SVD-derived Life Tables indexed by Child Mortality

Samuel J. Clark

Department of Sociology, The Ohio State University, Columbus, USA

SELECT: Climbing Mortality Models

August 30 – September 2, 2022
Misurina, Italy

Overview

Preliminaries

SVD Component Mortality Model

HMD SVD-Comp Indexed by ${}_5\mathsf{q}_0$ or $({}_5\mathsf{q}_0, {}_{45}\mathsf{q}_{15})$

Results

Discussion & Next Steps

Motivation & plan

Motivation

- Parsimonious, general, flexible model of age-specific mortality necessary in many applications
- Often want to relate covariates and/or time series of covariates to full age schedules of mortality
- Specific application in mind: UN Population Division needs to transform ${}_5q_0 \rightarrow {}_1q_x$, $x \in \{0, 110\}$ for many countries without age-specific **adult** mortality

Overview

- This builds on mortality models that utilize dimension reduction approaches, e.g. Bourgeois-Pichat (1962, 1990); Lee and Carter (1992); United Nations, Department of Economic and Social Affairs, Population Division (1982); Wilmoth (1990); Wilmoth et al. (2012), etc.
- Develop new way of using the singular value decomposition (SVD) to model demographic age schedules
- Demonstrate the new approach by creating a model relating ${}_5q_0$ or $({}_5q_0, {}_{45}q_{15})$ to mortality at all ages

The singular value decomposition

(Good, 1969; Stewart, 1993; Strang, 2009)

The singular value decomposition (SVD) of a generic matrix $\mathbf{X}$ is

$$\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathsf{T}} \tag{1}$$

- $\mathbf{U}$ is a matrix of ‘left singular vectors’ (LSVs) arranged in columns
- $\mathbf{V}$ is a matrix of ‘right singular vectors’ (RSVs) arranged in columns
- $\mathbf{S}$ is a diagonal matrix of ‘singular values’ (SVs)
- The LSVs and RSVs are orthogonal and have unit length

- The RSVs are a new orthonormal basis for the points defined by the rows of $\mathbf{X}$
- The LSVs scaled by their corresponding singular values are the projections of the points on the RSV-defined axes
- **By construction, the first RSV points in the direction that captures the greatest possible variation in the cloud of points, and subsequent RSVs sequentially capture as much of the remaining variation as possible**

The singular value decomposition

(Good, 1969; Stewart, 1993; Strang, 2009)

Equation 1 can be rearranged so that each column, $\mathbf{x}_\ell$, of $\mathbf{X}$ is represented as the weighted sum of LSVs:

$$\mathbf{x}_\ell = \sum_{i=1}^{\rho} s_i v_{\ell i} \mathbf{u}_i \qquad (2)$$

- ρ is the rank of $\mathbf{X}$
- i indexes components of the SVD of $\mathbf{X}$: $i \leq \rho$
- $\mathbf{u}$ are the LSVs of $\mathbf{X}$
- v are elements of the RSVs of $\mathbf{X}$
- s are the singular values of $\mathbf{X}$
- ℓ indexes columns of $\mathbf{X}$ and elements of the RSVs, $v_{\ell i}$
- The fact that the first RSV is associated with the direction of greatest variation in the cloud of points defined by $\mathbf{X}$ means that the first term in this sum accounts for the bulk of the variation among the columns $\mathbf{x}_\ell$ of $\mathbf{X}$ (Golub et al., 1987)
- **In general, a small number of terms is sufficient to closely approximate $\mathbf{x}_\ell$**

The SVD of mortality schedules

Clark (2016, 2015)

To closely approximate mortality schedules

- Given an A (age) $\times$ L (life table) matrix $\mathbf{Q}$ of mortality schedules for each sex, calculate the $\text{SVD}(\mathbf{Q}_z) = \mathbf{U}_z \mathbf{S}_z \mathbf{V}_z^\mathsf{T}$
- Using the resulting factors as in Equation 2, each mortality schedule $\mathbf{q}_{z\ell}$ is approximated as the c-term sum

$$\mathbf{q}_{z\ell} \approx \sum_{i=1}^{c} v_{z\ell i} \cdot s_{zi} \mathbf{u}_{zi} \qquad (3)$$

- $z \in \{\text{female}, \text{male}\}$
- i indexes components of the SVD of $\mathbf{Q}_z$
- $c \leq \rho$, the rank of $\mathbf{Q}_z$
- $\ell \in \{1 \ldots L\}$ indexes mortality schedules
- *The LSVs $\mathbf{u}_{zi}$ and the SVs s_{zi} are constant across all mortality schedules*
- Schedule-specific information only associated with weights $v_{z\ell i}$, elements of the RSVs
- Usually $c \leq 3$ is sufficient to very closely approximate demographic age schedules (Clark, 2015)

The SVD component mortality model

Clark (2016, 2015)

The SVD component model of mortality (SVD-Comp) is a generalization of Equation 3 for any mortality schedule $\mathbf{q}_z$ that is not necessarily in $\mathbf{Q}_z$

$$\mathbf{q}_z = \sum_{i=1}^{c} w_{zi} \cdot s_{zi}\mathbf{u}_{zi} + \mathbf{r}_z \tag{4}$$

- $z \in \{\text{female}, \text{male}\}$
- i indexes components of the SVD of $\mathbf{Q}_z$
- $c \leq \rho$, the rank of $\mathbf{Q}_z$
- The LSVs $\mathbf{u}_{zi}$ and the SVs s_{zi} come from the SVD of $\mathbf{Q}_z$ and are constant across all mortality schedules
- $\mathbf{r}_z$ is a residual
- w_{zi} are weights for each component $s_{zi}\mathbf{u}_{zi}$ that together minimize $\mathbf{r}_z$

SVD-Comp mortality model and covariates

If we have covariates p_ℓ related to the level and age pattern of the mortality schedules in $\mathbf{Q}_z$, then they will be related to the SVD-derived weights $v_{z\ell i}$ in Equation 3, and we can build models $m(\cdot)$ of these relationships.

$$v_{z\ell i} = m_{zi}(p_\ell)$$

Using the models we can predict weights from new values of the covariates p'.

$$\widehat{w}_{zi} = m_{zi}(p')$$

Substituting the predicted weights back into Equation 4, we can predict full age schedules of mortality.

$$\widehat{\mathbf{q}}_z = \sum_{i=1}^{c} \widehat{w}_{zi} \cdot s_{zi}\mathbf{u}_{zi} + \mathbf{r}_z \tag{5}$$

Calibrating SVD-Comp to ${}_5q_0$ or $({}_5q_0, {}_{45}q_{15})$ using the HMD

Data

- HMD downloaded 2016-11 (University of California, Berkeley and Max Planck Institute for Demographic Research, 2016)
- 4,486 1×1 life tables for each sex (see paper for few tables not used)
- ${}_1q_x$ values transformed to logit scale – model operates on full real line
- Calculate ${}_5q_0$ and ${}_{45}q_{15}$ for each life table – these are 'covariates' in this model
- For each sex, SVD calculated on 111 $\times$ 4,486 matrix of logit$({}_1q_x)$ values using the `svd()` function in the `base` package of the statistical programming environment `R`
- Retain first *four* components of the SVDs for each sex, accounts for greater than 99.9% of the total sum of squares for both sexes

Calibrating SVD-Comp to ${}_5\mathrm{q}_0$ or $({}_5\mathrm{q}_0,\, {}_{45}\mathrm{q}_{15})$ using the HMD

Calibrating to ${}_5\mathrm{q}_0$ or $({}_5\mathrm{q}_0,\, {}_{45}\mathrm{q}_{15})$

1. Linear regression models $v_{z\ell i} \sim {}^1\mathrm{m}_{zi}({}_5\mathrm{q}_{0,z\ell},\ {}_{45}\mathrm{q}_{15,z\ell})$ are estimated and the coefficients retained, $i \in \{1, 2, 3, 4\}$
2. A linear regression model ${}_{45}\mathrm{q}_{15,z\ell} \sim {}^2\mathrm{m}_z({}_5\mathrm{q}_{0,z\ell})$ is estimated and the coefficients retained
3. A linear regression model ${}_1\mathrm{q}_{0,z\ell} \sim {}^3\mathrm{m}_z({}_5\mathrm{q}_{0,z\ell})$ is estimated and the coefficients retained

The ${}_5\mathrm{q}_0$ or $({}_5\mathrm{q}_0,\, {}_{45}\mathrm{q}_{15})$ calibrated SVD-Comp consists of

- The first four scaled LSVs from the SVDs of the HMD 1×1 life tables: $s_{zi}\mathbf{u}_{zi}$, $z \in \{\text{female}, \text{male}\}$ and $i \in \{1, 2, 3, 4\}$
- The coefficients related to models ${}^1\mathrm{m}_{zi}$, ${}^2\mathrm{m}_z$, and ${}^3\mathrm{m}_z$

Using the ${}_5q_0$ or $({}_5q_0, {}_{45}q_{15})$ calibrated SVD-Comp

1. Identify input values for ${}_5q_{0,z}$ and optionally ${}_{45}q_{15,z}$. If ${}_{45}q_{15,z}$ is not available, predict ${}_{45}q_{15,z}$ using the input value for ${}_5q_{0,z}$ and the regression coefficients corresponding to model ${}^2m_z$.
2. Use the input values for ${}_5q_{0,z}$ and ${}_{45}q_{15,z}$ obtained in step 1 and the regression coefficients corresponding to models ${}^1m_{zi}$ to predict values for the weights $\widehat{w}_{zi}$ defined Equation 5.
3. Insert the weights predicted in step 2 into Equation 5 to calculate a predicted age schedule of mortality probabilities $\widehat{\mathbf{q}}_z$.
4. If desired, improve the prediction of ${}_1q_{0,z}$ using the regression coefficients corresponding to model ${}^3m_z$ to directly predict ${}_1q_{0,z}$ from the input value of ${}_5q_{0,z}$ from step 1. Replace the first element of $\widehat{\mathbf{q}}_z$ with this predicted value for ${}_1q_{0,z}$.

Preliminaries

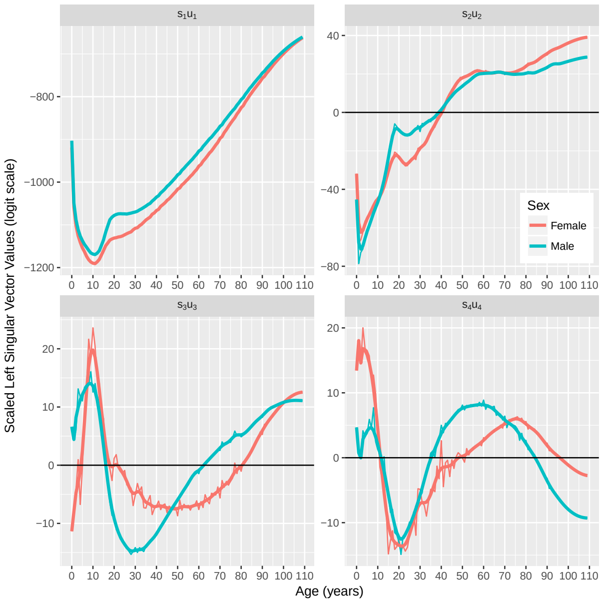
SVD Component Mortality Model

HMD SVD-Comp Indexed by ${}_5q_0$ or $({}_5q_0, {}_{45}q_{15})$

Results

Discussion & Next Steps

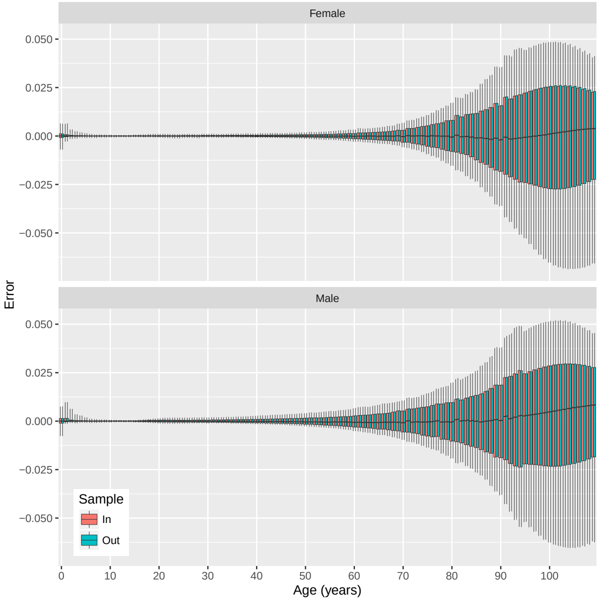
Scaled LSVs of HMD 1×1 mortality schedules logit$({}_1q_x)$: $s_{zi}\mathbf{u}_{zi}$



Sex	i	Prop. TSS
f	1	0.998
f	2	9.22×10^{-4}
f	3	6.73×10^{-5}
f	4	5.66×10^{-5}
f	1-4	0.999417
m	1	0.998
m	2	8.04×10^{-4}
m	3	9.76×10^{-5}
m	4	4.97×10^{-5}
m	1-4	0.999594

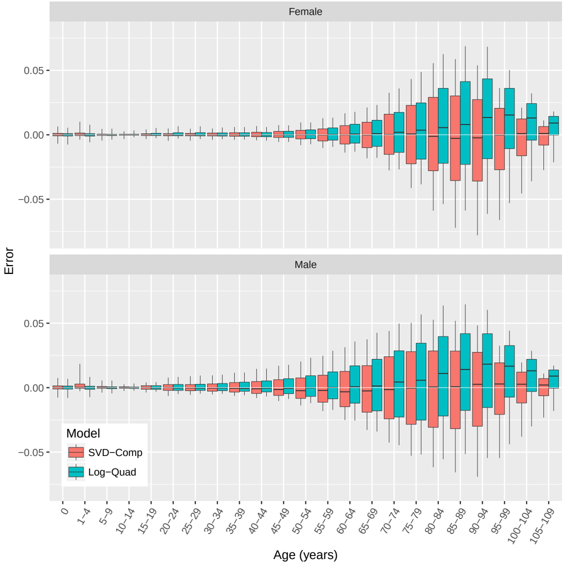
Sex	i	Prop. TSS
f	2	0.7185
f	3	0.0525
f	4	0.0442
f	2-4	0.8152
m	2	0.5925
m	3	0.0719
m	4	0.0366
m	2-4	0.7010

Validation



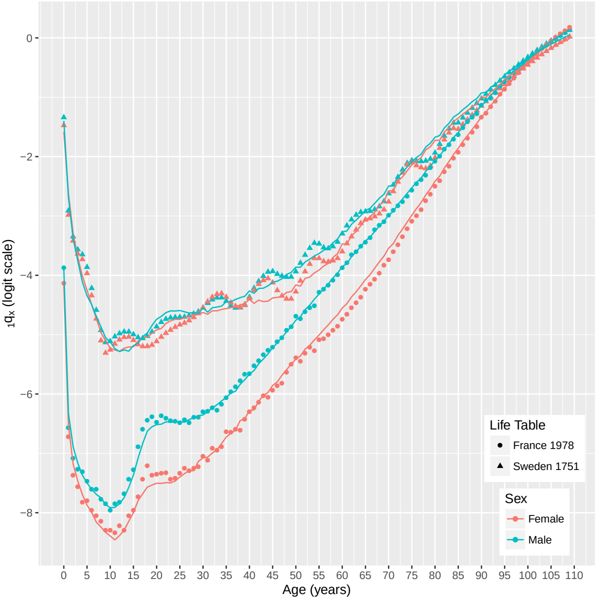
25 samples from HMD – 50% to calibrate; test on remaining 50%

Comparative performance: SVD-Comp vs. Log-Quad (Wilmoth et al., 2012)



Predict all ${}_5q_x$ from ${}_5q_0$; error distribution across HMD

Example predicted mortality schedules



Predict all $_1q_x$ from $_5q_0$

Summary & next steps

Summary

- Generalization of lots of previous work on dimension-reduction models for mortality
- Uncomplicated general framework for modeling demographic age structures, including both level and age structure
- Can easily include covariates in standard way
- Not sensitive to choice of age groups or time periods
- Can be applied to mortality, fertility, anything else with high degree of regularity

Next steps

- Developed version for UN Population Division World Population Prospects 2022: $\{{}_5q_0$, HIV prevalence, ART coverage$\} \to {}_1q_x$
- Working on age-specific fertility models, e.g. Pantazis and Clark (2018)

Contacts & Publication

- Contacts: samclark.net, work@samclark.net
- Slides: `http://samclark.net/site/talks/SVD-Comp-Italy-2022.pdf`
- Paper: A General Age-Specific Mortality Model With an Example Indexed by Child Mortality or Both Child and Adult Mortality. *Demography*, 2019.
- Reproducibility materials: `http://github.com/sinafala/svd-comp`
- R package
  - `https://github.com/sinafala/svd-comp/tree/master/package`
  - Install using *devtools*: `install_github(repo="sinafala/svdComp5q0")`

References I

Bourgeois-Pichat, J. (1962). Factor analysis and sex-age-specific death rates: a contribution to the study of the dimensions of mortality. *United Nations Population Bulletin* (6), 147–201.

Bourgeois-Pichat, J. (1990). Application de l'analyse factorielle à l'étude de la mortalité. *Population (french edition) 45*(4-5), 773–802.

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*.

CLARK†, S. J. (2019). A General Age-Specific Mortality Model with An Example Indexed by Child or Child/Adult Mortality (Reproducibility code). *Demography 56*(3), 1131–1159.

Golub, G. H., A. Hoffman, and G. W. Stewart (1987). A generalization of the eckart-young-mirsky matrix approximation theorem. *Linear Algebra and Its Applications 88*, 317–327.

Good, I. J. (1969). Some applications of the singular decomposition of a matrix. *Technometrics 11*(4), 823–831.

References II

Lee, R. D. and L. R. Carter (1992). Modeling and forecasting US mortality. *Journal of the American statistical association 87*(419), 659–671.

Pantazis, A. and S. J. Clark (2018). A parsimonious characterization of change in global age-specific and total fertility rates. *PloS one 13*(1), e0190574.

Stewart, G. W. (1993). On the early history of the singular value decomposition. *SIAM review 35*(4), 551–566.

Strang, G. (2009). *Introduction to Linear Algebra 4e*. Wellesley-Cambridge Press.

United Nations, Department of Economic and Social Affairs, Population Division (1982). *Model life tables for developing countries*. Number 77. New York: United Nations Department of International Economic and Social Affairs Population Division.

University of California, Berkeley and Max Planck Institute for Demographic Research (Downloaded November 2016). *Human Mortality Database*. `http://www.mortality.org` or `http://www.humanmortality.de`.

Wilmoth, J., S. Zureick, V. Canudas-Romo, M. Inoue, and C. Sawyer (2012). A flexible two-dimensional mortality model for use in indirect estimation. *Population Studies 66*(1), 1–28.

Wilmoth, J. R. (1990). Variation in vital rates by age, period, and cohort. *Sociological Methodology 20*, 295–335.