# Modelling and forecasting healthy life expectancy. A Compositional Data Analysis approach

Marie–Pier Bergeron–Boucher[1], Cosmo Strozza[1], Violetta Simonacci[2] and Jim Oeppen[1]

[1]Interdisciplinary Centre on Population Dynamics, University of Southern Denmark.
[2]University of Naples Federico II

EPC - Groningen, July 2 2022

# Introduction

- Knowing the quality of the extra years of life has implications for individuals and society.
- Clear rationale for forecasting healthy life expectancy.
- Very few available models to forecast healthy life expectancy!

# Objective

We aim to develop methods which can simultaneously forecast mortality and health prevalence.

We developed two new models, based on the two most common approach to calculate healthy life expectancy: Sullivan and multi-state life table.

# Data

1. Data on mortality: Human Mortality Database (HMD);
2. Data on health: Survey of Statistics on Income and Living Conditions (SILC);
3. Health measure: Activity limitations;
4. Populations: Spain (and Sweden);
5. Years: 2004 to 2019;
6. Ages: 50 to 80+;
7. Sex: Females (and males).

# Method 1: Sullivan's method

# Method

The Sullivan method estimates the number of years lived in a given health status as:

$$L_{tx}(s) = L_{tx} * \pi_{tx}(s) \tag{1}$$

- $L_{tx}$ is the person-years lived at time $t$ and age-interval $x : x + 1$;
- $\pi_{tx}(s)$ are the proportion of individuals with the health status $s$ at time $t$ and age $x$.

# Method

Rewriting the equation....

$$
\begin{aligned}
L_{tx}(s) &= \pi_{tx}(s)[l_{tx} - a_{tx}d_{tx}] \\
&= l_{tx}(s) - a_{tx}d_{tx}(s)
\end{aligned} \tag{2}
$$

- $d_{tx}$ is the life table deaths at age $x$ and time $t$;
- $l_{tx}$ is the survival probability to age $x$ and time $t$;
- $a_{tx}$ is the average number of person-years lived in the age-interval by those dying in the interval.

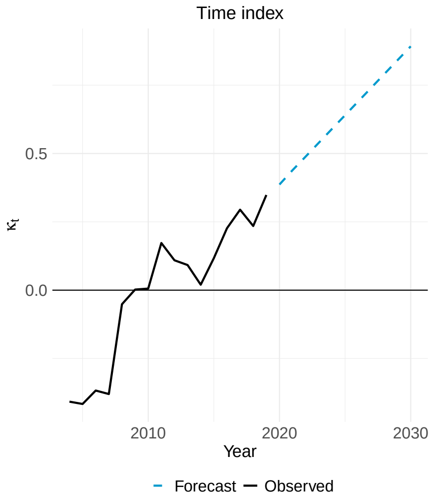
# Method

$d_{tx}(s)$ are compositional data, i.e. relative information constrained to sum to a constant. We can use the model of Oeppen (2008) to forecast $d_{tx}(s)$.

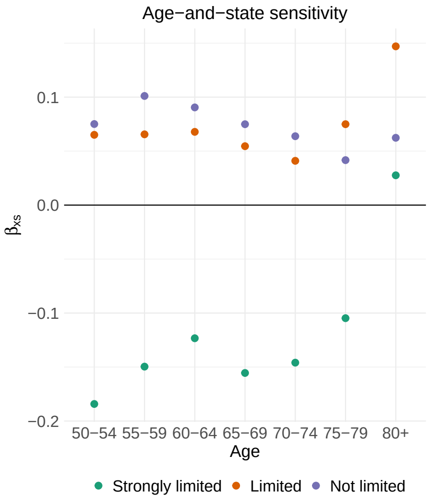
$$clr(d_{t,x*s} \ominus \alpha_{x*s}) = \kappa_t \beta_{x*s} + \epsilon_{t,x*s} \tag{3}$$

- $d_{t,x*s}$ is a matrix of life table deaths by time $t$ as rows and age-and-status $x*s$ as columns;
- $\alpha_{x*s}$ is the age-and-status-specific geometric mean;
- $\kappa_t$ and $\beta_{x*s}$ are the dominant components of a singular value decomposition.

# Parameters

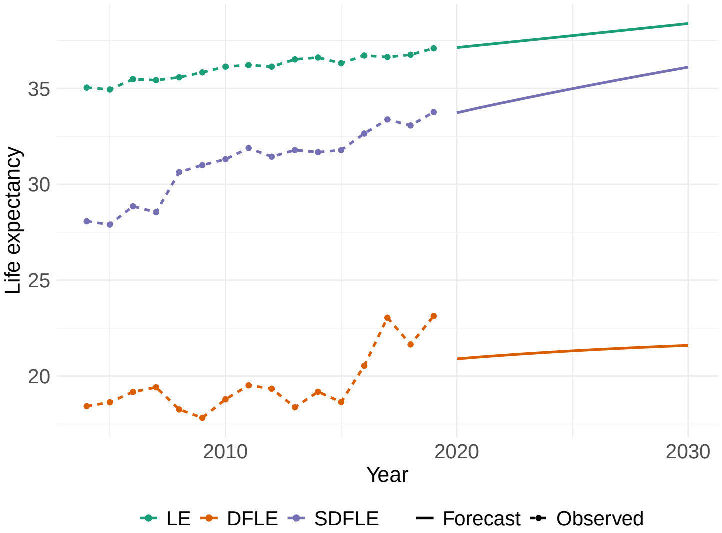


Time index
0.5
0.0
$\kappa_t$
2010
2020
2030
Year
Forecast
Observed

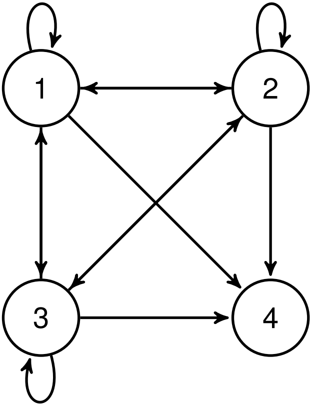
Age−and−state sensitivity
0.1
0.0
−0.1
−0.2
$\beta_{xs}$
50−54
55−59
60−64
65−69
70−74
75−79
80+
Age
Strongly limited
Limited
Not limited

# Forecast


Life expectancy
35
30
25
20
2010
2020
2030
Year
LE DFLE SDFLE
Forecast Observed

# Method 2: Multistate life table model

# Method

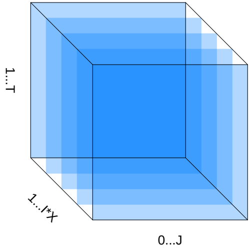



Transition matrix: $\begin{pmatrix} q_{11} & q_{12} & q_{13} & q_{14} \\ q_{21} & q_{22} & q_{23} & q_{24} \\ q_{31} & q_{32} & q_{33} & q_{34} \end{pmatrix}$

$$q_{ij} = \frac{D_{ij}}{N_i} \qquad (4)$$

# Method

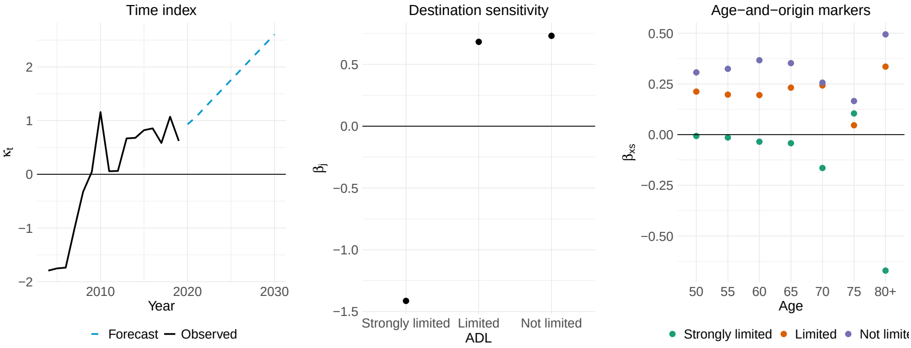
$q_{ij}$ are compositional data. We can use the 3D model of Bergeron-Boucher et al. (2018) to forecast $q_{t,j,i*x}$.




$$clr(q_{t,j,i*x} \ominus \alpha_{j,i*x}) = \kappa_t \beta_j \gamma_{i*x} + \epsilon_{t,j,i*x} \quad (5)$$

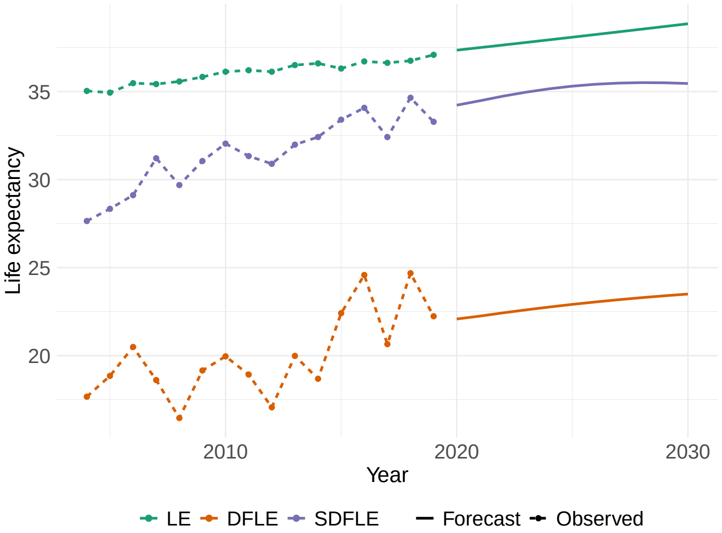
- $q_{t,j,i*x}$ is an array of transition probability by time $t$, destination-state $j$, age-and-origin $x * i$;
- $\alpha_{j,i*x}$ is the destination-specific geometric mean for each age and origin;
- $\kappa_t$, $\beta_j$ and $\gamma_{i*x}$ are the dominant components of a parafac model.

# Parameters



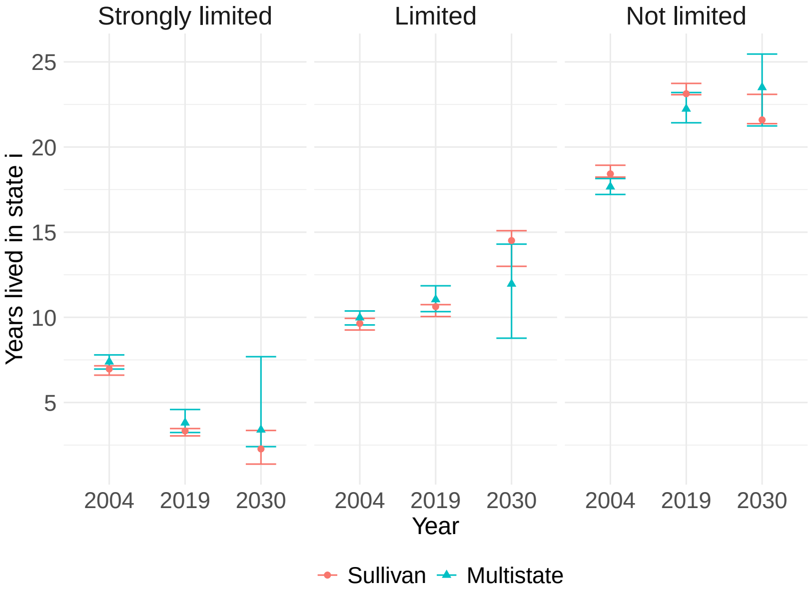
Time index
2
1
0
−1
−2
$\kappa_t$
2010
2020
2030
Year
Forecast
Observed
Destination sensitivity
0.5
0.0
−0.5
−1.0
−1.5
$\beta_j$
Strongly limited
Limited
Not limited
ADL
Age−and−origin markers
0.50
0.25
0.00
−0.25
−0.50
$\beta_{xs}$
50
55
60
65
70
75
80+
Age
Strongly limited
Limited
Not limite

# Forecast



Life expectancy
35
30
25
20
2010
2020
2030
Year
LE
DFLE
SDFLE
Forecast
Observed

# Comparison


Strongly limited
Limited
Not limited
Years lived in state i
25
20
15
10
5
2004
2019
2030
Year
Sullivan
Multistate

# Discussion

- First step!
- Correlation between components accounted for.
- Coherent forecasts of mortality and health status.
- Short time series.
- Out-of-sample evaluation over short forecast horizon only.
- Data quality?
- Covid-19 impact?
- Comparable estimates and forecasts with both the Sullivan and Multistate models.

Thank you!

mpbergeron@sdu.dk

Research project funded by the SCOR Foundation for Science.