SVD Component Mortality Model

Samuel Clark, Patrick Gerland,
Brian Houle, Sara Hertog, Jonathan Muir, Jason Thomas

UNPD Experts' Group Meeting

'2020 Methods for the World Population Prospects 2021 and Beyond'

April 6-8, 2020
Virtual via Zoom

SVD-Comp project

Motivation

- One standard modeling framework for ASMR
- Single year of age model life tables for UNPD
- Incorporate unusual effects of HIV, conflict, and natural disasters via summary indicators
- 'Rotation' at older ages and low overall mortality as in 'LiLeeCarter'
- Smoothing, interpolation, (some) extrapolation

Background

- Builds on *long* history of mortality models that utilize dimension reduction approaches, e.g. Bourgeois-Pichat (1962); United Nations, Department of Economic and Social Affairs, Population Division (1982); Bourgeois-Pichat (1990); Wilmoth (1990); Lee and Carter (1992); Clark (2001); Wilmoth et al. (2012); Sharrow et al. (2014); Alexander et al. (2016); Clark (2019), etc.

Plan

We want something like this

$$\text{Model}({}_5q_0, {}_{45}q_{15}, P_{HIV}, \dots) \rightarrow {}_1q_x$$

Overview of plan

- Comprehensive database of empirical single-year ASMR
- ASMR affected by HIV, conflict, and natural disaster
- Using data, calibrate and validate SVD-Comp
- Software to implement and (re)calibrate SVD-Comp in a (semi)automated way
- Maybe estimation method to use model in reverse, e.g. ${}_1q_x \rightarrow \widehat{P}_{HIV}$, ...

SVD-Comp idea: SVD

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

Singular value decomposition of matrix Q of mortality probabilities, age (rows) $\times$ life table (columns)

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

Each life table, ℓ, age schedule is exactly

$$\mathbf{q}_{\ell} = \sum_{i=1}^{\rho} s_i v_{\ell i} \mathbf{u}_i \tag{2}$$

Each life table, ℓ, age schedule very well approximated as

$$\mathbf{q}_{\ell} \approx \sum_{i=1}^{4} \underbrace{v_{\ell i}}_{w_i} \cdot \underbrace{s_i \mathbf{u}_i}_{\mathbf{c}_i} \tag{3}$$

SVD-Comp idea: model

(Clark, 2019)

Estimate a model f relating elements of right singular vectors to covariate indicators

$$v_{\ell i} \sim f(\text{indicators}) \tag{4}$$

For new mortality schedule, predict weights from estimated $\hat{f}$

$$\widehat{w}_i = \hat{f}(\text{indicators}) \tag{5}$$

Using the predicted weights in the component model, predict a full age schedule of mortality

$$\widehat{\mathsf{q}}_\ell = \sum_{i=1}^{4} \widehat{w}_i \cdot \mathsf{c}_i \tag{6}$$

Data 1

Data sources for the universe of life tables from vital registration

Source	Abridged and/or complete	Number of countries or areas represented	Earliest year	Latest year	Number of life tables
Human Mortality Database (HMD)	Complete	47	1751	2018	10,161
Human Life-Table Database (HLD)	Both	99	1778	2015	6,863
Global Burden of Disease (GBD) life tables	Abridged	112	1950	2016	11,958
Latin American Mortality Database (LAMBdA)	Abridged	17	1908	2008	234
UN Demographic Yearbook (DYB)	Both	181	1908	2018	5,984
World Health Organization Historical Life Tables Databank (WHO)*	Abridged	63	1900	1999	4,066
Total					39,266

For non-HMD sources:
Abridged life tables are graduated to single year of age using the pclm function of the ungroup library in R.
Life tables are extended to open-age group 110+ by fitting a Kannisto function to the age-specific mortality rates of the last 20 single years of age, implemented through the MortalityLaws library in R.

* The WHO compilation includes the life tables referenced for the 1982 UN Model Life Table system as well as those used to calibrate the WHO Modified Logit Life Table System (Murray et al (2003))

Data 2

Life tables selected for the preliminary reference dataset, by SDG region and source

	Number of countries or areas	Number of life tables by source						
		HMD	HLD	GBD	LAMBdA	DYB	WHO	Total
More developed regions								
Australia/New Zealand	2	404	0	0	0	0	0	404
Europe and Northern America	53	7,196	78	240	0	167	77	7,758
Less developed regions								
Central and Southern Asia	11	0	32	275	0	64	4	375
Eastern and South-Eastern Asia	13	328	52	197	0	70	9	656
Latin American and the Caribbean	35	34	29	1,549	80	221	49	1,962
Northern Africa and Western Asia	16	68	29	152	0	94	0	343
Oceania (excl Australia and New Zealand)	8	0	2	9	0	14	0	25
Sub-Saharan Africa	4	0	16	103	0	14	10	143
Total	142	8,030	238	2,525	80	644	149	11,666

Selected from the universe of life tables by:

1. Privileging HMD life tables for countries or areas represented in that compilation
2. Excluding life tables with child mortality or old-age mortality that is outside of the experience represented in the HMD
3. Excluding life tables with atypically large fluctuations in mortality over age (similar to criteria used for GBD 2017)
4. Excluding country-years for which IHME estimates of death registration completeness < 85%
5. Removing duplicates to ensure inclusion of only one life table per country-year-sex

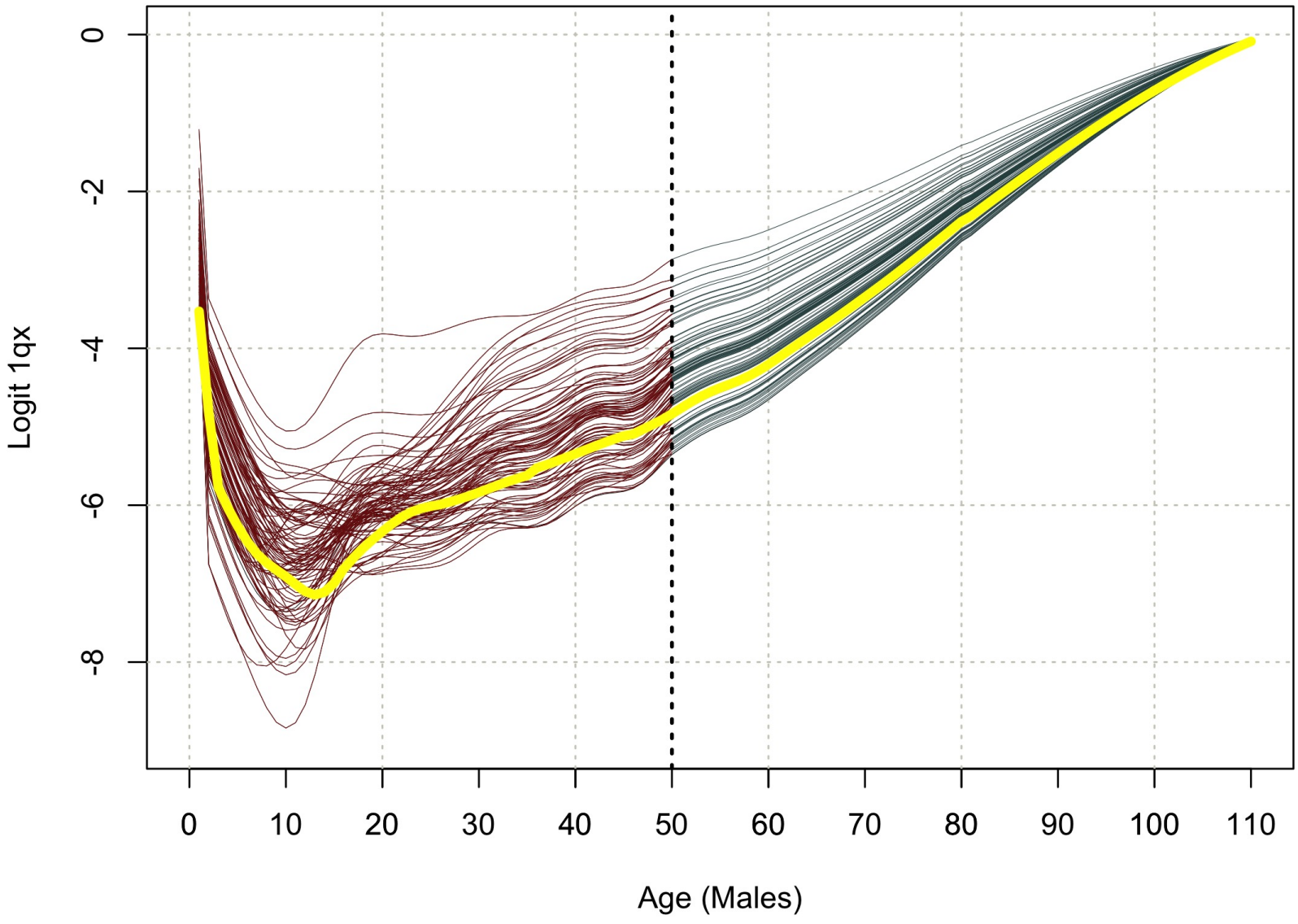
Data 3

Supplemental life tables from non-VR sources

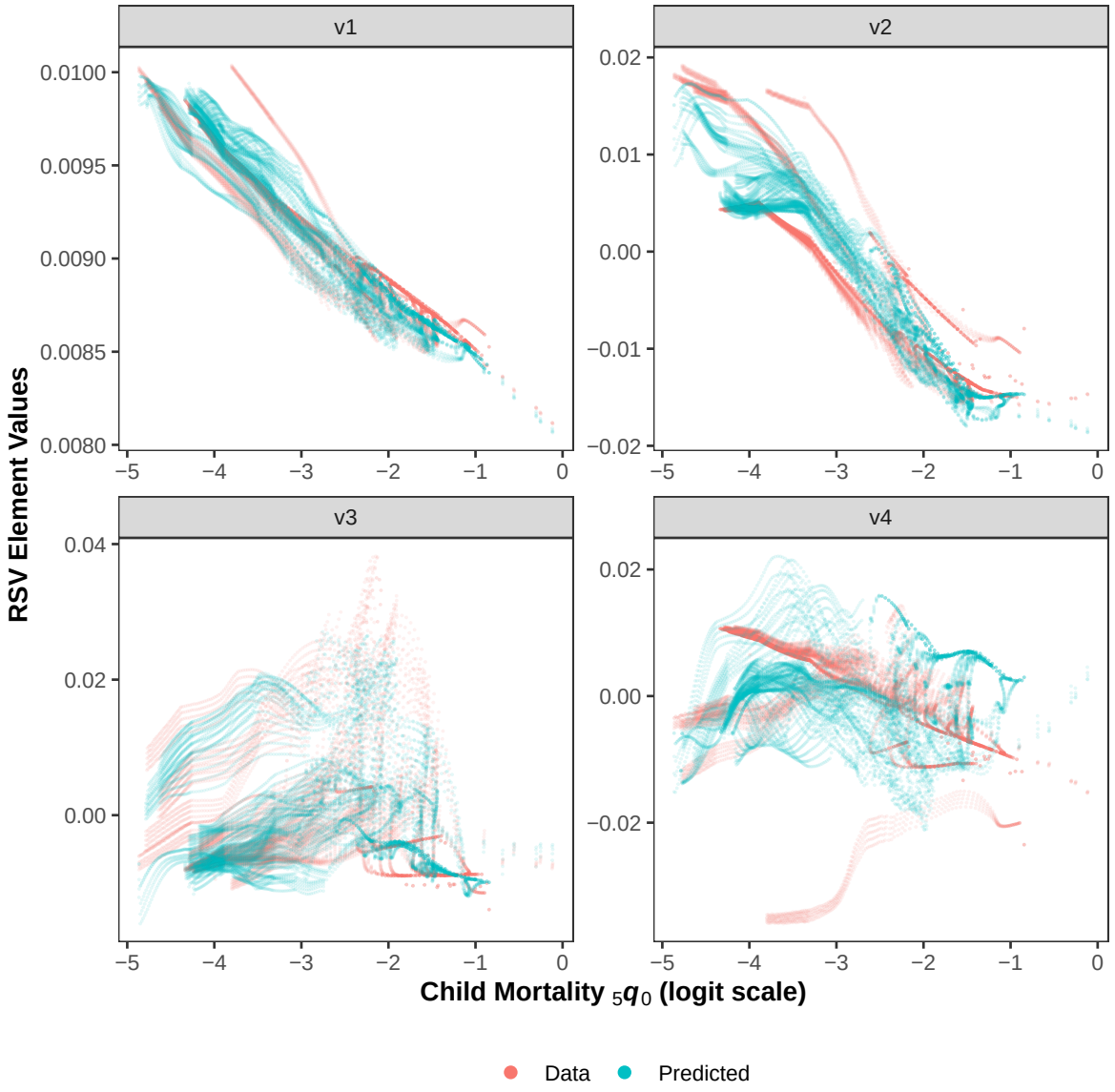
Source	Abridged and/or complete	Number of countries or areas represented	Earliest year	Latest year
Demographic and Health Surveys (DHS)	Truncated life tables constructed from: 1) full birth histories for ages 0-14; 2) sibling survival histories for ages 15-49 for the 10-year period preceding each survey	156 surveys from 61 countries*	1988	2018
Health and Demographic Surveillance System (HDSS) and urban VR for selected SSA	Life tables based on household deaths and population averaged over 5 or more years	42 sites across 19 countries*	1930	2016
Spectrum simulated life tables	Life tables simulated for each country-period with 5 scenarios for HIV incidence (observed, +/- 10% and 20%) and common ART assumptions	18 countries with HIV prevalence > 5% at any point during 1980-2018 per UNAIDS	1970	2100

* Inclusion criteria for bringing DHS and HDSS life tables into the reference dataset are still to be determined.

Extending truncated DHS life tables



Spectrum simulated data right singular vector (V) models, f



Model includes

- $_5q_0$
- $_{45}q_{15}$
- HIV prevalence
- ART coverage

Estimated using restricted cubic splines with 7 knots

Spectrum simulated data right singular vector (V) models, *f*

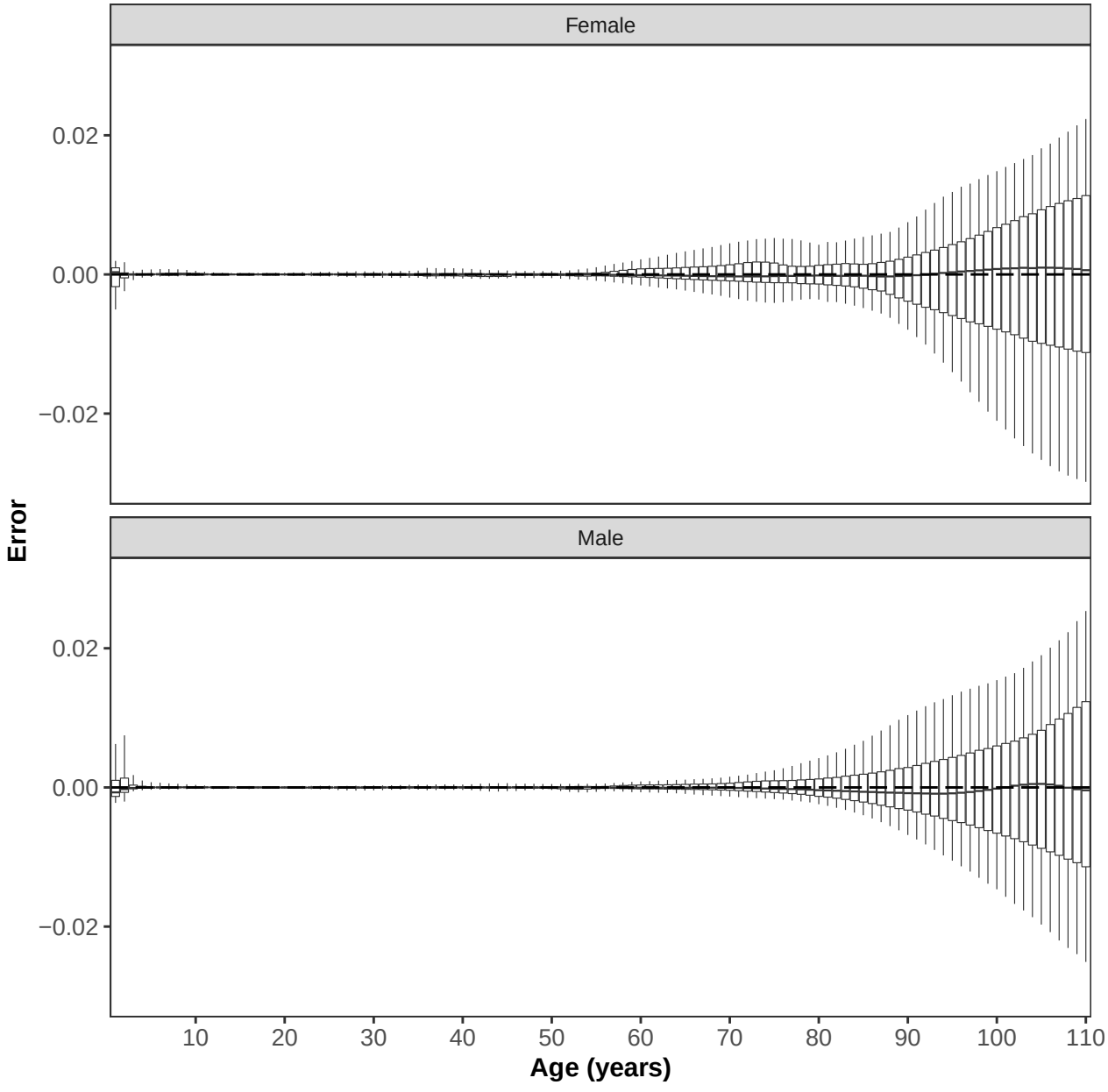
SS Explained by each Component

	Female		Male	
c	%	Cum. %	%	Cum. %
1	67.2	67.2	57.3	57.3
2	24.1	91.3	34.1	91.3
3	4.8	96.1	3.7	95.0
4	2.2	98.3	2.4	97.4

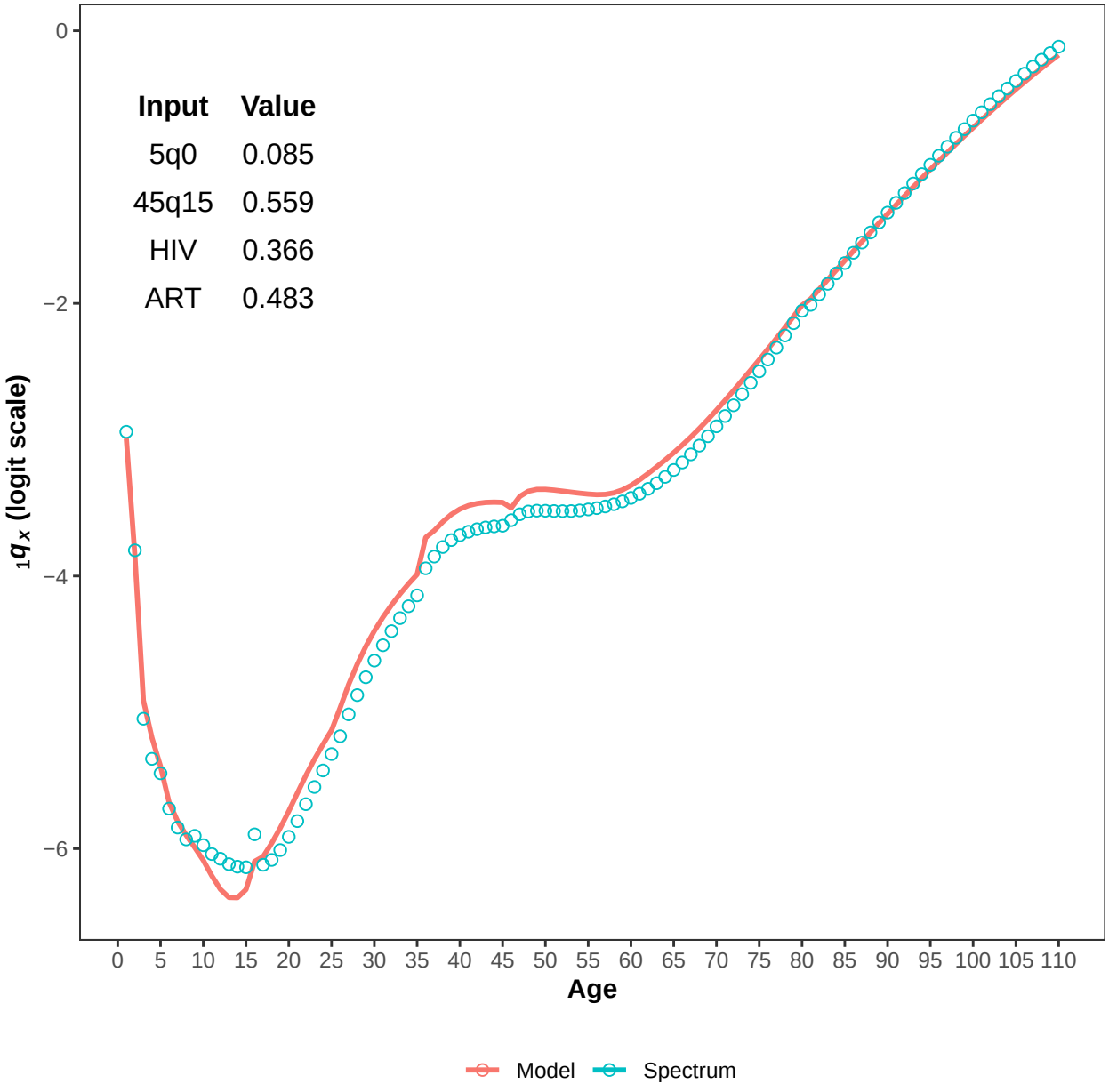
RSV Model Fits (female)

RSV	R^2	Adj R^2
1	0.999	0.999
2	0.976	0.976
3	0.988	0.988
4	0.910	0.909

Example prediction error distributions from Spectrum simulated data model



Example prediction from Spectrum simulated data model



Plan

Now

- Finish compiling empirical life table database
- Include detailed 1-year child mortality data, similar to truncated DHS
- Refine right singular vector models
- Build rotation model
- Characterize performance and validate overall SVD-Comp model
- Create R packages and R Shiny GUIs to implement in a user-friendly way

Medium term

- Replace/augment simulated Spectrum HIV mortality schedules with data from ALPHA Network of HIV surveillance HDSS sites
- Add conflict and natural disaster data
- Use SVD-Comp idea other applications: e.g. indirect methods, summary birth history estimates of ${}_5q_0$, and subnational estimates
- Refine R packages to allow semi-automated recalibration of overall model as empirical life table database grows

Publication fully describing this approach

(Clark, 2019)

Demography (2019) 56:1131–1159
https://doi.org/10.1007/s13524-019-00785-3

A General Age-Specific Mortality Model With an Example Indexed by Child Mortality or Both Child and Adult Mortality

Samuel J. Clark[1,2]

Published online: 28 May 2019

- Full reproducibility materials at https://github.com/sinafala/svd-comp
- `R` package that implements model
  - https://github.com/sinafala/svd-comp/tree/master/package
  - Install using *devtools*: `install_github(repo='sinafala/svdComp5q0')`

References I

Alexander, M., E. Zagheni, and M. Barbieri (2016). A flexible bayesian model for estimating subnational mortality. *arXiv preprint arXiv:1607.03534*.

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. (2001). *An Investigation into the Impact of HIV on Population Dynamics in Africa*. Ph. D. thesis, University of Pennsylvania, Philadelphia.

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

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

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

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

References II

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.

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.