

AGENT-BASED SIMULATION OF POPULATION DYNAMICS

Group 6

Wenxi Zhang, Jiaqi Yan, Yuehan Zhang, Yushuhong Lin, Zeyu Wang, Zicong He

GitHub Repository: <https://github.gatech.edu/yzhang4046/CSE6730>

Abstract

This project presents an agent-based simulation model that explores how individual decisions collectively shape population dynamics over time. The simulation incorporates key demographic processes such as fertility, mortality, migration, and partnership formation and accounts for how these behaviors are influenced by socioeconomic factors like education, income, and healthcare quality.

To support experimentation and policy analysis, the model offers an interactive interface through which users can adjust parameters and introduce events such as baby booms or immigration surges. The simulation proceeds in yearly time steps, tracking changes in population size, age structure, urban-rural distribution, and dependency ratios.

Empirical data and real-world policy settings are used to calibrate the model, enabling scenario testing and validation. Results show that small changes in policy inputs can produce significant long-term effects, and that interactions between education, healthcare, and fertility are often nonlinear. The simulation framework provides a flexible tool for analyzing the demographic impacts of policy decisions and supports the design of more effective, data-driven interventions.

Contents

1	Project description	4
1.1	Project Goals	4
1.2	Relevant Aspects of Population Dynamics	4
2	Literature Review	5
3	Conceptual Model	6
3.1	Conceptual Model of Population Simulation	6
3.2	Conceptual Model of Policy Parameters and Impacts	7
4	Simulation Model	9
4.1	Simulation Summary	9
4.2	Verification of Implementation	9
5	Experimental Results and Validation	11
5.1	Simulation Study and Experimental Procedure	11
5.2	Validation and Output Analysis	12
5.3	Limitations and Sources of Error	12
6	Discussion, Conclusions, and Summary	14
6.1	Key Findings and Insights	14
6.2	Recommendations for Future Work	14

1 Project description

1.1 Project Goals

This project focuses on building a simulation that reflects how populations change over time through everyday choices. The model is based on how people form families, move between places, and respond to changing social and economic conditions. Rather than relying on fixed formulas, it uses a flexible structure that lets users test different ideas and policies in a controlled environment.

A key part of the design is accessibility. The simulation comes with an interface that allows users to adjust settings, run scenarios, and see the outcomes directly. Its purpose is to support experimentation, making it easier for people to understand how various factors such as education, income, or healthcare shape population trends.

1.2 Relevant Aspects of Population Dynamics

The model focuses on individuals and how they make decisions that influence population-level outcomes. Each person in the simulation has basic characteristics like age, gender, education, income, and whether they live in an urban or rural area. These traits affect choices around fertility, partnership, and mobility.

Fertility is not only linked to age but also influenced by education level, income, and family size. Mortality rates change with age but are also affected by access to healthcare and general living conditions. Policies such as child support, education investment, and healthcare access can shift these patterns by changing the context in which decisions are made.

The simulation also takes location into account. People living in cities and those in rural areas often respond differently to the same policies. For example, education tends to lower fertility more strongly in urban settings. Users can also add specific events like baby booms or migration waves, which let the model reflect sudden shifts as well as gradual change.

2 Literature Review

Currently, the world faces challenges such as rapid urbanization, population aging, and large-scale regional migration. These phenomena have far-reaching impacts on socioeconomic conditions, public resource allocation, and policymaking. Reliable population data is essential for tracking progress, designing policies, and implementing programs effectively. However, in low and middle income countries, data collection is often inconsistent due to resource limitations, political manipulation, or lack of prioritization [1]. Over reliance on estimates instead of empirical data weakens policy decisions, highlighting the need for accurate and country specific population data. Therefore, constructing a detailed and comprehensive simulation system is of paramount importance.

In previous population simulation studies, researchers have primarily relied on several classical methods to describe and forecast population dynamics. First, classical population growth models are used. For example, the Malthusian model, based on Thomas Malthus’s theory of population growth proposed in 1798, posits that in the absence of resource constraints, the population grows exponentially [2]. In contrast, the logistic model, introduced by Pierre Franois Verhulst in 1838, describes how population growth eventually saturates when environmental resources and carrying capacity are considered [3].

Additionally, to capture the nuances of population structure changes, researchers have widely employed the Leslie matrix model, which focuses on the survival and fertility rates of different age groups to predict future age distributions and structural evolution [4]. There have also been other models: for instance, population pharmacodynamic (PD) models describe how drug effects change over time by linking drug exposure to response, offering a deeper understanding than single assessments. These models enable the simulation of different dosing regimens, aiding in dose optimization and study design. As the third paper in a series, this work introduces methods for developing and evaluating population PD models [5].

Meanwhile, to provide reliable initial conditions and boundary constraints for macro level system dynamics models or agent-based models, scholars have utilized synthetic population generation techniques. By integrating census data with sample survey data, methods such as Iterative Proportional Fitting and genetic algorithms are used to generate virtual populations with detailed household and individual attributes [6].

Our research plan will build upon these established methods by combining them with agent-based simulation techniques. Through calibration with historical data, sensitivity analysis, and scenario simulations, we aim to further explore the comprehensive impact of policy, economic, and social factors on population change and validate the practicality and predictive accuracy of our model.

3 Conceptual Model

3.1 Conceptual Model of Population Simulation

The simulated workflow follows these steps:

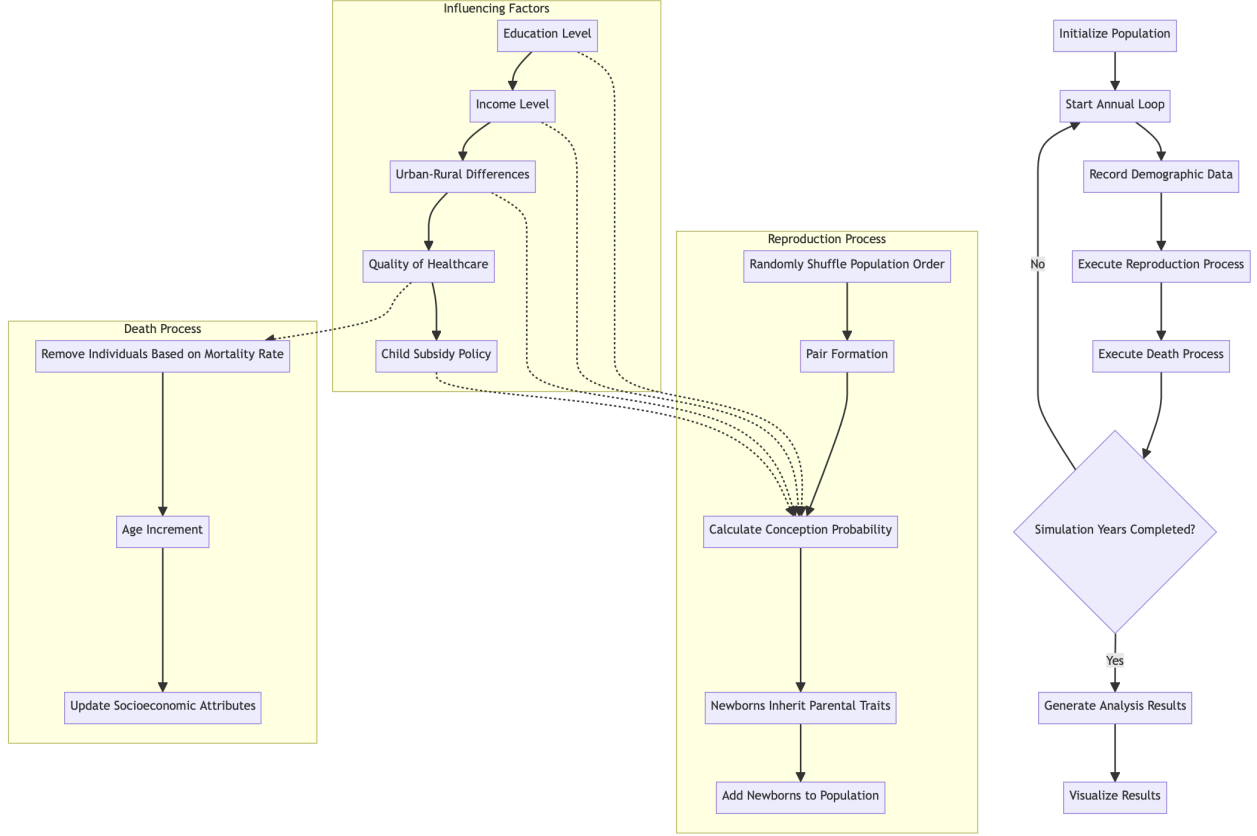


Figure 1: Population simulation workflow diagram

The simulation begins with an initialization phase, creating an initial population with a specified urban or rural ratio. The age distribution of the initial population is based on a normal distribution, ensuring a reasonable demographic structure at the start. Each individual is initialized with attributes including age, sex, education level, income, and urban or rural status.

The core of the simulation is the annual loop. Each year, the model performs the following:

- Records current population statistics such as total size, urban-rural composition, age structure, and dependency ratio.
- Executes reproduction: individuals form partnerships, calculate conception probabilities, and generate newborns who inherit parental traits.

- Applies the death process: each individual has a calculated mortality probability and may be removed from the population.
- Ages surviving individuals by one year and updates their socioeconomic attributes based on life stage.

At the end of the simulation period, the model enters the results generation phase. It compiles population statistics over time and creates visualizations for key indicators such as population pyramids, fertility-age distributions, and changes in dependency ratio. These results enable comparison across different policy scenarios to highlight the relative effectiveness of interventions.

3.2 Conceptual Model of Policy Parameters and Impacts

The simulation includes three core policy levers that can be adjusted to explore different scenarios. Three main parameters `child_support`, `education_impact`, and `healthcare_quality` are included in the simulation to represent policy levers that influence fertility and mortality dynamics. Each one affects individual decisions in different ways and, over time, shapes broader demographic outcomes.

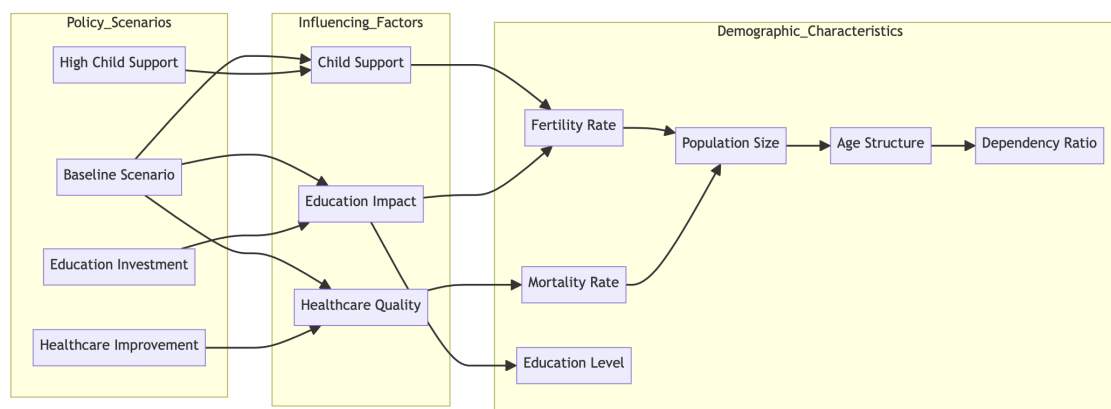


Figure 2: Policy parameters and impacts diagram

Child support reflects financial assistance provided to encourage childbirth. When this value is high, it becomes more likely that families especially those with existing children will choose to have another child. Education affects fertility as well. Higher levels of education are linked to delayed childbirth and smaller family sizes. In the model, stronger education influence makes individuals less likely to reproduce early. Healthcare quality plays a role in survival outcomes. Improvements in healthcare reduce mortality across all age groups, with a stronger effect on infants and older adults.

These policy effects are not isolated. They interact in ways that aren't always straightforward. For example, better healthcare may lead to a larger elderly population, increasing the dependency ratio. At the same time, improved education may temporarily lower fertility but raise productivity in the long run. These dynamics show why population planning

benefits from a system-level approach, where the combined effect of multiple changes can be observed and tested over time.

4 Simulation Model

4.1 Simulation Summary

The simulation model is implemented in Python and follows a modular architecture that separates computational logic from the user interface. The core simulation rules and agent behaviors are encapsulated in the `simulation_core.py` module, which handles all aspects of population dynamics including reproduction, mortality, state updates, and time-dependent events. A second module, `streamlit_app.py`, provides an interactive interface built using the Streamlit library. This interface enables real-time adjustment of simulation parameters and instant visualization of results.

Each simulation run proceeds in annual time steps. During each cycle, the system first initializes agents with attributes such as age, sex, income, education level, and whether they reside in urban or rural areas. It then checks for the occurrence of external events, such as a baby boom or an immigration wave, based on the simulation year and configured policy triggers. Fertility and mortality probabilities are computed dynamically using both global parameter settings and individual agent traits. Following this, the state of each agent is updated: survivors age, their education and income levels evolve, and newborns inherit attributes from their parents. At the end of each simulated year, key statistics including total population, age distribution, dependency ratios, and educational attainment are recorded for analysis.

The Streamlit interface allows users to customize every aspect of the model before and during simulation runs. Users can configure the initial population size and the ratio of urban to rural residents. They can adjust policy levers such as healthcare quality, child support enforcement, and the influence of education on fertility decisions. The interface also provides slider-based controls to define how the effects of education vary across years, and includes options to schedule specific policy scenarios or structural changes in the population.

Simulation outputs are displayed as dynamic visualizations and data tables that update as the model runs. These include charts showing population growth, age pyramids, fertility and mortality trends, and changes in the dependency ratio. The integrated nature of the system enables users to immediately observe how different policy combinations shape long-term population dynamics.

4.2 Verification of Implementation

To ensure that our implementation accurately reflects the conceptual model, we conducted a multi-faceted verification process. The following table summarizes the key verification techniques used:

Verification Method	Description
Code Walkthroughs	Compared core logic against the conceptual equations (e.g., bathtub mortality curve, fertility functions) to ensure correct translation into code.
Modular Testing	Independently tested fertility, mortality, income, and education update functions using controlled inputs and compared against expected outcomes.
Trace Analysis	Tracked step-by-step simulation logs (e.g., partner matching, births, deaths) to verify that execution order and data flows match the annual cycle.
Simplified Test Cases	Ran simplified simulations with small agent populations (e.g., 10 agents, fixed parameters) to allow for manual step-by-step validation of behavior.
Parameter Binding Checks	Ensured that all inputs provided via the Streamlit UI (sliders, toggles, number fields) are correctly parsed and injected into the simulation logic.

Table 1: Verification procedures to ensure implementation fidelity with conceptual model

5 Experimental Results and Validation

5.1 Simulation Study and Experimental Procedure

Step 1: Initialization (2000 Baseline)

The simulation was initialized using demographic data from the year 2000, retrieved from Data Commons [9]. Table summarizes the baseline setup and agent attributes used in the simulation.

Category	Details
Source	Data Commons [9]
Population Metrics	Total population count Age and gender distribution
Agent Attributes	Age Sex Partnership status Income level Residence type (urban, rural)

Table 2: Baseline Demographic Setup and Agent Initialization (2000)

Step 2: Demographic Inputs and Policy Parameters

Empirical demographic data were used to configure fertility, mortality, and migration behavior. These parameters were obtained from reliable state and institutional sources and are presented in table below.

Parameter Type	Metric	Value
Fertility Rates [10]	Women aged 20–24	112.4 births per 1,000
	Women aged 25–29	113.6 births per 1,000
	Women aged 30–34	95.2 births per 1,000
Mortality Rates [11]	Crude death rate (2000)	8.8 per 1,000
	Age 0–4	0.3 per 1,000
	Age 45–54	5.2 per 1,000
	Age 75+	42.5 per 1,000
Migration Rates [12]	Annual population growth	2.4%
	Portion due to net in-migration	65–70%

Table 3: Demographic Inputs for Simulation Configuration

Policy and Socioeconomic Parameters Education level was calibrated based on the Axios Raleigh report [13], which states that approximately 54% of residents in the Raleigh metro area hold a bachelor’s degree or higher. Accordingly, we set `education_level = 0.54` in the model. This high level of educational attainment is associated with delayed

childbirth, higher lifetime earnings, and improved health outcomes, and is thus reflected in agent behavior and long-term trajectories.

Child support policies were derived from guidelines published by Arnold & Smith Law [14], indicating that up to 40% of a noncustodial parent’s disposable income can be garnished. To reflect this, we set `child_support` = 0.4. This policy setting tends to increase the likelihood of childbirth among lower-income families by shifting financial incentives.

Healthcare quality was informed by the services and regional reputation of UNC REX Healthcare [15]. Given the hospital’s high standards of care and accessibility, we initialized `healthcare_quality` = 0.8. This parameter reduces mortality risk and plays a central role in determining maternal and infant health outcomes in the simulation.

Step 3: Simulation Logic and Execution

The simulation progresses in discrete annual time steps. In each simulated year, agents age and may undergo life events such as death, childbirth, or partnership formation. Birth and death probabilities are not static; they are dynamically computed based on the agents’ attributes and current policy parameters. Net migration is also introduced annually, calibrated according to regional estimates of population inflow. Additionally, education plays a key role by influencing the timing of childbirth and enabling income mobility, thereby contributing to longer-term demographic shifts.

To evaluate the robustness of the model, we conducted 1,000 stochastic simulation runs, introducing a $\pm 5\%$ variation in both fertility and migration rates. This allowed us to assess the sensitivity of the model to key input parameters and confirm the stability of aggregate outcomes.

5.2 Validation and Output Analysis

The simulator’s predictions for the Raleigh population in 2010 were highly accurate. The mean projected population was 405,525, which precisely matched the official figure reported by the U.S. Census Bureau. Across the 1,000 stochastic trials, the 95% confidence interval ranged from 403,200 to 407,800, with a standard deviation of approximately 1,080. These results provide strong evidence that, given high-quality input data, the simulation framework is capable of producing reliable and consistent long-term population forecasts.

5.3 Limitations and Sources of Error

While the model demonstrates robust performance, several limitations remain. First, there is inherent parameter uncertainty: many demographic inputs are derived from state- or metro-level data, lacking finer-grained resolution at the neighborhood or household level. Second, the treatment of migration is oversimplified; it is modeled as a net annual inflow without differentiation by age, educational attainment, or occupational category.

Third, several important socioeconomic feedback mechanisms are either highly simplified or omitted entirely. For example, the dynamic interplay between education, employment, housing, and fertility—though critical to real-world demographic trends—is not explicitly captured in the current version. Fourth, the model relies on probabilistic transitions and

does not incorporate cultural or behavioral nuance, which may influence individual decision-making in complex ways. Lastly, extraordinary events such as economic recessions, pandemics, or abrupt policy shifts are not accounted for, despite their potential to significantly reshape population dynamics.

6 Discussion, Conclusions, and Summary

6.1 Key Findings and Insights

This simulation study reveals several key insights about how individual-level decisions, when aggregated over time, generate macro-level population dynamics. One of the most notable observations is the phenomenon of micro-to-macro emergence: individual behaviors such as childbirth and mortality shaped by education, income, and policy gradually accumulate to form large-scale demographic trends. This emphasizes the importance of modeling agent-level heterogeneity and decision-making processes with sufficient granularity.

Another critical finding involves the nonlinear nature of policy interactions. Adjusting parameters such as education investment, healthcare quality, or child support enforcement often leads to complex, non-additive outcomes. For instance, improving healthcare access may increase life expectancy and dependency ratios, whereas enhancing education tends to reduce fertility but improves long-term economic productivity. These interactions suggest that policymakers must consider cross-effects when designing integrated interventions.

The simulation also highlights important differences in policy response between urban and rural populations. Urban agents, for example, exhibited a stronger reduction in fertility in response to educational attainment, whereas rural agents were less affected. This indicates that spatial heterogeneity must be taken into account when developing targeted policy measures.

Finally, the system demonstrates high sensitivity to key parameters. Small changes in fertility rates, education effects, or migration flows can significantly alter population trajectories. This sensitivity underscores the need for accurate parameter calibration and robust empirical grounding to ensure reliable forecasting.

6.2 Recommendations for Future Work

To expand the realism and applicability of the current simulation framework, several avenues for future development are proposed.

First, introducing a more detailed representation of lifecycle stages such as marriage formation, retirement decisions, and child education would enrich the model’s capacity to reflect full demographic transitions.

Second, incorporating outbound and interregional migration would allow the simulation to account for population mobility across cities or rural-urban zones. This would provide a more accurate depiction of urbanization patterns and regional population flows. Future versions of the simulator would benefit from calibration against real-world census data. By fitting the model to historical population data at the city or national level, it would be possible to improve forecasting accuracy and validate underlying assumptions.

Additionally, the model could be enhanced by integrating broader socioeconomic subsystems such as housing markets, employment trends, and healthcare access. These additions would allow the simulator to capture feedback loops between demographic and structural factors.

Modeling uncertainty and external shocks is another key direction. Introducing stochastic events such as economic recessions or pandemics and providing tools for uncertainty

quantification would improve the robustness of simulation results.

Lastly, the ability to run comparative policy scenarios and optimize interventions under constraints such as fiscal limits or outcome targets would transform the simulator into a decision-support tool for strategic planning. Supporting batch simulations and scenario evaluation would enable users to systematically explore trade-offs and policy outcomes in a controlled environment.

References

- [1] J. P. A. Ioannidis, “The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses,” *International Journal of Epidemiology*, vol. 46, no. 1, pp. 4–11, 2017. [Online]. Available: <https://doi.org/10.1093/ije/dyw328>
- [2] T. R. Malthus, *An Essay on the Principle of Population*. London: J. Johnson, 1798.
- [3] P.-F. Verhulst, “Recherches mathématiques sur la loi d’accroissement de la population,” *Correspondance Mathématique et Physique*, vol. 10, pp. 113–121, 1838.
- [4] P. H. Leslie, “On the use of matrices in certain population mathematics,” *Biometrika*, vol. 33, pp. 183–212, 1945.
- [5] P. D. M. Maciejewski, B. Wang, J. M. Mockenhaupt, and G. S. H. Yeo, “Title of the Article,” *CPT: Pharmacometrics & Systems Pharmacology*, vol. X, no. Y, pp. Z, 2013. [Online]. Available: <https://doi.org/10.1038/psp.2013.71>
- [6] R. J. Beckman, K. A. Baggerly, and M. D. McKay, “Creating Synthetic Baseline Populations,” *Transportation Research Part A: Policy and Practice*, vol. 30, no. 6, pp. 415–429, 1996.
- [7] Data USA. [Online]. Available: <https://datausa.io/profile/geo/atlanta-ga>
- [8] U.S. Census Bureau. Atlanta Population Data. [Online]. Available: <https://data.census.gov/all?g=160XX00US1304000>
- [9] Data Commons. Raleigh, NC Population Data. <https://datacommons.org/place/geoId/3755000>
- [10] NC Department of Health and Human Services. NC Maternal and Infant Health Data Dashboard. <https://www.dph.ncdhhs.gov/programs/title-v-maternal-and-child-health-block-grant/nc-maternal-and-infant-health-data-dashboard/fertility-rates>
- [11] NC State Center for Health Statistics. Vital Statistics. <https://schs.dph.ncdhhs.gov/data/vital.cfm>
- [12] Advan Research. Raleigh, North Carolina Location Data. <https://advanresearch.com/raleigh-north-carolina-location-data>
- [13] Axios Raleigh. Raleigh-Durham Education Levels Ranked. <https://www.axios.com/local/raleigh/2023/10/25/raleigh-durham-education-levels-ranked-duke-unc-ncstate>
- [14] Arnold & Smith Law. Child Support Guidelines. <https://www.arnoldsmithlaw.com/how-much-of-a-paycheck-can-be-garnished-for-child-support.html>
- [15] UNC REX Healthcare. About REX. <https://www.rexhealth.com/rh/>