Accounting for dependencies between fertility and mortality in population projections

Extended abstract submitted to IUSSP IPC 2025

Ever Hughes* Monica Alexander† Radu Craiu‡

Fertility and mortality outcomes are intrinsically linked, both at the individual and population level. However, methods of population projection generally do not consider these dependencies. This project investigates ways to carry out population projection accounting for these dependencies. In particular, we use copula models to account for the correlation between fertility and mortality in the past, and account for these correlations in the projection of the components of population change. Preliminary testing focusing on Sub Saharan African countries suggests that accounting for these dependencies could either increase or decrease the projected population size compared to WPP estimates, depending on the implied differences in the age structure of fertility.

1 Introduction and Motivation

Fertility and mortality outcomes are intrinsically linked. At the individual level, premature death in adulthood may lead to 'births forgone' due to potential parents not surviving through their reproductive years (Polizzi and Tilstra 2024). On the other hand, when child mortality is relatively high, parents may decide to bear more children to ensure a desirable number of surviving progeny (Preston 1978; Nobles, Frankenberg, and Thomas 2015). At the aggregate level, the population-level dependencies between fertility and mortality are clear from the Demographic Transition Theory: as overall mortality declines, fertility declines usually follow

*Statistical Sciences, University of Toronto
†Statistical Sciences and Sociology, University of Toronto. monica.alexander@utoronto.ca.
‡Statistical Sciences, University of Toronto

(Kirk 1996). This is driven not only by individual-level factors, but also shifts in the prevailing macro-level economic, epidemiological, and cultural conditions (Reher 1999). The connection between fertility and mortality from a behavioral and biological point of view has been studied previously, for example, in understanding the fertility implications of a mortality 'shock', such as the HIV/AIDS epidemic (Zagheni 2011) or a natural disaster (Nobles, Frankenberg, and Thomas 2015). Preston (1972) presented a series of mathematical relations to demonstrate the connection between the two quantities.

Despite these well-established dependencies, population projection models tend to treat fertility and mortality as independent. For example, population projections published by the United Nations as part of the World Population Prospects rely on two separate statistical models that project the total fertility rate (Alkema et al. 2011), and life expectancy (Raftery et al. 2013), respectively. These indicators are then disaggregated into age-specific rates and input into a cohort component model to produce population projections.

The aim of this work is to develop methods to account for the dependency between fertility and mortality in population projections. We use copula models to capture past correlations between country-level fertility and mortality, and project these components into the future, accounting for this correlation. This abstract presents some preliminary work and results focusing on Sub Saharan African countries.

2 Initial approach and results

As an initial exploration, we focused on producing population projections taking into account correlations between fertility and mortality in Sub Saharan African countries. Briefly, we estimate the marginal distributions of life expectancy and age-specific fertility rates across the region by decade, model the correlation between life expectancy and each of the age-specific fertility rates using Gaussian copula models, and then forecast the parameters of the marginals and copulae to produce projections of age-specific fertility rates. These are then used as input in a standard cohort component projection framework, combined with existing age-specific mortality projections from the World Population Prospects (WPP), to produce future projections of populations by age. More details are given below.

2.1 Data

We use estimates from the 2024 edition of the WPP (UNPD 2024) as a base for the estimation and as a comparison. Specifically, we use the median estimates of life expectancy from 1950 to 2023, estimates and projections of life table person-years lived (${}_nL_x$) up to 2100, and estimates and projections of the five-year age-specific fertility rates from 1950 to 2023. In this initial exploratory work, we focus on 17 countries in the Sub Saharan African region.

2.2 Outline of approach

We account for the dependency between mortality (as measured by life expectancy) and fertility rates using copulas models. A copula is mathematically convenient because it provides a functional link between a joint distribution and its marginals. Statistically, copulas are useful because they allow the dependence structure to be modeled separately from the marginals.

For the bivariate case, consider a continuous random vector (Y_1, Y_2) with marginal cumulative distribution functions (CDF) F_1 and F_2, respectively and joint CDF F. Set $U_j = F_j(Y_j)$ so that U_j follows a uniform distribution. The unique copula of (Y_1, Y_2), which we denote C, is the CDF of (U_1, U_2) so that

$$C(u_1, u_2) = \mathbb{P}(U_1 \leq u_1, U_2 \leq u_2).$$

The joint CDF can be then expressed as the composition of the copula and the marginal CDFs

$$\begin{aligned} F\left(y_1, y_2\right) &= \mathbb{P}\left(Y_1 \leq y_1, Y_2 \leq y_2\right) \\ &= \mathbb{P}\left(U_1 \leq F_1\left(x_1\right), U_2 \leq F_2\left(y_2\right)\right) \\ &= C\left(F_1\left(y_1\right), F_2\left(y_2\right)\right). \end{aligned}$$

Note that the copula is independent of the marginals, which means that it can be used to model dependence between variables that are in different distributional families.

In this application, we consider life expectancy Y_1 and each of the five-year age-specific fertility rates (age groups 10-14, 15-19, ..., 50-54) as $Y_2, Y_3, \ldots Y_{10}$. Specifically, we consider a separate copula to model the dependence between life expectancy, Y_1, the fertility rate in a given five-year period, Y_j, $2 \leq j \leq 10$. Different copulas are fitted to data collected in each decade. All the copulas used are Gaussian copulas which have the form

$$C\left(u_1, u_2\right) = \Phi_2\left(\Phi^{-1}\left(u_1\right), \Phi^{-1}\left(u_2\right); \rho\right)$$

where Φ is the CDF of a standard normal and $\Phi_2(.; \rho)$ is the joint CDF of a bivariate normal with mean vector zero, all variances equal to 1 and correlation ρ.

The population projection for a particular country was obtained as follows:

1. The marginal distributions for each variable $Y_1, Y_2, \ldots Y_{10}$ were estimated for each decade from 1950 to 2020. For life expectancy we assumed a normal marginal distribution, and for each of the fertility rates we assumed Gamma marginal distributions. Marginals were estimated separately by decade to account for changing mortality and fertility conditions.

2. A separate bivariate Gaussian copula model was estimated for life expectancy and each age-specific fertility rate for each decade from 1950 to 2020.

3. The parameters of each marginal distribution and each copula were projected for future decades, from 2030 to 2100. This was carried out assuming a linear change from 2010. For example, Figure 1 shows the estimates and projections of the mean and standard deviation of life expectancy across Sub Saharan Africa from 2050 to 2100. In general mean life expectancy is increasing and the standard deviation across countries is decreasing (with the exception of the height of the HIV/AIDS epidemic).

4. For each year from 2025 to 2100, 1000 conditional samples of each age-specific fertility rate were generated from the bivariate copula with correlation parameter as projected in

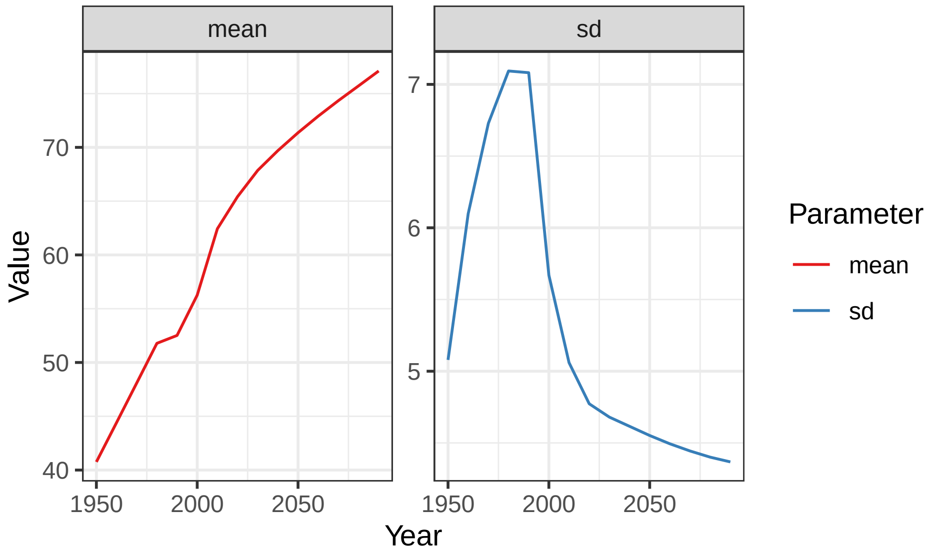


Figure 1: Estimates and projections of marginal normal distribution parameters for life expectancy in Sub Saharan African countries.

step 3, conditioning on the WPP life expectancy projection for the country of interest in that particular year. So for example, when generating fertility projections for Gambia in 2030, we sample from a Gaussian copula with correlation estimated in step 3, conditioning on a life expectancy value of 67.8 years, which is the medium variant projection from WPP. Note that the age-specific fertility rates for each projection year and sample are re-normalized such that the total fertility rate is equal to the total fertility rate projection estimate from WPP.

5. The 1000 trajectories of age-specific fertility rates, ${}_nF_x$, are then used to construct a standard Leslie matrix (Wachter 2014) for the female population in the country of interest for each projection year 2025-2100, along with WPP estimates of person years lived ${}_nL_x$. Here there are 21 age groups $x = 0, 5, 10, 15, \dots 100+$ with $n = 5$. Note that ${}_nF_x$ below age 10 and above age 55 are assumed to be zero. We project the population to 2100, starting with the initial population by age estimated for 2024 from WPP, denoted by $\mathbf{K}(0) = [{}_0K_5(0), {}_5K_5(0), \dots, {}_\omega K_{100}(0)]$. Population projection is carried out using the

cohort component projection method:

$$\mathbf{K}(t) = \mathbf{A}(t)\mathbf{K}(t-1)$$

for $t = 2025, 2030, \ldots, 2100$ where $\mathbf{A}(t)$ is a 21×21 matrix

$$\mathbf{A}(t) = \begin{bmatrix} 0 & 0 & {}_5\tilde{F}_{10}(t) & \ldots & 0 \\ \frac{{}_5L_{10}}{{}_5L_0}(t) & 0 & 0 & \ldots & 0 \\ 0 & \frac{{}_5L_{10}}{{}_5L_5}(t) & 0 & \ldots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \ldots & 0 \end{bmatrix}$$

and the top row non-zero entries are

$${}_n\tilde{F}_x(t) =_n L_0(t) \cdot \frac{1}{2}\left({}_nF_x(t) + {}_nF_{x+n}(t) \cdot \frac{{}_nL_{x+n}(t)}{{}_nL_x(t)}\right)$$

The projections using this approach are compared to projections calculated when using the medium variant projections of age-specific fertility published as part of WPP. Note that for the purposes of this preliminary work, migration as a component of population change is ignored.

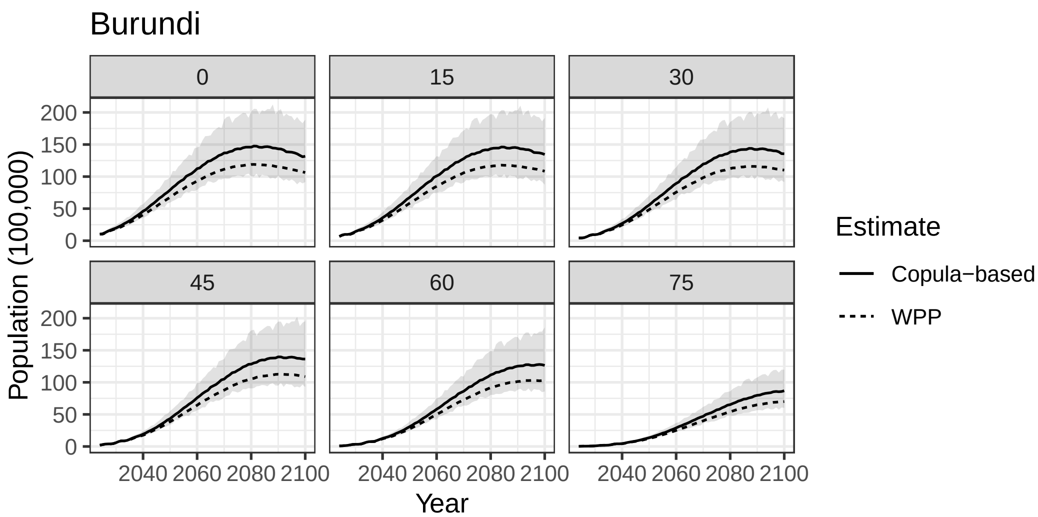
2.3 Preliminary results

The figures below show illustrative preliminary results for two countries, Burundi and Zimbabwe. Burundi has relatively high fertility (around 5 births per woman), whereas Zimbabwe has relatively low fertility for the region (around 3 births per woman); both have female life expectancy of around 65 years. For Burundi, the copula-based projections of population are higher than for WPP, with the discrepancy being larger and starting earlier for younger ages (Figure 2a). This has implications for the maximum population sized reached, as Burundi's population is projected to decrease from around 2080. The differences in projections lead to slightly different population age structures (Figure 2b); particularly in 2040, the copula-based methods imply a greater share of younger people. In Zimbabwe, the situation is reversed: the copula-based projections of population are lower than for WPP, although the difference is smaller than for Burundi (Figure 3a). This leads to a population pyramid that is slightly older (Figure 3b). Note that total fertility rate was standardized to be the same as WPP so differences are purely reflective of differences in the age structure of fertility.

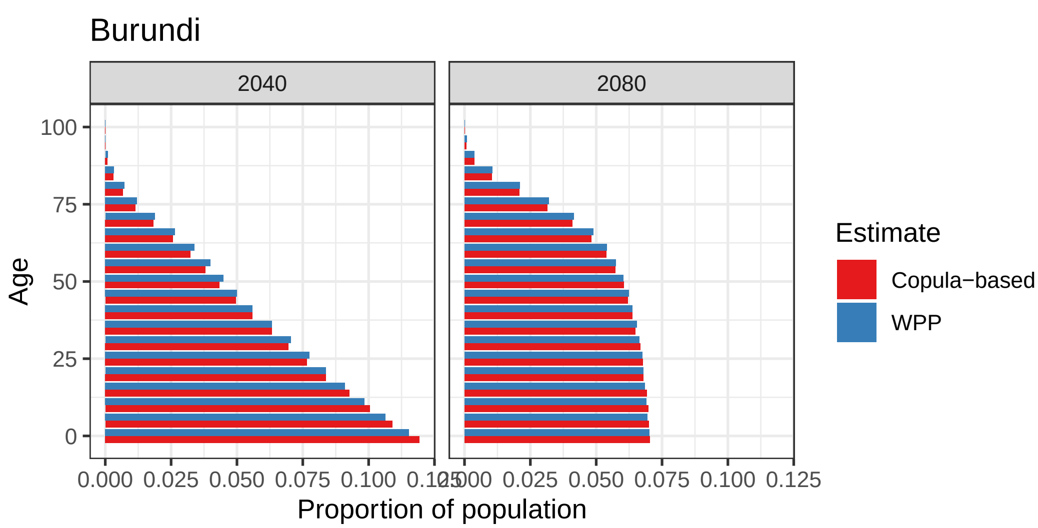
3 Summary and future work

The goal of this project is to incorporate the dependency between fertility and mortality rates in statistical models used for population projections. Preliminary work suggests that copula models may be well-suited to achieve this goal.

Future directions stemming from this project are manifold. Other copula models will be explored and methods for copula selection examined or developed. So far we have not taken into account that successive decades retain similarities that warrant an alternative to independence. Similarly, the countries included in the analysis are similar but perhaps not enough to warrant an identical copula model. To account for these issues, we will develop hierarchical Bayesian models which allow the borrowing of information across decades and countries and allow more specific predictions. In the Sub Saharan context, we are using a linear trend to project parameter values in the future. We plan to consider models that accommodate more complex trends and use information from other parts of the world where fertility levels have already dropped to replacement level or below. This will allow us to learn about dependency patterns from those regions and apply to Sub Saharan Africa. In order to achieve this aim, we must develop models for the copula and marginal parameters that incorporate auxiliary information or other relevant covariates.

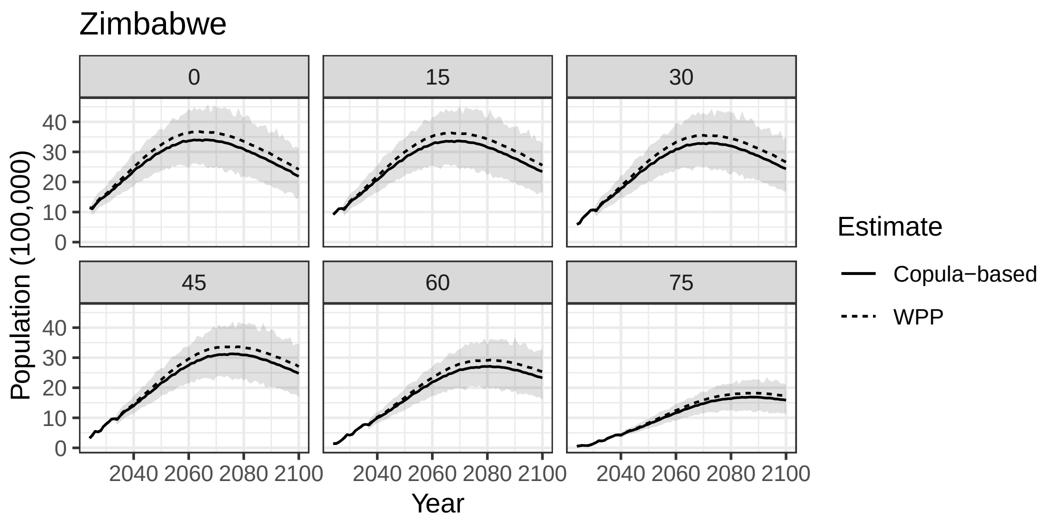


(a) Projections for Burundi, WPP versus copula-based. Each subplot is a different five year age group. The shaded area represents 2.5th and 97.5th quantiles of 1,000 fertility trajectories in copula-based approach.

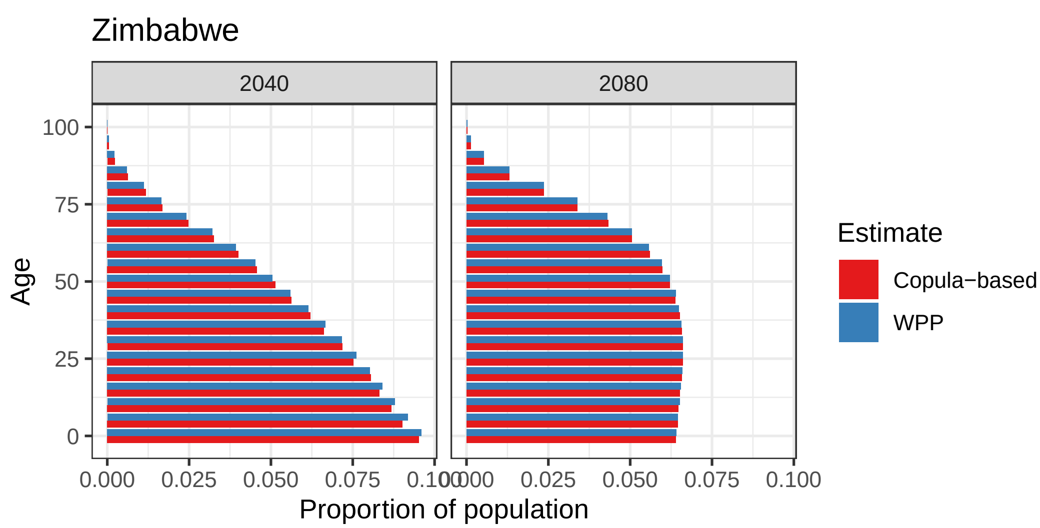


(b) Population pyramids for Burundi, 2040 and 2080, WPP versus copula-based.

Figure 2: Results for Burundi.



(a) Projections for Zimbabwe, WPP versus copula-based. Each subplot is a different five year age group. The shaded area represents 2.5th and 97.5th quantiles of 1,000 fertility trajectories in copula-based approach.



(b) Population pyramids for Zimbabwe, 2040 and 2080, WPP versus copula-based.

Figure 3: Results for Zimbabwe.

References

Alkema, Leontine, Adrian E Raftery, Patrick Gerland, Samuel J Clark, François Pelletier, Thomas Buettner, and Gerhard K Heilig. 2011. "Probabilistic Projections of the Total Fertility Rate for All Countries." *Demography* 48 (3): 815–39.

Kirk, Dudley. 1996. "Demographic Transition Theory." *Population Studies* 50 (3): 361–87.

Nobles, Jenna, Elizabeth Frankenberg, and Duncan Thomas. 2015. "The Effects of Mortality on Fertility: Population Dynamics After a Natural Disaster." *Demography* 52 (1): 15–38.

Polizzi, Antonino, and Andrea M Tilstra. 2024. "The Impact of Early Death on Birth Counts in the United States, 1950 to 2019." *PNAS Nexus* 3 (6).

Preston, Samuel H. 1972. "Interrelations Between Death Rates and Birth Rates." *Theoretical Population Biology* 3 (2): 162–85.

———. 1978. *The Effects of Infant and Child Mortality on Fertility.*

Raftery, Adrian E, Jennifer L Chunn, Patrick Gerland, and Hana Ševčíková. 2013. "Bayesian Probabilistic Projections of Life Expectancy for All Countries." *Demography* 50 (3): 777–801.

Reher, David. 1999. "Back to the Basics: Mortality and Fertility Interactions During the Demographic Transition." *Continuity and Change* 14 (1): 9–31.

UNPD. 2024. "World Population Prospects 2024." https://population.un.org/wpp/.

Wachter, Kenneth W. 2014. *Essential Demographic Methods*. Harvard University Press.

Zagheni, Emilio. 2011. "The Impact of the HIV/AIDS Epidemic on Kinship Resources for Orphans in Zimbabwe." *Population and Development Review* 37 (4): 761–83.