Smoothing trends and seasonality in short-term mortality forecasting

Ainhoa-Elena Leger[1]
Silvia Rizzi[1]
Ugofilippo Basellini[2]

[1] *Interdisciplinary Centre on Population Dynamics (Cpop), Odense*
[2] *Max Planck Institute for Demographic Research (MPIDR), Rostock*

Extended abstract for the EPC 2024 conference

October 31, 2023

Abstract

Excess mortality is a useful measure to quantify the death toll of various health shocks, from the impact of seasonal influenza to the effects of heatwaves, and pandemics of infectious diseases. Methodological choices about the model to forecast baseline mortality in the absence of a shock can lead to varying estimates of excess death. Starting from modulation models for seasonal death counts developed by Eilers et al. 2008 and combining them with the P-splines forecasting approach developed by Currie et al. 2004, a forecasting strategy is proposed for baseline seasonal patterns. Different specifications of the model are compared. An illustration is shown for Sweden using data from Statistics Sweden from 2008 to February 2020 to forecast mortality from March 2020 to 2021 by age groups. Preliminary results show the forecasts obtained modelling smooth trends and smooth seasonality components.

Introduction

Measuring the mortality burden related to natural and health shocks provide fundamental information to aid public health responses by guiding policy-making decisions. Excess mortality is a useful indicator that has long been used to assess the impact of influenza outbreaks (Mazick et al. 2012; Mølbak et al. 2015), pandemic of infectious diseases, such as the 1918–1920 H1N1 influenza pandemic (Ansart et al. 2009), the 2019 coronavirus pandemic (Kontis et al. 2020), and heat waves (Fouillet et al. 2006). Studies of the mortality experience during shocks permit to evaluate the vulnerability of population subgroups, e.g., of different age groups, the geographical gradient and the effect of different policies adopted in response to the shock.
Excess deaths is computed as the difference between the expected deaths in absence of the mortality shock and the reported deaths on the same period. Various models are available for estimating the expected deaths, which show considerable seasonal variation, mostly striking harder in winter than in summer. First attempts to model a "standard curve of expected seasonal mortality" accounting for secular trends and seasonal variation go back to the contribution of Serfling (Serfling 1963). More recent modelling consider Poisson Serfling regressions (Thompson et al. 2009) considering the count nature of mortality data. Although widely used, a limitation of the Poisson Serfling model is the assumption of the linearity of the trend on the logarithmic scale, which can lead to the under or over estimation of excess mortality, if the trend is not linear.
In this article, we propose a way of predicting the baseline mortality for the estimation of excess deaths that relaxes the liearity assumption of the Serfling Poisson model. Our work builds on the works of Eilers et al. (2008) and Currie, Durban, and Eilers (2004). First, we model seasonal mortality using modulation models (Eilers et al. 2008) that consider smooth long-term trends and varying seasonal effects over time. Second, we forecast a baseline mortality in the short-term via a P-splines smoothing approach. P-splines forecasting has been developed to forecast long-term trends over several years (Currie, Durban, and Eilers 2004) and is here adapted to forecast mortality during an epidemiological year. For comparison purposes, we propose a modification of the modulation models to account for smooth long-term trends and fixed seasonality. In this extended abstract, we show an application to Swedish female mortality.

Methods

Poisson Serfling regression

Let Y_t be non-negative random variable denoting the death counts in a population at the months t, with $t = 1, ..., T$. The realizations of Y_t are the observed number of deaths y_t. We assume that the random variable Y_t follows a Poisson distribution with expected values μ_t.

$$Y_t \sim Poi(\mu_t), \quad \mu_t = E[Y_t]$$

The log link function relates the mean μ_t to the linear predictor $log(\mu_t) = \eta_t$.
The first model that we consider is the Poisson Serfling model, which models the seasonality using sine and cosine functions.

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

where $t = 1, \ldots, T$ and $w = 2\pi/p$ is p is the period. Analyses in this paper are performed on monthly death counts and rates, so $p = 12$. Estimation of the regression coefficients $\hat{\beta}$ can be performed with the Iterated Weighted Least Squares (IWLS) for GLM models.

Modulation model and smooth trend model

Secondly, we consider the modulation models developed by Eilers et al. (2008) and extend them to forecast the baseline mortality by adapting the P-splines approach of Currie, Durban, and Eilers (2004) to seasonal data. The modulation models introduce a smooth trend function and time-varying coefficients. The structure is the following

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

The smooth trend is represented by υ, while f and g are smooth series that describe the local amplitudes of the cosine and sine waves. The model allows for exposures e, when the objective is to model mortality rates (in which case $e_t = 1$). The smooth trend function and the modulation series f and g are constructed by approximating B-splines basis. Specifically, $\upsilon_t = \sum_j \alpha_j B_j(t)$, $f_t = \sum_j \beta_j B_j(t)$ and $g_t = \sum_j \gamma_j B_j(t)$, with $B = [b_{tj}] = [B_j(t)]$ B-splines basis, $t = 1, \ldots, T$ the time index, and $j = 1, \ldots, J$ the B-splines index.
By introducing the matrices $C = diag\{cos(wt)\}$ and $S = diag\{sin(wt)\}$, the model can be written in the matrix-vector notation

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

where $\upsilon = B\alpha$, $f = B\beta$, $g = B\gamma$. The linear predictor can then be re-arranged

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

In addition to employing the Poisson Serfling model and the modulation model, we propose an alternative approach to account for smooth trend component and fixed seasonality. In the following, we will call this approach smooth trend model. Specifically, 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.

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

P-splines forecasting

For both models, estimation of the models' parameters is achieved using the P-splines procedure proposed by Eilers and Marx (1996) that aims to optimize the amount of smoothing. P-splines use B-spline bases to model series such as υ, f and g and different penalties on the B-spline coefficients to increase smoothness. Estimation of the coefficients is then performed via the penalized version of the Iterated Weighted Least Squares (IWLS).

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

where $\boldsymbol{B}$ is the regression matrix, $\widetilde{\boldsymbol{M}} = diag(\mu)$ is the matrix of weights, and $\boldsymbol{P} = \Lambda\boldsymbol{D'D}$ is the penalty term. The positive penalty hyper-parameter Λ balance smoothness against fit to the data, and $\boldsymbol{D}$ is a matrix that forms second-order differences. The modulation model allows for different penalty weights $\Lambda = diag(\lambda_1, \lambda_2, \lambda_2)$ for the trend (λ_1) and modulation functions (λ_2), while the smooth trend model allows a penalty weight for the trend $\Lambda = \lambda_1$.

Following Currie, Durban, and Eilers (2004), the forecasting of future values can be treated as a missing value problem and the fitted and forecast values can be estimated simultaneously. Consider that we have data on death counts y_1 for n_1 months and we forecast n_2 months into the future. The IWLS algorithm is adapted

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

where the regression matrix $\boldsymbol{B}$ is extended for $n_1 + n_2$ months and the matrix $\boldsymbol{V} = blockdiag(\boldsymbol{I}; \boldsymbol{0})$ weights 1 the observations ($\boldsymbol{I}$ has size n_1) and 0 the forecasts ($\boldsymbol{0}$ has size n_2).

Application

We present the results of employing the traditional Poisson Serfling model, the modulation model and the smooth trend model on Swedish female mortality experience by four age groups. We use the monthly data from Statistics Sweden on death counts (Data collection 2023a) and yearly exposures (Data collection 2023b) from epidemic year 2008 to February 2020, to forecast the expected death counts from March 2020 to 2021 if the coronavirus pandemic did not occur. We also provide the 95% prediction intervals. The modulation model and the smooth trend model use a basis of 10 cubic P-splines and smoothing parameters $\lambda_1 = 10^{2.8}$ for the trend and $\lambda_2 = 10^{1.6}$ for the seasonal component (values of λ_1 and λ_2 from Eilers et al. 2008).

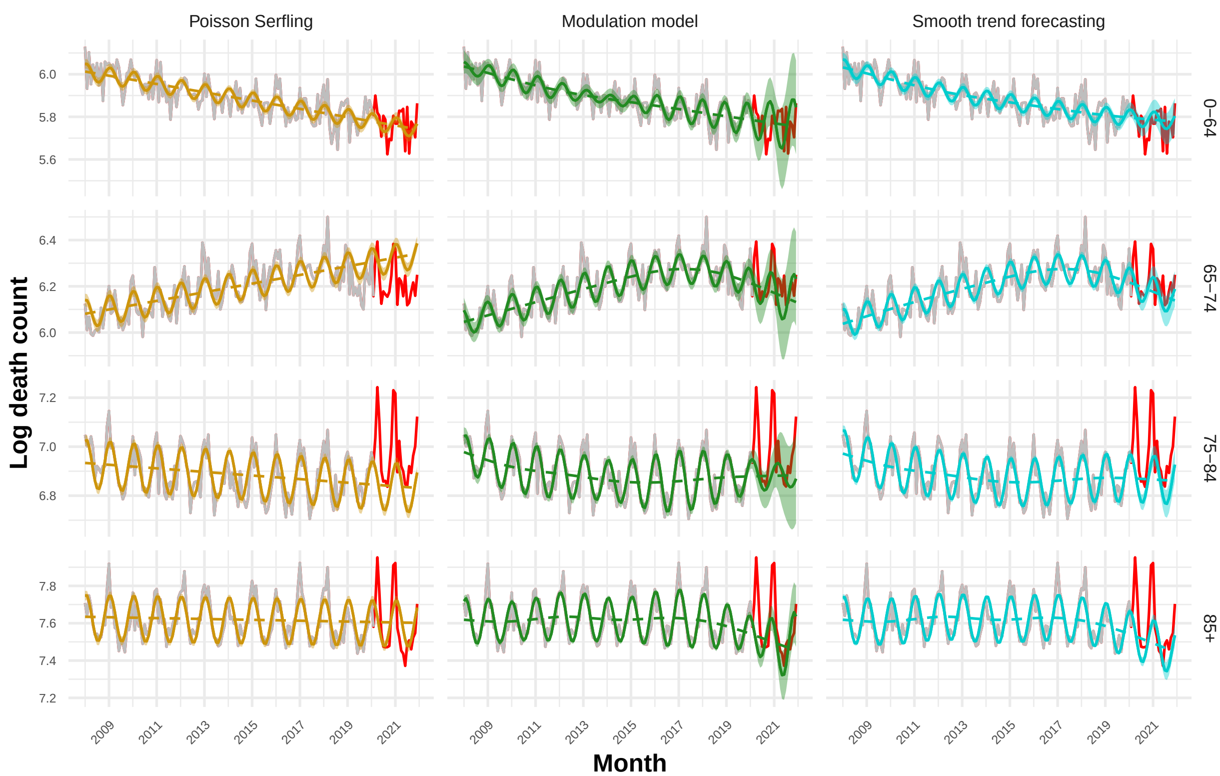
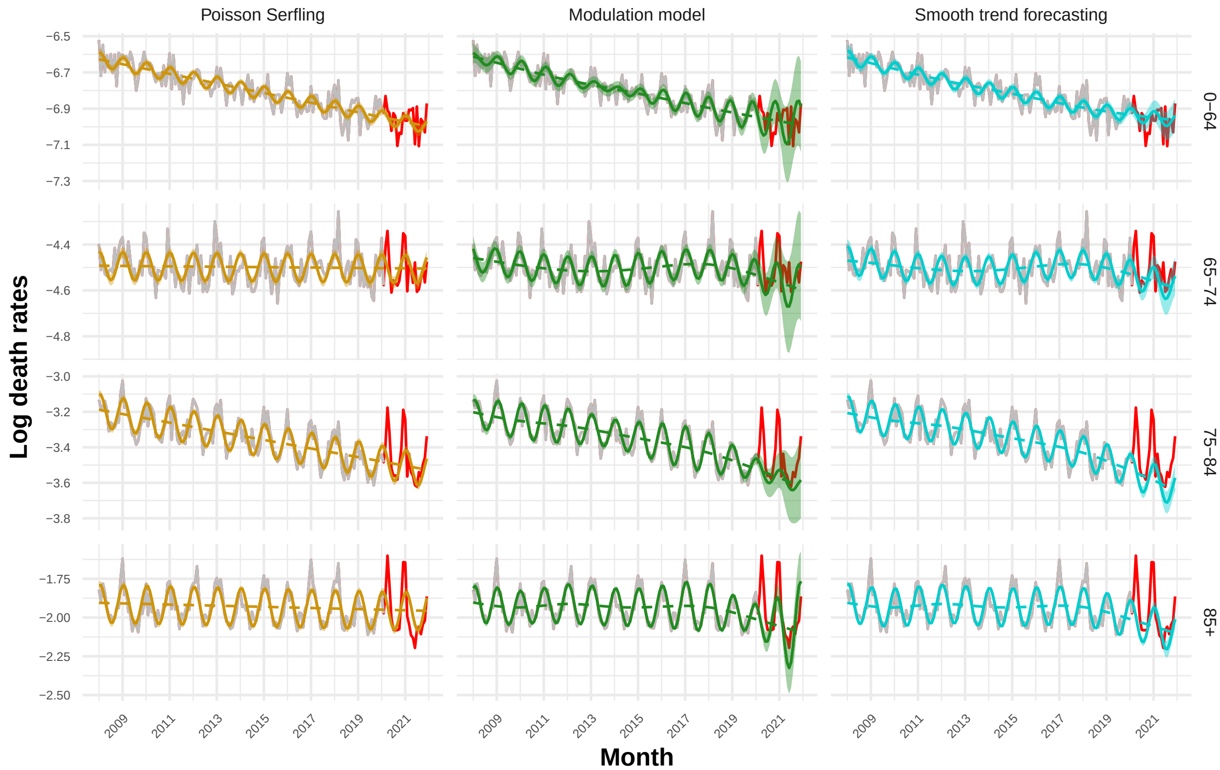


Figure 1 shows the observed death counts for Swedish females by four age groups on the fitting (grey) and forecasting period (red). The forecasts of the smooth trend model (right) seem more plausible than those of the Poisson Serfling model (left) and the modulation models (middle). The trends appear to be non-linear for many age groups, particularly the central ones. The seasonal component shows larger fluctuations and the prediction intervals are wider for modulation model than for the smooth trend model.

Figure 2 present the same results for the death rates. The prediction intervals of the Poisson Serfling are so narrow that the variability in mortality of the age group 0-64 result as excess/deficit mortality. On the other hand, the prediction intervals of the modulation models are too wide to discriminate any excess mortality after the first year of forecasts. The smooth trend model estimates a significant excess mortality during the coronavirus pandemic in the age groups above 65.



Conclusion

The aim of this paper is to propose a new methodology to model and forecast the baseline mortality that is more flexible compared to the widely used Poisson Serfling model. Specifically, we compare the forecast of the modulation model and a model with smooth trend and fixed seasonal component. We apply our approach to forecast mortality of Swedish females during the coronavirus pandemic. Preliminary results show that the smoothness of the trend is a desirable feature, while the Poisson Serfling and the modulation models provide respectively too rigid and too flexible results.

Further analyses are foreseen in the next months. The tuning of the penalty parameters λ_1 and λ_2 will be achieved though AIC minimization on a grid search. Out-of-sample validation will be performed to measure forecast accuracy in terms of point forecasts and prediction intervals.

References

Ansart, Séverine, Camille Pelat, Pierre-Yves Boelle, Fabrice Carrat, Antoine Flahault, and Alain-Jacques Valleron. 2009. "Mortality Burden of the 1918–1919 Influenza Pandemic in Europe." *Influenza and Other Respiratory Viruses* 3 (3): 99–106.

Currie, Iain D, Maria Durban, and Paul HC Eilers. 2004. "Smoothing and Forecasting Mortality Rates." *Statistical Modelling* 4 (4): 279–98.

Data collection, Statistics, Sweden. 2023a. "Deaths Per Month by Region, Region of Birth, Age and Sex. Year 2000 - 2022."

———. 2023b. "Population by Region, Marital Status, Age and Sex. Year 1968 - 2022."

Eilers, Paul HC, Jutta Gampe, Brian D Marx, and Roland Rau. 2008. "Modulation Models for Seasonal Time Series and Incidence Tables." *Statistics in Medicine* 27 (17): 3430–41.

Eilers, Paul HC, and Brian D Marx. 1996. "Flexible Smoothing with b-Splines and Penalties." *Statistical Science* 11 (2): 89–121.

Fouillet, Anne, Grégoire Rey, Françoise Laurent, Gérard Pavillon, Stéphanie Bellec, Chantal Guihenneuc-Jouyaux, Jacqueline Clavel, Eric Jougla, and Denis Hémon. 2006. "Excess Mortality Related to the August 2003 Heat Wave in France." *International Archives of Occupational and Environmental Health* 80: 16–24.

Kontis, Vasilis, James E Bennett, Theo Rashid, Robbie M Parks, Jonathan Pearson-Stuttard, Michel Guillot, Perviz Asaria, et al. 2020. "Magnitude, Demographics and Dynamics of the Effect of the First Wave of the COVID-19 Pandemic on All-Cause Mortality in 21 Industrialized Countries." *Nature Medicine* 26 (12): 1919–28.

Mazick, A, B Gergonne, J Nielsen, F Wuillaume, MJ Virtanen, A Fouillet, H Uphoff, et al. 2012. "Excess Mortality Among the Elderly in 12 European Countries, February and March 2012." *Eurosurveillance* 17 (14): 20138.

Mølbak, K, L Espenhain, J Nielsen, K Tersago, N Bossuyt, G Denissov, A Baburin, et al. 2015. "Excess Mortality Among the Elderly in European Countries, December 2014 to February 2015." *Eurosurveillance* 20 (11): 21065.

Serfling, Robert E. 1963. "Methods for Current Statistical Analysis of Excess Pneumonia-Influenza Deaths." *Public Health Reports* 78 (6): 494.

Thompson, William W, Eric Weintraub, Praveen Dhankhar, Po-Yung Cheng, Lynnette Brammer, Martin I Meltzer, Joseph S Bresee, and David K Shay. 2009. "Estimates of US Influenza-Associated Deaths Made Using Four Different Methods." *Influenza and Other Respiratory Viruses* 3 (1): 37–49.