Smoothing trends and seasonality in short-term mortality forecasting

Ainhoa Elena Léger[1], Silvia Rizzi[1], Ugofilippo Basellini[2]

[1]Interdisciplinary Centre on Population Dynamics
University of Southern Denmark
[2] Max Planck Institute for Demographic Research (MPIDR)

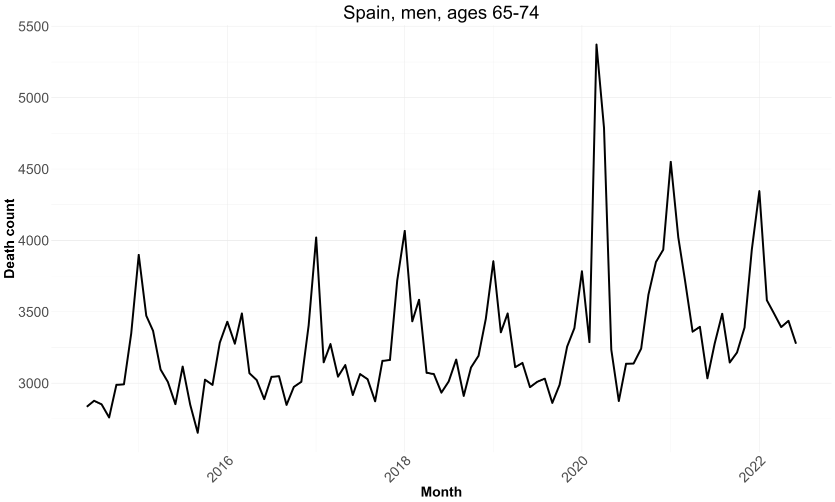
June 14, 2024

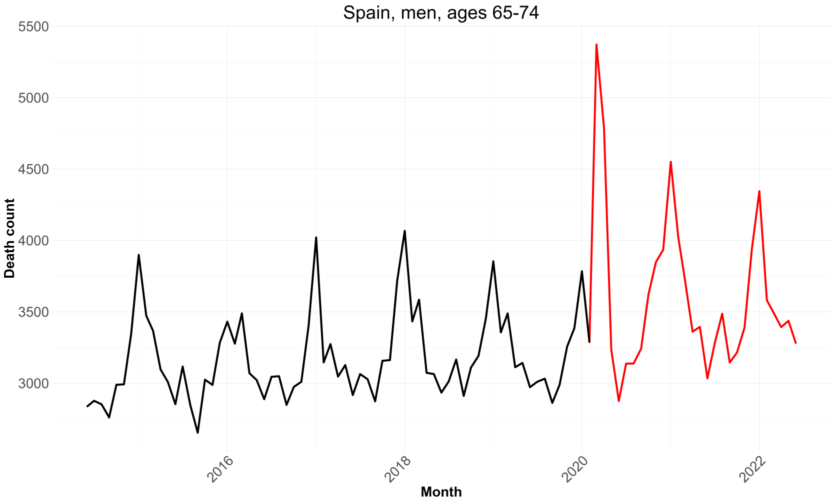
Excess mortality in the literature

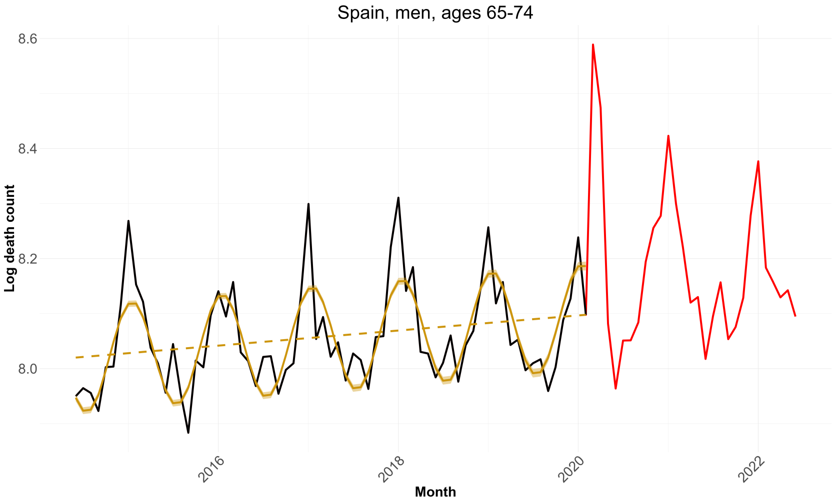
- Difference between the observed and expected deaths (Beaney et al., 2020; Karanikolos and McKee, 2020; Garcia et al., 2021).
- Assessment of the mortality burdens related to **mortality shocks**.
- Seasonal influenza and respiratory diseases (Mazick et al., 2012; Mølbak et al., 2015; Nielsen et al., 2018).
- Effects of heatwaves (Fouillet et al., 2006; Toulemon and Barbieri, 2008).
- Previous pandemics such as the Spanish flu of 1918 (Ansart et al., 2009), the 2009 influenza A(H1N1) virus (Simonsen et al., 2013).

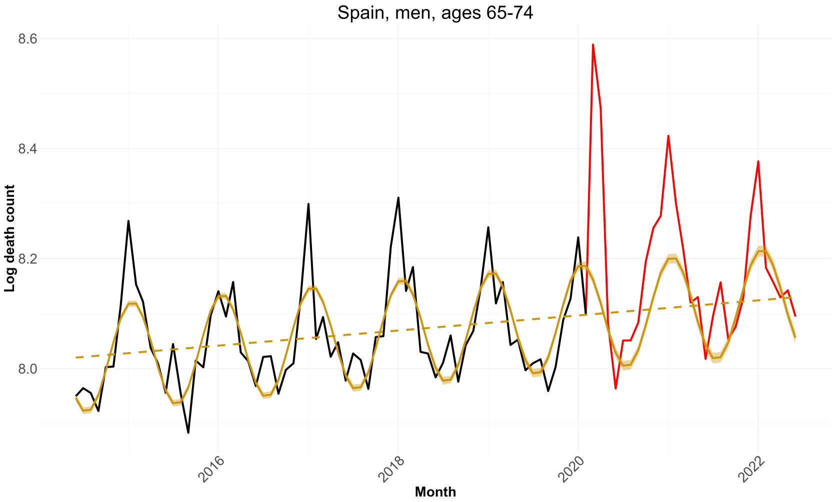
Excess mortality more recently

- Excess mortality during the coronavirus pandemic.
- During the first wave (Kontis et al., 2020; Morgan et al., 2020)
  - France: 29,778 COVID-19 deaths, 21,849 excess death
  - Spain: 28,346 COVID-19 deaths, 44,5050 excess death
- More **comprehensive assessment** of the impact of the health shock than the reported COVID-19 deaths.
- Study subgroups of the population, geographical gradient, and effect of the response policies.
- European mortality monitoring activity (EuroMOMO) hosted by Statens Serum Institute and supported by the European Center for Disease Prevention and Control and the World Health Organization.


Spain, men, ages 65-74
Death count
5500
5000
4500
4000
3500
3000
2016
2018
2020
2022
Month


Spain, men, ages 65-74
5500
5000
4500
4000
3500
3000
Death count
2016
2018
2020
2022
Month


Spain, men, ages 65-74
8.6
8.4
8.2
8.0
Log death count
2016
2018
2020
2022
Month


Spain, men, ages 65-74
8.6
8.4
8.2
8.0
Log death count
2016
2018
2020
2022
Month

A "standard curve of expected seasonal mortality"

- Serfling (1963) introduced a linear regression with a cyclical component.
- The number of deaths y_i is a realization of a Poisson distribution.

$$Y_i \sim Poi(\mu_i), \quad \mu_i = E[Y_i]$$

- The log link function relates the mean μ_i to the linear predictor

$$log(\mu_i) = \eta_i = \tilde{x}_i' \beta$$

- The **Poisson Serfling** regression

$$log(\mu_t) = \beta_0 + \beta_1 x_t + \beta_3 cos(wt) + \beta_4 sin(wt)$$

- Estimation is performed with the Iterated Weighted Least Squares (IWLS).

Short-term mortality forecasts and excess mortality

- Methodological choices influence the estimates of excess death.
- Our **aim** is to propose and compare several apporaches with varying degree of flexibility to forecasting mortality in the short term
  - The **Poisson Serfling** impose a rigid structure on the trend and seasonality
  - The **modulation model** assumes a smooth trend and varying seasonality. We combine the modulation models for seasonal data (Eilers et al., 2008) and forecasting with P-splines smoothing developed in the long term (Currie et al., 2004).
  - We propose a **smooth trend model** with fixed seasonality.

Estimation of a modulation model

- We seek a smooth estimate of $\boldsymbol{\mu} = (\mu_i)$.
- Eilers et al. (2008) developed the modulation models that introduce a smooth trend function and time-varying coefficients.

$$log(\mu_t) = \upsilon_t + f_t cos(wt) + g_t sin(wt)$$

where

- $B = [b_{tj}] = [B_j(t)]$ B-spline basis, $t = 1, \ldots, T$ is time index of the observations, and $j = 1, \ldots, J$ index of the B-splines.
- $\upsilon_t = \sum_j \alpha_j B_j(t)$
- $f_t = \sum_j \beta_j B_j(t)$, $g_t = \sum_j \gamma_j B_j(t)$

Estimation of a modulation model

- In the matrix-vector notation

$$log(\mu_t) = B\alpha + CB\beta + SB\gamma = \eta$$

where

  - $\upsilon = B\alpha$, $f = B\beta$, $g = B\gamma$
  - $C = diag\{cos(wt)\}$, $S = diag\{sin(wt)\}$

- The linear predictor of the modulation model can be expressed as

$$\eta = [B|CB|SB][\alpha'|\beta'|\gamma'] = \boldsymbol{B\theta}$$

Estimation of a modulation model

- Estimation via a penalized version of the IWLS. The estimate $\hat{\theta}$ is obtained iteratively

$$(\boldsymbol{B}'\widetilde{\boldsymbol{M}}^{(t)}\boldsymbol{B} + \boldsymbol{P})\theta^{(t+1)} = \boldsymbol{B}'\widetilde{\boldsymbol{M}}^{(t)}\boldsymbol{B}\theta^{(t)} + \boldsymbol{B}'(y - \tilde{\mu})$$

where

- $\theta^{(t+1)}$ regression coefficients at the t iteration
- $\widetilde{\boldsymbol{M}} = diag(\mu)$ matrix of weights at the t iteration
- $\boldsymbol{P} = \Lambda \boldsymbol{D}'\boldsymbol{D}$ penalty matrix with $\Lambda = diag(\lambda_1, \lambda_2, \lambda_2)$. If different orders are considered for λ_1 and λ_2 then $\boldsymbol{P} = blockdiag(\lambda_1 \boldsymbol{D}'\boldsymbol{D}, \lambda_2 \boldsymbol{D}'\boldsymbol{D}, \lambda_2 \boldsymbol{D}'\boldsymbol{D})$ (Carballo, 2019).

Smooth trend model

- We propose a model with smooth trend and fixed seasonality (ST model).
- We employ the same structure

$$log(\mu_t) = log(e_t) + \upsilon_t + \beta_1 cos(wt) + \beta_2 sin(wt).$$

- The linear predictor models the trend component with the B-splines matrix B and time varying coefficients α, and the seasonal component with the vectors $c_t = cos(wt)$ and $s_t = sin(wt)$ and the coefficients β_1 and β_2

$$\eta = [B|c_t'|s_t'][\alpha'|\beta_1|\beta_2] = \boldsymbol{B\theta}.$$

Forecasting with P-splines

- Forecasting of future values can be treated as a missing value problem (Currie et al., 2004).
- Estimation and forecasting can be estimated simultaneously extending the B-spline basis for the trend and the modulation components.
- We have data y_1 for n_1 months and we forecast n_2 months into the future.
  - We extend the regression matrix B for $n_1 + n_2$ months: $\boldsymbol{B}_+ = [B_+|C_+B_+|S_+B_+]$.
  - We define a weight matrix $\boldsymbol{V} = blockdiag(\boldsymbol{I};0)$ that weights 1 the observations ($\boldsymbol{I}$ has size n_1) and 0 the forecasts (0 has size n_2).
- The IRLS algorithm has a convenient form for fitting and forecasting simultaneously.

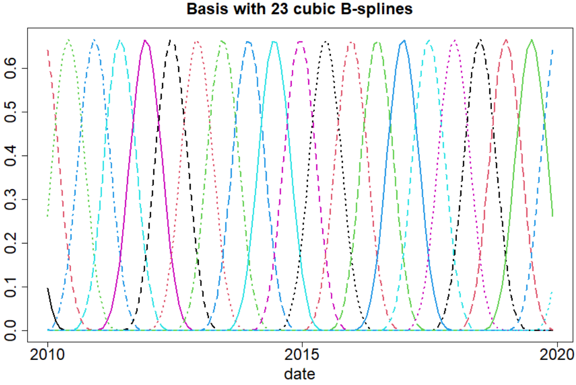
$$(\boldsymbol{B}_+'\boldsymbol{V}\widetilde{\boldsymbol{M}}^{(t)}\boldsymbol{B}_+' + \boldsymbol{P})\theta^{(t+1)} = \boldsymbol{B}_+'\boldsymbol{V}\widetilde{\boldsymbol{M}}^{(t)}\boldsymbol{B}_+\theta^{(t)} + \boldsymbol{B}_+'\boldsymbol{V}(y - \tilde{\mu})$$

Data and application

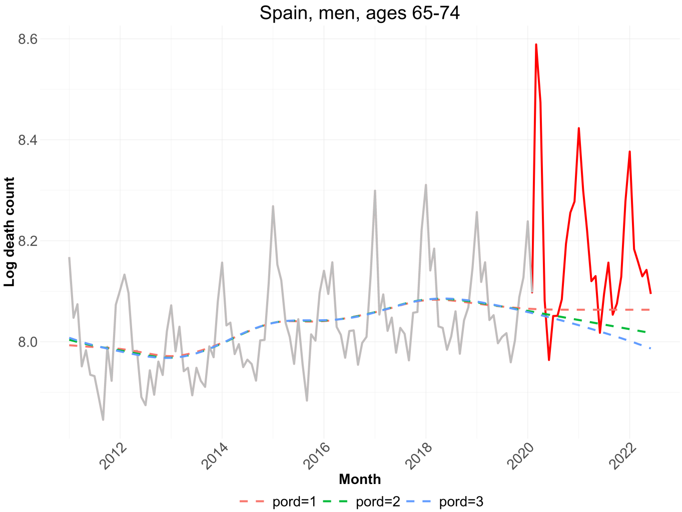
- Data on monthly death counts by sex and age-group for
  - Denmark from 2007 to 2022 (Statistics Denmark),
  - Spain from 2009 to 2022 (Instituto Nacional de Estadística),
  - Sweden from 2000 to 2022 (Statistics Sweden).
- Obtain monthly mortality forecasts for 1-3 epidemic years, comparing
  - the Poisson Serfling (PS),
  - the modulation models (MM),
  - and smooth trend model (ST).

Choice of the parameters

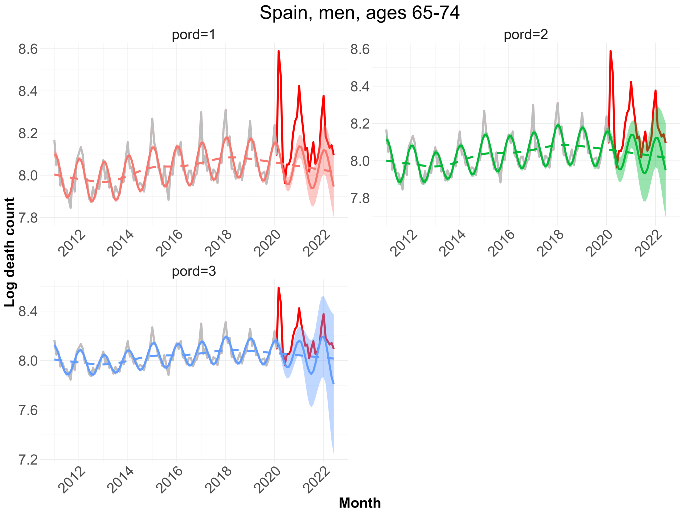
- The model uses a basis of **cubic P-splines** with **2 knots per year**
- Tuning the **smoothness parameters** λ_1 and λ_2 for the smoothness of the model components is achieved using BIC minimization.



- Penalty for the **trend**: second order both for the ST and MM models.



- Penalty for the **seasonal component**: first order for the MM model.



Modelling

Table: Mean BIC resulting of the fit for men. A rolling window of 5 years (from first available year by country to 2019) is used.

Age	PS	MM	ST
Denmark			
0-64	101.623	**101.086**	101.185
65-74	**95.506**	95.512	95.550
75-84	147.339	**143.463**	143.642
85	153.729	**151.290**	153.135
Sweden			
0-64	114.892	**114.835**	114.937
65-74	121.092	**119.878**	120.715
75-84	199.890	**196.062**	198.249
85	255.956	**243.882**	251.276
Spain			
0-64	430.782	**422.837**	426.768
65-74	470.676	**432.816**	440.236
75-84	1226.209	**1030.918**	1105.309
85	1944.721	**1546.566**	1689.830

Historical forecasts

Table: Mean RMSE and MAPE of the forecasts for men. A rolling window of 5 years (from first available year by country to 2019) is used to forecast one year ahead.

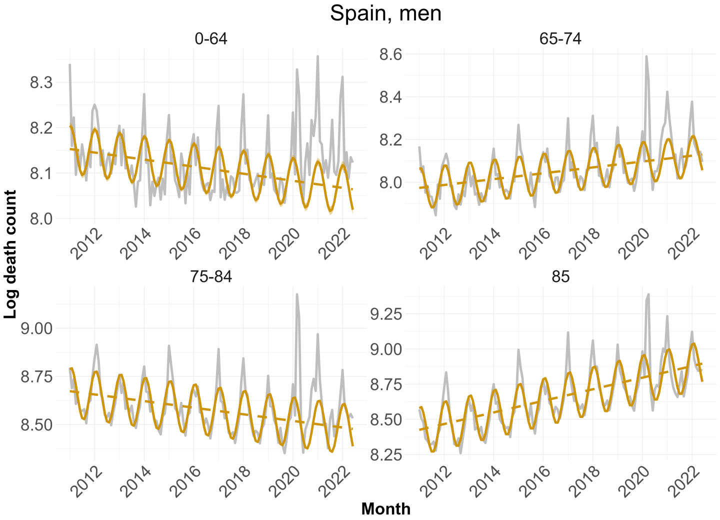
	RMSE			MAPE		
Age	PS	MM	ST	PS	MM	ST
Denmark						
0-64	28.580	28.638	**28.355**	5.301	5.184	**5.115**
65-74	28.819	28.819	**28.818**	4.274	4.274	**4.273**
75-84	43.970	40.439	**40.056**	5.015	4.619	**4.570**
85	41.791	43.224	**41.650**	5.700	5.900	**5.674**
Sweden						
0-64	33.888	33.895	**33.872**	4.765	4.768	4.763
65-74	40.856	40.106	**38.971**	4.781	4.714	**4.575**
75-84	62.899	65.750	**62.458**	4.279	4.421	**4.240**
85	**77.620**	81.240	77.634	**4.858**	5.217	4.963
Spain						
0-64	162.242	160.511	**157.134**	3.584	3.634	**3.508**
65-74	199.894	**185.587**	186.931	4.923	**4.409**	4.507
75-84	**371.580**	399.674	398.937	**5.108**	5.390	5.330
85	**535.402**	552.901	566.327	**6.376**	6.394	6.825

Historical forecasts

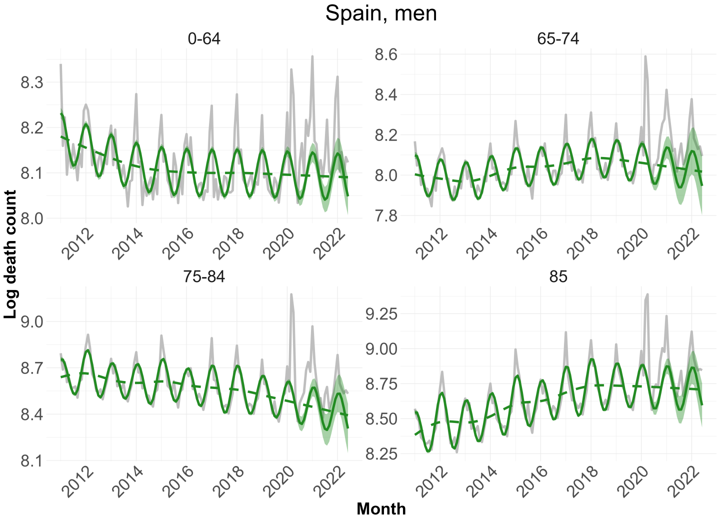
Table: Mean RMSE and MAPE of the forecasts for men. A rolling window of 10 years (from first available year by country to 2019) is used to forecast one year ahead.

	RMSE			MAPE		
Age	PS	MM	ST	PS	MM	ST
Denmark						
0-64	33.986	30.096	**30.010**	6.619	5.654	**5.620**
65-74	32.237	32.230	**32.174**	4.819	4.817	**4.805**
75-84	70.564	45.545	**45.259**	8.452	4.606	**4.576**
85	**44.749**	47.119	47.251	**5.739**	6.076	6.121
Sweden						
0-64	35.721	33.096	**32.922**	5.289	4.999	**4.975**
65-74	54.612	42.246	**42.093**	6.415	4.941	**4.910**
75-84	73.175	**67.649**	67.653	5.192	**4.758**	4.763
85	76.701	79.017	**76.034**	4.959	4.916	**4.802**
Spain						
0-64	142.694	128.477	**125.774**	**2.887**	3.341	3.236
65-74	143.544	137.736	**137.351**	3.498	3.418	**3.284**
75-84	340.270	**293.678**	330.081	6.379	**3.829**	4.495
85	554.042	**479.439**	569.952	7.785	**5.409**	7.023

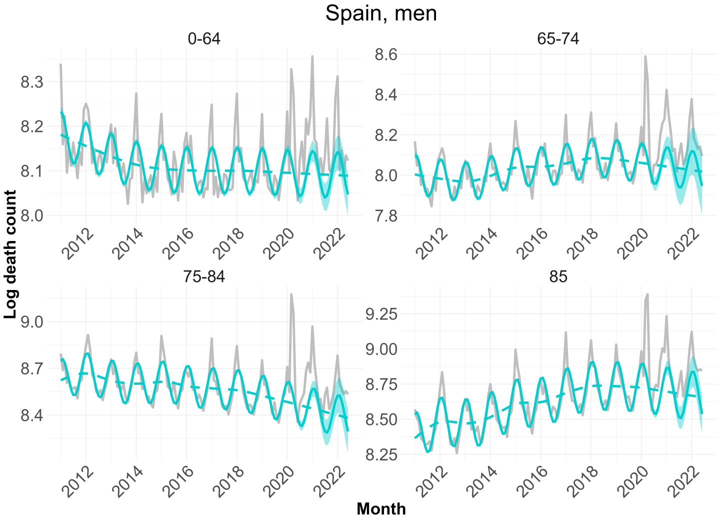
Poisson Serfling model - fit and forecasts



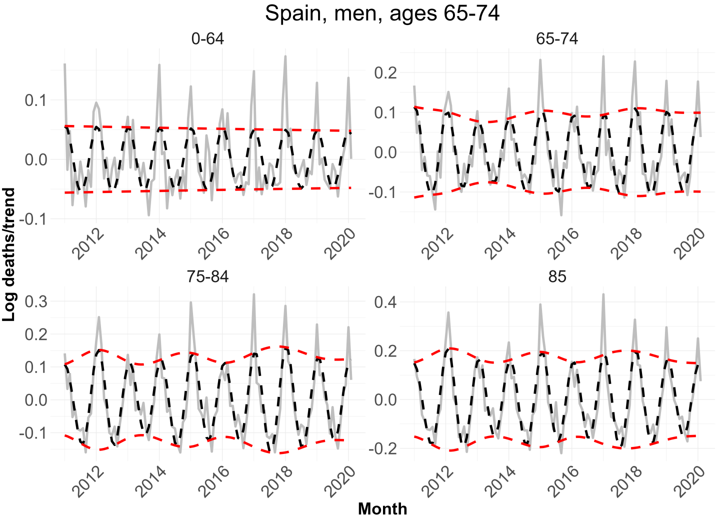
Modulation model - fit and forecasts



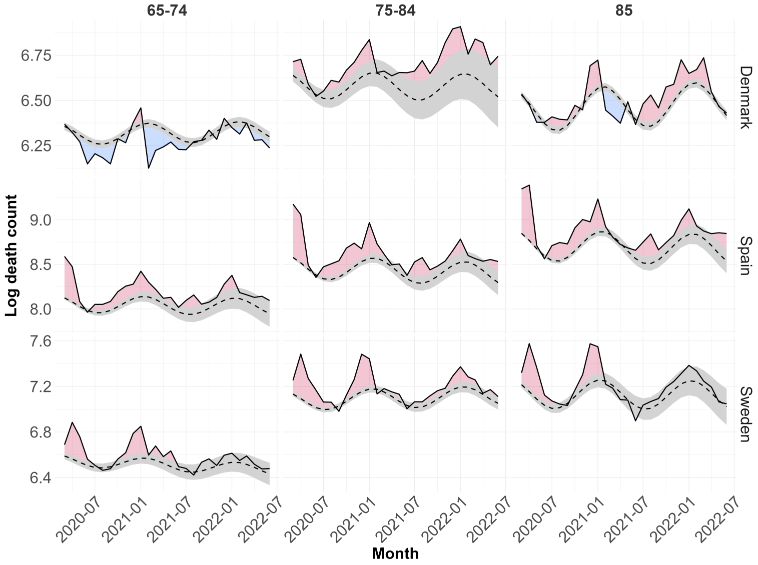
Smooth trend model - fit and forecasts



Amplitude of the modulation component



Excess mortality (ST model), men



Contribution and Further steps

- Contribution
  - Forecasts of baseline seasonality in mortality are useful for quantifying the impact of seasonal influenza, heatwaves, past and future pandemics.
  - **Flexibility** of the method compared to the Poisson Serfling, but be careful about too much flexibility!
  - **Smoothness of the trend** is a desirable feature when modeling the expected seasonality for **all-cause death data**.
- Further steps
  - Improve the **forecasting accuracy** of the smooth trend model, excluding the months of the winter peak and years with severe influenza season.
  - Replicate the study for **death rates** instead of death counts.
  - Verify the model providing more accurate forecasts for **cause-of-death data**.
  - Consider **mortality surface** to model jointly age and years/months.

Thank you

Ainhoa-Elena Leger
aleg@sdu.dk @LegerAinhoa
Silvia Rizzi
srizzi@sdu.dk @silviriz
Ugofilippo Basellini
basellini@demog.mpg.de @ugobas

SDU CPop SCOR FOUNDATION FOR SCIENCE