Age Heaping in Population Data of Emerging Countries

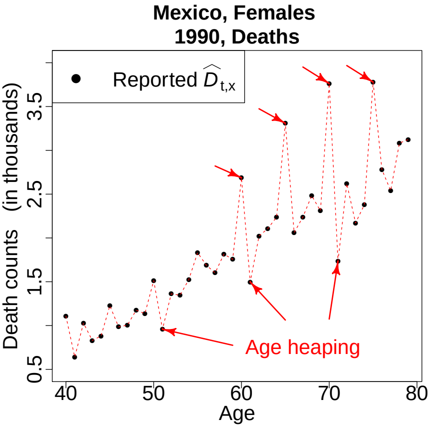
Andres Barajas Paz, Andrew Cairns, Torsten Kleinow

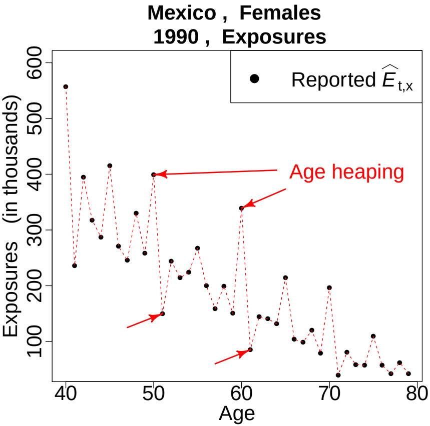
Heriot-Watt University, Edinburgh

ab108@hw.ac.uk

5th HMD symposium,
Harnack Haus, Berlin, May, 2019

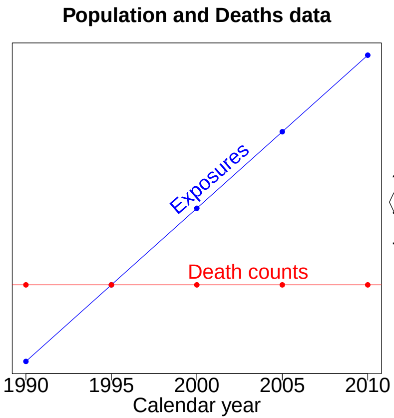
Motivation.



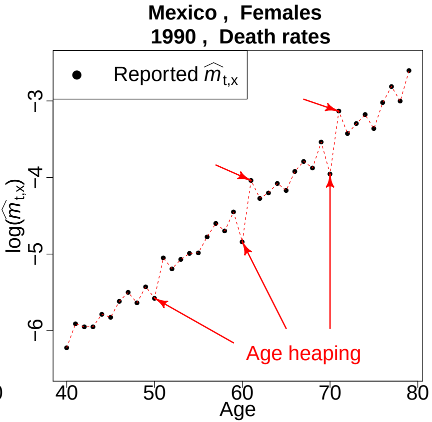


Age Heaping occurs when people misreport age.

Motivation.



Population and Deaths data
Exposures
Death counts
1990
1995
2000
2005
2010
Calendar year



Mexico , Females
1990 , Death rates
Reported $\widehat{m}_{t,x}$
$\log(\widehat{m}_{t,x})$
−3
−4
−5
−6
Age heaping
40
50
60
70
80
Age

Motivation.

- Mortality analyses $\xrightarrow{\text{High quality mortality data}}$ HMD.

However, in many countries population and deaths data can be somewhat unreliable.

- Population $\xrightarrow{\text{Misreporting of age}}$ census.
- Deaths $\xrightarrow{\text{Misreporting of age}}$ vital statistics.

Main Objective

- Develop mortality models for countries where population data is affected by age heaping.

Application: Reported data $\rightarrow$ Smoothed HMD $\rightarrow$ International Reinsurance.

We design a model taking into account two dimensional data. Hence, we consider the data by age x and calendar years $t \in \mathcal{T}$ for deaths, and census years $t \in \mathcal{T}_c$.

- $\mathcal{T} = \{1990, \cdots, 2010\}$,
- $\mathcal{T}_c = \{1990, 1995, 2000, 2005, 2010\}$.

MLE and Bayesian approaches

For any calendar year $|t|$, $n_t = |t|$ is the length of the years in our data set. Where x represents the reported age and y the true age.

$$E_{t,y} \xrightarrow{\text{Age heaping}} \widehat{E}_{t,x}$$

$$D_{t,y} \xrightarrow{\text{Age heaping}} \widehat{D}_{t,x}$$

Model and Notation

$$\widehat{D}(t,x)\Big|\underline{m},\underline{E},\theta \sim Poisson\Big(\sum_{y} m(t,y)E(t,y)g^{D}(t,x,y)\Big),\ \forall t \in \mathcal{T}$$

$$\widehat{E}(t,x)\Big|\underline{E},\theta \sim Poisson\Big(\sum_{y} E(t,y)g^{E}(t,x,y)\Big),\ \forall t \in \mathcal{T}_c.$$

$$m(t,y) = \exp\Big[a(t) + b(t)\big(y-\overline{y}\big) + c(t)\Big(\big(y-\overline{y}\big)^2 - \sigma_y^2\Big)\Big].$$

log-likelihood function

$$\ell_E(\theta) = \sum_x \sum_{t \in \mathcal{T}_c} \left[\widehat{E}(t,x) \log \left(\sum_y E(t,y) g^E(t,x,y) \right) - \sum_y E(t,y) g^E(t,x,y) \right] + \text{Const}$$

$g^* \rightarrow$ Probability that an individual report their true age (y) given that they have reported age (x) in calendar year (t),

$\eta^* \rightarrow$ Captures the increasing improvements across years,

$H^* \rightarrow$ Describes "popular" ages.

HERIOT WATT UNIVERSITY

Actuarial Research Centre
Institute and Faculty of Actuaries

Where,

$$\theta = \{\underline{E}, \underline{\eta}^D, \underline{\eta}^E, \underline{H}^D, \underline{H}^E, \underline{a}, \underline{b}, \underline{c}, \delta_A, \mu_b, \mu_c\}$$

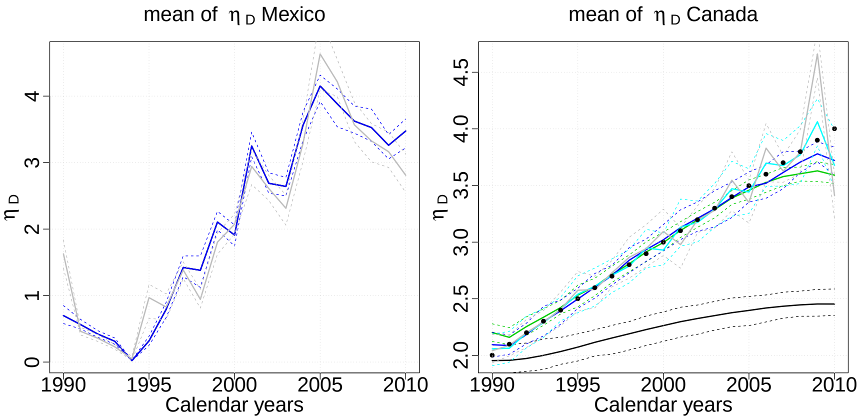
Prior distributions

$$E(t, y) \sim U(0, \infty) \textbf{ iid},$$

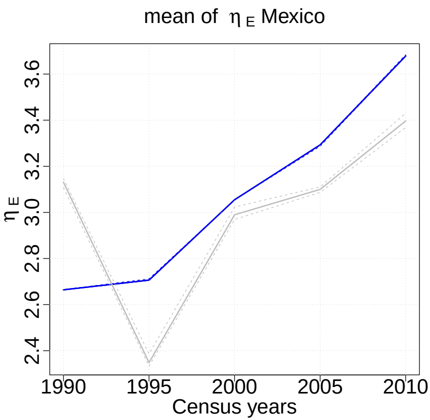
$$\eta^D(t) \sim \textbf{ random walk}, \qquad \eta^E(t) \sim \textbf{ random walk},$$

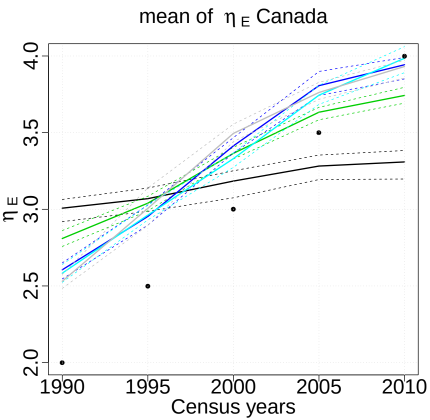
$$\left|H^D(x)\right| \sim \textbf{Exp}(\lambda) \textbf{ iid}, \qquad \left|H^E(x)\right| \sim \textbf{Exp}(\lambda) \textbf{ iid}.$$

Parameter $\underline{\eta^D}$ M-H, Mexico & Canada.

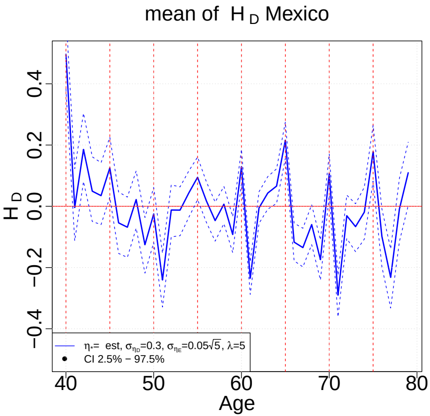


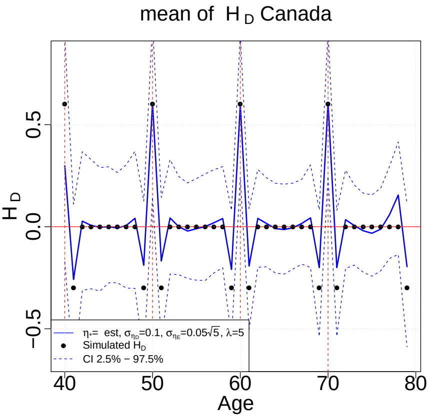
Parameters η^E M-H, Mexico & Canada.



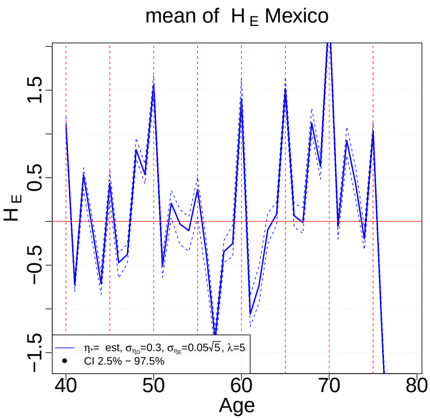


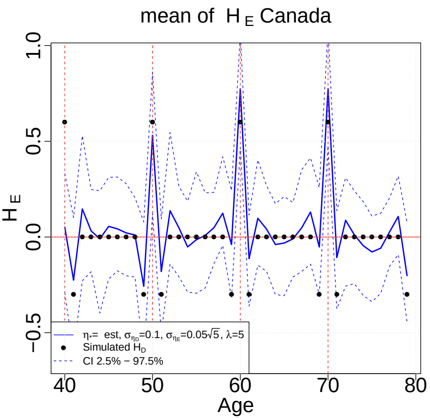
Parameter H^D M-H, Mexico & Canada.



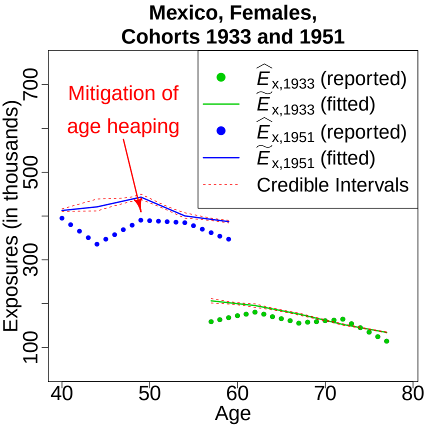


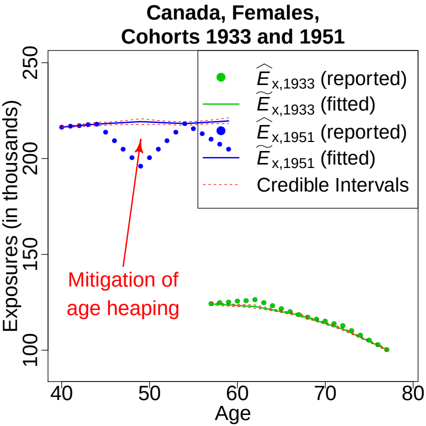
Parameter $\underline{H^E}$ M-H, Mexico & Canada.





Fitted exposures by Cohort, Mexico & Canada.

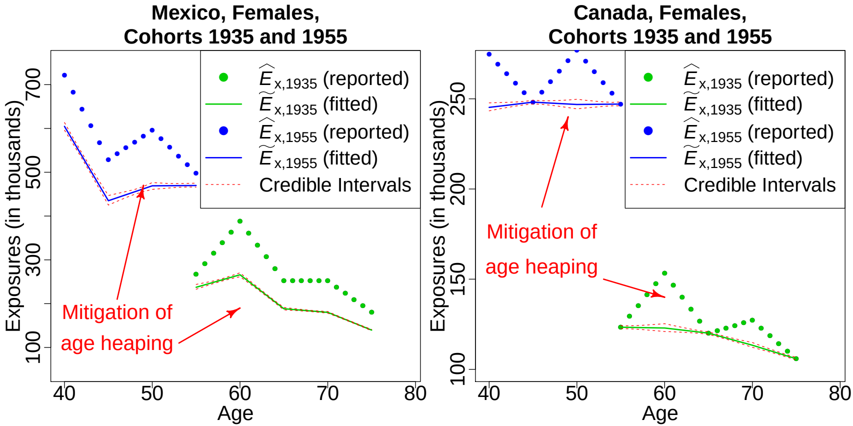




HERIOT WATT UNIVERSITY

Actuarial Research Centre
Institute and Faculty of Actuaries

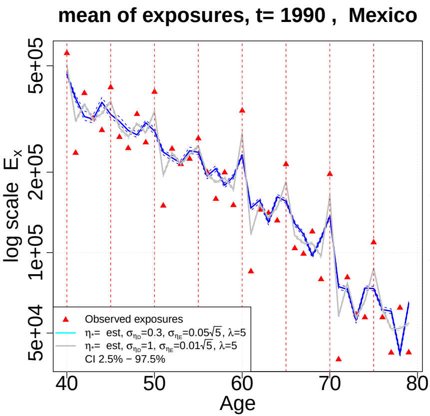
Fitted exposures by Cohort, Mexico & Canada.

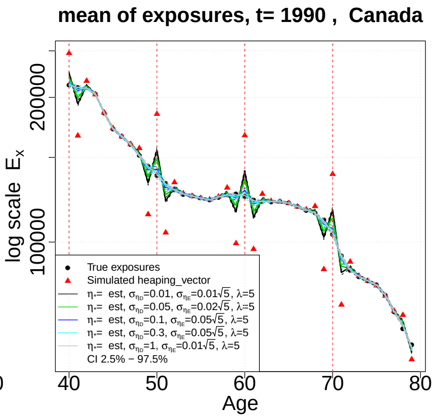


HERIOT WATT UNIVERSITY

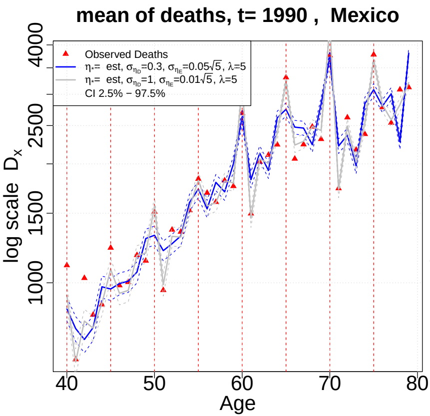
Actuarial Research Centre
Institute and Faculty of Actuaries

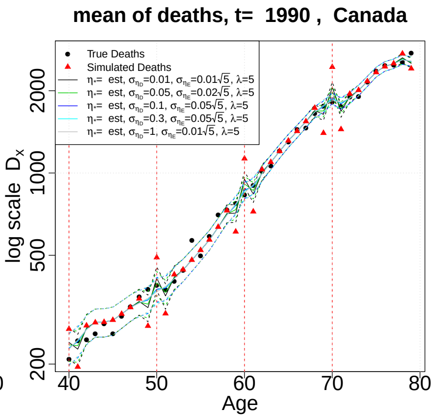
Fitted exposures $E_{t,y}$, Mexico & Canada 1990.





Fitted deaths $D_{t,y}$, Mexico & Canada 1990.

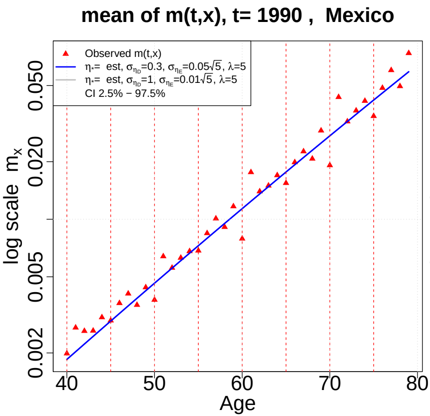


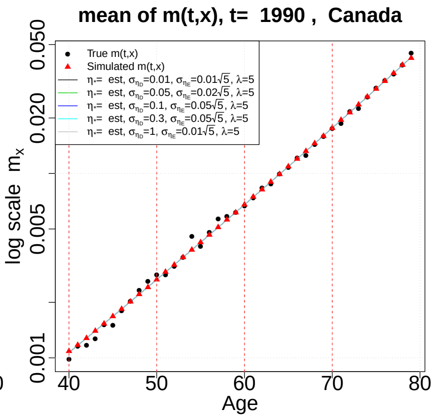


HERIOT WATT UNIVERSITY

Actuarial Research Centre
Institute and Faculty of Actuaries

Fitted exposures $E_{t,y}$, Mexico & Canada 1990.





Conclusions.

- Smooth time series $\underline{a}$, $\underline{b}$ and $\underline{c}$ $\xrightarrow{\text{Reduce age heaping}}$ $m(t, y)$ and population. However, we do not want to smooth too much because it would destroy the natural volatility from the data.
- This model improves the quality of the Mexican data by reducing age heaping across all calendar years.
- The remaining volatility in the fitted exposures comes from the death counts.

Conclusions

- Parameter η^* reflects the improvement in the quality of the data over time. Hence, we expect η^* to increase over time. In other words, there is less age heaping that there use to be, say in 1990.

Forthcoming research

- Include priors for parameters regarding the true force of mortality.
- Sensitivity of choosing different priors.
- We will keep collaborating with HMD to see how their approach can be adapted to Mexican data for producing complete life table series, which is also relevant to international reinsurance.

Thank You!

Questions?