

# Reproduce of the Paper A General Age-Specific Mortality Model with An Example Indexed by Child or Both Child and Adult Mortality\*

Luyuan Hu

2020/12/10

## Abstract

Most African countries and nearly one third of all countries need to use mortality models to infer a complete age table of mortality rates for population estimates, projections/projections and other demographic and epidemiological tasks. Models that link child mortality to mortality at other ages are important because almost all countries have measures of child mortality. In this article, our goal is to absorb the essence of “A General Age-Specific Mortality Model with An Example Indexed by Child or Both Child and Adult Mortality” (Clark 2016) by replicating it. The article of Clark (2016) defines a generic parameterized mortality component model (SVD-COMP) using singular value decomposition and calibrates the relationship between child or child/adult mortality and mortality rates at other ages in the mortality table observed in the human mortality database. It also verifies the model by cross-validation and compares the predictive performance of the model with that of the log-quaternion model, which indicates the efficiency of the proposed method.

**keywords:** Mortality; SVD; HMD; SVD-Comp

## 1 Introduction

A complete age-disaggregated mortality table is an essential input for the various formal demographic and epidemiological approaches. A key example is the biennial World Population Outlook prepared by the United Nations Population Division (United Nations, Department of Economic and Social Affairs, Population Division 2015). These indicators are generally considered to be reference demographic indicators and are widely used by other national and international agencies as inputs to estimation and modelling efforts. The data contain estimates of mortality, fertility and population size by sex and age from 1950 to the present, as well as projections for the same number of countries in the world by 2100. Therefore, each WPP update must include a complete timeline of deaths at a particular age between 1950 and 2100.

The paper focuses on mortality, finding that Some countries in the developing world, particularly in Africa, where there is no information on child or adult mortality. there are 50 countries in the world (with a total population of nearly 1 billion) with no information on adult mortality, most of them in Africa – 33 countries with a total population of 666 million. Li (2015) shows detailed household surveys on fertility and maternal and child health, and there is essentially at least some up-to-date information on child mortality around the world.

In the absence of a full mortality record in many poor countries, we consider using the mortality models to make up for this loss.

---

\*Codes are available at: [github.com/LuyuanHu/Stat304-PS5](https://github.com/LuyuanHu/Stat304-PS5). The data come from the Human Mortality Database (HMD) and are available online at [www.mortality.org](http://www.mortality.org).

The standard approach for producing complete age tables for countries and areas with inadequate data is to take advantage of the fact that they do have information on child mortality. Typically, the model life table is used to infer a full death schedule from  ${}_5q_0$ . The United Nations Population Division has also (made extensive use of the traditional Coale and Demeny (1966) model life table} and updated {Li (2015)} with incomplete mortality data for many countries and regions of the world), and the Institute for Health Indicators and Assessment (Health) has done the same using the variant Modified logarithm (Mod-Logit) model Murray et al. (2003).

Life table of commonly used model system – zone model Coale and Demeny (1966) population life table and stability and developing country life table United Nations, Department of Economic and Social Affairs, Population Division (1982), modify the Logit life table system (Mod - Logit) (Murray et al. (2003)). All of these are empirical models in that they summarize observed mortality rates and use that summary to produce mortality prediction tables that are consistent with observed mortality rates. They are both regional and continuous. Regional models identify and replicate commonly observed mortality patterns associated with geographic regions and allow mortality rates to vary within specific patterns within each region. In contrast, the pattern of mortality produced by the continuous model varied steadily. Both methods are essentially two-parameter models. The regional model first identifies a discrete region and then effectively adjusts the level of mortality in a particular region by using continuously changing life expectancy within each region. The continuum model has two parameters that change continuously, such as life expectancy, child mortality or adult mortality.

Murray et al. (2003) enumerate three characteristics required of mortality models: 1) simplicity and ease of use, 2) comprehensive representation of the true variability in sex-age-specific mortality observed in real populations, and 3) validity that is well quantified by comparing age schedules of mortality predicted by the model to corresponding observed life tables. This paper added 1) generality with respect to the underlying model structure, 2) flexibility in terms of input parameters, and 3) an ability to handle a wide range of age groups, including very fine-grained, without having to fundamentally alter the structure of the model.

Clark (2016) defines and describes a new SVD component-based mortality modeling framework that satisfies all of those requirements. The SVD-component framework provides a general, flexible way to model any demographic age schedule as a function of covariates or predictors that are related to age-specific variation in the age schedule. Here the SVD-component framework is demonstrated by creating a mortality model that predicts single-year of age mortality schedules using either  ${}_5q_0$  or both ( ${}_5q_0$ ,  ${}_{45}q_{15}$ ) as predictors, similar to both the Mod-Logit and Log-Quad models. The resulting model can be used to produce single year of age mortality schedules from  ${}_5q_0$  alone that are consistent with observed mortality schedules, and this could be useful for those like the UN Population Division who must manipulate full age schedules of mortality but only have observed values for  ${}_5q_0$ . The resulting SVD-component model performs better than the current state of the art two-parameter model (Log-Quad), provides predictions by single-year of age, and is easily extensible to include additional predictors beyond child and adult mortality.

## 2 Data

### 2.1 Human Mortality Database Life Tables - HMD

The Human Mortality Database (HMD) University of California, Berkeley and Max Planck Institute for Demographic Research (n.d.) contains rigorously cleaned, checked and validated information on deaths and exposure from a number of developed countries where death registration and census data are virtually complete.

The data are aggregated and presented in a wide variety of formats. The objective of this analysis is to capture and characterize as much variability in age-specific mortality as possible, and consequently I chose to use the  $1 \times 1$  HMD life tables for each sex. Those provide all columns of a standard life table for single calendar years by single year of age from  $0 \rightarrow 110+$ . Each country provides data for different historical periods, and some countries are subdivided into more specific subpopulations. In the latter situation a ‘national population’ life table is typically provided that aggregates across the subgroups. Both the national

and subgroup populations are included in this analysis to maximize the variability in age-specific mortality schedules in the overall dataset. A few of the  $1 \times 1$  life tables from the HMD contain problems: The excluded tables have been updated to reflect the updates in HMD through August 21, 2018.

HMD nomenclature describes age $\times$ period life tables. For example  $1 \times 1$  are single calendar year by single year of age, and  $5 \times 5$  are five-year age groups by five-year periods, with the first age group broken into 0 and 1–4 years.

## 2.2 Clean HMD and Additional Indicator Calculation

There are two persistent problems with the HMD  $1 \times 1$  life tables: 1. some of the Belarus life tables are empty, and 2. the  ${}_1q_x$  values for some life tables are ‘flat’ at older ages, i.e. are constant.

In both cases, these life tables need to be removed. We’ll get rid of the Belarus tables first. The strategy is general: identify life tables with ‘NA’ values and remove those. They turn out to be Belarus 1914–1918.

After checking how many life tables are left, we are sure all the data objects have the same number of life tables and age groups.

We need child,  ${}_5q_0$ , and adult,  ${}_{45}q_{15}$ , mortality values for females and males, which can be Calculated from the  $1 \times 1$   ${}_nq_x$  values and store in separate matrices, including the log and logit transformed values.

## 3 Mortality Model

Traditional model life tables (e.g. United Nations, Department of Economic and Social Affairs, Population Division 1955) take an inductive, empirically-driven approach to identify and parsimoniously express the regularity of mortality with age based on observed relationships in large collections of high quality life tables. An alternative, sometimes deductive approach, can be found in the wide variety of parametric or functional-form mortality models that define age-specific measures of mortality in an analytical form, sometimes with interpretable parameters. Brass (1971) developed an innovative new approach with his two-parameter ‘relational’ model that has been extended and refined in many ways. More recently the Log-Quad model combines empirical and functional-form approaches to mortality models.

Population forecasting has motivated another important family of related mortality models. Forecasting generates many iterations of both age-specific mortality and fertility into the future, and those are usually based on a summary of the corresponding age-specific mortality and fertility in the past. Hence there is an immediate need to represent full age schedules and their dynamics compactly. This led to the widespread use of dimension-reduction or data compression techniques to reduce the dimensionality of the problem so that only a few parameters are necessary to represent age schedules and their dynamics.

Ledermann and Breas (1959) appear to have been the first to use principal components analysis (PCA) to summarize age-specific mortality and generate model life tables, and this approach was refined by many subsequent investigators. Following the early use of PCA to build model life tables, PCA and related methods like the singular value decomposition (SVD) have been widely used and refined by forecasters to create time series models of mortality and fertility. Bell (1997) provides a comprehensive summary of this line of development in various fields, dominated by actuarial science and applications in forecasting.

### 3.1 Model Scales

This analysis is conducted on life table probabilities of dying for those who survive to the beginning of each one-year age group. Single year probabilities  ${}_1q_x$  are taken directly from the HMD life tables, five-year probabilities  ${}_5q_x$  are calculated as  ${}_5q_x = 1 - \prod_{a=x}^{x+4} (1 - {}_1q_a)$ , and  ${}_{45}q_{15}$  is calculated as  ${}_{45}q_{15} = 1 - \prod_{a=15}^{59} (1 - {}_1q_a)$ . Child mortality refers to  ${}_5q_0$  and ‘adult mortality’ refers to  ${}_{45}q_{15}$ .

The natural scale of the models described below is the full real line, so life table probabilities of dying  $q$  are transformed using the *logit* function

$$\text{logit}(x) = \ln\left(\frac{x}{1-x}\right)$$

so that their transformed values occupy the full real line. Outputs from the models are transformed back to the probability scale with range  $[0,1]$  using the *expit* function, inverse of the *logit*

$$\text{expit}(x) = \frac{e^x}{1 + e^x}.$$

### 3.2 SVD Component Model of Mortality

This section is about the most of the operations needed to calculate and validate SVD-Comp models. The major procedure is:

- Calculate/estimate an SVD-component mortality model using a set of age-specific  ${}_nq_x$  as inputs
- Calculate/estimate a smoothed SVD-component mortality model using a set of age-specific  ${}_nq_x$  as inputs
- Randomly sample a set of age-specific  ${}_nq_x$ , calculate an SVD-component model of mortality (smoothed or not), predict  ${}_nq_x$  for the not-sampled age-specific  ${}_nq_x$ , and summarize the prediction errors
- All of this can be repeated a specified number of times
- The return object contains very detailed results for everything that was requested

$\mathbf{U}$  is a matrix of ‘left singular vectors’ (LSVs) arranged in columns,  $\mathbf{V}$  is a matrix of ‘right singular vectors’ (RSVs) arranged in columns, and  $\mathbf{S}$  is a diagonal matrix of ‘singular values’ (SVs). The LSVs and RSVs are independent and have unit length. If one views the columns of  $\mathbf{X}$  as a set of dimensions, then the rows of  $\mathbf{X}$  locate points defined along those dimensions – the data cloud. The RSVs define a new set of dimensions that line up with the axes of most variation in the data cloud. The first RSV points from the origin to the data cloud, or if the cloud is around the origin, then it points along the line of maximum variation within the cloud. The remaining RSVs are orthogonal to the first and each other and line up with successively less variable dimensions within the cloud. The elements of the LSVs are values that correspond to the projection of each point along the new dimensions defined by the RSVs. The SVs effectively stretch the new dimensions defined by the RSVs in accordance with the variation in the cloud along each RSV. The numeric value of each SV is the square root of the sum of squared distances from the origin to each point along the corresponding SVD dimension, and their squares sum to the total sum of squared distances from the origin to each point along all of the original dimensions.

The basic form of the SVD can be rearranged to yield two new useful expressions

$$\mathbf{X} = \sum_{i=1}^{\rho} s_i \mathbf{u}_i \mathbf{v}_i^T$$

$$\mathbf{x}_\ell = \sum_{i=1}^{\rho} s_i \mathbf{v}_{\ell i} \mathbf{u}_i$$

where  $\mathbf{u}_i$  are LSVs,  $\mathbf{v}_i$  are RSVs,  $s_i$  are SVs,  $\rho$  is the rank of  $\mathbf{X}$ ,  $\mathbf{x}_\ell$  are columns of  $\mathbf{X}$ , and  $v_{\ell i}$  are the elements of RSV  $\mathbf{v}_i$ .  $\mathbf{X}$  can be written as a sum of rank-1 matrices, each created from one of the LSVs by applying weights in the form of the elements of the corresponding RSV. Equivalently, each column  $\mathbf{x}_\ell$  of  $\mathbf{X}$  can be written as the weighted sum of the LSVs with the weight for each being the  $\ell$  element of the corresponding RSV. The LSVs and SVs are constant, so the the weights are the ‘variables’ in these expressions, and their values determine how much of each LSV is added to the mixture to represent the original data. Finally because the LSVs are independent, OLS regression can be used to estimate models that relate  $\mathbf{x}_\ell$  to the LSVs. If the constant is constrained to be zero, then the coefficients are equal to  $s_i v_{\ell i}$ .

Because the RSVs define successively less variable dimensions in the data cloud, the first term in Equations contains the most information and subsequent terms contain less and less. Including all  $\rho$  terms replicates the original data matrix  $\mathbf{X}$  or any of its columns  $\mathbf{x}_\ell$  exactly, while including only the first few terms provides a good approximation.

Additionally, the input age-specific  ${}_nq_x$  must all be logit-transformed, the function assumes this and uses an *expit* transformation to do predictions on the natural scale. The ‘mods’ return object is very useful for doing predictions and building additional modeling features using the return object of this function. The ‘retAll’ option is included because full results can be very large, and returning the summaries is a much more compact way to do things if you need to run many times and don’t need the detailed results each time.

Using SVD Component Model of Mortality, we can also obtain a dataframe containing the predicted life tables.

### 3.3 Validation

Use one iteration and a 100% sample in both ‘base’ and ‘smoothed’ form using only child mortality as a direct input, which will yield SVD-Comp models calibrated on the entire HMD data set. The process provides a little feedback, here indicating that two-component models were run on one sample of 100% with the child mortality-only model, either base or smoothed.

## 4 Results

To compare the predicted results from the SVD-comp model calibrated with the entire HMD to results produced by Wilmoth et al.’s Log Quad model, we must calculate five-year age group probabilities of dying,  ${}_tq_x$ , because Log Quad operates only with five-year age groups. The following code uses the single-year age group predictions from SVD-comp to calculate five-year age group probabilities of dying.

Then compare the models first using only child mortality  ${}_5q_0$  to predict and then using both child  ${}_5q_0$  and adult  ${}_{45}q_{15}$  mortality to predict.

Regression models are defined that relate the RSVs  $\mathbf{v}_{zi}$  to  ${}_5q_0$  and  ${}_{45}q_{15}$ . Scatterplots of the elements of the RSVs versus  $\text{logit}({}_5q_0)$  in Figures 1 and 2 (on-line appendices) make it clear that the relationships are not linear or simple. With no theory to guide the choice of predictors, I tried all combinations of simple transformations of  $\text{logit}({}_5q_0)$  and  $\text{logit}({}_{45}q_{15})$  and their interactions. The resulting models explain almost all the variance in the elements of  $\mathbf{v}_1$  ( $R^2 \approx 97\%$  for both sexes), the vast majority of the variance in the elements of  $\mathbf{v}_2$  ( $R^2 \approx 86\%$  for both sexes), and about one third of the variance in the elements of  $\mathbf{v}_3$  and  $\mathbf{v}_4$ . Additionally, I tried to avoid overfitting or creating odd boundary effects in the predicted values that would have made out-of-sample predictions immediately implausible. These models behave sensibly up to the edges of the sample. The final models are

$$v_{z\ell i} = c_{zi} + \beta_{z1i} \cdot {}_5q_0 + \beta_{z2i} \cdot \text{logit}({}_5q_0) + \beta_{z3i} \cdot \text{logit}({}_5q_0)^2 + \beta_{z4i} \cdot \text{logit}({}_5q_0)^3 + \beta_{z5i} \cdot {}_{45}q_{15} + \beta_{z6i} \cdot \text{logit}({}_{45}q_{15})^2 + \beta_{z7i} \cdot \text{logit}({}_{45}q_{15})^3$$

where  $i \in \{1 : 4\}$  indexes the SVD dimensions and  $\ell$  indexes mortality schedules and elements of  $\mathbf{v}_{zi}$ . OLS regression is used to estimate coefficients for the eight regression models. Using new values for both  ${}_5q_0$  and  ${}_{45}q_{15}$  as inputs, these models are used to predict values for the weights.

Figure 1 is  ${}_1q_x$  for very high mortality early in Sweden’s time series and low mortality for a more recent year in Austria and predicted values produced using  ${}_5q_0$  alone as an input. Data as symbols and predicted values as lines, which displays  ${}_1q_x$  on the logit scale for Sweden in 1751 and Austria in 1990, with both data and predicted values produced by SVD-Comp using  ${}_5q_0$  alone as an input.

First four LSVs scaled by their corresponding SVs from the SVD of the mortality schedules in the HMD. The more variable lines are raw components and less variable lines have been smoothed with a kernel smoother.

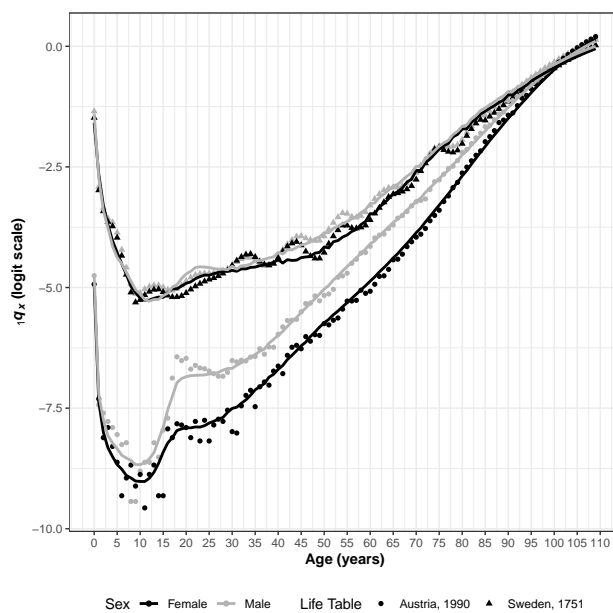


Figure 1: Example data and predictions.

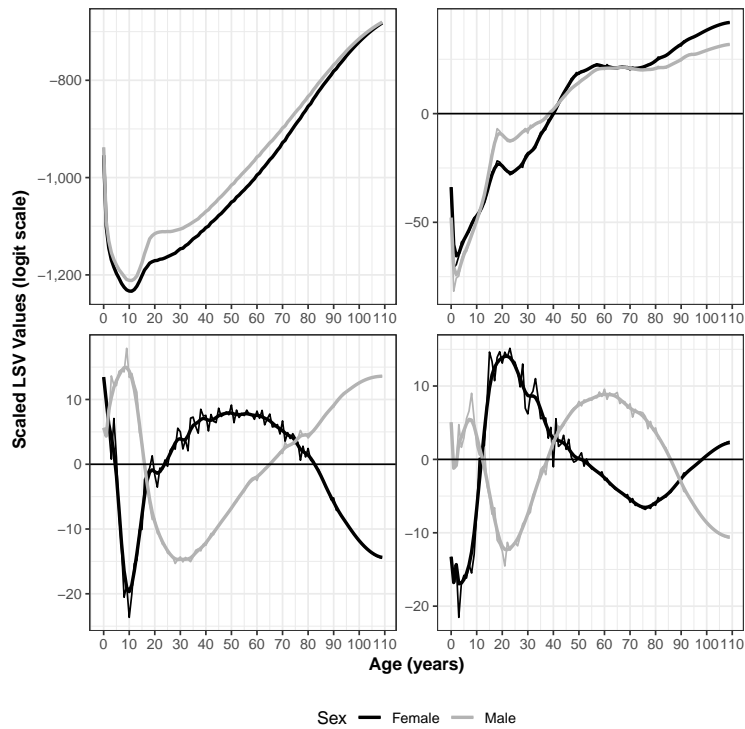


Figure 2: Scaled Left Singular Vectors.

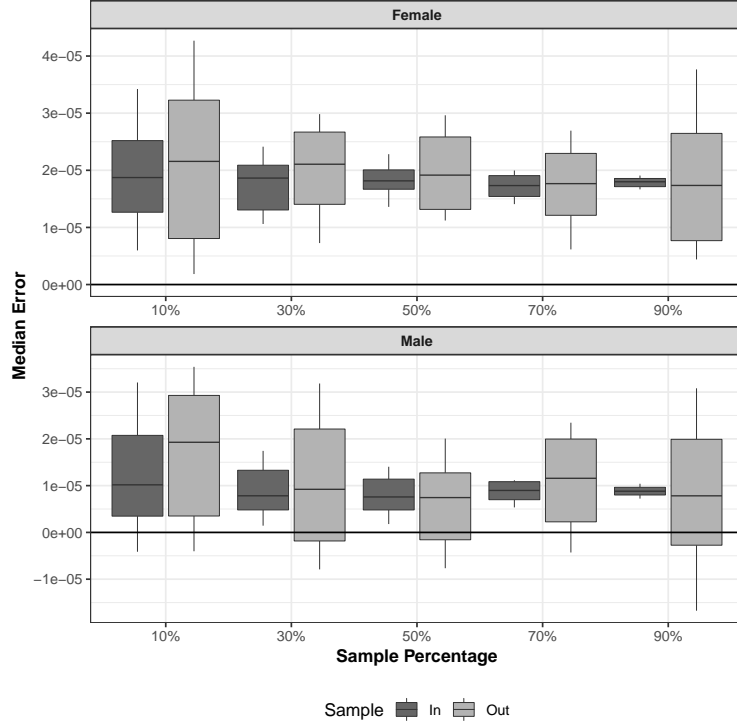


Figure 3: Median Prediction Error by Sample Fraction.

The raw values are used throughout this work. Figure 2 also displays smoothed8 versions of the scaled LSVs. The smoothed versions can be used to make the predicted mortality schedules smoother.

Figure 3 demonstrate that 50 samples for each sample fraction. For each sample, median calculated across all ages and all mortality schedules in each sample category (in/out). Whiskers extend to 10% and 90% quantiles.

Figure 4 is the application to Mexico and South Africa, in which data and predicted values in standard five-year age groups produced by Log-Quad and SVD-Comp models using both child and adult mortality as predictors.

Table 1: Life tables

Region	Abbre.	Years Covered	Total Life Tables
Australia	AUS	1921–2014	94
Austria	AUT	1947–2017	71
Belgium	BEL	1841–1913	73
Belgium	BEL	1919–2015	97
Bulgaria	BGR	1947–2010	64
Belarus	BLR	1959–2016	58
Canada	CAN	1921–2011	91
Switzerland	CHE	1876–2016	141
Chile	CHL	1992–2008	17
Czechia	CZE	1950–2016	67
East-Germany	DEUTE	1956–2015	60
Germany	DEUTNP	1990–2015	26
West-Germany	DEUTW	1956–2015	60
Denmark	DNK	1835–2016	182
Spain	ESP	1908–2016	109
Estonia	EST	1959–2017	59
Finland	FIN	1878–2015	138
France-Civilian-Population	FRACNP	1816–2016	201

Table 1: Life tables (*continued*)

Region	Abbre.	Years Covered	Total Life Tables
France-Total-Population	FRATNP	1816–2016	201
England-and-Wales–Civilian-National-Population	GBRCENW	1841–2016	176
England-and-Wales–Total-Population	GBRTENW	1841–2016	176
Northern-Ireland	GBR\_NIR	1922–2016	95
United-Kingdom	GBR\_NP	1922–2016	95
Scotland	GBR\_SCO	1855–2016	162
Greece	GRC	1981–2013	33
Croatia	HRV	2002–2016	15
Hungary	HUN	1950–2017	68
Ireland	IRL	1950–2014	65
Iceland	ISL	1838–1851	14
Iceland	ISL	1853–2016	164
Israel	ISR	1983–2014	32
Italy	ITA	1872–2014	143
Japan	JPN	1947–2016	70
Korea	KOR	2003–2016	14
Lithuania	LTU	1959–2017	59
Luxembourg	LUX	1960–2014	55
Latvia	LVA	1959–2017	59
Netherlands	NLD	1850–2016	167
Norway	NOR	1846–2014	169
New-Zealand–Maori	NZL\_MA	1948–1948	1
New-Zealand–Maori	NZL\_MA	1950–1955	6
New-Zealand–Maori	NZL\_MA	1957–1958	2
New-Zealand–Maori	NZL\_MA	1960–2008	49
New-Zealand–Non-Maori	NZL\_NM	1901–2008	108
New-Zealand	NZL\_NP	1948–2013	66
Poland	POL	1958–2016	59
Portugal	PRT	1940–2015	76
Russia	RUS	1959–2014	56
Slovakia	SVK	1950–2014	65
Slovenia	SVN	1983–2014	32
Sweden	SWE	1751–2016	266
Taiwan	TWN	1970–2014	45
Ukraine	UKR	1959–2013	55
USA	USA	1933–2016	84

Table 1 contains an organized list of the life tables included in this analysis: 4,610 life tables for each sex and 9,220 in total.

Table 2 & 3 displays the total absolute errors on the natural scale for the SVD-Comp and LogQuad models for predictions based on either  ${}_5q_0$  alone or both  ${}_5q_0$  and  ${}_45q_{15}$ . The table also presents differences between the total absolute errors for the two models in both additive (Log-Quad - SVD-Comp) and proportional form  $((\text{Log-Quad} - \text{SVD-Comp}) / \text{SVD-Comp})$ . In all cases, the SVD-Comp model predictions are globally closer to the HMD life tables

Table 2: Summary of prediction errors for SVD-Como and Lou-Quad of female

Female	model	C1	C2	C3 C2-C1
R1	SVD-Comp	1,496	1,336	-159.0
R2	Log-Quad	1,554	1,449	-105.0
R3	R2-R1	59	113	54.0
R4	R3/R1 (\%)	3.9	8.4	-33.8



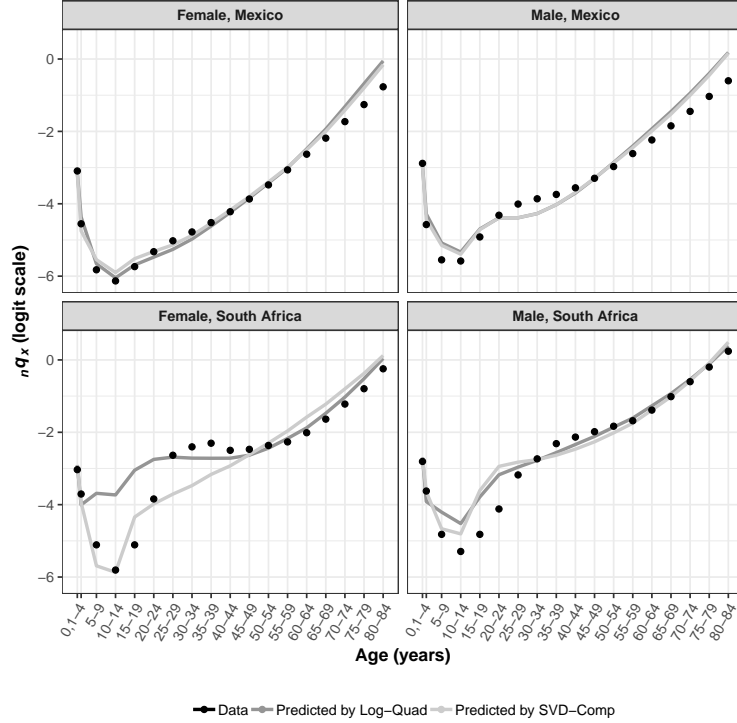


Figure 4: Application to Mexico and South Africa.

Table 3: Summary of prediction errors for SVD-Como and Lou-Ouad of male

Male	model	C1	C2	C3	C2-C1
R1	SVD-Comp	1,722	1,415	-308.0	
R2	Log-Quad	1,838	1,520	-318.0	
R3	R2-R1	116	106	-11.0	
R4	R3/R1 (\%)	6.7	7.5	3.4	

## 5 Discussion

Mortality models often have inbuilt identification issues challenging the statistician. The statistician can choose to work with well-defined freely varying parameters, derived as maximal invariants in this paper, or with ad hoc identified parameters which at first glance seem more intuitive, but which can introduce a number of unnecessary challenges.

This article describes how to use SVD to develop a generic modeling framework for population age tables. Based on the mortality age patterns contained in the HMD, this article demonstrate this framework by creating and verifying an accurate mortality model of one or two variables. Several important advantages of this framework is below.

### 5.1 Straightforward and easy to understand and use

This approach is general and allows all-age (in arbitrarily fine age groups) mortality schedules to be predicted from any covariates that are related to age-specific mortality. This general relationship is quantified in the models that relate the weights to the covariates. Allowing this is the fact that the relationship of each age

to all others is maintained through the constant components derived from the SVD, and those intra-age relationships are affected all together through the weights on the components. This constrains the intra-age relationships and relates them to the covariates in a simple, flexible way.

When the weights are modeled as functions of child mortality and calibrated using the relationship between the empirical weights ( $v_{z\ell i}$ ) and child mortality in the HMD, the model serves the same purpose as the Log-Quad model, and it performs slightly better in a direct comparison, while having the advantage of producing mortality schedules by single year of age. It is important to note that this comparison is conducted with the Log-Quad as presented and that in that article the authors explicitly favored an estimation technique that would, they claimed, reduce estimation bias at the cost of having (slightly) larger prediction errors when evaluated against the historical dataset, a fact that is apparent. The published Log-Quad was calibrated to the slightly different and smaller set of HMD life tables that existed at the time and met the authors' criteria for inclusion. Consequently the results of the comparison will likely change if the Log-Quad were recalibrated using the same set of HMD life tables described and used here. However, given how robust the SVD-Comp is to the set of life tables used in calibration, this potential difference is unlikely to be large.

## 5.2 General and applicable to any demographic age schedule

Concerning calibration and complexity, the cross validation results clearly demonstrate that the calibration to the HMD is robust with respect to exactly which and how many mortality schedules are used, and SVD-Comp is no more complex than Log-Quad. SVD-Comp {uses} {one SVD calculation and} six regression models for each sex to capture the relationship between child mortality and mortality at other ages in the HMD {- 12 regression models in total}. Log-Quad {uses} {one SVD calculation and} one log-quadratic model of the general form  $\log({}_5m_x) \sim \log({}_5q_0) + \log({}_5q_0)^2$  for each five-year age group and another to refine the prediction of  ${}_1q_0$  for each sex - 46 regression models in total. {The total number of regression coefficients required by each model (for each sex) is: 44 for SVD-Comp and 70 for Log-Quad. The total number of discrete values required for prediction (for each sex) is: SVD-Comp - 484 (4.4/age group), and Log-Quad - 92 (3.8/age group). SVD-Comp makes predictions in single-year age groups, and Log-Quad makes predictions in five-year age groups. Comparing the complexity of the models is not easy and depends where one focuses, but it is clear that neither is obviously more/less complex than the other. Perhaps the only important difference in this respect is that there is nothing in the overall Log-Quad model to directly constrain the relationship of mortality at one age to another except for the quadratic form of the relationship between mortality at each age and  ${}_5q_0$ , whereas SVD-Comp manipulates a linear combination of age-specific vectors, so the relationships between ages are constrained to fall within the four-dimensional space defined by the four components used by SVD-Comp.}

Together with our earlier work on an HIV-calibrated version of SVD-Comp, this demonstration suggests that it is reasonable to expect that SVD-Comp could be calibrated in a variety of additional ways to produce useful models that relate age-specific mortality to, for example, life expectancy at birth (or some other age), GDP, geographic region, time period, epidemiological indicators, a combination of any of these, or something else. Moreover, subtle effects on the age structure of mortality such as the 'rotation' in age-specific mortality could be incorporated by adding the necessary elements to the models for the weights. The same approach could be applied to develop models for the difference between underlying age-specific mortality and age-specific mortality affected by specific shocks such as natural disasters, conflict or epidemic disease such as HIV. It is even possible to refine the Lee-Carter model in Equation by adding more components to the SVD-derived  $\mathbf{b}_x \mathbf{k}_t$  term so that the enhanced model could represent a wide range of age patterns instead of the constant age pattern included in the existing formulation. This would add more parameters to the model, but the payoff might be sufficient to make that worthwhile. Going further, the entire Lee-Carter model could be replaced by the SVD-Comp model which would give it the ability to model changing levels and age patterns of mortality independently and generally be more flexible.

### 5.3 Able to incorporate covariates or predictors in a unified way

The general SVD-Comp model can be used in another way to interpolate or smooth incomplete or noisy age schedules by simply using OLS regression of the incomplete mortality schedule against the corresponding elements of the first few components  $s_{zi}\mathbf{u}_{zi}$  with the constant constrained to be zero, and then predicting the full mortality schedule from all elements of the components and the coefficients estimated by the regression. Bayesian estimation can also be used to estimate the weights and their uncertainty.

### 5.4 Able to handle age groups of any granularity

The application to Mexico and South Africa confirmed that the HMD-calibrated SVD-Comp works at least as well as Log-Quad when applied to mortality schedules in populations well outside of the HMD. For South Africa neither model was able to reproduce the HIV/AIDS-related mortality bulge at adult ages. SVD-Comp produced plausible mortality schedules for both sexes that were as close as possible to South Africa's, given that it could not reproduce the bulge. In contrast, Log-Quad produced a plausible mortality schedule for males but a nonsensical schedule for females. These results reveal an urgent need to increase the diversity of mortality schedules available in freely-accessible archives like HMD, and in particular, an important need to compile much better mortality data for Africa and other developing world regions where age schedules of mortality are different from what has been observed in the developed world. Additionally, the South Africa application suggests that SVD-Comp may provide a stable framework to begin building mortality models that include epidemiological (e.g. HIV prevalence and ART coverage) and other predictors. Our earlier work using modeled data is a start, but building models using modeled data is of limited value so we must assemble reasonable large, high quality empirical mortality data sets from the places where models such as Log-Quad and SVD-Comp are most useful.

## References

- Bell, William R. 1997. “Comparing and Assessing Time Series Methods for Forecasting Age-Specific Fertility and Mortality Rates.” *Journal of Official Statistics* 13.
- Brass, William. 1971. “On the Scale of Mortality.” In *Biological Aspects of Demography*, edited by William Brass, 69–110. Taylor; Francis: London, UK.
- Clark, Samuel J. 2016. “A General Age-Specific Mortality Model with an Example Indexed by Child or Child/Adult Mortality.” *arXiv Preprint arXiv:1612.01408*.
- Coale, Ansley J., and Paul Demeny. 1966. *Regional Model Life Tables and Stable Populations*. Princeton University Press.
- Ledermann, Sully, and Jean Breas. 1959. “Les Dimensions de La Mortalité.” *Population* 14 (4): 637–82.
- Li, Nan. 2015. “Estimating Life Tables for Developing Countries.” 2014/4. <http://www.un.org/en/development/desa/population/publications/pdf/technical/TP2014-4.pdf>: United Nations Department of Economic; Social Affairs Population Division; United Nations.
- Murray, Christopher JL, Brodie D Ferguson, Alan D Lopez, Michel Guillot, Joshua A Salomon, and Omar Ahmad. 2003. “Modified Logit Life Table System: Principles, Empirical Validation, and Application.” *Population Studies* 57 (2): 165–82.
- United Nations, Department of Economic and Social Affairs, Population Division. 1955. *Age and Sex Patterns of Mortality: Model Life-Tables for Under-Developed Countries*. New York: United Nations Department of International Economic; Social Affairs Population Division.
- . 1982. *Model Life Tables for Developing Countries*. 77. New York: United Nations Department of International Economic; Social Affairs Population Division.
- . 2015. *World Population Prospects: The 2015 Revision*. New York: United Nations.
- University of California, Berkeley, and Max Planck Institute for Demographic Research. n.d. *Human Mortality Database*. <http://www.mortality.org> or <http://www.humanmortality.de>.