Application of a Singular Value Decomposition-based Factorization and Parsimonious Component Model of Mortality to HIV Epidemics in Africa

Jonathan A. Muir[1, *] and Samuel J. Clark[1, 2, 3, 4]

[1]The Ohio State University, 238 Townshend Hall, 1885 Neil Ave Mall, Columbus, OH 43210
[2]MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt), School of Public Health, Faculty of Health Sciences, University of the Witwatersrand
[3]ALPHA Network, London School of Hygience and Tropical Medicine, London, UK
[4]INDEPTH Network, Accra, Ghana
[*]Correspondence to: work@jamuir.net

September 19, 2018

Abstract

Many countries around the world, particularly in Africa, require the modeling of mortality in order to derive complete age schedules of mortality. These age schedules are required for population estimation, forecasting, and projection and other tasks in Demography and Epidemiology. One method for constructing these models that is garnering increasing interest is the use of singular value decomposition-base factorization (SVD). Using simulated demographic projection data of HIV epidemics calibrated and organized in SPECTRUM, this study applies SVD to (1) derive a parsimonious component model of demographic age schedules using age-specific mortality and (2) predict age-specific mortality using HIV indicators and summary measures of age-specific mortality. The component model of age-specific mortality successfully reproduces the data with four inputs, and acting through those four inputs, HIV related covariates are able to accurately predict full age schedules.

1 Introduction

Many countries in Africa lack well functioning civil registration and/or vital statistics systems. This results in a paucity of accurate information on mortality. Mortality models are thus required to derive complete age schedules of mortality for these countries. One novel method for constructing these models is the use of singular value decomposition (SVD) (Clark 2015, 2016; Sharrow et al. 2014), which "is a matrix factorization method that decomposes a matrix X into three matrix factors with special properties:

$$X = USV^T \tag{1}$$

U is a matrix of left singular vectors (LSVs) arranged in columns, V is a matrix of right singular vectors (RSVs) arranged in columns, and S is a diagonal matrix of singular values (SVs)" (Clark 2016, 8-9). Using all of the information contained in the three matrix factors it is possible to reconstitute the original matrix X using matrix algebra. However, only the first few terms of each of the matrix factors are necessary to produce a close approximation of the original matrix (Clark 2016). The LSVs and SVs are constant and their contribution to reconstituting the original matrix is determined by the RSVs, which function as weights. Of interest to this study, it is possible to relate these weights (the v terms of the RSVs) to select covariates (e.g., child mortality or HIV related indicators) using OLS regression. Clark (2015, 2016) and Sharrow et al. (2014) present detailed theoretical explanations of SVD as well as several examples of its application. This study further demonstrates the potential of SVD in constructing models of demographic age schedules by using it to (1) predict a parsimonious component model of demographic age schedules using age-specific mortality (2) predict a parsimonious component model of demographic age schedules using using HIV indicators and age-specific mortality.

2 Materials and Methods

The primary objective of this study is to utilize SVD to derive age-schedules of mortality (${}_nq_x$). To this end, demographic simulation data were obtained using SPECTRUM (Stover et al. 2010). Developed by the Futures Institute in collaboration with Futures Group International, USAID, UNAIDS, WHO, and UNICEF, SPECTRUM contains a database of population estimates for 193 countries and regions as well as a database of family planning information drawn from Demographic and Health Surveys (Stover et al. 2010). At its core is a demographic projection model–DemProj–which projects population by age and sex. Of importance to this study, SPECTRUM also includes an AIDS Impact Model (AIM), which interacts with the demographic projections from DemProj to evaluate the impacts of HIV/AIDS (Stover 2003).

The 11 African countries selected for inclusion in this study were those identified as having an HIV epidemic, defined as having an adult HIV prevalence higher than 6.0% (Table 1). Project simulations were carried out in SPECTRUM for each country by systematically varying adult and child ART coverage as well as HIV incidence with 275 unique simulations produced for each country, resulting 3,025 distinct simulations of HIV epidemics.

Table 1: Adult HIV Prevalence in Select African Countries

Country	**HIV Prevalence**
Swaziland	27.4%
Lesotho	22.9%
Botswana	21.9%
South Africa	19.1%
Zimbabwe	15.0%
Namibia	14.3%
Zambia	12.5%
Mozambique	10.8%
Malawi	10.3%
Uganda	7.4%
Kenya	6.0%

*Note: Prevalence data obtained from Kharsany and Karim (2016)

The first objective is to assess the analytic potential of SVD in its most basic form by performing a matrix factorization of ${}_nq_x$ for the 3,025 distinct simulations of HIV epidemics in the data using the SVD package in the R statistical software package (R Foundation for Statistical Computing 2014). Initial steps in this process indicated that the first four terms of the SVD matrix factors account for 99.9% of the variation in the original data, therefore only these initial terms were used to reconstitute the original matrix.

The second objective is to then relate the SVD weights to key covariates; i.e., under five child mortality (${}_5q_0$), adult mortality (${}_{45}q_{15}$), and Δ_{HIV}, which is a measure of the fraction of the population that is HIV$^+$ but not on ART and therefore likely to die as a result of HIV (Δ_{HIV} = HIV prevalence - ART coverage). As an extension of the SVD component model from objective 1, these covariates were used to predict the SVD model weights using the following formulas, which are based upon models developed by Clark (2016):

$$\begin{aligned} v_{zli} = c_{zi} + \beta_{z1i} * {}_5q_{0\,zl} + \beta_{z2i} * logit({}_5q_0)_{zl} + \beta_{z3i} * logit({}_5q_0)^2_{zl} + \beta_{z4i} * logit({}_5q_0)^3_{zl} \\ + \beta_{z5i} * {}_{45}q_{15\,zl} + \beta_{z6i} * logit({}_{45}q_{15})^2_{zl} + \beta_{z7i} * logit({}_{45}q_{15})^3_{zl} \\ + \beta_{z8i} * [logit({}_5q_0)_{zl} * logit({}_{45}q_{15})_{zl}] + \epsilon_{zli} \end{aligned} \tag{2}$$

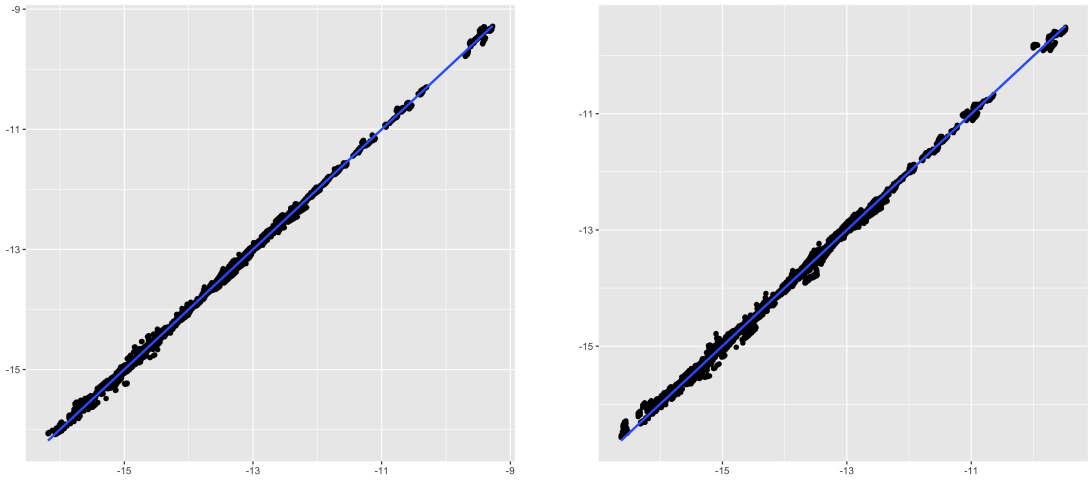
$$
\begin{aligned}
v_{zli} = c_{zi} + \beta_{z1i} * {}_5q_{0\,zl} + \beta_{z2i} * logit({}_5q_0)_{zl} + \beta_{z3i} * logit({}_5q_0)^2_{zl} + \beta_{z4i} * logit({}_5q_0)^3_{zl} \\
+\beta_{z5i} * {}_{45}q_{15\,zl} + \beta_{z6i} * logit({}_{45}q_{15})^2_{zl} + \beta_{z7i} * logit({}_{45}q_{15})^3_{zl} \\
+\beta_{z8i} * [logit({}_5q_0)_{zl} * logit({}_{45}q_{15})_{zl}] + \Delta_{HIV} + \epsilon_{zli}
\end{aligned} \tag{3}
$$

Results

The SVD component model of age-specific mortality successfully reproduces the original data. Figures 1(a) and 1(b) visualizes the overall prediction error for males and females within the data when attempting to predict ${}_nq_x$ using only first four terms of the matrix factors produced by SVD. The values for ${}_nq_x$ from the original data correspond to the x-axis and the values for ${}_nq_x$ predicted by the SVD derived component model correspond to the y-axis. Overall, the distribution indicates that SVD successfully reproduces estimates of ${}_nq_x$ with minimal prediction error. Figures 2(a) and 2(b) present simulation specific examples of age specific ${}_nq_x$ for males and females living in Lesotho, which again demonstrates the close approximation of SVD derived estimates compared to the original data (i.e., The orange dots are the original data juxtapose the solid black line that indicates predicted values). These distributions reflect the "humped" age pattern of mortality that has been attributed to the large number of deaths at very young and adult ages that occur in HIV-affected populations (Sharrow et al. 2014).

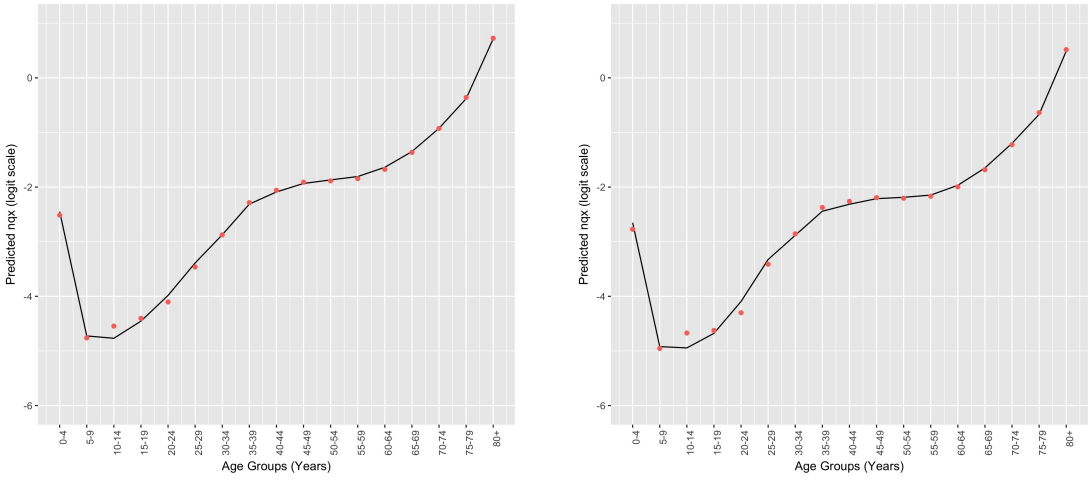
Similarly, SVD is generally successful at reproducing the original data when ${}_5q_0$ and ${}_{45}q_{15}$ are used to predict the RSV model weights. The prediction error is somewhat larger when using ${}_5q_0$ and ${}_{45}q_{15}$ (see Figures 3(a) and 3(b)), but overall suggests that SVD successfully reproduces estimates of ${}_nq_x$ with minimal prediction error. Figures 4(a) and 4(b) use the same simulation specific examples for males and females living in Lesotho that are presented in Figures 2(a) and 2(b), but uses the estimates of ${}_nq_x$ obtained by using ${}_5q_0$ and ${}_{45}q_{15}$ to predict the model weights. The results are largely consistent.

SVD also successfully reproduces the original data when Δ_{HIV} is used to predict model weights in addition to ${}_5q_0$ and ${}_{45}q_{15}$. The prediction error for males in Figure 5(a) is consistent with that from Figure 2(a). Likewise, the prediction error for females in Figure 5(b) is consistent with that from Figure 2(b). Moreover, the distributions of the predict vs. original data presented in Figures 6(a) and 6(b) are largely consistent with that in Figures 4(a) and 4(b). This consistency is maintained across epidemics in which there is high or low prevalence of persons who are HIV$^+$ but not on ART for men, but there is some divergence in the epidemic for which there was a low prevalence of females who are HIV$^+$ but not on ART.



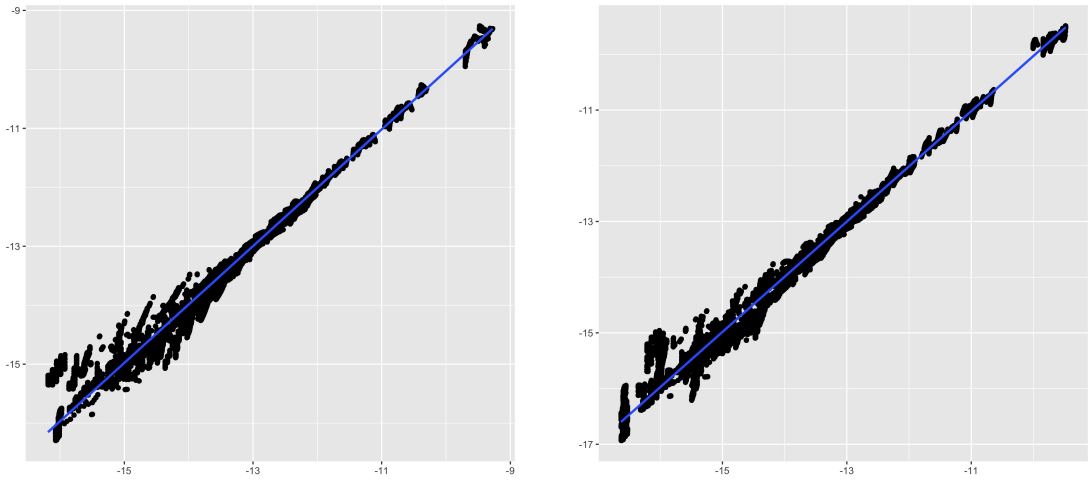
(a) Male (b) Female

Figure 1: Prediction error in SVD estimation of Spectrum simulation data in which the first four terms of the RSVs were used as model weights. Original Spectrum data are plotted along the x-axis. SVD predicted data are plotted along the y-axis.



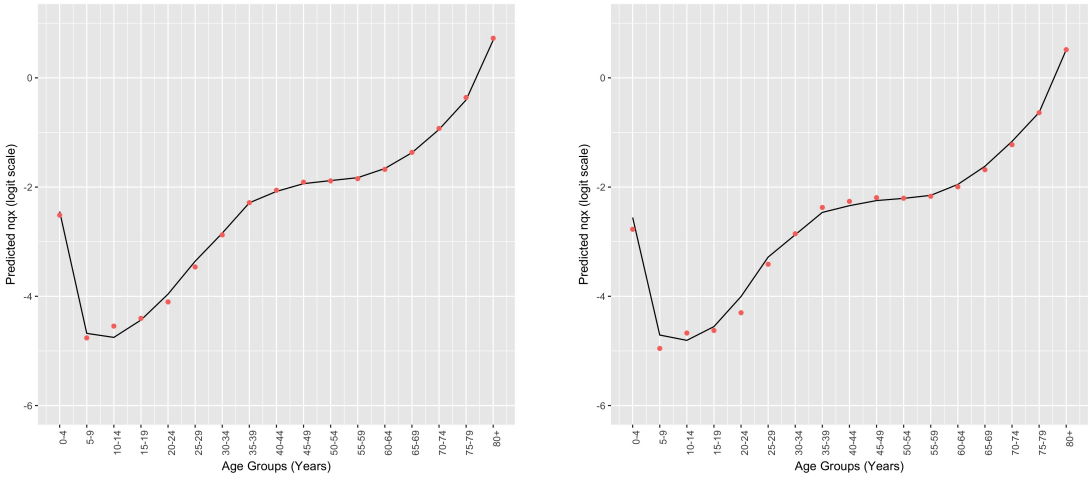
(a) Male (b) Female

Figure 2: Lesotho Example of Age-specific ${}_nq_x$ in which the first four terms of the RSVs were used as model weights. The orange dots are the original data juxtapose the solid black line that indicates predicted values.



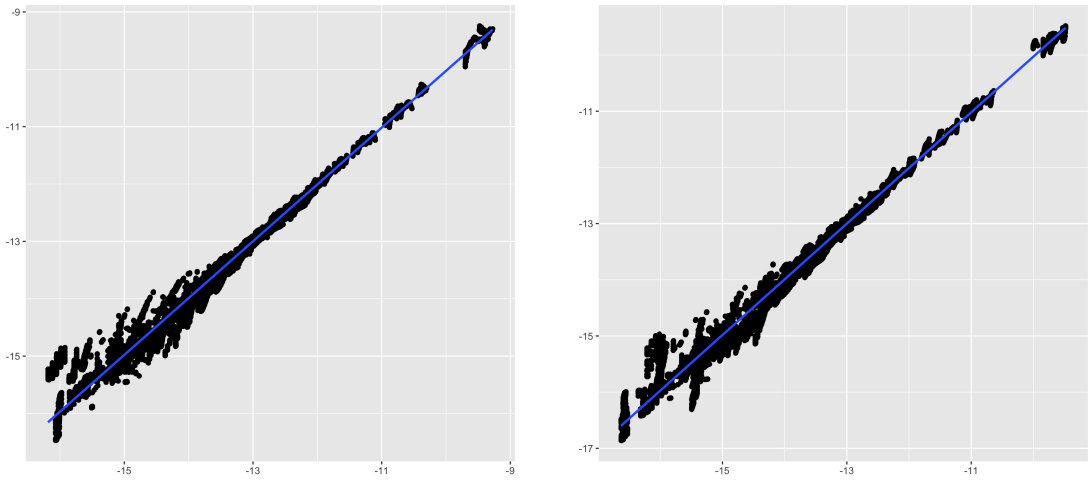
(a) Male (b) Female

Figure 3: Prediction error in SVD estimation of Spectrum simulation data in which SVD model weights were predicted using covariates ${}_5q_0$ and ${}_{45}q_{15}$. Original Spectrum data are plotted along the x-axis. SVD predicted data are plotted along the y-axis.



(a) RSV Weights Predicted from 5q0 and 45q15 (b) RSV Weights Predicted from 5q0 and 45q15

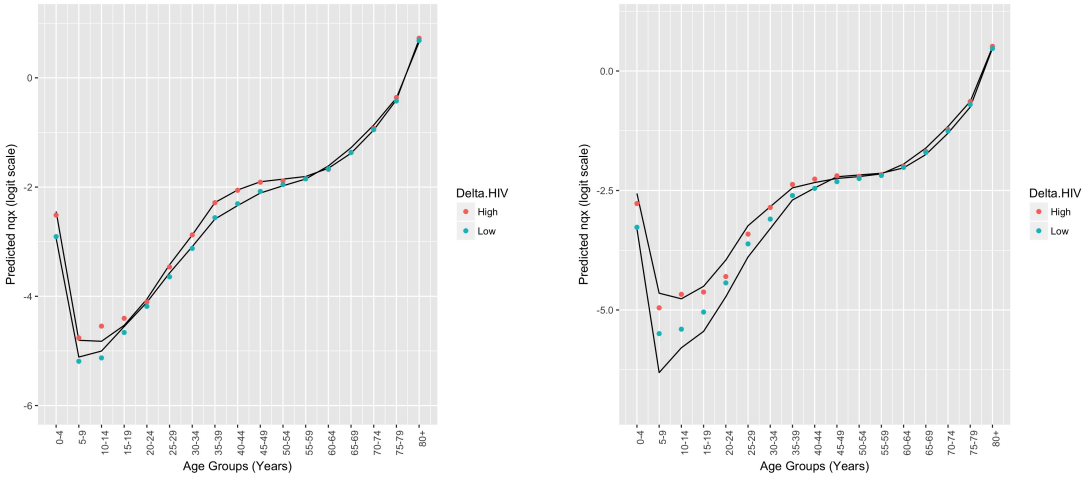
Figure 4: Lesotho example of age-specific ${}_nq_x$ in which SVD model weights were predicted using covariates ${}_5q_0$ and ${}_{45}q_{15}$. Orange dots are original data. Solid black lines indicate predicted values.



(a) RSV (c=4)

(b) RSV Weights Predicted from 5q0 and 45q15

Figure 5: Prediction error in SVD estimation of Spectrum simulation data in which SVD model weights were predicted using Δ_{HIV} in addition to ${}_5q_0$ and ${}_{45}q_{15}$. Original Spectrum data are plotted along the x-axis. SVD predicted data are plotted along the y-axis.



(a) Male

(b) Female

Figure 6: Lesotho example of age-specific ${}_nq_x$ for high and low prevalence of persons who are HIV$^+$ but not on ART. Covariates Δ_{HIV}, ${}_5q_0$, and ${}_{45}q_{15}$ were used to predict SVD model weights. Orange dots are original data for high Δ_{HIV}. Green dots are original data for low Δ_{HIV}. Solid black lines indicate predicted values.

References

Clark, S. 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*.

Clark, S. J. (2016). A general age-specific mortality model with an example indexed by child or child/adult mortality. *arXiv preprint arXiv:1612.01408*.

Kharsany, A. B. and Karim, Q. A. (2016). Hiv infection and aids in sub-saharan africa: current status, challenges and opportunities. *The open AIDS journal*, 10:34.

R Foundation for Statistical Computing (2014). The r project for statistical computing.

Sharrow, D. J., Clark, S. J., and Raftery, A. E. (2014). Modeling age-specific mortality for countries with generalized hiv epidemics. *PloS one*, 9(5):e96447.

Stover, J. (2003). Aim. version 4. a computer program for making hiv/aids projections and examining the social and economic impacts of aids.

Stover, J., McKinnon, R., and Winfrey, B. (2010). Spectrum: a model platform for linking maternal and child survival interventions with aids, family planning and demographic projections. *International journal of epidemiology*, 39(suppl_1):i7–i10.