Incorporating Mortality Estimation into Population Reconstruction

April 10, 2021

Brief Introduction to Mortality Modelling

Let T be the time (number of years) to death of a newborn (i.e. aged 0), we have

$$ {}_tq_x = P(T \leq t + x | T > x), $$

i.e. the probability that the newborn will die *within* t years given that she has survived for x years, then

$$ {}_tp_x = 1 - {}_tq_x = P(T \geq t + x | T > x) $$

is the probability that an individual aged x will survive *at least* t years.

Brief Introduction to Mortality Modelling

Some common demographic measures:

- ${}_5q_0$: child mortality, i.e. probability of a newborn dying *within* 5 years, before reaching age 5
- ${}_{45}q_{15}$: adult mortality, i.e. probability of an individual aged 15 dying *within* 45 years, before reaching age 60

Force of Mortality *vs* Death Probabilities

Force of mortality is defined as:

$$\begin{aligned}
\mu_{x+t} &= \lim_{\delta t \to 0} \frac{P(T \leq x+t+\delta t | T > x+t)}{\delta t}, \\
&= -\frac{1}{{}_tp_x} \frac{d}{dt} {}_tp_x, \\
&= -\frac{d}{dt} \log {}_tp_x, \\
\implies {}_tq_x &= 1 - {}_tp_x = 1 - e^{-\int_0^t \mu_{x+s}\, ds}.
\end{aligned}$$

Intuitively it can be thought as a measure of the instantaneous "death probability" at age $x + t$.

- the force of mortality is a *rate*, $\mu \in [0, \infty)$
- the death probability is a *probability*, $\in [0, 1]$

Central Mortality Rates

It is often impossible to model directly on μ.

A usual practice is to aggregate data into separate age groups (usually 1 year), with mortality measured by the *central mortality rates*:

$$ {}_1m_x = m_x = \frac{\int_0^1 E_{x+t}\mu_{x+t}\,dt}{\int_0^1 E_{x+t}\,dt} = \frac{\mathbb{E}({}_1d_x)}{{}_1E_x^C} $$

where ${}_1E_x^C$ and ${}_1d_x$ are the central exposures and number of deaths in age $[x, x+1)$ in 1 year.

- Intuitively the central mortality rate can be viewed as the average force of mortality weighted by the corresponding exposures
- Most mortality modelling is done on m_x.

Conversion between μ_x, m_x and q_x

As mentioned before, the exact relationship between μ_x and q_x is ${}_tq_x = 1 - e^{-\int_0^t \mu_{x+s}\, ds}$, however, this requires integration over the age range. Often simplifying assumptions are made, two common ones are:

1) Constant force of mortality → the force of mortality is constant within the age group

   $\Longrightarrow \; q_x = 1 - e^{-\mu_{x+0.5}} \approx 1 - e^{-m_x}$

2) Uniform distribution of deaths → deaths are uniformly distributed over the age range

   $\Longrightarrow \; q_x = \frac{m_x}{1+0.5m_x}$

Mortality Models

Recall

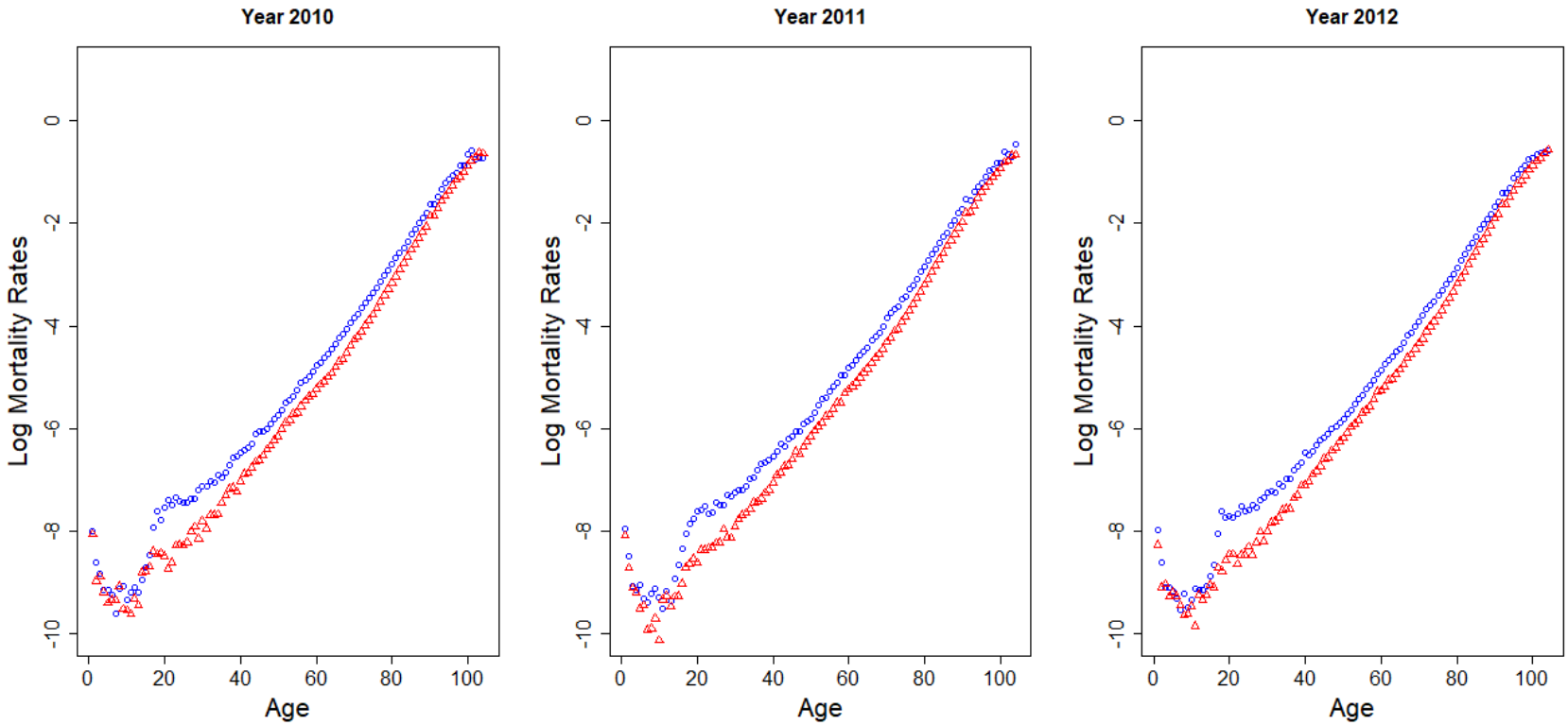
$$m_x = \frac{\mathbb{E}(d_x)}{E_x^C} \implies \mathbb{E}(d_x) = E_x^C \, m_x$$

Since d_x is a count variable, it is often modelled as Poisson distributions, i.e. $d_x \sim Poisson(E_x^C \, m_x)$. Alternatives include

- $d_x \sim Binomial(E_x^0, q_x)$, where E_x^0 is the initial exposures
- $\log(m_x) \sim N(\log(\hat{m}_x), \sigma)$, where $\hat{m}_x$ is the crude mortality rates
- the Poisson distribution is often too restrictive as $Var(d_x) = \mathbb{E}(d_x)$, and over-dispersion usually presents in mortality data.
  - ↪ introduce an extra parameter of specific form to capture the excessive variance, one arrives at a Negative Binomial distribution, i.e. $d_x \sim NegBinom(E_x^C \, m_x, s)$

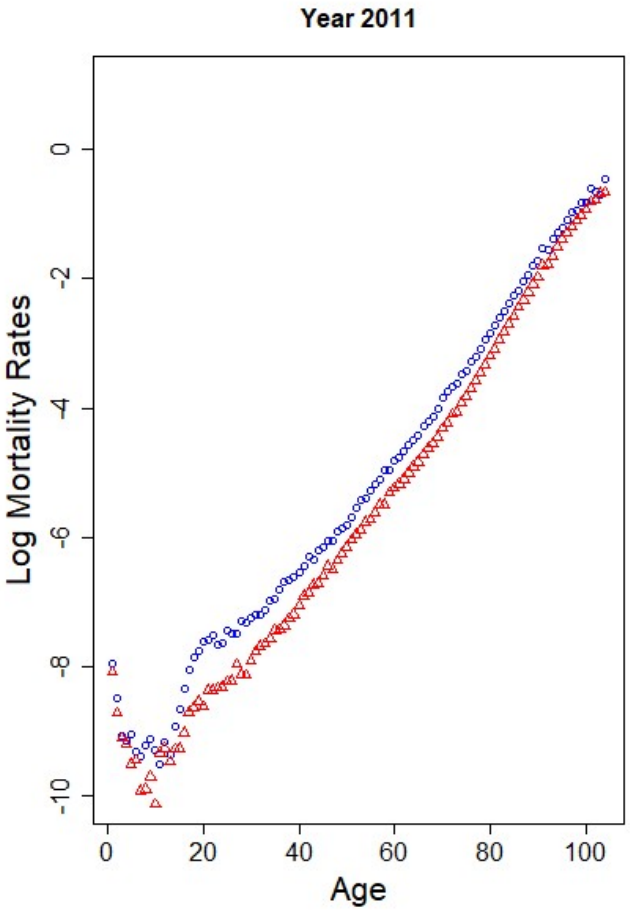
Mortality Models

Typical non-HIV infected mortality schedules, England and Wales males and females [1]



[1] Data obtained from the Office for National Statistics

Mortality Models



- male mortality (blue) is higher than female mortality (red) most of the time
- infant mortality is exceptionally high, followed by a sharp decrease into early teens
- humps can be observed at late teens to early adulthood, more prominent for males (accident hump / maternal hump)
- mortality increases steadily and relatively log-linearly after the mortality humps into older ages due to senescence

Mortality Models

One of the earliest attempts in mortality modelling dates back to 1825, the Gompertz law of mortality (Gompertz, 1825),

$$\log(m_x) = a + bx,$$

which is then modified to the Gompertz-Makeham law (Makeham, 1860) with the addition of an age invariant background mortality constant (age-unrelated deaths such as accidents)

$$m_x = A + BC^x$$

- relatively simple, with only few parameters
- provides adequate fit to mortality at older ages
- only applicable to older ages, due to the log-linearity assumption

Mortality Models

The mortality schedules often deviate from the log-linearity and level-off (mortality plateau) at the oldest ages (approx after 92, Carriere, 1992). Perks (1932), Beard (1959) and Thatcher (1999) proposed logistic models of the form

$$m_x = \frac{cz}{1+z} + \gamma$$

- at younger ages the logistic curve behaves similarly to a log-linear curve
- at the oldest ages provides a level-off to an asymptote
- again only applicable to older ages
- other attempts to capture the mortality plateau include Coale and Kisker (1990), Lindbergson (2001)

Mortality Models

Full age range models require relatively complicated forms due to the peculiar shape of mortality age patterns.

Several parametric models have been proposed which share a similar concept - decomposition of mortality into different stages: child mortality, young adult mortality and senescence, e.g. Thiele (1871), Heligman and Pollard (1980), Siler (1983) and Carriere (1992).

Some non-parameteric models have also been developed as well as a vast number of mortality models designed with focuses on projection, but I will not be discussing them here.

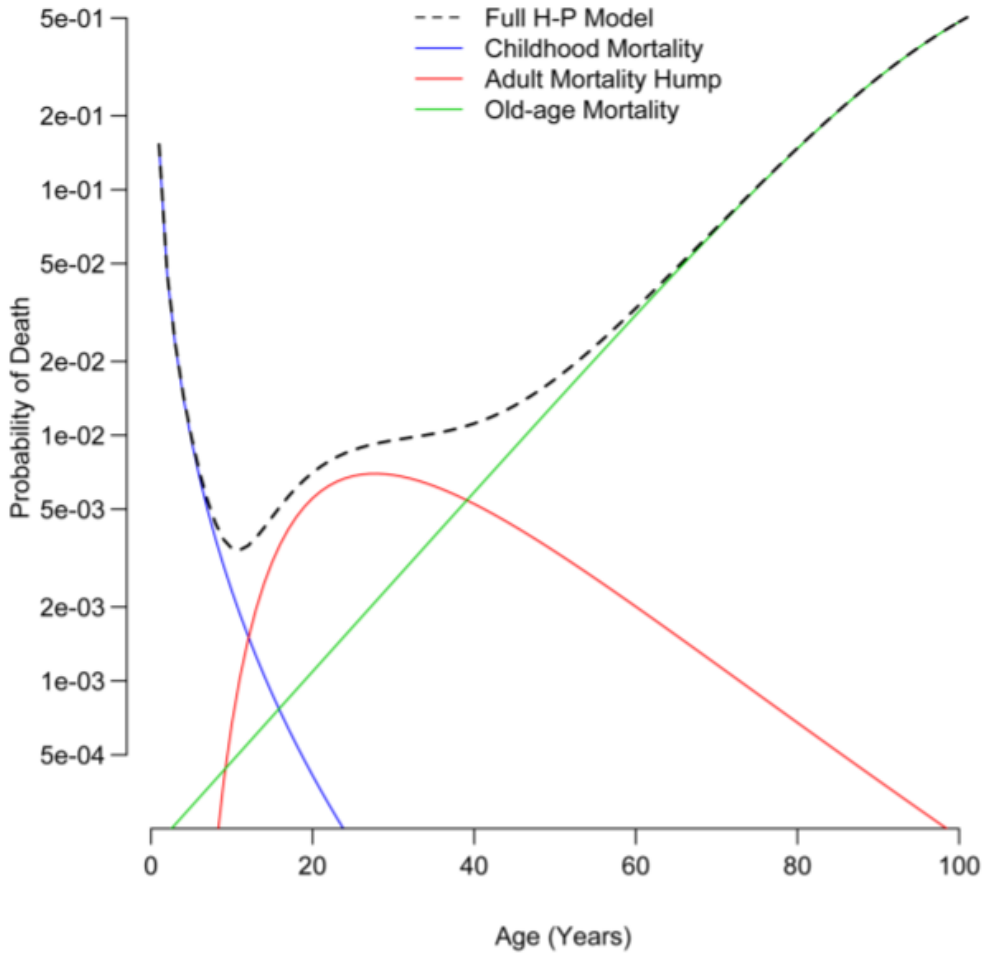
Mortality Models

Among them, the Heligman-Pollard model (Heligman and Pollard, 1980) is probably the most popular:

$$\frac{q_x}{1-q_x} = A^{(x+B)^C} + De^{-E(lnx-lnF)^2} + GH^x,$$

- 8 interpretable parameters
- suitable for the whole age range
- some variants also proposed in Heligman and Pollard (1980)
- very high correlation among parameters during estimation, adding difficulties
- the high correlation also compromises the interpretability

Mortality Models



Decomposition of the HP model [2]

[2] Source: Sharrow et al. (2013)

Mortality Models

Alternative to parametric mortality laws, several model life table systems have also been developed and widely used, such as the Coale-Demeny family (Coale and Demeny, 1966). Some organisations have also produced their own life table system, e.g. UN, GBD.

These systems often use a few summary indices such as the child mortality ${}_5q_0$, life expectancy at birth $\mathring{e}_0$, adult mortality ${}_{45}q_{15}$ etc. to generate full mortality schedules.

Mortality Models

A more recent model life table system is developed by Wilmoth et al. (2012) which has 2 parameters, namely the Log Quadratic (LogQuad) model (still somehow parameteric I guess...):

$$\log(m_x) = a_x + b_x\, h + c_x\, h^2 + v_x\, k,$$

where a_x, b_x, c_x and v_x are constants derived from a set of life tables using SVD and h and k are the 2 effective parameters to vary. In addition, $h = \log({}_5q_0)$, therefore given a fixed level of child mortality, a full schedule can be generated.

Life Tables

ELT15 (Males)

x	l_x	d_x	q_x	μ_x	$\mathring{e}_x$
0	100 000	814	0.008 14		73.413
1	99 186	62	0.000 62	0.000 80	73.019
2	99 124	38	0.000 38	0.000 43	72.064
3	99 086	30	0.000 30	0.000 33	71.091
4	99 056	24	0.000 24	0.000 27	70.113
5	99 032	22	0.000 22	0.000 23	69.130
6	99 010	20	0.000 20	0.000 21	68.145
7	98 990	18	0.000 19	0.000 19	67.158
8	98 972	19	0.000 18	0.000 18	66.171
9	98 953	18	0.000 18	0.000 18	65.183
10	98 935	18	0.000 18	0.000 18	64.195
11	98 917	18	0.000 18	0.000 18	63.206
12	98 899	19	0.000 19	0.000 19	62.218
13	98 880	23	0.000 23	0.000 21	61.230
14	98 857	29	0.000 29	0.000 26	60.244

English Life Table 15 males [3]

x	ℓ_x	q_x	m_x	x	ℓ_x	q_x
0	100,000	0.004746	0.004766	58	92,261	0.006687
1	99,525	0.000306	0.000306	59	91,644	0.007299
2	99,495	0.000207	0.000207	60	90,975	0.008016
3	99,474	0.000147	0.000147	61	90,246	0.008773
4	99,460	0.000115	0.000115	62	89,454	0.009492
5	99,448	0.000099	0.000099	63	88,605	0.010259
6	99,438	0.000091	0.000091	64	87,696	0.011227
7	99,429	0.000088	0.000088	65	86,712	0.012404

English Life Table 17 males [4]

- information included may vary from tables to tables
- columns can be derived from each other

[3] Source: Formulae and Tables for The Institute and Faculty of Actuaries

[4] Source: Dodd et al. (2018)

Bayesian Population Reconstruction

Two main approaches:

- prospective reconstruction ("inverse projection")
  - ↪ start with a "baseline population" at the initial time period and project and reconstruct the population forward with corresponding demographics rates
- retrospective reconstruction ("back projection")
  - ↪ start with a final population at the terminal time period and propagate backwards in time
  - ↪ however, there was debates about the efficacy of this approach as it is *'trying to "resurrect" members of the open ended age group and simultaneously estimate fertility, mortality and migration rates.'* (Wheldon et al., 2016)

Bayesian Population Reconstruction

I will be adopting the Bayesian population reconstruction framework by Wheldon et al. (2016)

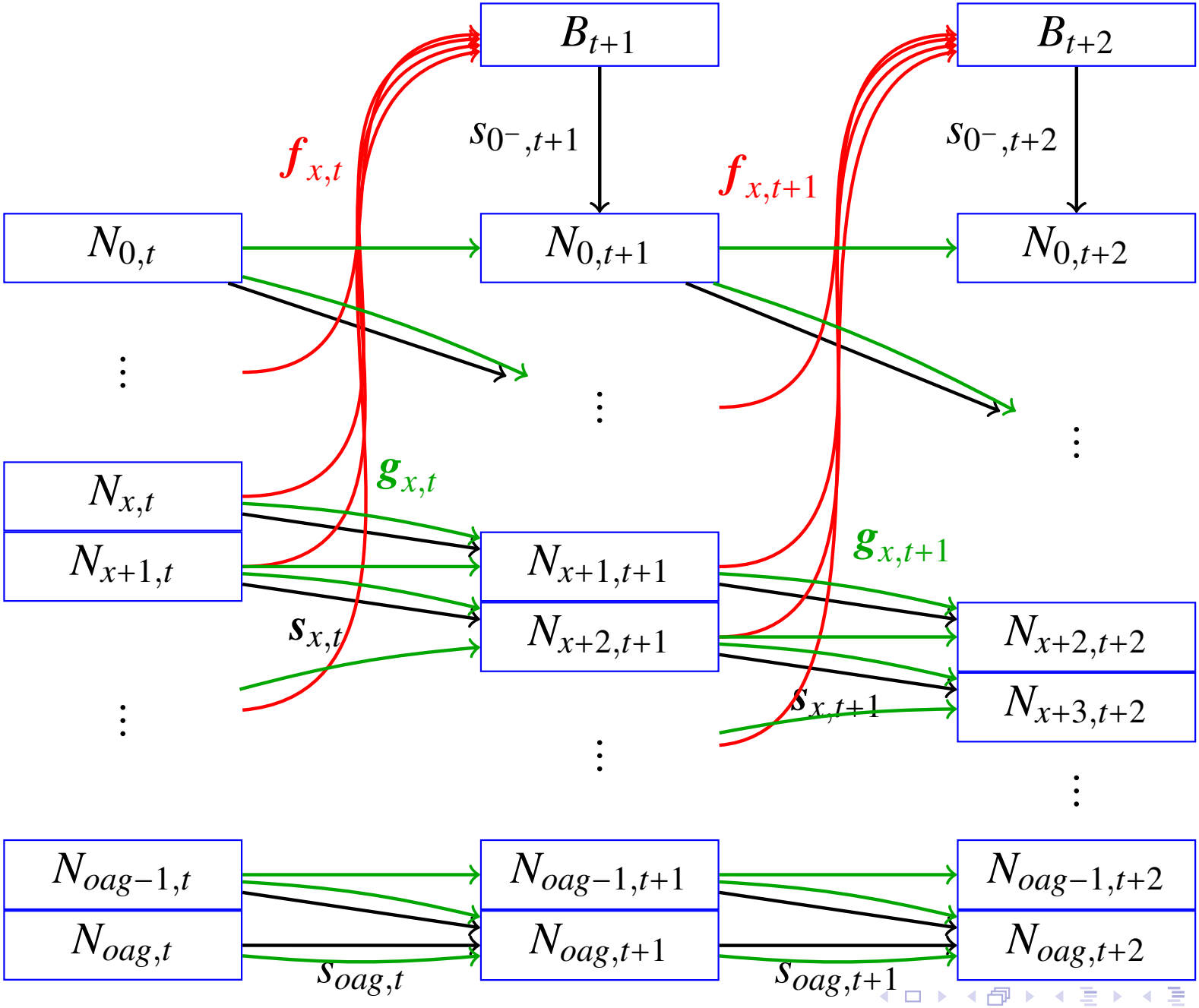
- prospective approach
- Bayesian approach, gives associated uncertainty ranges of the reconstructed population counts

The core of this framework is the cohort component method of population projection (CCMPP) which utilises the concept of the demographic balancing equation

$$\underbrace{N_{t+1}}_{\text{population at time } t+1} = \underbrace{N_t}_{\text{population at time } t} - \underbrace{D_t}_{\text{deaths in period } t} + \underbrace{B_t}_{\text{births in period } t} + \underbrace{G_t}_{\text{net migration in period } t}$$

- continuous process, need to discretise time steps
- population projected along the cohort direction

CCMPP


B_{t+1}
B_{t+2}
$s_{0^-,t+1}$
$s_{0^-,t+2}$
$f_{x,t}$
$f_{x,t+1}$
$N_{0,t}$
$N_{0,t+1}$
$N_{0,t+2}$
$N_{x,t}$
$N_{x+1,t}$
$g_{x,t}$
$N_{x+1,t+1}$
$N_{x+2,t+1}$
$s_{x,t}$
$g_{x,t+1}$
$N_{x+2,t+2}$
$N_{x+3,t+2}$
$s_{x,t+1}$
$N_{oag-1,t}$
$N_{oag,t}$
$N_{oag-1,t+1}$
$N_{oag,t+1}$
$N_{oag-1,t+2}$
$N_{oag,t+2}$
$s_{oag,t}$
$s_{oag,t+1}$

Bayesian Population Reconstruction

Require demographics rates (fertility rates, survival proportions, net migration proportions) to project the populations

$$
\begin{aligned}
N^*_{x,t_0}|\sigma^2_N &\sim logNormal(\log(N_{x,t_0}), \sigma^2_N) \\
f^*_{x,t}|\sigma^2_f &\sim logNormal(\log(f_{x,t}), \sigma^2_f) \quad \forall x \in [15, 50) \\
s^*_{x,t}|\sigma^2_s &\sim logNormal(\log(s_{x,t}), \sigma^2_s) \\
g^*_{x,t}|\sigma^2_g &\sim AR1 : AR1(\rho_x, \rho_t, \sigma^2_g) \\
N^*_{x,t}|f^*_{x,t}, s^*_{x,t}, g^*_{x,t} &= CCMPP(N^*_{x,t_0}, f^*_{x,t}, s^*_{x,t}, g^*_{x,t}, srb) \quad \forall t \neq t_0 \\
N_{x,t} &\sim logNormal(\log(N^*_{x,t}), \sigma^2_N) \quad \forall t \neq t_0 \\
\sigma^2_N, \sigma^2_f, \sigma^2_s, \sigma^2_g, \rho_x, \rho_t &\propto h(\cdot)
\end{aligned}
$$

Bayesian Population Reconstruction

Instead of using estimates of the survival proportions from other data source as priors, incorporate DHS data into population reconstruction

- DHS data supports estimation of mortality rates m_x, and most models are on m_x instead of survival proportions s_x
  - ↪ convert m_x to s_x
- DHS data limited age range (15 - 60), where as population reconstruction spans the full age range (0 - open age group)
  - ↪ utilise laws of mortality / model life tables, e.g. LogQuad, HP-8, Thiele

Conversion between m_x and s_x

s_x, the survivorship between age groups, are then calculated by $\frac{L_{x+1}}{L_x}$, where L_x denotes the exposures in the period at age $[x, x+1)$. Under UDD, $L_x = l_x - \frac{1}{2}d_x$,

$$s_x \overset{\text{UDD}}{=} {}_{0.5}p_{x+0.5}\ {}_{0.5}p_{x+1}$$

$$= {}_{1}p_{x+0.5}$$

Open Age Group Mortality Rate

Aggregating mortality rates of the last age groups into an open age group also needs attention.

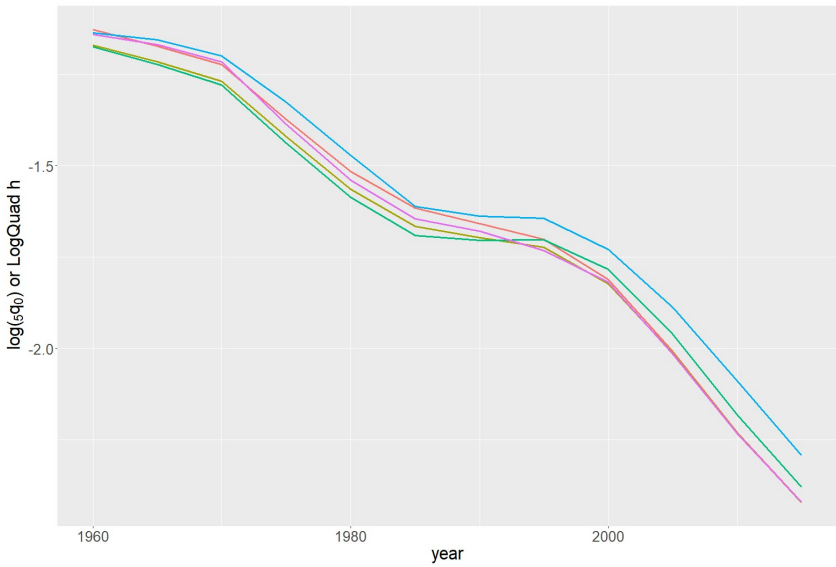
$$m_{oag} = \frac{\sum_{i \geq oag} w_i m_i}{\sum_{i \geq oag} w_i}$$

where, under UDD,

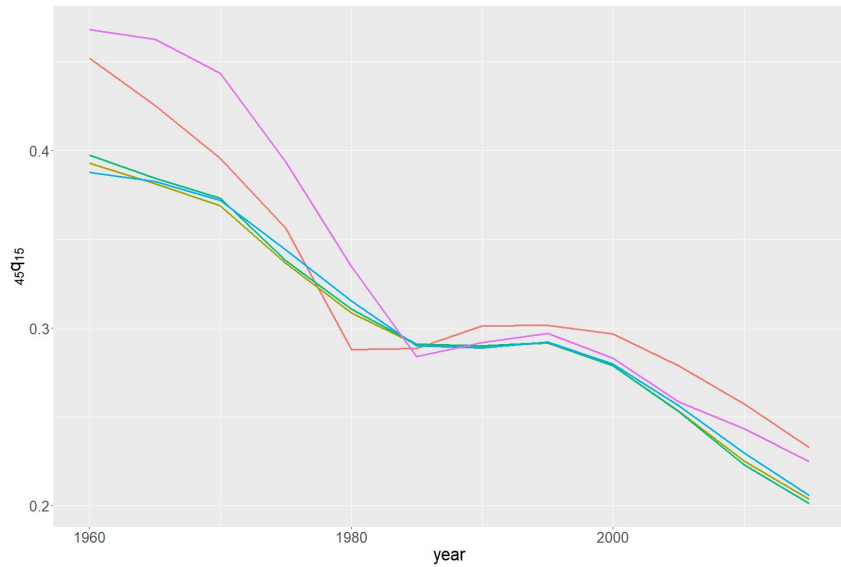
$$w_i = (1 - 0.5\, q_i) \prod_{oag \leq j < i} p_j$$

LogQuad Model - Females

Using IGME child moratlity estimates as priors for the LogQuad parameters

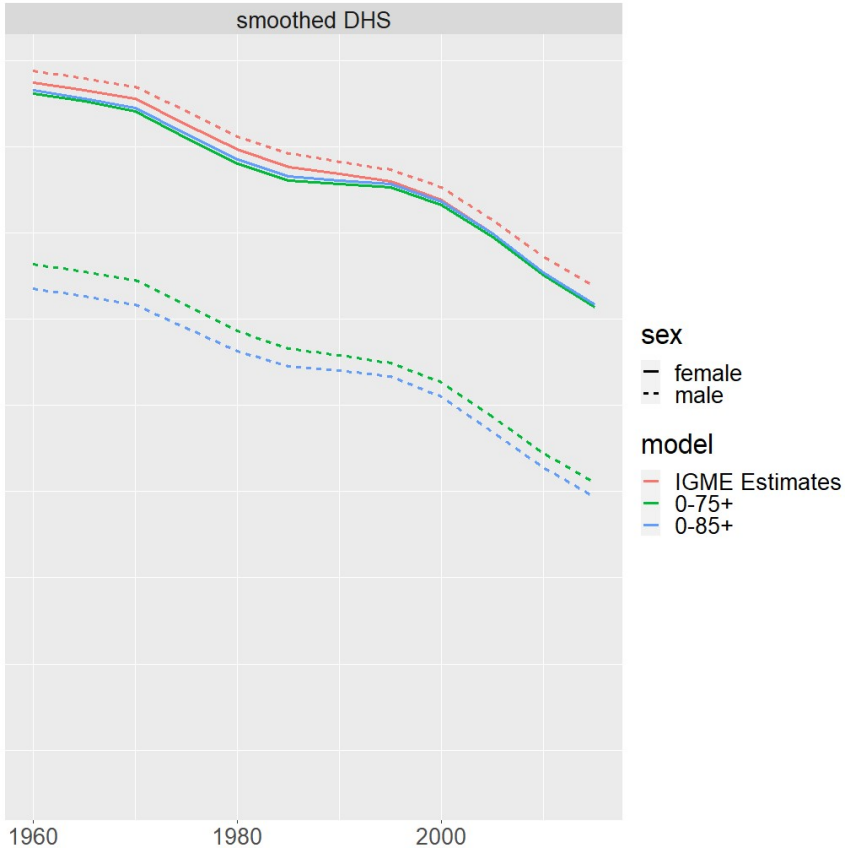


Estimated log child mortality.
IGME estimates in red.



Estimated ${}_{45}q_{15}$.WPP estimates in red.

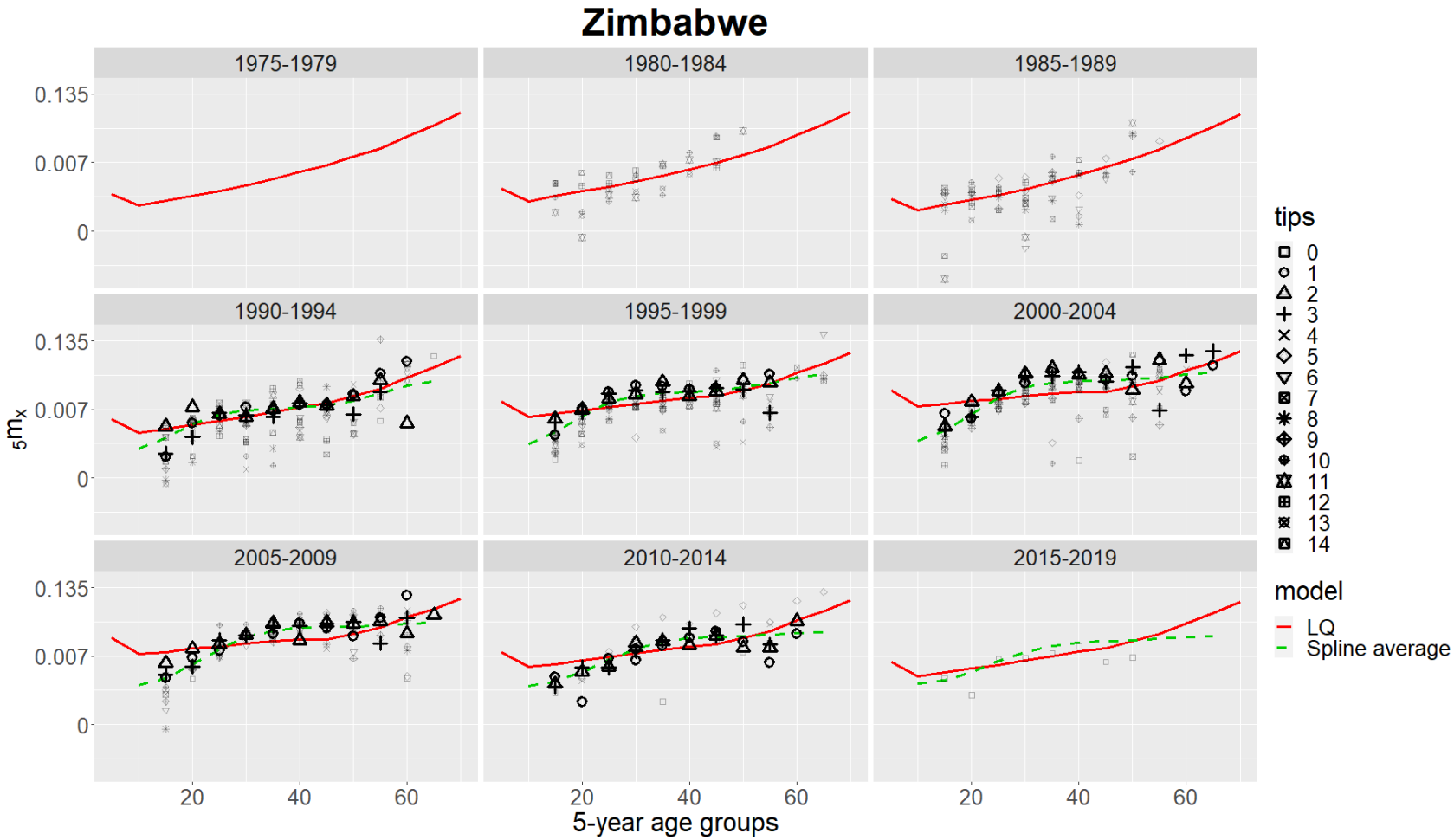
LogQuad Model - Joint Sex



Estimated log child mortality. IGME estimates in red.

- Estimated male child mortality is much lower than the IGME estimates
- LogQuad developed for countries without HIV
- HIV-hump in mortality in adult ages (age range with DHS data) affects estimated child mortality due to the in-flexibility of the LogQuad model for HIV-infected populations

LogQuad Model - Zimbabwe



LQ obviously not meant for naive implementation in high-HIV countries, but this illustrates the LQ fit in this context

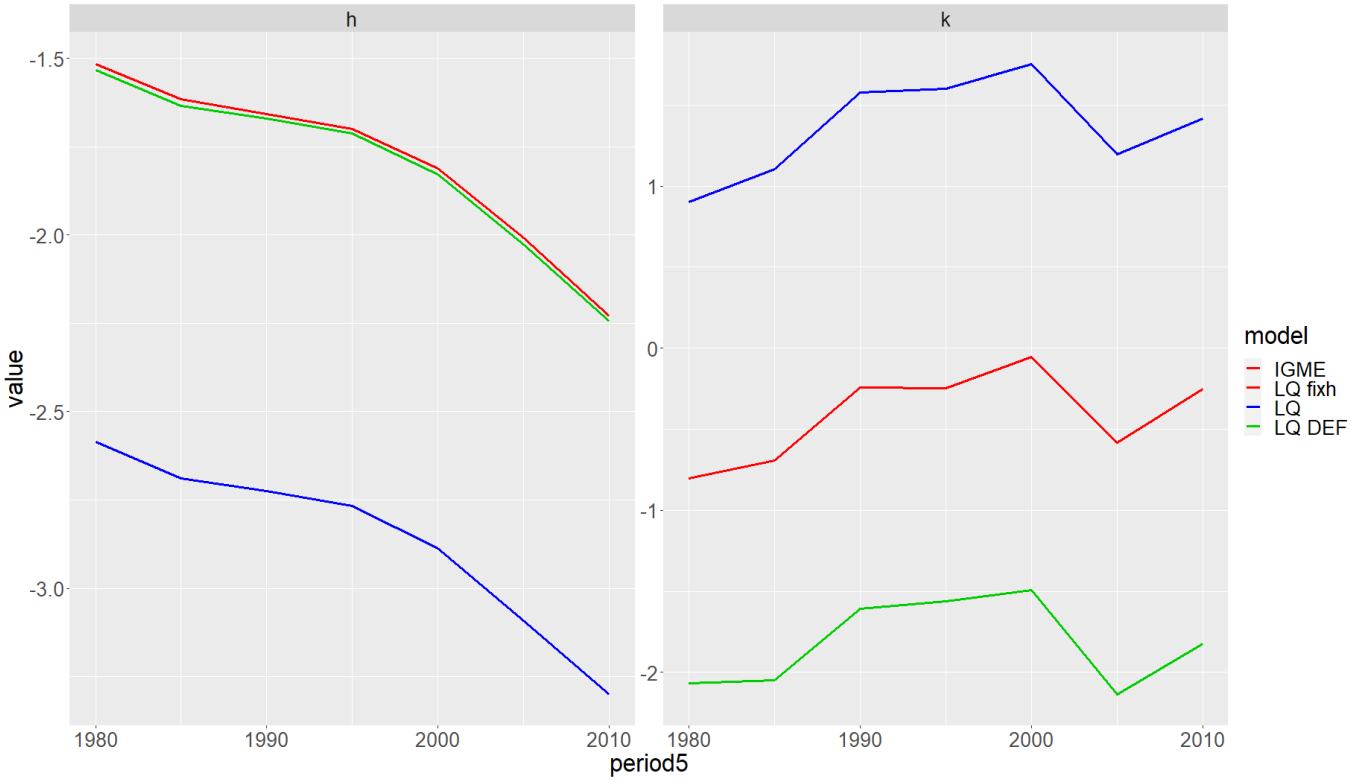
LogQuad Model + DEF Hump from the HP-8 Model

Recall the HP-8 Model decomposes the age pattern into 3 stages

$$\frac{q_x}{1-q_x} = A^{(x+B)^C} + De^{-E(lnx-lnF)^2} + GH^x,$$

Attempted to extract the middle hump and add it on top of the LQ model and fit only to the DHS data, still using IGME estimates as priors.

LogQuad Model + DEF Hump from the HP-8



- LQ+DEF allows the estimated child mortality to be more consistent for males (green and red)
- difficult to converge for females and gives insensible results, possibly due to the less prominent hump in Burkina Faso as well as the corerlation between the DEF and the LQ k

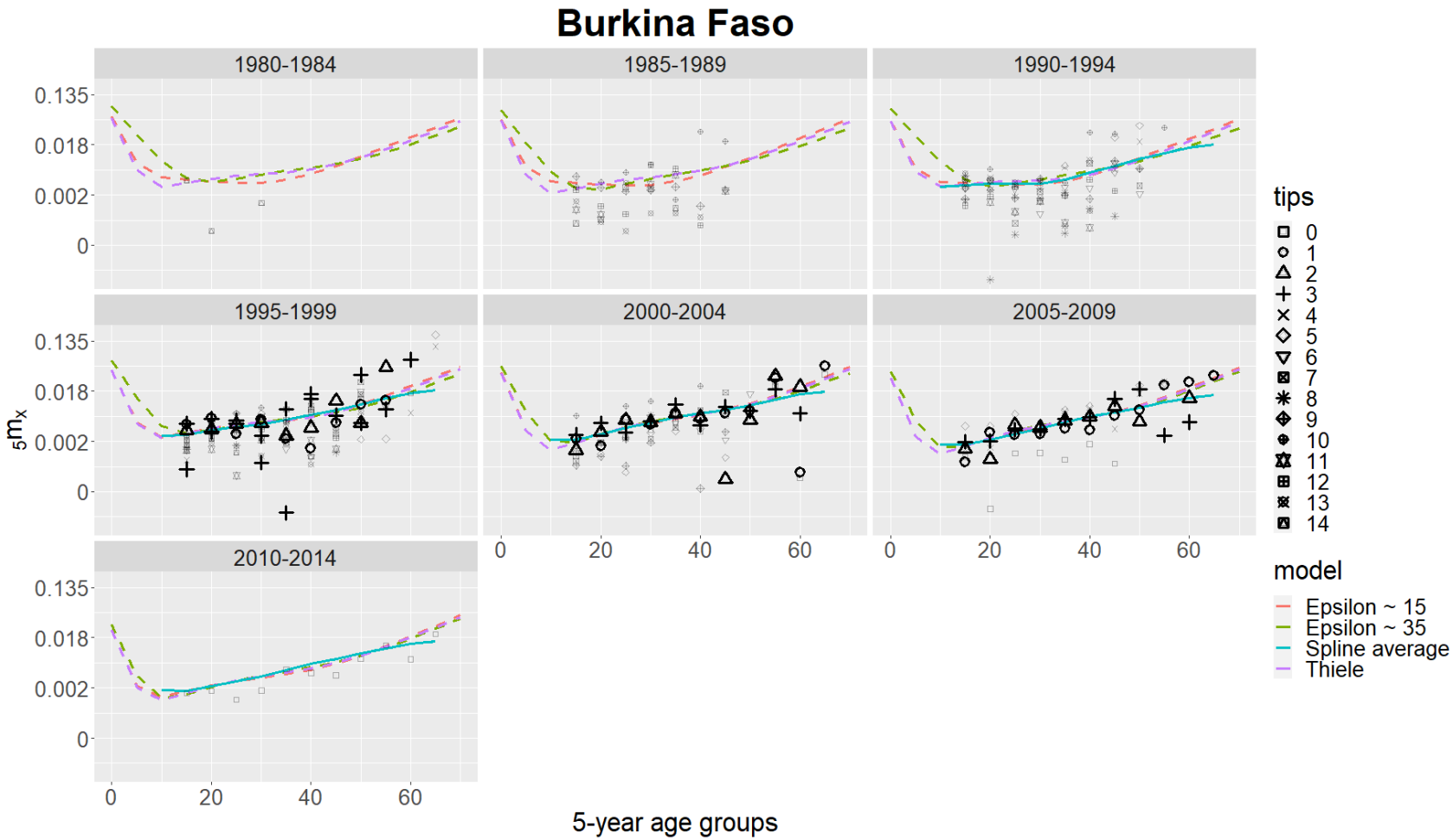
Thiele

$$m(x) = \underbrace{\varphi e^{-\psi x}}_{\text{negative exponential}} + \underbrace{\lambda e^{-\delta(x-\epsilon)^2}}_{\text{normal}} + \underbrace{A e^{Bx}}_{\text{Gompertz}} \qquad \varphi, \psi, \lambda, \delta, \epsilon > 0$$

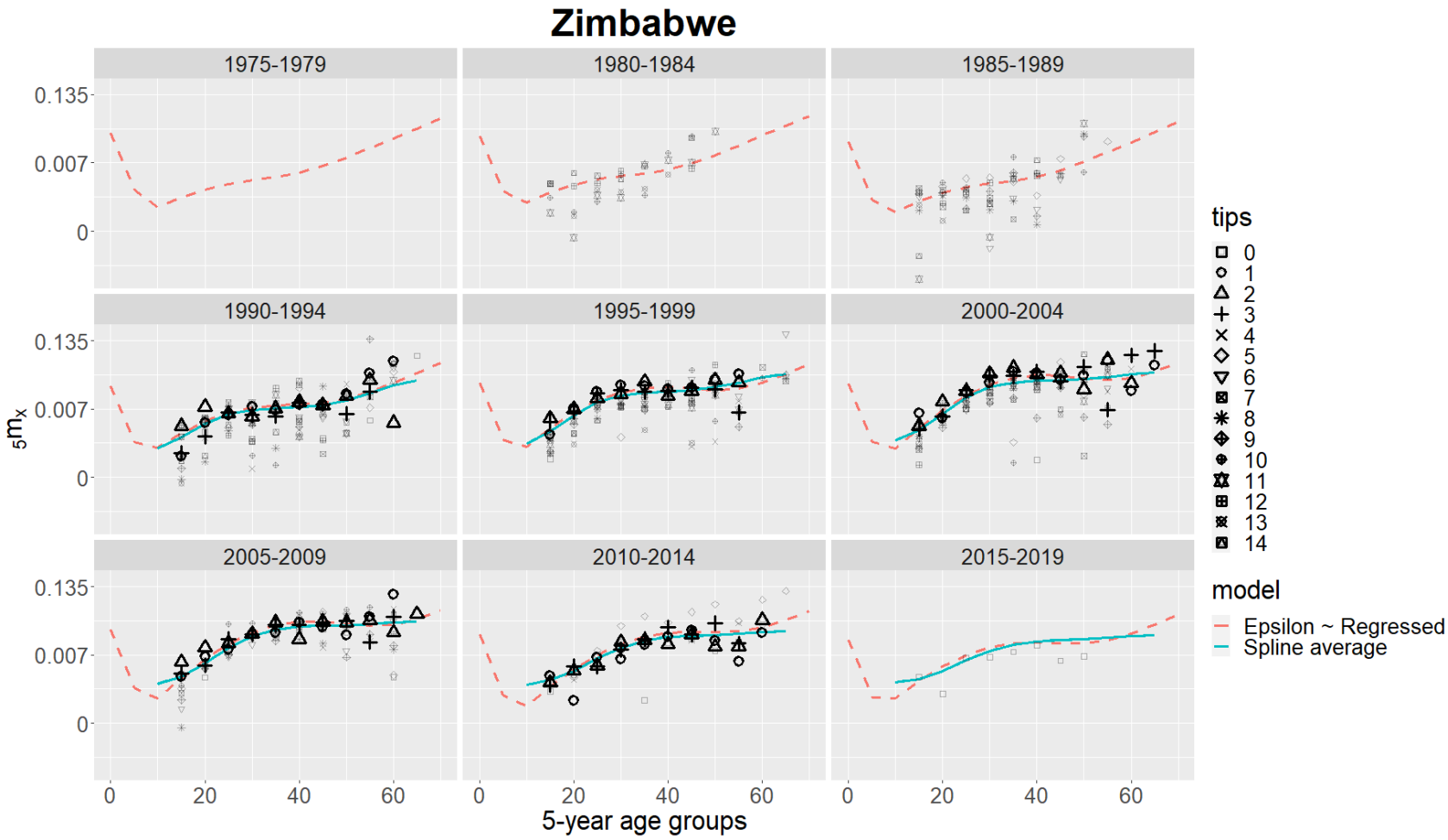
Used IGME estimates and LogQuad model to derive priors for the 7 parameters of Thiele

- consistent child mortality estimates with the IGME estimates
- sensible results
- converged quite stably for different countries (so far)

Thiele

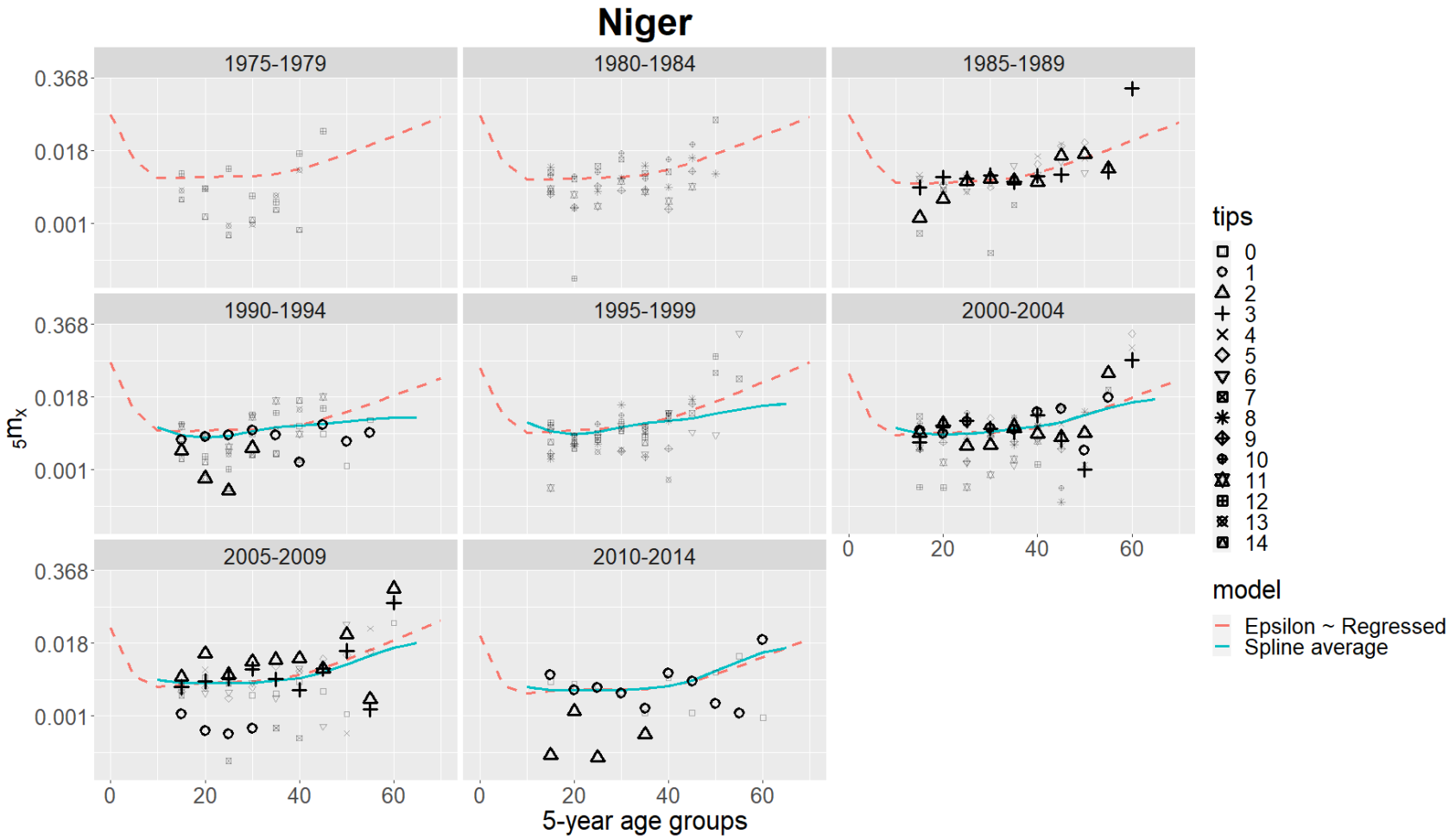

Burkina Faso
1980-1984
1985-1989
1990-1994
1995-1999
2000-2004
2005-2009
2010-2014
0.135
0.018
0.002
0
0
20
40
60
$_5m_x$
5-year age groups
tips
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
model
Epsilon ~ 15
Epsilon ~ 35
Spline average
Thiele

Thiele



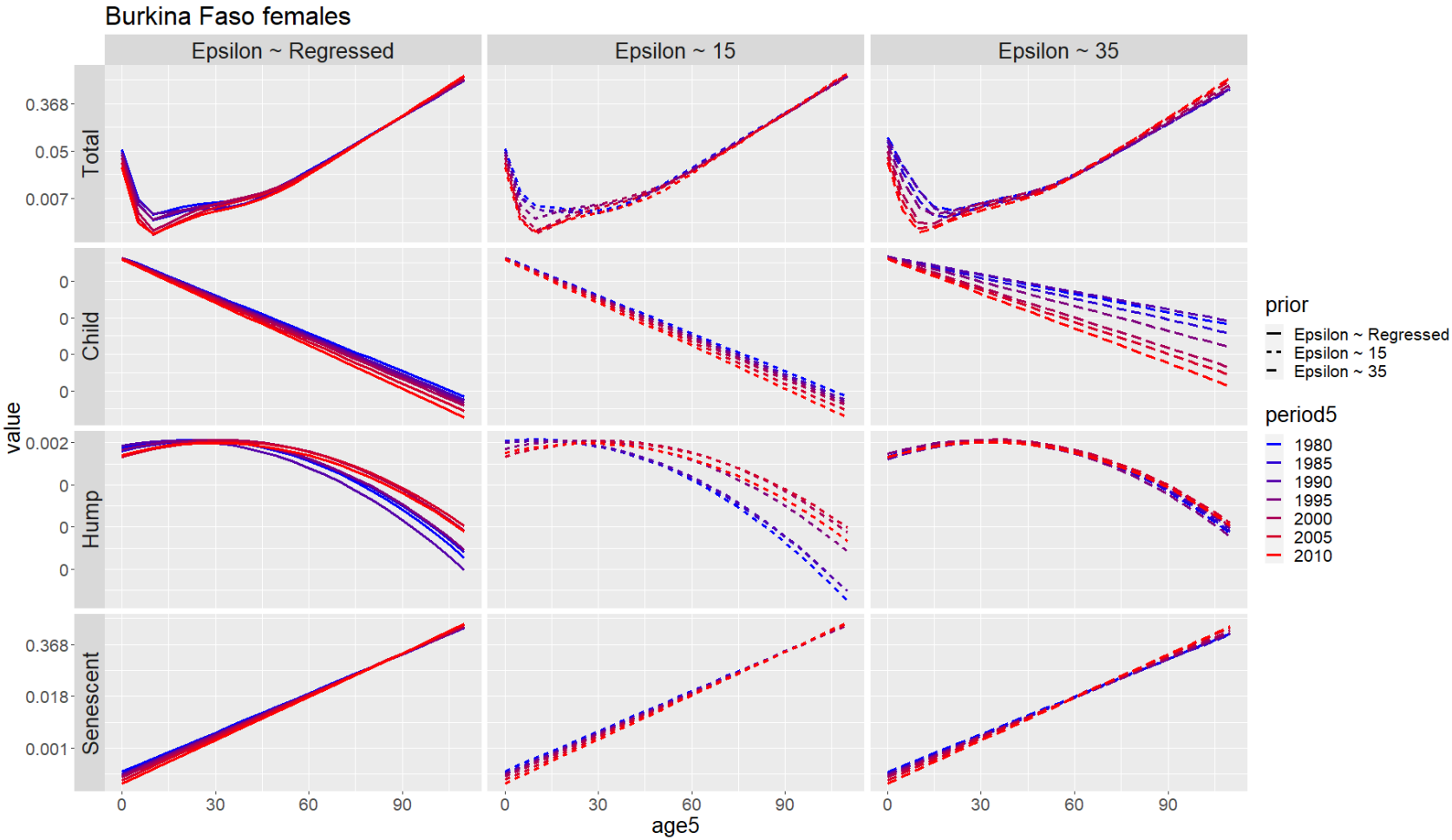
Adequate fit even to high HIV countries

Thiele



Also in very low HIV countries

Thiele



Thiele estimated m_x and decomposed

Next Steps

- So far, because of the issues encountered with the LogQuad model, took a step back and modelled only DHS data
- Incorporate Thiele model into population reconstruction
- Try to modify the hump component of the model such that it is on log-scale and compare fit
- Extend to other countries

Beard, R. E. (1959). Note on some mathematical mortality models. In *Ciba Foundation Symposium-The Lifespan of Animals (Colloquia on Ageing)*, volume 5, pages 302–311. Wiley Online Library.

Carriere, J. F. (1992). Parametric models for life tables. *Transactions of the Society of Actuaries*, 44:77–99.

Coale, A. and Demeny, P. (1966). Regional model life tables and stable populations princeton university press. *Prínceton, New Jersey*.

Coale, A. J. and Kisker, E. E. (1990). Defects in data on old-age mortality in the united states: new procedures for calculating mortality schedules and life tables at the highest ages.

Dodd, E., Forster, J. J., Bijak, J., and Smith, P. W. (2018). Smoothing mortality data: the english life tables, 2010–2012. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 181(3):717–735.

Gompertz, B. (1825). On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. in a letter to francis baily, esq. frs &c. *Philosophical transactions of the Royal Society of London*, (115):513–583.

Heligman, L. and Pollard, J. H. (1980). The age pattern of mortality. *Journal of the Institute of Actuaries*, 107(01):49–80.

Lindbergson, M. (2001). Mortality among the elderly in sweden 1988–1997. *Scandinavian Actuarial Journal*, 2001(1):79–94.

Makeham, W. M. (1860). On the law of mortality and the construction of annuity tables. *Journal of the Institute of Actuaries*, 8(6):301–310.

Perks, W. (1932). On some experiments in the graduation of mortality statistics. *Journal of the Institute of Actuaries*, 63(1):12–57.

Sharrow, D. J., Clark, S. J., Collinson, M. A., Kahn, K., and Tollman, S. M. (2013). The age pattern of increases in mortality affected by

hiv: Bayesian fit of the heligman-pollard model to data from the agincourt hdss field site in rural northeast south africa. *Demographic research*, 29:1039.

Siler, W. (1983). Parameters of mortality in human populations with widely varying life spans. *Statistics in medicine*, 2(3):373–380.

Thatcher, A. R. (1999). The long-term pattern of adult mortality and the highest attained age. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 162(1):5–43.

Thiele, T. N. (1871). On a mathematical formula to express the rate of mortality throughout the whole of life, tested by a series of observations made use of by the danish life insurance company of 1871. *Journal of the Institute of Actuaries*, 16(5):313–329.

Wheldon, M. C., Raftery, A. E., Clark, S. J., and Gerland, P. (2016). Bayesian population reconstruction of female populations for less developed and more developed countries. *Population studies*, 70(1):21–37.

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