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

Paris, April 6 2022

Introduction

- There is a demand for forecasts for some components of mortality.
- Knowing the quality of the extra years of life has implications for individuals and society.
- Most efforts to improve forecasts are directed towards life expectancy.
- Only few available methods to forecast healthy life expectancy:
  - Based on scenarios;
  - Separate forecasts of transition rates and incidence rates.

Objective

Develop methods which can simultaneously forecast mortality and health prevalence.

Healthy life expectancy

Three main theoretical frameworks:

- “Failure of success" or expansion of morbidity: individuals surviving to older ages with more health problems. (Gruenberg 1977)
- Compression of morbidity: individuals surviving to older ages with fewer health problems. (Fries 1980)
- Dynamic equilibrium: disability is redistributed towards less severe states resulting in more years of life with moderate disability and equal or fewer years with severe disabilities. (Manton 1982)

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

By rewriting the equation, we can express the number of years lived in a given health status as a function of the death distribution by age and health status ($d_{tx}(s)$):

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

- 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

By modelling and forecasting $d_{tx}(s)$, we can derive the total mortality (d_{tx} and l_{tx}) and health prevalence ($\pi_{tx}(s)$):

$$\pi_{tx}(s) = \frac{d_{tx}(s)}{d_{tx}} \tag{3a}$$

$$d_{tx} = \sum_{s=1}^{S} d_{tx}(s). \tag{3b}$$

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 $d_{tx}(s)$.

$$clr(d_{t,x*s} \ominus \alpha_{x*s}) = \kappa_t \beta_{x*s} + \epsilon_{t,x*s} \tag{4}$$

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

Data

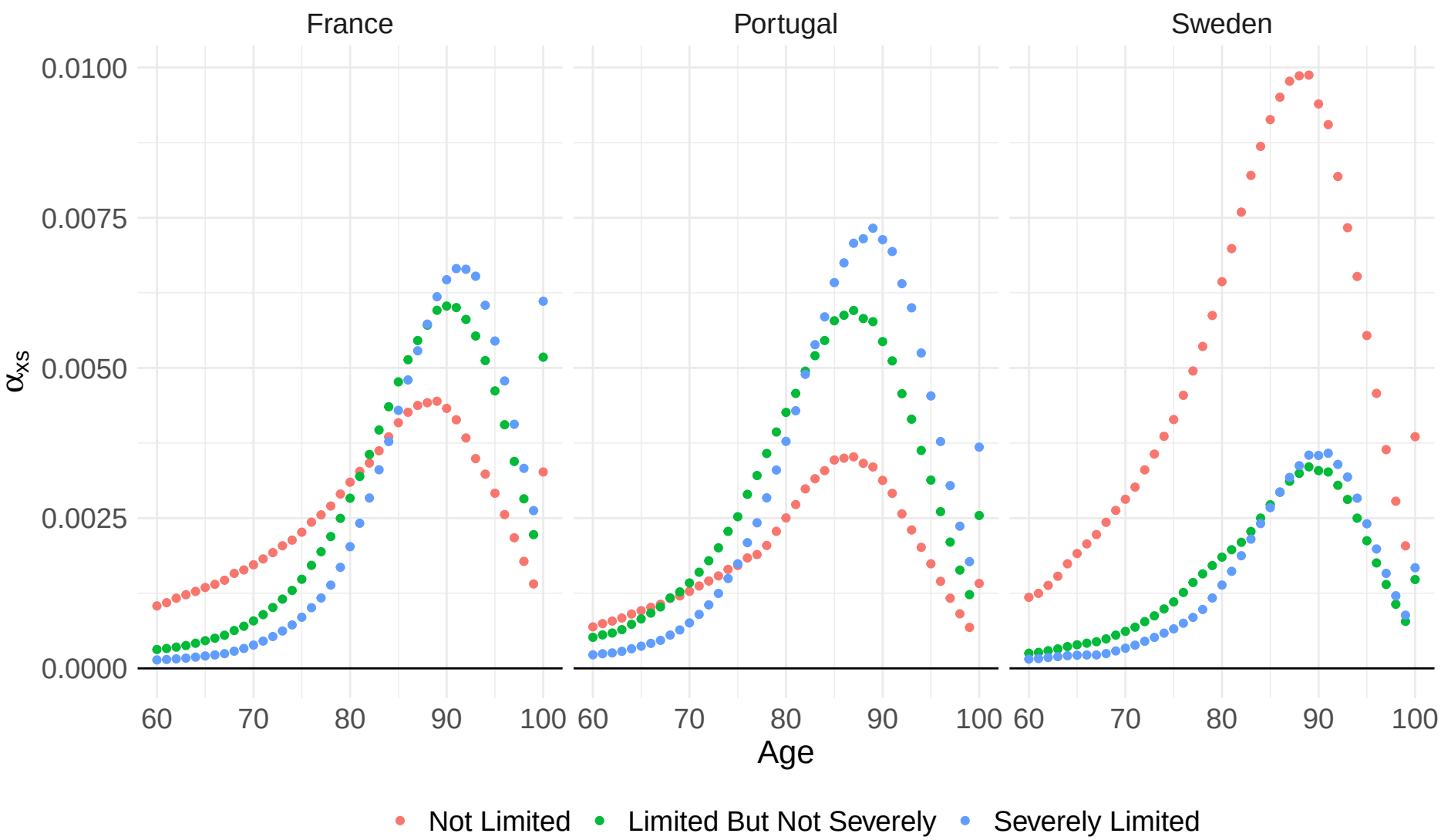
Data from EuroHex database based on the Statistics on Income and Living Conditions (SILC).

- Years 2004 to 2016;
- Age 60+ by 5-year age-group (linear interpolation to obtain 1-year age-groups);
- Health measured by activity limitations;
- Females in France, Portugal and Sweden.

Parameters: α_{xs}

α_{xs} is the mean age-and-health-status death distribution over time. As seen in the next slide, French and Portuguese females who die at younger ages do so without limitations. But the health status at the time of death transit towards severely limited as individuals reach older ages. On average, Swedish females mainly died as not-limited at all ages.

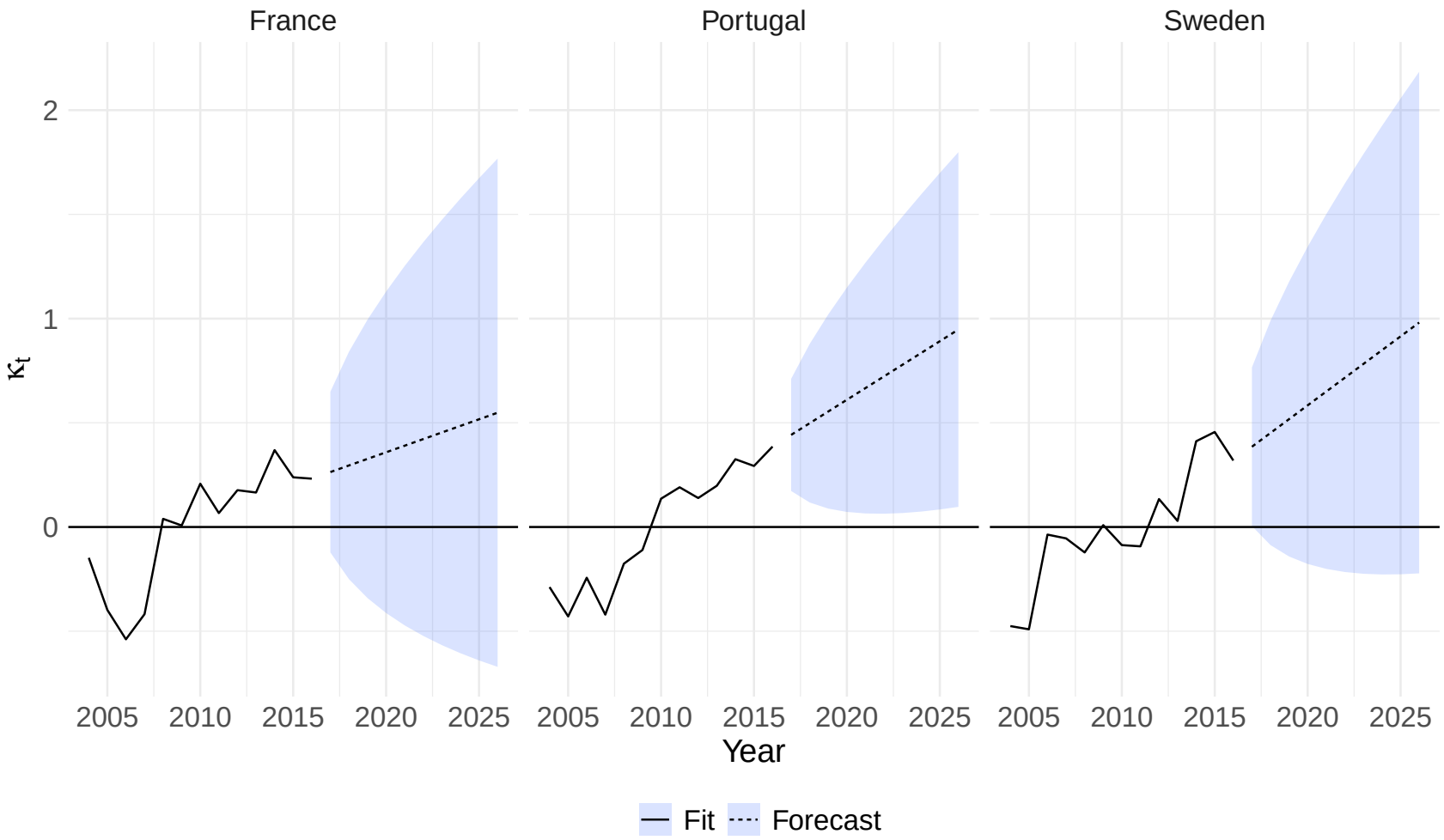
Parameters: α_{xs}



Parameters: κ_t

κ_t is an index of the general level of mortality over time. κ_t was linear for all selected countries and was forecast with a random-walk with drift.

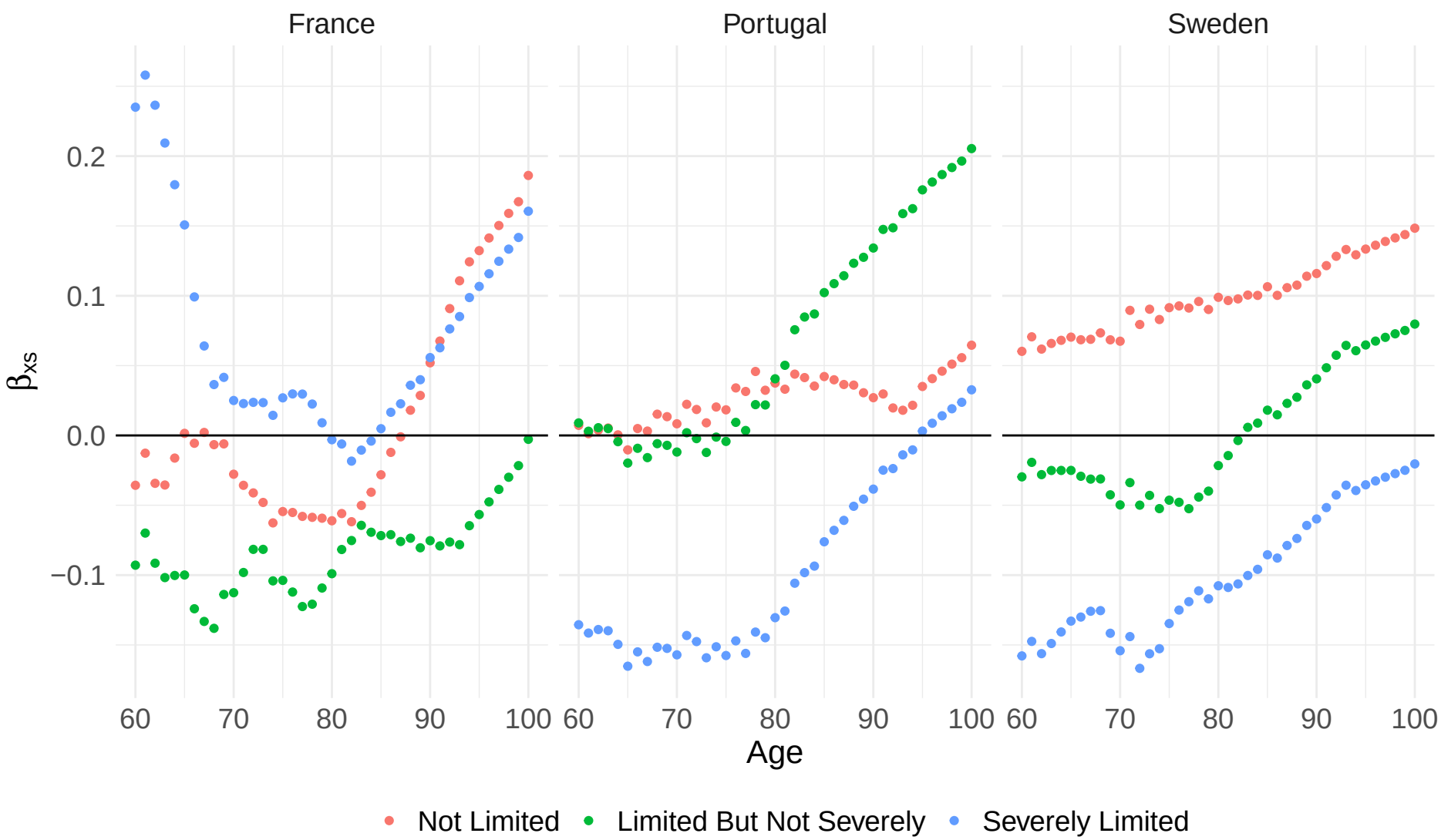
Parameters: κ_t



Parameters: β_{xs}

β_{xs} is an age-and-status-pattern indicating how deaths shifted across ages and health-status over time, in relative terms. When κ_t increases over time, there is a shift from negative towards positive β_{xs}. In Sweden, deaths have been shifted towards not-limited at all ages, but also towards older ages within all statuses. In Portugal, deaths has been shifted towards older ages and towards not-severe limitations. In France, there was an increase in severely limited deaths at younger ages, but a shift towards not limited and severely limited at older ages.

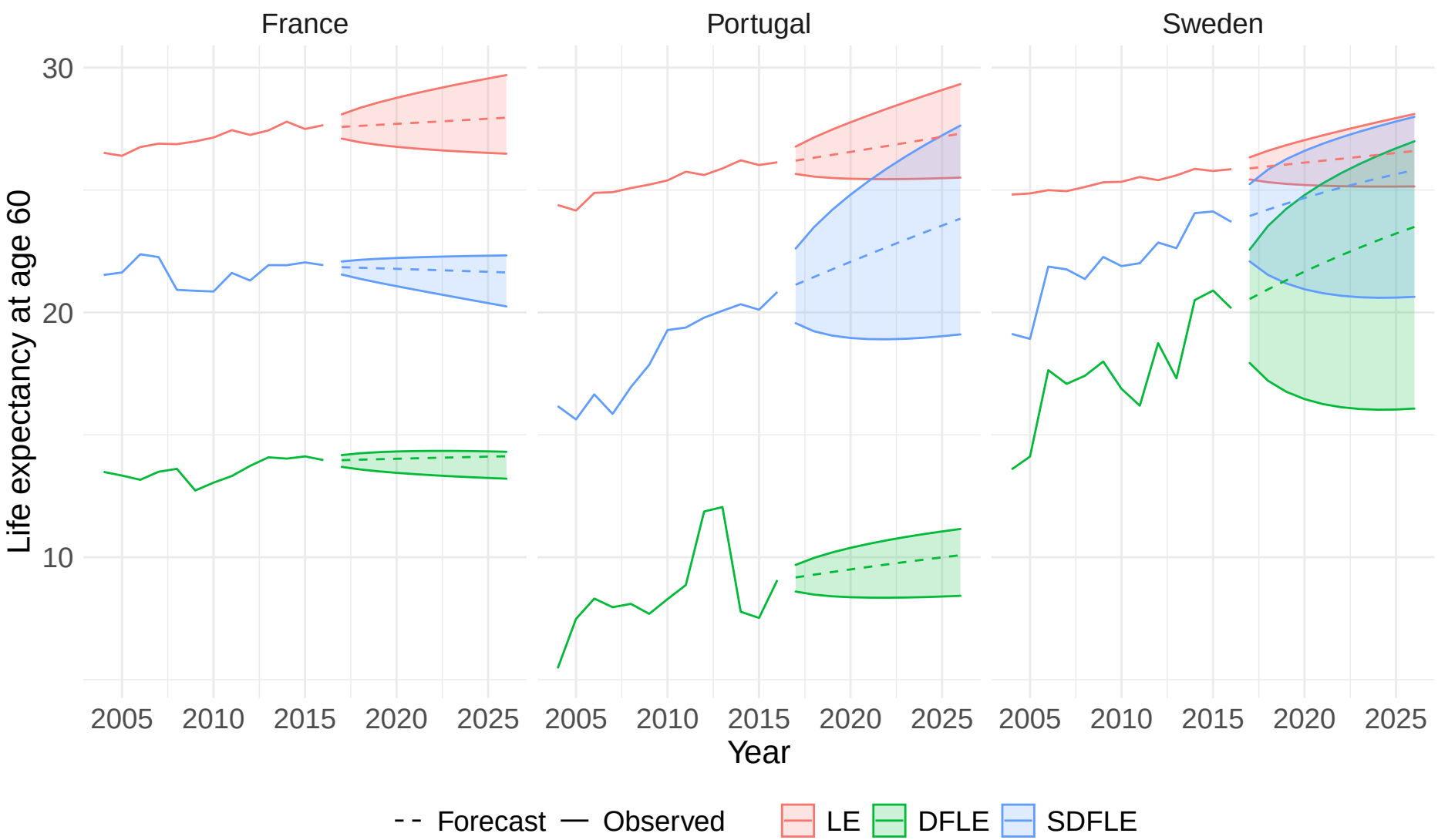
Parameters: β_{xs}



Life expectancy at age 60

The figure on the next slide shows the life expectancy (LE), disability-free life expectancy (DFLE) and severe disability-free life expectancy (SDFLE) at age 60, observed and forecast. In Sweden, we observed and forecast an increase in all three indicators, but a more rapid increase for DFLE and SDFLE, suggesting a compression of morbidity. In Portugal, the dynamic equilibrium hypothesis seems more appropriate, with more years of life lived with not-severe limitations, but less with severe limitations (the gap between LE and SDFLE narrowed). In France, an expansion of morbidity is observed and forecast.

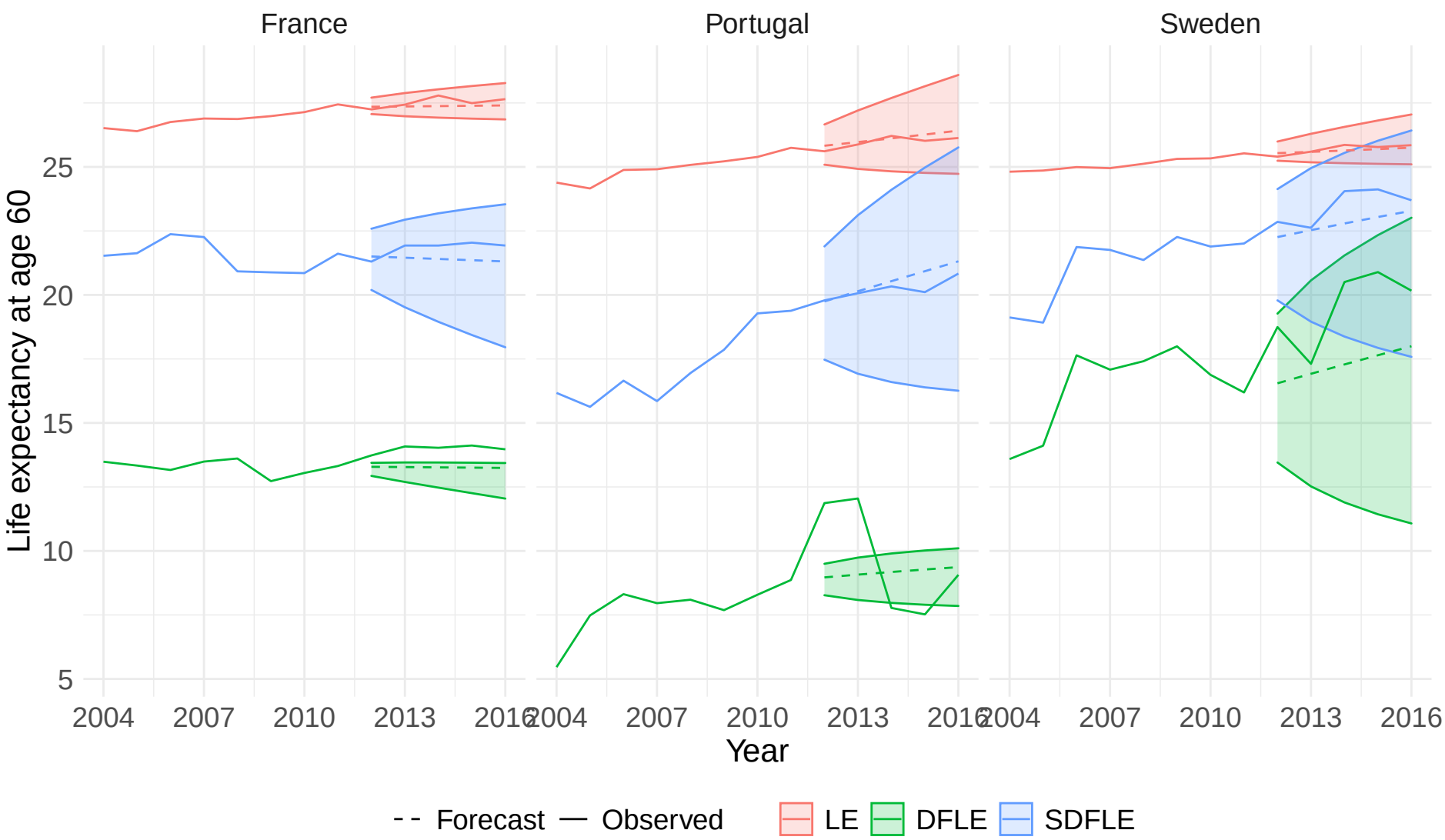
Life expectancy at age 60



Out-of-sample analysis

To test the accuracy of the model, we used data from 2004 to 2011 to forecast LE, DFLE and SDFLE from 2012 to 2016. The model predicted well the LE and SDFLE in all three countries. However, it tends to underpredict DFLE in France and Sweden.

Out-of-sample analysis



Next steps

- Include uncertainty around point estimates and parameters.
- Out-of-sample analysis using multiple fitting periods and forecast horizons.
- Include covariates, such as smoking, to improve the forecast.

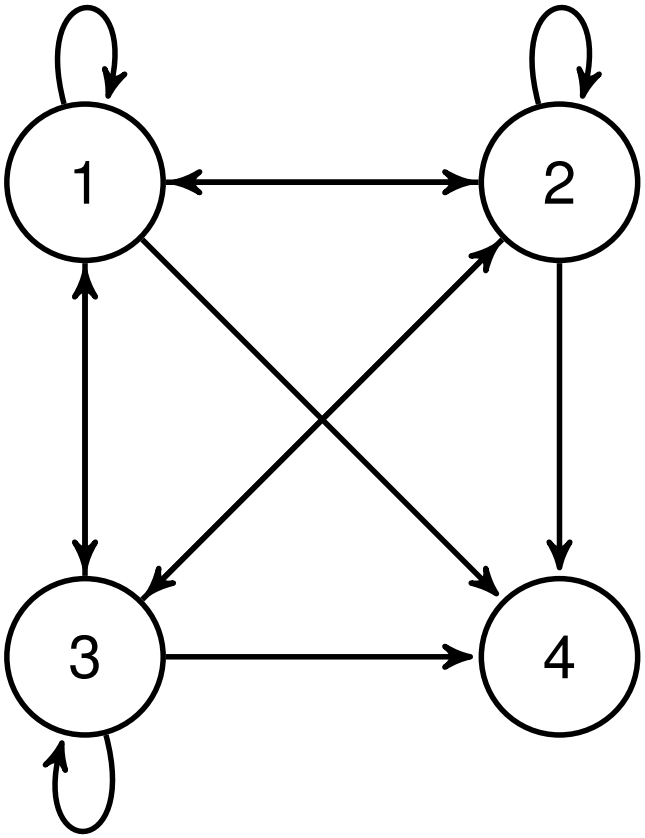
Discussion

1. Correlation between age and status accounted for, by using a CoDA approach.
2. No heavy data required: cross-sectional survey data are enough.
3. Limitations from the Sullivan method:
    1. No health transitions;
    2. Underestimate disability!

Method 2: Multistate model
(Preliminary)

Method

Multistate models are models for a process. Here we are interested in the transition over time across health status: 1. Not limited; 2. Limited but not severely; 3. Severely limited and 4. Dead, being an absorbing state.



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

$$q_{ij} = \frac{D_{ij}}{\sum_j^4 D_{ij}} \tag{5}$$

Method

q_{ij} are compositional data!

When one of the states is death, increment-decrement life tables can be calculated from q_{ij} and obtained the number of years lived in different states!

Method

We can then use the model of Bergeron-Boucher et al. (2018) to forecast the transition rates:

$$clr(q_{t,j,i*x} \ominus \alpha_{j,i*x}) = \kappa_t \beta_j \gamma_{i*x} + \epsilon_{t,j,i*x} \tag{6}$$

- $q_{t,j,i*x}$ are the transition rates by time t as rows; destination state j as columns; and age x and origin state i as the 3rd dimension;
- $\alpha_{j,i*x}$ is the destination-specific geometric mean for each age x and origin state i;
- κ_t, β_j and γ_{i*x} are the dominant components of a Tucker3 decomposition.

Data

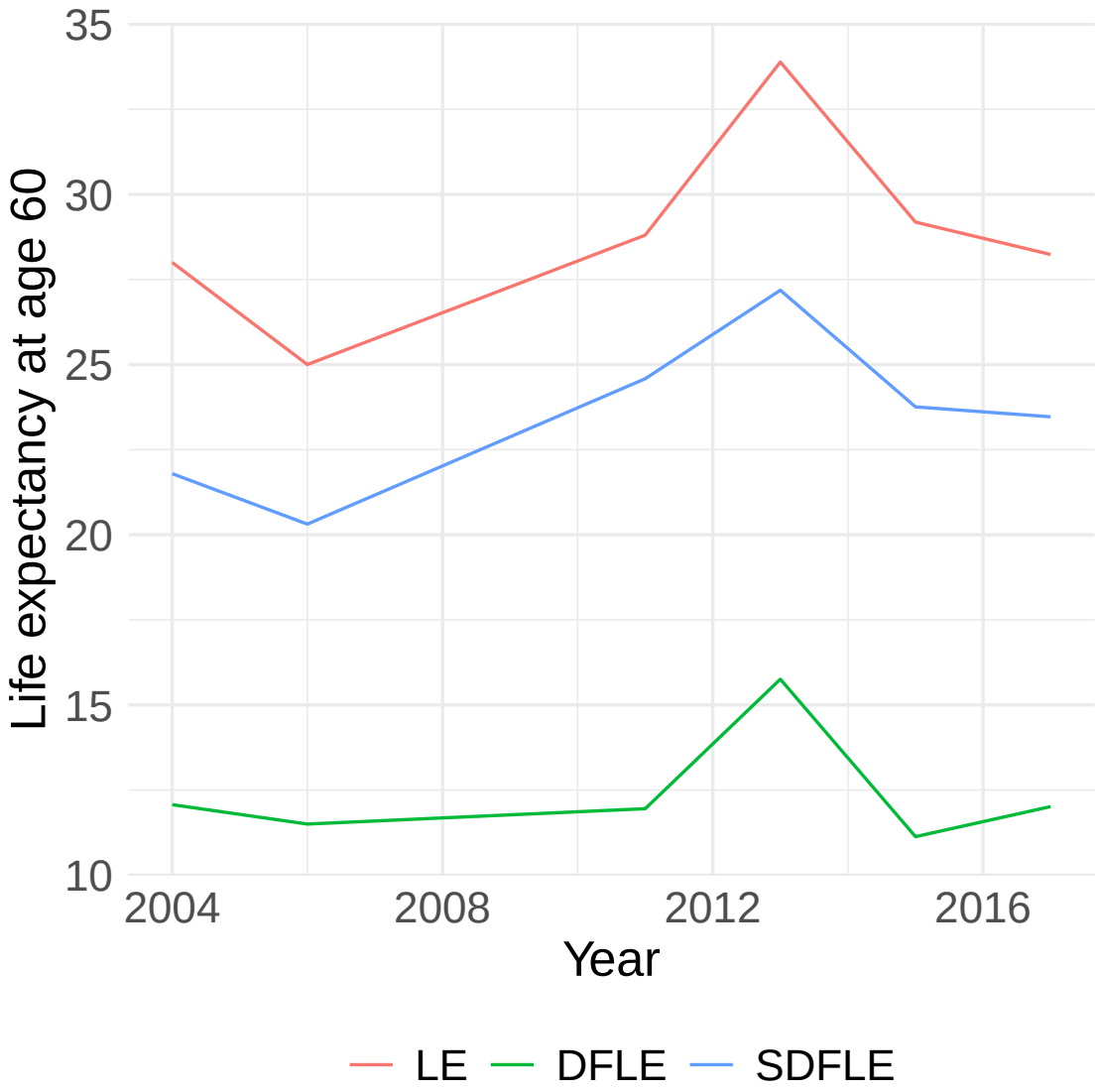
SHARE data for Swedish females

- Waves 2004, 2006, 2011, 2013, 2015, 2017;
- Age 60+ by 5-year age-groups;
- Health measured by activity limitations.

Healthy life expectancy

Unlike the findings based on the Sullivan method and the SILC data, we here observed an increasing number of years lived with not-severe limitations for Swedish females (differences between LE and DFLE increased). The dynamic equilibrium scenario seems more appropriate here.

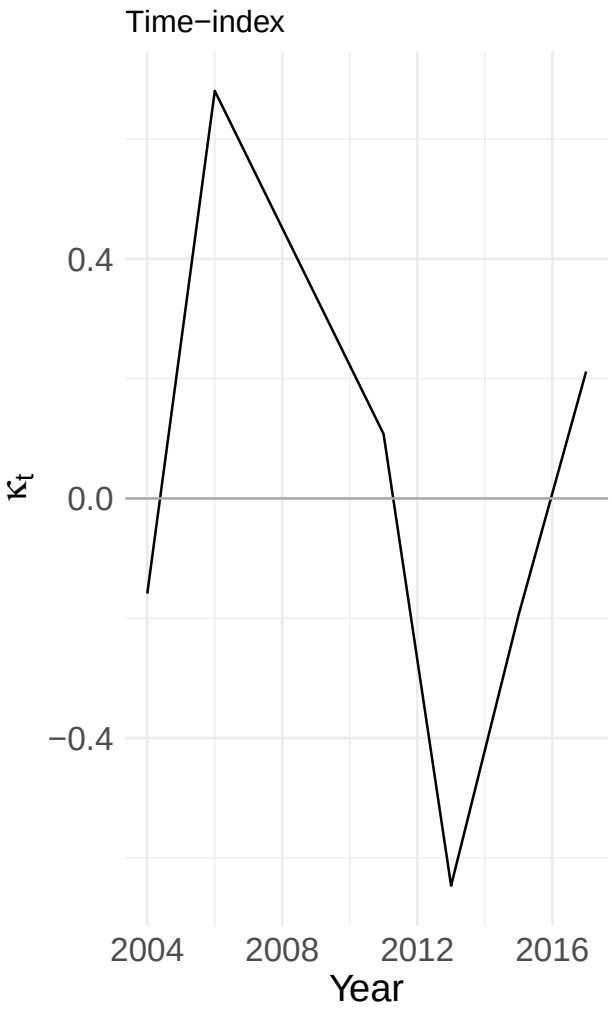
Healthy life expectancy

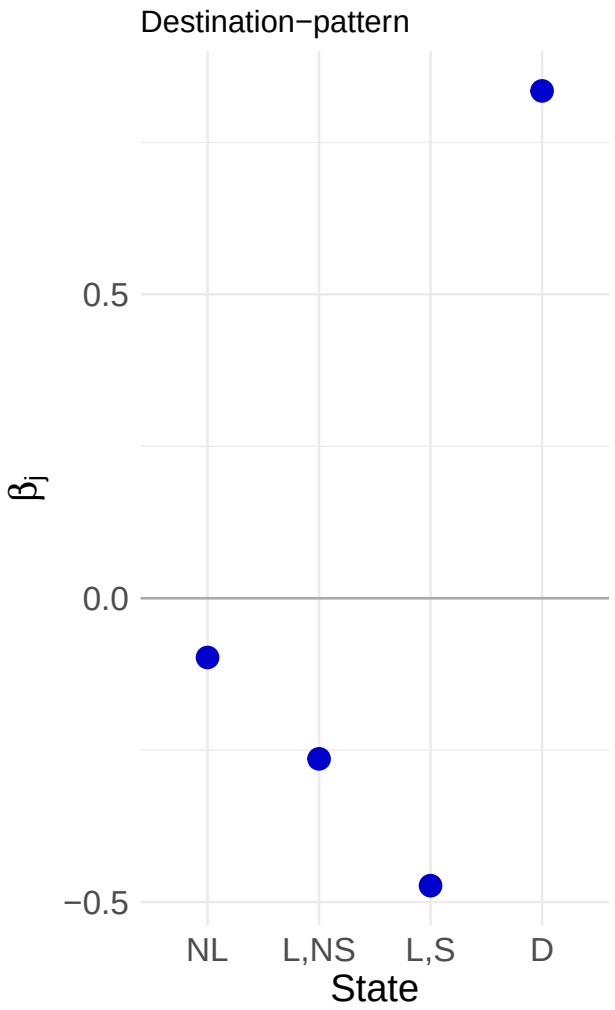


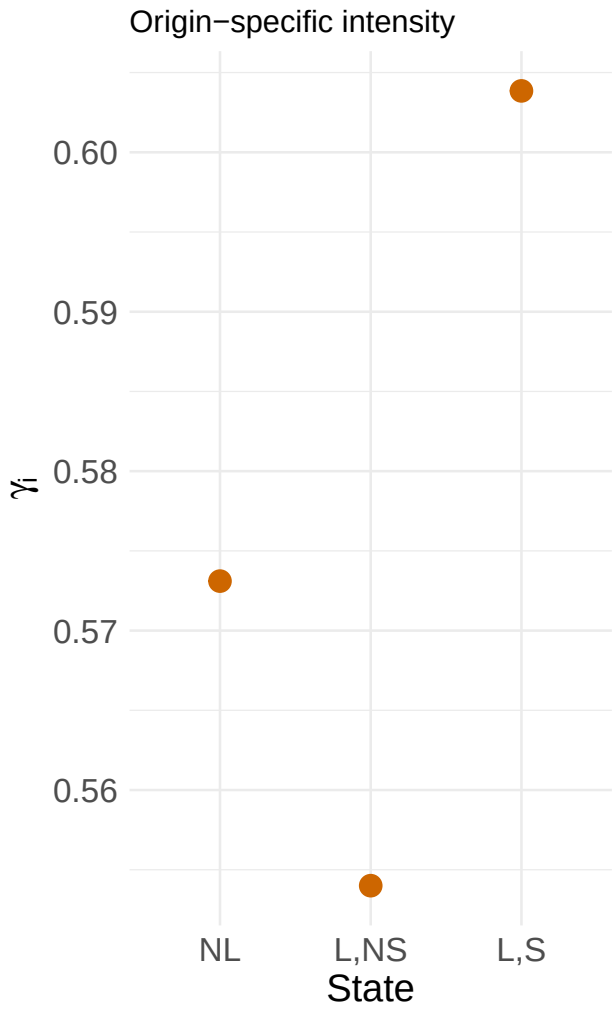
Parameters

The parameters interpretation is similar to that of model based on the Sullivan estimation. Origin-specific intensity indicates how strongly individuals originating in a giving health status experience the transition described in panel A and B in the following figure.
The time-index κ_t has no clear trend. β_j indicates that individuals tend to increasingly die within the origin-status and transition less and less towards other states, in particular less and less towards severely limited.

Parameters


Time−index
0.4
0.0
−0.4
κ_t
2004
2008
2012
2016
Year


Destination−pattern
0.5
0.0
−0.5
β_j
NL
L,NS
L,S
D
State


Origin−specific intensity
0.60
0.59
0.58
0.57
0.56
γ_i
NL
L,NS
L,S
State

Next steps

- Forecast with uneven lags, i.e. when the number of years between the waves is unequal;
- Include uncertainty around point estimates and parameters;
- Out-of-sample analysis to test the model's accuracy;
- Test for separate forecasts by origin and age as a sensitivity analysis.

Discussion

- Correlation between destination status accounted for by using a CoDA approach.
- Data demanding: individuals have to be followed over time.
- Small number of observations, increasing the uncertainty.

Merci! Thank you!

mpbergeron@sdu.dk

Research project funded by the SCOR Foundation for Science.

References

- Aitchison, J. (1986). The Statistical Analysis of Compositional Data. Chapman & Hall, Ltd.,416p.
- Bergeron-Boucher, M.-P., V. Simonacci, J. Oeppen, and M. Gallo (2018). Coherent modeling and forecasting of mortality patterns for subpopulations using multiway analysis of compositions: An application to canadian provinces and territories. North American Actuarial Journal 22 (1), 92–118.
- Borsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmacher, J., Malter, F., et al. (2013). Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE). International Journal of Epidemiology, 42(4), 992–1001. https://doi.org/10.1093/ije/dyt088.
- EHLEIS (2021, October). European Health & Life Expectancy Information System, EuroHex, url: `http://www.eurohex.eu/`
- Fries, J. F. (1980). Aging, natural death, and the compression of morbidity. The New England Journal of Medicine 303 (3), 130–250.
- Gruenberg, E. M. (1977). The failures of success. The Milbank Memorial Fund Quarterly. Health and Society, 3–24.
- Manton, K. G. (1982). Changing concepts of morbidity and mortality in the elderly population. The Milbank Memorial Fund Quarterly. Health and Society, 183–244.
- Oeppen, J. (2008). Coherent forecasting of multiple-decrement life tables: a test using Japanese cause of death data. In Population Association of America (PAA) 2008 Annual Meeting.

Acknowledgment

This presentation uses data from SHARE Waves 1, 2, 4, 5, 6 and 7; see Börsch-Supan et al. (2013) for methodological details. The SHARE data collection has been primarily funded by the European Commission through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARELIFE: CIT4-CT-2006-028812) and FP7 (SHARE-PREP: No. 211909, SHARE-LEAP: No. 227822, SHARE M4: No. 261982).