Bachelor Thesis

Data-driven Lag-lead Selection in the Context of Exposure-Lag-Response-Associations

Author

Lisa Xu

Supervisor

Dr. Andreas Bender
Department of Statistics

Department of Statistics

Ludwig-Maximilians-Universität München

Munich, 4th of July 2022

Abstract

Special methods are needed to model the impact of time-dependent covariates in longitudinal data. For instance, the outcome can depend on multiple past observations with potentially time-varying exposures and time delays. Therefore the intensity and time of past exposures have to be taken into account for modeling cumulative effects. Since the time window, denoting all exposures affecting the hazard rate at a given time, is mostly unclear a priori, we apply a proposed method in which a fairly wide time window is chosen, in which past observations are penalized more strongly the further they are in the past (Obermeier et al., 2015). We consider the WCE model proposed by Sylvestre and Abrahamowicz (2009) to model cumulative effects, where a weight function for past exposures is estimated. By extending the method proposed by Obermeier et al. (2015) to the WCE model, the weight function is estimated using additional ridge penalties where either small or great latencies or both are penalized more strongly, making it possible to choose a fairly wide time window. We evaluate the correctness of the approach by performing simulation studies and comparing them to the method proposed so far. Lastly, we apply the extended method to data from the Colorado Plateau uranium miners cohort, where the association between radon exposures and mortality is modeled, to show the beneficial performance of the proposed methods.

Contents

List of Figures

List of Tables

1 Introduction

In the past different approaches were made to model the impact of *time-dependent covariates* (TDC) in longitudinal data. These special methods are needed because exposure status and its intensity may vary over time (Sylvestre and Abrahamowicz, 2009). Additionally, the effects of covariates can be lagged and last beyond the exposure period itself which means that the outcome is influenced by multiple exposure events with potentially time-varying exposures in the past. To measure the effect at a given time the intensity and the time of past exposures have to be taken into account when modeling cumulative effects. These dependencies are defined as *exposure-lag-response associations* (ELRA) (Gasparrini, 2014). To incorporate a variety of covariate effects we fit a *piece-wise exponential additive mixed model* (PAMM) (Bender et al., 2018b). It is unclear how the time window has to be chosen to consider all exposures that can affect the hazard rate at a given time. Moreover the choice of the maximal time lag is crucial and challenging since the true number of past observations is often unknown. In most existing implementations the time window for cumulative effects must be set a priori, which limits the functional utility of the approach. One proposed idea was to choose a fairly wide time window, including the true unknown maximal time lag, where partial effects of past observations are penalized depending on how far in the past the exposures occur. That requires the use of an additional ridge penalty (Obermeier et al., 2015).

In this thesis we strive to extend that approach to fit cumulative effects of PAMMs by constructing penalty matrices where the choice of the time window is not crucial anymore. In addition to penalizing partial effects that are further in the past, we also introduce another penalty to penalize partial effects of recent observations. These two penalties can be combined to penalize partial effects of observations that are either far in the past or occurred recently.

The thesis is structured as follows. First, in Section 1 the current problems of estimating time-dependent covariates in time-to-event data and its limitations are shown to motivate the proposed methods. In Section 2, we introduce the piece-wise exponential additive model and its ability to incorporate a wide variety of effects like time-varying effects or cumulative effects. For exposure-lag-response associations we present one possible model for modeling cumulative effects, the WCE model. After that we describe the idea of penalized splines for non-linear effects, providing the background knowledge for the use of additional ridge penalties. To conclude the methodology section, we introduce the methods extended to the cumulative effects setting and the use of two further penalty matrices. In Section 3, we present a simulation study to evaluate the proposed modeling approach and its estimation for two different shapes of the weight function. That section also includes the implementation of the use of penalties in **R**. To evaluate the performance, the proposed methods are applied to a data set from the Colorado Plateau uranium miners cohort in Section 4. The discussion in Section 5 deals with the main findings and addresses limitations and suggestions for further research. Finally, we conclude in Section 6.

2 Methodology

Section 1 showed the importance of special methods for modeling time-to-event data and its limitations. Therefore we start with the definition and notation of the used model, the piece-wise exponential additive mixed model (Section 2.1), and introduce the different effects that can be modeled with PAMMs (Bender et al., 2018b). Since PAMMs can include non-linear effects, splines are needed for modeling those effects. For that we introduce P-splines and penalty matrices proposed by Eilers and Marx (1996) which are used for the penalization (Section 2.2).

As mentioned before, the practical use of the PAMM is limited because of the lack of knowledge about the number of past observations to choose. Obermeier et al. (2015) proposed a method for modeling linear effects of covariates that have a lagged effect on the response variable, where the mentioned limitation is removed (Section 2.3). This approach is described and is then applied and extended to the case of complex cumulative effects (Section 2.4).

2.1 Piece-wise exponential additive mixed models

Survival time analysis focuses on the time T until a certain event or the censoring of an individual occurs (Cox, 1972). When dealing with time-to-event data the focus lies on the hazard rate which is defined as $\lambda(t)$. It represents the risk for an event occurring at the time t under the condition that the event has not yet occurred by time t. The hazard rate is mathematical given by

$$\lambda(t) := \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | t \leq T)}{\Delta t}. \tag{2.1}$$

In the following we are using Piece-wise exponential additive mixed models for modeling time-to-event data as proposed in Bender et al. (2018b). The PAMM is an extension of the piece-wise exponential model (PEM), which allows us to include diverse kind of effects like non-linear, multivariate or random effects by incorporating some of the methodology and algorithm which are already developed for Generalized additive mixed models. First the model specifications and framework for the PEM are described, which also apply to the PAMM. After defining those requirements the complete specification of the PAMM is introduced.

When using a PEM or a PAMM the follow-up time $(0, t_{\max}]$ has to be divided into J intervals with interval cut-points $0 = \kappa_0 < \cdots < \kappa_J = t_{\max}$, where $t_{\max}$ denotes the maximal follow-up time. Additionally we assume that the baseline hazard rate $\lambda_0(t)$ is constant for each interval j, so that $\lambda_0(t) = \lambda_{0j}, \forall t \in (\kappa_{j-1}, \kappa_j], t > 0$. The general proportional hazard model is given by

$$\log(\lambda_i(t|\boldsymbol{x}_i)) = \log(\lambda_0(t)) + \boldsymbol{x}_i^T \boldsymbol{\beta}, \quad \forall t \in (\kappa_{j-1}, \kappa_j] \tag{2.2}$$

where $\boldsymbol{x}_i^T = (x_{i,1}, ..., x_{i,P})$ is a row-vector of P time-constant covariates for subjects $i = 1, \dots, n$. Under the assumption that the basline hazard rate is constant, we can simplify (2.2) to

$$\log(\lambda_i(t|\boldsymbol{x}_i)) = \log(\lambda_{0j}) + \boldsymbol{x}_i^T \boldsymbol{\beta}, \quad \forall t \in (\kappa_{j-1}, \kappa_j]. \tag{2.3}$$

As described in Bender et al. (2018a) the data must include event indicators δ_{ij} for subject i for each interval $j = 1, \ldots, J$ to fit the required data structure. For that we define $\delta_{ij} = I(t_i \in (\kappa_{j-1}, \kappa_j] \wedge t_i = T_i)$ with $t_i := \min(T_i, C_i)$, which either stands for the event time T_i or the right-censored time C_i. Additionally, an offset $\log(t_{ij})$, in which $t_{ij} = \min(t_i - \kappa_{j-1}, \kappa_j - \kappa_{j-1})$ stands for the time a subject i was under risk in interval j, has to be incorporated into the likelihood of the model (Bender et al., 2018b). Finally, the likelihood of the PEM can be denoted as

$$\log(\mathbb{E}(\delta_{ij}|\boldsymbol{x}_i)) = \log(\lambda_{0j}) + \boldsymbol{x}_i^T\boldsymbol{\beta} + \log(t_{ij}) \tag{2.4}$$

and will be extended in the following to the PAMM. Note that the likelihood of a PEM corresponds to the likelihood of a poisson generalized linear model (Bender et al., 2018b).

The extension of a PEM to a PAMM is analogous to the extension of a generalized linear model to a generalized additive mixed model. Model (2.4) can then be extended to include other effects besides linear ones and to get rid of some restrictions of the PEM. The complete specification of the piece-wise exponential additive mixed model is given by

$$\log(\lambda_i(t|\boldsymbol{x}_i, \boldsymbol{z}_i, l_i)) = \log(\lambda_0(t)) + \sum_{p=1}^{P} f_p(x_{i,p}, t) + g(\boldsymbol{z}_i, t) + b_{l_i} + \log(t_{ij}) \tag{2.5}$$

where $\log(\lambda_0(t))$ describes the log-baseline hazard rates, $f_p(x_{i,p}, t)$ flexible covariate effects, $g(\boldsymbol{z}_i, t)$ exposure-lag-response-associations and b_{l_i} describes random effects (Bender et al., 2018b). Those individual components are discussed in more detail below.

Baseline hazard

In contrast to PEM where the baseline hazard, denoted as $\lambda_0(t) = \lambda_{0j}$, was a step function it is now estimated as a smooth, non-linear function. This is advantageous since the step function often led to instability and large changes between adjacent intervals. To overcome this issue the baseline hazard is now estimated with a spline at every interval midpoint $\tilde{t}_j$, so that big changes in the hazard rate between adjacent intervals are appropriately penalized. Consequently $\log(\lambda_{0j})$ can be rewritten to $f_0(\tilde{t}_j)$ (Bender et al., 2018b).

Flexible covariate effects

In addition to model (2.4), PAMMs can contain non-linear and time-varying effects which are included by $f_p(x_{i,p}, t)$ in the model specification and allows more flexibility. The term $f_p(\cdot)$ can represent linear, time-constant effects, like $\boldsymbol{x}_i^T\boldsymbol{\beta}$ in (2.4), as well as time-varying effects, which can be linearly or smoothly included. The smooth function is described with the help of splines $f_p(\cdot) = \sum_{m=1}^{M} \gamma_{m,p} B_{m,p}(\cdot)$, with the covariate specific basis functions $B_{m,p}$ and the belonging spline coefficients $\gamma_{m,p}$ (Bender et al., 2018b).

Random effects

By incorporating random effects b_{l_i} into the model, group-specific effects can be taken into account where l_i indicates the group $l = 1, \dots, L$ to which subject i belongs (Bender et al., 2018b).

Exposure-Lag-Response-Associations

When computing the risk at time t time-varying exposures and the timing of past exposures have to be taken into account as proposed in Gasparrini (2014). The outcome can then depend on multiple exposure events with potentially different intensities over the follow-up time t. Before describing one possible specification of exposure-lag-response-associations $g(\boldsymbol{z}_i, t)$, which is described in Sylvestre and Abrahamowicz (2009), we introduce some notation.

Variables that have time-varying exposures are denoted as *time-dependent covariates* (TDC). These time-varying exposures can be described through a subject's exposure history $\boldsymbol{z} = (z(t_{e_1}), z(t_{e_2}, ..))$ where the value of the TDC at exposure time t_e is denoted as $z(t_e)$. In order to model *cumulative effects* over the follow-up time, it is necessary to indicate a time window, which is defined by $t_{\text{lag}}(t_e)$ and $t_{\text{lead}}(t_e)$. The *lag* time $t_{\text{lag}}(t_e)$ describes the time delay until the TDC registered at time t_e affects the hazard. The *lead* time $t_{\text{lead}}(t_e)$ describes the time interval in which the TDC registered at time t_e has an effect on the hazard. These definitions lead to considering the following exposures $\{z(t_e) : t_e \in [t - t_{\text{lag}}(t_e) - t_{\text{lead}}(t_e), t - t_{\text{lag}}(t_e)]\}$ when computing the hazard at time t. To model cumulative effects we compute the so-called *partial effects* over the *lag-lead window* $\mathcal{T}_e(j) := \{t_e : (\kappa_{j-1}, \kappa_j] \in \mathcal{J}(t, t_e)\}$ with $\mathcal{J}(t, t_e) := \{((\kappa_{j-1}, \kappa_j] : \kappa_{j-1} > t_e + t_{\text{lag}} \wedge \kappa_j \leq t_e + t_{\text{lag}} + t_{\text{lead}}\}$ indicating those intervals j in which the exposure at time t_e can affect the hazard rates (Bender et al., 2018b).

One possible specification of ELRA is the *weighted cumulative exposure* (WCE) model and was introduced by Sylvestre and Abrahamowicz (2009) where partial effects are assumed to be linear in $z(t_e)$ and are only dependent on the latency $t - t_e$. Using the notation given in Bender et al. (2018a) the WCE is defined as

$$g(\boldsymbol{z}, t) = \int_{\mathcal{T}_e(t)} h(t - t_e) z(t_e) \mathrm{d}t_e. \tag{2.6}$$

The WCE model computes the cumulative effects of time-varying exposures weighted by recency. Sylvestre and Abrahamowicz (2009) proposed to estimate the weights $h(t - t_e)$ with the help of cubic regression B-splines.[1]

[1] Sylvestre and Abrahamowicz (2009) provides a more detailed discussion of the WCE.

2.2 Penalized splines

As mentioned earlier non-linear effects as well as time-varying effects can be included in the smooth function $f_p(\cdot)$ with the help of splines. There are different kind of splines, but we will focus on the P-splines, which were defined by Eilers and Marx (1996) and represent a combination of B-splines and difference penalties. For B-splines the choice of knots is crucial: if too many knots are chosen it will lead to overfitting, but if too few knots are used it will lead to underfitting the data. This problem is avoided when using P-splines because the number of knots that must be selected is chosen relatively large. In order to prevent overfitting a difference penalty is used on the coefficients of adjacent B-splines. This method is described in more detail in the following for univariate smooth functions (Eilers and Marx, 1996).

Without any penalization, the aim of a regression, with m observations (x_i, y_i) on n B-splines $B_j(\cdot)$ with associated spline basis coefficients γ_j, is to minimize the least squares objective function

$$LS = \sum_{i=1}^{m}(y_i - \sum_{j=1}^{n}\gamma_j B_j(x_i))^2. \tag{2.7}$$

In order to avoid overfitting when choosing a relatively large number of knots Eilers and Marx (1996) recommended a penalty based on finite differences of the coefficients of adjacent B-splines. Then (2.7) can be extended to the penalized least squares criterion

$$PLS = \sum_{i=1}^{m}(y_i - \sum_{j=1}^{n}\gamma_j B_j(x_i))^2 + \lambda \sum_{j=k+1}^{n}(\Delta^k\gamma_j)^2 \tag{2.8}$$

where differences of order k are used, with Δ^k as the k^{th} difference of the B-spline coefficients and parameter λ for controlling the smoothness of the fit (Eilers and Marx, 1996, Fahrmeir et al., 2013). The difference operator Δ^k in (2.8)

$$\begin{aligned}
\Delta^1\gamma_j &= \gamma_j - \gamma_{j-1} \\
\Delta^2\gamma_j &= \Delta^1\Delta^1\gamma_j = \Delta^1\gamma_j - \Delta^1\gamma_{j-1} = \gamma_j - 2\gamma_{j-1} + \gamma_{j-2} \\
&\vdots \\
\Delta^k\gamma_j &= \Delta^{k-1}\gamma_j - \Delta^{k-1}\gamma_{j-1}
\end{aligned} \tag{2.9}$$

can be recursively defined (Fahrmeir et al., 2013). With the help of these smoothing splines the dimensionality of the problem is reduced from the number of data points m to the number of B-splines n. Depending on the model different orders of difference penalties can be used where the difference penalty get more complex the higher the order of the penalties get (Eilers and Marx, 1996). The first-order difference matrix, where D_k denotes the matrix notation of the difference operator Δ^k, is given by

$$\underset{(n-1)\times n}{D_1} = \begin{pmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{pmatrix} \tag{2.10}$$

which leads to

$$D_1\boldsymbol{\gamma} = \begin{pmatrix} \gamma_2 - \gamma_1 \\ \vdots \\ \gamma_n - \gamma_{n-1} \end{pmatrix}. \tag{2.11}$$

Differences of higher orders are recursively defined:

$$\underset{(n-k)\times n}{D_k} = D_1 D_{k-1}. \tag{2.12}$$

This results in the penalty

$$\lambda \sum_{j=k+1}^{n} (\Delta^k \gamma_j)^2 = \lambda \gamma^T D_k^T D_k \gamma = \lambda \gamma^T K_k \gamma \tag{2.13}$$

with $D_k^T D_k = K_k$ (Fahrmeir et al., 2013). Analogous to the least squares method on B-splines we now have to minimize (2.8). Therefore a system of equations, where the number of equations are equal to the number of splines, is given by

$$B^T y = (B^T B + \lambda K_k)\gamma \tag{2.14}$$

with B consisting of the elements $b_{ij} = B_j(x_i)$. When $\lambda = 0$ we get the least squares objective function with a B-spline basis as in (2.7). When $\lambda > 0$ only the main diagonal and k subdiagonals of the system of equations are affected by the penalty. After introducing the penalty the penalized likelihood function is given by

$$l_p(\gamma) = l(\gamma) - \frac{1}{2}\lambda \gamma^T K_k \gamma \tag{2.15}$$

with the log-likelihood $l(\gamma)$ (Eilers and Marx, 1996).

2.3 Flexible distributed lag model

In many situations it is of interest to model the association between a covariate x and a response variable y, where the association is likely to be cumulative over a time period. Therefore preceding time points $t-1, t-2, .., t-L$ must be taken into account for modeling the effect at time t. The association between a covariate x_{t-l} and a response variable y_t at time t is denoted

as the *lag effect* of x on y with time lag $l = 0, .., L$. There have been several approaches to model this effect especially since the simple inclusion of all lagged covariates $x_{t-l}, l = 0, .., L$ leads to an estimation with many parameters that are likely to be collinear. Furthermore the choice of L is crucial and challenging since it is often unclear how many past time points have an effect on the response variable. Obermeier et al. (2015) proposed a lag modeling approach, the so-called *flexible distributed lag (FDL) model*, for the linear influence of lagged covariates on the response variable based on B-splines in which, in addition, the crucial choice of L is no longer critical. A difference penalty as well as a ridge penalty are used in the FDL model. The former is used for smoothing the shape of the lag effects and the latter is used to make sure that the last lag coefficient is shrunk towards 0. Since a ridge penalty is used, a large lag length L can be chosen where those coefficients, which are redundant, are estimated close to 0. In the following we will discuss the mathematical background of the FDL model and the use of difference and ridge penalties (Obermeier et al., 2015).
First, consider a lag model where the lag coefficients β_L are developed in B-splines of degree d:

$$\underset{(L+1)\times 1}{\beta} = \underset{(L+1)\times m}{B} \underset{m\times 1}{\gamma} \tag{2.16}$$

with

$$\underset{(L+1)\times m}{B} = \begin{pmatrix} B_1(0) & \dots & B_m(0) \\ \vdots & & \vdots \\ B_1(L) & \dots & B_m(L) \end{pmatrix} \tag{2.17}$$

with $m = d + K$ basis coefficients and K equidistant knots κ_k. We will refer to this model as the *lag model based on basis functions*. To this model a smoothing and a shrinkage penalty is added. For penalizing adjacent coefficients a difference penalty, as described in Section 2.2, is used. The lag coefficients $\boldsymbol{\beta}$ are penalized by penalizing the basis coefficients $\boldsymbol{\gamma}$ with the $m \times m$ smoothing penalty matrix $K_d = D_2^T D_2$ when using differences of second order (Obermeier et al., 2015). Additionally a ridge penalty is defined to penalize the last lag coefficient β_L. Since $d+1$ d-degree B-spline basis functions are non-zero at any point $l \in [0, L]$, especially at point L, this can be used to make sure that the last lag coefficient β_L is shrunk towards 0 by penalizing the last $d + 1$ γ-coefficients with the ridge penalty matrix, also called shrinkage matrix, which is given by

$$\underset{m\times m}{K_r} = \begin{pmatrix} 0 & & & & & \\ & \ddots & & & & \\ & & 0 & & & \\ & & & 1 & & \\ & & & & \ddots & \\ & & & & & 1 \end{pmatrix}. \tag{2.18}$$

After defining the smoothing and the ridge penalty the penalized likelihood can be defined by

$$l_p(\gamma) = l(\gamma) - \frac{1}{2}\gamma^T(\lambda_d K_d + \lambda_r K_r)\gamma \tag{2.19}$$

with smoothing parameter $\lambda_d > 0$ and shrinkage parameter $\lambda_r > 0$. The smoothing parameter controls the smoothness of the fit by penalizing large differences between adjacent coefficients whereas the shrinkage parameter penalizes large values in the last lag coefficient. The estimation can be performed with known methods like the maximization of the penalized likelihood. In addition to the difference and ridge penalty the model can be further limited by applying the constraint matrix C to make sure that the last lag coefficient is not only shrunk towards 0 but is equal to 0. C consists of the first row of the model matrix X (Obermeier et al., 2015).

Besides this use of difference and ridge penalties, another variation can also be considered, the *full-rank* penalty. Instead of (2.19) the penalized likelihood can be denoted as

$$l_p(\gamma) = l(\gamma) - \frac{1}{2}\lambda\gamma^T(K_d + K_r)\gamma \tag{2.20}$$

when using a full-rank penalty, where the ridge penalty is added with the same variance as the difference penalty and therefore only one coefficient λ is estimated.

2.4 Extension and application of the FDL model to PAMMs

In the following section we will apply the methods proposed by Obermeier et al. (2015) to the estimation of exposure-lag-response-associations, which were presented in Section 2.1, and extend the approach to more complex cumulative effects. One limitation when modeling cumulative effects is the choice of the lag-lead window $\mathcal{T}_e(t)$, which must be set a priori. To remove this limitation a relative wide window is chosen, in which partial effects of past observations are penalized more strongly. Further on, when modeling ELRA, we will focus on the WCE model. We are interested in penalizing the latency $t - t_e$ since the partial effects in (2.6) are dependent on it. Therefore we introduce penalty matrices to penalize the latency. For the WCE model regular P-splines are used to penalize the basis coefficients instead of cubic regression B-splines. Analogous to the FDL model we add a smoothing penalty matrix $K_d = D_2^T D_2$, when using differences of second order with D_2 as defined in (2.12). Additionally we also add a shrinkage matrix as in (2.18) to the model, which we refer to as *lag-penalty* $K_{r,1}$:

$$\underset{n\times n}{K_{r,1}} = \begin{pmatrix} 0 & & & & & \\ & \ddots & & & & \\ & & 0 & & & \\ & & & 1 & & \\ & & & & \ddots & \\ & & & & & 1 \end{pmatrix} \tag{2.21}$$

with n as the number of basis coefficients. This ensures that great values for $t-t_e$ are penalized. In addition also small latencies can be penalized by adjusting $K_{r,1}$ to

$$\underset{n\times n}{K_{r,2}} = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{pmatrix} \tag{2.22}$$

called *lead-penalty*. Now, we can also combine these two shrinkage penalties to penalize the greatest values and the smallest values for $t - t_e$ with the *laglead-penalty* matrix $K_{r,3}$:

$$\underset{n\times n}{K_{r,3}} = \begin{pmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 1 & & & & & & \\ & & & 0 & & & & & \\ & & & & \ddots & & & & \\ & & & & & 0 & & & \\ & & & & & & 1 & & \\ & & & & & & & \ddots & \\ & & & & & & & & 1 \end{pmatrix}. \tag{2.23}$$

The penalized likelihood can then be defined by

$$l_p(\gamma) = l(\gamma) - \frac{1}{2}\gamma^T(\lambda_d K_d + \lambda_{r,i} K_{r,i})\gamma \tag{2.24}$$

for $i = 1, 2, 3$, where the latency coefficients are penalized by penalizing the basis coefficients $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_n)^T$ for the WCE model. As mentioned in Section 2.3 the shrinkage matrix penalizes the last $d+1$ basis coefficients in the case of $K_{r,1}$ and the first $d+1$ basis coefficients in the case of $K_{r,2}$. For $K_{r,3}$ the first $d+1$ and the last $d+1$ basis coefficients are penalized, however, it should be noted that the number of coefficients that are penalized is $\min(2\cdot(d+1), n)$.

3 Simulation study

We perform a simulation study to evaluate the performance of the proposed modeling approach of cumulative effects. The simulation focuses on the investigation of ELRAs and its estimation of partial effects based on the WCE model as described in (2.6), where the weights $h(t-t_e)$ are estimated with the help of P-splines and the use of the different penalty matrices defined in Section 2.4. Two different simulation studies are performed to evaluate the ability of the proposed approach in Section 2.4 for two different distributions of the weight function $h(t-t_e)$.

In *Simulation Part A* a normal distribution is used to simulate $h(t-t_e)$ whereas in *Simulation Part B* a half-normal distribution was chosen. Simulation Part A (Section 3.2) demonstrates the effect of variables where their impact on the outcome is lagged. Therefore relative small latencies, as well as great latencies, are weighted weaker than intermediate latencies. In contrast, Simulation Part B (Section 3.3) examines the performance of variables whose effect is only slightly lagged or not lagged at all. The results are evaluated using graphical comparisons of the estimated weight function $\hat{h}(t-t_e)$ to the respective true weight function $h(t-t_e)$. Additionally, the simulation is evaluated with the *root mean square error* statistic RMSE $= \sqrt{\frac{1}{R}\sum_{i=1}^{R}(h(t-t_e)-\hat{h}(t-t_e))^2}$ (Fahrmeir et al., 2013) and with coverage$_\alpha$ $= \frac{1}{R}\sum_{i=1}^{R} I\big(h(t-t_e) \in [\hat{h}(t-t_e) \pm \zeta_{1-\frac{\alpha}{2}}\hat{\sigma}_{\hat{h}(t-t_e)}]\big)$, where $\zeta_{1-\frac{\alpha}{2}}$ is the $(1-\frac{\alpha}{2})$-quantile of the standard normal distribution and $\hat{\sigma}_{\hat{h}(t-t_e)}$ is the estimated standard deviation of the estimated weight function, for R replications (Bender et al., 2018b).

3.1 Data generation

In both parts random survival times are simulated from the piece-wise exponential distribution ($t \sim \text{PEXP}(\boldsymbol{\lambda}, \boldsymbol{\kappa})$) with piece-wise constant hazards $\boldsymbol{\lambda}$ for each interval and time-points $\boldsymbol{\kappa}$. We aim to simulate data with hazard rate

$$\log(\lambda(t|x_1, \boldsymbol{z})) = -3.5 + f_0(\tilde{t}_j) - 0.5x_1 + \int_{\mathcal{T}_e(t)} h(t-t_e)z(t_e)\mathrm{d}t_e \tag{3.1}$$

where $f_0(\tilde{t}_j)$ is a gamma density function $\mathcal{G}(8,2)$. The window of effectiveness $\mathcal{T}_e(t)$ is shown in Figure 1 and partial effects are integrated over the preceding 12 time units. The shape of $h(t-t_e)$ differs between Part A and Part B

$$h(t-t_e)z(t_e) = \begin{cases} 2 \cdot \Phi_{6,1.7^2}(t-t_e) \cdot z(t_e) & \text{for Part A} \\ 2 \cdot HN_{0.45}(t-t_e) \cdot z(t_e) & \text{for Part B} \end{cases} \tag{3.2}$$

with Φ_{μ,σ^2} density function of the normal distribution with mean μ and variance σ^2. HN_θ denotes the density function of the half-normal distribution with parameter θ. Figure 2 shows the two different shapes of $h(t-t_e)$. Data is generated for different number of observations $n \in \{500, 1000, 1500, 3000\}$ to evaluate the impact of the number of observation on the estimation.

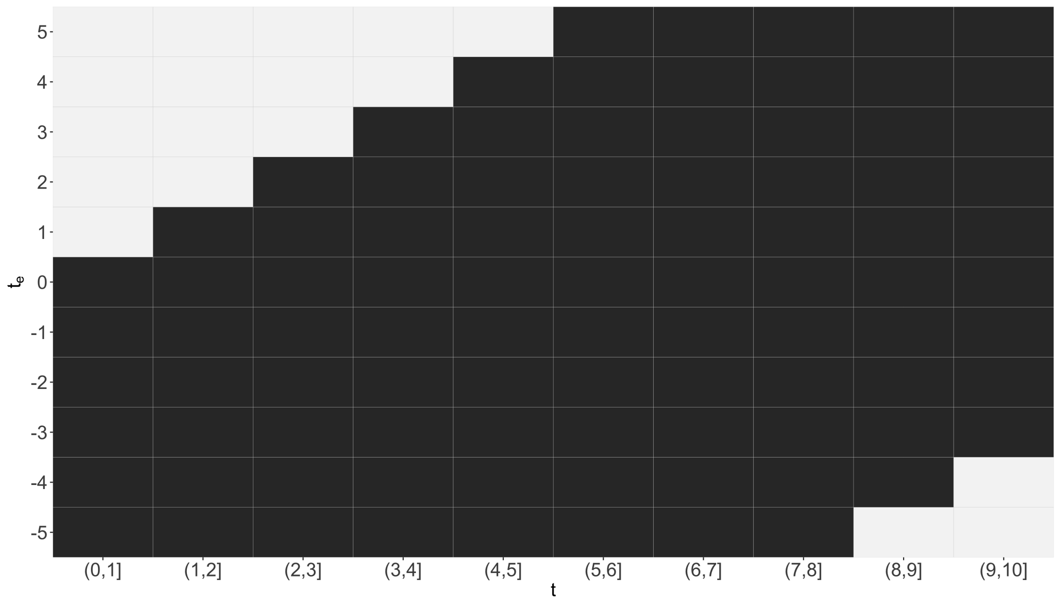

Figure 1: Illustration of the lag-lead window $\mathcal{T}_e(t)$. Viewed column-wise the black squares indicate which exposure times t_e are affecting the hazard at time t. For example, the hazard at $t = (1, 2]$ is affected by exposures which were recorded at $t_e \in [-5, 1]$.

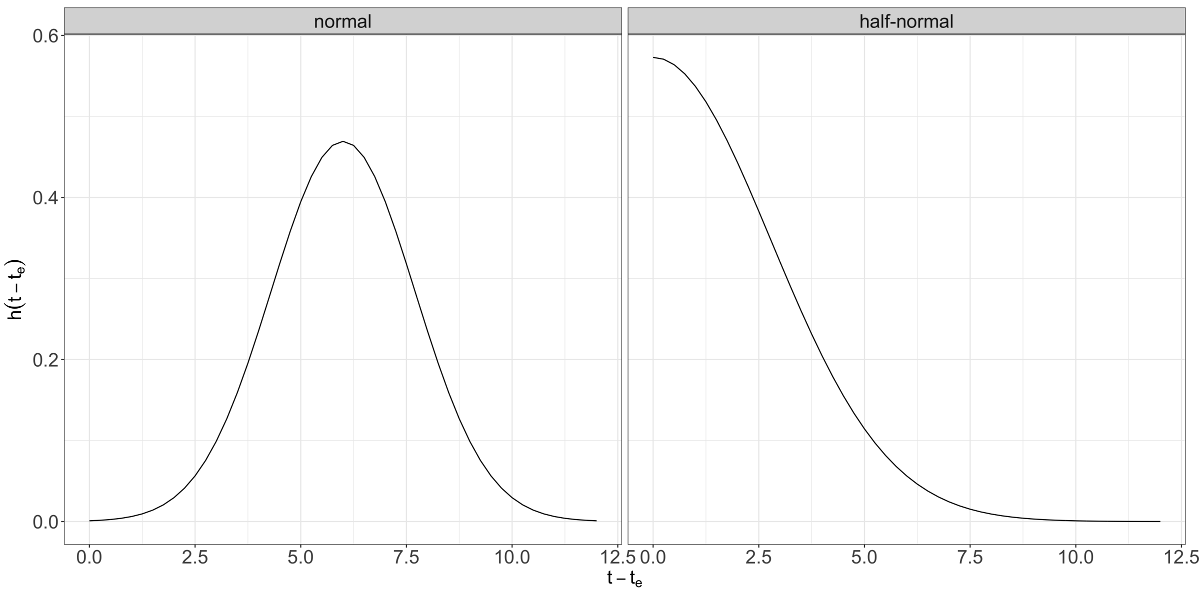

Figure 2: $h(t - t_e)$ defined for Simulation Part A (left) and Simulation Part B (right).

After simulating the data all models are estimated by PAMMs. We differ between four methods for estimating the weight function $h(t - t_e)$. To investigate the performance of the three ridge penalties $K_{r,i}$ for $i = 1, 2, 3$ as defined in Section 2.4, three different models are estimated using one penalty matrix each. To compare their performance, a fourth model is fitted, where the latency is penalized using cubic regression B-splines as originally proposed for WCE models by Sylvestre and Abrahamowicz (2009). For each setting, where we use different number of observations and different penalization approaches, $R = 100$ replications were run.

3.2 Simulation Part A

The results for Simulation Part A are shown in Figure 3. It gives an overview of the performance of the different methods in comparison to the proposed use of cubic regression B-splines by Sylvestre and Abrahamowicz (2009). As mentioned before, we simulated data for four different numbers of observations. Note that in every run different survival times, and therefore different datasets, are simulated.

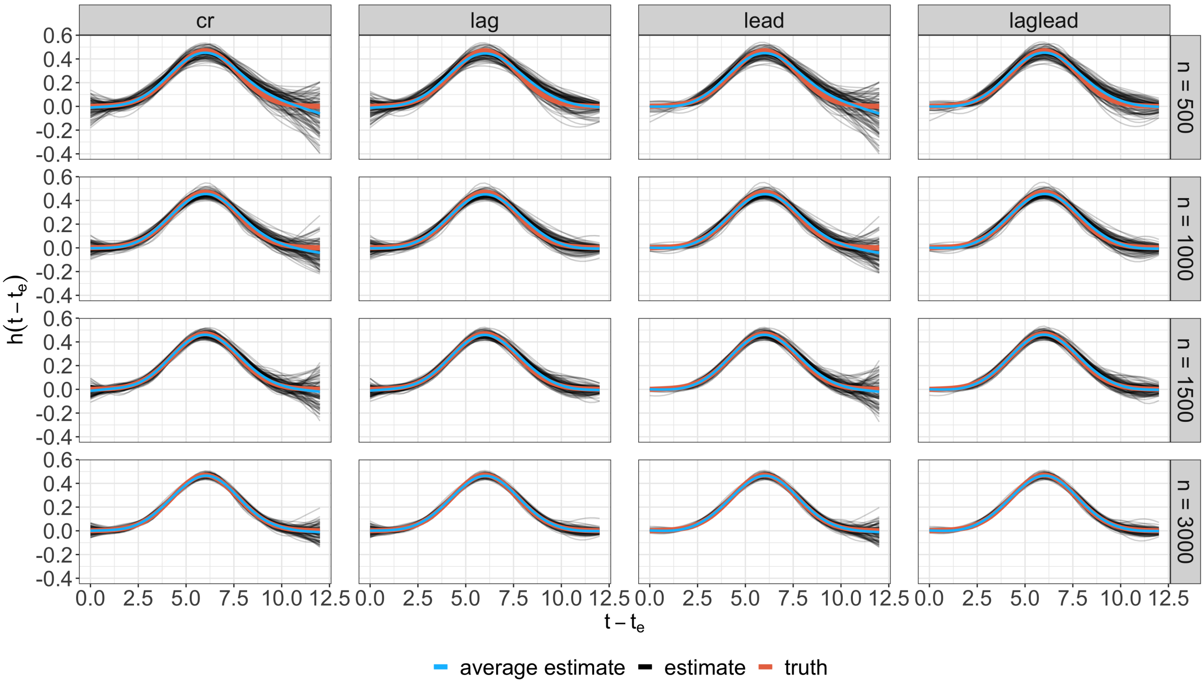

Figure 3: Simulation Part A for $h(t-t_e)$ with the shape of a normal distribution: estimated WCE and true simulated WCE for different number of observations n and different penalty methods: Cubic regression B-splines (first column), FDL model with lag-penalty $K_{r,1}$ (second column), FDL model with lead-penalty $K_{r,2}$ (third column) and FDL model with laglead-penalty $K_{r,3}$ (fourth column).

First, we consider the second column in Figure 3 for penalization with additional lag-penalty $K_{r,1}$: in comparison to the estimation with cubic regression B-splines great latencies are successfully shrunk towards 0 whereas small latencies remain unaffected. This applies to all n but the quality of the shrinkage varies between different numbers of observations. Graphically, the estimation based on the largest dataset with $n = 3000$ delivers the best result as the estimated $h(t-t_e)$ are the closest to the true weight function. If we additionally look at the RMSE in Table 1, we see that the FDL model with lag-penalty peformed well since it has a noticeably lower RMSE for great latencies compared to the model using cubic regression B-splines. Comparing the coverage displayed in Table 2, the model using the lag-penalty yields, especially for great latencies, a higher coverage than using cubic regression B-splines. This applies to all settings except for $n = 500$ with $t - t_e = 6.0$, where the coverage decreases compared to the usual method.

The FDL model with lead-penalty $K_{r,2}$ ensures that small latencies are properly penalized.

RMSE					
number observations	$t - t_e$	cr	lag	lead	laglead
	0.0	0.07	0.07	0.02	0.02
$n = 500$	6.0	0.04	0.04	0.04	0.04
	12.0	0.15	0.04	0.15	0.03
	0.0	0.04	0.04	0.01	0.01
$n = 1000$	6.0	0.03	0.03	0.03	0.03
	12.0	0.10	0.02	0.11	0.02
	0.0	0.04	0.04	0.01	0.01
$n = 1500$	6.0	0.03	0.03	0.03	0.03
	12.0	0.11	0.03	0.11	0.03
	0.0	0.03	0.03	0.01	0.01
$n = 3000$	6.0	0.02	0.02	0.02	0.02
	12.0	0.06	0.02	0.07	0.02

Table 1: Evaluation of Simulation Part A: RMSE for different number of observations n and different latencies $t - t_e$ using cubic regression B-splines (third column), FDL model with lag-penalty $K_{r,1}$ (fourth column), FDL model with lead-penalty $K_{r,2}$ (fifth column) or FDL model with laglead-penalty $K_{r,3}$ (sixth column).

coverage					
number observations	$t - t_e$	cr	lag	lead	laglead
	0.0	0.94	0.93	0.79	0.79
$n = 500$	6.0	0.90	0.86	0.88	0.90
	12.0	0.94	1.00	0.94	1.00
	0.0	0.99	1.00	0.99	0.95
$n = 1000$	6.0	0.88	0.89	0.91	0.93
	12.0	0.98	1.00	0.99	1.00
	0.0	0.97	0.99	1.00	0.91
$n = 1500$	6.0	0.91	0.94	0.95	0.95
	12.0	0.97	1.00	0.96	1.00
	0.0	0.97	0.99	1.00	0.81
$n = 3000$	6.0	0.94	0.95	0.96	0.96
	12.0	0.99	0.99	0.99	0.99

Table 2: Evaluation of Simulation Part A: $\text{coverage}_{0.05}$ for different number of observations n and different latencies $t - t_e$ using cubic regression B-splines (third column), FDL model with lag-penalty $K_{r,1}$ (fourth column), FDL model with lead-penalty $K_{r,2}$ (fifth column) or FDL model with laglead-penalty $K_{r,3}$ (sixth column).

This is supported by consideration of the simulation study as depicted in the third column in Figure 3. Small latencies are shrunk towards 0 for all numbers of observations. For small latencies the simulated weight functions were estimated very well and the RMSE for the FDL model with lead-penalty has been successfully reduced. Note that the coverage decreased notably for small latencies for $n = 500$.

Finally, we examine the estimation when using the third penalty, the laglead-penalty $K_{r,3}$. The laglead-penalty combines $K_{r,1}$ and $K_{r,2}$ by penalizing small and great values for $t-t_e$. Small latencies were penalized similarly as when using the lead-penalty $K_{r,2}$ and great latencies were penalized similarly to the second case, where an additional lag-penalty $K_{r,1}$ was used. As we see in Figure 3 the estimated weight functions in the fourth column represent a combination of the weight functions seen in the second and third columns. For all settings the RMSE decreased as supposed for $t - t_e = 0.0$ and $t - t_e = 12.0$ when using the laglead-penalty. Considering the coverage for that case, it only increased for $t - t_e = 6.0$ and $t - t_e = 12.0$ compared to the model with cubic regression B-splines, whereas the coverage for $t - t_e = 0.0$ does not correspond to the coverage yielded when using the lead-penalty.

3.3 Simulation Part B

The results of the simulation study with weight function $h(t-t_e)$ from a half-normal distribution are shown in Figure 4. Considering the estimation with cubic regression B-splines, we see that in general small latencies are already better estimated than great latencies where the estimated weight functions differs more from the true weight function.

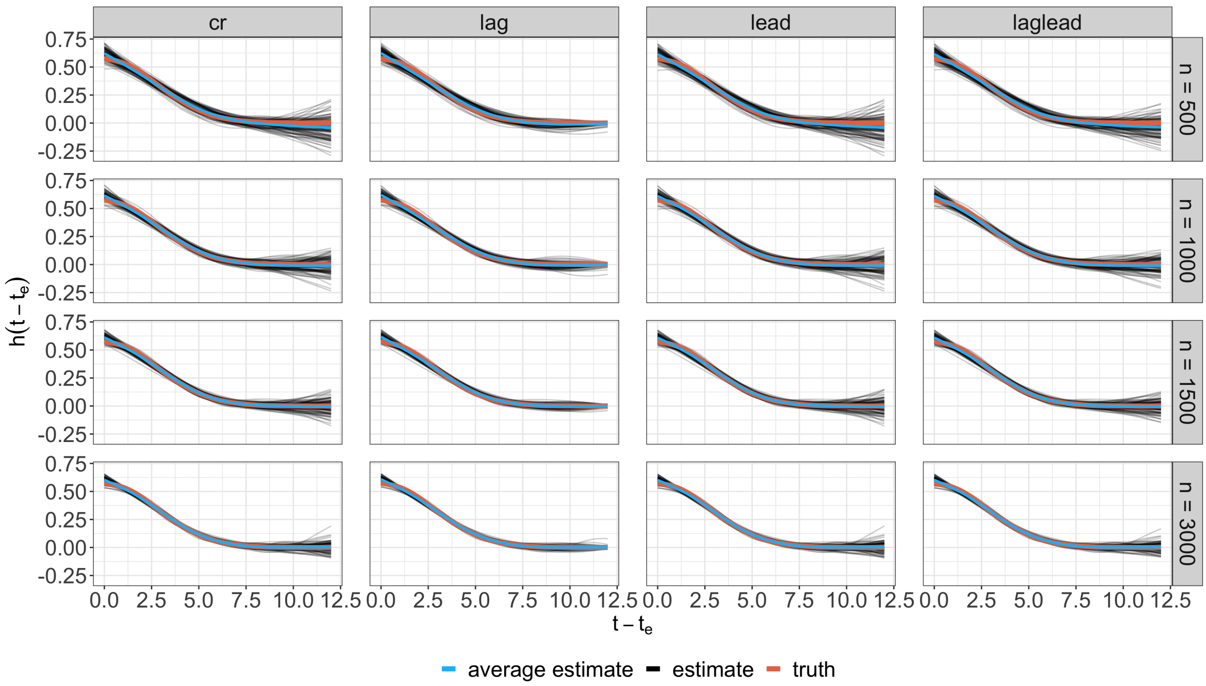

Figure 4: Simulation Part B for $h(t-t_e)$ with the shape of a half-normal distribution: estimated WCE and true simulated WCE for different number of observations n and different penalty methods: Cubic regression B-splines (first column), FDL model with lag-penalty $K_{r,1}$ (second column), FDL model with lead-penalty $K_{r,2}$ (third column) and FDL model with laglead-penalty $K_{r,3}$ (fourth column).

For the FDL model with additional lag-penalty $K_{r,1}$ great latencies were successfully shrunk towards 0 as displayed in Figure 4. Comparing the FDL model with lag-penalty to the model using cubic regression B-splines, the RMSE notably decreased for $t - t_e = 12.0$ for all number of observations (Table 4).

RMSE					
number observations	$t - t_e$	cr	lag	lead	laglead
	0.0	0.07	0.07	0.06	0.06
$n = 500$	6.0	0.03	0.03	0.03	0.03
	12.0	0.11	0.01	0.11	0.09
	0.0	0.06	0.06	0.05	0.05
$n = 1000$	6.0	0.02	0.02	0.02	0.02
	12.0	0.08	0.01	0.08	0.07
	0.0	0.04	0.05	0.05	0.05
$n = 1500$	6.0	0.02	0.02	0.02	0.02
	12.0	0.06	0.01	0.06	0.06
	0.0	0.04	0.04	0.04	0.04
$n = 3000$	6.0	0.01	0.01	0.01	0.01
	12.0	0.05	0.01	0.05	0.04

Table 3: Evaluation of Simulation Part B: RMSE for different number of observations n and different latencies $t - t_e$ using cubic regression B-splines (third column), FDL model with lag-penalty $K_{r,1}$ (fourth column), FDL model with lead-penalty $K_{r,2}$ (fifth column) or FDL model with laglead-penalty $K_{r,3}$ (sixth column).

coverage					
number observations	$t - t_e$	cr	lag	lead	laglead
	0.0	0.90	0.88	0.95	0.94
$n = 500$	6.0	0.91	0.90	0.91	0.91
	12.0	0.95	1.00	0.95	0.97
	0.0	0.85	0.81	0.86	0.86
$n = 1000$	6.0	0.95	0.96	0.97	0.96
	12.0	0.98	1.00	0.98	0.98
	0.0	0.80	0.71	0.83	0.83
$n = 1500$	6.0	0.96	0.95	0.96	0.96
	12.0	0.99	1.00	1.00	1.00
	0.0	0.70	0.65	0.74	0.73
$n = 3000$	6.0	0.97	0.98	0.98	0.98
	12.0	0.99	1.00	0.99	0.99

Table 4: Evaluation of Simulation Part B: coverage$_{0.05}$ for different number of observations n and different latencies $t - t_e$ using cubic regression B-splines (third column), FDL model with lag-penalty $K_{r,1}$ (fourth column), FDL model with lead-penalty $K_{r,2}$ (fifth column) or FDL model with laglead-penalty $K_{r,3}$ (sixth column).

When considering the use of the lead-penalty $K_{r,2}$ we see in Figure 4 that the estimated weight functions looks very similar to the estimated weight functions using cubic regression B-splines. Due to the distribution of the weight function, $h(t - t_e)$ for small latencies should not be shrunk towards 0 since the true weight function has its maximum at $t - t_e = 0$. Since small latencies have the greatest values for the weight function, recent exposures get weighted more strongly. Therefore an additional penalty matrix has hardly any influence on the estimation of the weight function since $K_{r,2}$ was designed to shrink small latencies that are redundant.

Considering Table 3 and Table 4 we see that the RMSE for the model using cubic regression B-splines is nearly equivalent to the RMSE for the model using $K_{r,2}$, however, the use of the lead-penalty resulted in a higher coverage.

Lastly, if we look at the fourth column in Figure 4, we see that the use of an additional ridge penalty $K_{r,3}$ leads to almost the same estimation as we saw for cubic regression B-splines and for the lead-penalty $K_{r,2}$. Unlike Simulation Part A, the use of $K_{r,3}$ does not lead to a combination of the estimated weight functions seen in the second and third columns. The RMSE in Table 3 show the same result, the RMSE approximately corresponds to the RMSE of the model with cubic regression B-splines or of the model with lead-penalty. Further, the RMSE for $t - t_e = 12.0$ did not decrease as well as with the lag-penalty. Considering Table 4, the coverage for the model using $K_{r,3}$ roughly corresponds to the coverage when using $K_{r,2}$ and is slightly higher than the coverage when using cubic regression B-splines.

3.4 Computational Details

The simulation study was performed using the software **R** (Version 4.1.0). In general the **R** package **pammtools** (Version 0.5.8) was used for data generation of random survival times from the piece-wise exponential distribution and for model post-processing (Bender et al., 2022). For estimation the package **mgcv** (Version 1.8-40) was used for fitting GAMs with smooth terms (Wood, 2017). To apply the FDL model approach, Obermeier et al. (2015) used a constructor function to specify smooth terms in a GAM, which was already embedded in the package **mgcv**. That constructor function was then extended to add two further penalties $K_{r,2}$ and $K_{r,3}$. Data and code used for both simulation study parts and for the following application in Section 4 are provided on GitLab.

4 Application

We apply our method in a survival analysis of time-to-event data. The analysis is based on data from the Colorado Plateau uranium miners cohort, that was collected by the National Institute for Occupational Safety and Health (Gasparrini, 2014). Since the exposure histories for radon for each subject vary during the follow-up time, it represents a setting where we can apply our proposed methods to model cumulative effects. We are interested in the performance of the proposed methods compared to the use of cubic regression B-splines when estimating the weight function $h(t - t_e)$.

4.1 Data

If miners worked within the Colorado Plateau area between 1950 and 1960 they were suitable to enter the cohort. In this example we used data referring to the follow-up on December 31, 1982. Exposure histories for radon are available from the time the subjects started to work in the mines, even if this was before entering the cohort. The radon exposures were expressed in working-level months (WLM) and reconstructed for each year the subject worked in the mines (Gasparrini, 2014). A working level is defined as 1.3×105 MeV of alpha energy/l air, which is emitted by short lived radon progeny. A WLM equals exposure to 1 WL for 170 hours (Kreuzer et al., 2011). We are interested in the association between radon exposure and mortality. If a subject entered the cohort on January 01, 1950, we considered the maximal follow-up time $t = 33$ years after study entry. Therefore all miners still alive after December 31, 1982, were censored and the follow-up time t was set to $t = 33$. The cohort includes $n = 3323$ miners (after pre-processing), from which 2639 (79.4%) miners were smokers and 1245 (37.5%) miners died during the follow-up time. A summary of the data is provided in Table 5.

	Min	Q_1	Median	Q_3	Max
Follow-up time (years)	0.1	19.6	23.8	25.5	32.5
Age at study entry	15.8	25.8	34.0	44.0	80.0
Smoking starting age	3.0	15.0	17.0	19.0	60.0
Radon exposure starting age	6.9	22.5	29.5	40.3	73.0
Total cumulative radon exposure (WLM/year)	0.0	153.0	429.0	1015.5	10000.0

Table 5: Descriptive statistics of the Colorado Plateau uranium miners cohort.

4.2 Modeling

Beside the radon exposures for each subject we included the age at which the subject started smoking. The effect of radon is represented in the model by ELRA $g(\boldsymbol{z}_{\text{radon}}, t)$ as structured in (2.6). The model specification is given by

$$\log(\lambda_i(t|\boldsymbol{x}_i, \boldsymbol{z}_i, l_i)) = \log(\lambda_0(t)) + f_{\text{smk}}(x_{i,\text{smk}}, t) + g(\boldsymbol{z}_{\text{radon}}, t) + \log(t_{ij}) \tag{4.1}$$

where $f_{\text{smk}}(x_{i,\text{smk}}, t)$ incorporates the non-linear effect of the age at which subject i started smoking. The weight function $h(t - t_e)$ was estimated using four different ELRA specifications each: cubic regression B-splines, lag-penalty, lead-penalty or laglead-penalty. The lag-lead window $\mathcal{T}_e(j)$ was chosen relatively wide with maximal latency $t - t_e = 40$.

4.3 Model selection

We are mainly interested in the performance of the proposed methods when estimating the weight function $h(t - t_e)$ for radon exposures. Therefore, we focus on model selection criteria . One criterion we will first consider is the *Akaike Information Criterion* (AIC) which is defined as $\text{AIC} = -2 \cdot l(\hat{\boldsymbol{\beta}}_M, \hat{\sigma}^2) + 2(|M|+1)$ where $l(\hat{\boldsymbol{\beta}}_M, \hat{\sigma}^2)$ is the maximum value of the log-likelihood and $|M| + 1$ is the total number of parameters. Smaller values of the AIC represent a better model fit. Secondly, we look at the *Bayesian Information Criterion* (BIC) which is defined by $\text{BIC} = -2 \cdot l(\hat{\boldsymbol{\beta}}_M, \hat{\sigma}^2) + \log(n)(|M| + 1)$, where $l(\hat{\boldsymbol{\beta}}_M, \hat{\sigma}^2)$ and $|M| + 1$ are defined in the same way as for the AIC and n is the number of recorded measurements. Here, too, smaller values of BIC represent a better model fit. The main difference between AIC and BIC is that complex models are more strongly penalized by BIC than by AIC (Fahrmeir et al., 2013).

ELRA specification	AIC	BIC
laglead	14882.8	15042.0
lag	14883.2	15043.5
lead	14883.4	15044.1
cr	14883.6	15050.3

Table 6: Evaluation of the application to data from the Colorado Plateau uranium miners cohort: Values for AIC and BIC for different models of ELRA between radon and mortality.

The fit of different ELRA specifications is shown in Table 6 by AIC and BIC. Both criteria came to the same result, identifying the model where $h(t - t_e)$ was estimated using a laglead-penalty as the model with the best performance. That model penalizes the weight function for small latencies as well as for great latencies more strongly. All our proposed methods obtained smaller values for AIC and BIC, although the differences can be marginal, than the originally proposed method using cubic regression B-splines. The findings suggest that for estimating ELRA an additional laglead-penalty should be used.

5 Discussion

In this thesis, we explored how to model time-to-event data using PAMMs without the constraint of defining the time window for cumulative effects a priori. We successfully applied the FDL model for modeling effects of lagged linear covariates proposed by Obermeier et al. (2015) to model cumulative effects of TDCs (ELRA) (Bender et al., 2018b). For estimating ELRAs we focused on the WCE model, where partial effects are assumed to be linear in $z(t_e)$ and are only dependent on the latency $t - t_e$ (Sylvestre and Abrahamowicz, 2009). By applying the additional ridge penalty proposed by Obermeier et al. (2015), the weight function $h(t - t_e)$ for great latencies is properly estimated by penalizing the last $d+1$ basis coefficients. Additionally we extended the FDL model by introducing two additional ridge penalties, the lead-penalty $K_{r,2}$ and the laglead-penalty $K_{r,3}$, to penalize the first $d + 1$ basis coefficients or the first and the last basis coefficients. By penalizing those latencies, the time window for cumulative effects can be chosen quite wide since $h(t-t_e)$ for superfluous latencies are correctly estimated close to 0. Another advantage over the originally proposed use of cubic regression B-splines to estimate the weight function is that the choice of knots is not further critical when using the FDL model.

Since the weight function $h(t-t_e)$ can have different shapes we have looked at two different distributions. Regarding the shape of a normal distribution, which represents the lagged effect of TDCs, we observe the following: the use of the three proposed ridge penalties leads to the intended results, where the weight function for either small latencies, great latencies or both are successfully shrunk towards 0. Compared to the originally proposed use of cubic regression B-spline (Sylvestre and Abrahamowicz, 2009), the use of the lag-penalty led to an notable smaller RMSE for great latencies, whereas the model, in which the lead-penalty was used, yielded in a smaller RMSE for small latencies. The use of an additional laglead-penalty led to an smaller RMSE for small and great latencies. For each of the four different penalization methods, the RMSE decreased or at least remained the same as the number of observations increased. In general, a high coverage was achieved when using the FDL model with either the lag-penalty or lead-penalty. Although the laglead-penalty showed the smallest RMSE, it did not achieved the highest coverage among the other methods. For increasing number of observations, the FDL models with lag-penalty or lead-penalty showed the highest coverage.

In Simulation Part B we observed that, compared to the use of cubic regression B-splines for estimating $h(t - t_e)$, for the FDL model with lag-penalty the weight function was shrunk towards 0 for great latencies. For all number of observations and for $t - t_e = 12$ the RMSE were decreased to 0.01 and coverage was increased to 1.00. For the FDL model using the lead-penalty, the RMSE could hardly or not at all be reduced for small latencies, however, the coverage could be increased. Considering the last ridge penalty, we saw that, unlike Simulation Part A, the use of the laglead-penalty does not lead to a combined result of the lag-penalty and the lead-penalty. Especially, RMSE for great latencies were minimally reduced but not as well as with lag-penalty. The coverage of the FDL model with laglead-penalty could slightly be increased for small and great latencies, compared to the model using cubic regression B-splines.

After applying the proposed methods to data from the Colorado Plateau uranium miners cohort, we have detected that the use of a additional laglead-penalty is suggested, especially since we do not know the true window of effectiveness $\mathcal{T}_e(t)$. The application also showed that the use of an additional ridge penalty, regardless of which of the three proposed penalty matrices is used, led to better performance, regarding AIC and BIC, than using cubic regression B-splines.

While we focused on the WCE model as one possible specification of ELRA there are other specifications of those associations. Since the assumption that partial effects are linear in $z(t_e)$ and are only dependent on the latency $t-t_e$ is restricted, we present two further specifications of ELRAs: One further specification is the *distributed lag non-linear model* (DLNM) where partial effects are specified by Berhane et al. (2008) and the framework was proposed by Gasparrini (2014). The DLNM is given by

$$g(\boldsymbol{z}, t) = \int_{\mathcal{T}_e(t)} h(t - t_e, z(t_e))\mathrm{d}t_e \tag{5.1}$$

where the partial effects are also assumed to depend on the latency $t - t_e$. In contrast to the WCE model (2.6), function $h(\cdot)$ is now defined as a two-dimensional function. Therefore special tensor product P-splines are used to extend the one-dimensional function $h(\cdot)$ to a two-dimensional function, where tensor product basis coefficients are penalized, depending on the latency $t - t_e$ and the exposure $z(t_e)$.

Another specification was proposed by Bender et al. (2018b) where the assumption made for the WCE model or DLNM is relaxed. The general exposure-lag-response-association is given by

$$g(\boldsymbol{z}, t) = \int_{\mathcal{T}_e(t)} h(t, t_e, z(t_e))\mathrm{d}t_e \tag{5.2}$$

where $h(t, t_e, z(t_e))$ ensures that partial effects can also depend on the time t and on the exposure time t_e, and not only on the latency $t - t_e$. Therefore specific combinations of t and t_e can be considered. For example, for the WCE model or for DLNM $h(t = 30, t_e = 3) \overset{!}{=} h(t = 40, t_e = 13) \overset{!}{=} \tilde{h}(t - t_e = 27)$ applies, whereas for the general ELRA partial effects for $t = 30$ and $t_e = 3$ are not the same as for $t = 40$ and $t_e = 13$ (Bender et al., 2018b).

Based on the results from our analysis, further research should examine the extension and construction of penalty matrices to estimate more complex cumulative effects like the presented DLNM (5.1) or the general ELRA (5.2). A special focus will be on embedding the additional penalties into the **mgcv** environment since for both specifications tensor product P-splines are used for penalizing.

6 Conclusion

In this thesis an approach was shown to loosen the limitation of setting the width of the time window for cumulative effects a priori when using the WCE model. By applying the approach made by Obermeier et al. (2015) to cumulative effects, the estimation of the weight function was shrunk towards 0 for great latencies. Thus, past observations are penalized more strongly the further they are in the past by using the additional lag-penalty. Additionally, we introduced two further penalty matrices, one to penalize past observations more strongly the closer they are to the time of interest t, by shrinking the weight function for small latencies towards 0. We refer to this penalty matrix as the lead-penalty. The second introduced penalty matrix, the so-called laglead-penalty, combines these two penalty matrices. When using the laglead-penalty for estimating the weight function, small latencies and great latencies are penalized more strongly, which results in the weight function being pushed towards 0 at both ends. Moreover, the choice of knots, which was crucial when using cubic regression B-splines, is no longer crucial when using penalty matrices. Especially for weight functions with shape of a normal distribution, the use of one of the proposed penalty matrices leads to an advanced estimation in comparison to the originally proposed use of cubic regression B-splines for the WCE model.

Bibliography

Andreas Bender, Andreas Groll, and Fabian Scheipl. A generalized additive model approach to time-to-event analysis. *Statistical Modelling*, 18(3-4):299–321, 2018a. doi: 10.1177/1471082X17748083.

Andreas Bender, Fabian Scheipl, Wolfgang Hartl, Andrew G. Day, and Helmut Küchenhoff. Penalized estimation of complex, non-linear exposure-lag-response association. *Biostatistics*, 20(2):315–331, 2018b. doi: 10.1093/biostatistics/kxy003.

Andreas Bender, Fabian Scheipl, Philipp Kopper, and Lukas Burk. *pammtools: Piece-Wise Exponential Additive Mixed Modeling Tools for Survival Analysis*, 2022. URL https://adibender.github.io/pammtools/. R package version 0.5.8.

Kiros Berhane, Michael Hauptmann, and Bryan Langholz. Using tensor product splines in modeling exposure–time–response relationships: Application to the colorado plateau uranium miners cohort. *Statistics in Medicine*, 27:5484–5496, 2008. doi: 10.1002/sim.3354.

D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34:187–220, 1972. ISSN 00359246.

Paul H. C. Eilers and Brian D. Marx. Flexible smoothing with b-splines and penalties. *Statistical Science*, 11:89–121, 1996. doi: 10.1214/ss/1038425655.

Ludwig Fahrmeir, Thomas Kneib, Stefan Lang, and Brian Marx. *Regression: Models, Methods and Applications*. Springer-Verlag Berlin Heidelberg, 2013. ISBN 978-3-642-34332-2.

Antonio Gasparrini. Modeling exposure-lag-response associations with distributed lag non-linear models. *Statistics in medicine*, 33(5):881–899, 2014. doi: 10.1002/sim.5963.

Michaela Kreuzer, Bernd Grosche, Florian Dufey, Maria Schnelzer, Annemarie Tschense, and Linda Walsh. The german uranium miners cohort study (wismut cohort). *Technical Report*, 2011. URL http://nbn-resolving.de/urn:nbn:de:0221-201102185211.

Viola Obermeier, Fabian Scheipl, Christian Heumann, Joachim Wassermann, and Helmut Küchenhoff. Flexible distributed lags for modelling earthquake data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 64(2):395–412, 2015. doi: 10.1111/rssc.12077.

Marie-Pierre Sylvestre and Michal Abrahamowicz. Flexible modeling of the cumulative effects of time-dependent exposures on the hazard. *Statistics in Medicine*, 28(27):3437–3453, 2009. doi: 10.1002/sim.3701.

Simon N. Wood. *Generalized Additive Models: An Introduction with R (2nd ed.)*, 2017.

Declaration of Authenticity

The work contained in this thesis is original and has not been previously submitted for examination which has led to the award of a degree.
To the best of my knowledge and belief, this thesis contains no material previously published or written by another person except where due reference is made.

Lisa Xu