Henry Hodell
Professor Handel
Economics 191
11 May 2023

Can Agent-Based Economic Simulations Model Wealth Disparities?
How do Wealth Disparities Originate?

In my research paper I will be exploring how wealth disparities originate and evolve in economies using economic simulations. I will be exploring if it is possible to understand how these wealth inequalities arise in a quasi-realistic economic simulation or is it something that does not fit with traditional economic models, or in other words a market failure. I will also be exploring if it is possible to use economic simulation to understand how wealth inequalities change over time in order to determine whether wealthy families stay wealthy or is it entirely random? Finally, I will be comparing different policies and assessing how well they mitigate existing wealth inequalities.

In order to answer these questions, I will be using an individualized heterogeneous agent model in which wealth is modeled in different periods and not only will individuals be heterogeneous from one another but also to themselves in different periods. Essentially, agents will be housed under a family like structure in which agents interact and act heterogeneously and wealth changes based on numerous parameters such as altruism, charity, patience, etc. This model will also include events such as death, birth, inheritance, and intertemporal choice each of which will create time inconsistency and discreteness in time periods. Meaning that in each period each individual is checked to ensure that they have not exceeded the maximum number of periods to remain in the simulation and that they have a nonzero number of goods to use in the next period. If either of these conditions are not satisfied then the individual is removed and any goods left are equally distributed to offspring, if there are any. This model will also include interactions such as taxation, and charitable donations. This model will scale to form entire economic systems that can be tailored using data such as population growth, consumption rates, saving rates, and existing tax structures. This model will be used to analyze how pareto wealth distributions originate in different initial conditions such as equal wealth, random wealth, etc. This model will also be used to analyze how wealth distributions can evolve over time given inputs and or changes meaning that this model could be used to analyze policy changes and how they can impact relative wealth distributions.

The model will be built using Java and the benefits of Object-Oriented Programming in Java to create the underlying familial and economic systems that will be the basis for the heterogeneous agent model. This program will be able to handle scale into the millions of agents which will allow the model to have as realistic outcomes as possible and will attempt to mirror the smoothness of the continuous time heterogeneous agent models. Due to the implication of familial ties and locality connections, or local web of connections, the models will not follow a Brownian motion approach to generalize a large economic system but rather simulate direct actions and interactions between individuals. This will create similar results to the generalization approach but will have somewhat more noise due to the nature and randomness of the model, but the overall trends should be equivalent. The model’s randomness which creates its heterogeneity will be used in the instantiation of agents and in the actions the agents undertake. This will ensure enough heterogeneity to match the generalization approach and will allow the model to have familial structure that will be extremely beneficial in understanding wealth transfers and generational wealth distributions in discrete time.

Since the start of economic research economists have attempted to bridge the gap between micro and macroeconomics in an effort to provide a link from foundational economic theory and tangible data. More recently with the advent of personal computers and the massive amounts of computational ability available economists have sought to expand microeconomic theory into macroeconomic models, which previously had not been possible due to the limitations of paper and pencil. These expansions led to the construction of DSGE models in which dynamic stochasticity is used to represent a more realistic model of a macroeconomic model which can be tailored given certain inputs. DSGE models, while providing a fair amount of realism, do not account for long term adaptability such as permanent demand shocks or permanent supply shocks which are problematic for common events such as monetary policy. In an effort to solve this challenge economists theorized the use of building economies from the ground up using agents that can interact and act rationally, but this required large amounts of computation which was not as widely available as it is today.

Over the last decade computational accessibility has skyrocketed resulting in the smartphone in one's pocket containing more computational ability than most computers that existed a decade ago. This rapid expansion in computation has drastically changed how economists view microeconomics in that modeling individual behavior has long been an arguably intuitive task that involved more psychological than statistical analysis but with the advent of modern computers the gap between micro and macroeconomics has never been closer. Agent Based Models have existed since the 1990s and have widely been used to model economic systems at a realistic scale using individual microeconomic analysis and computation to aggregate economic behavior at the individual level. This was theorized in the 1990s by Epstein and Axtell who wrote, "fundamental social structures and aggregate behaviors emerge from the interaction of individual agent operating on artificial environments under rules that place only bounded demands on each agent's information and computation capacity" (Epstein and Axtell (1996), henceforth E&A 96, p.6). This aggregate microeconomic analysis has been expanded with the advent of increasing computational capacity and agent-based models have gotten increasingly complex as the underlying models have been expanded in programs that can handle increasing levels of complexity.

Considering the existing computational power available to them, economists further developed the agent-based model (ABM) into a model which allows for interaction and rationality in economic decisions at the individual level. ABM models have been extensively used to model macroeconomic and biological systems and do a very good job of modeling homogenous agents that only differ in the conditions in which they exist. ABM models, however, do not perform as well in environments with heterogeneous agents that are different from one another and are different from themselves at different points in time. The limitations of ABM models were widely discovered as a result of the financial collapse in which agents had permanent shifts in their demand and utility parameters, and in which the system which contained the agents had permanent shifts in policies that forced the individuals to adapt. These limitations led to the development of the heterogeneous agent models (HAM) in which individuals are all different and it is the very differences in individuals that create the realism of the model. HAM models often follow a series of differential equations which act as a systematic stochasticity that ensures heterogeneity between agents and a basic structure for the underlying economic system. These HAM models are very computationally efficient and effective methods for estimating heterogeneous economic systems and even to some degree have somewhat accurate results.

Despite the efficiency, simplism, and accuracy, there exists an underlying assumption in all HAM models which is continuous time. This means that all agents exist at all times in the system (Gabaix, Lasry, Lions, Moll 2016). This assumption might not be far off when estimating business competition, or entire economic systems acting as agents, or very short time periods in which all agents exist throughout the period. This assumption however is not as clear when estimating how individual and generational wealth changes over time because wealth changes hands frequently over long periods of time. While it could be argued that the individual or agent that receives the wealth, say in form of an inheritance, will have similar enough preferences to the originator that the effect will be negligible, this violates the heterogeneity of the model and would yield inaccurate long-term results. These frequent wealth transfers are not only essential to modeling wealth inequality but act as a vital way in mitigating future wealth inequality, and yet are not included in existing HAM models. An example of which is inheritance in which wealth is transferred to offspring. This is a crucial part of wealth inequality that is not accounted for in HAM models and yet is an essential part of modeling pareto wealth distributions.

In order to fully understand the evolution of pareto distributions, especially over long periods of time, these factors must be accounted for. These factors, while on the surface may seem to be easily implementable, due to the aggregation and generalization of most HAM models, relationships are hard to account for at the individual level. In order to solve this problem a more individualized approach is required in which the aggregation is directly built from the ground up and modeled as individual behavior. This individualized model will not only solve the issues discussed earlier, in that it will directly model wealth transfers and creation of new agents but will create similar aggregation measures used to analyze wealth distributions in HAM models. This individualized approach will also allow us to analyze agent specific traits such as patience and altruism and how they impact relative wealth. This model will follow a general equilibrium approach in which rational individuals maximize utility given a number of constraints in a single good system. This model will also include additional considerations discussed above including offspring and time inconsistency. The utility of each individual will be as follows:

$$\mathrm{U}(C) = \sum_{i=0}^{n_p} \beta^i u\left(c_{s_i}(1+r)^i\right) + \alpha \sum_{j=0}^{n_f} u_j\left(c_{f_j}\right) + \delta \sum_{k=0}^{n_d} u_k\left(c_{d_k}\right) + \epsilon$$

The first part of this equation represents the total utility gained from consumption over all periods n_p where c_{s_i} is the number of goods consumed in period i including the return received when the goods are consumed. The individual’s patience parameter β discounts the utility in each period and is scaled by the number of periods into the future the goods are allocated. The second part represents the total utility gained from donations to family members over all family members n_f where u_j is the utility function of the jth family member discounted by the individual's altruism α where c_{f_j} is the number of goods donated to the jth family member. The final part of the utility function represents the total utility gained from donations given to a number charity cases n_d where u_k is the kth charity case's utility function which is discounted by a charity parameter δ and where c_{d_k} is the number of goods given to the kth charity case. A charity case in the model is defined as any random individual that is in any other family, in order to replicate asymmetry of information and make fewer assumptions about the individuals the model will assume random uniformity in who is eligible for charity. This random uniformity means that as in everyday life individuals can give to anyone so long as they want to regardless of the individual receiving the gift. Each of these "transactions" are then summed as an aggregate utility, consuming a total number of goods C, for the individual in addition to some random factor ϵ which will represent random shocks in individual preferences. This utility function is also constrained by initial wealth in that period ω. Where each utility function follows a power law utility function as shown below:

$$\mathrm{u}(c) = \frac{c^{1-\gamma}}{1 - \gamma} \quad 0 \leq \gamma \leq 1$$

Where gamma represents returns to scale of the good and in the model will be a positive value from zero to one meaning that the curve is concave and therefore will have a solution to the constrained maximization below. This utility function was chosen due to the nature of its second derivative and how slowly it approaches zero or in other words the difference in utility from one level of goods to the next is more distinguishable than using a Cobb Douglas or natural log function. This distinguishability will be crucial to the model's ability to make decisions without making additional assumptions as described below. With all this in mind the constrained maximization is as follows:

$$\max_{\mathrm{c_s},\mathrm{c_f},\mathrm{c_d}} \mathrm{U}(C) = \sum_{i=0}^{n_p} \beta^i u\left(c_{s_i}(1+r)^i\right) + \alpha \sum_{j=0}^{n_f} u_j\left(c_{f_j}\right) + \delta \sum_{k=0}^{n_d} u_k\left(c_{d_k}\right) + \epsilon$$

$$\text{s.t.} \;\; \mathrm{C} = c_s + c_f + c_d \leq \omega$$

$$FOC_{c_s}: U'(C) = \sum_{i=0}^{n_p} \beta^i (1+r)^i u'\left(c_{s_i}(1+r)^i\right) = 0$$

$$FOC_{c_f}: U'(C) = \alpha \sum_{j=0}^{n_f} u_j{}'\left(c_{f_j}\right) = 0$$

$$FOC_{c_d}: U'(C) = \delta \sum_{k=0}^{n_d} u_k'\left(c_{d_k}\right) = 0$$

While mathematically this can be solved it would require additional assumptions about the individuals, families, or underlying economy in the maximization which arise from the random individual selection in each action. Another reason additional assumptions would need to be made to solve this problem is the heterogeneity that exists between individuals that precludes agents from determining another agent's utility prior to an agent performing an action. An example of this is that an action between agents one and two may not yield the same utility if agent two receives a lot of goods prior to the action which in turn may make this action no longer optimal for agent one. The program will refrain from making these and circumnavigates these problems by having the individual determine the optimality of each good sequentially in the time they occur. The computational approach to solving this maximization problem is less intuitive but more realistic and involves iterating through each good rather than determining the optimal use of all goods before any transactions occur. Iterating through each good that each agent has allows an agent to maximize utility given an initial wealth by determining the best use of each good sequentially which is a somewhat more realistic solution seeing that this is how real agents would act given a number of goods. Using this approach in the example above agent one would not participate in the action because they have taken agent two's gift into account and determined the action suboptimal. One problem with this process, however, is that it is impossible to have several actions or several individuals acting at the same time due to computational constraints. The model will be solving this problem by having a random individual act and then another until all goods are used.

This iteration process goes through several choices, each of which resembles the first order constraints but are only concerned with how the specific action compares to the utility gained from the individual consuming the good directly. The first step in this process is calculating personal utility, or the utility gained from consuming c goods directly which is then compared to a series of choices to determine which will result in the highest marginal utility. Each comparison requires that for an action to occur the resulting marginal utility the individual gains must be the highest of each action with each comparison is shown below as inequalities. With c being some number of goods that is passed into a given marginal utility function for the possible individual, including some random variable ϵ representing preference shocks. The inequalities are as follows:

Offspring:

$$\mathrm{u}'(\bar{c}) < \alpha \hat{u}'(\bar{c})$$

This inequality represents some estimated marginal utility that the parent will receive from a set number of goods $\bar{c}$, which represent total cost of childcare being given to their child. In other words, a possible child's utility from all forms of childcare, which is then weighed by the parent's altruism parameter. If this inequality holds then the individual will create a new child and give $\bar{c}$ goods to that child, which is meant to represent the tradeoff of having a child or using the childcare cost equivalent for other utility generating sources. This in mind the offspring equation has an estimated marginal utility which is symmetric or similar to the individual who would be the parent meaning that the γ or returns to scale of the possible child is very similar to the parent's value. The rationale behind this is that parents might view their child's preferences as that of their own with some variation. One thing to note about this inequality is that this is the only inequality that does not compare all actions but rather just that of individual consumption. This is because in order to create a sustainable population growth we must make a single assumption that individuals have some sort of priority to produce children over other utility generating sources which goes beyond the altruism parameter. It might be helpful to think of this assumption as other societal norms and factors that influence individuals to produce offspring beyond the utility it provides. While this assumption is somewhat weak and would be hard to prove, it was a choice that I thought would provide better results and would have little effect on the maximization problem. This is because the individual is still maximizing between two choices rather than five.

Family Donation:

$$\max(\mathrm{u}'(c), \delta u_k{}'(c), \beta^i(1+r)^i u'\left(c(1+r)^i\right)) < \alpha u_j{}'(c)$$

This inequality represents the marginal utility that the jth random family member would receive from the individual weighted by the individual's altruism parameter. If the inequality holds then the utility gained from the family donation is higher than any other action and the random family member will receive the goods from the individual.

Charity:

$$\max(\mathrm{u}'(c), \alpha u_j{}'(c), \beta^i(1+r)^i u'\left(c(1+r)^i\right)) < \delta u_k{}'(c)$$

This inequality represents the marginal utility gained from the kth random individual that is not within their family and works like the family donation inequality but is weighted by a charity parameter delta.

Saving:

$$\max(u'(c), \alpha u_j{}'(c), \delta u_k{}'(c)) \; < \beta^i(1+r)^i u'\left(c(1+r)^i\right)$$

This inequality represents the tradeoff between consuming a number of goods in the current period or saving the goods for the ith period in addition to receiving some return r compounded by the number of periods it is in savings which is all weighted by some patience parameter beta. If the inequality holds the individual will save the goods for the ith period and receive a return. This model assumes the individual will only consider the next period, this is to represent the uncertainty individuals have in their lifespans and to reduce unnecessary computation.

Despite its simplicity this iteration poses two problems, the first of which is that the agent is using whole goods, meaning that goods cannot be split or divided, and that this is a very computationally intensive process, especially when considering the scale of realistic economies. Despite these shortcomings the model should produce a level of accuracy similar to the generalized HAM models given similar inputs but should reveal some insight into generational wealth distributions that previously was not possible.

This model also requires a fair number of assumptions which are not entirely unrealistic. The most important of which is decreasing returns to scale. This assumption is required because indifference in the number of goods in a bundle should not exist in our model meaning that the millionth good should not have equal utility to the million and first good but rather will have a slightly lower utility. This assumption is crucial to the iteration process described above because if an individual is indifferent to multiple decisions additional assumptions would need to be made such as selfish behavior, charitable nature, or altruistic giving. In other words, each good must be used and if an individual is indifferent, they must decide where to put that good which requires additional information about the individual, which will be making further assumptions making the model more susceptible to inaccuracies. By eradicating indifference, the model will be able to iterate through each good and place each good according to its optimal use which reduces the number of assumptions made and reduces computation. This additionally adds realistic expectations to our model because in reality humans experience indifference and yet still make decisions due to scarcity of goods and their temporary nature.

Another important aspect of the model is the introduction of randomness and stochasticity into parameters, events, and interactions. In continuation of realistic modeling this mode will add randomness to almost every parameter, action, and interaction in order to make the model more accurate. This randomness will create a more realistic approach to the seemingly stochastic nature of economies and human interaction and will create the heterogeneity in individuals and action that the model hopes to replicate. In individuals, parameters will be randomized with some following uniform distributions and others normally distributed in order to create heterogeneity as shown below:

$$\gamma \sim \text{Uniform}(0,\ 1)\ ,\ \mathrm{r} \sim \text{Uniform}(0,\ 0.2)$$
$$\beta,\ \alpha,\ \delta \sim \mathrm{N}(\mu,\ \sigma^2)\ ,\ \epsilon \sim \mathrm{N}(0,\ \sigma^2)$$

Within the family structure new agents will have somewhat random parameters with traits such as altruism and impatience following values closer to their parents in order to replicate the learned values and characteristics that are taught to us by our immediate family members. In the general economy each action will be randomly selected or in other words an agent will be selected randomly and will do a single “turn”. These “turns” are what the individual will do next according to the inequalities above. In all actions a random variable will be inserted which will simulate the random nature of events and stochasticity of preferences that occur daily. The randomness inserted into each facet of the model will create realistic noise in the data which will highlight the stochastic nature of the model and underlying economic system.

In each simulation data shall be collected from each individual in each time period. This dataset will include all parameters of the individual including altruism, charity, preferences, returns to scale, etc. as well as a number of family characteristics such as wealth, size, and average characteristics. This dataset will also include economic aggregates such as total wealth, population, and additional average family totals such as wealth and size. This dataset will also include the number of goods used in each choice including the number of goods that the individual started with, goods to be received in the next period, charitable and family donations, and goods used in child production.

In order to ensure that the model creates realistic economic outcomes a simplistic simulation was run to illustrate how different individuals act given a set number of goods. This simulation was run by first generating 100,000 individuals each of which have randomly regenerated altruism, charity, and patience each of which is uniformly distributed from zero to one. Each individual is also given a random number of goods as a function of a set childcare cost which for this specific simulation is twenty goods, so each individual is given a randomly generated integer from one to ten, which is uniformly distributed. This integer is multiplied by the childcare cost to create a starting wealth. This simulation is only run for a single period because we are only interested in how these individuals act holding all other factors constant such as family size, family wealth, and inheritance. After running this simulation ten times four charts were created to highlight subsets of the population which are shown in figure one.

Figure one displays two charts for high and low returns to scale groups, roughly 0.2 and 0.8, and within those plots highlights wealthy and poorer individuals, meaning those in the bottom ten percent income percentiles and those in the top ten, which are highlighted in gold and blue respectively. Within these plots are aggregated measures for individual subsets containing different altruism and patience with the left plot aggregating the ratio of wealth spent on producing new children and the right aggregating the ratio of wealth spent on consumption. In addition to these visualizations the plots display the aggregate savings rate for individuals in each subset and the darker a region is the higher the savings rate is. While this plot contains an immeasurable amount of information, several characteristics highlight the model's features and how we will be using it with real data. The first of which is that individuals with higher altruism spend more of their wealth creating children and individuals with higher patience have higher savings rate, while these results may seem trivial, they provide some proof that the individuals in the model act as expected. These results also mean that given data such as population growth and personal consumption expenditures will allow the model to triangulate the subset of individuals that fit best with the subset population the statistics result from. This subset triangulation method will allow us to form the model to different populations which will be the key to understanding localized wealth disparities and how they arise and evolve.

Another important conclusion from figure one is that different returns to scales affect the level of the ratio of wealth spent on producing children and the ration of wealth spent on consumption. This result once again validates the expectation of the model, in particular individuals with high returns to scale and a low level of altruism have very high levels of consumption and vice versa for those with low returns to scale and high altruism. Despite returns to scales' impact on our model it is not possible to calculate with current data so in this model regardless of the population it will assume that returns to scale are uniformly distributed. While this assumption is mandatory due to lack of data it is not too far-fetched an assumption given apparent uniqueness and uniformity of individual utility distributions regardless of locality.

From the model expectation validations in figure one we now will be able to model an economy based on its data and use this tailored model to determine how wealth inequalities occur and how they change under different policies. So given a set of inputs which include population growth, savings rate, consumption rate, and charity rate the program will run a series of simulations for any number of periods and produce a very large database of results. Using these data points the model will triangulate the closest altruism, patience, and charity parameters for the estimated subset that the population exists in, this will be used to fit a model to the given data, which will then be used for analysis on its wealth distributions. Using population growth, the model will normally distribute altruism parameters to produce a similar rate of growth given a population size which can remain constant or change with the model. Consumption and Savings rates, however, will be used to create normally distributed patience values that represent the tradeoff between saving and spending. Finally, the charity rate is used to normally distribute the charity parameter to fit the level of charity in a given economy. These inputs will create a tailored model which should be specific enough to highlight distinctive features in economies and will show specific evolutions of wealth distributions.

In order to create triangulation data for possible input values a large number of simulations must be run in order to account for all possible subsets of populations. The simulations will go as follows; economies will be simulated according to different starting parameters such as altruism, patience, and charity of which there are ten groups each or 100 total variations. For each group a different standard deviation value is passed in order to create varying amounts of heterogeneity and each of these subgroups is simulated five times in order to mitigate outliers in sample data. In addition, there will be twenty subgroups which will each receive a different starting wealth as a function of childcare cost ranging from one to twenty which will be used to determine the optimal level of income for accuracy. Each sample consists of a thousand individuals which will be in an economy over the course of fifty periods in order to ensure the population growth in the economy has stabilized. In addition to these simulations the model will be slightly more complex due to the introduction of time inconsistency and now each individual will have a maximum lifespan of six periods and cannot produce children in the first two or last periods. While these may seem arbitrary, they are meant to simulate the stages of life that humans undertake ranging from early childhood to elderly and how these stages in life create heterogeneity between the same individual in different time periods. Another addition to the model is that of an interest rate which will be available at the economy level, which works similar to that of a fixed rate from a bond, and this will be uniformly distributed from zero to twenty percent in order to create a nonzero interest rate.

Using the simulated data, the program will be able to estimate parameters for real economies and for this paper I will be primarily analyzing the US economy, which I chose mainly due to the availability of data and notorious wealth inequalities. The data collected to estimate the US model includes population growth, personal consumption expenditures, personal income, charitable contributions, and savings rates from the 1980s to now all of which is from the FRED database. The consumption rate is calculated by dividing personal consumption expenditures (PCE) by personal incomes (PI) which represents the percent of income spent on consumption. The charity rate is calculated by dividing charitable contributions (CC) by PI and represents how much income is given to charity. The savings rate is accessible as a calculated statistic and requires no additional calculation. One thing to note about the US rates is that taxes are not factored into any of the rates, this is because of the withholding process in the US tax system that forces individuals to automatically pay a percentage of taxes out of any income they receive. This means that because individuals will not consider the portion of their income that is withheld in their consumption, savings, or charity actions meaning for the US we can ignore taxes in these rates. Using this data, the program will find closest economic group for each year which is estimated using a normalized Euclidean distance formula shown below:

$$e_t = \sqrt{\left(\frac{x_t - x_i}{\sigma_x}\right)^2 + \left(\frac{y_t - y_i}{\sigma_y}\right)^2 + \left(\frac{z_t - z_i}{\sigma_z}\right)^2 + \left(\frac{p_t - p_i}{\sigma_p}\right)^2}$$

This formula represents the "error" so to speak of a simulated economy to an economy in time t and represents how close a simulated economy is to the real data where x, y, z, and p represent the savings rate, consumption rate, charity rate, and population growth respectively. In addition, the subscript t represents the true US data at time t and the subscript i represents the simulated value for a given economic simulation. This formula was chosen because of the "error" term that it outputs and due to the normalizing factor, which weights the datapoints equally. This means that if there is more variation in savings rates than population growth simply due to the nature of the economy then this formula will take this into account and not forgo population growth accuracy to be give a higher accuracy to the savings rate. This equation is then used to create an average error for all time periods for each simulated economy and the closest fifty points are used to estimate the true parameters.

Using the simulated data, US data, and the normalized Euclidean distance formula the program can now determine the optimal parameters which best fit the real economy in question, which for this paper will be the US. Figure two presents one factor to consider when performing these simulations which is how many goods should individuals start with? While this question may seem arbitrary because we are measuring economies in relative terms, the starting income will change every individual's thresholds for certain actions. What I mean is that if individuals have more income while all other factors remain the same then the choices up until the previous income level will remain the same but those beyond will be different and therefore create different consumption, saving, charity, and population growth rates. This figure attempts to solve this problem by calculating the "error" of the closest fifty economies under different initial income levels plotting the error on the left side and the estimated parameters on the right. Some things to note about this plot is that the left side asymptotes to some level which could be thought of as a systematic difference between the economic simulation and real economy this model is based off. This systematic difference could be thought of as underlying features in the US economy that could not be captured by the model but due to the relatively low level of error, we can infer that as initial income increases the error will asymptotically decrease to the systematic difference. Another thing to not about this plot is that the parameters seem to asymptote to their values very quickly and have no variation after sixty which is because the simulation groups increment each parameter by about 0.1 and so each parameter essentially finds which of the groups fit best. That being said, it is still fairly clear that the parameters asymptote to certain values as initial income increases.

Using the results from figure two the program will be using sixty as the initial level of wealth. This is because the error appears to flatten at this point, and this is the leftmost point that shares all the asymptotic parameter values which will allow the model to as little computation required as possible while still benefitting from the asymptotic properties described above. This value will be crucial in all future simulations and could possibly even be accurate enough for other economies but that will be left for ponderance. Using this level of goods, we will filter out the economic simulations to just those with this level of initial wealth and this data will be used to calculate the optimal parameters. This filtered data is plotted along with the real US data in figure three. This figure highlights several key pieces of information, the first of which is that the scatter plot is flat meaning that the tradeoff between consumption, charity, and savings creates a triangular surface which can be seen in this scatter plot. Another very crucial thing to note in this plot is the stagnant nature of the US economy meaning that despite nearly forty years of data which have spanned several recessions and each year of which represents an arguably vastly different economy from the next. This feature of the US data will be crucial to our analysis of longer time horizons because as shown in this plot despite ever changing economic conditions the underlying norms and structure of the US economy remains somewhat stagnant around this area. Which means that the parameters which govern the underlying economic system in our simulations need not change as well even though the economic conditions within them may change. This in essence proves that this model can work over very long periods of time so long as the parameters are accurate enough to capture the stagnant features of the economy in question. In addition to the US data Canadian data has been plotted which shows a similar stagnant nature with dramatically different consumption, saving, and charity rates which can primarily be due to the US tax system incentivizing charity as a means of decreasing tax burden. The existing US tax structure with the addition of the large amount of wealth in the US creates a much higher charity rate in the US than in other countries which is what creates the large difference between the US and Canada in this figure.

While this plot contains an immense amount of information, the only part we are interested in going forward is where the US data lie. So, figure four has inflated the area around the US data to once again how stagnant the data is and to show the fifty best model estimates which are highlighted in green. This figure shows that the model estimates seem to fit the data well with the majority of the points within the cloud of US data but due to the unsustainability of some simulations some areas are sparser than others. This problem is overcome however by the structure of the normalized Euclidean distance formula which will weigh more sustainable economies, or those with population growth closer to the US, as high as the other inputs. This means that the skew of the best fit points is likely because it is skewing to more sustainable economies. Despite the skew it is still clear that there are enough datapoints to create an accurate enough picture of the US economy to understand the cause of the existing wealth inequalities in the US.

Using this model and the dataset resulting from the simulations it runs it will be possible to follow individuals with certain characteristics in order to understand how certain aspects of an individual affect individual and generational wealth. Some examples of this which will be highly revealing will be how highly skilled poor individuals generate sustainable wealth and how low skilled wealthy individuals degrade or subsist on existing generational wealth. These two subsets of individuals will be extremely important in the understanding of how sustainable generational wealth is created and how it can be depleted. Another subset of individuals that will allow us to determine the role that offspring has in wealth accumulation and depletion is highly altruistic individuals which have a lot of offspring, this is because an important component of generational wealth is the tradeoff between family size and success. These subsets of individuals are crucial to understanding how some families remain wealthy for decades and others deplete wealth in a few generations and will be highly beneficial in understanding how to mitigate wealth inequality.

In order to understand wealth inequalities and how they arise we need to determine if existing wealth inequalities happen out of random processes or happen due to specific policy changes or something non-random. In order to compare the simulated economies to existing ones we will be using the Gini coefficient which is a measure of how much inequality exists between individuals and can be calculated using the formula below:

$$G = \frac{\sum_{j=1}^{N}\sum_{i=1}^{N}\left|x_j - x_i\right|}{2N^2\bar{x}}$$

This formula will output a value from zero to one and essentially represents the absolute total difference in incomes divided by the average and scaled by the number of individuals in the economy. From the formula above a coefficient value of one will mean perfect inequality in the economy and a value of zero will be perfect equality or that every individual has the same income. Using this formula, we will calculate the coefficient of each economy in each period using the optimal parameters calculated above and this will allow us to determine a confidence interval for coefficient values which will be used to determine the significance of an existing wealth inequality. If this value is significant at an $\alpha = 0.05$ level among the simulations than it could be concluded that the existing inequality has been created by something other than random processes but if not, we will accept the null hypothesis of random occurrence. In order to create the confidence interval, we will first need to create the simulated economies. These simulations will each have the optimal parameters calculated above and will exist for two hundred periods and the coefficient will be calculated for each period and economy. There will also be two hundred of these simulations run to ensure there are no outliers in the starting conditions which may produce different results further into the periods.

After running these simulations for two hundred periods each, the resulting Gini coefficient distribution is shown in figure five. This figure displays the distribution of the average Gini coefficient for each simulation meaning that there are two hundred observations in this distribution. This figure also plots the existing US Gini coefficients since 1984 and a confidence interval has been calculated at the 95% level which is shown in the subtitle. This plot shows that not only is there no overlap between distributions but that each US coefficient is highly significant. This result means that the existing Gini coefficients, or existing wealth inequalities, are the result of something other than random process. What that result is will require further analysis, but it could be due to policy in favor of wealthy individuals, tax changes that help those with means, or just a social pressure for those with wealth to maintain it. While this result is significant it is not surprising in that for decades the US has been the prime example of wealth inequality driven by a movement of other wealthy countries to redistribute wealth to more equitable levels. This result only furthers theories regarding wealthy individuals manipulating the current US system to increase wealth when considering that not only are all the US coefficients highly significant but have been increasing steadily since the 1980s. This means that the current system creates nonrandom wealth inequalities, and that the existing system has increased the inequality to even more significant levels since the 1980s.

So, given all these results the question becomes how do we bring these inequalities to more equitable levels? This question has been pondered by many researchers and economists since the it first entered the academic space and while much progress has been made at implementing policies that alleviate some wealth inequality many countries still experience high levels of inequality. There are several reasons for this which most important are that drafting and implementing policies is a tough pill to swallow despite the goal, especially in politically gridlocked countries. This problem becomes even more exacerbated when the policy is about tax increases, tax loophole fixes, or any form of wealth redistribution and one could speculate the many reasons why policymakers might not back a bill pertaining to any of these. The most common reason, however, is that it costs time and money to implement any bill with as sweeping changes as those listed especially when considering the difficulty of assessing policy efficacy or the impossibility of comparing efficacies. Both reasons have hindered the economic progress of nations and until the second is solved improving the existing inequalities will be nearly impossible.

My approach to this problem will be using the model to implement different policies and compare the results as we did when comparing simulated economies to real data. Specifically, the model will be used to implement different tax structures such as wealth taxes, progressive and wealth redistribution and determine efficacies and drawbacks of each. In order to implement these tax structures, however, the model will need several additional assumptions. The first of which is that no individuals cheat on their taxes or essentially that there are no loopholes in the implemented tax systems. While it would be possible to eradicate this assumption by adding a random variable to each tax structure that would weigh the individual's participation in such a tax structure, the model will take the assumption of the legislators that individuals do not cheat on their taxes. Another assumption needed for implementing these tax structures is perfect implementation in that these policies are perfectly implemented in the economy in that the governing body knows every individual's true income and require no costs to implement. This is a fairly weak assumption in that it would be impossible for any government, no matter how efficient, to perfectly implement any policy regardless of what it entails, let alone one the size of most wealthy countries. Despite this assumption's lack of standing, it is required to analyze the policies described above.

Using the data and assumptions described above the model will first implement a progressive income tax rate. This will be implemented by taking a part of the return on saved goods in the period that they are given back to the individual. This is because when an individual saves part of their wealth this is returned to them along with their principal in the next period as income. The model will be implementing progressive tax brackets which will change as income (or the return on saved wealth) increases. The tax bracket thresholds will be percentiles of the income distribution starting from 70% and will be taxing 5% for the first bracket and an additional ten percent for each higher bracket. This means that the highest tax bracket, or those in the top 10% of the income distribution, will be paying 15% in taxes. One important part to implementing this policy is that the tax revenue gained will not be redistributed meaning that we will assume that the governing body uses the tax revenue for public utilities which do not directly change the individual's consumption habits but will increase the individual's total utility. The program will run the same number of simulations as in our simulated data comparison, with two hundred total simulations which each run for a maximum of two hundred periods. For each simulation the average Gini coefficient will be taken of each economy over the course of those two hundred periods as in the simulation comparison above. The results of these simulations are found in figure six which plots the distribution of economies with the policy and those without in addition to the median of both distributions.

From this plot it is clear that this policy has little effect and even when calculating the average treatment effect, the resulting coefficient is not significant reinforcing the lack of effect displayed in the figure. While this lack of treatment may be due to the policy implemented in the code or the setup of the model, one cause is that the tax revenue collected is not reintroduced into the economy. In essence this is creating a much larger deadweight loss than if the revenues were used for welfare programs as in the US and might be masking the true effects of the policy on the existing inequalities. In the next policy we will see this same problem, but we will not attempt to solve it until the final policy implementation where the model will redistribute tax revenues.

The wealth tax will be implemented by taking a part of the revolving savings or in other words the goods individuals have saved minus the return they have received which is meant to capture the initial principal in the savings not including the income. This will be a progressive wealth tax in that the higher savings the individual has the higher the individual will be taxed, and it will be following the same bracketing and taxing as the income tax. This means that individuals in the top 70% of wealth will pay 1% of their wealth in taxes which increments by an additional 5% every additional bracket to a maximum of 3% for those in the top 10% of the wealth distribution. The results for this policy implementation are shown in figure seven which follows the same structure as the previous figure. This plot illustrates the exacerbated problem of deadweight loss which now has created two distinct distributions with little to no overlap meaning that if the deadweight loss problem did not exist the policy would increase inequality. Which from economic intuition makes no sense meaning that this result is likely due to a larger deadweight loss from the increased tax revenue being removed from the economy since in this model wealth is typically higher than income. In order to address this problem, the model will be reintroducing the wealth in the final policy implementation described below.

The final policy implementation will be direct wealth transfers in which a percentage of an individual's wealth will directly go to a random individual in need out of a pool of “needy” individuals. This pool of individuals is meant to relax the assumption of perfect implementation and perfect information on the part of the governing body in that the government has information on income but does not know the utility gained from the additional goods. Therefore, it will randomly distribute these goods uniformly to the individuals in this pool in an effort to reach the individuals who need it most. This pool of individuals will be made up of the bottom quantile of the economy or the bottom 25% of the wealth distribution and out of this pool half of the individuals will receive an equal payment of the total tax revenue. This wealth redistribution will be implemented under the two tax policies analyzed above, progressive income and wealth taxes, to determine the optimal method of wealth inequality mitigation. The results of which are shown in figure eight which displays both policies in addition to the no policy economies. This figure shows a major improvement from the problems discovered in the previous policies in that each policy shows positive improvements in mitigating wealth inequalities due to the reintroduction of the tax revenues into the economy. Using this plot we can infer two things, the first of which is obvious and is that both policies reduce wealth inequalities at a significant level, as determined with an average treatment effect. The second is that wealth tax clearly is more effective in mitigating wealth inequalities in that the distribution is closer to zero and has a much higher average treatment effect. This is because the amount of goods going to the pool of individuals is much higher than in the income tax and therefore has a larger impact on the wealth distribution. While one could look at this plot and determine that the wealth tax is the best solution because it has the largest effect and while this isn't entirely correct in the context of this model when considering the cost and social impact of this tax the answer is less clear. The wealth tax does create the largest effect it would also be the costliest socially and fiscally of the two policies seeing that it would be costly to not only keep track of individual wealth but to then distribute that same wealth which in essence recycles the same wealth. This policy would also be very costly socially in that there has been a stigma attached to any form of wealth distribution which stems from the red scare. The income tax on the other does seem to provide a fair amount of mitigation and does not suffer from the same issues the wealth tax would encounter but may not be going far enough to address the issue. Further analysis of both policies and the effects they have on the economy is required before any definitive solutions are implemented but this paper attempts to shed light on some of the discussion in the field, and in no way provides solutions.

With all that in mind this model this paper does not attempt to calculate any accuracy measures because there is no data to compare our results with. This is because the model is tailored to the US economy which has only implemented a progressive income tax with some redistribution so we would not be able to determine the effects of the other policies until the US has done so. Despite the US's implementation of a progressive income tax because of when this policy was implemented and the lack of data in that time period it would not only be difficult to determine the effect of the policy. In addition to this problem, it is highly likely that the tailored model needed to compare the effect of the policy would be completely different seeing that the progressive tax system was implemented nearly a hundred years ago. So, not only would we need to create a measure of efficacy for that time period, of which there is far less data than today, but we would also need consumption, savings, and charity rates for the same time period, which does not exist. This is likely the biggest obstacle this model faces and is what creates a fair amount of hesitancy towards using these models because it is difficult to translate findings into reality especially when the stakes are as high as reforming the US tax system. Additional research is still needed in this field to determine a method of calculating accuracy using these models and how to apply these models to time frames in which we do not have data. These problems however do not entirely invalidate this model or our findings around the source of wealth inequality because all the data is within the same time frame and the significance level is high enough that clearly something nonrandom.

These problems do, however, poke some holes in our policy analysis because despite all the policy comparisons being in relative terms the model does not account for any changes in the culture. This means that if a policy is implemented because of a cultural shift, which the model will not account for, the model will yield incorrect results in the level of efficacy. So, if no major cultural shifts occur in the timeframe that the policy is implemented the comparison analysis holds for the policies described. While this may appear to be a weak assumption, if one considers the timeframe in which policies are implemented given that no cultural shift created the policy, it is fairly likely that none will occur in this span. This assumption makes it possible to conduct policy comparisons as done above, so long as they are in relative terms, and makes the relative efficacy calculations meaningful given that they are significant. Which means that the results found in this paper should hold, even though no accuracy calculations were made, due to the structure of the analysis in relative terms, the length of time of our data, and the probable time duration the policy implementation takes.

All the code used to create the simulations and data visualizations are available at https://github.com/HenryHod/EconomySim. The simulations are coded with Java, the database and data wrangling are done in SQL including the underlying database, and all the data analysis and visualizations are done in Python. The simulations are created by instantiating an economy given a number of individuals that are each placed as the founder of a family and are given a random number of goods as a function of the current childcare cost, which currently is a fixed number but, in the future, could be a function of wealth. Individuals in the simulations are instantiated as connected agents in that they are part of a larger "clan" which represents immediate family, including siblings, parents, etc., which is part of a larger family which represents a number of clans. This family structure is the underlying system that creates the connectivity that separates this model from most HAM models and will allow us to view individuals as part of a web rather than individual agents. This model heavily benefits from the Objective Oriented Programming (OOP) nature of Java and in future research could be extended into further questions. Such as how does a web of "influence" so to speak or an individual's connections affect their success? or how do an individual's inherited skills affect their ability to maintain wealth? This model setup has a seemingly infinite number of applications and should continue to be improved and back tested to create the most accurate model possible so that the predictions the model creates become increasingly accurate. It is possible to imagine given the correct model setup and a fair amount of model complexity that the accuracy of the model could become useful enough to conduct meaningful policy analysis. If this accuracy becomes a reality policymakers will be able to pass policies that have the highest efficacies and the tradeoff between resources and return will no longer be discussed.

In this paper we have shown that the next step in the evolution of macroeconomics is coupling microeconomic theory and computational prowess to create macroeconomic models which can be accurate enough to make meaningful predictions. The model described in this paper relies heavily on the research done into HAM models and aims to approach a critical problem in the applications of HAM models to wealth distribution analysis. The model directly simulates the economic actions that are generalized most HAM models and by doing so this allows the model to view perform analysis at the individual level. This change also allows the model to perform actions previously impossible in HAM models such as inheritance and the ability to model an individual's web of connections through familial and locale inheritance. These changes required much more structure in the model than in many HAM models and required many assumptions about the nature of individuals and tendencies of utility generating sources. While many of the assumptions are common in many economic models a few required additional explanation and analysis but had little effect in the overall model and were more of an educated preference in order to make the simulations and underlying economic system more realistic. Using the model created the program ran many simulations in order to determine if the outcomes that the model produced were economically feasible, then used those results to create data used for determining the optimal parameters, and finally used those results to simulate the US economy. In the final simulations we ran an experiment to determine if existing wealth inequalities were the result of a random process and found every US value since 1984 was highly significant. This means that we can reject the notion that these inequalities resulted from something random but did not identify the true source which is likely a combination of multiple factors. In the final part of this paper, we focused on policies that could mitigate the existing inequalities and how one could compare policies to determine the best use of taxpayer time and money. Which we found that the wealth tax proves most effective but has more drawbacks and that income taxes do have some effect but appear to not go far enough in fixing the problem, but that both policies have clear positive effects on mitigating wealth inequalities.

Clearly there are a lot of considerations when formulating policy surrounding wealth inequalities especially when that policy is implemented in one of the wealthiest nations in the world. These considerations often leave policymakers and voters to make no decision at all and live with the existing system that they know and understand, which leaves little room for progress for those who cannot afford the current system. Although there may not be one clear cut solution to this problem decades of researchers have chipped away at various arguments, which this paper attempts to further, and will continue to do so as more data and methods are created. As long as inequality exists, research will continue to further our understanding of why they exist, how they evolve, and how to mitigate them so future generations will not have to deal with the same inequality they dealt with.

Figures

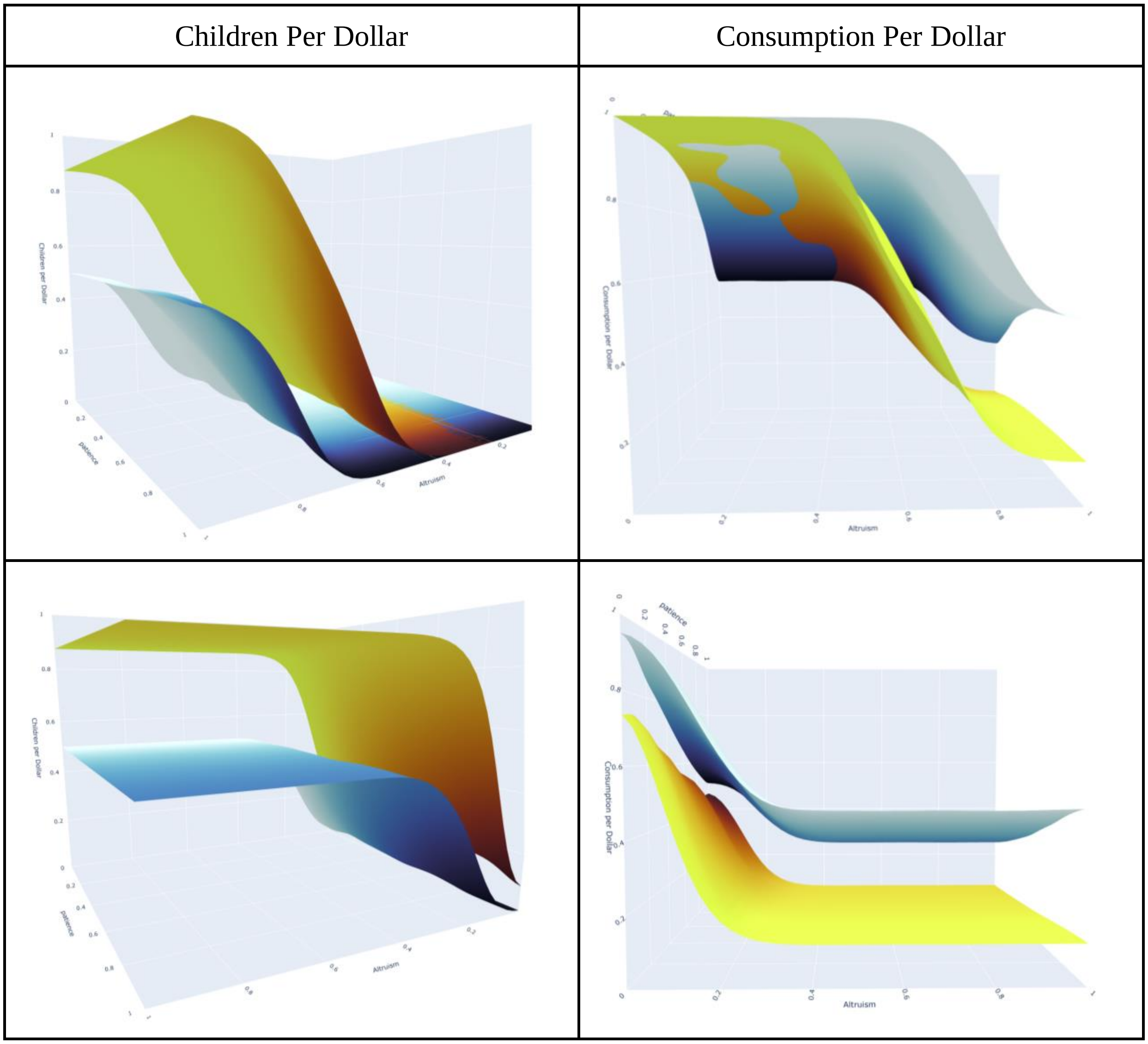


Fig. 1 Highlighting several subsets of potential individual characteristics and how they affect consumption and offspring production. The left column represents the percentage of income spent on producing offspring and the right displays the consumption rate of the individual. The top row represents those with low returns to scale γ and the bottom represents those high returns to scale. In addition, each plot is grouped by income where the gold surface represents higher income individuals and blue represents lower income where the x and y axis represent respective altruism and patience parameters. Finally, the saturation of the surface represents the level of the savings rate for the individual.

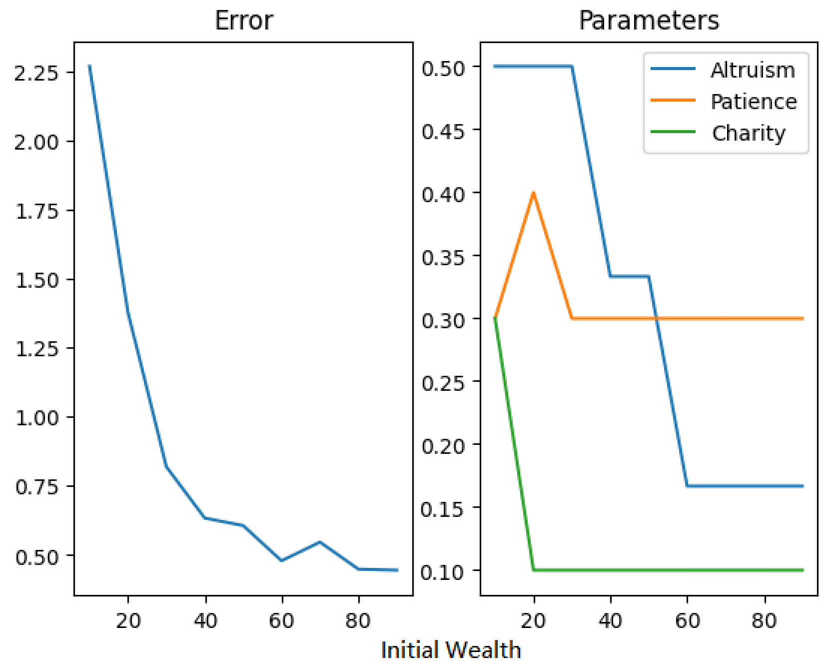


Fig. 2 Displaying the tradeoff between accuracy and computation and how increased income creates more accurate results. The left chart shows the normalized Euclidean distance or "error" for the best fitting economies for each level of starting wealth and the right displays the best fitting parameter estimate for each level of initial wealth.

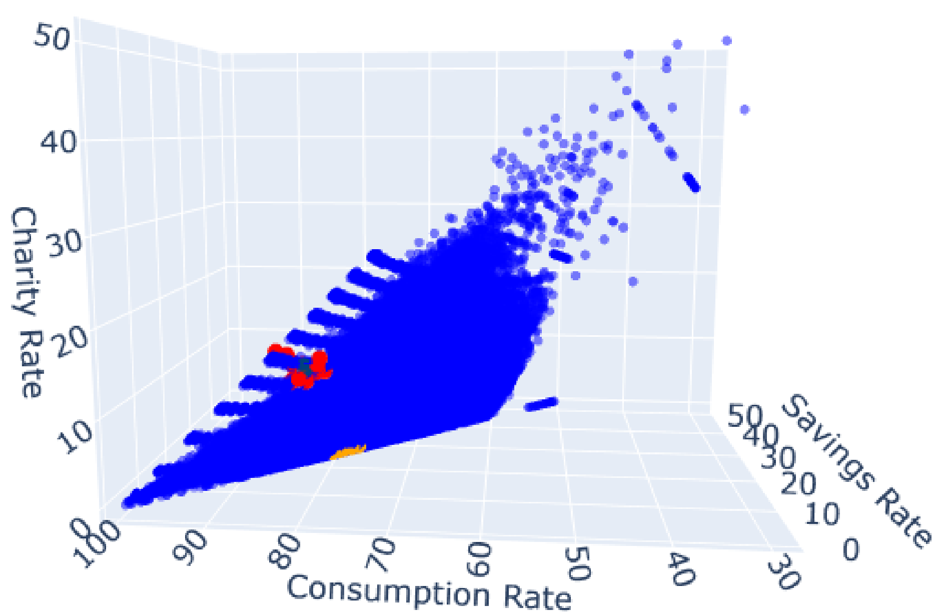


Fig 3. Compares the simulated economic data to existing economies and highlights the stagnant nature of economies despite changes over time. This chart displays the US highlighted in red and Canada highlighted in yellow where x, y, and z are the consumption, savings, and charity rates respectively. This chart also shows the linear nature of the tradeoff between consumption, savings, and charity which is primarily due to the fact that the three rates must sum to one.

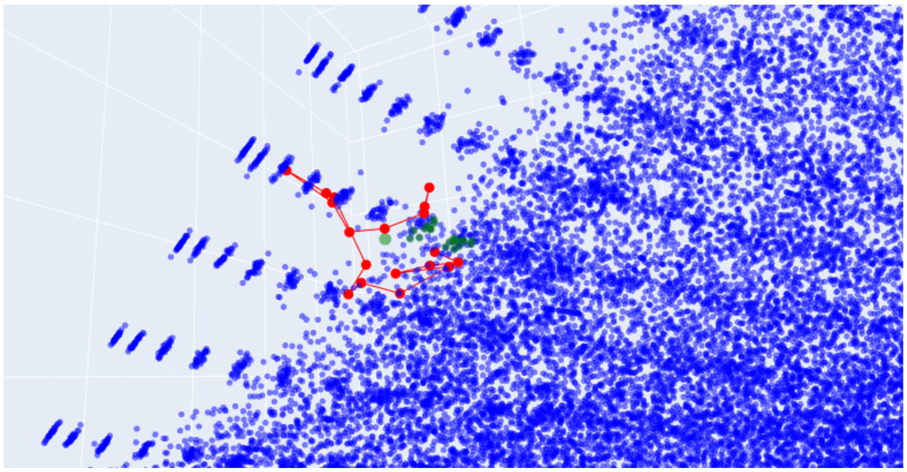


Fig 4. This is an inflated view of Figure but illustrates how close the US data is despite the time horizon of the data and that the estimates which are highlighted in green fit the economy well but is not perfect.

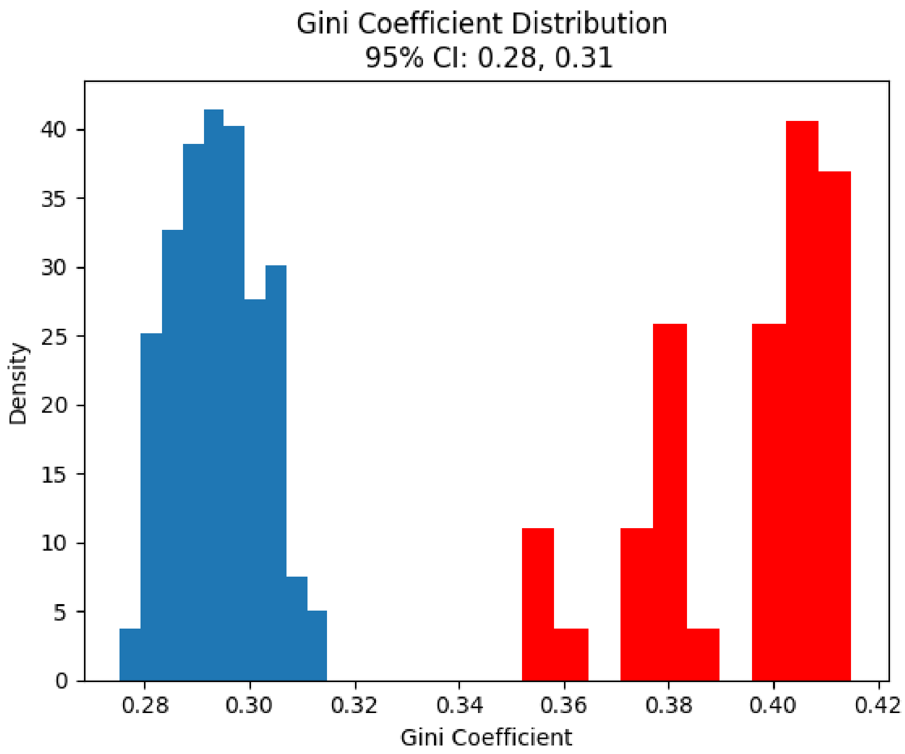


Fig 5. This chart shows the distribution of simulated Gini coefficients in blue and the actual US estimated Gini coefficients since 1984 in red. The 95% confidence interval for the simulated coefficients is included in the title which makes every data point highly statistically significant.

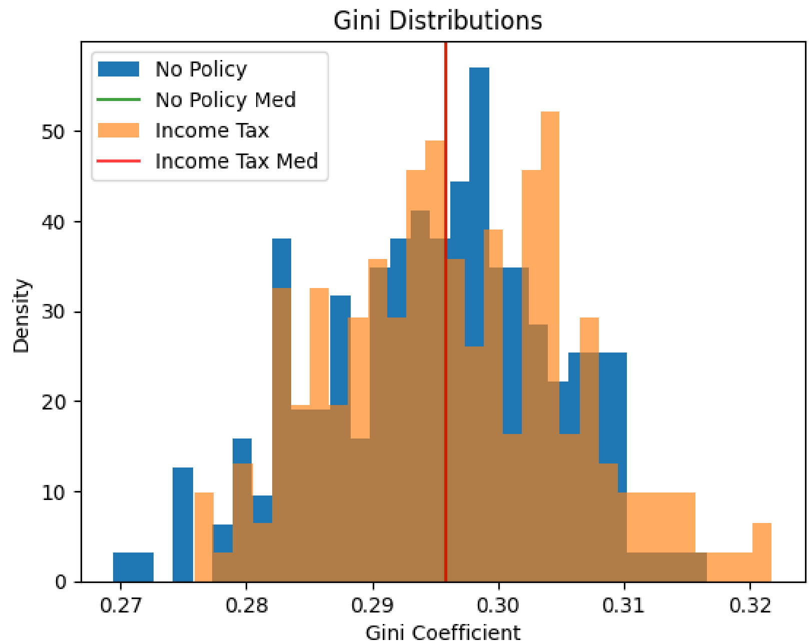


Fig. 6 This chart plots the distribution of Gini coefficients for economies with a progressive income tax and economies with no policies in place. This chart plots the overall distributions in addition to the median of each to highlight the similarities between the distributions.

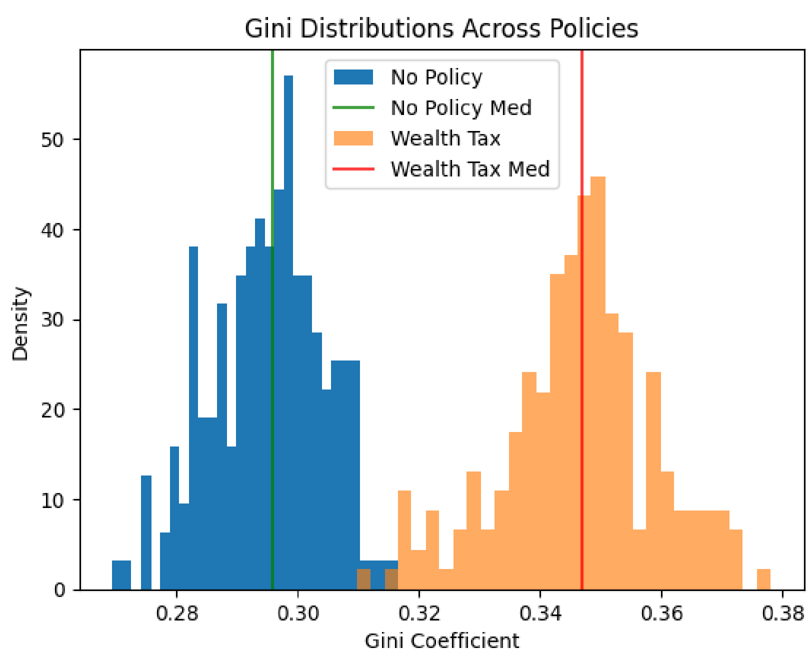
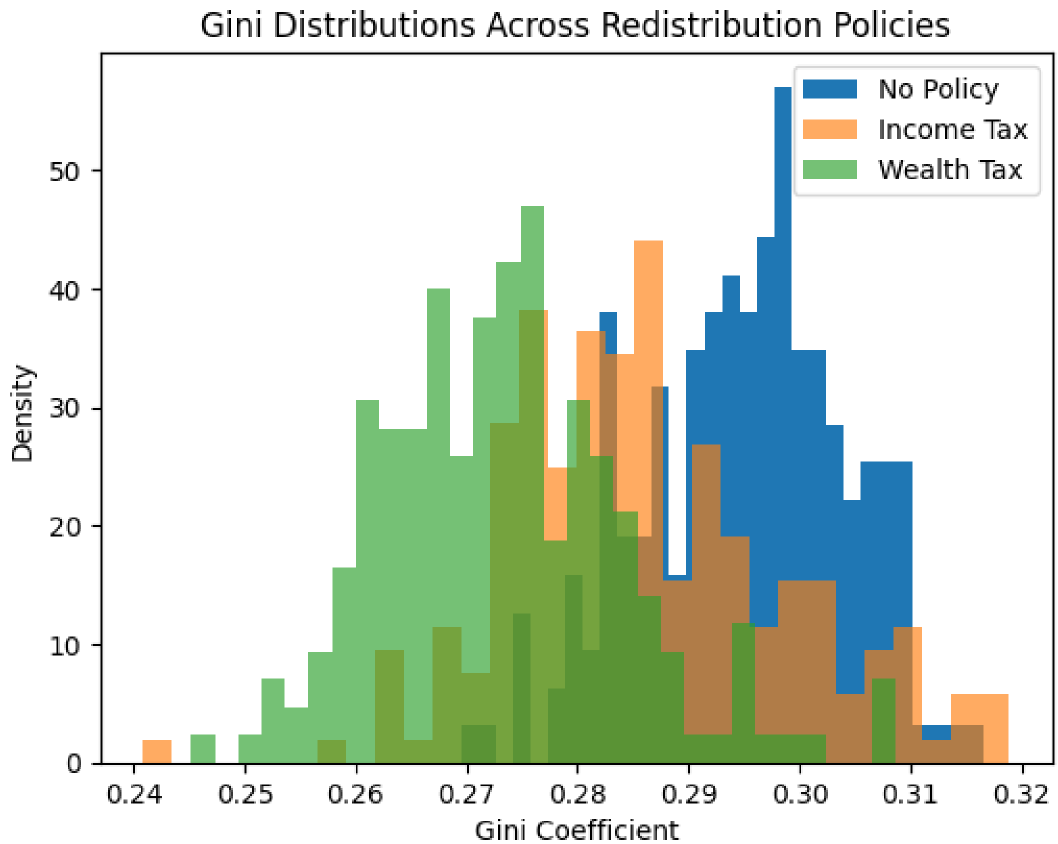


Fig. 7 This chart plots the distributions of economies that have implemented a wealth tax along with those with no policy implemented with the medians of each highlighted in red and green respectively.



Income Tax Average Treatment Effect

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.0090	0.001	7.820	0.000	0.007	0.011

Wealth Tax Average Treatment Effect

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.0218	0.001	21.260	0.000	0.020	0.024

Fig. 8 This figure displays the two redistribution tax policies in addition to the no policy economies along with the fixed effect estimate for each policy below the chart including errors and confidence intervals.