8802 PROJECT: SHORT DEMO FOR INDIRECT ESTIMATION OF CHILD MORTALITY

Yue Chu[*]

[1]Department of Sociology, Ohio State University
[*]Correspondance to chu.282@osu.edu

Dec.01, 2019

1 Introduction

Estimation of under-five mortality rates (U5MR) is one of the essential health indicators monitoring and evaluating the population health status, as well as the impact of health policies and programs. [1] In low- and middle income countries (LMICs) lacking reliable cival registration vital statistics (CRVS) systems, births and deaths are not prospectively and routinely recorded and investigated. Population-based surveys, such as Demographic and Health Surveys, are the primary approach to retrospectively collect information on births histories and survival status of children among women at reproductive age (usually aged 15 to 49 years), in order to calculate probability of deaths for children. [2]

In full births history surveys, information on birth dates and survival status for all children ever born alive would be collected. However in many surveys where resources are limited, such as censuses or Multiple Indicator Cluster surveys (MICS), only summary birth history surveys are conducted. In summary birth history, only the total number of children ever born alive and the total number of surviving children would be ask, without information on survival status of each child.

With summary birth history, U5MR estimates would be based on indirect estimation approach, initially developed by William Brass. [3] The core idea of Brass method is the relationship between the fraction dead among children born to mother of a particular age group, and the population-level child mortality rates. It requires some underlying assumptions on trends and age-patterns of mortality and fertility to fill the data gaps.

Over the years, demographers have proposed a few refinements to the original Brass method to better address or relax the underlying assumptions[24]-[30], and coming up with better uncertainty estimates.[4] For example, Rajaratnam et al. [4] developed methods using maternal age or time since first birth and utilizing data from ohter countries, to better capture trends in most recent time periods and generate uncertainty.

In a recent unpublished work,Carl Schemertmann[5] proposed a new Bayesian approach, exploiting the same regularitis as classic Brass methods but with relaxed requirements on demographic stability in age-spepcific rates. It derives better uncertainty measures that incorporates both sampling error and potential errors in demographic assumptions, and allowing for changing fertility rates over time. In his approach, period fertility were estimated as a weighted sum of 4 archetype shapes of fertility patterns, constructed from empirical data from 2003 Census International Database. Prior distribution for weights assigned probabilities to schedule shapes following an distribution of Dirichlet(1,1,1,1). Period mortality were modeled using Wilmoth's log-quadratic model, a flexible two dimensional mortality model base on $_0q_5$.[6] Normal priors were used for fertility and mortality changes. Posterior distribution for U5MR were estimated using data from Demographic and Health Surveys, and modeled results were validated against direct estimations and conventional Brass method. Huge variation in model performance were seen across countries. For countries like Niger, the model was not catpuring the U5MR trend well.

In this study, we are proposing a new approach, using the Singular Value Decomposition (SVD) models to estimate age schedules for mortality and fertility, which could then be used as priors in Bayesian model for indirect estimation of child mortality. SVD has been shown to be a reliable approach to model and predict age-specific mortality and fertility schedules, with less requirements on inputs and greater flexibility incooporating covariates. [7] And Bayesian approach is known for its flexibility with prior assumptions and uncertainty estiamtes. Combining the strength of

both Bayesian and SVD models, we aimed to develop a simple model to improve the model performance and uncertainty estimation of indirect estimates of U5MR based on summary birth history data.

2 Method

Overview

We first constructed schedule of age-specific probability of deaths (ASDP) by month for children under-five from full birth histories from DHS surveys. We also constructed schedule of age-specific probability of births (ASFP) by months for women aged 15 to 49 years from women's individual records from DHS.

We then used the SVD component model to fit the survival schedule for probability of deaths among children under-five, and the fertility schedule for probability of giving births among women aged 15-49 years.

With the two schedules, we microsimulated a woman's life trajectory of giving birth throughout the reproductive ages, and for each child born to the women, the life trajectory of survival through age five. Based on the results of the simulation, we calculated the distribution of total number of children ever born as well as total number of children survived. We then could develop a predictive model of child mortality as a function of total number of children ever born and total number of children survived, assuming that the woman live through the ASFP schedule for fertility during her reproductive years, and her children follows the ASDP schedule for survival.

Data source

Among all DHS surveys conducted after year 2000, a total of 177 surveys had publically available full birth histories data as well as women's records. For the ease of code-runnning for this demo project, results presented here were based on a subset of available DHS surveys, including 10 births records and 10 women datasets. Data were accessed at DHS program website. (Table 1)

Country Name	Survey Year
Afghanistan	2015
Albania	2008
Albania	2017
Angola	2015
Armenia	2000
Armenia	2005
Armenia	2010
Armenia	2016
Azerbaijan	2006

Table 1: List of survyes used in the demo analysis

Probabilities of deaths and births
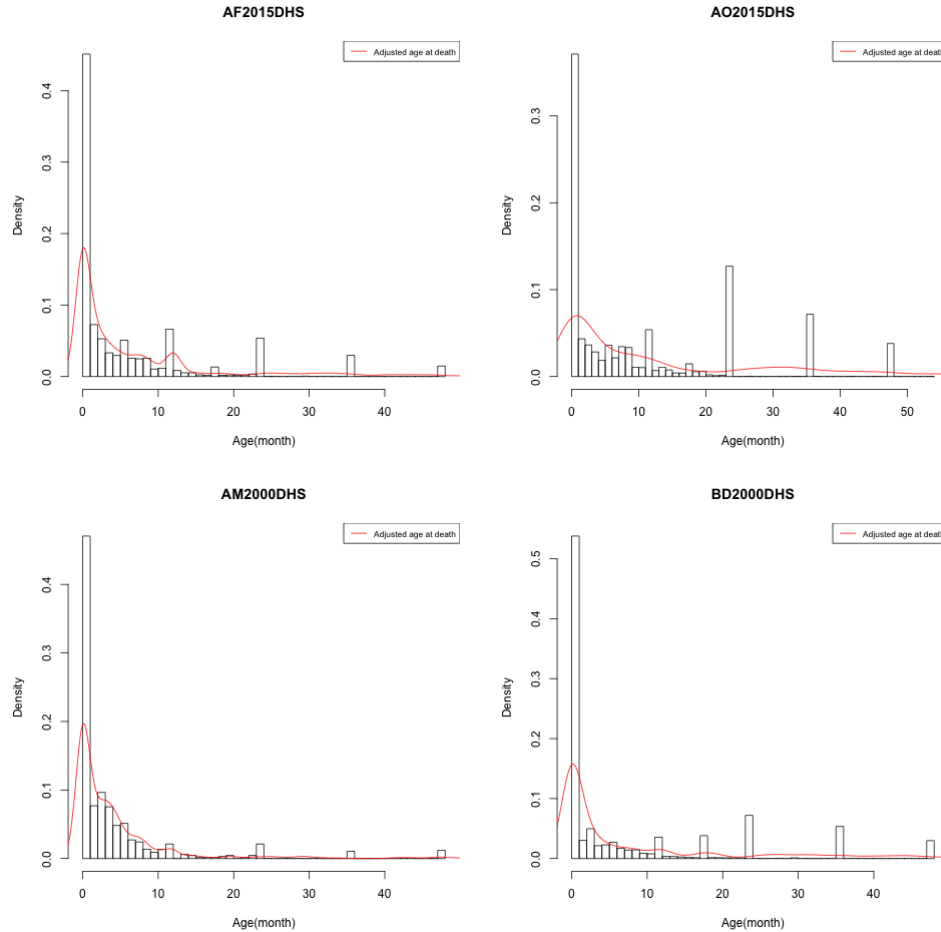
Age-specific probability of deaths for children under-five were calculated by children's month of age, for a total of 60 single-month age segments.

Age heaping and adjustment of age at death

Since age at deaths were reported by month only for children under 24-months of age.1 For children aged 24 and above, age at deaths were reported by unit of years instead. In order to calculate probability of deaths by month up till 59 months, all deaths with age at death reported as multiples of 12 months for 24 months and above were randomly redistributed to subsequent months within the year, assuming uniform distribution of death through out the year.

Age heaping also existed in DHS birth histories.1 People tended to round the age at death to 6, 12 and 18 months, resulting in significantly more deaths in these months comparing to the neighboring months. Therefore we redistributed deaths occurred at 6, 12 and 18 months to neighboring months (± 2 months) with symmetrically graduated probabilities P(0.1, 0.2, 0.4, 0.2, 0.1).

The calculation of component probability of births then followed the classifcal DHS approach, only with higher resolution of age groups. The methods were described in details in DHS methodology reports.[8]

Figure 1: Density distribution of age at death among children under-five

Similarly, age-specific probability of giving birth for women of reproductive age (aged 15 to 49 years) were also calcualted by women's month of age, for a total of 419 single-month age segments. The calculation of component probability of births also followed the classifcal DHS approach, with

higher resolution of age groups.[8]

SVD modeled age schedules for ASDP and ASFP

We used Singular Value Decomposition (SVD) approach to construct general, parsimonious component models of age schedules for age-specific mortality and fertility rates. The method was described in details in Clark, 2015 publication. [7] Briefly speaking, the SVD approach first factorizes a matrix of demographic estimates (denoted as X) into three matrices - namely a matrix of 'left singular vectors' (LSVs) arranged in columns (denoted as U), a matrix of 'right singular vectors' (RSVs) arranged in columns (denoted as V), and a diagonal matrix of 'singular values' (SVs) (denoted as S) (equation demostrated as below). Then the matrix of demographic age schedules could be reconstructed with weighted-sum of much fewer components from the component model, yet still yielding a reasonably realiable estimation.

$$X = USV^{-1}$$

The ASDP and ASFP matrices constructed above from sample DHS surveys were used as the input data sources to train the models. Four of the new dimensions identified by each SVD were retained.

Microsimulation

We used microsimulation to estimate the distribution of total number of children ever born and total number of children survived.

For a hypothetical woman, based on the SVD-modeled ASFP schedule for fertility during her reproductive years, the woman was simulated through the model month-by-month while keeping track of her status transition trajectories in each time segment - whether giving birth, not giving birth or death. And if the woman give birth to a child, the survival trajectory of the child will also be simulated assuming that s/he lived through the ASDP schedule for survival.

By aggregating all the transition values over the model's cycles from the simulation, we will be able to get the simulated distribution of total number of births among women of reproductive age, as well as the total number of children surviving to age 5, given the SVD-modeled ASFP and ASDP schedules.

Prediction and validation

We then could develop a predictive model of child mortality as a function of total number of children ever born and total number of children survived. To assess the model performace, we randomly selected 80% of data for model fitting and used the remaining 20% for out-of-sample predictions. Root mean squared error were used to evelute accuracy of predictions. Then we tested the model with empirical population data, and compared modeled estimates to mortality estimates coming from direct estimates and classical Brass method.

3 Preliminary results

SVD modeled schedules for ASDP and ASFP

Figure 2 compares the SVD modeled age schedule for probability of deaths by month among children under-five to the empirical probability of deaths from DHS surveys used as model input. And Figure 3 shows the schedules for the first 4 components in the SVD models, $s_i u_i$, which are the LSVs from the SVD of DHS probability of deaths schedules scaled by their corresponding singular values.
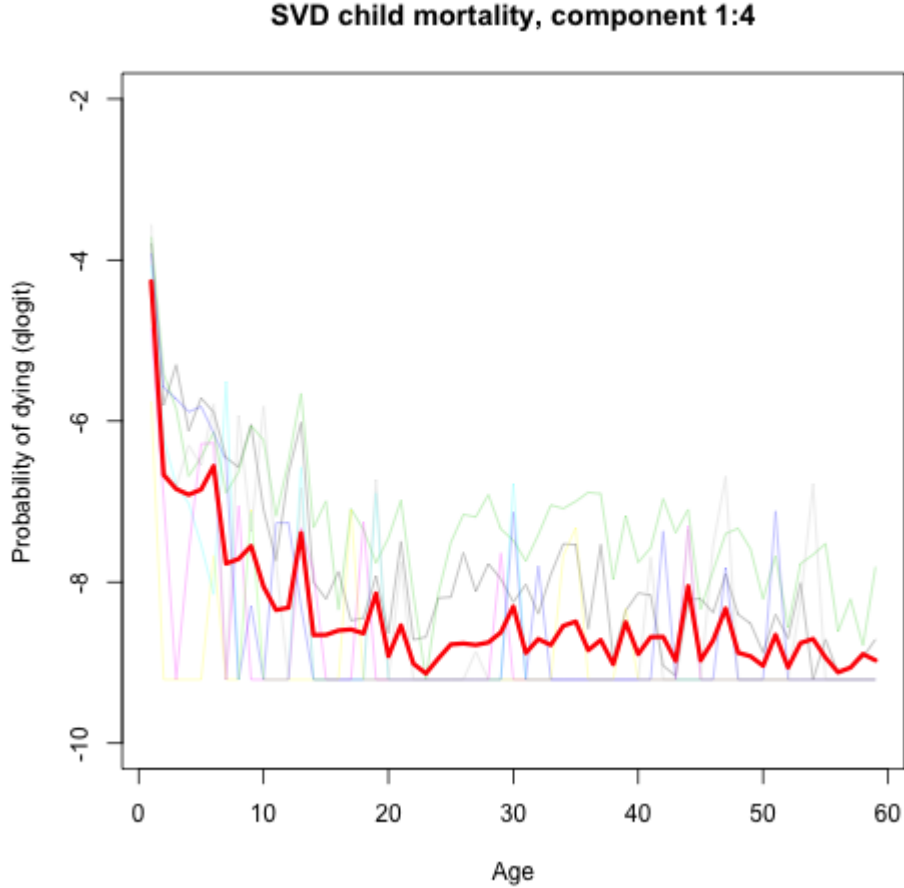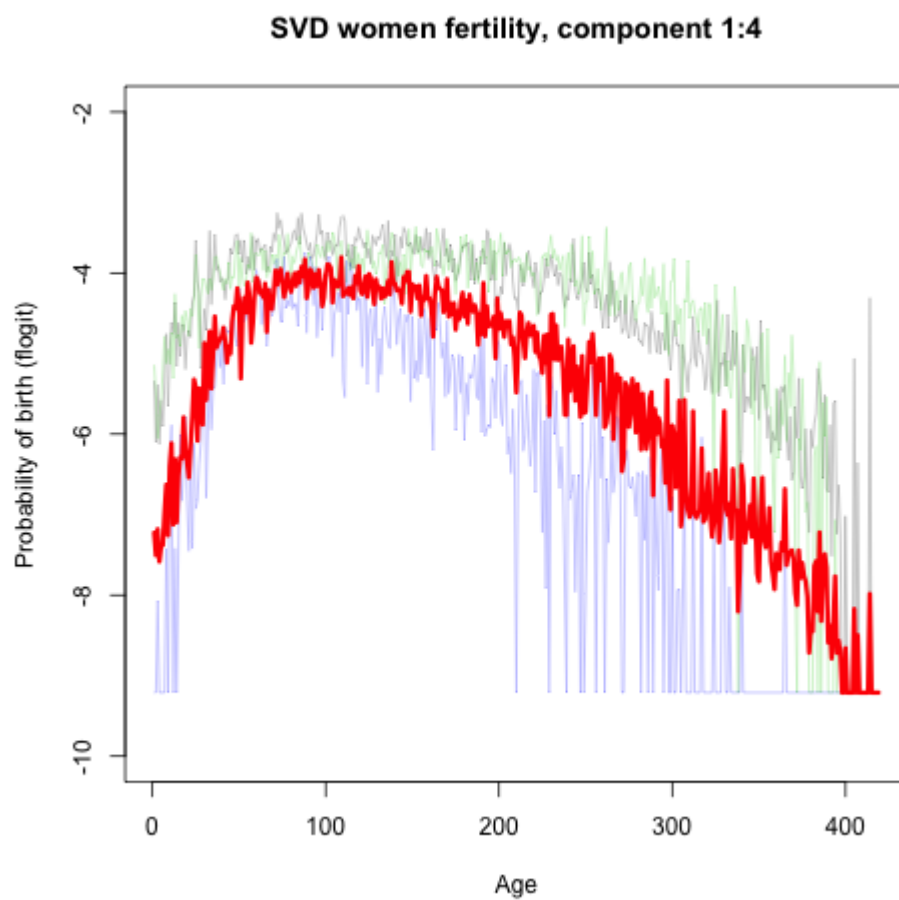

Figure 2: SVD for log-transformed age-specific probability of death, modeled vs empirical

Similar to the figures above, Figure 4 compares the SVD modeled age schedule for probability of giving births by month among women of reproductive age to the empirical probability of births from DHS surveys used as model input. And Figure 5 shows the schedules for the first 4 components in the SVD models, $s_i u_i$, which are the LSVs from the SVD of DHS probability of births schedules scaled by their corresponding singular values.

In general, the SVD models captured the shape of monthly ASDP and ASFP well. However, we also noticed that the modeled trends were very bumpy for both outcomes. Even after redistribution of deaths and adjustment for age heaping, the log-scaled probability of deaths had large fluctuation on monthly basis at individual survey level. Increasing the number of input survey data points
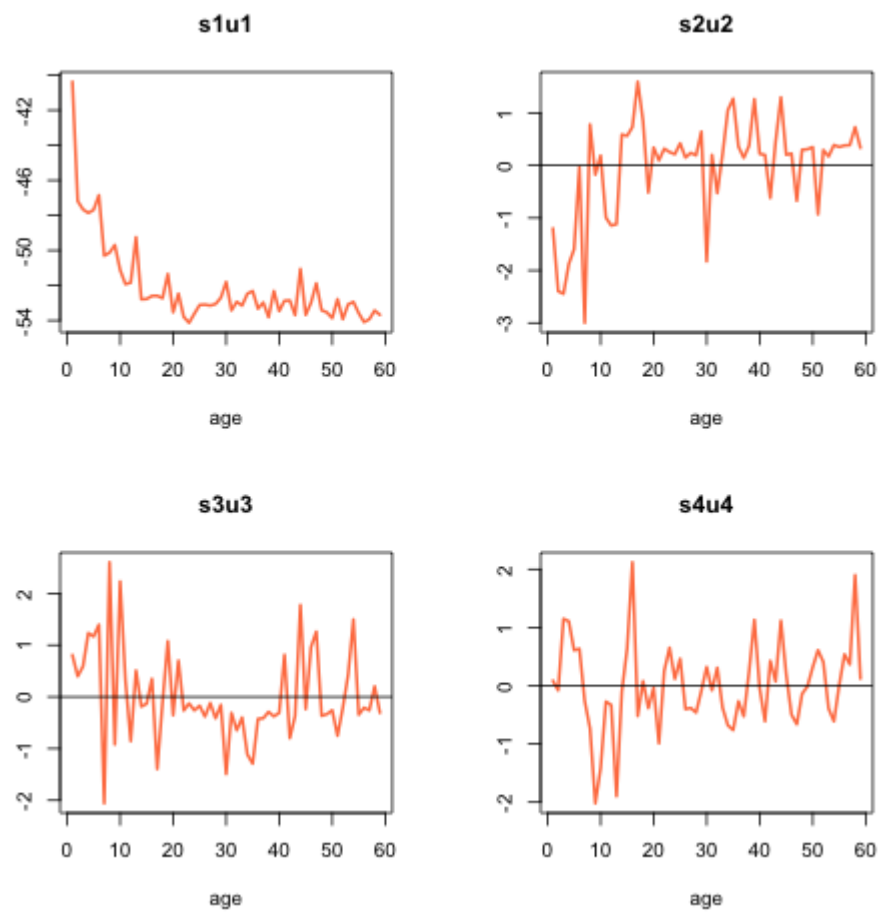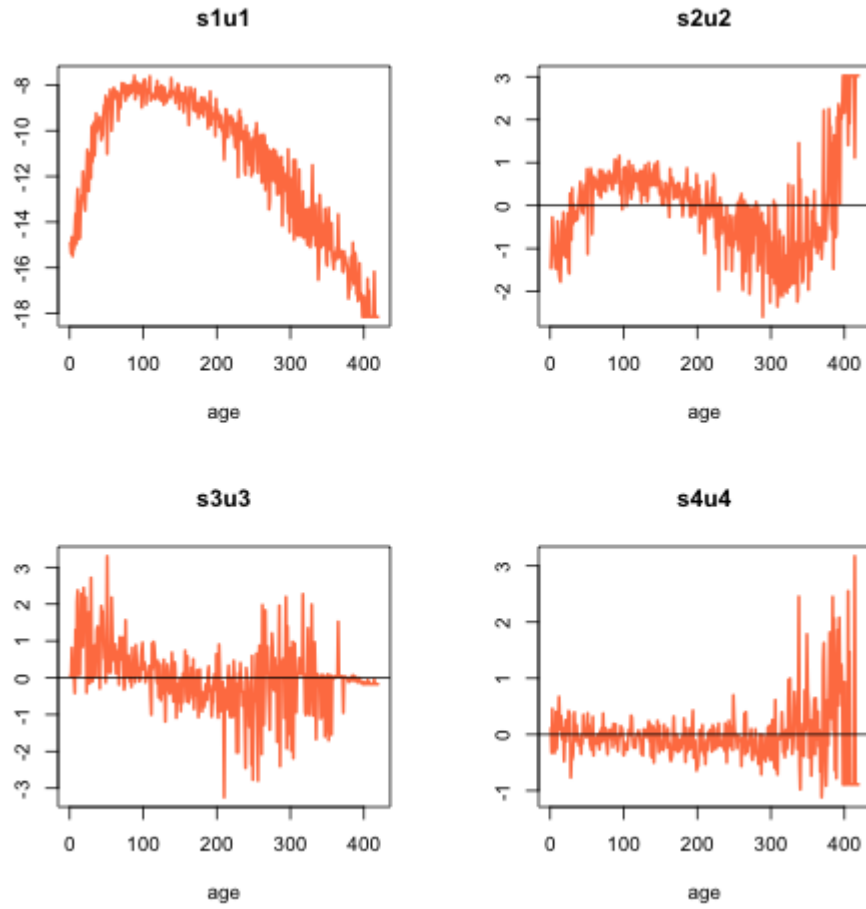
Figure 3: SVD components for log-transformed age-specific probability of death

Figure 4: SVD for log-transformed age-specific probability of giving birth, modeled vs empirical

Figure 5: SVD components for log-transformed age-specific probability of giving birth

would help smoothing the trend to some extent.

Upcoming results and next steps

- Results for microsimulation (Upcoming)

  – Account for maternal mortality: currently assumed no maternal daeths in this population.

  – Adjust probability of births during pregnancy.

  – Account for association between maternal and child mortality: currently assumed independence between mother and child, but evidence suggests that among high HIV-prevalence population, maternal mortality is significantly associated with child mortality and could impact the validity of indirect estimation of child mortality. [9]

  – Incoorporating trend in fertility and mortality schedules over time.

- Test SVD with smoothed ASDP (/and ASFP) inputs

- Fit SVD under Bayesian framework

- Uncertainty estimation

- Validation of modeled results

References

[1] United Nations General Assembly. Transforming our world: the 2030 Agenda for Sustainable Development; 2015.

[2] Mathers C, Boerma T. Mortality measurement matters: improving data collection and estimation methods for child and adult mortality. Plos Medicine. 2010;7(4):e1000265.

[3] Brass W. Methods for estimating fertility and mortality from limited and defective data. Methods for estimating fertility and mortality from limited and defective data. 1975;.

[4] Rajaratnam JK, Tran LN, Lopez AD, Murray CJL. Measuring Under-Five Mortality: Validation of New Low-Cost Methods. Plos Medicine. 2010 Apr;7(4):e1000253–24.

[5] Schmertmann CP. Probabilistic Indirect Estimation of Under-Five Mortality. 2019 Sep;p. 1–22.

[6] Wilmoth J, Zureick S, Canudas-Romo V, Inoue M, Sawyer C. A flexible two-dimensional mortality model for use in indirect estimation. Population studies. 2012;66(1):1–28.

[7] Clark SJ. A singular value decomposition-based factorization and parsimonious component model of demographic quantities correlated by age: Predicting complete demographic age schedules with few parameters. arXiv preprint arXiv:150402057. 2015;.

[8] Croft TN, Marshall AM, Allen CK, et al. Guide to DHS statistics. Rockville, Maryland, USA: ICF. 2018;.

[9] Quattrochi J, Salomon JA, Hill K, Castro MC. Measuring and correcting bias in indirect estimates of under-5 mortality in populations affected by HIV/AIDS: a simulation study. BMC public health. 2019;19(1):1516.