Population (Back) Projections and Scarring Events

Limits to Prediction: Final Assignment

Nick Kaashoek, Christina Pao, and Andrew Zhao

April 2024

1 Abstract

The most common models for demographic population projections often use contemporary population rates as the best estimate for rates in the future. However, we know from history that the key population rate inputs—i.e., for births, deaths, and migration—are constantly in flux. Moreover, we know that there are *specific* times in the history of the human population that are uniquely poor representations of the average rates of a certain population: famines and wars, for instance, can create unrepresentative pictures of the overall mortality and migration rates of a population. These events are sometimes called "scarring" events, since they can leave a 'mark' or 'scar' on the "population pyramid" (i.e., the distribution of individuals in a population by sex and age). That being said, these 'scars' are not permanent: Populations, after rebounding to a state of quasi-stable rates, are thought to quickly "forget their pasts" and can often, within a few generations, have a fully new and unrecognizable population pyramid. In the case of scarring events, standard deterministic demographic methods for projections fall apart: The population rates during that event cannot be used to give a good estimate of the population in the future, and the future population can forget its past.

In this project, we study two interconnected research questions: 1) How quickly does a population forget its past? And 2) knowing a present population, can we back out a population scarring event? We study scarring events because these are moments where demographic methods often fail; using machine learning, we hope to see if we are able to study these questions without being reliant on deterministic projections forward using rates at the time of the scarring event. We select three case studies to explore these two questions: Rwanda, Cambodia, and Russia. For the first question, we project the population forward from the time of the scarring event with our globally-trained population forecasting model to evaluate how quickly the actual (scarred) population begins to diverge. For the second question, we start with what we know from the population today, and we attempt to project backward to see how our models believe the population looked at the time of the scarring event and before. In other words, we are interested in whether models can pick up on residualized evidence of scarring, or whether these scars are completely forgotten in the eyes of machine learning models. In all, we find that our forward (backward) projection methods worked well for Cambodia and Rwanda (with two shorter timescale scarring events)—though with lead (lag) time around the predicted scarring event depending on the direction of the projection. We conclude with reflections on the process of predicting population size from scarring using machine learning.

2 Background

2.1 Introduction to Demographic Population Projections

In demography, there are only five numbers that one needs to know to estimate a population size at time point T. Represented by the 'Demographic Balancing Equation,' the population $N(T)$ is calculated as follows (Preston, 2000):

$$N(T) = N(0) + B[0,T] - D[0,T] + I[0,T] - O[0,T]$$

where

$N(T)$ = number of persons alive in the population at time T

$N(0)$ = number of persons alive in the population at time 0

$B[0,T]$ = number of births in the population between time 0 and time T

$D[0,T]$ = number of deaths in the population between time 0 and time T

$I[0,T]$ = number of in-migrations in the population between time 0 and time T

$O[0,T]$ = number of out-migrations in the population between time 0 and time T

For this reason, much of demography is focused on these three components of change: namely, births, deaths, and net migration. And, due to the relative ease of disaggregating these aforementioned components of population change, the methods for predicting the population forward and backward are relatively straightforward. As long as we have good counts of the population at times 0 and T, we may then be able to know what the feasible *ranges* of potential population sizes could be in the future (and potentially in the past).[1] Nonetheless, it is often difficult to discern within these ranges of potential population sizes what the *actual* population will be in the future. Population projections, when done with traditional demographic methods (e.g., the cohort-component method—see more below), often use the contemporary rates (e.g., of births and deaths) to project forward how the population will look in the future. However, we know that these population rates are unstable and often subject to change.

Take the example of the demographic transition—one of the most widely studied population phenomena (Bongaarts, 2009; Lee, 2003). The demographic transition is defined by the rapid *changes* in first mortality and then fertility rates. All countries in the world have undergone the demographic transition, but the exact pace of the transition and its diffusion across countries was unpredictable: Africa has been said to have undergone a "unique" transition (that was characterized by very different paces and rates of fertility decline) (Bongaarts, 2017, S1), and the *diffusion* of fertility decline in Europe (at the start of the demographic transition) seemed to spread more erratically than was initially hypothesized (Coale, 2017). Like what we read about the unpredictability of social media cascades (Martin, Hofman, Sharma, Anderson, & Watts, 2016), there seemed to be something somewhat unpredictable about how the fertility transition (i.e., the change

[1] i.e., Due to other characteristics of the human population, we often have a good upper and lower bound about population change in some future time point T due to properties we know about humans. For instance, we know that the fecundity period for people assigned female at birth is often limited between teenage-years and mid-life, and there are specific birth-spacings that are more or less possible. People inevitably die at the end of their lives, and there is an upper bound for the time of death in known human history. The number of immigrants into a population is bounded, at the maximum, by the rest of the world population outside of the specific population of interest, and the maximum outmigration is the size of the population of interest, itself.

in fertility rates) would spread across neighboring countries; though so-called classical theory anticipated that *within* country mechanisms, like industrialization and urbanization (Notestein, 1953), would set off the process of population growth change, later theories recognized that there seemed to be some inexplicable effect *between* countries (Bongaarts & Watkins, 1996; Mason, 1997). Though many other *theories* have since proliferated about the mechanisms of changing population rates, *methods* have stayed relatively stable.

The cohort-component model of projecting forward the population is still one of the most widely-used methods and is still used for US Census Bureau projections today; using just life-tables (i.e., the distribution of the population broken down by age and sex) and age-specific fertility rates for females, the population is projected forward deterministically at set-intervals. For example, today's population distribution (at T_0) could be projected forward 15 years based on the fertility rates of today's 15-25, 26-35, 36-45, and 46-55 year-olds. That population distribution (at T_{15}) would then be projected forward using the *same rates*. This could continue on as such.

Nonetheless, as can be imagined, these cohort-component models do not work very well beyond a short time-frame into the future. The furthest forward that the Census has published projections out to using the cohort-component method is around 75 years (less than one generation's life expectancy today!), and one of the declared weaknesses of the method is its assumption of a "homogenous population, with no allowance for the changing internal composition" (Wilson & Rees, 2021). This is particularly problematic when the internal composition of a population changes in a sudden and/or unpatterened way—for instance, with a war, a new national policy, famine, and/or a disease/epidemic. These events, known as scarring events, can change the composition of the population at the time, and the corresponding fertility, mortality, and migration rates needed to project the population forward. This is where traditional demographic methods fall apart: As stated by McCaa, "Crisis mortality, due to epidemics, famine or other demographic catastrophe, is poorly approximated by model life tables" (McCaa, 2001). In the following, we describe these scarring events in more detail before discussing the intervention of machine learning.

2.2 Population Pyramids and Scarring Events

"Scarring" events are often called as such because they make a mark on the "population pyramid." A population pyramid is a sex- and age-disaggregated way of visually representing a population distribution, like seen in Figure 1 below. The "indents," as stated in the annotations in Figure 1, are shown to be areas on the population pyramid that experienced heightened mortality rates. And, as time moves on, these "indents" (or scars, in the case of large increases of mortality) move up the population pyramid. This can be seen from the real-world case of France's population pyramid after the first World War in Figure 2.

As can be seen in Figure 2 from Pison (2014), the panel from the top left (January 1, 1914) has an "indent" at age 42 corresponding to "the birth deficit due to the Franco-Prussian war of 1870-1871, aggravated by a sudden increase in infant mortality: some 23% of babies born in 1871 died before age one, versus 17 % on average in the second half of the nineteenth century"; further, "a very hot summer" in 1911 "produced another infant mortality peak due to severe infantile diarrhoea, and this explains the dent in the 1914 pyramid at age 2." Nonetheless, these minor indents are nothing similar to the larger ones to coming from the 1914-1918 war and WWII. These time periods were ones that were not only marked by higher mortality

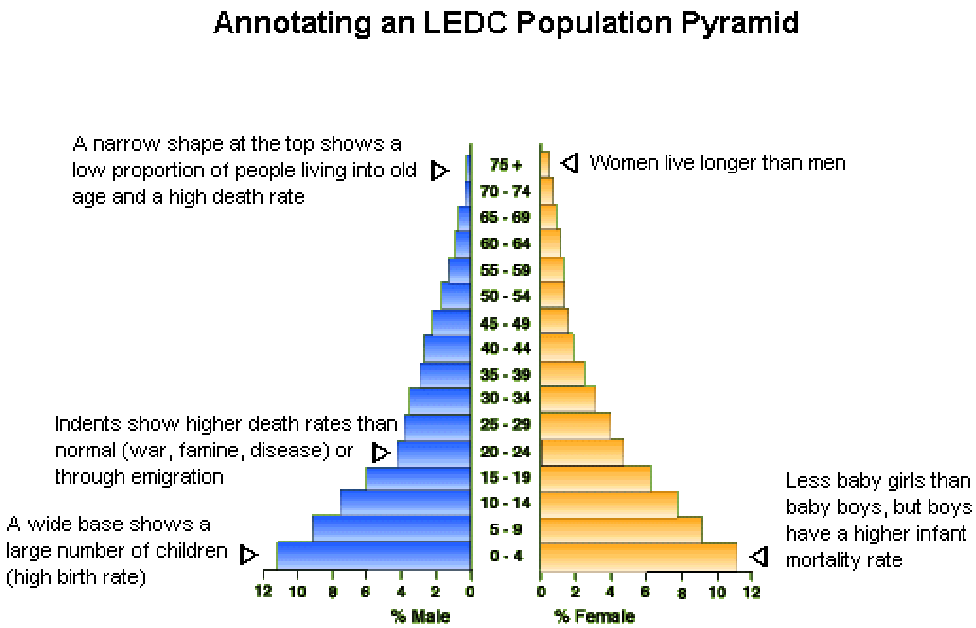

Figure 1: Annotated Population Pyramid, https://www.buddinggeographers.com/population-pyramid/

(particularly for males of military age), but also lowered fertility (i.e., smaller birth cohorts).

These types of population scarring events are, unfortunately, not uncommon and have been seen throughout the course of history. Nonetheless, despite the fact that these scarring events happen often, current population methods primarily rely on the same philosophies of standard demographic models (e.g., life tables and cohort-component methods) 1) to reconstruct hypothetical versions of the population as we might have expected it *without* the scarring event (e.g., calculating "excess mortality") and 2) to understand the population moving forward using estimated rates from that period through a process of *back* or *inverse* projection. This type of analysis was done in the case of Cambodia by Neupert and Prum (2005). To reconstruct the demographic history of Cambodia, the authors 1) analyzed the scars in the population pyramid stemming from increased mortality occurring during the reign of the Khmer Rouge, 2) estimated major demographic events during the 1970s, and 3) compared the size and composition of their simulated "normal" population to the real population. The authors published this paper at the turn of the millennium, and only projected the population forward to 2020 using both standard methods and simulation, based on their back projection methods. Nonetheless, even their standard model (which was supposed to accurately project forward Cambodia's population as it exists with its scarring event), was off by 2.3 million people, or about 14 % of Cambodia's overall population, in relation to the actual population of Cambodia in 2020. In the following section, we discuss this process of demographic back projection and why we believe the current methods may have failed for projecting the Cambodian population forward accurately after the scarring events.

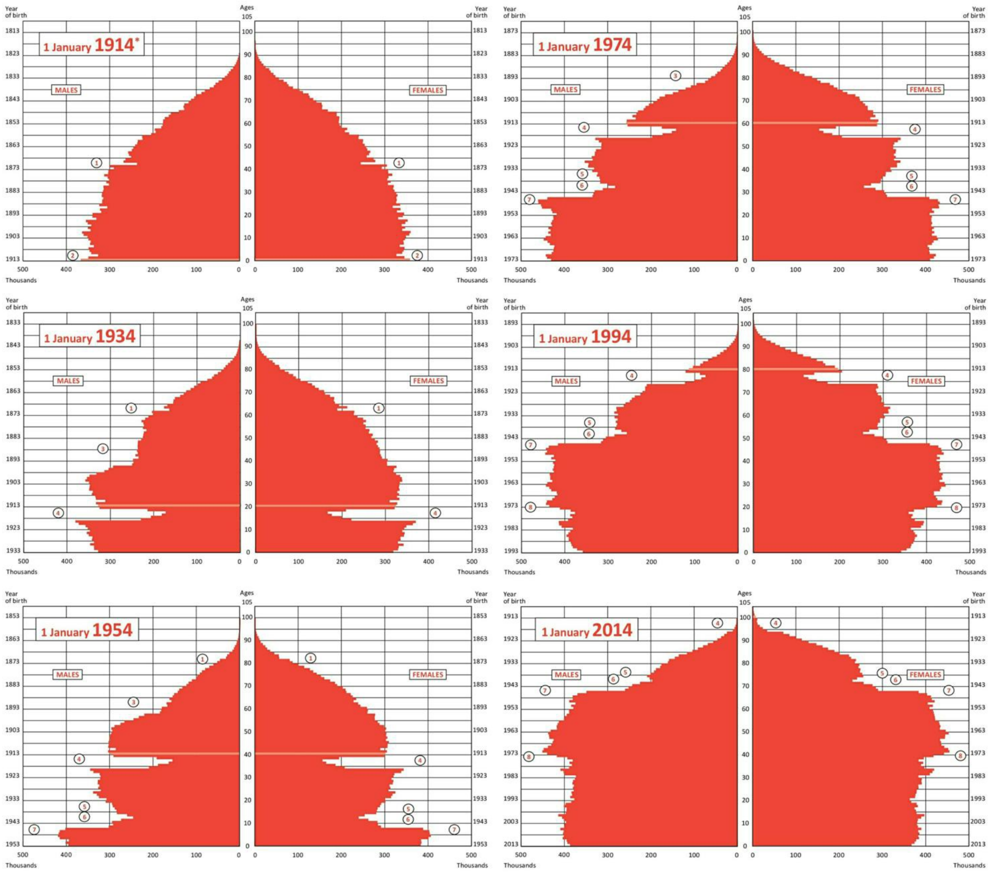

Figure 2: French Population Pyramid, (Pison, 2014)

2.3 Understanding Demographic Inverse/Back Projection: "Vital Rates, Not Initial States"

As stated before, demographic methods perform forward projection by taking the population rates of a certain population pyramid and age that population into the future. However, in times that there are not good estimates for rates (e.g., when data are sparse in historical data), there are a series of methods used for "inverse" and back projection. Inverse projection is seen as the "logical inversion of conventional population projection techniques" since it infers demographic statistics from crude data (McCaa, 2001). To start an inverse projection, one must start with an initial age structure and then apportion deaths, births, and migration: In the absence of a known initial age structure, scholars have shown that "even an arbitrarily chosen model age structure yields remarkably robust estimates"—leading to the broder "mantra" in inverse projections: "vital rates, not initial states" (McCaa, 2001). Due to the principal of 'weak ergodicity' in demography (Preston, 2000), which states that "the age distribution of a closed population is asymptotically independent of the initial distribution" (Inaba, 1989), populations are thought to quickly "forget their past." As seen in Figure 3, taken from Cohen (1979), populations with vastly different starting age structures (such as Thailand in 1955 and Eastern Germany in 1957) can look the same given the same hypothetical vital rates within 100 years. Given the importance of understanding population rates, the question remains: what happens in times where these rates experience shocks (e.g., with scarring events)?

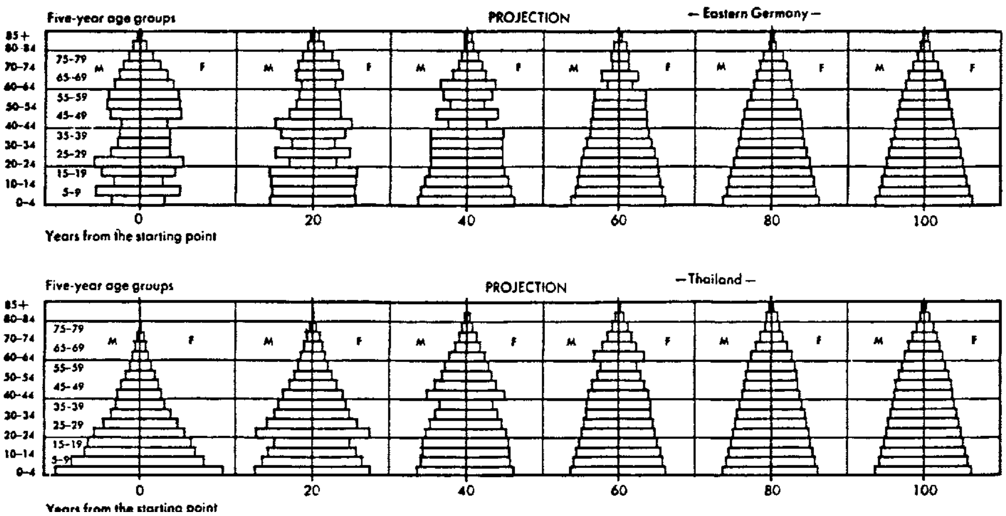

Figure 3: Caption from Cohen (1979): "Two sets of projections computed on the basis of the population of Eastern Gennany in 1957 and of an estimate of the population of Thailand in 1955, respectively; age distribution by five-year age groups. M * male; F = female. Hypothetical vital rates used in both projections assume an expectation of life at birth for both sexes of 60.4 years and a gross reproduction rate of 1.50. Source: Bourgeois-Pichat 1968"

Many of the best methods for dealing with changes in these sudden changes in rates is essentially expert guessing. A summary of the methods most commonly used for crisis mortality is stated in McCaa (2001) as follows (our own emphasis added in **bold**):

> Crisis mortality, due to epidemics, famine or other demographic catastrophe, is poorly approximated by model life tables. **For the inverse projectionist the alternative is to either apply a special set of empirical or model age-specific death rates for the crisis years or fix a crude death rate ceiling or threshold to mimic the impact of crises on the population age structure** (Bonneuil 1990). In crisis years, mortality domain interpolation will assign too many deaths to age groups where variability is greatest, principally for infants, children and the oldest or open-ended age group. The optimal solution is to use empirically derived age-specific death rates peculiar to the type of crisis (Bertino and Sonnino, 1995; Rosini, 1996). Less desirable, although widely used due to the scarcity of age-specific data for these situations, is to set a mortality threshold, say a crude death rate of 40 per thousand population (Bonneuil, 1990; McCaa and Vaupel 1992; McCaa 1993). Deaths below the threshold are apportioned according to standard inverse methodology. Any excess is distributed at a flat rate for each age group. **Although an arbitrary fix, the effect is to produce more robust estimates than that provided by the standard allocation algorithm.**

This then creates an interesting challenge for demographers. Presently, we often have decent estimates of population sizes (i.e., "initial states"), but under certain circumstances like scarring events, we might not have good estimates of "vital rates." Nonetheless, it is these vital rates that are key for determining the future. Vital rates have already been extremely fickle over time, as shown by the demographic transition, and this is particularly heightened when populations undergo traumatic events (e.g., increased mortality rates for certain age groups or a decrease in fertility due to sex-ration imbalances, etc.).

In the following project, we attempt to explore two interconnected issues that may help decompose some of the unpredictability of a scarred population pyramid. Specifically, we are interested in empirically testing two parts of the assumption of weak ergodicity—i.e., the idea that a population undergoing specific rates can "forget its past" overtime, despite an initial scarring event—with machine learning by studying the time horizons of "forgetfulness": In one case, we are interested in projecting forward from the time of the scarring event (i.e., moving from T_0 and projecting forward), and in the other case, we are interested in projecting backward from T_n back to T_0 to see whether we can back out the impact of the scar on a population pyramid. In this case, we want to move beyond standard demographic methods, which we have seen are often poor at determining the vital rates that project forward, such as in the case of Cambodia (Neupert & Prum, 2005). For this, we use basic machine learning tools to try to avoid the pitfalls of standard demographic population projection methods which necessitate the use of well-known (and at least quasi-stable) demographic rates.

3 Our Intervention: Machine Learning for Population Back Projection

Our general approach to the project was to use both forward and back prediction in the context of demographic scarring events. Using forward prediction, we attempted to answer the question of how long it takes for a population to "forget" about a scarring event; with back prediction we tried to determine the initial state of a population *prior* to the scarring event using data from after the event.

In the following sections, we discuss our case studies, the data we used to train the models, the specific models we used for our evaluation, and the results of our evaluation.

3.1 Case Studies

In order to leverage the richness of modern data, we selected three countries as case studies that each experienced demographic scarring events in the relatively recent past. The Russian Federation, Rwanda, and Cambodia all fit this criteria: Russia experienced a dramatic population decline in the 1990s and early 2000s, driven by a variety of factors including declines in both birth and immigration rates (Eberstadt, 2004); Rwanda in the 1990s experienced a series of armed conflicts and a civil war punctuated by a genocide against the Tutsi population in 1994 (Forges, (Organization), & internationale des droits de l'homme, 1999); and the brutal Khmer regime in Cambodia killed nearly a quarter of the country's population (Kiernan, 2003).

Each of these events had a notable impact on the population of the country. From 1996 to 2007, the population of Russia decreased by more than 6.3 million people — just over a 4% decrease in total population. The decline was steady: no single year had an outsize contribution to the final result. For comparison's sake, the rest of the world experienced a marked increase of more than a billion people during the same time period.

Rwanda, from 1990 to 1995, lost $\sim 1,000,000$ citizens to the combined effect of all the conflict happening in and around the country: more than a 14% decrease. The year 1994 accounts for a large portion of this: close to three-quarters of a million Rwandans were killed during the genocide and civil war that took place during that year. In this time period, the world population increased by more than 400 million.

The Khmer regime, which ruled Cambodia from 1975 to 1979 carried out a systematic program of persecution and execution that resulted in the deaths of a million or more Cambodians. The dataset we used is more conservative in its estimate than other sources we read are: it shows a decrease of close to nine-hundred thousand. Still, this represents more than 11% of the population in 1974. Much like the Russian case study, no single year accounted for a majority of the decrease, although the trend did taper off toward the end of the Khmer regime. The rest of the world increased in population during this time period as well.

3.2 Data

We combined data from two sources for our analysis. First, we used the International Database (IDB). This dataset, collected by the United States Census Bureau, reports the population of each country in the world as well as some summary statistics about birth rate, death rate, and other factors related to population growth and decline. The data are commonly used to build population pyramids similar to the ones presented earlier in this paper. Additionally, Census includes projections of populations up to the year 2100 in the dataset; we ignore these projections, and only examine the segment of the dataset covering the years 1960 to 2023. We chose the year 1960 as a starting point as it gives us a reasonable runway up to each of the scarring events, and matches the start of the other dataset we used.

As the International Database is commonly used by demographers and other scientists, we believe it to be reliable. However, we did notice some conflicting values in it when compared with other research and estimates about the countries we selected for our case studies. As mentioned above, the measurements presented for Cambodia seem on the low end. As any population measurements during that time period are estimates at best, this can be explained simply by a conservative estimate on the part of the Census Bureau. Still, we felt it important to mention explicitly, even though we do not believe it had any serious impact on our results. On the other hand, the numbers for Rwanda seemed on the high end of reported estimates, which range from 400,000 to 800,000; the IDB reports a decline of 744,339 during 1994. Again, we don't believe this affected or altered the conclusions we reached as part of our research, but believe it is worth mentioning.

To augment the IDB with further covariates, we turned to the World Bank's World Development Indicators. This dataset includes more than 1400 variables representing indicators of development on a country-by-country basis. Examples include how much of the population has access to clean water, what kind of natural resources are available, and what the average income in a country was. Downloading the entire dataset was impossible, as the online portal errors on too large of a request. Instead, we selected a sample of 155 indicator variables covering individual financial indicators, health indicators, and national-level financial indicators. Please refer to the dataset included in our Github Repository for more details.

Like the IDB, we deemed the data provided by the World Bank to be reliable. Unfortunately, we had fewer options for cross-referencing the data with external sources, so we cannot characterize any potential discrepancies with detail. However, given the source, we believe it to be fit for our analyses.

To combine the datasets, we performed some basic cleaning on the World Bank dataset. This was not necessary for the IDB, as it is already complete. We removed covariates where more than 10% of the data was missing, manipulated the dataset to match the IDB format, and then merged the two together on country and year. The code for performing these steps can be found in `analysis.py` and `join.py`. Documentation for using each of these scripts can be found in our README file. Finally, we used min-max normalization on the resulting dataset to ensure that each column was on the same scale. To avoid leakage, we removed Russia, Rwanda, and Cambodia from the dataset for training.

3.3 Method

We tested out three types of regression models for time series: a decision tree, a random forest regressor, and a gradient boosted regressor. All three are tree models: they construct a flowchart that automatically processes the data. Tree models are state of the art when predicting tabular or time series data. The main difference between the models is the number of trees used and how the trees are trained: a decision tree only creates one tree, while the random forest and gradient boosted trees generate multiple trees. The difference between the two is that random forests aggregate individually trained trees, and gradient boosted trees modify the loss function after training one tree.

Concretely, our task for forward prediction is to predict the population at a time $t+1$ given the existing yearly data and predicted population data at time $t, t-1, t-2$. In contrast, our task for backwards prediction is to predict the population at time $t-1$ given the existing yearly data and prediction population at time $t, t+1, t+2$. By autoregressively utilizing the outputs of the model as inputs for earlier times, the regression models allows time series forecasting.

We used scikit-learn's implementations of all 3 models in our prediction. We trained on the entire dataset after removing Russia, Rwanda and Cambodia and then autoregressively generated backpredictions for all 3 countries.

All the code we used to train and test our models can be found in `model.py` in our Github repository. Documentation for this file can be found in our README as well.

3.4 Findings

To evaluate each of the regression models, we trained them on the global, joined dataset produced by merging the IDB data with that of the World Bank. Prior to training, we removed the three countries we were interested in from the dataset to avoid leakage. Rather than separating the data into a dedicated train and test split, we used the data from the rest of the world for training and our three case studies as the test set.

After training, we used the models to produce a population prediction for each of our case studies, which we compare with the actual data withheld from the training data. The results for forward prediction are shown in Figure 4 and Figure 5. We applied the same methodology for back predictions, altering the way the data is fed into the models to flip things so that past years are unknown rather than future years. These results are shown in Figure 6 and Figure 7. The "centered on scarring" graphs, Figure 5 and Figure 7 show the same results as the other graphs, but present only the subsection of years that capture the scarring periods we are interested in. We use them to highlight differences between the models and actual data around the

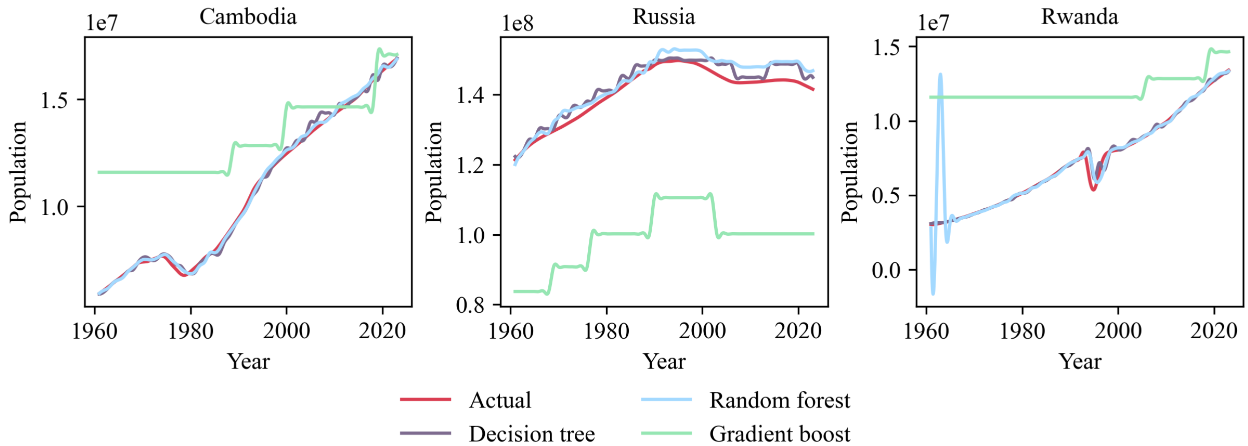

Figure 4: Forward Prediction Results

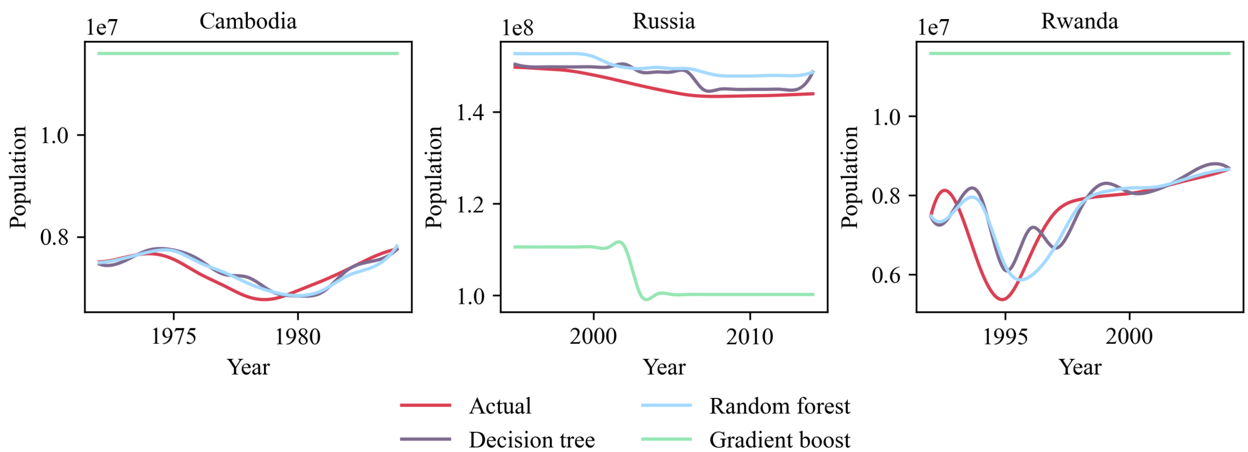

Figure 5: Forward Prediction Results (centered on scarring)

scarring specifically.

We draw similar conclusions from both sets of graphs. First, we note that the performance of the gradient boost model is, in comparison to other two regressors, notably poor. The mean squared error for gradient boost is multiple orders of magnitude worse than that of the others. This difference is particularly pronounced in the case of Rwanda, where gradient boost fails to account for *any* change in population up until the early 2000s. The poor performance likely stems from a lack of optimization on our parts; as the other two models performed relatively well, we chose not to pursue this task for the purposes of the project. For the rest of our findings, we ignore the Gradient Boost model.

Random forest and decision tree both perform relatively well on the data. They capture the general trends of population growth and decline for all of our case studies, although there are some strange characteristics present in each of the predictions. For one, the models produce a more "jittery" (for lack of a better term) population curve than is present in the actual data. This is likely due to the limited number of points of prediction: data is only present from 1960 to 2023, limiting the number of predictions we can make. This

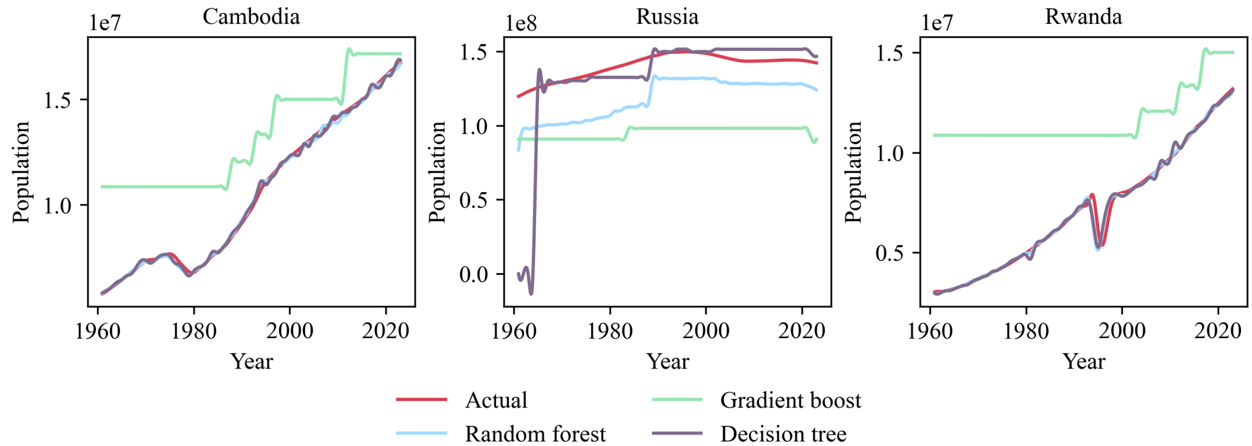

Figure 6: Back Prediction Results

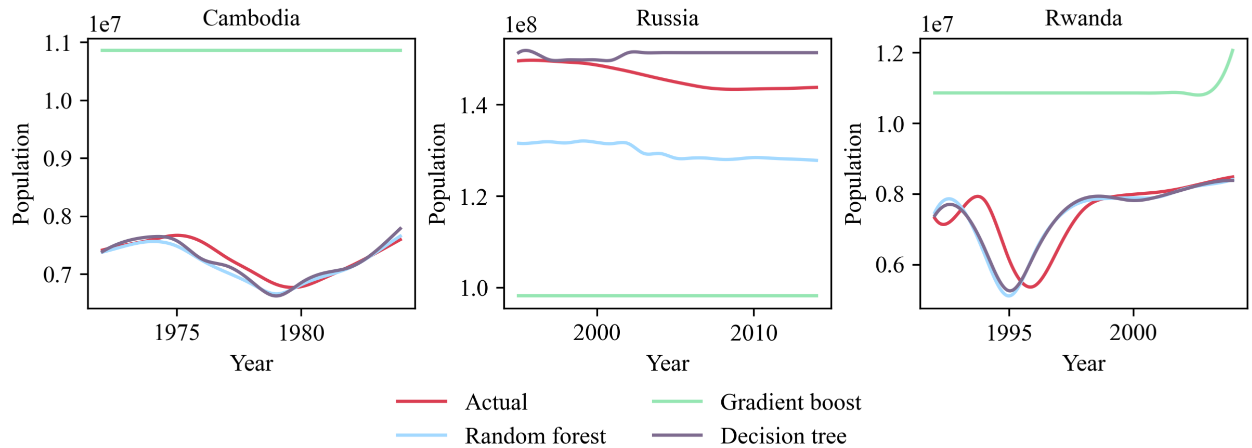

Figure 7: Back Prediction Results (centered on scarring)

likely explains the odd spike in the random forest's prediction for the early sixties on the Rwandan case study — we believe this is just noise.

Comparing model performance on forward and backward prediction reveals an interesting fact: the predictions are nearly identical, with certain characteristics flipped depending on the order of the prediction. Note that for both Cambodia and Rwanda, our models are slightly off in when they predict the population decline to start. For forward prediction they are too late, and for back prediction they are too early. In effect, this is mirrored: being late on forward prediction and early on back prediction are functionally the same thing.

We turn now to the scarring periods in each case study. At a high level, the two models that perform well on the dataset (i.e., the decision tree and random forest) do markedly better on the Cambodian and Rwandan case studies than on the Russian one. This is somewhat surprising: the Russian scar takes place over a long period of time, and is more gradual than the other two. At the start of our research, we believed that the models would perform *better* on this case study than the other two, as long, gradual trends tend to be handled well by regressors. The opposite was true in this case. Our best explanation for this odd behavior is that the covariates for Russia provide a more muddied perspective on the state of the country than they do for Cambodia and Rwanda: both of those nations experienced massive economic downturns and rises in violence, indicators not present in the Russian data. The other hypothesis is that our globally-trained model may give us better estimates of population growth rates for countries that experience a one-off shock and otherwise resume to quasi-stable rates; in the Russian case, the longer time-scale of rate change may be somewhat anomalous in comparison to what the model was trained on.

For the Cambodian and Rwandan case studies, the two performant models accurately captured the general trends of population rise and decline. However, they exhibit a curious “lag” in both cases — regardless of prediction direction, the actual data always starts declining earlier (in terms of prediction direction) than the models do. The shape of the decline is consistent across models and actual data, so this is intriguing. While we can’t be sure exactly why the models exhibit this trend, we believe it likely comes down to the covariates as well: macro-scale economic factors such as GDP can change in response to population shifts. As the models rely on these factors to make their predictions, they are slower to predict the scarring than is accurate.

The most surprising result, however, is how quickly the models return to alignment with the actual data after the scarring concludes. Our hypothesis at the outset of the project was that they would be slow to adjust — this would suggest that populations hold their scars for a longer period of time. Considering the cultural impact of events such as the Khmer regime, this made logical sense to us. However, the machine learning models suggest that this isn’t the case: within just five years of the scarring event’s conclusion, they return to alignment with the actual data and resume the general trajectory exhibited by the data prior to the scarring event. This is a novel finding: Our model was able to (with a window of error) essentially realign with actual population numbers after a scarring event. Whereas other demographic projections have struggled to create future alignment after scarring, even on short time-scales—e.g., Neupert and Prum (2005)—we have shown that machine learning models may be able to quickly project forward the population using global data. The true test of this would be if we could use these models to predict the effects of population scarring events that are happening now on future (unseen) data on populations. We discuss this more in our conclusion section.

Overall, our results suggest that populations do, in fact, “forget their past” quickly from their scarring events in demographic terms, and, importantly for our research question, populations recover in a somewhat predictable way. We emphasize the use of demographic terms here — our intent is not to suggest that these countries have wholesale “recovered” from the terrible effects of their respective scarring events, just that the changes in their population *rates* have. By the year 1982, the Cambodian population was growing at close to the same rate it was before the Khmer Rouge seized power, and with a similar trajectory as the rest of the world. The same is true of Rwanda, although we cannot make any such claims about Russia. The slower decline has not truly slowed down by the current time, no doubt thanks to the conflict with Ukraine and other associated factors. As the scarring event is, to some extent, still ongoing, it is impossible to say how long it will take Russia to recover. In all, our exploration shows that population projections, at least for shorter timescales, may be feasible on a globally-trained model even when populations experience a scarring event.

More explicitly expressed to our research questions, we show that two things can exist simultaneously: a population can "forget its past" relatively quickly, but it is still possible to roughly back out a population scarring event—albeit with some lags and leads around our own model estimates.

4 Conclusion

To conclude our project, we discuss the impact of our findings on the two research questions we set out to address, as well as some limitations to our findings.

4.1 Discussion

For this project, we evaluated two main research questions: 1) How quickly does a population forget its past? And 2) knowing a present population, can we back out a population scarring event? In doing so, we find that our machine learning models (with the exception of the gradient boost) tend to generally perform well in predicting population trends forward and backward for Cambodia and Rwanda—two countries that experienced a shorter-timescale scarring event. In these cases, we find that there is a couple year lead (for forward projections) or lag (for backward projections) around the scarring period itself, where our models are slightly off on the exact timing of the scarring event; we were surprised to see that these models were relatively invertible and seemed to work similarly well forward and backward. Nonetheless, for Russia, our models struggled with creating accurate predictions. In part, we believe that this is due to a combination of the timescale of the scarring event and the particularities of the event; because the scarring event took place for over a decade and had a less dramatic impact on covariates, we believe that the models struggled to pick up the changes in *rates* over a longer period of time.

With just publicly available data and simple models, we show that relatively simple machine learning approaches can create reasonable population forecasts for populations that experienced scarring events. This adds to our understandings of what is (and importantly, isn't) possible in applying new methods (i.e., machine learning) to an old problem (i.e., changing rates) in standard demographic tools. We use the example of demographic scarring, since these are moments that demographic models explicitly fall apart—specifically due to the changes in rates that occur *after* the population scarring takes place. Instead of using expert-guessed rates or a model-life table, we show that globally-trained machine learning models can help predict rates of population increase over time after scarring can take place—at least for populations that experienced scarring events on a shorter time-scale.

4.2 Limitations

We have several limitations to note. First, we only chose three country case studies for which we had relatively recent and well-founded data. Given our findings with Russia and our hypothesis about longer-term scarring events, we would have wanted to test further country case studies on a range of different time-scales.

Second, we used data from two well-known sources, and our choices for covariates were theoretically (and logistically) motivated but nonetheless somewhat arbitrary. If we were doing this project as more than a pilot, we would want to be more rigorous about feature selection.

Third, though we trained our models on global covariates excluding Russia, Cambodia, and Rwanda, we included the country of interest's (i.e., Russia, Cambodia, or Rwanda) covariates as part of the test prediction. Though this may be considered leakage, we note that the covariates that are included do not necessarily make the model better at predicting population *rates* (since that's not what it's trained to do). However, it is these rates that are the most interesting often for making traditional projections into the future with demographic methods. In other the words, we could have imagined a world in which this addition of Russia, Cambodia, or Rwanda's covariates in the model would not have increased its ability to produce good projections into the future given their scarring events. We discuss more what future would should do in the following section.

4.3 Future Work

Potential extensions to this project include: 1) analyzing the relative importance of predictors in our model, 2) training models that do *not* include the covariates of the country of interest as predictors of the population, and 3) creating predictions from our model for present-day scarring events to test the utility of our model in forward projection. For the first extension, we note in our data section that we chose ~ 150 features from the World Bank's database—in part due to their theoretical importance, and in part due to the limitations of the download size. We wonder what our models will have looked like if we had included more covariates (e.g., all available covariates from the World Bank) or no covariates at all besides population rates from other countries (i.e., just what was contained in the IDB database). We would be interested in seeing *which* combination of covariates seem to best predict populations in general, but also for scarring in particular. Further, we would be interested in seeing how parsimonious we could make the model to have relatively similar predictive accuracy.

Second, as mentioned in our limitations section, we may be interested in how well our models work when the country of interest's covariates are *not* included for producing the prediction. We had considered using imputation methods based on global means or creating some other theoretically-informed way of creating "hypothetical" covariates under a non-scarred population regime, but we believed that this would be subject to the equivalent of random guessing that current back/inverse projections use. In the future, it would be interesting to see how far and how helpful machine learning models would be at producing estimates of rates for determining population growth *without* the use of other country characteristics at the point of time of scarring.

Third, we would want to see whether our model(s) actually work for unseen data—e.g., using the cases of modern-day scarring events such as in Palestine or Ukraine. This may be a pathway to determining the utility of a model of this sort in the future (i.e., when the model is trained on past global data). This would be the best test of whether ML models can truly help with the demographer's future-oriented tasks, but this was outside of the scope of our initial exploration.

References

Bongaarts, J. (2009). Human population growth and the demographic transition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1532), 2985–2990. doi:10.1098/rstb.2009.0137

Bongaarts, J. (2017). Africa's unique fertility transition. *Population and Development Review*, *43*, 39–58. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1728-4457.2016.00164.x. doi:10 . 1111 / j . 1728-4457.2016.00164.x

Bongaarts, J., & Watkins, S. C. (1996). Social interactions and contemporary fertility transitions. *Population and Development Review*, *22*(4), 639–682. Publisher: [Population Council, Wiley]. doi:10.2307/2137804

Coale, A. J. (2017, March 14). The decline of fertility in europe. In *The decline of fertility in europe*. doi:10.1515/9781400886692

Cohen, J. E. (1979). Ergodic theorems in demography. *Bulletin of the American Mathematical Society*, *1*(2), 275–295. doi:10.1090/S0273-0979-1979-14594-4

Eberstadt, N. (2004). Russia's demographic straightjacket. *The SAIS Review of International Affairs*, *24*(2), 9–25. Retrieved May 6, 2024, from https://www.jstor.org/stable/26999221

Forges, A., (Organization), H. R. W., & internationale des droits de l'homme, F. (1999). *"leave none to tell the story": Genocide in rwanda*. Human Rights Watch. Retrieved from https://books.google.com/books?id=ZlpyAAAAMAAJ

Inaba, H. (1989). Weak ergodicity of population evolution processes. *Mathematical Biosciences*, *96*(2), 195–219. doi:10.1016/0025-5564(89)90059-x

Kiernan, B. (2003). The demography of genocide in southeast asia: The death tolls in cambodia, 1975-79, and east timor, 1975-80. *Critical Asian Studies*, *35*. doi:10.1080/1467271032000147041

Lee, R. (2003). The demographic transition: Three centuries of fundamental change. *Journal of Economic Perspectives*, *17*(4), 167–190. doi:10.1257/089533003772034943

Martin, T., Hofman, J. M., Sharma, A., Anderson, A., & Watts, D. J. (2016, February 2). Exploring limits to prediction in complex social systems [arXiv.org]. doi:10.1145/2872427.2883001

Mason, K. O. (1997). Explaining fertility transitions. *Demography*, *34*(4), 443–454. Publisher: Springer. doi:10.2307/3038299

McCaa, R. (2001). An essay on inverse projection. Retrieved May 3, 2024, from https://users.pop.umn.edu/~rmccaa/populate/ipessay.htm

Neupert, R. F., & Prum, V. (2005). Cambodia: Reconstructing the demographic stab of the past and forecasting the demographic scar of the future. *European Journal of Population / Revue Européenne de Démographie*, *21*(2), 217–246. Publisher: Springer. Retrieved May 3, 2024, from https://www.jstor.org/stable/20164303

Notestein, F. W. (1953). Economic problems of population change. In *Proceedings of the eighth international conference of agricultural economics* (pp. 13–31). London: Oxford University Press.

Pison, G. (2014). 1914-2014: A century of change in the French population pyramid. *Population Societies*, *509*(3), 1–4. Bibliographie_available: 1 Cairndomain: www.cairn-int.info Cite Par_available:

1 Publisher: Ined Editions. Retrieved May 3, 2024, from https://www.cairn-int.info/article-E_POPSOC_509_0001--1914-2014-a-century-of-change-in-the.htm

Preston, S. H. (2000, September 18). *Demography: Measuring and modeling population processes* (Illustrated edition). Malden, MA: WB.

Wilson, T., & Rees, P. (2021). A brief guide to producing a national population projection. *Australian Population Studies*, *5*, 77–100. doi:10.37970/aps.v5i1.84