Quantifying Harvesting:
Estimating and forecasting the effect of mortality shocks on population dynamics

Trifon I. Missov

Interdisciplinary Center on Population Dynamics
University of Southern Denmark

FOUNDATION FOR SCIENCE

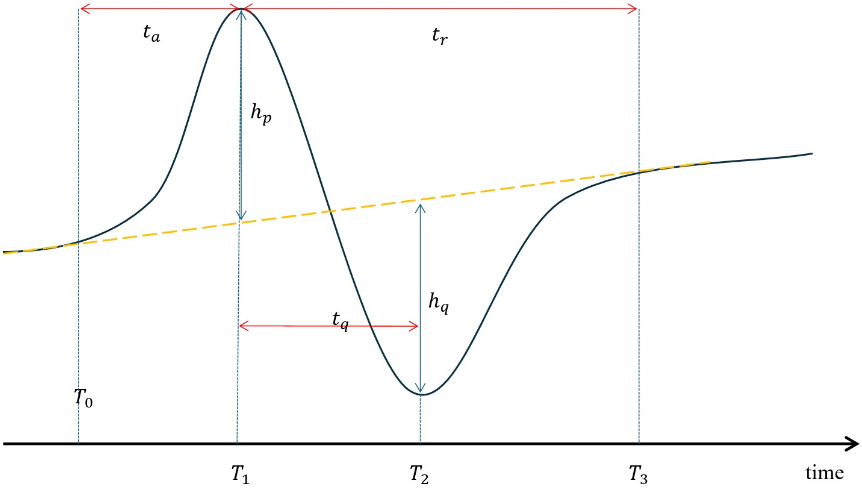
2025 SCOR Chair on Mortality Research, Workshop 3
Odense, 03.04.2025

Harvesting

Mortality displacement when a sudden shock eliminates imminently a higher than anticipated number of individuals, especially those that are expected to die soon.

The outcome, in terms of death counts or death rates, is a curve with a sharp peak, corresponding to the excess deaths caused by the shock, followed by a trough, reflecting the lack of deaths as the expected individuals at risk have already been eliminated.

Harvesting: a simple visualization



CPop COVID-19 Pandemic: A Natural Setup

- Mortality shock: pandemic (multiple waves)
- Data availability: daily/weekly/monthly

  STMF (www.mortality.org)
  MoCy (github.com/jschoeley/mocy)
  COVerAGE-DB (www.coverage-db.org)

- Main task: calculate expected deaths

Expected deaths

Weekly death counts $D_i \sim \text{Poisson}(\lambda_i, \varphi)$ with $\lambda_i = \exp(\xi_i)$.

1. 5-year average model (simple, but with low accuracy)
   $\xi_i = \alpha_i + \ln E_i$

2. Serfling method (accounts for time trends, seasonality)
   $\xi_i = \alpha_i + \beta_i t_i + \gamma_{1,i} sin\frac{2\pi}{52} w_i + \gamma_{2,i} cos\frac{2\pi}{52} w_i + \delta_{i,k[i]} + \ln E_i$

   Extensions (used by Léger, Rizzi & Basellini 2025):
   smooth trend & seasonality (SP-STSS, Eilers et al. 2008)
   smooth trend & fixed seasonality (SP-STFS)

3. Latent Gaussian model
   $\xi_i = \alpha + \beta t_i + \zeta_{t[i]} + \gamma_{t[i]} + \delta_{k[i]} + \varepsilon_i + \ln E_i$

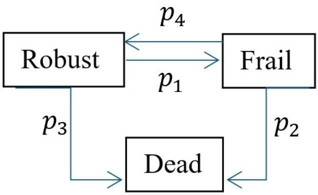
Expected deaths: Accuracy

Schöley (2021), using weekly data, and Léger, Rizzi & Basellini (2025), using monthly data, compared the accuracy of forecasts on multiple countries. They arrived at a similar conclusion in favor of a Serfling model with relaxed specification (assuming smoothness as in SP-STFS or something else).

Both studies find, in general, little evidence in favor of harvesting, no matter how expected deaths are obtained.

Delayed Effects (Work in Progress)

Is there something more we can do? Are there any delayed effects?



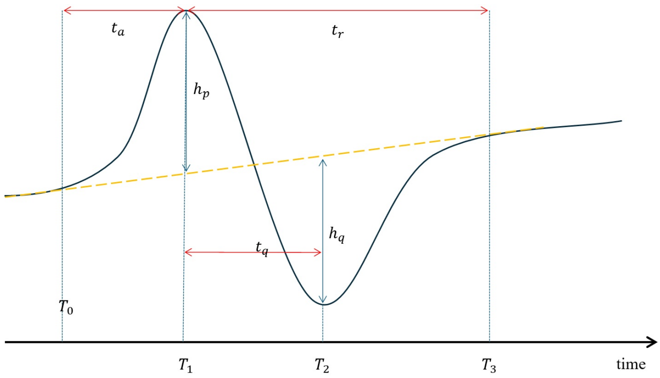
Idea: Add another AR series to the linear predictor with order equal to time elapsed from reaching peak, e.g., death counts $D_i \sim \text{Poisson}(\lambda_i, \varphi)$ with $\lambda_i = \exp(\xi_i)$ and

$$\xi_i = \alpha_i + \beta_i t_i + \gamma_{1,i} sin\frac{2\pi}{52} w_i + \gamma_{2,i} cos\frac{2\pi}{52} w_i + \delta_{i,k[i]} + \sum_{n=1}^{i-T} \psi_i \xi_{i-n} + \ln E_i$$

What if we consider densities?

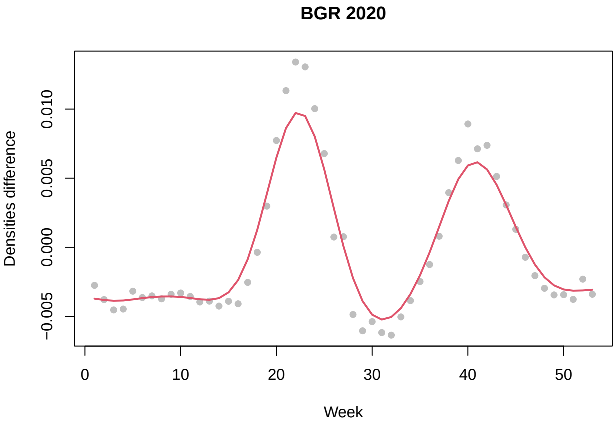
Suppose

- $D_{w,y}$ = death counts in week w in epi-year y
- $d_{w,y} = D_{w,y} / \sum_y D_{w,y}$ is the density of deaths for that week, normalized by the total deaths across years
- $\bar{d}_{w,y}$ represents the five-year average of densities for week w from year from year y to $y + 4$
- In the context of the COVID year, y^*, we calculated the difference (), we calculate the difference $diff_w$ between the density for week w in year y^* and the five-year average density (year average density, d_{w,y^*-5}, for the same week.
- To smooth the variations in $diff_w$, we apply the Nadaraya–Watson kernel regression.



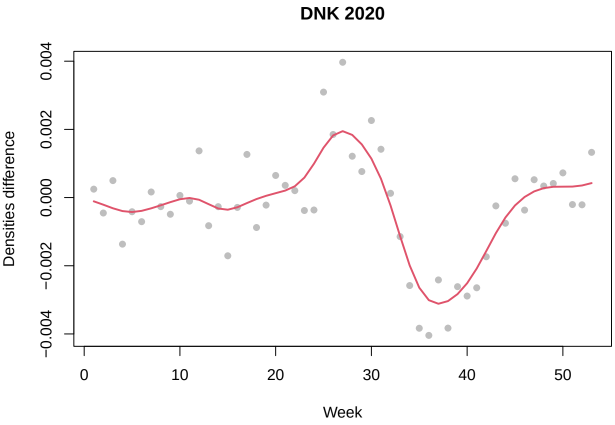
To determine T_1 and T_2, we carry out global optimization by the differential evolution technique (R library DEoptim).

T_3, the inflection point occurring after T_2, is calculated by using a combination of numerically estimated Hessians and applying the Richardson extrapolation method, implemented through the numDeriv package in R.

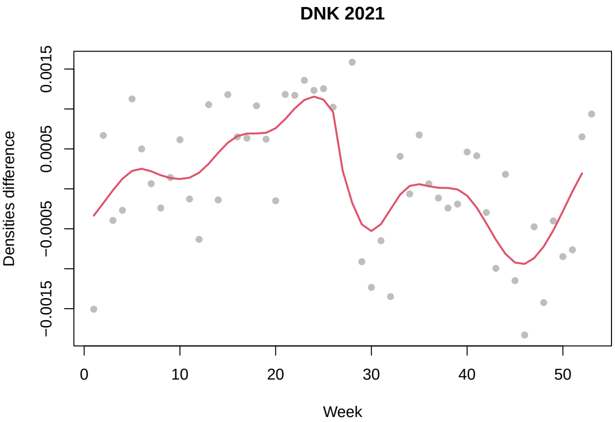


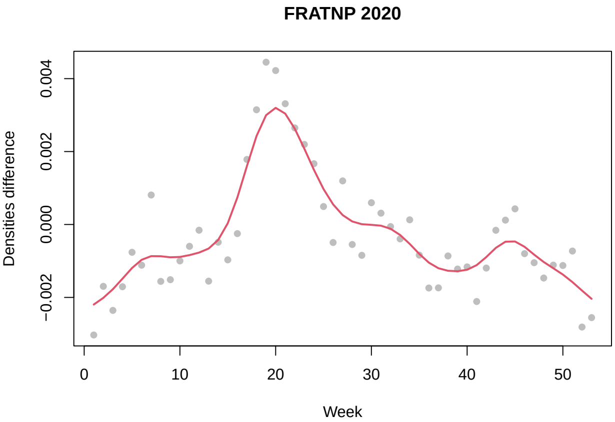
BGR 2020
Densities difference
0.010
0.005
0.000
−0.005
0
10
20
30
40
50
Week

CPop

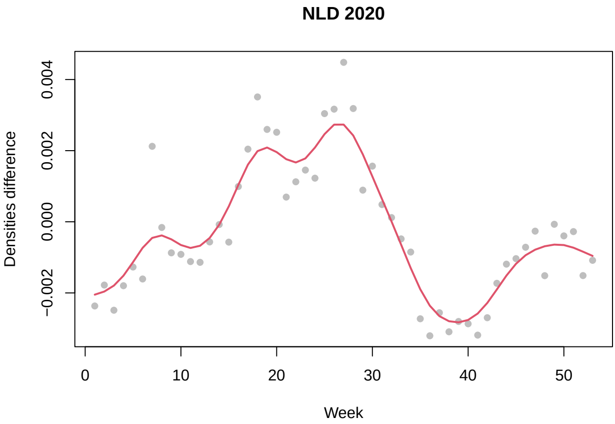

DNK 2020
Densities difference
0.004
0.002
0.000
−0.002
−0.004
0
10
20
30
40
50
Week

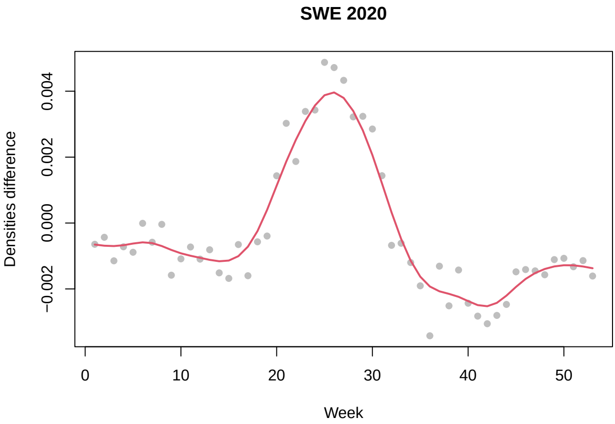
CPop


DNK 2021
Densities difference
0.0015
0.0005
−0.0005
−0.0015
0
10
20
30
40
50
Week

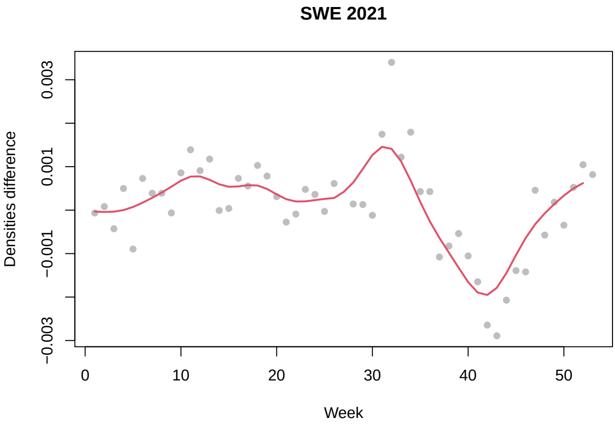


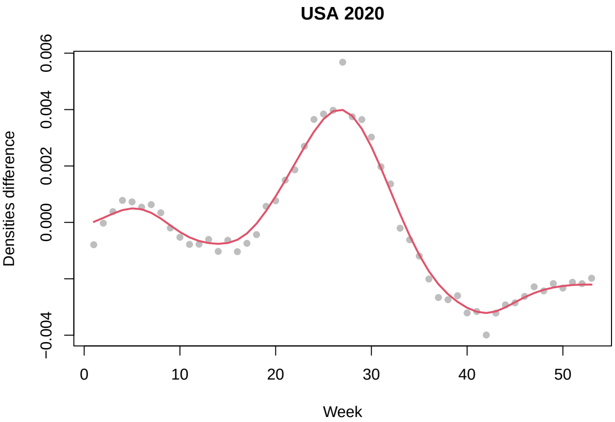
FRATNP 2020
Densities difference
0.004
0.002
0.000
−0.002
0
10
20
30
40
50
Week

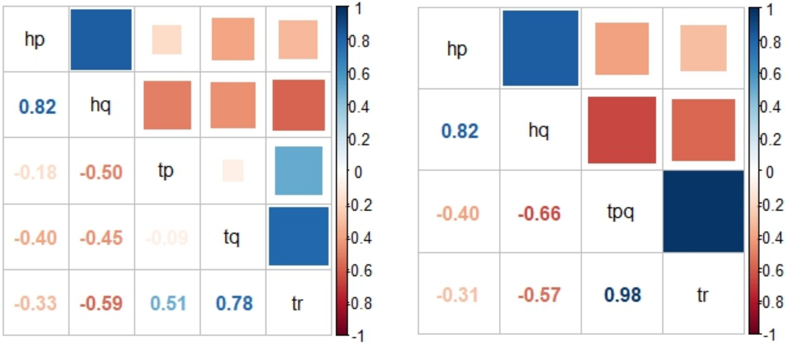

NLD 2020
Densities difference
0.004
0.002
0.000
−0.002
0
10
20
30
40
50
Week



SWE 2020
Densities difference
0.004
0.002
0.000
−0.002
0
10
20
30
40
50
Week


SWE 2021
Densities difference
0.003
0.001
−0.001
−0.003
0
10
20
30
40
50
Week


USA 2020
Densities difference
0.006
0.004
0.002
0.000
−0.004
0
10
20
30
40
50
Week



Correlation between pairs of h_p, h_q, t_p, t_q, and t_r (left panel), as well as between pairs of h_p, h_q, $t_{pq} = t_p + t_q$, and t_r (right panel) based on the 14 studied populations.

Further Steps

- Study model
  $D_i \sim \text{Poisson}(\lambda_i, \varphi)$ with $\lambda_i = \exp(\xi_i)$ and ξ_i given by

$$\alpha_i + \beta_i t_i + \gamma_{1,i} sin \frac{2\pi}{52} w_i + \gamma_{2,i} cos \frac{2\pi}{52} w_i + \delta_{i,k[i]} + \sum_{n=1}^{i-T} \psi_i \xi_{i-n} + \ln E_i$$

- Using individual data, estimate shares of people in different states (Robust, Frail, Dead)
- Study what causes of death have been affected by COVID-19
- Study impact of COVID-19 on other subgroups (e.g., by residence)

THANK YOU