Example: Stomach cancer in Slovenia

Data: Slovenia is partitioned into 194 municipalities

- For each municipality, we have (years 1995-2001)
  - observed count of stomach cancer cases ($O_i$)
  - expected # cases ($E_i$), using indirect standardization

- A measure of socioeconomic status, centered and scaled ($SEC_i$)

Reich et al. (2006)

Model: $O_i \sim \text{Poisson} (E_i e^{M_i})$

$M_i = \beta_{SEC_i} \cdot SEC_i + S_i + H_i \sim \text{iid } \mathcal{N}(0, \frac{1}{C_h})$

$CAR$, precision $T_3$; "neighbors" = share a boundary

<table>
<thead>
<tr>
<th>Model</th>
<th>DIC</th>
<th>Po</th>
<th>$\beta_{SEC}$ post% 95% interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>no CAR, no hetero</td>
<td>1153</td>
<td>2.0</td>
<td>(-0.175, -0.098)</td>
</tr>
<tr>
<td>CAR, hetero</td>
<td>1081</td>
<td>62.3</td>
<td>(-0.100, 0.057)</td>
</tr>
</tbody>
</table>

With 62.3 DF in the fit (out of 194), the fit must be pretty close to the data

VAM T2 add 1  8/1/08
Here's a picture of the data:

\[ \text{Oil} \]

\[ \text{Sec} \]
Penalized spline model (Nick Saltkowske class project):

\[ O_i \sim \text{Poisson} \left( E_i \right) \]

\[ \mu_i = \beta_0 + \beta_{SE} SE_i + \beta_{X1c} X1c_i + \beta_{X2c} X2c_i \]

- Centroid of municipality \( i \)
- \( E \times W \), centered
- \( N \times S \), centered

\[ + \sum_{k=1}^{16} \lambda_k \| x - K_k \|^2 \log \| x - K_k \| \]

Result (using SemiPar package):

\[ f(X1c, X2c) \quad \text{df} \quad \text{knots} \]

\[ 2 \quad 48 \quad \text{i.e. the } \lambda_k \text{ are all smoothed to zero!} \]

Here's the fit with 2 df in the spline, and one forced to have 26 df in the spline (8 50 max)
Figure 8: Plots of the bivariate smoothing function with 2,001 and 26 degrees of freedom.
If you force the Ul to have about 3 df collectively about 7
of the 48 knots get "large" Ul.

Figure 9: Ordered absolute estimated random effects with 4.952 degrees of freedom (top). Location of the knots associated with the 7 largest absolute estimated random effects.