Inference and forecasting in the age-period-cohort model with unknown exposure with an application to Mesothelioma mortality

M.M. Martínez Miranda, B. Nielsen & J.P. Nielsen

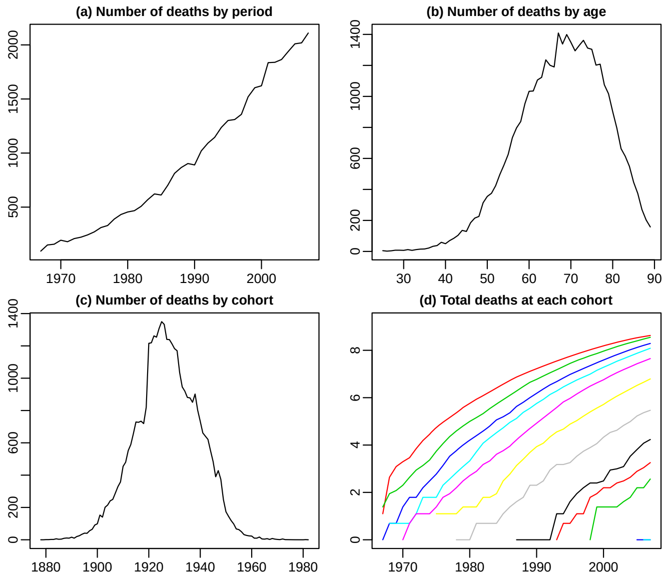
To appear in JRSS-A
9 April 2014

Aim: forecast Mesothelioma mortality.

Methodological issues:

- Data: counts of deaths, but no exposure.
- Model number of deaths not mortality ratios
- Age-period-cohort model has identification problem
- Apply time series methods for forecast

Data



Data: counts of deaths in GB by age for 1967–2007.

Model: Distribution

Dose-Response Model

Y_{ij} counts of deaths, Z_{ij} exposure, by age i and period j.
Apply Poisson model with

$$\mathsf{E}(Y_{ij}|Z_{ij}) = \exp(\nu_{ij})Z_{ij} = \exp(\nu_{ij} + \log Z_{ij})$$

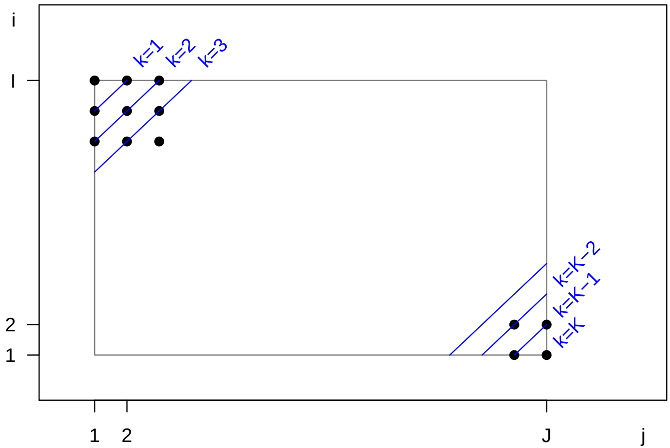
Modelling without exposure

Y_{ij} counts of deaths, by age i,j.
Apply Poisson model with

$$\mathsf{E}(Y_{ij}) = \exp(\mu_{ij})$$

In Generalized Linear Model literature μ_{ij} is called *predictor*.

Model: Indices



$i = 1, \ldots, I$ is age
$j = 1, \ldots, J$ is period
$k = 1, \ldots, K$ is cohort
where $k = I - i + J$ and $K = I + J - 1$.

Model: Age-Period-Cohort predictor

$$\mu_{ij} = \alpha_i + \beta_j + \gamma_k + \delta$$

Identification problem:

$$\mu_{ij} = (\alpha_i + a + di) + (\beta_j + b - dj) + (\gamma_k + c + dk) + (\delta - a - b - c - dI)$$

Traditional solution:

ad hoc identify by allocating **2** linear trends to **3** time effects.

Parsimonenous parametrisation:

$$\begin{aligned}
\mu_{ij} &= \mu_{I1} + (i - I)(\mu_{I1} - \mu_{I-1,1}) + (j - 1)(\mu_{I2} - \mu_{I1}) \\
&\quad + \sum_{t=i}^{I-2} \sum_{s=t}^{I-2} \Delta^2 \alpha_{s+2} + \sum_{t=3}^{j} \sum_{s=3}^{t} \Delta^2 \beta_s + \sum_{t=3}^{k} \sum_{s=3}^{t} \Delta^2 \gamma_s \\
&= X'_{ij}\xi \qquad \text{so } X'X \text{ invertible and } \xi \text{ freely varying}
\end{aligned}$$

Analysis of model

With $\mu_{ij} = X_{ij}'\xi$ the likelihood is

$$\ell(\xi; Y) = \sum_{i,j} Y_{ij} X_{ij}'\xi - \sum_{i,j} \exp(X_{ij}'\xi).$$

ξ varies freely so model is *regular exponential family*.

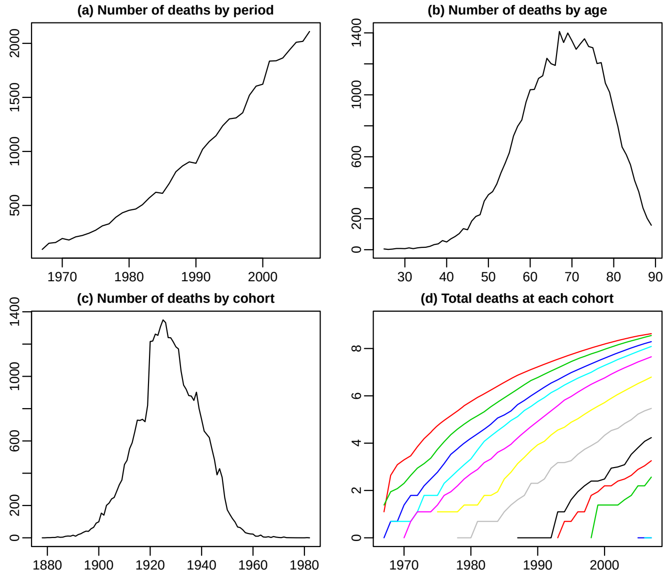
Multinomial sampling

Let $\tau = \sum_{ij} \exp(X_{ij}'\xi)$, $\quad Y_\bullet = \sum_{ij} Y_{ij}$, $\quad \pi_{ij} = \exp(X_{ij}'\xi)/\tau$.
Then likelihood satisfies

$$\ell(\xi; Y) = \underbrace{Y_\bullet \log \tau - \tau}_{Poisson} + \underbrace{\sum_{i,j} Y_{ij} \log \pi_{ij}}_{Multinomial}.$$

Conduct inference about π_{ij} in
conditional multinomial distribution of Y_{ij} given $Y_\bullet$

Data

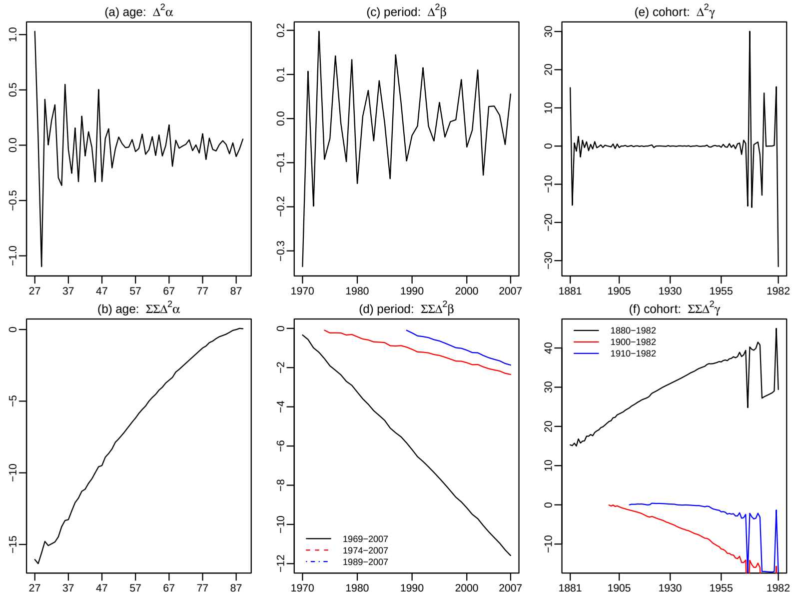


Data: counts of deaths in GB by age for 1967–2007.

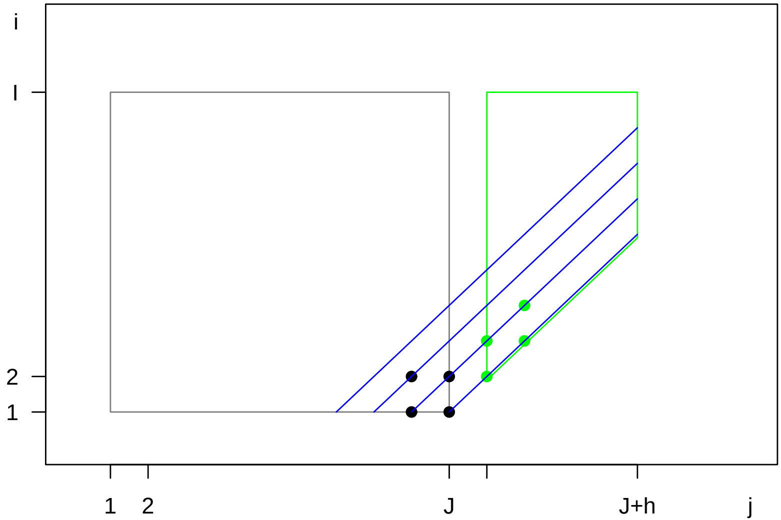
Deviances: APC or AC?

Model	Deviance	df	p
APC	2384.9	2457	0.852
AC vs. APC	56.8	39	0.033

Plots of Estimates



Forecast region



With $\mu_{ij} = \alpha_i + \beta_j + \gamma_k + \delta$
need to extrapolate β_j.

Extrapolation

With ad hoc identified parameters

forecasts should not depend on ad hoc identification.

This holds *iff* extrapolation is linear trend preserving.

With parsimoneous parameter this is not constraint.

Extrapolation method: Let

$$x_j = \sum_{t=3}^{j} \sum_{s=3}^{t} \Delta^2 \beta_s \quad \text{for } j = 3, \ldots, J.$$

Apply time series model $x_j = \nu_c + \nu_\ell j + \varepsilon_j$

Extrapolation

$$\widetilde{x}_{J+h} = \widehat{\nu}_c + \widehat{\nu}_\ell (J + h) + \underbrace{\widehat{\varepsilon}_J}_{\substack{intercept \\ correction}} = x_J + \widehat{\nu}_\ell h.$$

Forecast

Point forecast

$$\widetilde{\mu}_{i,J+h} = \widehat{\mu}_{I1} + (i - I)(\widehat{\mu}_{I1} - \widehat{\mu}_{I-1,1}) + (J + h - 1)(\widehat{\mu}_{I2} - \widehat{\mu}_{I1})$$
$$+ \sum_{t=i}^{I-2} \sum_{s=t}^{I-2} \Delta^2 \widehat{\alpha}_{s+2} + \sum_{t=3}^{J} \sum_{s=3}^{t} \Delta^2 \widehat{\beta}_s + \sum_{t=3}^{k} \sum_{s=3}^{t} \Delta^2 \widehat{\gamma}_s + \widehat{\nu}_{\ell} h$$

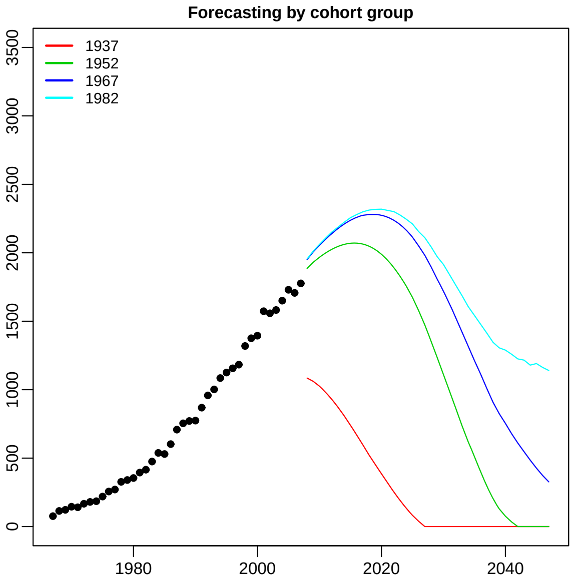
Distribution forecast

$$\widetilde{Y}_{i,J+h}^{distribution} = \exp(\widetilde{\mu}_{i,J+h}) + \mathsf{N}(0, s_{i,J+h}^2)$$

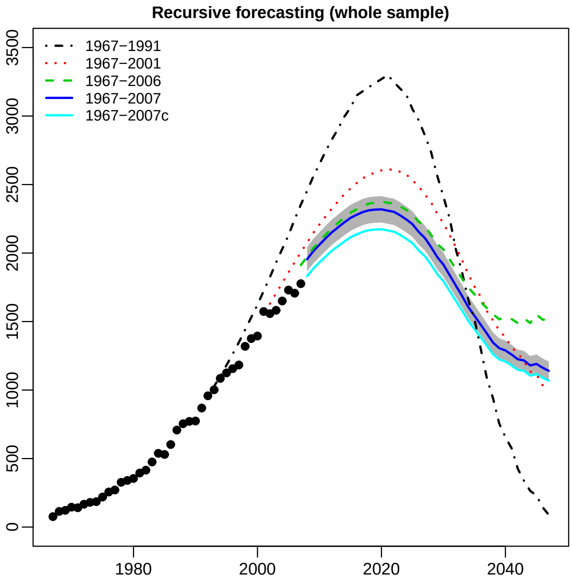
where

$$s_{i,J+h}^2 = s_{extrapolation\,error}^2 + \underbrace{s_{estimation\,error}^2}_{\text{negligible for near future}} .$$

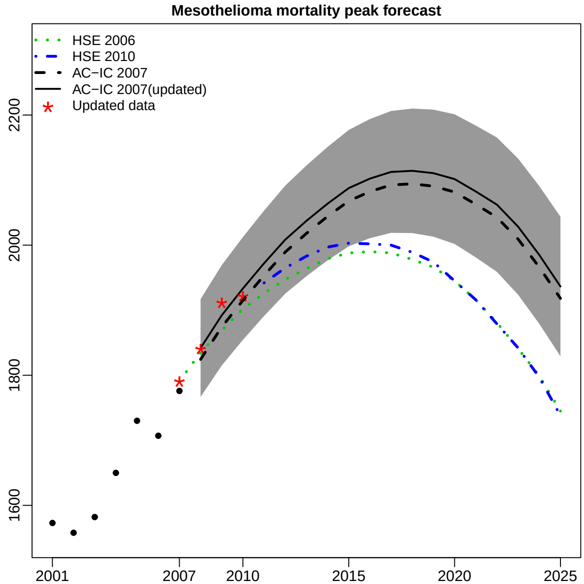
Forecast: Cohort components



Forecast: Recursive



Forecast: Recent data



References

Exponential families

Barndorff-Nielsen ('78)

McCullagh & Nelder ('89)

Age-period-cohort models

Clayton & Schifflers ('87) *Statist. Medicine*

Kuang, Nielsen, Nielsen ('08a,b) *Biometrika*

Mesothelioma

Peto, Matthews, Hodgson, Jones ('95) *Lancet*

Hodgson, McElvenny, Darnton, Price, Peto ('05) *Br. J. Cancer*

Tan & Warren ('09) Health & Safety Executive