Estimation and extrapolation of time trends in multivariate registry data using Bayesian age-period-cohort models

Andrea Riebler

Joint work with Leonhard Held and Håvard Rue

Research Seminar, Vienna, November 23, 2012
Registry data

- National health care registers typically collect incidence and/or mortality counts **stratified by age and period.**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>50–54</td>
<td>No. Deaths</td>
<td>177</td>
<td>271</td>
<td>312</td>
<td>382</td>
<td>321</td>
<td>305</td>
<td>308</td>
<td>304</td>
<td>274</td>
<td>278</td>
</tr>
<tr>
<td></td>
<td>No. at Risk</td>
<td>301000</td>
<td>317000</td>
<td>353000</td>
<td>395000</td>
<td>426000</td>
<td>473000</td>
<td>498000</td>
<td>552000</td>
<td>598000</td>
<td>629000</td>
</tr>
<tr>
<td>55–59</td>
<td>Cohort</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>No. Deaths</td>
<td>262</td>
<td>350</td>
<td>552</td>
<td>620</td>
<td>714</td>
<td>649</td>
<td>738</td>
<td>718</td>
<td>780</td>
<td>759</td>
</tr>
<tr>
<td></td>
<td>No. at Risk</td>
<td>212000</td>
<td>248000</td>
<td>279000</td>
<td>301000</td>
<td>358000</td>
<td>411000</td>
<td>443000</td>
<td>435000</td>
<td>510000</td>
<td>583000</td>
</tr>
<tr>
<td>60–64</td>
<td>Cohort</td>
<td>6</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>No. Deaths</td>
<td>360</td>
<td>479</td>
<td>644</td>
<td>949</td>
<td>932</td>
<td>1292</td>
<td>1327</td>
<td>1507</td>
<td>1602</td>
<td>1712</td>
</tr>
<tr>
<td></td>
<td>No. at Risk</td>
<td>159000</td>
<td>194000</td>
<td>222000</td>
<td>222000</td>
<td>258000</td>
<td>304000</td>
<td>341000</td>
<td>404000</td>
<td>403000</td>
<td>482000</td>
</tr>
<tr>
<td>65–69</td>
<td>Cohort</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>No. Deaths</td>
<td>409</td>
<td>544</td>
<td>812</td>
<td>1150</td>
<td>1668</td>
<td>1958</td>
<td>2153</td>
<td>2375</td>
<td>2742</td>
<td>2973</td>
</tr>
<tr>
<td></td>
<td>No. at Risk</td>
<td>132000</td>
<td>144000</td>
<td>169000</td>
<td>210000</td>
<td>230000</td>
<td>264000</td>
<td>297000</td>
<td>322000</td>
<td>396000</td>
<td>401000</td>
</tr>
<tr>
<td>70–74</td>
<td>Cohort</td>
<td>4</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>No. Deaths</td>
<td>328</td>
<td>509</td>
<td>763</td>
<td>1097</td>
<td>1593</td>
<td>2039</td>
<td>2433</td>
<td>3066</td>
<td>3432</td>
<td>3939</td>
</tr>
<tr>
<td></td>
<td>No. at Risk</td>
<td>76000</td>
<td>94000</td>
<td>110000</td>
<td>125000</td>
<td>149000</td>
<td>180000</td>
<td>197000</td>
<td>213000</td>
<td>233000</td>
<td>293000</td>
</tr>
<tr>
<td>75–79</td>
<td>Cohort</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>No. Deaths</td>
<td>222</td>
<td>359</td>
<td>584</td>
<td>845</td>
<td>1192</td>
<td>1638</td>
<td>2068</td>
<td>2671</td>
<td>3356</td>
<td>3928</td>
</tr>
<tr>
<td></td>
<td>No. at Risk</td>
<td>37000</td>
<td>47000</td>
<td>59000</td>
<td>71000</td>
<td>91000</td>
<td>108000</td>
<td>118000</td>
<td>132000</td>
<td>141000</td>
<td>193000</td>
</tr>
<tr>
<td>80–84</td>
<td>Cohort</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>No. Deaths</td>
<td>108</td>
<td>178</td>
<td>285</td>
<td>475</td>
<td>742</td>
<td>992</td>
<td>1374</td>
<td>1833</td>
<td>2353</td>
<td>3184</td>
</tr>
<tr>
<td></td>
<td>No. at Risk</td>
<td>19000</td>
<td>22000</td>
<td>32000</td>
<td>39000</td>
<td>44000</td>
<td>56000</td>
<td>66000</td>
<td>77000</td>
<td>93000</td>
<td>94000</td>
</tr>
</tbody>
</table>
Age-period-cohort analysis

- **Age-period-cohort (APC) model** is commonly used to describe vital rates using three time scales:
  - **Age**: age at diagnosis/death
  - **Period**: date of diagnosis/death
  - **Cohort**: date of birth

- **Goals**:
  - Detecting temporal patterns in such data as they could provide important clues for disease etiology.
  - Extrapolation and **prediction**.
The univariate age-period-cohort model

\( y_{ij} \): Number of cases in age group \( i \) at period \( j \)

\( n_{ij} \): Number of persons at risk in age group \( i \) at period \( j \)

\[ y_{ij} \mid \eta_{ij} \sim \text{Poisson}(n_{ij} \exp(\eta_{ij})) \]

\[ \eta_{ij} = \mu + \alpha_i + \beta_j + \gamma_k \]

with age effect \( \alpha_i \), period effect \( \beta_j \) and cohort effect \( \gamma_k \).
The univariate age-period-cohort model

\( y_{ij} \): Number of cases in age group \( i \) at period \( j \)

\( n_{ij} \): Number of persons at risk in age group \( i \) at period \( j \)

\[
y_{ij} | \eta_{ij} \sim \text{Poisson}(n_{ij} \exp(\eta_{ij}))
\]

\[
\eta_{ij} = \mu + \alpha_i + \beta_j + \gamma_k
\]

with age effect \( \alpha_i \), period effect \( \beta_j \) and cohort effect \( \gamma_k \).

- The cohort index \( k = M \times (I - i) + j \) is a linear function of \( i \) and \( j \), where \( M \) is the ratio of age group to period interval.
- To assure identifiability of the intercept \( \mu \) we apply sum-to-zero constraints for each parameter vector \( \alpha, \beta, \gamma \).
Identifiability problems

Time trends are not identifiable due to the linear dependence between age, period and cohort indices.

For any value of $a \in \mathbb{R}$, the linear transformations

$$\alpha_i \rightarrow \alpha_i + M \times a \left( i - \frac{(I + 1)}{2} \right),$$

$$\beta_j \rightarrow \beta_j - a \left( j - \frac{(J + 1)}{2} \right),$$

$$\gamma_k \rightarrow \gamma_k + a \left( k - \frac{(K + 1)}{2} \right)$$

will leave $\eta_{ij}$ unchanged and maintain the sum-to-zero constraints.

Holford, 1983, Biometrics
Identifiability problems: Illustration (M=1)

Note: Second differences are identifiable, but hard to interpret.
Identifiability problems: Illustration (M=1)

Note: Second differences are identifiable, but hard to interpret.
For unequally spaced data, a second identifiability problem induces artificial cyclical patterns (saw-tooth-pattern) in the period and cohort estimates. 

Remember, $M$ is the ratio of age group to period interval length.

For any value of $b_1, \ldots, b_M \in \mathbb{R}$ (subject to $b_1 + \ldots + b_M = 0$), the transformations

$$
\beta_j = \beta_j + b_1 + (j-1) \mod M
$$

$$
\gamma_k = \gamma_k - b_1 + (k-1) \mod M
$$

will leave the linear predictor $\eta_{ij}$ unchanged.
Identifiability problems: Illustration (M=5)

Age effects

Period effects

Cohort effects
Likelihood Inference

Mainly classical maximum likelihood (ML) estimation has been used for APC models.

Disadvantages:

- Additional constraints are necessary for identifiability.
- The model overfits cohorts for which only a single observation exists.
- For unequally spaced data, ML estimates become very unstable resulting in artificial saw-tooth pattern.

⇒ We go for a Bayesian approach.
The Bayesian APC model

- A natural choice is to penalise second differences.

- The second order random walk (RW2) prior for $\alpha$ is

$$
\pi(\alpha | \tau_\alpha) \propto \tau^\frac{l-2}{2} \exp \left( -\frac{\tau_\alpha}{2} \sum_{i=3}^{l} (\alpha_i - 2\alpha_{i-1} + \alpha_{i-2})^2 \right) = \tau^\frac{l-2}{2} \exp \left( -\frac{1}{2} \alpha^\top R \alpha \right),
$$

where $R$ is the precision matrix:

$$
R = \tau_\alpha \begin{pmatrix}
1 & -2 & 1 & \\
-2 & 5 & -4 & 1 \\
1 & -4 & 6 & -4 & 1 \\
& & & & & \\
& & & & & \\
& & & & & \\
1 & -4 & 6 & -4 & 1 \\
1 & -4 & 5 & -2 \\
1 & -2 & 1
\end{pmatrix}.
$$

⇒ Latent Gaussian Markov random field (GMRF) model.
The Bayesian APC model (II)

- Use independent RW2 priors for $\alpha, \beta, \gamma$.
- To account for overdispersion add $z_{ij} \sim \mathcal{N}(0, \tau_z^{-1})$ to $\eta_{ij}$:

$$\eta_{ij} = \mu + \alpha_i + \beta_j + \gamma_k + z_{ij}.$$  

- All precision parameters are treated as unknown and suitable gamma-hyperpriors are assigned.
Inference in the Bayesian APC model

- Besag et al. (1995) propose a sophisticated MCMC algorithm using suitable reparameterisation and block sampling.

- However, they also note:
  
  "We anticipate that analytical approximations should work well on our model and on others similar to it, especially for the present data where there appears not to be any significant multimodality in the posterior distribution."

- Today, the INLA methodology can be used for routine application using an R interface (www.r-inla.org).
Integrated nested Laplace approximations (INLA)
(Rue et al, 2009, JRSS-B)

INLA is a fast alternative to inference via MCMC in latent Gaussian models. The methodology is particularly attractive if the latent Gaussian model is a GMRF.

The INLA approach

- incorporates posterior uncertainty with respect to hyperparameters,
- can be used for out-of-sample prediction,
- can be used for model assessment and comparison based on leave-one-out cross-validation.

INLA can be called in a modular way, just as glm() or lme(), say, using an R interface (www.r-inla.org).
The general setting

Three-stage Bayesian hierarchical model:

- **Observation model:** \( \pi(y|x) = \prod_u \pi(y_u|x_u, \theta) \).
- **Parameter model:** \( \pi(x|\theta) \), usually a GMRF.
- **Hyperprior:** \( \pi(\theta) \).
The general setting

Three-stage Bayesian hierarchical model:

- Observation model: \( \pi(y|x) = \prod_u \pi(y_u|x_u, \theta) \).
- Parameter model: \( \pi(x|\theta) \), usually a GMRF.
- Hyperprior: \( \pi(\theta) \).

The posterior distribution is

\[
\pi(x, \theta | y) \propto \pi(\theta) \pi(x|\theta) \prod_u \pi(y_u|x_u, \theta).
\]

\( \text{Dim}(x) \) is large, while \( \text{dim}(\theta) \) is small.
INLA: non-Gaussian observations

Main goal: Compute the posterior marginals

\[
\pi(x_u|y) = \int_\theta \int_{x_{-u}} \pi(x, \theta|y) dx_{-u} \ d\theta,
\]

\[
\pi(x_u, \theta|y) = \pi(x_u|\theta, y) \pi(\theta|y)
\]

\[
\pi(\theta_v|y) = \int_{\theta_{-v}} \int_x \pi(x, \theta|y) dx \ d\theta_{-v}.
\]

INLA uses nested Laplace approximations for this purpose.

In our model: \( x = (\mu, \alpha^T, \beta^T, \gamma^T, z^T)^T, \theta = (\tau_\alpha, \tau_\beta, \tau_\gamma, \tau_z)^T. \)
Using INLA

```r
> head(ProstateCancer, 4)
   deaths  pop age.group period cohort index
 1    177 301000       1     1     7     1
 2    262 212000       2     1     6     2
 3    360 159000       3     1     5     3
 4    409 132000       4     1     4     4

> library(INLA)
> my.hyper <- list(prec=list(param=c(1, 0.005)))
> model <- deaths ~ f(age.group, model="rw2", hyper=my.hyper) +
            f(period, model="rw2", hyper=my.hyper) +
            f(cohort, model="rw2", hyper=my.hyper) +
            f(index, model="iid", hyper=my.hyper)
> results <- inla(model, family="poisson", data=ProstateCancer,
                  E=pop, quantiles=c(0.1, 0.5, 0.9))
```
Comparing INLA with MCMC

Age effects

Period effects

Cohort effects

Overdispersion
Predictions in INLA

- Prediction of future deaths rates was one of the major goals in Besag et al. (1995).
- This can be also done in INLA by setting the observations to be predicted to NA.
- Post-processing of the posterior predictive distribution of the linear predictor $\eta_{ij}$ gives the predictive distribution of $y_{ij}$.
- Even simultaneous credible bands can be computed.

Sørbye and Rue, 2010
Prediction: Prostate cancer


Observed and predicted number of cases within 80% point-wise credible bands.
Multivariate APC models

- Data now also depend on strata $r = 1, \ldots, R$.
- Most general formulation (apc model):

$$y_{ijr} | \eta_{ijr} \sim \text{Poisson}(n_{ijr} \exp(\eta_{ijr}))$$

$$\eta_{ijr} = \mu_r + \alpha_{i,r} + \beta_{j,r} + \gamma_{k,r} + z_{ijr}$$

with independent $z_{ijr} \sim \mathcal{N}(0, \kappa_z^{-1})$, say.
Multivariate APC models

- Data now also depend on strata $r = 1, \ldots, R$.
- Most general formulation (apc model):
  \[ y_{ijr} \mid \eta_{ijr} \sim \text{Poisson}(n_{ijr} \exp(\eta_{ijr})) \]
  \[ \eta_{ijr} = \mu_r + \alpha_{i,r} + \beta_{j,r} + \gamma_{k,r} + z_{ijr} \]
  with independent $z_{ijr} \sim \mathcal{N}(0, \kappa_z^{-1})$, say.
- Simpler models can be obtained, e.g. assuming shared age effects (Apc model):
  \[ y_{ijr} \mid \eta_{ijr} \sim \text{Poisson}(n_{ijr} \exp(\eta_{ijr})) \]
  \[ \eta_{ijr} = \mu_r + \alpha_i + \beta_{j,r} + \gamma_{k,r} + z_{ijr} \]
- As a start: independent RW2 priors for $\alpha, \beta_r, \gamma_r, r = 1, \ldots, R$. 
Identifiability of relative risks

- The multivariate APC model inherits all identifiability problems from the univariate APC model.
- However, differences
  \[
  \Delta_j^{(r)} = \beta_{j,r} - \beta_{j,R} \quad \text{in the ApC model}
  \]
  \[
  \Delta_k^{(r)} = \gamma_{k,r} - \gamma_{k,R} \quad \text{in the APc model}
  \]
  \[
  \Delta_{jk}^{(r)} = \Delta_j + \Delta_k \quad \text{in the Apc model}
  \]
  are identifiable.
- Let \( \Delta_{\mu}^{(r)} = \mu_r - \mu_R \). The adjusted differences
  \[
  \tilde{\Delta}_j^{(r)} = \Delta_{\mu}^{(r)} + \Delta_j^{(r)}
  \]
  \[
  \tilde{\Delta}_k^{(r)} = \Delta_{\mu}^{(r)} + \Delta_k^{(r)}
  \]
  can be interpreted as (average) log relative risk.
Analysing heterogeneous time trends: Apc model

Data: COPD mortality counts among males in England & Wales

Hansell et al., 2003, Epidemiology

- $I = 7$ age groups:
  - 15–24, 25–34, ..., 75+
- $J = 50$ periods:
  - 1950–1999
- $K = 110$ birth cohorts.
- $R = 3$ regions
  - Greater London
  - Other conurbations
  - Rural areas

Riebler & Held, 2010, Biostatistics.
Analysing heterogeneous time trends: Apc model

Data: COPD mortality counts among males in England & Wales

Hansell et al., 2003, Epidemiology

- $I = 7$ age groups: 15–24, 25–34, . . . , 75+.
- $K = 110$ birth cohorts.
- $R = 3$ regions
  - Greater London
  - Other conurbations
  - Rural areas

Riebler & Held, 2010, Biostatistics.
A side comment: A conditional approach

Let \( y_{ij} = y_{ij_1} + \ldots + y_{ij_R} \).

It is easy to see that the Apc model for \( y_{ijr} \) implies that \( y_{ij} \mid y_{ij} \) is multinomial with individual success probability

\[
\pi_{ijr} = \frac{\exp \left( \log \left( \frac{n_{ijr}}{n_{ijR}} \right) + \Delta_j^{(r)} + \Delta_k^{(r)} \right)}{1 + \sum_{s=1}^{R-1} \exp \left( \log \left( \frac{n_{ij}s}{n_{ijR}} \right) + \Delta_j^{(s)} + \Delta_k^{(s)} + \Delta_j^{(s)} \right)}.
\]

Note that through conditioning, the original parameters are replaced by the differences \( \Delta_j^{(r)} \) and \( \Delta_k^{(r)} \).

Age effects are no longer present in this formulation.

All parameters are identifiable and can be estimated with ML with suitable smoothing, if necessary (R-package VGAM).

Correlate separate random walks

When assuming separate time effects across strata it might nevertheless be plausible to assume some correlation.

⇒ Use of correlated random walks

Illustration for $R = 3$: 

![Diagram showing correlated random walks with $R = 3$.]
Correlate separate random walks

When assuming separate time effects across strata it might nevertheless be plausible to assume some correlation.

⇒ Use of correlated random walks

Illustration for $R = 3$: 

![Graph showing correlated random walks](image-url)
Correlate separate random walks

When assuming separate time effects across strata it might nevertheless be plausible to assume some correlation.

⇒ Use of correlated random walks

Illustration for $R = 3$: 

[Diagram of interconnected nodes and arrows representing the correlation structure]
Correlate separate random walks

When assuming separate time effects across strata it might nevertheless be plausible to assume some correlation.

⇒ Use of correlated random walks

Illustration for $R = 3$: 
Correlate separate random walks

When assuming separate time effects across strata it might nevertheless be plausible to assume some correlation.

⇒ Use of correlated random walks

Illustration for $R = 3$: 
Correlated GMRF priors

- For simplicity: $R = 3$.
- Let $\mathbf{C} = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}$ denote a uniform correlation matrix.

The random walks $\beta_1, \beta_2, \beta_3$ can be correlated using the stacked vector $\tilde{\beta} = (\beta_1^T, \beta_2^T, \beta_3^T)^T$:

$$
 f(\tilde{\beta}|\mathbf{C}_\beta, \tau_\beta) \propto \tau_\beta |\tau_\beta \mathbf{C}_\beta^{-1}|^{(J-2)/2} \exp \left( -\frac{1}{2} \tilde{\beta}^T \left\{ \mathbf{C}_\beta^{-1} \otimes \mathbf{R} \right\} \tilde{\beta} \right).
$$

- Multivariate RW2 with correlated increments.
- Correlated overdispersion can also be incorporated:

$$
 \mathbf{z}_{ij} = (z_{ij1}, z_{ij2}, z_{ij3})^T \sim \mathcal{N}(\mathbf{0}, \tau_z^{-1} \mathbf{C}_z).
$$

- All correlations are treated as unknown.
Prior on correlation parameters

Reparameterise $\rho$ using the general Fisher’s z-transformation:

$$
\rho = \frac{\exp(\rho^*) - 1}{\exp(\rho^*) + R - 1} \quad \quad \rho^* = \log \left( \frac{1 + \rho \cdot (R - 1)}{1 - \rho} \right),
$$

Fisher, 1958, page 219

and assign a $\mathcal{N}(0, \tau^{-1})$ prior to $\rho^*$.

- This prior automatically ensures that $\rho \in (-1/(R - 1), 1)$, which is required to ensure positive definiteness of $\mathbf{C}$, is fulfilled.

- In addition, $P(\rho > 0) = 0.5$, independent of $R$. 
Illustration of prior on correlation parameters
INLA call in \texttt{R-INLA} package

\begin{verbatim}
> library(INLA)
> ## define the grouping index
> g <- rep(c(1,2,3), each=I*J)
> ## Apc model with correlated time effects & overdispersion.
> my.hyper.rho <- list(rho=list(param=c(0, 0.2)))
> model_Apc <- y ~ -1 + mu1 + mu2 + mu3 +
>   f(age.group, model="rw2", hyper=my.hyper) +
>   f(period, model="rw2", hyper=my.hyper, constr=TRUE, rankdef=2,
>     control.group=list(hyper=my.hyper.rho, model="exchangeable")),
>   group=g) +
>   f(cohort, model="rw2", ...) +
>   f(index, model="iid", hyper=my.hyper,
>     control.group=list(hyper=my.hyper.rho, model="exchangeable"),
>     group=g)

> results <- inla(model_Apc, family="poisson", E=n, data=data,
>                   control.compute=list(hyperpar =TRUE))
\end{verbatim}
Relative risks accounting for correlation

Wide credible intervals for younger birth cohorts.

Adjusting for correlation improves precision of the relative risk estimates!
Imputation/extrapolation of rates

Imputation of missing data for one stratum by taking advantage of corresponding observations in other strata.

- In particular interesting for short term projections, historic data or to adjust for varying collection periods.
- We are able to consider the most flexible \textit{apc} model.
Mortality extrapolation

Data: Overall mortality counts among females stratified by

- $R = 3$ regions: Denmark, Sweden and Norway.
- $I = 17$ age groups: $0 - 4, 5 - 9, \ldots, 75 - 79, 80 - 84$.
- $J = 20$ periods from 1900 – 1999 for Denmark and Sweden (only 10 periods for Norway)

\begin{figure}
\centering
\begin{tikzpicture}
\fill[red!50] (0,0) rectangle (6,2); \node at (3,1) {Denmark};
\fill[blue!50] (7,0) rectangle (13,2); \node at (10,1) {Sweden};
\fill[red!50] (14,0) rectangle (20,2); \node at (17,1) {Norway};
\draw (0,0) -- (0,-1) -- (20,-1) -- (20,0);
\foreach \x in {0,1,2,...,20} \draw (\x,-1.2) -- (\x,-1.7); \foreach \y in {0,1,2} \draw (-0.2,\y) -- (0.2,\y);
\node at (0.5,-0.5) {1900}; \node at (1.5,-0.5) {1910}; \node at (2.5,-0.5) {1920}; \node at (3.5,-0.5) {1930}; \node at (4.5,-0.5) {1940}; \node at (5.5,-0.5) {1950}; \node at (6.5,-0.5) {1960}; \node at (7.5,-0.5) {1970}; \node at (8.5,-0.5) {1980}; \node at (9.5,-0.5) {1990}; \node at (10.5,-0.5) {2000}; \node at (11.5,-0.5) {Periods};
\end{tikzpicture}
\end{figure}
Mortality extrapolation II

- Projection for Norwegian women for 1900–1949 by borrowing strength of full mortality rates of Sweden and Denmark.

- Comparison to a univariate APC model and an extended Lee-Carter demographic forecasting approach.

- The correlated model performs better in terms of mean squared error, coverage probabilities and proper scoring rules.
Summary

- Applying APC models you need to be aware of identifiability problems.
- (Multivariate) APC models are well suited to estimate and project time trends in registry data.
- A Bayesian APC analysis can easily be done using INLA. No MCMC necessary.
Acknowledgements

Leonhard Held & the Biostatistics group, UZH.

Håvard Rue & the Department of Mathematical Sciences, NTNU.

Financial support by the “Swiss National Science Foundation (SNSF)” and the “Research Council of Norway” is gratefully acknowledged.

Thank you for your attention!
INLA: non-Gaussian observations

Main goal: Compute the posterior marginals

\[
\pi(x_i | y) = \int_\theta \int_{x_{-i}} \pi(x, \theta | y) \, dx_{-i} \, d\theta,
\]

\[
\pi(x_i, \theta | y) = \pi(x_i | \theta, y) \pi(\theta | y)
\]

\[
\pi(\theta_j | y) = \int_{\theta_{-j}} \int_x \pi(x, \theta | y) \, dx \, d\theta_{-j}.
\]

From \( \pi(x, \theta, y) = \pi(x | \theta, y) \times \pi(\theta | y) \times \pi(y) \) it follows that:

\[
\pi(\theta | y) \propto \frac{\pi(x, \theta, y)}{\pi(x | \theta, y)} \approx \frac{\pi(x, \theta, y)}{\tilde{\pi}_G(x | \theta, y)} \propto \tilde{\pi}(\theta | y) \quad \text{(Laplace approximation)}
\]

with \( \tilde{\pi}_G(x | \theta, y) \) Gaussian approximation to \( \pi(x | \theta, y) \).
INLA

Step I  Build a Laplace approximation to

\[ \pi(\theta|y) \propto \frac{\pi(x, \theta, y)}{\pi(x|\theta, y)} \approx \frac{\pi(x, \theta, y)}{\tilde{\pi}_G(x|\theta, y)} \propto \tilde{\pi}(\theta|y) \]

and “explore it numerically” to obtain good support points \( \theta_k \).

Step II  Approximate \( \pi(x_i|y, \theta_k) \) for each \( \theta_k \).

Step III  For each \( i \), sum out \( \theta_k \)

\[ \tilde{\pi}(x_i|y) = \sum_k \tilde{\pi}(x_i|\theta_k, y) \times \tilde{\pi}(\theta_k|y) \times \Delta_k. \]

Step IV  Approximate \( \pi(\theta_k|y) \) using numerical integration.
Posterior correlation in Apc model

Correlation: period

Correlation: cohort

Correlation: overdispersion

Andrea Riebler (University of Zurich)