8802 Final R Markdown
Short demo for indirect estimation of child mortality

Yue Chu | chu.282@osu.edu

Document version: last modified on Tue Dec 10 22:43:38 2019

Overview

This document shows the analysis conducted to generate the results, figures and tables in the "8802Final_tex.tex" document.

The data analysis follows the following steps:
1. Extract and prepare the data sets from Demographic and Health Surveys (DHS)
* Birth records (a subsample of 10 surveys after 2000)
* Women's records (a subsample of 10 surveys after 2000)
2. Calculate component probability of births an deaths
* Calculate age-specific probability of deaths (ASDP) for children under-five by months, from 0 month up till 59 months * Calculate age-specific probability of fertility (ASFP) for women aged 15-49 years at the time of interview
3. Use Singular Value Decomposition (SVD) approach to model schedules of ASDP and ASFP
4. Use Microsimulation to estimate the distribution 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.

More details are provided as follows.

1. Prepare births and women's datasets

1.1 Get available DHS datasets

Among all DHS surveys conducted after year 2000, a total of 177 surveys have publically available birth records, while a total of 177 surveys have publically available women's records. For the ease of code-runnning for this demo project, we used the only a subset of available DHS surveys, including 10 births records and 10 women datasets. Data were accessed at [DHS program website] (https://dhsprogram.com/data/available-datasets.cfm). (Table 1)

Table 1: List of surveys used in the analysis

	CountryName	SurveyYear	SurveyId
1	Afghanistan	2015	AF2015DHS
44	Albania	2008	AL2008DHS
48	Albania	2017	AL2017DHS
108	Angola	2015	AO2015DHS
172	Armenia	2000	AM2000DHS
176	Armenia	2005	AM2005DHS
180	Armenia	2010	AM2010DHS
184	Armenia	2016	AM2016DHS
307	Azerbaijan	2006	AZ2006DHS
345	Bangladesh	2000	BD2000DHS

2 Calculate component probability of births and deaths

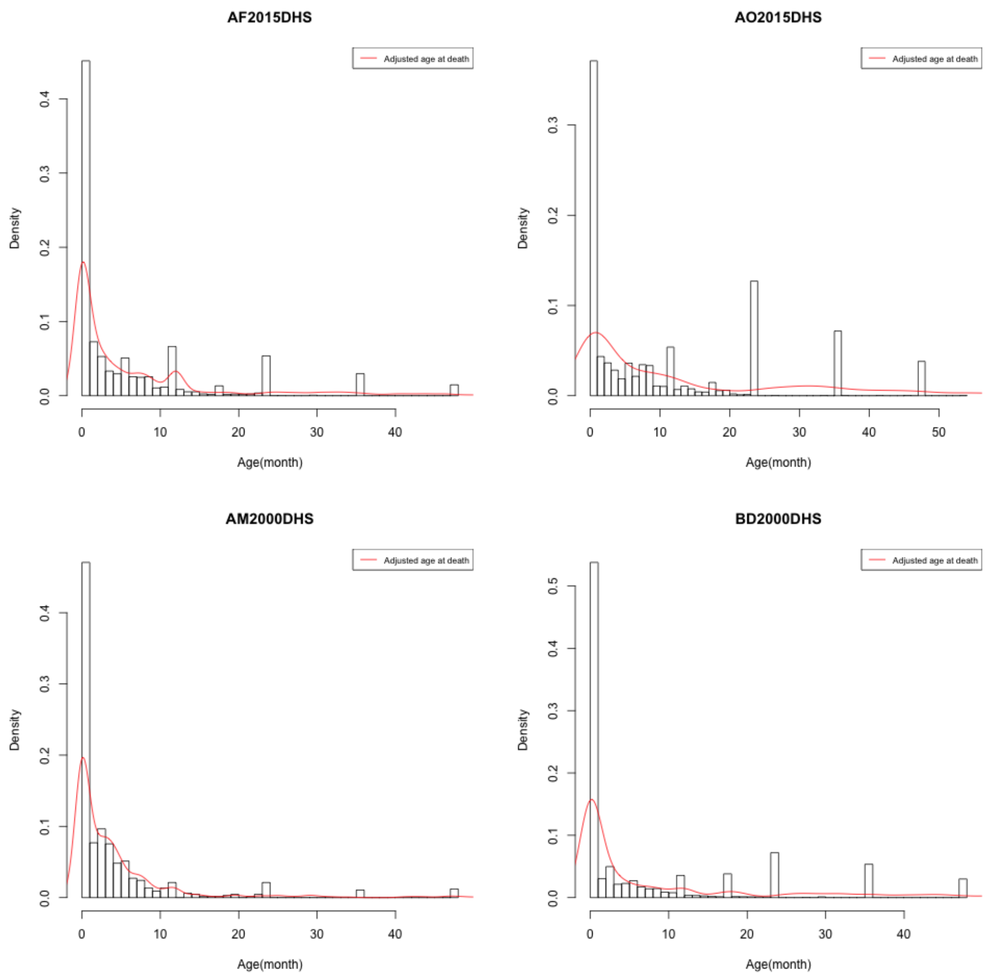
2.1 Component probability of deaths by month of age

Age-specific probability of deaths for children under-five were calculated by children's month of age, for a total of 60 age segments. The calculation of component probability of births follows the classifcal DHS approach, only with higher resolution of age groups. The methods were described in details in DHS methodology reports.(Croft et al. 2018)

2.1.1 Age heaping and adjustment of age at death

There were two major challenges in the process of generating probability of deaths estimates by months for children under-five:
First, age at deaths are reported by month only for children under 24-months of age, and are reported by full years completed instead for children aged 24 and above. In this paper, deaths with age at death reported in full years (as multiples of 12 months for 24 months and above) are redistributed to subsequent months, assuming uniform distribution of death through out the year.
Second, age heaping also exists in DHS birth histories. People tend to round the age at death to 6, 12 and 18 months, resulting in significantly more deaths in these two months comparing to the neighboring months. In this paper, deaths occurred at 6, 12 and 18 months are also redistributed to neighboring months ($\pm$ 2 months) with symmetrically graduated probabilities P(0.1, 0.2, 0.4, 0.2, 0.1).

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

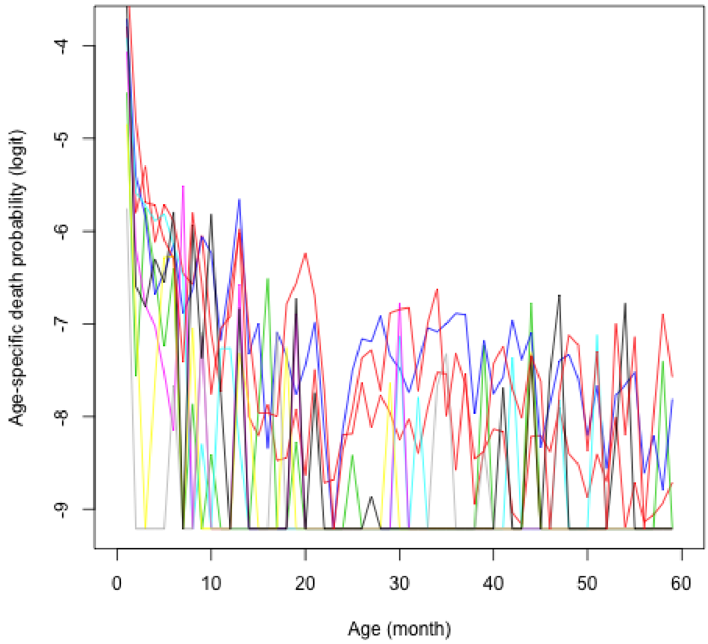


After the adjustments, the density distribution of age at deaths are overall smooth without obvious clustering at certain post-neonatal age groups. (Figure 1) Thus we decide not to perform further smoothing at this point. However we should also keep in mind that the bumpiness could be more obvious after log-transformation, thus affect the model performance for fitting age schedules. This will be demostrated later.

2.1.2 Age-specific death probabilities by month of age

Figure 2 demonstrates the age-specific death probabilities by month of age from selected surveys.

Figure 2. Log-transformed ASDP



2.2 Component probability of births by month of age

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 age segments. The calculation of component probability of births also followed the classifcal DHS approach, with higher resolution of age groups. The methods were described in details in DHS methodology reports.(Croft et al. 2018)

Figure 3 demonstrates the age-specific birth probabilities by month of age from selected surveys. We noticed high level of fluctuation for monthly-estimation for individual surveys. The trend became smoother and more stabelized when we aggregated multiple surveys together.

3. Modeling age patterns for deaths and births using

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. (Clark 2015) 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.

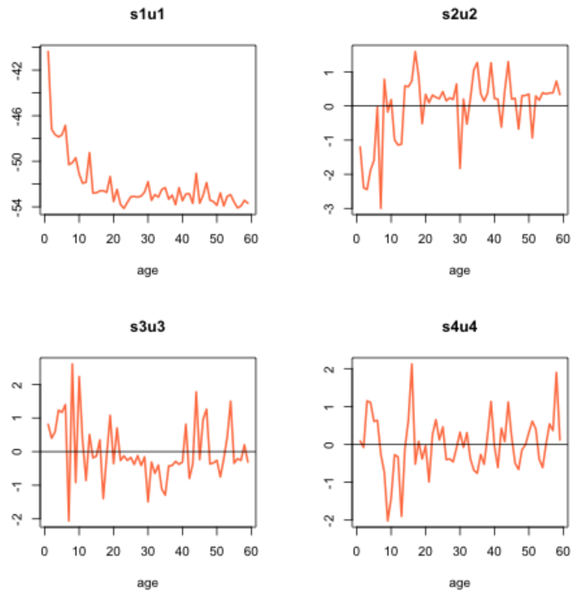
$$X = USV^{-1}$$

The ASDP and ASFP matrices constructed above from sample DHS surveys were used as the input data sources to train the models. We took the first four components from the SVD for estimation of age-specific probability of death, and took the first four components from the SVD for estimation of age-specific probability of birth.

Figure 4.a 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 Figure4.b 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 4.a SVD for log-transformed age-specific probability of death, modeled vs empirical

Figure 4.b SVD components for log-transformed age-specific probability of death



Similar to the figures above, Figure 5.a 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 Figure5.b 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.

Figure 5.a SVD for log-transformed age-specific probability of death, modeled vs empirical

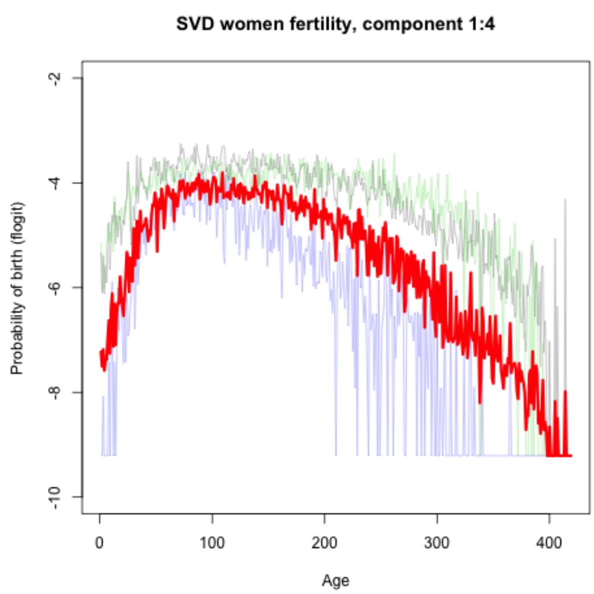
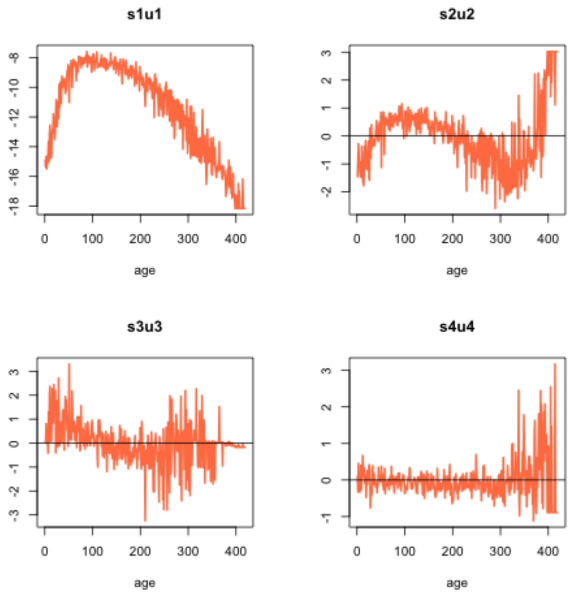


Figure 5.b SVD components for log-transformed age-specific probability of death



Microsimulation (Upcoming)

Then we used microsimulation to estimate the distribution 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.

References

Clark, Samuel J. 2015. "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:1504.02057*.

Croft, Trevor N, Aileen MJ Marshall, Courtney K Allen, and others. 2018. "Guide to Dhs Statistics." *Rockville, Maryland, USA: ICF*.