LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

INSTITUT FÜR STATISTIK

Andreas Bender, Fabian Scheipl, Helmut Küchenhoff

Modeling Exposure-Lag-Response Associations with Penalized Piece-wise Exponential Models

Technical Report Number 192, 2016
Department of Statistics
University of Munich

http://www.stat.uni-muenchen.de

Abstract

We present a novel approach for the flexible modeling of exposure-lag-response associations, i.e., time-to-event data where multiple past exposures are cumulatively associated with the hazard after a certain temporal delay. Our method is based on piece-wise exponential models and allows estimation of a wide variety of effects, including potentially smooth and time-varying effects as well as cumulative effects with leads and lags, taking advantage of the advanced inference methods that have recently been developed for generalized additive mixed models.

1 Introduction

In many applications of survival analysis, study objects are exposed to different levels of an external covariate over the course of the follow up, for example different amounts of caloric intake received by critically ill patients during their stay on the intensive care unit. Modeling the association between such protracted exposures and outcome is difficult as, for each subject, the level of exposure may vary over time. Moreover, the effect of exposure on outcome is also likely to vary over the course of the follow up, and hazard rates at a particular point in time may depend on multiple past exposures. Lastly, the impact of a concrete exposure may have a delayed impact on the outcome and decline depending on the gap between time of exposure and evaluation time.

In more technical terms, such complex exposures imply the need for an approach that can incorporate time-dependent covariates (TDC) and model their possibly non-linear, possibly time-varying, cumulative effects on survival with lead and lag times. Additionally, we need to adjust for cluster or individual specific heterogeneity (frailty) and other possibly non-linear, possibly smoothly time-varying effects of confounders recorded at baseline. In previous work in this field, Berhane et al. (2008) used tensor product smooths to model the association between survival and protracted exposure to radiation. Sylvestre and Abrahamowicz (2009) presented the *weighted cumulative exposure* (WCE) model, where the effect of exposure at time t is the sum of weighted past exposures and the weight function is estimated smoothly using B-Splines. Smoothness is controlled through comparison of models based on different number of interior knots with respect to the BIC. Xiao et al. (2014) extended the WCE approach to marginal structural Cox models. Gasparrini (2014) introduced an approach based on distributed lag non-linear models and coined the term *exposure-lag-response associations* (ELRA) for the type of relationship described above, which we will adopt in this article.

We propose a flexible, novel approach for the modeling of the aforementioned *exposure-lag-response associations*. The method, an extension of the piece-wise exponential model (PEM), is described in detail in section 2. By embedding the concept of PEMs into the framework of generalized additive mixed models (GAMM) (cf. section 2.2), we can define a flexible model class for survival analysis and ELRA in particular, that *inherits* most of the flexible tools for modeling, estimation and validation of GAMMs. Practical usefulness of this approach is further increased due to readily available, robust and efficient implementations of these methods (Wood, 2006, 2011). We extend existing methodology regarding confidence intervals and testing procedures for smooth terms, to derive respective measures and test statistics for ELRAs and particularly for the comparison of hazard differences resulting from different patterns of a TDC. In section 3, we review the proposed method and discuss advantages as well as disadvantages of the approach.

2 Methods and Model

2.1 Piece-wise Exponential Models

We define

$$\lambda_i(t|\mathbf{x}_i) = \lambda_0(t)\exp(\mathbf{x}_i'\boldsymbol{\beta}), \tag{1}$$

a general proportional hazards model with $i = 1, \dots, n$, n the number of subjects under study and $\mathbf{x}_i' = (x_i^1, \dots, x_i^P)$ the row-vector of time-constant covariates $\mathbf{x}^p, p = 1, \dots P$.

A piece-wise exponential model (PEM) is obtained by partitioning the follow up period $(0, t_{\max}]$ into J intervals with $J+1$ cut-points $0 = \kappa_0 < \dots < \kappa_J = t_{\max}$. The j-th interval is given by $(\kappa_{j-1}, \kappa_j]$, where $t_{\max}$ is the maximal follow up time. Assuming the hazard rate in each interval j to be constant, such that $\lambda_0(t) = \lambda_j, \forall t \in (\kappa_{j-1}, \kappa_j], t > 0$, equation (1), in log-linear form, simplifies to

$$\log(\lambda_i(t|\mathbf{x}_i)) = \log(\lambda_j) + \mathbf{x}_i'\boldsymbol{\beta} \ \forall \ t \in (\kappa_{j-1}, \kappa_j]. \tag{2}$$

Let $t_i = \min(T_i, C_i)$ the right-censored time under risk for subject i. Given intervals $1, \dots, J$, Holford (1980) and Laird and Olivier (1981) first established the link between the likelihood of the model in (2) and the likelihood of the Poisson GLM (3) with

(a) one observation for each interval j under risk for each subject i,

(b) responses $y_{ij} = 1$ if $t_i \in (\kappa_{j-1}, \kappa_j] \wedge t_i = T_i$, else $y_{ij} = 0$ as event indicators for subject i for interval j, and

(c) offsets $t_{ij} = \min(t_i - \kappa_{j-1}, \kappa_j - \kappa_{j-1})$, the time subject i spends under risk in interval j (Friedman, 1982):

$$\log(\mathbb{E}(y_{ij}|\mathbf{x}_i)) = \log(\lambda_{ij}t_{ij}) = \log(\lambda_j) + \mathbf{x}_i'\boldsymbol{\beta} + \log(t_{ij}), \tag{3}$$

or, with $\lambda_i(t|\mathbf{x}_i) := \lambda_{ij}$,

$$\log(\lambda_i(t|\mathbf{x}_i)) = \log\left(\frac{\lambda_{ij}t_{ij}}{t_{ij}}\right) = \log(\lambda_j) + \mathbf{x}_i'\boldsymbol{\beta}. \tag{4}$$

The likelihood of model (3) is proportional to the likelihood of the PEM (2), thus the two models are equivalent with respect to the ML estimation of the model parameters $\boldsymbol{\beta}$. In practice, when fitting the according Poisson regression, $\log(\lambda_j)$ is incorporated in the linear predictor $\mathbf{x}_i'\boldsymbol{\beta}$ and $\log(t_{ij})$ enters as an offset.

A major advantage of this model structure is that it lends itself easily to include TDC, as a covariate can change its value in each interval. Alternatively, the interval cut-points could be chosen as the time-points at which a change in the TDC is recorded. Then (4) can be extended to $\log(\lambda_i(t|\mathbf{x}_{ij})) = \log(\lambda_j) + \mathbf{x}_{ij}'\boldsymbol{\beta}$. Additionally, time-varying effects can be incorporated by creating a TDC for time itself, e.g. by using the interval midpoints $\tilde{t} := (\kappa_j - \kappa_{j-1})/2$, and including interaction terms of selected covariates with time $\tilde{t}$ in the linear predictor.

2.2 Piece-wise Exponential Additive Model

Transitioning from the framework of GLMs to the framework of generalized additive mixed models (GAMM), model (4) can be further extended to include smoothly time-varying effects of time-constant and time-dependent covariates. For the sake of notational simplicity, here we present a model with one TDC. An extension to multiple ELRAs, however, is straight forward. In reference to the idioms known for piece-wise exponential models (PEM) and generalized additive models (GAM), we will refer to this model class as PAM. We first present the general model specification and discuss individual terms in subsequent sections.

Let $\mathcal{Z}_i(t)$ denote a subset of past exposures that affect the hazard at time t (cf. section 2.2.3 for more details), $\ell = 1, \dots, L$ the index for different clusters and ℓ_i the cluster associated with subject i.

We model the hazard rate λ at time t for individual i from cluster ℓ as:

$$\log\left(\lambda_i(t|\mathbf{x}_i, \boldsymbol{z}_i, \ell_i)\right) = f_0(t) + \sum_{p=1}^{P} f_p(x_i^p, t) + g(\mathcal{Z}_i(t), t) + b_{\ell_i} \qquad (5)$$

where

- $f_0(t)$ represents the baseline hazard rate (cf. section 2.2.1),
- $f_p(x_i^p, t)$, $p = 1, \dots, P$, are potentially smooth, smoothly time-varying effects (cf. section 2.2.2) of time-constant confounders $\mathbf{x}^p$,
- $g(\mathcal{Z}_i(t), t)$ denotes the *exposure-lag-response association* and will be discussed in detail in section 2.2.3.

- b_{ℓ_i} is a Gaussian random effect (frailty) for subject i.

2.2.1 Baseline hazard

In the original definition of PEMs (4), the baseline hazard is a step function and interval-specific hazards λ_j are estimated by including dummy variables for the individual intervals in the model matrix. One problem with this approach is the, more or less, arbitrary choice of interval cut-points (Demarqui et al., 2008), which affects the estimation of interval-specific baseline hazards λ_j. By representing the baseline hazard as a regression spline over the interval mid-points $\tilde{t}$, we can ameliorate this issue. Given a sufficiently large number of knots, the hazard can be estimated flexibly, while overfitting is avoided due to penalization (cf. section 2.3). As hazards in clinical studies tend to change quickly in the beginning of the follow up and become more stable towards the end of the observation period, adaptive spline smooths (Wood, 2011, section 5.1) can be employed to allow the smoothness of the function to vary over time.

2.2.2 Smooth, smoothly time-varying effects

The summands $f_p(x_i^p, t)$ in the second term in (5) represent possibly nonlinear, possibly time-varying effects of time-constant covariates. In the simplest case, when effects are assumed to be linear and not time-varying, this would reduce to a linear effect $x_i^p \beta_p$. Time-varying effects are modeled as interaction terms between the variable of interest $\mathbf{x}^p$ and time t. Table 1 shows possible representations of time-varying effects. Depending on the specification of the interaction term, flexibility can increase from linear effects with linear time-variation $\beta_p x_i^p + \beta_{p:t}(x_i^p \cdot t)$, to varying coefficients $x_i^p f_p(t)$ or $f_p(x_i^p)t$ (Hastie and Tibshirani, 1993), to nonlinear, smoothly time-varying covariate effects $f_p(x_i^p, t)$ modeled as bivariate function surfaces, e.g. tensor product smooths (Wood et al., 2012). The smooth functions $f_p(\cdot)$ can be represented as splines of the form $\sum_{m=1}^{M} \gamma_m^p B_m^p(\cdot)$, where B_m^p are covariate specific basis functions. The specification $x_i^p f_p(t)$ is particularly useful when $\mathbf{x}_p$ is a dummy variable coding for a certain level of a categorical variable, in which case a smoothly time-varying effect $f_p(t)$ is estimated for each category. One possible application is the evaluation of the effects of different treatment arms in clinical trials, when the proportional hazards assumption is not fulfilled. Specification $f_p(x_i^p, t)$ is the most flexible and should be employed whenever prior information or domain specific knowledge regarding the relationship is absent.

However, this latter option is also the most computationally demanding.

Effect specification	Description
$\beta_p x_i^p + \beta_{p:t}(x_i^p \cdot t)$:	Linear, linearly time-varying effect
$f_p(x_i^p) \cdot t$:	Smooth, linearly time-varying effect
$x_i^p \cdot f_p(t)$:	Linear, smoothly time-varying effect
$f_p(x_i^p, t)$:	Smooth, smoothly time-varying effect

Table 1: Overview of possible time-varying effect specifications.

In general, due to the model definition and respective estimation routine, the number of parameters to be estimated needs to be considerably lower than the number of subjects n under study. In addition, depending on the number of such components and their specification, identifiability issues may arise, especially since, in contrast to "standard" additive regression models, time t will typically appear in multiple model terms in PAMs (5).

2.2.3 Exposure-lag-response Associations

For the specification of the ELRA $g(\mathcal{Z}_i(t), t)$ in (5) it is important to distinguish between time at risk t and time of exposure t_e, i.e. the time at which the hazard is evaluated and the time at which the value of the TDC is observed, respectively.

Let $z_i(t_e)$ denote the value of the TDC at *exposure*-time t_e. To model the time-varying, cumulative effects of exposure histories $\mathcal{Z}_i(t)$, we:

1. Define a time window $\mathcal{T}(j)$ of exposure-times t_e for which the time-dependent covariate $z(t_e)$ is assumed to affect survival in interval j, such that the exposure-history affecting the hazard at time t is defined by
$$\mathcal{Z}_i(t) := \{z_i(t_e) : t_e \in \mathcal{T}(j)\}. \tag{6}$$
This window can be specified by setting variables t_{lag} (delay before exposure at time t_e can affect hazard) and t_{lead} (maximal time after $t_e + t_{\text{lag}}$ after which the exposure still affects the hazard), such that $t_e \in \mathcal{T}(j)$ if $\kappa_{j-1} < t_e + t_{\text{lag}} + t_{\text{lead}}$ and $\kappa_j \geq t_e + t_{\text{lag}}$.

2. Specify the shape of partial effects $g(z_i(t_e), t)$ representing the ELRA
$$g(\mathcal{Z}_i(t), t) = \int_{t_e \in \mathcal{T}(j)} g(z_i(u), t)\mathrm{d}u \approx \sum_{k: t_{ek} \in \mathcal{T}(j)} \Delta_k g(z_i(t_{ek}), t), \tag{7}$$
with $\Delta_k = t_{ek} - t_{e(k-1)}$ the time between two consecutive exposures.

We represent the relationship as a bivariate smooth function in t_e and t

$$g(z_i(t_e), t) = f(t_e, t) \cdot w_{ij}, \tag{8}$$

where

$$w_{ij} = \begin{cases} z_i(t_e) & \text{if } t_e \in \mathcal{T}(j) \\ 0 & \text{else,} \end{cases} \tag{9}$$

and

$$f(t_e, t) = \sum_{m=1}^{M} \sum_{k=1}^{K} \gamma_{mk} B_m(t_e) B_k(t) = \sum_{m,k} \gamma_{mk} B_{mk}(t_e, t) \tag{10}$$

is modeled as a tensor product spline smooth, with marginal bases $B_m(\cdot)$, $B_k(\cdot)$ evaluated at the respective values of t_e and t, $B_{mk}(\cdot,\cdot) = B_m(\cdot)B_k(\cdot)$, and spline coefficients γ_{mk} controlling the shape of $f(t_e, t)$. The penalized estimation of the smooth terms (cf. section 2.3) implies the assumption of smoothness for $f(t_e, t)$, which ensures that effects of exposures on consecutive time points t_e, t'_e are similar and that effects of exposure $z(t_e)$ on the hazards in neighboring intervals j, j' are similar as well.

Note that the information regarding the amount of exposure $z_i(t_e)$ is not included in the construction of the marginal bases $B(\cdot)$. This information is added to the design matrix through weights (9) w_{ij}, specified beforehand (and therefore known). The leads and lags are also specified using these weights, setting the partial effects for exposures outside the relevant window $\mathcal{T}(j)$ to zero.

The above specification of the ELRA implies that effects of the TDC are smooth regarding the timing of exposure t_e and their effect over time t but not with respect to the value of $z_i(t_e)$, which enters linearly. An extension of the presented framework to non-linear ELRAs via three-dimensional smooths of the form $f(t_e, t, z_i(t_e))$ is straight forward (Wood, 2006, sec. 4.1.8), but was not pursued in this work.

2.3 Estimation and Inference

Stable likelihood-based methods for the parameter estimation of the proposed model have been recently developed in Wood (2011) in the context of penalized models of the form $D(\boldsymbol{\gamma}) + \sum_p \lambda_p \boldsymbol{\gamma}' \boldsymbol{K}_p \boldsymbol{\gamma}$, where $D(\boldsymbol{\gamma})$ is the model deviance, $\boldsymbol{\gamma}$ contains all spline basis coefficients representing model (5), and λ_p and K_p are the smoothing parameters and penalty matrices for the individual smooths $f_p(\cdot)$, respectively. Given $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_p)$, parameter estimates can be obtained by penalized iteratively reweighted least squares (P-IRLS). To guarantee convergence, Wood (2011) employs P-IRLS based on nested iterations, i.e. after each P-IRLS step, estimation of $\boldsymbol{\lambda}$ is updated given the current $\boldsymbol{\gamma}$ estimates.

In subsequent papers Marra and Wood (2011, 2012), Wood (2013) develop shrinkage based procedures for simultaneous smoothness and variable selection and methods for confidence intervals and significance tests for smooth components, which can largely be applied to the context of PAM.

For example, confidence intervals (CI) with good coverage properties for smooth terms are developed in Marra and Wood (2012) and are applicable to the smooth components in (5) and particularly the ELRA (7). Let $\hat{\boldsymbol{\gamma}}_q$ the vector of parameter estimates associated with $f(t_e, t)$ in (10), and $\mathbf{V}_{\hat{\gamma}_q}$ the empirical Bayesian covariance matrix of the estimated parameters $\hat{\boldsymbol{\gamma}}_q$. Let further $\mathbf{X}^q$ the $n_J \times n_e$ design matrix for a specific exposure history $\mathcal{Z}(t)$, where n_J is the number of intervals in which the follow up period has been partitioned, and n_e is the number of columns associated with the tensor-product smooth of the ELRA term. The confidence intervals are given by

$$\mathbf{X}^q\hat{\boldsymbol{\gamma}}_q \pm z_{1-\alpha/2}\sqrt{\text{diag}(\mathbf{X}^q\mathbf{V}_{\hat{\gamma}_q}\mathbf{X}^{q^T})} = \hat{\mathbf{f}}_q \pm z_{1-\alpha/2}\widehat{\mathbf{SE}}_q \tag{11}$$

In (11), $\hat{\mathbf{f}}_q$ as well as $\widehat{\mathbf{SE}}_q$ are vectors of length n_J, representing the estimated cumulative effect and standard errors in intervals $j = 1, \ldots, J$. By defining $\mathbf{X}^q := \mathbf{X}^{q_2} - \mathbf{X}^{q_1}$ in (11) we can obtain estimated *differences* in cumulative effects (and a respective CI) given different exposure histories $\mathcal{Z}^2(t)$ and $\mathcal{Z}^1(t)$.

3 Summary

By embedding the concept of PEMs into the framework of penalized GAMMs (cf. section 2.2), we were able to establish a very flexible model class for survival analysis in general and exposure-lag-response associations (ELRA) in particular. This model class inherits the robust and flexible tools for modeling, estimation and validation of the penalized GAMMs, as discussed in section 2. In comparison to the classical PEM, major advantages include the semi-parametric, possibly adaptive, estimation of the baseline hazard, which ameliorates the problem of *arbitrary* choice of cut-points (cf. section 2.2.1), and the smooth, penalized estimation of time-varying effects.

References

Berhane, K., M. Hauptmann, and B. Langholz (2008). Using tensor product splines in modeling exposure-time-response relationships: Application to the colorado plateau uranium miners cohort. *Statistics in Medicine 27*(26), 5484–5496.

Demarqui, F. N., R. H. Loschi, and E. A. Colosimo (2008). Estimating the grid of time-points for the piecewise exponential model. *Lifetime Data Analysis 14*(3), 333–356.

Friedman, M. (1982). Piecewise exponential models for survival data with covariates. *The Annals of Statistics 10*(1), 101–113.

Gasparrini, A. (2014). Modeling exposure-lag-response associations with distributed lag non-linear models. *Statistics in Medicine 33*(5), 881–899.

Hastie, T. and R. Tibshirani (1993). Varying-coefficient models. *Journal of the Royal Statistical Society. Series B (Methodological) 55*(4), 757–796.

Holford, T. R. (1980). The analysis of rates and of survivorship using log-linear models. *Biometrics 36*(2), 299–305.

Laird, N. and D. Olivier (1981). Covariance analysis of censored survival data using log-linear analysis techniques. *Journal of the American Statistical Association 76*(374), 231–240.

Marra, G. and S. N. Wood (2011). Practical variable selection for generalized additive models. *Computational Statistics & Data Analysis 55*(7), 2372–2387.

Marra, G. and S. N. Wood (2012). Coverage properties of confidence intervals for generalized additive model components. *Scandinavian Journal of Statistics 39*(1), 53–74.

Sylvestre, M.-P. and M. Abrahamowicz (2009). Flexible modeling of the cumulative effects of time-dependent exposures on the hazard. *Statistics in Medicine 28*(27), 3437–3453.

Wood, S. N. (2006). *Generalized additive models: An introduction with R.* Boca Raton and FL: Chapman & Hall/CRC.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal* *of the Royal Statistical Society: Series B (Statistical Methodology) 73*(1), 3–36.

Wood, S. N. (2013). On p-values for smooth components of an extended generalized additive model. *Biometrika 100*(1), 221–228.

Wood, S. N., F. Scheipl, and J. J. Faraway (2012). Straightforward intermediate rank tensor product smoothing in mixed models. *Statistics and Computing 23*(3), 341–360.

Xiao, Y., M. Abrahamowicz, E. E. M. Moodie, R. Weber, and J. Young (2014). Flexible marginal structural models for estimating the cumulative effect of a time-dependent treatment on the hazard: Reassessing the cardiovascular risks of didanosine treatment in the swiss hiv cohort study. *Journal of the American Statistical Association 109*(506), 455–464.