Social Security Sustainability

Rohan Thavarajah, Courtney Cochrane, and Sarah Mohamed

December 14, 2016

1 Background and Motivation

There are two ways a government might structure a public pension scheme. One option would be to assign every individual a retirement account. As they earn during their working years, a portion of a person's income is mandatorily siphoned into their account to support them when they retire. However, such a scheme will yield larger payments to wealthier individuals who are precisely the least likely to depend on government support when they retire. Furthermore, retirees at the time of institution of such a scheme would have empty accounts; it would take several years before a generation is encountered that had accumulated non-trivial balances.

For both of these reasons, when the United States kicked off its Social Security scheme in 1935, legislators structured it so that **current** workers directly support **existing** retirees. Rather than maintain an account for each individual, the Social Security portion of one's taxes are first used to meet the government's payment obligations to existing retirees and any remainder is channeled into a collective trust.

Current workers funding existing retirees made a lot of sense between 1946-1964. In these years the US was experiencing a post-World War II baby boom. Compositionally, with many more workers than retirees, social security receipts were enough to meet payments and even yield a surplus to swell the trust. However, like many developed nations, birth rates in the US have plunged since then and stagnated. As the large cohort of baby boomers begin to retire they are supported by a compositionally smaller working population. The Social Security Administration expects that social security payments will very soon outstrip receipts and that the trust will be depleted in the next 20-30 years.

In this paper, we seek to estimate years till the social security trust depletes. We believe quantifying depletion timing precisely is valuable for two reasons. First, economic behavior becomes aberrant when a system is expected to grind to a halt. For instance, expectations that a bank is on the cusp of becoming insolvent leads to bank runs (which perversely expedite insolvency). Supposing social security is indeed slated to deplete in 30 years, social security depletion is not a problem for 30 years from now. When the precipice approaches and a generation becomes more and more certain that a fund to which they contribute will not be one from which they benefit, political pressure will mount.

Second, by quantifying depletion timing we are afforded a rubric with which to score alternative strategies for addressing unsustainability. In this paper we focus on two strategies, raising the retirement age and changing the proportion of taxes directed to social security.

2 Approach

We formulate the problem of estimating social security sustainability as consisting of two steps that echo the structure of this paper. In step one, we generate population projections by age group. We consider three projection techniques: (1) the Leslie matrix which harnesses matrix algebra (2) a compartmental model founded on a system of ODEs and (3) the Gurtin-MacCamy age-structured population model which utilizes a hyperbolic PDE.

In step two, we use these population projections to estimate time till trust depletion under a variety of scenarios including raising the retirement age to 67 and changing the proportion of income directed to social security.

3 Part I: Population Projection

We apply three numerical techniques in increasing order of sophistication. We look for the following two features in age-structured population models to facilitate sustainability analysis:

1. Fertility and mortality rates should be allowed to differ by age group.
2. Projections should occur in annual time steps (in preference to say 10-year time steps).

The Leslie is appealing both for its simplicity and for the fact that fertility rates and mortality rates are allowed to differ by age group. However, our data is chunked into 10-year age groups necessitating that projections occur at 10-year intervals. We deem this step size too coarse.

Compartmental models, wherein we study dynamics between child, worker and retiree groupings, yield projections with suitably fine time steps. However, fertility rates are considered uniform across the entire working population. As will be demonstrated, this assumption is at odds with the data.

The Gurtin-MacCamy model, which advects the initial population distribution with attrition due to mortality and growth due to birth, yields the best of both the Leslie and compartmental model. We are able to allow rates to differ by age group and yield projections in annual time steps. Furthermore, we yield population projections in annual age increments as opposed to less malleable child/worker/retiree groups. This allows us to experiment with retirement age and its effect on sustainability.

3.1 Leslie Matrix

The Leslie matrix is an age-structured population growth model often used in ecology to model changes in animal populations. The Leslie matrix encodes information from the fertility and mortality rates within each class of the population. The sizes of the age classes at successive time steps are constructed by iteratively applying multiplication of the Leslie matrix to an age-structured vector of the current population [2].

At time t, the age-structured distribution of the population may be represented by the vector X^t, and the Leslie matrix by L :

$$X^t = \begin{bmatrix} x_1^t \\ x_2^t \\ x_3^t \\ \vdots \\ x_n^t \end{bmatrix} \qquad L = \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \dots & \alpha_n \\ \mu_1 & 0 & 0 & \dots & 0 \\ 0 & \mu_2 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \dots & \vdots \\ 0 & \dots & 0 & \mu_{n-1} & 0 \end{bmatrix}$$

In this scheme, integers $1 \dots n$ denote the age classes, α denote per-female birth rates and μ denote survival rates within each class.

The long-term stability of the population is governed by the dominant eigenvalue of the Leslie matrix. It can be shown that, for dominant eigenvalue λ, the convergence of the population follows [2]:

- $0 < \lambda < 1$: the population collapses
- $\lambda = 1$: the population stabilizes
- $\lambda > 1$: the population increases boundlessly

3.2 Compartmental Models

In compartmental models, we find another framework describing the changes among class-structured groups in a population. Often used in epidemiology to model the transmission of disease, these models describe the transfer of units among a set of compartments within a system [3]. This strategy may be extended to population dynamics in order to describe the movement of individuals among age classes within the population.

Gonzalez-Parra and Arenas[4] describe a system of differential equations in order to model the effect of increasing Social Security's retirement age on the populations of workers and retirees. They divide the population among three discrete populations (children, workers, and retirees) and model the movement of the population from group to group in accordance with population-wide birth and death rates.

$$\begin{aligned} C'(t) &= bP(t) - \beta C(t) \\ W'(t) &= \beta C(t) - \gamma W(t) \\ R'(t) &= \gamma W(t) - P(t)\mu \end{aligned}$$

Gonzalez-Parra and Arenas provide a compartmental diagram to demonstrate the internal and external population traffic in the three groups (Figure 1)).In this system, C, W, and R denote children, workers, and retirees, respectively. The constant b represents the rate of inflow to the population, μ represents the death rate among retirees, β is proportionally inverse to the mean of the starting work age, and γ is proportionally inverse to the mean of the worker age range.

In simulation and analysis, Gonzalez-Parra and Arenas use a scaled version of this model in order to simplify the system. In our version of this model, we modify the unscaled ordinary differential system above, by allowing each class to be affected by an associated death rate. μ_C, μ_W, and μ_R denote the death rates for each respective age class below:

Figure 1: The Gonzalez-Para and Arenas compartmental model

$$
\begin{aligned}
C'(t) &= bP(t) - (\beta + \mu_C)C(t) \\
W'(t) &= \beta C(t) - (\gamma + \mu_W)W(t) \\
R'(t) &= \gamma W(t) - \mu_R R(t)
\end{aligned}
$$

3.3 Gurtin-MacCamy

3.3.1 Overview

Adopting notation from Abia et al [1] to describe the model originally outlined by Gurtin and MacCamy [5], we wish to solve the initial boundary value problem for the following hyperbolic PDE:

Let $u(x,t)$ represent age specific density of individuals of age x at time t. Let A and T represent the upper bounds on age and time respectively. Finally let β and and μ represent fertility and mortality rates respectively both of which are expressed as functions of age, time and the density function I. Then we have:

$$u_t + u_x = -\mu(x, I(t))u, \text{ for } 0 \leq x \leq A, 0 \leq t \leq T \tag{1}$$

where

$$I(t) = \int_0^A u(x,t)dt$$

The initial population density by age group at time 0 is given by:

$$u(x,0) = \phi(x), \text{ for } 0 \leq x \leq A \tag{2}$$

Similarly, newborn density at time t is given by:

$$u(0,t) = \int_0^A \beta(x, I(t))u(x,t)dx, \text{ for } 0 \leq t \leq T \tag{3}$$

We solve this problem using the following nonlinear upwind scheme simplified from Abia et al. After discretizing age-time space into a J x N grid, let U_j^n represent our estimate of the age specific density, $u(x_j, t_n)$. Then the explicit scheme for Equations 1-3 are given by Equations 4-6 respectively:

$$\frac{U_j^{n+1} - U_j^n}{\Delta t} + \frac{U_j^n - U_{j-1}^n}{\Delta x} = -\mu(x_j, I^h(\mathbf{U}^n))U_j^n \tag{4}$$

Where, in line with Abia et al., $I^h(\mathbf{U}^n)$ is our composite rectangle quadrature estimate of $I(t_n)$,

$$I^h(\mathbf{U}^n) = \sum_{j=1}^{J} hU_j^n$$

The first column of our grid describes the initial population density by age group:

$$\mathbf{U}^0 = (U_1^0, U_2^0, \ldots, U_J^0) \tag{5}$$

Similarly, newborn density at time t is estimated by:

$$U_0^{n+1} = \sum_{j=1}^{J} h\beta(x_j, I^h(\mathbf{U}^{n+1}))U_j^{n+1}, \text{ for } 0 \leq n \leq N-1 \tag{6}$$

In order to recover population levels from densities, we employ a multi-dimensional composite rectangle quadrature rule.

3.3.2 Error Analysis

We test our implementation by applying it to a problem with known analytical solution. The problem is from Abia et al. who choose mortality and birth rates defined as follows:

$$\mu(x,z,t) = z \tag{7}$$

$$\beta(x,z,t) = \frac{4xze^{-x}(2-2e^{-A}+e^{-t})^2}{(1+z)^2(1-e^{-A})(1-(1+2A)e^{-2A})(1-e^{-A}+e^{-t})} \tag{8}$$

Let the initial population density by age be given by:

$$u_0(x) = \frac{e^{-x}}{2-e^{-A}}$$

Then the analytical solution is

$$u(x,t) = \frac{e^{-x}}{1-e^{-A}+e^{-t}}$$

From Figure 2, our implementation of the nonlinear upwind scheme appears to yield a good approximation.

We also experiment with step sizes to affirm the claim made in Abia et al. that the scheme is first order accurate in time and age. As an example, in Table 1 we are interested in the relative error associated with population density projections ten years out. We find doubling age step size increases error by a factor of four; tripling age step size increases error by a factor of 9. The scheme appears to behave as though it is second-order accurate in age, out-performing expectations. Conversely, error appears to be unchanged with respect to choice of reasonable time-step. Given the scheme converges on the analytical solution, we accept it and move on to real data, however we flag this inconsistency in convergence properties.

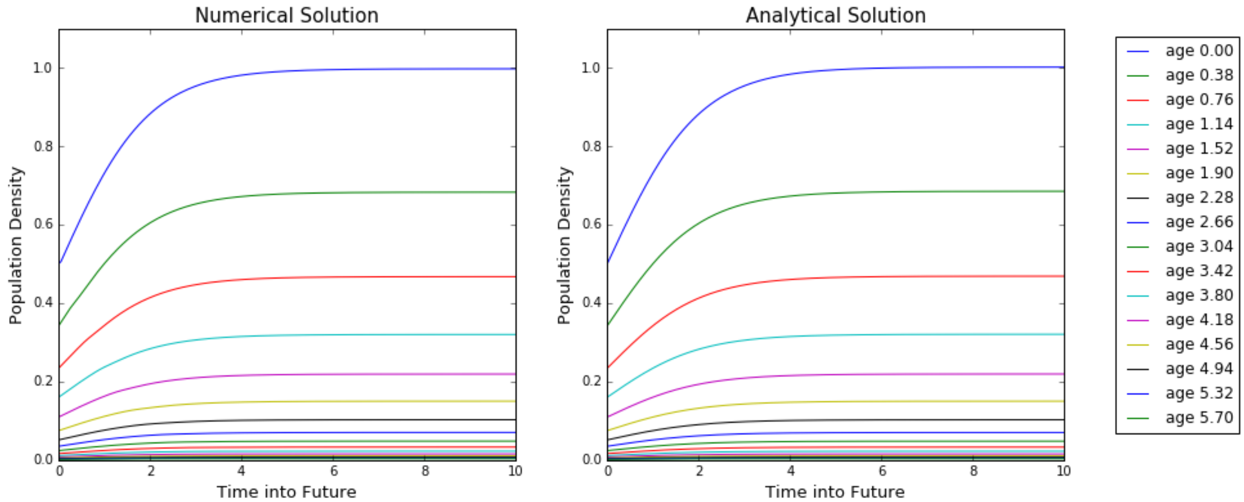


Figure 2: Cross-sections of Population Density by Age - Numerical vs Analytical Solution

Table 1: Error analysis on population projection ten years out

t_steps	a_steps	Relative Error
200	20	0.040
200	40	0.009
200	80	0.002
400	20	0.040
400	40	0.009
400	80	0.002
800	20	0.040
800	40	0.009
800	80	0.002

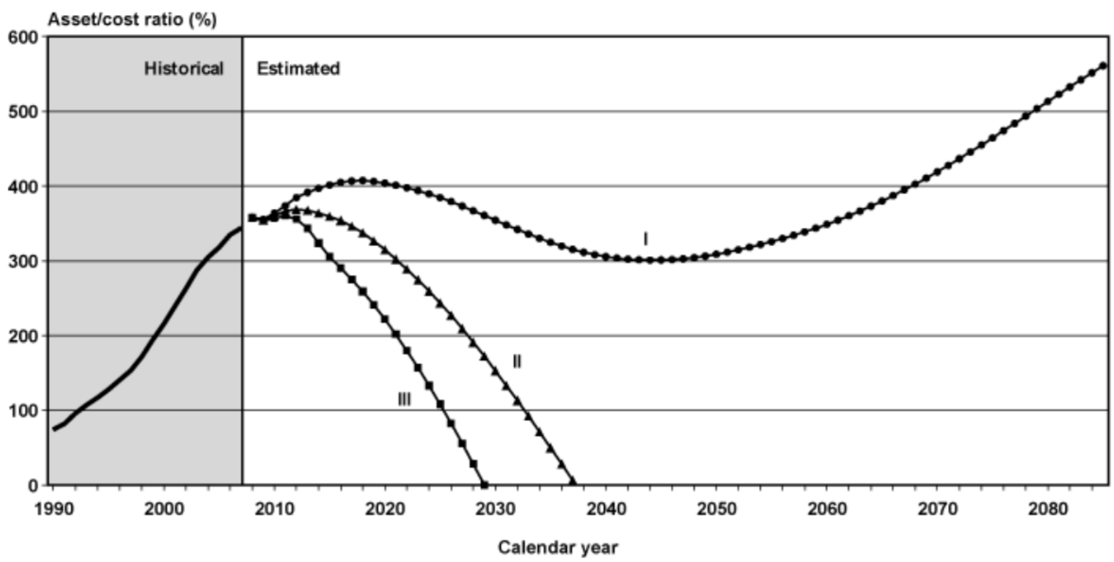
4 Part II: OASI Fund Estimation

4.1 Literature Review

There is an abundance of literature concerning the estimation and prediction of fund depletion. Throughout our research we specifically used the Social Security Administration (SSA) Office of Retirement and Disability Policy's bulletins [6] as well as a paper by the Organization for Economic Co-Operation and Development (OECD) [7].The OECD paper provides a solid background into simulating age-populations and pension systems in 20 countries including the U.S. They consider population projections out until 2070 and predict the government primary balance. Of all the countries considered, they predict that that United States and United Kingdom will have the smallest decrease in government primary balance over the period from 2000-2030.

Even through the OECD's predictions are favorable compared to other countries, the SSA makes the projection that the fund will be exhausted in 2037. They consider different cost assumptions and plot the predicted fund depletion under them (see Figure 3). We strive to produce fund estimates with our own models that mirror these predicted trends. Based on their simulations, they suggest that an immediate increase in the tax rate to 14.4% or increasing the retirement age (they consider an increase from 65 to 67), would mitigate the predicted fund collapse.

Chart 1.
Combined OASI and DI Trust Fund assets as a percentage of program cost, 1990–2008, projected under alternative assumptions, 2009–2085



SOURCE: *2009 Social Security Trustees Report,* Figure II.D6 and Table IV.B3.
NOTES: Alternative I = low-cost assumptions; alternative II = intermediate assumptions; alternative III = high-cost assumptions.

Figure 3: SSA's Estimates for Depletion of Fund

4.2 Our Models

In our work we apply our population estimate results to the task of predicting social security fund depletion and exploring how government interventions could mitigate this depletion. As a baseline we use the actual population data collected from the U.S. Census Bureau for 1990-2015. All our collected data contains information, on population, expenditures, etc., from 1990-2015; thus we needed to come up with an extrapolation scheme. We decided that it was logical to predict a future year's, Y, data by creating a line of best fit using the previous two years of data and then evaluating the line at Y. We performed this extrapolation scheme for any years where we did not have "true" data for expenditures and incomes. In addition, for our baseline model we also used the extrapolation scheme for the retiree and worker populations. The OASI tax rate baseline was defined to be 10.52%, although we varied it later on.

We create a script to simulate and plot the changes in the social security fund iteratively over a given number of years. The amount of money in the OASI fund is calculated by this recursive formula:

$$Fund\_Amount(t+1) = Fund\_Amount(t) + rIW - ER$$

where r = OASI tax rate, I = average income for working population, W = worker population, E = average expenditure per retiree, and R = retiree population.

5 Data

5.1 Sources

Our project required the collection of data on different topics: population, death rate, birth rate, median income, average social security benefit, and social security fund amount. We collected the population and income data from the U.S. Census Bureau. [1] The average social security fund amount and expenditures were collected from the Social Security Administration's website. [2] Finally the birth rates were collected from the Child Trends organization [3] and the death rate data from the CDC. [4] These statistics were collected for each year from 1990 through 2015 and for each age bracket we considered within the respective year.

5.2 Assumptions

Each row in our input data is at the level of year and age group. However, some methods (in particular the PDE approach) demand that we express the initial population density by age, fertility and mortality rates as continuous functions. We must therefore make two kinds of assumption. Between existing data points (the interpolation region), we must produce a smooth function that interpolates our data. Beyond these data points (the extrapolation region), we must ensure that our assumptions are realistic. An underlying theme of our assumptions is that we wish to extrapolate as conservatively as possible.

1. **Initial population density by age group**

   We employ a cubic spline and further necessitate that density is non-negative and zero whenever age is less than 0 or greater than 100. The latter assumption ensures that the cubic spline is restricted to the interpolation region where its performance is optimal (i.e. that we extrapolate very little). Figure 4 illustrates one such spline for initial population density in 1990 alongside real data; another is employed in 2015.

2. **Fertility rates**

   The PDE approach demands we express fertility rates as a continuous function of age and time. To yield this function we employ scipy's radial basis interpolation function. We do two things to curtail extrapolation: (1) we necessitate that fertility rates be zero for ages outside the range 12-45 (2) for years greater than 2015, we assume fertility rates are constant and equal to the mean of their values in 2014 and 2015. Assumption (2) is particularly important because the rbf interpolant quickly degrades in shape after 2015.

[1] https://www.census.gov/popest/data/ for population data and http://www. census.gov/data/tables/time-series/demo/income-poverty/historical-income-people.html for income data

[2] https://www.ssa.gov/policy/docs/statcomps/supplement/index.html for expenditures and https://www.ssa.gov/OACT/STATS/table4a3.html for fund amount

[3] http://www.childtrends.org/indicators/fertility-and-birth-rates/

[4] https://wonder.cdc.gov

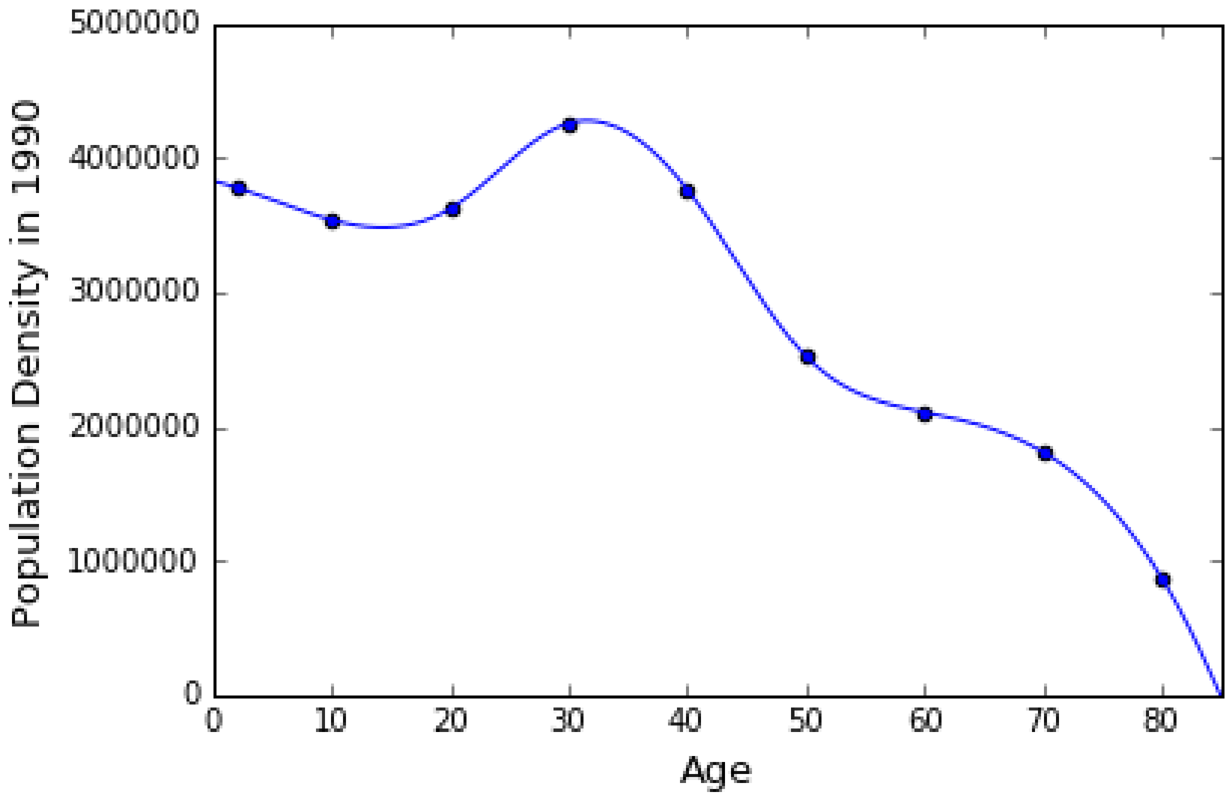


Figure 4: Initial Population Density Cubic Spline

In Figure 5 we plot the fertility function on the left and cross-sections by year with real data on the right. From the rightward shift of the mass of the fertility function, we observe that women in the US are having children later than they did in 1990. Furthermore, they are having fewer children demonstrated by the decline in peak fertility. It is easy to see why we are uncomfortable allowing a numerical method to extrapolate these trends. Fertility tails off after 45 by biological necessity so that the trend of aging motherhood will eventually plateau. Similarly, it is uncertain how peak fertility rates will evolve but assuming they will continue to decline indefinitely is unrealistic.

3. **Mortality rates**

   Likewise, the PDE approach demands we express mortality rates as a continuous function of age and time. We employ the same methods used for fertility, again assuming rates are constant after 2015. It is slightly surprising to us that mortality rates have barely budged since 1990 but this providing further justification for holding mortality constant after 2015.

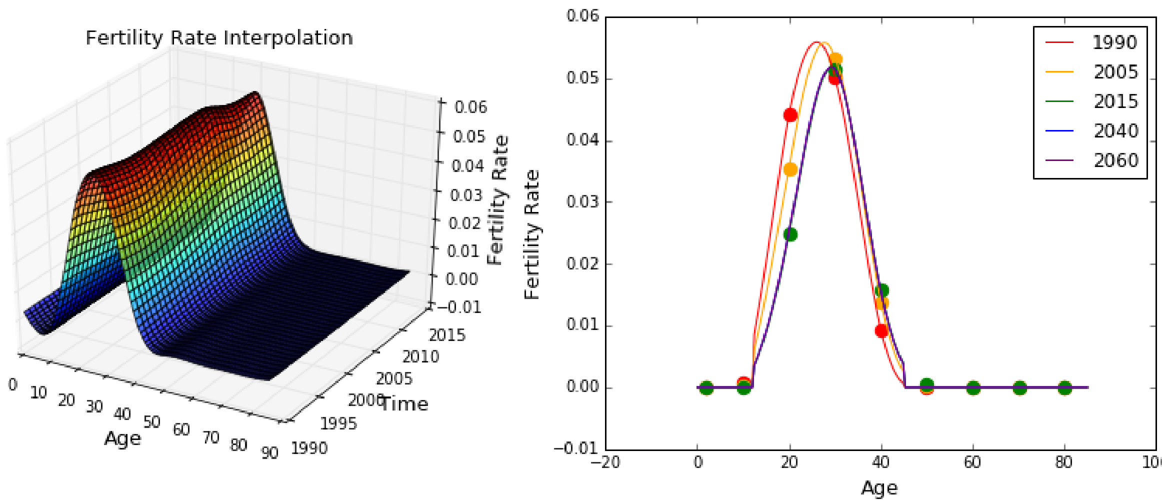


Figure 5: Fertility Rate Interpolation Using Cubic Radial Basis Function

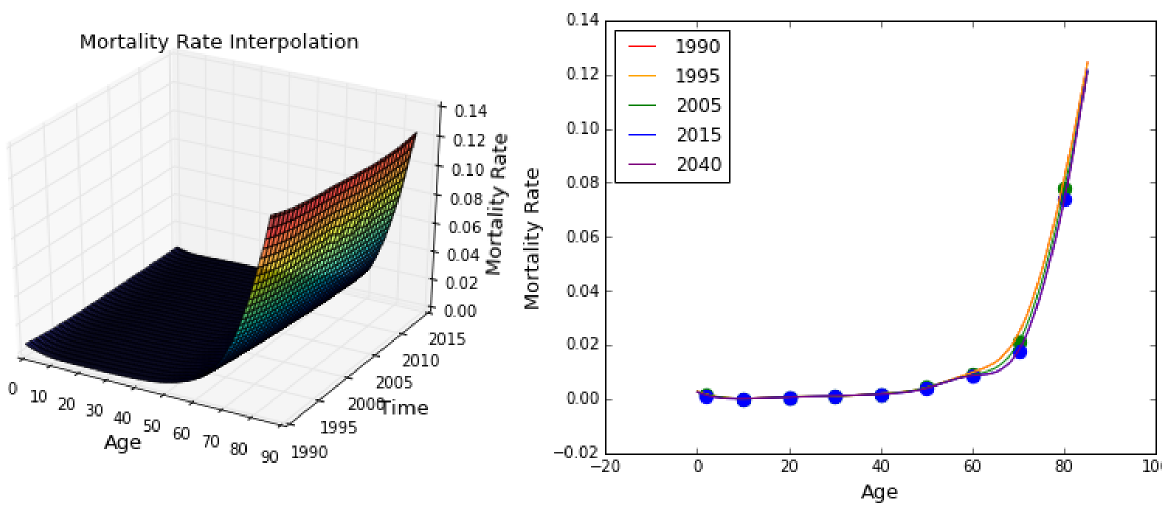


Figure 6: Mortality Rate Interpolation Using Cubic Radial Basis Function

6 Results

6.1 Numerical Schemes

We present population projections from two start years. We choose one year to be 1990 and retrospectively occlude population density data between 1991-2015. We then apply a projection strategy and compare it to true population density to evaluate our performance. With this error metric in mind, we then project from 2015 furnishing input for social security sustainability analysis.

6.1.1 Leslie Model

1. Evaluation - Projection from 1990
   The projected child, worker, and retiree populations beginning in 1990 are shown in Figure 7. While the retiree population changes in tune with the true population over this interval, there is a notable divergence in the projected child population.

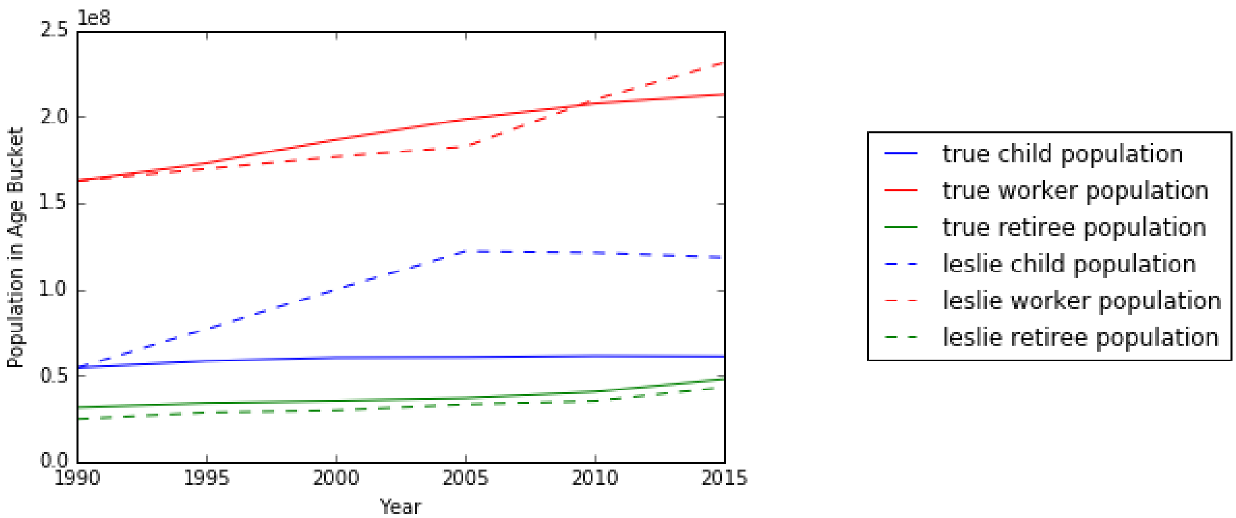


Figure 7: Leslie Matrix Population Projection 1990

   While this result seems reasonable in the short term, the trend of the Leslie matrix over time suggests boundless population increase among all age classes (Figure 8) over 25 time steps. In accordance with the age ranges, each time step spans 5 years.

2. Analysis - Projection from 2015
   Projections for the year 2015 follow a similar trend as those initialized in 1990; the birth, death, and population estimates from this year produce a very similar Leslie matrix. It therefore follows that the projections from 2015 onward also follow a boundless, exponential trend, and that confidence in the long-term projections is low.

   Reasons for this model's failure are twofold; first, the Leslie matrix is particularly structured to model populations of females rather than mixed-sex groups. We attempted a heuristic to ameliorate this effect by using a multiplicative factor that halves the population. Secondly, birth rates are measured on a per-woman scale, and the data provided is divided among buckets. We also apply a multiplicative factor here in order to offset the effect, but these two considerations offer one explanation for the unreliable result.

6.1.2 Compartmental Model

1. Evaluation - Projection from 1990
   The ODE approach produces population projections that much more closely follow the trend expected in the data. The largest divergence occurs in the worker population, which the ODE predicts should gradually increase over time at a higher rate.

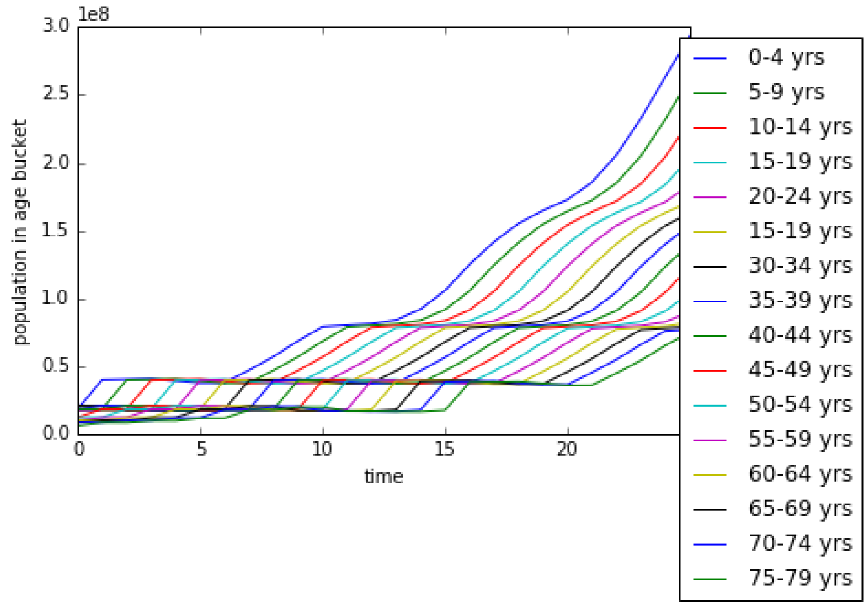


Figure 8: Leslie Matrix Long-term Population Projection

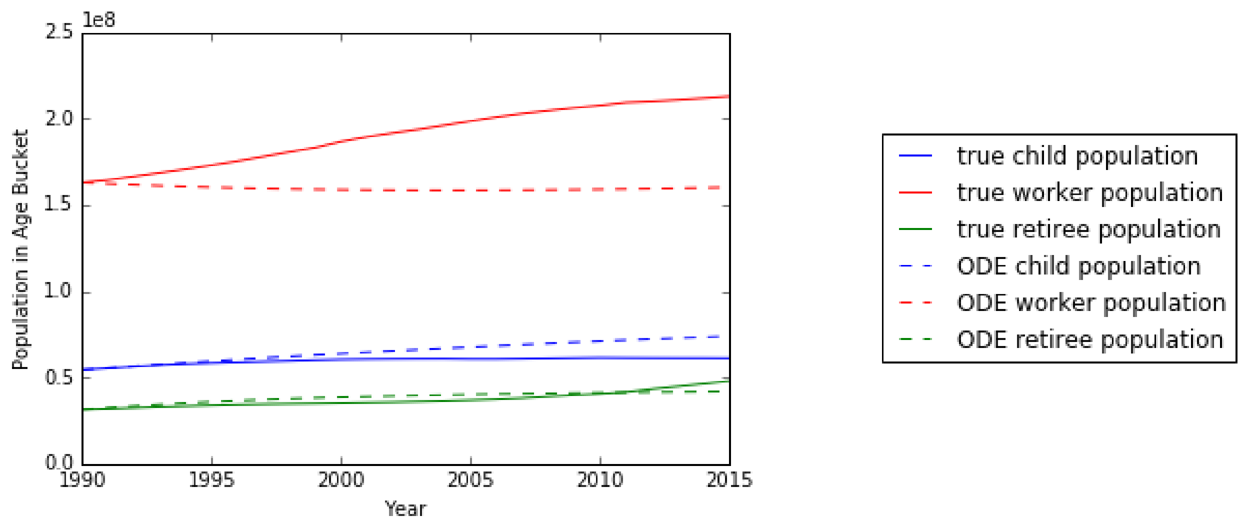


Figure 9: ODE Population Projection 1990

To interpret this result, we can examine the system of equations comprising the ODE. The retiree equation might be expected to experience an overall increase over time if the combination of the death rate and the rate of retirement is exceeded by the movement of children into the working group.

6.1.3 Gurtin MacCamy Model

1. Evaluation - Projection from 1990: As illustrated in Figure 7, we find that the PDE approach yields logical population projections. As expected, the further we predict into the future, the less accurate our estimates become.

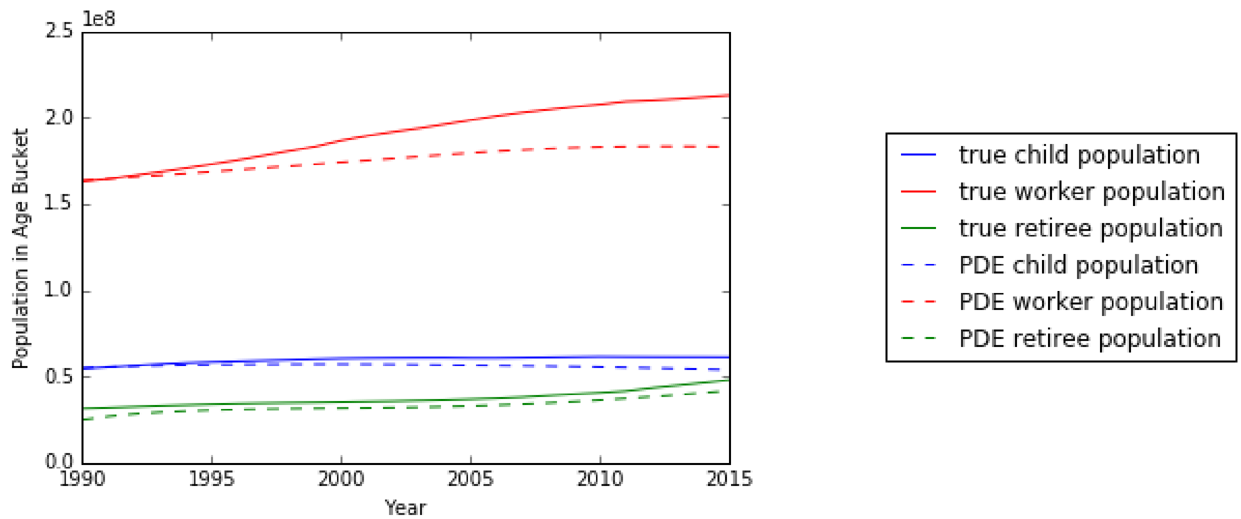


Figure 10: PDE Population Projection 1990

2. Analysis - Projection from 2015: In Figure 8, we plot population density projections from 2015 set beside projections from 1990 for comparison. The shapes are similar suggesting our population predictions from 2015 are reasonable. Note that the graph on the left allows us to visualize the advection of the peak of baby boomers into retirement.

6.2 OASI Fund

6.2.1 Fund Depletion Estimates by Model

We find that our baseline model, which uses rolling lines of best fit to predict population, predicts that the fund will run out in 2038 (see Figure 12), which is close to the Social Security Office's prediction of fund depletion in 2037 [6]. We used a tax rate of 10.52%, and a working age range of 15-64 with retirement starting at 65.

We then predicted fund depletion using the population predictions generated by the Leslie, ODE, and PDE methods. We started in 2015 and used the predicted data to generate future OASI fund amounts. The Leslie method had trouble predicting reasonable future populations as discussed above. The ODE method did a better job at estimating future population numbers, but fund simulations are inaccurately optimistic. It predicts that the fund declines for twenty years but then monotonically increases. This may be justified if for instance the social security trust is able to recover after the uncharacteristic pressure exerted by baby boomer retirement has eased. However, we find that worker population increases monotonically suggesting that fund behavior is in this case a vestige of the ODE's overestimation of the worker population. Due to our PDE approach producing the most reasonable population estimates, it is no surprise that it also best

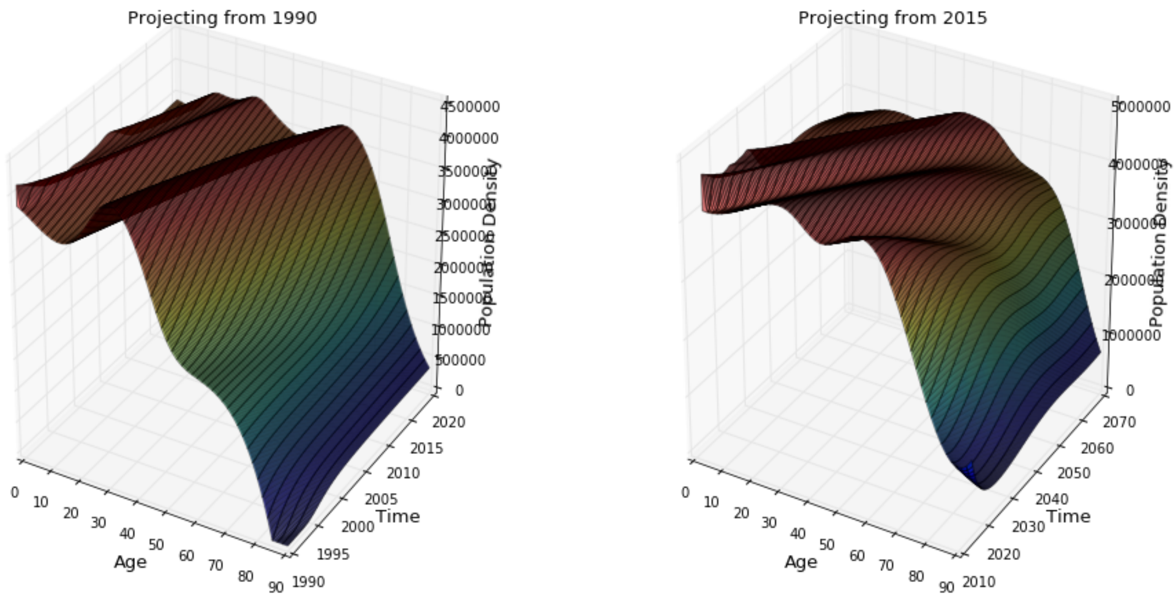


Figure 11: Population Density Projections

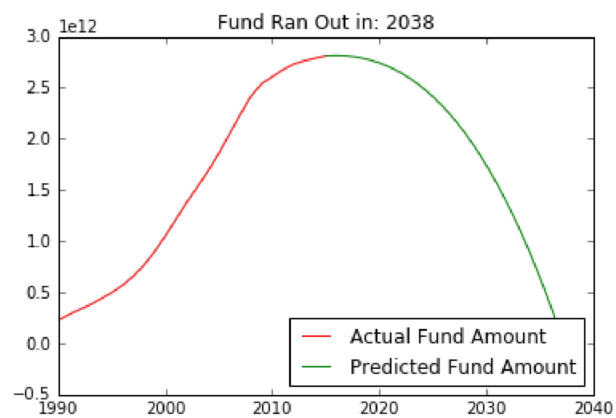


Figure 12: Expected Fund Depletion with True Data Extrapolated

predicted fund depletion. It predicts that the fund will run out in 2040, again very close to the Social Security Office's predicted depletion data (see Figure 13). For this simulation, we used a tax rate of 10.52%, and a working age range of 15-64 with retirement starting at 65.

6.2.2 Effect of Government Interventions on Depletion

We used the PDE model, our most accurate population projection model, to explore fund depletion with different government interventions. Specifically we considered the two most realistic government interventions: 1. altering the OASI tax rate and 2. increasing the retirement rate to 67. Here we briefly discuss our findings for three different cases:

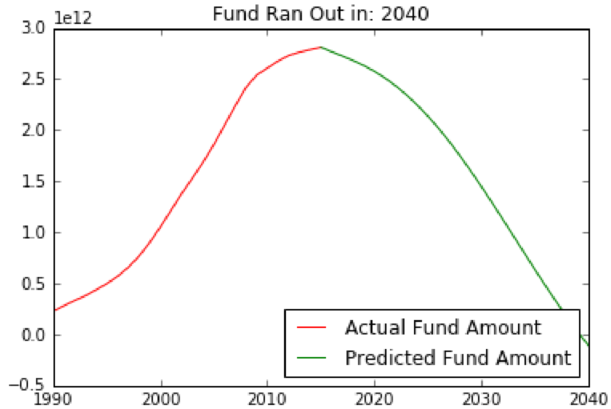


Figure 13: Expected Fund Depletion with PDE Data

1. Increasing the OASI tax rate while maintaining a retirement age of 65

   As expected, increasing the tax rate leads to a longer period until predicted fund depletion. Increasing the tax rate by just 3% leads to fund depletion being delayed by 20 more years. When the rate is increased by 5% we no longer see fund depletion within the next 50 years.

2. Increasing the retirement age to 67 while maintaining the same OASI tax rate

   Increasing the retirement age to 67 leads to the fund continuously increasing over the next 50 years. This shows the importance retirement age has on fund depletion.

3. Increasing the retirement age to 67 and changing the OASI tax rate

   The larger worker to retiree ratio due to the older retirement age (of 67) means that taxes devoted to social security could be lowered without incident. Interestingly enough, the combination of increasing the retirement age and decreasing the tax rate by 20% yields fund depletion in 2037, our business-as-usual predicted fund depletion date. While the government is unlikely to decrease the tax rate, we found that a decrease by 10% or less will not harm the fund's solvency for the next 50 years at least provided it is shielded by an increase in the retirement age.

We find that the PDE method's population predictions provide the best way to estimate fund depletion. We discover that small increases in the tax rate lead to much longer fund stability, and that increasing the retirement age from 65 to 67 will significantly alleviate aging population induced stress.

7 Conclusion

In the absence of government intervention, the our most promising model predicts that the OASI fund will run out in 2040. This differs from the Social Security Adminstration's estimate by only 3 years which implies: 1. our PDE method is estimating population well, as it gleans results that are comparable to the SSA's population estimates and 2. our crude method of extrapolating income, expenditures, etc. in order to predict fund depletion is fairly effective. In the case of both the ODE and PDE approaches (and superficially, the Leslie matrix approach), the retirement population is expected to increase gradually in the coming years relative to the worker population. For this reason, an increase in the retirement age to 67 is anticipated to defer the depletion of the OASI fund. Based on our simulations for fund depletion, our results confirm that both raising the retirement age to 67 (already scheduled to happen by 2022) or slightly increasing the tax rate will help preserve the fund's stability. We determine that raising the retirement age has an even larger effect on delaying fund depletion than small tax increases.

Possible future refinements to our models are two fold. Firstly on the population estimation front, we could create more complex models by doing away with the simplifying assumptions we made. Adding in immigration would make our predictions much more realistic. In addition, to tether our estimates to the interpolation region, we assume max age is 90, an assumption that is perhaps exceedingly pessimistic. Similarly, when interpolating initial ages, we place a data point for the 75+ age group at 80, which again pessimistically squeezes the population distribution at the right tail. Secondly, with fund modeling we could explore better methods for predicting future average expenditures and incomes that take into account inflation and other economic predictions.

References

[1] LM Abia, O Angulo, JC Lopez-Marcos, et al. Numerical solution of structured population models, 2004.

[2] Howard Anton. *Elementary linear algebra*. John Wiley & Sons, 2010.

[3] F Brauer, P van den Driessche, and J Wu. Compartmental models in epidemiology. *Mathematical Epidemiology*, 18:71, 2008.

[4] Gilberto Gonzalez-Parra and Abraham J Arenas. A mathematical model for social security systems with dynamical systems. *Ingeniería y ciencia*, 10(19):33–53, 2014.

[5] Morton E Gurtin and Richard C MacCamy. Non-linear age-dependent population dynamics. *Archive for Rational Mechanics and Analysis*, 54(3):281–300, 1974.

[6] Stephen C. Goss-Social Security Office of Retirement and Disability Policy. The future financial status of the social security program. *Social Security Bulletin*, 70(3), 2010.

[7] Deborah Roseveare, Willi Leibfritz, Douglas Fore, and Eckhard Wurzel. Ageing populations, pension systems and government budgets. 1996.