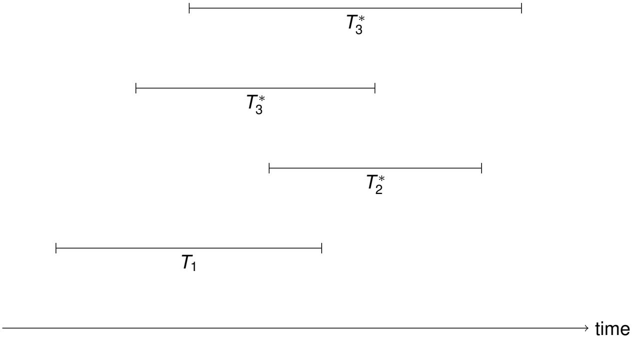
Regularized Hazard Estimation for Age-Period-Cohort Analysis

V. Goepp[†], G. Nuel[‡], O. Bouaziz[†]

†: MAP5 (CNRS 8145), Université Paris Descartes
‡: LPSM (CNRS 8001), Sorbonne Université

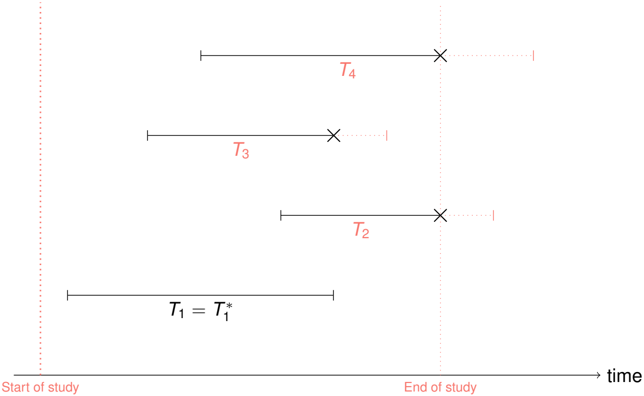
IWAP 2018, Budapest

Introduction: right-censored data



- We want to infer T^*, time before an event of interest.

Introduction: right-censored data



- We want to infer T^*, time before an event of interest.
- But we don't observe the T_i^*s, but

$$T_i = \min\left(T_i^*, C_i\right).$$

Introduction: survival analysis

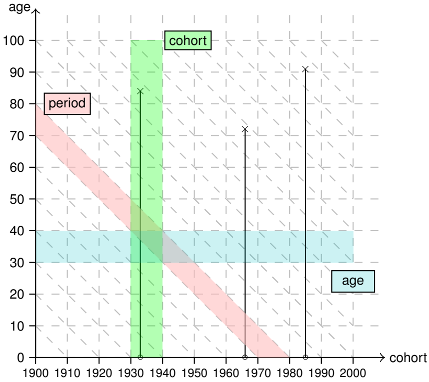
Framework of survival analysis:

- C is the *censoring* variable.
- We also observe $\Delta_i = \mathbb{1}_{T_i^* = T_i}$.
- We infer the conditional density called *hazard rate*:

$$\lambda(t) = \lim_{\delta t \to 0} \frac{\mathbb{P}\left(t \leq T^* \leq t + \delta t | T^* > t\right)}{\delta t}$$

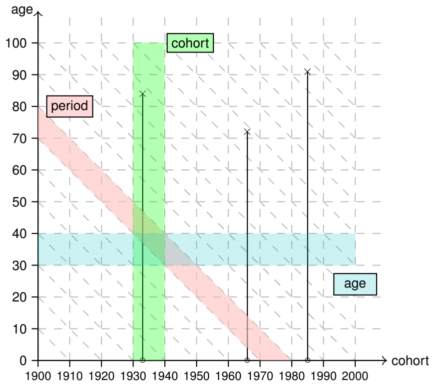
- If $C \perp\!\!\!\perp T^*$, the likelihood writes:

$$L_n = \sum_{i=1}^{n} \left(\Delta_i \log\left(\lambda\left(T_i\right)\right) - \int_0^{T_i} \lambda\left(t\right) dt \right).$$



Age-Cohort Diagram

- age effect: menopause
- cohort effect: carcinogenic baby food
- period effect: nuclear incident



Age-Cohort Diagram

- age effect: menopause
- cohort effect: carcinogenic baby food
- period effect: nuclear incident

Additional variable u.

$\rightarrow$ Bi-dimensional hazard $\lambda(t|u)$

Parametric estimation

- The hazard λ is discretized into J age intervals and K cohort intervals:

$$\lambda(t|u) = \sum_{j=1}^{J} \sum_{k=1}^{K} \lambda_{j,k} \mathbb{1}_{[c_{j-1}, c_j) \times [d_{k-1}, d_k)}(t, u)$$

Goal: infer $\lambda_{j,k}$

Parametric estimation

- The hazard λ is discretized into J age intervals and K cohort intervals:

$$\lambda(t|u) = \sum_{j=1}^{J} \sum_{k=1}^{K} \lambda_{j,k} \mathbb{1}_{[c_{j-1},c_j) \times [d_{k-1},d_k)}(t,u)$$

Goal: infer $\lambda_{j,k}$

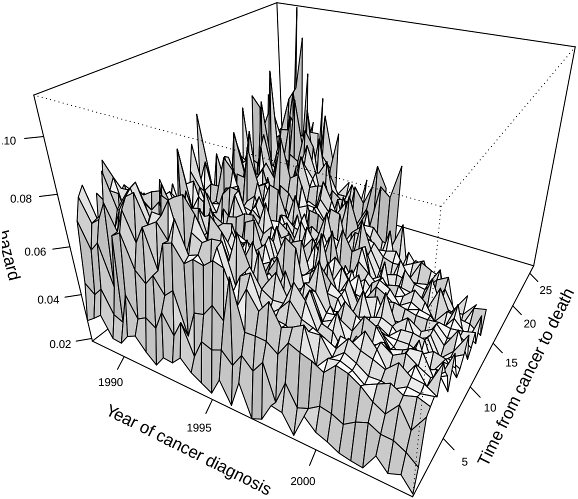
- Log-likelihood: $\ell_n = \sum O_{j,k} \log\left(\lambda_{j,k}\right) - R_{j,k}\lambda_{j,k}$, with exhaustive statistics

$$\begin{cases} O_{j,k} = \text{number of events} \\ R_{j,k} = \text{time at risk.} \end{cases}$$

Explicit MLE: $\lambda_{j,k} = \log \frac{O_{j,k}}{R_{j,k}} \quad \rightarrow$ **overfitting**.

Maximum likelihood estimate

Illustration of overfitting



Our approach: penalized likelihood

Reparametrization:

$$\log \lambda_{j,k} = \eta_{j,k},$$

Our approach: penalized likelihood

Reparametrization:

$$\log \lambda_{j,k} = \eta_{j,k},$$

And η is infered by *penalized* maximum likelihood:

$$\ell_n^{\text{pen}}(\boldsymbol{\eta}) = \underbrace{\ell_n(\boldsymbol{\eta})}_{\substack{\text{goodness}\\ \text{of fit}}}$$

Our approach: penalized likelihood

Reparametrization:

$$\log \lambda_{j,k} = \eta_{j,k},$$

And η is infered by *penalized* maximum likelihood:

$$\ell_n^{\text{pen}}(\boldsymbol{\eta}) = \underbrace{\ell_n(\boldsymbol{\eta})}_{\substack{\text{goodness}\\ \text{of fit}}} - \frac{\text{pen}}{2} \underbrace{\sum_{j,k} \left(\eta_{j+1,k} - \eta_{j,k}\right)^2 + \left(\eta_{j,k+1} - \eta_{j,k}\right)^2}_{\text{regularization}},$$

Our approach: penalized likelihood

Reparametrization:

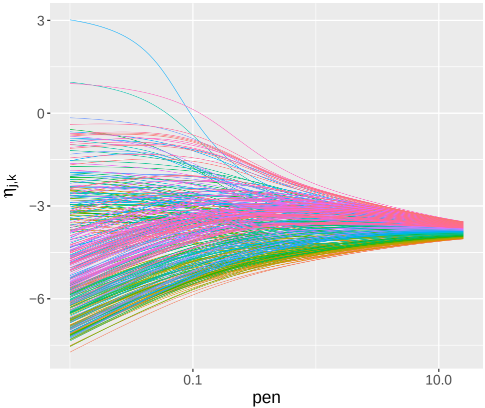
$$\log \lambda_{j,k} = \eta_{j,k},$$

And η is infered by *penalized* maximum likelihood:

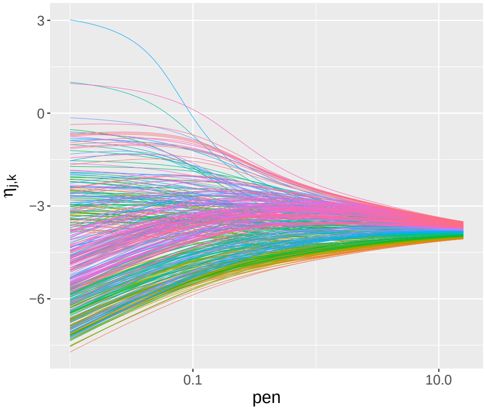
$$\ell_n^{\text{pen}}(\boldsymbol{\eta}) = \underbrace{\ell_n(\boldsymbol{\eta})}_{\substack{\text{goodness}\\\text{of fit}}} - \underbrace{\frac{\text{pen}}{2} \sum_{j,k} \left(\eta_{j+1,k} - \eta_{j,k}\right)^2 + \left(\eta_{j,k+1} - \eta_{j,k}\right)^2}_{\text{regularization}},$$

- pen $\to 0$; $\hat{\eta} \to \hat{\eta}^{\text{mle}}$; bias $\searrow$; variance $\nearrow$
- pen $\to \infty$; $\hat{\eta} \to$ constant; bias $\nearrow$; variance $\searrow$
- pen is a **bias-variance trade-off** parameter.

Regularization path:

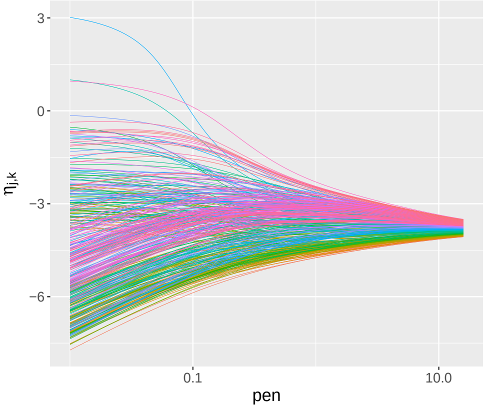


Regularization path:



pen $\to 0$: $\widehat{\lambda} \to \widehat{\lambda}^{\text{mle}}$
pen $\to \infty$: $\widehat{\lambda}$ constant

Regularization path:



$$\text{pen} \to 0 \quad : \quad \widehat{\lambda} \to \widehat{\lambda}^{\text{mle}}$$
$$\text{pen} \to \infty \quad : \quad \widehat{\lambda} \text{ constant}$$

The estimated hazard is **smoothed**.

We want to detect breakpoint moments of λ.

We want to detect breakpoint moments of λ.

The differences in the penalty are weighted:

$$\ell_n^{\text{pen}}(\boldsymbol{\eta}) = \ell_n(\boldsymbol{\eta}) - \frac{\text{pen}}{2} \sum_{j,k} v_{j,k} \left(\eta_{j+1,k} - \eta_{j,k}\right)^2 + w_{j,k} \left(\eta_{j,k+1} - \eta_{j,k}\right)^2 ,$$

We want to detect breakpoint moments of λ.

The differences in the penalty are weighted:

$$\ell_n^{\text{pen}}(\boldsymbol{\eta}) = \ell_n(\boldsymbol{\eta}) - \frac{\text{pen}}{2} \sum_{j,k} v_{j,k} \left(\eta_{j+1,k} - \eta_{j,k}\right)^2 + w_{j,k} \left(\eta_{j,k+1} - \eta_{j,k}\right)^2,$$

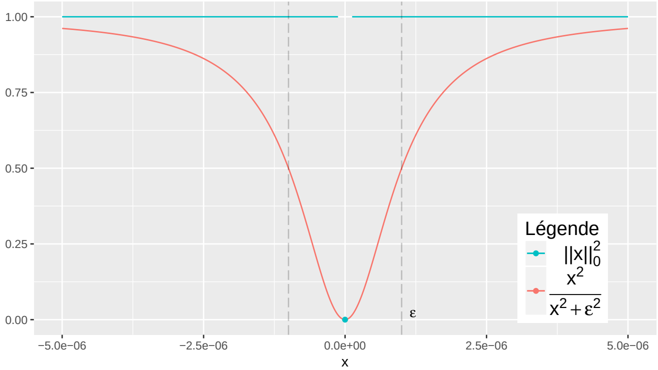
Weights are iteratively adapted:

$$\begin{cases} v_{j,k} = \left(\left(\eta_{j+1,k} - \eta_{j,k}\right)^2 + \varepsilon^2\right)^{-1} \\ w_{j,k} = \left(\left(\eta_{j,k} - \eta_{j,k-1}\right)^2 + \varepsilon^2\right)^{-1} \end{cases}, \quad \text{with} \quad \varepsilon \ll 1.$$

[1] F. Frommlet and G. Nuel, An Adaptive Ridge Procedure for L0 Regularization, *Public Library of Science*, 2016.

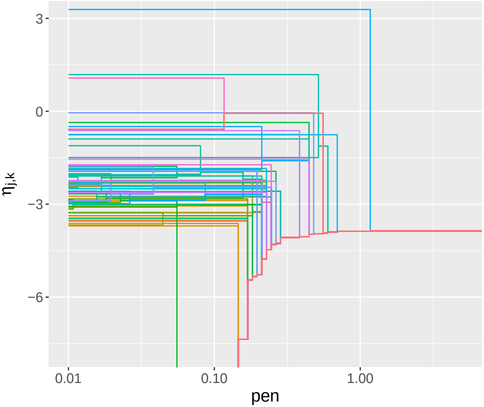
When $\varepsilon \ll 1$:

$$v_{j,k}\left(\eta_{j+1,k}-\eta_{j,k}\right)^2 \simeq \|\eta_{j+1,k}-\eta_{j,k}\|_0^2 = \begin{cases} 0 & \text{si } \eta_{j+1,k}=\eta_{j,k} \\ 1 & \text{si } \eta_{j+1,k}\neq\eta_{j,k} \end{cases}$$



L_0 regularization

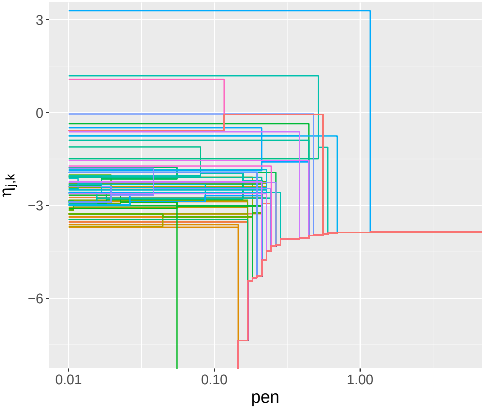
Regularization path:



$$\text{pen} \to 0 \quad : \quad \widehat{\lambda} \to \widehat{\lambda}^{\text{mle}}$$
$$\text{pen} \to \infty \quad : \quad \widehat{\lambda} \text{ constant}$$

L_0 regularization

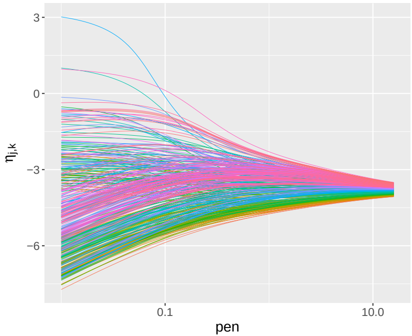
Regularization path:

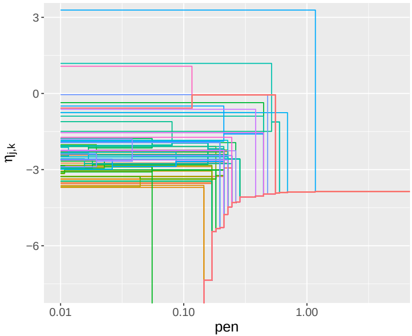


$$\text{pen} \to 0 \quad : \quad \widehat{\lambda} \to \widehat{\lambda}^{\text{mle}}$$
$$\text{pen} \to \infty \quad : \quad \widehat{\lambda} \text{ constant}$$

The estimated hazard is **piecewise constant**.

Comparision: smoothed *vs* segmented estimate

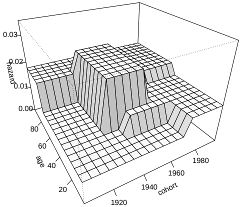




L_2 regularization:
Each penalty yields an estimate

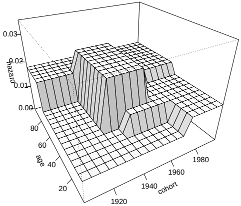
L_0 regularization:
Each penalty yields a *model*

Simulation: with piecewise constant true hazard

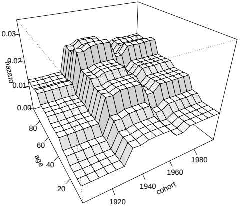


True λ

Simulation: with piecewise constant true hazard

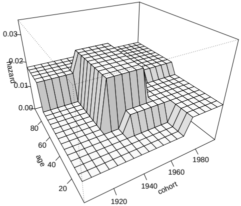


True λ

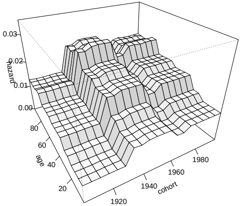


AGE-COHORT model:
$\lambda_{j,k} = \exp\left(\alpha_j + \beta_k\right)$

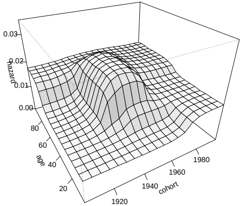
Simulation: with piecewise constant true hazard



True λ

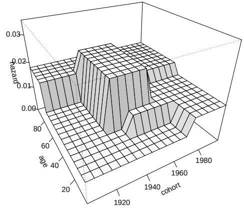


AGE-COHORT model:
$\lambda_{j,k} = \exp\left(\alpha_j + \beta_k\right)$

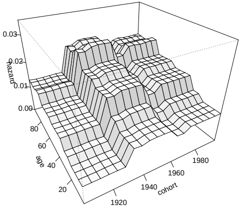


L_2 Regularization

Simulation: with piecewise constant true hazard

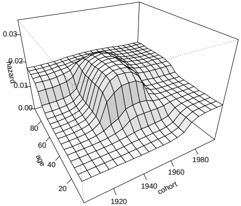


True λ

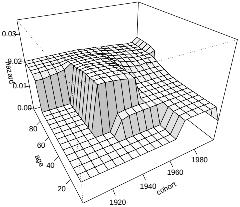


AGE-COHORT model:

$$\lambda_{j,k} = \exp\left(\alpha_j + \beta_k\right)$$

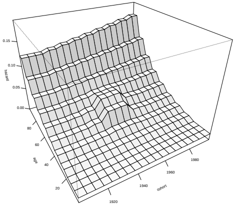


L_2 Regularization



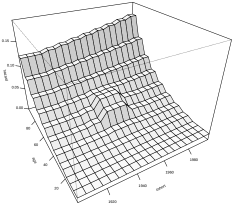
L_0 Regularization

Simulation: with smooth true hazard

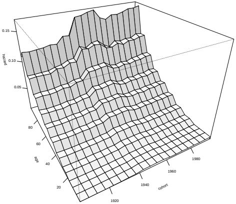


True λ

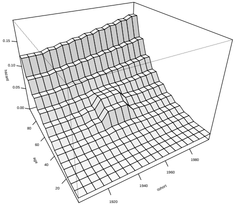
Simulation: with smooth true hazard



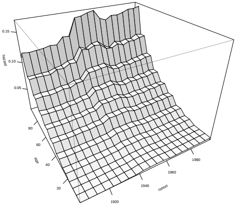
True λ



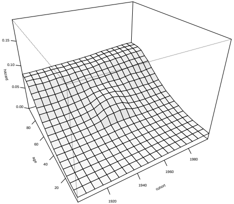
AGE-COHORT model:
$\lambda_{j,k} = \exp\left(\alpha_j + \beta_k\right)$



True λ

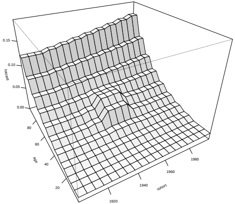


AGE-COHORT model:
$\lambda_{j,k} = \exp\left(\alpha_j + \beta_k\right)$

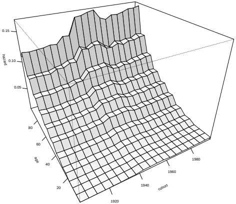


L_2 Regularization

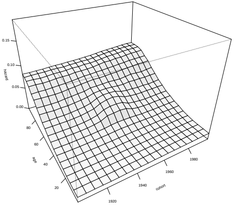
Simulation: with smooth true hazard



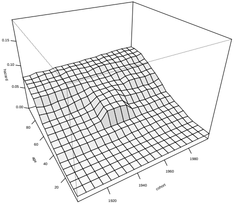
True λ



AGE-COHORT model: $\lambda_{j,k} = \exp\left(\alpha_j + \beta_k\right)$



L_2 Regularization

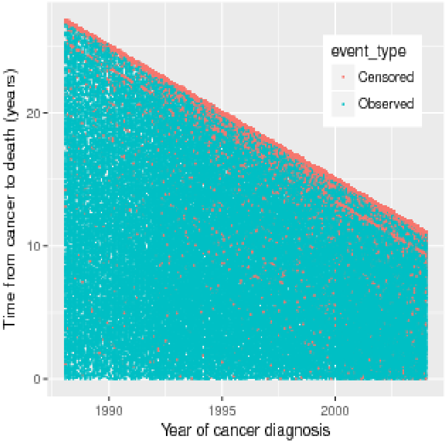


L_0 Regularization

SEER breast cancer mortality data:

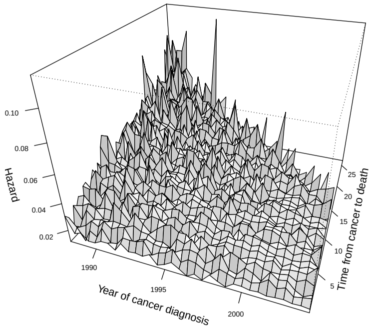
- US cancer survey
- Period: 1986 –.
- Sample size $\simeq 400,000$
- **Cohort** = Time of diagnosis
- **Age** = Time after diagnosis
- Cancer stage is registered

Question: has the mortality of breast cancer evolved with time?



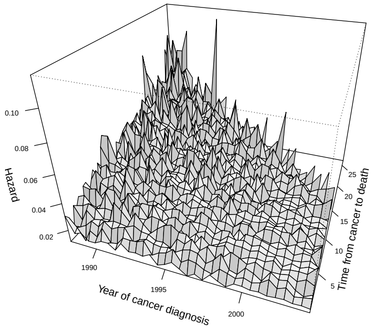
Death after diagnosis of stage 1 breast cancer

Application: Breast cancer morality (SEER data)

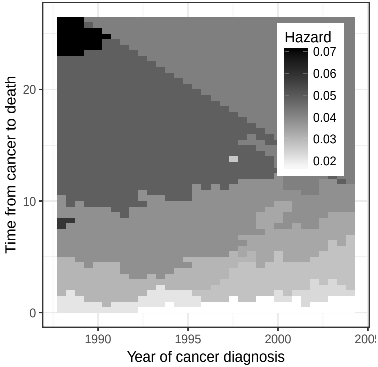


MLE: stage 1 cancers

Application: Breast cancer morality (SEER data)

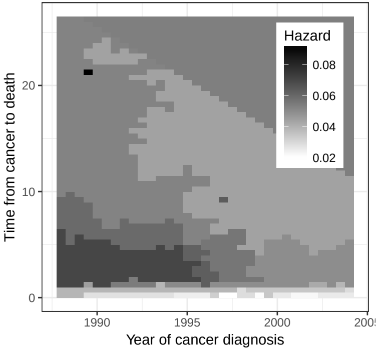


MLE: stage 1 cancers



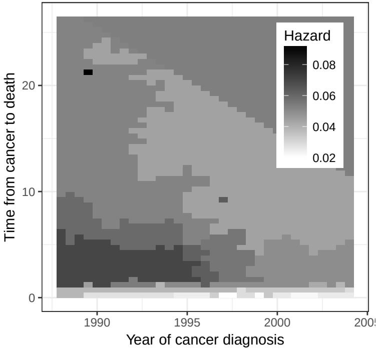
L_0 Regularization: stage 1 cancers

Application: Breast cancer morality (SEER data)

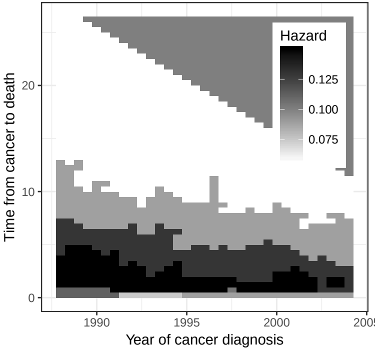


L_0 Regularization: stage 2 cancers

Application: Breast cancer morality (SEER data)



L_0 Regularization: stage 2 cancers



L_0 Regularization: stage 3 cancers

Summary and Perspectives

Summary:

- The hazard is estimated as a piecewise constant function
- Our model performs well even when the true hazard is not piecewise constant.

Perspectives:

- We can add age and cohort effects:

$$\log\left(\lambda_{j,k}\right) = \alpha_j + \beta_k + \delta_{j,k},$$

  with regularization of δ.

- Piecewise linear hazard estimation with penalty $\propto \left(\Delta^2 \eta_{j,k}\right)^2$.