ,], type = "o", col = "blue",
     xlab = "Year", ylab = "Population", ylim = range(p3) ,
     main = "Population Prediction from year 1955 to 3,000 for combined model", xaxt = "n")
lines(t_vals, p3[2 ,], type = "o", col = "red")
axis(1, at = t_vals, labels = t_vals, las = 2)
legend("topleft", legend = c("Population of California", "Population Rest of U.S."),
       col = c("blue", "red"), lty = 1, pch = c(1, 1))

#In 2020
p3[,14]
```

Extension R-code:

```
#1955 ca
I <- matrix(c(1, 0, 0, 1), nrow = 2)
B <- matrix(c(0.02483992, 0, 0, 0.024929591), nrow = 2)
D <- matrix(c(0.00849997, 0, 0, 0.00850001), nrow = 2)
G_new <- I+B-D
T <- matrix(c(0.970076896, 0.029923103,0.015999946, 0.984000053), nrow = 2)
t_vals <- seq(1955, 2100, by = 5)
p_init <- c(5487.9, 11142.1)
p3 <- matrix(0, nrow = 2, ncol = length(t_vals))
p3[, 1] <- p_init
for (t in 2:length(t_vals)) {
  p3[,t] <- G_new%*%T %*% p3[,t - 1]
}
plot(t_vals, p3[1 ,], type = "o", col = "blue",
     xlab = "Year", ylab = "Population", ylim = range(p3),
     main = "Population Prediction from 1955 to 2100 for combined model",xaxt = "n")
lines(t_vals, p3[2 ,], type = "o", col = "red")
axis(1, at = t_vals, labels = t_vals, las = 2)
legend("topleft", legend = c("Population of Ontario", "Population Rest of Canada"),
       col = c("blue", "red"), lty = 1, pch = c(1, 1))
```

```
#1970
I <- matrix(c(1, 0, 0, 1), nrow = 2)
B <- matrix(c(0.013663438, 0, 0, 0.015557205), nrow = 2)
D <- matrix(c(0.007521999, 0, 0, 0.007470255), nrow = 2)
G_new <- I+B-D
T <- matrix(c(0.959999957, 0.040000042,0.015993776, 0.984006223), nrow = 2)
t_vals <- seq(1970, 2100, by = 5)
p_init <- c(5803.658, 11658.023)
p3 <- matrix(0, nrow = 2, ncol = length(t_vals))
p3[, 1] <- p_init
for (t in 2:length(t_vals)) {
  p3[,t] <- G_new%*%T %*% p3[,t - 1]
}
plot(t_vals, p3[1 ,], type = "o", col = "blue",
     xlab = "Year", ylab = "Population", ylim = range(p3),
     main = "Population Prediction from 1970 to 2100 for combined model",xaxt = "n")
lines(t_vals, p3[2 ,], type = "o", col = "red")
axis(1, at = t_vals, labels = t_vals, las = 2)
legend("topleft", legend = c("Population of Ontario", "Population Rest of Canada"),
       col = c("blue", "red"), lty = 1, pch = c(1, 1))
```

```
#1970-3000
I <- matrix(c(1, 0, 0, 1), nrow = 2)
B <- matrix(c(0.013663438, 0, 0, 0.015557205), nrow = 2)
D <- matrix(c(0.007521999, 0, 0, 0.007470255), nrow = 2)
G_new <- I+B-D
T <- matrix(c(0.959999957, 0.040000042,0.015993776, 0.984006223), nrow = 2)
t_vals <- seq(1970, 3000, by = 5)
p_init <- c(5803.658, 11658.023)
p3 <- matrix(0, nrow = 2, ncol = length(t_vals))
p3[, 1] <- p_init
for (t in 2:length(t_vals)) {
  p3[,t] <- G_new%*%T %*% p3[,t - 1]
}
plot(t_vals, p3[1 ,], type = "o", col = "blue",
     xlab = "Year", ylab = "Population", ylim = range(p3),
     main = "Population Prediction from 1970 to 3000 for combined model",xaxt = "n")
lines(t_vals, p3[2 ,], type = "o", col = "red")
axis(1, at = t_vals, labels = t_vals, las = 2)
legend("topleft", legend = c("Population of Ontario", "Population Rest of Canada"),
       col = c("blue", "red"), lty = 1, pch = c(1, 1))
```

7. Appendix: Individual Contributions

Report:

1. Introduction: Jingyan Li

2.1. Base Model: Hanning Qi, Shakar Saeed

2.2. Analysis of Base Model: Shakar Saeed

2.3. Extension: Tianchi Ai

3. Result: Tianchi Ai

4. Discussion: Xinzhe Wang

5. References: Hanning Qi, Tianchi Ai

6.1. R Code: Hanning Qi

6.2. R Code: Hanning Qi, Shakar Saeed

Extension R-code: Tianchi Ai

Presentation Slides: Jingyan Li, Xinzhe Wang