The Chinese University of Hong Kong, Shenzhen
MAT3300 2024-2025 Fall

Final Project 1

-

Global Population Forecasting
Based on **Logistic**, **Exponential Decay** and **Polynomial Model**

Yinqi (Marcus) Zhong
marazhong1@gmail.com
+(86) 13710210206

January 12, 2025

1 Abstract

Global population dynamics are essential for understanding future challenges in economic development, environmental sustainability, and social stability. This study used historical data from 1970 to 2022 for 234 countries to forecast population changes up to 2050 using logistic model, decay model and polynomial regression. The logistic model accounted for growth or stable patterns. The decay model is used when the population is declining, while polynomial regression suited for countries where logistic and decay fits fail. By categorizing countries into rapid growth, stable, or declining trends, and integrating key factors like population density and land area, the analysis includes regional and global patterns. Visualizations of the findings may offer policymakers and planners suggestions into the implications of population growth and decline at both regional and global levels.

2 Intro

Understanding global population trends is essential for addressing future challenges in economic development, resource allocation, and social stability. By 2050, population dynamics will determine aspects of infrastructure, environmental sustainability, and public policy. However, these trends vary significantly across countries due to factors such as land area, population density, and economic development.

This project used historical population data from 1970 to 2022 for over 200 countries to forecast population changes up to 2050. The logistic growth model which considers saturation and geographic limitations was used for initial fitting. The decay model is used when the population is declining. However, in cases where logistic model and decay model failed to converge, polynomial regression was used as an alternative. This dual-model approach ensured forecasting even in complex or irregular growth situations.

3 Literature Review

Population forecasting is a vital field of research that intersects with demography, economics, and environmental studies. Various models and methodologies have been employed to project population dynamics and trends.

One significant study is by Verhulst (1838), who introduced the logistic growth model, a foundational framework for analyzing population growth with resource constraints. This model effectively captures saturation effects, making it suitable for populations approaching carrying capacity. More recent research, such as the work by Lee and Tuljapurkar (1994), explores probabilistic forecasting, which accounts for uncertainties in fertility and mortality rates. These methods offer a broader understanding of long-term population trajectories.

For populations exhibiting consistent decline, decay models have emerged as a practical approach for forecasting future trends. The decay model, often modified with a baseline to reflect a minimum stable population, captures the gradual nature of population decrease while accounting for potential stabilization effects. This model provides a flexible framework for analyzing and projecting declining population trends.

Cohen (2003) provides an extensive review of population models, emphasizing the importance of incorporating demographic variables and economic indicators. Similarly, Preston, Heuveline, and Guillot (2001) highlight the role of age-specific fertility and mortality rates in global population trends. Their findings have been instrumental in refining logistic and polynomial models for country-level analyses.

For cases where logistic models fail to converge, alternative techniques such as polynomial regression have proven effective. Research by Hyndman and Booth (2008) demonstrates the application of these models in irregular or declining population scenarios, show their flexibility in addressing complex growth patterns.

Overall, the literature establishes the importance of combining multiple modeling techniques, including logistic model, decay model and polynomial regression, to address diverse population dynamics and regional variations.

4 Visualization of Historical Trends

4.1 Data Preprocessing

The dataset is complete, containing no missing values across all columns. This allows for direct analysis without the need for imputation or data cleaning.

4.2 Population Growth Patterns

The average annual growth rate (AAGR) for each country was calculated using the

Formula:

$$r = \left(\frac{P_1}{P_0}\right)^{\frac{1}{t_1 - t_0}} - 1$$

Where:

P_0: Population in 1970.

P_1: Population in 2022.

t_0: Initial year (1970).

t_1: Final year (2022).

Countries were classified into five categories based on their average annual growth rates. Most of the countries are still experiencing a growing population. Only few countries have population declining.

Growth Category	Number of Countries
Rapid Growth (> 0.02)	91
Moderate Growth (0.01 – 0.02)	64
Slow Growth (0 – 0.01)	64
Stable (-0.005 – 0)	12
Declining (< -0.005)	3

4.3 Correlation Analysis

4.3.1 Growth Rate vs. Population Density

The scatter plot highlights that population density alone does not determine growth, as countries with similar densities can have widely varying growth rates. The correlation coefficient is -0.01.

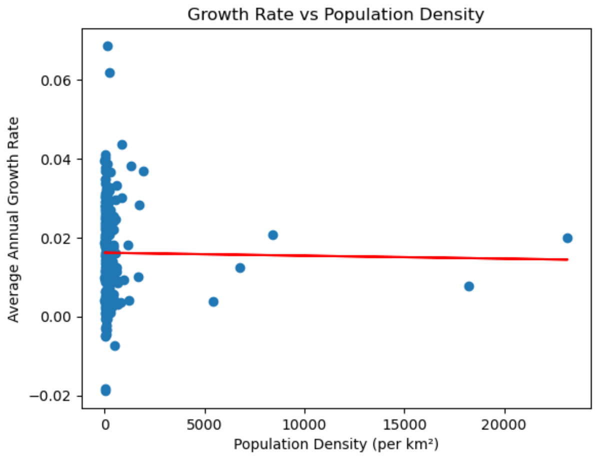

Figure 1: Growth Rate vs. Population Density

4.3.2 Growth Rate vs. Land Area

The scatter plot highlights that land area alone does not determine growth, as countries with similar land areas can have widely varying growth rates. The correlation coefficient is -0.02.

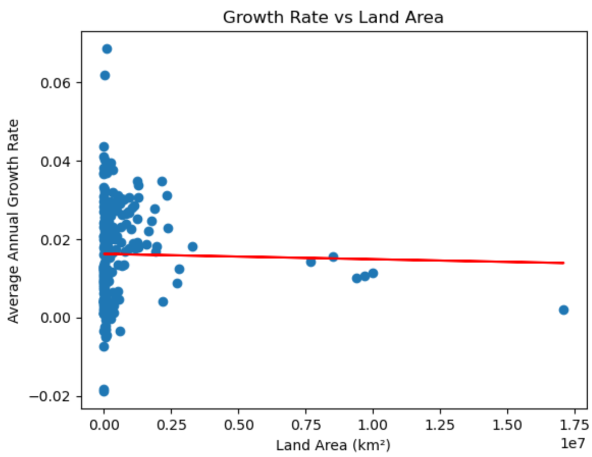

Figure 2: Growth Rate vs. Land Area

4.3.3 Growth Rate vs. Population Percentage

The scatter plot highlights that population percentage of the corresponding country alone does not determine growth, as countries with similar percentage can have widely varying growth rates. The correlation coefficient is 0.01.

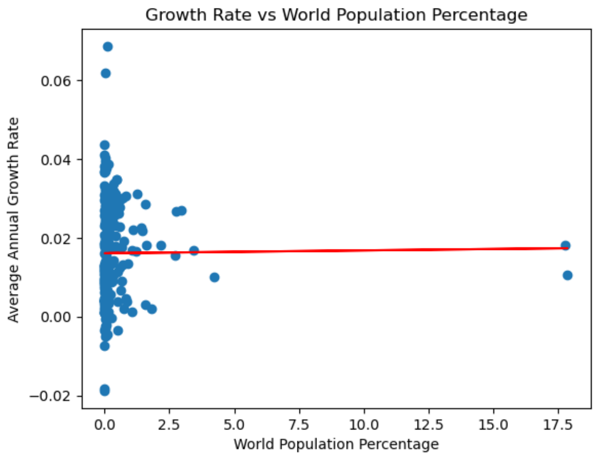

Figure 3: Growth Rate vs. Land Area

4.3.4 Growth Rate by Continent

The average annual growth rates were calculated by continent, revealing significant regional differences.

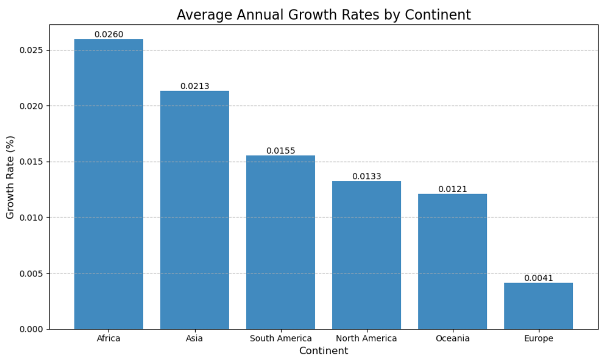

Figure 4: Growth Rate by Continent

From the graph, we can conclude that Africa has the largest growth rate, reaching 0.026 and Europe has the smallest growth rate at 0.0041.

5 Model Development

Based on the literature, the project will forecast the future trend of population by the following models.

5.1 Logistic Growth Model

The Logistic Model was formulated as:

$$P(t) = \frac{K}{1 + \exp(-r(t - t_0))}$$

Where:

- $P(t)$: Population at time t
- K: Carrying capacity
- r: Growth rate
- t_0: Inflection point

5.2 Exponential Decay Model

The Decay Model was formulated as:

$$P(t) = P_0 \cdot \exp(-r \cdot (t - t_0)) + C$$

Where:

- $P(t)$: Population at time t
- P_0: Initial population at the start of the decline
- r: Decay rate, representing the rate of population decrease
- t_0: Start year of the decline
- C: Baseline population, representing the population after the decline stabilizes

5.3 Polynomial Regression Model

For countries where logistic models failed (e.g., due to irregular or declining growth), polynomial regression was used. The model is expressed as:

$$P(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_n t^n$$

Where n=3 (cubic polynomial). The model was chosen for its flexibility in capturing non-standard growth patterns, such as declining populations or populations with irregular trends.

5.4 Train and Validation Splits

The dataset is divided into training data using population ranges from 1970 to 2010, and data from 2015 to 2022 as validation data.

5.5 Initial Parameter Estimation

5.5.1 Logistic Model

- K_0: Set to $1.5 \times$ max training population, assuming potential for future growth.
- r_0: Initial growth rate set to 0.02.
- t_0: Middle of the training period (1990).

5.5.2 Exponential Decay Model

- P_0: Set to the first declining population data point.
- r_0: Initial decay rate set to 0.05 for a gradual decline.
- t_0: Start year of the decline, typically the first year with a recorded decrease.
- C_0: Baseline population set to 50% of the last known declining population for smoothing.

5.5.3 Polynomial Model

- n: The degree of polynomial was set to be 3, providing a balance between flexibility and model simplicity.

6 Result

6.1 Countries Categories

Number of countries fitting the logistic model: 190, examples: Austria, China, France, India, etc.

Number of countries fitting the Decay model: 26, examples: Greece, Japan, Serbia, Ukraine, etc.

Number of countries fitting Polynomial Model: 12, examples: New Zealand, Nigeria, Austria, Denmark, etc.

6.2 Global Population Projection (2022-2050)

The global population was projected using historical data (1970-2022) and predictions every 5 years up to 2050. The combined logistic model, decay model and polynomial model captured the overall growth trend.

The graph below demonstrates the global population growth trend, from historical data to future projections. By 2050, the world population is projected to be **near 10 billion**.

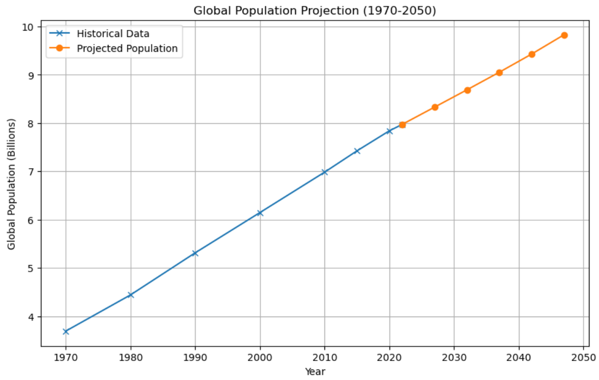

Figure 5: Global Population Projection (1970–2050)

6.3 Regional Forecast Comparisons

Population projections were also analyzed at the continental level for regional trends. The detailed breakdown of trends is plotted.

Africa: The most significant growth is projected in Africa, where population increases are driven by high fertility rates and improving healthcare.

Asia: Asia remains the region with the largest population and is expected to increase steadily.

Europe: Europe shows minimal growth, and some countries experience a population decline.

North and South America: Both regions are projected to grow steadily, albeit at slower rates compared to Africa and Asia.

Oceania: While population growth in Oceania is modest, growth remains positive.

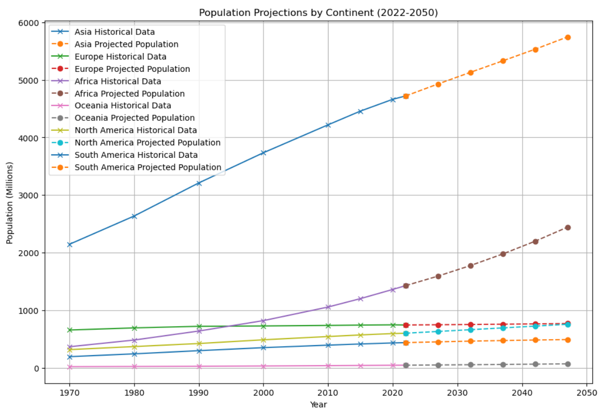

Figure 6: Population Projections by Continent (2022-2050)

6.4 Country-wise Forecast Comparisons

Individual country trends highlight the diversity in population. The followings show forecasting results of representative countries fitting different models.

6.4.1 Logistic Growth Model - China & Australia

Historical Data: Both China and Australia experienced rapid growth from 1970 to 2020.

Prediction Analysis: The population growth in both countries is expected to stabilize, with China's population projected to reach 1.5 billion and Australia's population to reach 35 million by 2050.

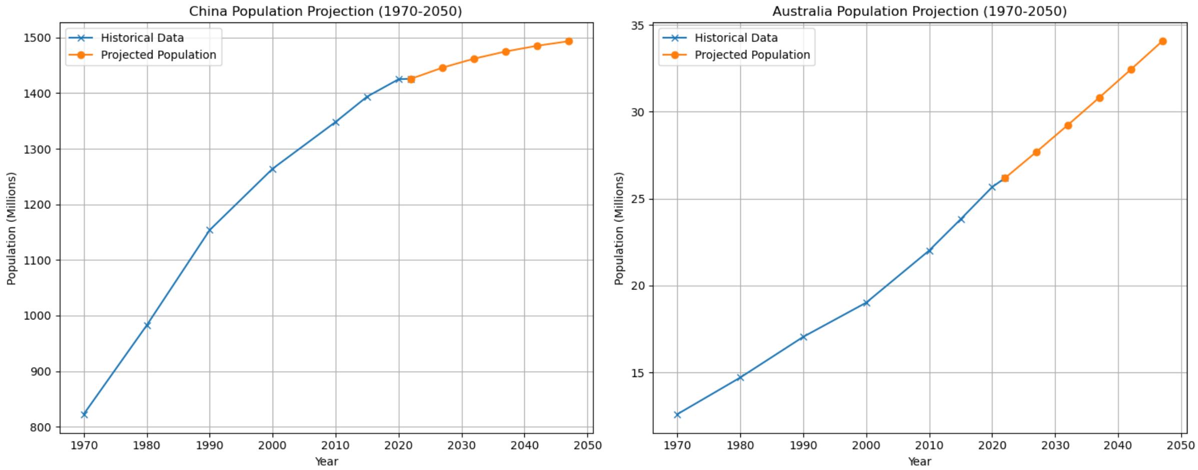

Figure 7: Logistic Growth Model - China & Australia

6.4.2 Exponential Decay Model - Japan & Greece

Historical Data: Both Japan and Greece experienced consistent growth from 1970 to 2000, reaching its peak at 2010. A noticeable decline began after 2010.

Prediction Analysis: The decay model predicts a gradual population decline from 2022 onward. This aligns with demographic trends and economic challenges faced by Japan, including labor shortages.

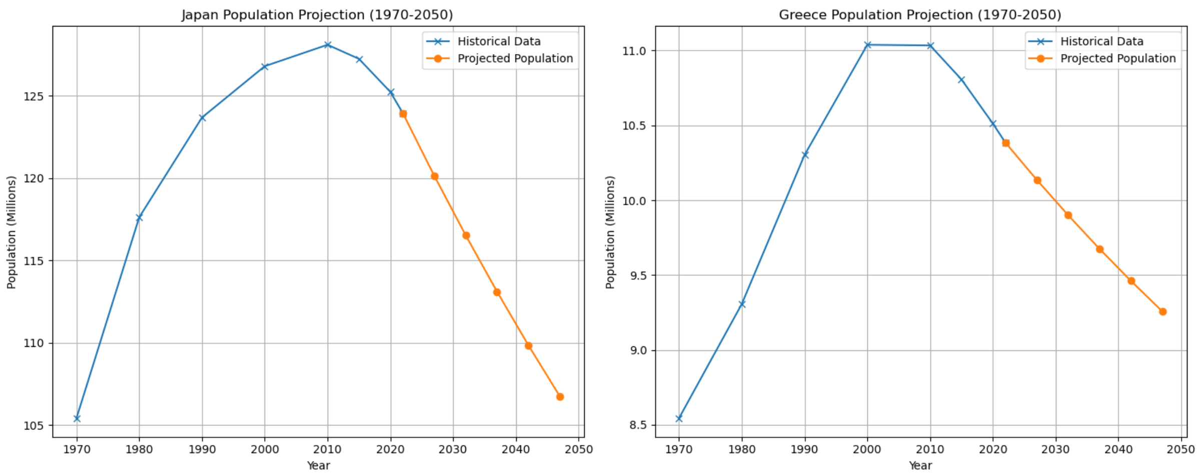

Figure 8: Exponential Decay Model - Japan & Greece

6.4.3 Polynomial Regression Model - New Zealand & Austria

The polynomial regression model is used as a fallback solution when logistic & decay model does not fit. New Zealand and Austria are chosen to visualize.

Historical Data: Both countries display rapid growth from 1970 to 2020, and the growth speed is increasing.

Prediction Analysis: The rapid growth is expected to continue until 2050.

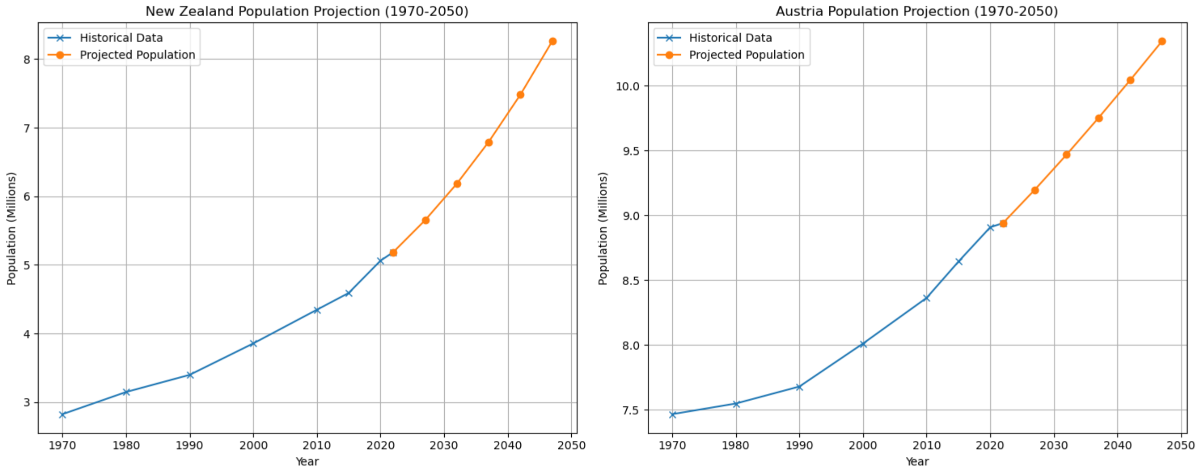

Figure 9: Polynomial Regression Model - New Zealand & Austria

7 Future Improvements

While the current model provides accurate forecasts for most countries, there are several opportunities to dive deeper and further enhance its performance.

External Factors: Incorporating external factors is highly recommended to provide a more comprehensive understanding of growth trends. For instance, the rapid population growth in New Zealand and Austria may be closely linked to immigration rates. Additionally, factors such as government policies, economic conditions, and environmental challenges could significantly influence long-term population trends.

Smoother Time Slots: Furthermore, collecting more detailed data at smaller time intervals could help create smoother trend lines. For countries with fluctuating population data, using more detailed temporal data could make patterns of variation clearer, contributing to improve the model's predictive performance.

Model Selection: In terms of model selection, this study employs the logistic growth model, exponential decay model, and polynomial model for forecasting. While these models are effective, their parameters could be more carefully fine-tuned to meet specific improvement needs. Exploring more complex models might also deliver more precise and accurate results.

8 Conclusion

This study analyzed global population trends using historical data (1970–2022) and projected future populations up to 2050. By using a combination of logistic model, decay model and polynomial regression, the analysis provided predictions across diverse growth patterns and regions. The average relative RMSE of 6.69% indicates that the models used are reliable for long-term population forecasting. The following key conclusions were drawn:

- The world population is projected to be near 10 billion by 2050, driven primarily by growth in regions like Africa and parts of Asia.

- Africa is expected to experience the highest growth, highlighting the need for significant investments in infrastructure, education, and healthcare.
- Europe and some developed nations are facing population declines, emphasizing the need for policies to counteract aging populations.
- North and South America are expected to grow steadily, highlighting their balanced demographic trends and relatively stable fertility rates.
- Stable countries such as China and Australia suits for logistic model, as the population will stabilize in the long run.
- Declining countries such as Japan and Greece can be analyzed using decay model.
- Growing countries such as New Zealand and Austria exhibits exponential-like growth, which suit for polynomial model.
- External factors, smoother time slots and more complex models may be employed to improve model performance.

References

[1] Cohen, J. E. (2003). *Human Population: The Next Half Century*. Science, *302*(5648), 1172–1175. https://doi.org/10.1126/science.1088665

[2] Hyndman, R. J., & Booth, H. (2008). *Stochastic population forecasts using functional data models for mortality, fertility and migration*. International Journal of Forecasting, *24*(3), 323–342. https://doi.org/10.1016/j.ijforecast.2008.02.009

[3] Lee, R., & Tuljapurkar, S. (1994). *Stochastic population forecasts for the United States: Beyond high, medium, and low*. Journal of the American Statistical Association, *89*(428), 1175–1189. https://doi.org/10.1080/01621459.1994.10476857

[4] Preston, S. H., Heuveline, P., & Guillot, M. (2001). *Demography: Measuring and Modeling Population Processes*. Wiley-Blackwell.

[5] Verhulst, P. F. (1838). *Notice sur la loi que la population suit dans son accroissement*. Correspondance Mathématique et Physique, *10*, 113–121.