Biostatistics (2014), **0**, 0, *pp.* 1–13
doi:10.1093/biostatistics/Supplement_Mat_revision-1

Supplementary material for A penalized algorithm for event-specific rate models for recurrent events

O. Bouaziz

MAP5, UMR CNRS 8145

University Paris Descartes,

Paris, France,

olivier.bouaziz@parisdescartes.fr

and

A. Guilloux

LSTA, University Pierre et Marie Curie,

4 place Jussieu,

Paris, France,

agathe.guilloux@upmc.fr

1. Discussion of Assumptions 1 and 2

In the main paper, it is said that the processes Y_i^s are almost surely null on the complementary of A_s in $[0, \tau]$. This is easily shown by the inequality

$$\mathbb{E}[Y^s(t)] = \mathbb{P}[D \geqslant t, C \geqslant t, \tilde{N}(t-) = s-1] \leqslant \mathbb{P}[D \geqslant t, \tilde{N}(t-) = s-1],$$

since the right term is null on the complementary of A_s in $[0, \tau]$. This explains why the domains of integration of the limiting distribution of our estimators can equivalently be on A_s^τ or $[0, \tau]$

(see the results in Theorems 3.1 and 3.2).

We now present two sufficient conditions for Assumptions 1 and 2 to hold.

Condition 1 For all t in $[0,\tau]$, $\forall x \in \mathrm{Supp}(X(t))$, $\mathbb{P}[C \geqslant t | X(t) = x] \geqslant c > 0$ and $\mathbb{P}[E(B) \leqslant \tau] > 0$.

Condition 2 For all $t \in [0,\tau]$, C and $(D, dN^*(t), N^*(t-))$ are independent conditionally to $X(t)$.

On the one hand, under these conditions, for all $i = 1,\ldots,n$, and all $s = 1,2,\ldots,B$, one has

$$\begin{aligned}\mathbb{E}[Y^s(t)] &= \mathbb{E}\big[\mathbb{P}[D \geqslant t, C \geqslant t, \tilde{N}(t-) = s-1 | X(t)]\big] \\ &= \mathbb{E}\big[\mathbb{P}[D \geqslant t, \tilde{N}(t-) = s-1 | X(t)]\,\mathbb{P}[C \geqslant t | X(t)]\big] \geqslant c\,\mathbb{P}[D \geqslant t, \tilde{N}(t-) = s-1],\end{aligned}$$

which implies that for all $i = 1,\ldots,n$, and all $s = 1,2,\ldots,B$, the process $Y^s(t)$ is non null with positive probability for $t \in A_s^\tau$. On the other hand, Condition 2 trivially implies Assumption 2 of the main paper.

2. Proofs

Proofs of Theorems 3.1 and 3.2 of the main paper rely on the following lemmas.

A key relation

Lemma 2.1 Under Assumption 2, for all $i = 1,\ldots,n$ and all $s = 1,\ldots,B$

$$\mathbb{E}\big[Y_i^s(t)dN_i(t) \mid X_i(t), T_i \geqslant t, N_i(t-) = s-1\big] = Y_i^s(t)\rho_0(t,s,X_i(t))dt.$$

Proof. By its definition, $Y_i^s(t)dN_i(t) = Y_i^s(t)dN_i^*(t)$ and as a consequence

$$\begin{aligned}&\mathbb{E}[Y_i^s(t)dN_i(t) \mid X_i(t), T_i \geqslant t, N_i(t-) = s-1] = \mathbb{E}[Y_i^s(t)dN_i(t) \mid X_i(t), T_i \geqslant t, N_i^*(t-) = s-1] \\ &= \mathbb{E}[Y_i^s(t)dN_i^*(t) \mid X_i(t), T_i \geqslant t, N_i^*(t-) = s-1] = Y_i^s(t)\rho_0(t,s,X_i(t))dt,\end{aligned}$$

where the last equality comes from Assumption 2. □

Decomposition of the least squares criterion in the additive model

The next proposition gives the details of the construction of the partial least squares in the additive model. One has to notice that the processes $Z_n(s)$ introduced below are centered which implies that finding a minimizer of L_n^{PLS} is a natural way of estimating β_0 in the additive model.

LEMMA 2.2 In the additive event-specific model (2.2), the partial least squares criterion (2.5) can be rewritten as

$$L_n^{PLS}(\beta) = \sum_{s=1}^{B} \left\{ \beta(s)^\top \mathbf{H}_n(s)\beta(s) - 2\beta(s)^\top \mathbf{H}_n(s)\beta_0(s) - 2Z_n(s)\beta(s) \right\}, \tag{2.1}$$

where

$$Z_n(s) = \frac{1}{n}\sum_{i=1}^{n}\sum_{s=1}^{B}\int_{[0,\tau]} \{X_i(t) - \bar{X}^s(t)\} Y_i^s(t) dM_i^s(t).$$

Proof. From the definition of the partial least-squares criterion, write

$$\begin{aligned} L_n^{PLS}(\beta) = {} & \frac{1}{n}\sum_{i=1}^{n}\sum_{s=1}^{B}\int_{[0,\tau]} \{X_i(t)\beta(s) - \bar{X}^s(t)\beta(s)\}^2 Y_i^s(t)\, dt \\ & - \frac{2}{n}\sum_{i=1}^{n}\sum_{s=1}^{B}\int_{[0,\tau]} \{X_i(t)\beta(s) - \bar{X}^s(t)\beta(s)\} Y_i^s(t) dN_i(t). \end{aligned} \tag{2.2}$$

Apply Lemma 2.1 in the equation $Y_i^s(t)dM_i^s(t) = Y_i^s(t)(dN_i(t) - \mathbb{E}\big[dN_i(t) \mid X_i(t), T_i \geqslant t, N_i(t-) = s-1\big])$ and conclude the proof using the relation

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{s=1}^{B}\int_{[0,\tau]} \{X_i(t)\beta(s) - \bar{X}^s(t)\beta(s)\}\{\alpha_0(t,s) + \bar{X}^s(t)\beta_0(s)\} Y_i^s(t)\, dt = 0.$$

□

A technical lemma

LEMMA 2.3 Let $\mathcal{D}[0,\tau]$ denotes the set of càdlàg functions on $[0,\tau]$ and let $F_n(\cdot, s)$ and $f(T, \delta, X(\cdot), N(\cdot), s)$ be two random processes of bounded variation on $[0,\tau]$. Suppose that for all z in $[0,\tau]$,

$$\mathbb{E}\left[\Big(\int_0^z f(T,\delta,X(t),N(t),s) dM^s(t)\Big)^2\right] < \infty.$$

We then have the following properties:

(i) If $f(T,\delta,X(\cdot),N(\cdot),s)$ is a random variable of bounded variation on $[0,\tau]$, then

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^z f(T_i,\delta_i,X_i(t),N_i(t),s)dM_i^s(t)$$

converges weakly in $\mathcal{D}[0,\tau]$ to a centered gaussian process with variance equal to

$$\mathbb{E}\left[\Big(\int_0^z f(T,\delta,X(t),N(t),s)dM^s(t)\Big)^2\right].$$

(ii) If $\sup_{t\in[0,\tau]}|F_n(t,s)-F(t,s)|=o_{\mathbb{P}}(1)$, where $F(\cdot,s)$ is a random process on $[0,\tau]$, then

$$\sup_{z\in[0,\tau]}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^z (F_n(t,s)-F(t,s))f(T_i,\delta_i,X_i(t),N_i(t),s)dM_i^s(t)\right\}=o_{\mathbb{P}}(1).$$

Proof. Since a function of bounded variation can be decomposed into the difference between two nondecreasing functions, the process $1/\sqrt{n}\sum_i\int_0^z f(T_i,\delta_i,X_i(t),N_i(t),s)dM_i^s(t)$ can be written as the difference between two nondecreasing càdlàg empirical processes. Therefore, (i) follows from example 2.11.16 in van der Vaart and Wellner (1996). To prove (ii), write

$$\sup_{z\in[0,\tau]}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^z (F_n(t,s)-F(t,s))f(T_i,\delta_i,X_i(t),N_i(t),s)dM_i^s(t)\right\}$$
$$\leqslant \sup_{t\in[0,\tau]}|F_n(t,s)-F(t,s)|\sup_{z\in[0,\tau]}\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^z f(T_i,\delta_i,X_i(t),N_i(t),s)dM_i^s(t)\right\}.$$

The second term is a $O_{\mathbb{P}}(1)$ from (i) of this lemma, while the first term converges to 0 in probability.

□

Proof of Theorem 3.1

Proof of 1. Let $\Gamma_n^{add}(\beta)$ be the quantity minimized by $\hat\beta_{\mathrm{TV}/add}$ and introduce $\Gamma_{add}(\beta)=\sum_{s=1}^{B}\left[\beta(s)^\top\mathbf{H}(s)\beta(s)-2\boldsymbol{h}(s)\beta(s)\right]$ where

$$\boldsymbol{h}(s):=\int_{A_s^\tau}\mathbb{E}\left[Y^s(t)X(t)dN(t)\right]-\int_{A_s^\tau}\frac{\mathbb{E}[Y^s(t)X(t)]}{\mathbb{E}[Y^s(t)]}\mathbb{E}[Y^s(t)dN(t)].$$

Using Lemma 2.1 notice that $\boldsymbol{h}(s) = \beta_0(s)^\top \mathbf{H}(s)$ and consequently, $\text{argmin}_\beta \Gamma_{add} = \beta_0$. Since the criterion to minimize is convex, the convergence in probability of $\hat{\beta}_{\text{TV}/add}$ to β_0 follows from the pointwise convergence of $\Gamma_n^{add}(\beta)$ towards $\Gamma_{add}(\beta)$. Now write:

$$\begin{aligned}\left|\Gamma_n^{add}(\beta) - \Gamma_{add}(\beta)\right| \quad &\leqslant \left|L_n^{PLS}(\beta) - \Gamma(\beta)\right| + \frac{\lambda_n}{n} Bp \max_{s,j} |\beta^j(s) - \beta^j(s-1)| \\ \leqslant Bp^2 \max_{j,k,s} |\beta^j(s)\beta^k(s)(\mathbf{H}_n^{j,k}(s) - \mathbf{H}^{j,k}(s))| &+ 2Bp \max_{j,s} |\boldsymbol{h}_n^j(s) - \boldsymbol{h}^j(s)||\beta^j(s)| + \frac{\lambda_n}{n} Bp\end{aligned}$$

and the result follows from the law of large numbers and the fact that $\lambda_n/n \to 0$ as n tends to infinity.

Proof of 2. Define

$$\Lambda_n^{add}(u) = \sum_{s=1}^{B} u(s)^\top \mathbf{H}_n(s) u(s) - 2\sqrt{n} \sum_{s=1}^{B} Z_n(s) u(s) + \lambda_n \sum_{j=1}^{p} \Big(\text{TV}(\beta_0^j + u^j/\sqrt{n}) - \text{TV}(\beta_0^j) \Big)$$

and notice that $\Lambda_n^{add}(u)$ is minimum at $u = \sqrt{n}(\hat{\beta}_{\text{TV}/add} - \beta_0)$. For $s = 1, \dots, B$, write

$$\begin{aligned}\sqrt{n} Z_n(s) u(s) = &\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_{A_s^\tau} \left(X_i(t) - \frac{\mathbb{E}[Y^s(t)X(t)]}{\mathbb{E}[Y^s(t)]} \right) u(s) Y_i^s(t) dM_i^s(t) \\ &- \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_{A_s^\tau} \left(\bar{X}^s(t) - \frac{\mathbb{E}[Y^s(t)X(t)]}{\mathbb{E}[Y^s(t)]} \right) u(s) Y_i^s(t) dM_i^s(t).\end{aligned}$$

Let $F_n(t,s) = (\bar{X}^s(t) - \mathbb{E}[Y^s(t)X(t)]/\mathbb{E}[Y^s(t)])u(s)\mathbf{1}(t \in A_s^\tau)$ and $F(t,s) = 0$. $F_n(t,s)$ has bounded variation for t in $[0,\tau]$ and from Lemma 2.3 (ii), the second term converges to 0 in probability. Now, take $f(T_i, \delta_i, X_i(t), N_i(t), s) = (X_i(t) - \mathbb{E}[Y^s(t)X(t)]/\mathbb{E}[Y^s(t)])u(s)Y_i^s(t)\mathbf{1}(t \in A_s^\tau)$ which is also a function of bounded variation for t in $[0,\tau]$. From Lemma 2.3 (i), the first term converges weakly towards a centered gaussian variable with variance equal to

$$\begin{aligned}&\mathbb{E}\left[\Big(\int_{A_s^\tau} (X(t) - \mathbb{E}[Y^s(t)X(t)]/\mathbb{E}[Y^s(t)]) u(s) Y^s(t) dM^s(t) \Big)^2 \right] \\ &\quad = u(s)^\top \mathbb{E}\left[\Big(\int_{A_s^\tau} (X(t) - \mathbb{E}[Y^s(t)X(t)]/\mathbb{E}[Y^s(t)]) Y^s(t) dM^s(t) \Big)^{\otimes 2} \right] u(s).\end{aligned}$$

Then, note that $\sum_{s=1}^{B} u(s)^\top \mathbf{H}_n(s) u(s)$ converges to $\sum_{s=1}^{B} u(s)^\top \mathbf{H}(s) u(s)$, in probability and

$\lambda_n \sum_j \left(\text{TV}(\beta_0^j + u^j/\sqrt{n}) - \text{TV}(\beta_0^j)\right)/\lambda_0$ converges to

$$\sum_{j=1}^{p}\sum_{s=2}^{B}\left\{|\Delta u^j(s)|\mathbf{1}(\Delta\beta^j(s)=0) + \text{sgn}(\Delta\beta_0^j(s))(\Delta u^j(s))\mathbf{1}(\Delta\beta^j(s)\neq 0)\right\}.$$

Thus $\Lambda_n^{add}(u)$ converges to $\Lambda_{add}(u)$ in distribution. Since Λ_n^{add} is convex and Λ_{add} has a unique minimum, it follows that $\sqrt{n}(\hat{\beta}_{\text{TV}/add} - \beta_0)$ converges to $\text{argmin}_u \Lambda_{add}(u)$ in distribution.

Proof of Theorem 3.2

First define for $l = 0, 1$ or 2

$$S_n^{(l)}(s,t,\beta) = \frac{1}{n}\sum_{i=1}^{n} Y_i^s(t) X_i(t)^{\otimes l}\exp(X_i(t)\beta(s)).$$

Following the arguments in example VII.2.7 page 502 of Andersen et al. (1993), it can easily be shown that

$$\sup_{t\in A_s^\tau} |S_n^{(l)}(s,t,\beta_0) - s^{(l)}(s,t,\beta_0)| \xrightarrow[n\to\infty]{\mathbb{P}} 0,\ \forall\, l = 0,1,2,$$

using the fact that the covariate process is of bounded variation (in particular, this assumption guarantees that $s^{(l)}(s,t,\beta_0)$ has a countable number of jumps).

Proof of 1. Let $\Gamma_n^{mult}(\beta)$ be the quantity minimized by $\hat{\beta}_{\text{TV}/mult}$ and introduce

$$\begin{aligned}\Gamma_{mult}(\beta) &= -\sum_{s=1}^{B}\int_{A_s^\tau}\mathbb{E}\left[X(t)\beta(s)Y^s(t)dN(t)\right] + \sum_{s=1}^{B}\int_{A_s^\tau}\log(s^{(0)}(s,t,\beta))\mathbb{E}\left[Y^s(t)dN(t)\right]\\ &= -\sum_{s=1}^{B}\int_{A_s^\tau}\alpha_0(t,s)\left(s^{(1)}(s,t,\beta_0)\beta(s) - \log(s^{(0)}(s,t,\beta))s^{(0)}(s,t,\beta_0)\right)dt,\end{aligned}$$

where the last equality follows from Lemma 2.1. From similar arguments as in proof 1. of Theorem 3.1 and the uniform convergence with respect to t of $S_n^{(0)}(s,t,\beta_0)$ towards $s^{(0)}(s,t,\beta_0)$, we get the pointwise convergence in probability of $\Gamma_n^{mult}(\beta)$ to $\Gamma_{mult}(\beta)$. Then, the consistency of $\hat{\beta}_{\text{TV}/mult}$ follows from the convexity of $\Gamma_n^{mult}(\beta)$ and the fact that $\text{argmin}_\beta\,\Gamma_{mult}(\beta) = \beta_0$.

Proof of 2. Consider the convex function

$$\Lambda_n^{mult}(u) = n\Gamma_n(\beta_0 + u/\sqrt{n}) - n\Gamma_n(\beta_0) + \lambda_n\sum_{j=1}^{p}\left(\text{TV}(\beta_0^j + u^j/\sqrt{n}) - \text{TV}(\beta_0^j)\right)$$

which is minimum at $u = \sqrt{n}(\hat{\beta}_{\text{TV}/mult} - \beta_0)$. Then from a Taylor expansion, one gets

$$\Lambda_n^{mult}(u) = -\frac{\sqrt{n}}{n}\sum_{s=1}^{B}\sum_{i=1}^{n}\int_{[0,\tau]}\left(X_i(t) - \mathbf{E}_n(s,t,\beta_0)\right)Y_i^s(t)dN_i(t)u(s)$$
$$+\frac{1}{2n}\sum_{s=1}^{B}u(s)^\top\sum_{i=1}^{n}\int_{[0,\tau]}\mathbf{V}_n(s,t,\beta_0)Y_i^s(t)dN_i(t)u(s) + \lambda_n\sum_{j=1}^{p}\Big(\text{TV}(\beta_0^j + u^j/\sqrt{n}) - \text{TV}(\beta_0^j)\Big) + o_{\mathbb{P}}(1),$$

where

$$\mathbf{E}_n(s,t,\beta) = \frac{S_n^{(1)}(s,t,\beta)}{S_n^{(0)}(s,t,\beta)}, \quad \mathbf{V}_n(s,t,\beta) = \frac{S_n^{(2)}(s,t,\beta)}{S_n^{(0)}(s,t,\beta)} - \mathbf{E}_n(s,t,\beta)^{\otimes 2}.$$

The uniform convergence with respect to t of $S_n^{(0)}(s,t,\beta)$ and $S_n^{(2)}(s,t,\beta)$ towards $s^{(0)}(s,t,\beta_0)$ and $s^{(2)}(s,t,\beta_0)$, respectively, and the law of large numbers give the convergence in probability of the term

$$\frac{1}{2n}\sum_{s=1}^{B}u(s)^\top\sum_{i=1}^{n}\int_{[0,\tau]}\mathbf{V}_n(s,t,\beta_0)Y_i^s(t)dN_i(t)u(s)$$

towards

$$\frac{1}{2}\sum_{s=1}^{B}u(s)^\top\int_{[0,\tau]}\mathbf{v}(s,t,\beta_0)\mathbb{E}[Y^s(t)dN(t)]u(s).$$

Notice that

$$\sum_{i=1}^{n}\left(X_i(t) - \mathbf{E}_n(s,t,\beta_0)\right)Y_i^s(t)\alpha_0(t,s)\exp(X(t)\beta_0(t))dt = 0$$

in order to rewrite the first term of $\Lambda_n^{mult}(u)$ as

$$-\frac{\sqrt{n}}{n}\sum_{s=1}^{B}\sum_{i=1}^{n}\int_{[0,\tau]}\left(X_i(t) - \mathbf{E}_n(s,t,\beta_0)\right)Y_i^s(t)dM_i^s(t)u(s).$$

From Lemma 2.3, the same kind of arguments as in the proof of Theorem 3.1 can be applied to conclude the proof.

3. Characterization of the asymptotic distribution in a particular case

In Theorem 3.1, we state that if $\lambda_n/\sqrt{n} \to \lambda_0 \geqslant 0$ as $n \to \infty$ then $\sqrt{n}(\hat{\beta}_{\mathrm{TV}/add} - \beta_0)$ converges in distribution to

$$\begin{aligned}\underset{u\in\mathbb{R}^{p\times B}}{\operatorname{argmin}}\, \Lambda_{add}(u) &= \underset{u\in\mathbb{R}^{p\times B}}{\operatorname{argmin}} \Big[\sum_{s=1}^{B} \big\{ u(s)^\top \mathbf{H}(s) u(s) - 2u(s)^\top \xi_{add}(s) \big\} \\ &+ \lambda_0 \sum_{j=1}^{p}\sum_{s=2}^{B} \big\{ |\Delta u^j(s)| \mathbf{1}(\Delta\beta^j(s) = 0) + \operatorname{sgn}(\Delta\beta^j(s)) \Delta u^j(s) \mathbf{1}(\Delta\beta^j(s) \neq 0) \big\} \Big], \end{aligned} \tag{3.3}$$

and for each s, $\xi_{add}(s)$ is defined in Theorem 3.1.

The minimum in Equation (3.3) equals, after permutation, the optimum

$$\begin{aligned}\underset{v\in\mathbb{R}^{p\times B}}{\operatorname{argmin}}\, \Lambda_{add}(v) &= \underset{v\in\mathbb{R}^{p\times B}}{\operatorname{argmin}} \Big[\Big\{ v^\top \mathbf{HH} v - 2v^\top \xi\xi add^\top \Big\} \\ &+ \lambda_0 \sum_{j=1}^{p}\sum_{s=2}^{B} \big\{ |\Delta v^j(s)| \mathbf{1}(\Delta\beta^j(s) = 0) + \operatorname{sgn}(\Delta\beta^j(s)) \Delta v^j(s) \mathbf{1}(\Delta\beta^j(s) \neq 0) \big\} \Big], \end{aligned} \tag{3.4}$$

where, in v, the coordinates are arranged by covariates, then by recurrent event number

$$v = (v^1(1), \dots, v^1(B), v^2(1), \dots, v^2(B), \dots, v^p(1), \dots, v^p(B))^\top,$$

and $\mathbf{HH}$ and $\xi\xi_{add}$ are the matrix and vector obtained after the same permutation applied to, respectively, the block diagonal matrix $\operatorname{diag}\big(\mathbf{H}(1), \dots, \mathbf{H}(B)\big)$ and the vector $\big(\xi_{add}^\top(1), \dots, \xi_{add}^\top(B)\big)^\top$. In the case where the matrices $\mathbf{H}(s)$ are diagonal for all $s = 1, \dots, B$, the matrix $\mathbf{HH}$ is also diagonal and the search for the argument minimum can be separated in each covariate:

$$\begin{aligned}\underset{v\in\mathbb{R}^{p\times B}}{\operatorname{argmin}}\, \Lambda_{add}(v) &= \underset{v\in\mathbb{R}^{p\times B}}{\operatorname{argmin}} \Big[\sum_{j=1}^{p} \Big\{ {v^j}^\top \mathbf{HH}^j v^j - 2{v^j}^\top \xi\xi_{add}^j \\ &+ \lambda_0 \sum_{s=2}^{B} \big\{ |\Delta v^j(s)| \mathbf{1}(\Delta\beta^j(s) = 0) + \operatorname{sgn}(\Delta\beta^j(s)) \Delta v^j(s) \mathbf{1}(\Delta\beta^j(s) \neq 0) \big\} \Big\} \Big], \end{aligned}$$

where $\mathbf{HH}^j$ is the jth block in $\mathbf{HH}$ and $\xi\xi^j$ is the jth block of B coordinates in $\xi\xi$. Hence it suffices to make explicit the solution v^* for one j in $\{1, \dots, p\}$ of:

$$\underset{v\in\mathbb{R}^{B}}{\operatorname{argmin}} \Big[v^\top \mathbf{HH}^j v - 2v^\top \xi\xi_{add}^j + \lambda_0 \sum_{s=2}^{B} \big\{ |\Delta v(s)| \mathbf{1}(\Delta\beta^j(s) = 0) + \operatorname{sgn}(\Delta\beta^j(s)) \Delta v(s) \mathbf{1}(\Delta\beta^j(s) \neq 0) \big\} \Big\} \Big].$$

Now write $T_B\gamma = v$, where T_B is the $B \times B$ lower triangular matrix defined on page 8 of the main paper. We then have $v^* = T_B\gamma^*$ and γ^* is the argument minimum of

$$\gamma^\top T_B^\top \mathbf{H}\mathbf{H}^j T_B \gamma - 2\gamma^\top T_B^\top \xi\xi_{add}^j + \lambda_0 \sum_{s=2}^{B} \left\{ |\gamma(s)| \mathbf{1}(\mathbf{\Delta}\beta^{\mathbf{j}}(\mathbf{s}) = \mathbf{0}) + \mathrm{sgn}(\Delta\beta^j(s))\gamma(s)\mathbf{1}(\mathbf{\Delta}\beta^{\mathbf{j}}(\mathbf{s}) \neq \mathbf{0}) \right\}.$$

By definition of the sub differential, the solution γ^* verifies

$$\begin{aligned}\vec{0} \in\ & 2T_B^\top \mathbf{H}\mathbf{H}^j T_B\gamma^* - 2T_B^\top \xi\xi_{add}^j + \lambda_0 \Big(0, \partial|\gamma^*(2)|\mathbf{1}(\mathbf{\Delta}\beta^{\mathbf{j}}(\mathbf{2}) = \mathbf{0}), \ldots, \partial|\gamma^*(\mathbf{B})|\mathbf{1}(\mathbf{\Delta}\beta^{\mathbf{j}}(\mathbf{B}) = \mathbf{0})\Big)^\top \\ & + \lambda_0 \Big(0, \mathrm{sgn}(\Delta\beta^j(2)), \ldots, \mathrm{sgn}(\Delta\beta^j(B))\Big)^\top\end{aligned}$$

where $\partial|u| = \{\mathrm{sgn}(u) + \nu(1 - |\mathrm{sgn}(u)|) : \nu \in [-1, 1]\}$. This is equivalent to

$$\begin{aligned}\vec{0} \in\ & 2\mathbf{H}\mathbf{H}^j T_B\gamma^* - 2\xi\xi_{add}^j + \lambda_0 \big(T_B^\top\big)^{-1} \Big(0, \partial|\gamma^*(2)|\mathbf{1}(\mathbf{\Delta}\beta^{\mathbf{j}}(\mathbf{2}) = \mathbf{0}), \ldots, \partial|\gamma^*(\mathbf{B})|\mathbf{1}(\mathbf{\Delta}\beta^{\mathbf{j}}(\mathbf{B}) = \mathbf{0})\Big)^\top \\ & + \lambda_0 \big(T_B^\top\big)^{-1} \Big(0, \mathrm{sgn}(\Delta\beta^j(2)), \ldots, \mathrm{sgn}(\Delta\beta^j(B))\Big)^\top.\end{aligned}$$

To illustrate this example, consider the case $j = 1$ and $\beta^1 = (0, 0, b_1, b_1, 0)$ as in the simulation study (see Section 4.2) and set for the sake of simplicity, $\mathbf{H}\mathbf{H}^j = \mathbf{I}$. We have:

$$\vec{0} \in \begin{pmatrix} \gamma^*(1) - \xi\xi_{add}^1(1) - \frac{\lambda_0}{2}\partial|\gamma^*(2)| \\ \gamma^*(1) + \gamma^*(2) - \xi\xi_{add}^1(2) + \frac{\lambda_0}{2}(\partial|\gamma^*(2)| - 1) \\ \gamma^*(1) + \gamma^*(2) + \gamma^*(3) - \xi\xi_{add}^1(3) - \frac{\lambda_0}{2}(\partial|\gamma^*(4)| - 1)) \\ \gamma^*(1) + \gamma^*(2) + \gamma^*(3) + \gamma^*(4) - \xi\xi_{add}^1(4) + \frac{\lambda_0}{2}(\partial|\gamma^*(4)| + 1)) \\ \gamma^*(1) + \gamma^*(2) + \gamma^*(3) + \gamma^*(4) + \gamma^*(5) - \xi\xi_{add}^1(5) - \frac{\lambda_0}{2} \end{pmatrix}.$$

We recall that the soft-thresholding operator η_b, introduced in Donoho and Johnstone (1995) and defined as $\eta_b(a) = \mathrm{sign}(a)(|a| - b)_+$ for a in $\mathbb{R}$, verifies the equation:

$$0 \in x - a + b\partial|x| \Leftrightarrow x = \eta_b(a).$$

Therefore, $\gamma^*(2) = \eta_{\lambda_0}(\xi\xi_{add}^1(2) - \xi\xi_{add}^1(1) + \frac{\lambda_0}{2})$ and $\gamma^*(4) = \eta_{\lambda_0}(\xi\xi_{add}^1(4) - \xi\xi_{add}^1(3) - \lambda_0)$. If $\gamma^*(2) = 0$, which happens with probability $\mathbb{P}[|\xi\xi_{add}^1(2) - \xi\xi_{add}^1(1) - \lambda_0/2| \leqslant \lambda_0]$, then

$$v^*(1) = \gamma^*(1) = (\xi\xi_{add}^1(1) + \xi\xi_{add}^1(2))/2 + \lambda_0/4 = v^*(2).$$

Otherwise, with probability $\mathbb{P}[\xi\xi_{add}^1(2) - \xi\xi_{add}^1(1) - \lambda_0/2 \geqslant \lambda_0]$, $v^*(1) = \gamma^*(1) = \xi\xi_{add}^1(1) + \lambda_0/2$, $v^*(2) = \xi\xi_{add}^1(2) + \lambda_0$, etc.

4. Simulation study (continued)

4.1 *Simulation of recurrent events*

Recurrent events are simulated via an accept-reject method. First, for $i = 1, \dots, n$, $s = 1, \dots, B_{\max}$ and $k = 1, 2, \dots,$ denote by $E_i^{(k)}(s)$ the k^{th} sequence of variables, simulated in the following way:

$$E_i^{(k)}(s) = \Big(-\log(\mathcal{U}) \exp(-X_i \beta_0(s)) \Big)^{1/a_{\mathcal{W}}} \quad \text{for the Weibull case and}$$
$$E_i^{(k)}(s) = \frac{1}{a_{\mathcal{G}}} \log \Big(1 - a_{\mathcal{G}} \log(\mathcal{U}) \exp(-X_i \beta_0(s)) \Big) \quad \text{for the Gompertz case,}$$

in the event-specific multiplicative rate model and

$$E_i^{(k)}(s) = \mathcal{E}(X\beta_0(s)) \wedge \mathcal{W}(a_{\mathcal{W}}, 1) \quad \text{for the Weibull case and}$$
$$E_i^{(k)}(s) = \mathcal{E}(X\beta_0(s)) \wedge \mathcal{G}(a_{\mathcal{G}}, 1) \quad \text{for the Gompertz case,}$$

in the event-specific additive rate model. In the latter formulas, $\mathcal{U}$ is as an uniform distribution on $(0, 1)$, $\mathcal{E}(r)$ an exponential distribution with rate r, $\mathcal{W}(\phi, 1)$ a Weibull distribution with shape parameter ϕ and $\mathcal{G}(\phi, 1)$ a Gompertz distribution with shape parameter ϕ and hazard rate $\exp(\phi t)$ for $t > 0$.

For each individual, we simulated $B_{\max}$ recurrent event times, where $B_{\max}$ has been empirically determined, in such a way that the probability that

$$E_i^{(k)}(B_{\max}) \leqslant D_i \wedge C_i$$

is negligible, here D_i is the time of terminal event and C_i the censoring time of individual i.

Then, simulate the recurrent event times $E_i(s)$, $s = 1, \dots, B_{\max}$, with the following algorithm: start with k=1 and do

step 1. for all $s = 1, \dots, B_{\max}$, simulate $E_i^{(k)}(s)$ in the way described above.

step 2. if $E_i^{(k)}(1) < E_i^{(k)}(2) < \dots < E_i^{(k)}(B_{\max})$ then $E_i(s) = E_i^{(k)}(s)$, $s = 1, \dots, B_{\max}$. Otherwise, $k = k + 1$ and return to step 1.

We also explain, hereafter, why this simulation scheme gives recurrent events with the correct rate function studied in the paper. Introduce $\rho = \inf\{k : E^{(k)}(1) < E^{(k)}(2) < \ldots < E^{(k)}(B_{\max})\}$ such that $E(s) = E^{(\rho)}(s)$, $s = 1, \ldots, B_{\max}$. For $s = 1, \ldots, B_{\max}$, we have, for $t < b$:

$$
\begin{aligned}
&\mathbb{P}[E^{(\rho)}(1) < \ldots < E^{(\rho)}(s-1) < t \leqslant E^{(\rho)}(s) < b, X(t), D \geqslant t] \\
&\quad = \sum_{k=1}^{\infty} \mathbb{P}[E^{(k)}(1) < \ldots < E^{(k)}(s-1) < t \leqslant E^{(k)}(s) < b, X(t), D \geqslant t, \rho = k] \\
&\quad = \mathbb{P}[E^{(1)}(1) < \ldots < E^{(1)}(s-1) < t \leqslant E^{(1)}(s) < b, X(t), D \geqslant t] \\
&\qquad \times \sum_{k=1}^{\infty} \Big(1 - \mathbb{P}[E^{(1)}(1) < \ldots < E^{(1)}(B_{max})]\Big)^{k-1} \\
&\quad = \frac{\mathbb{P}[E^{(1)}(1) < \ldots < E^{(1)}(s-1) < t \leqslant E^{(1)}(s) < b, X(t), D \geqslant t]}{\mathbb{P}[E^{(1)}(1) < \ldots < E^{(1)}(B_{max})]}
\end{aligned}
$$

Then, applying this equality twice, once with $b = t + dt$ and once with $b = \infty$, it can easily be shown that:

$$
\begin{aligned}
&\mathbb{P}[t \leqslant E^{(\rho)}(s) < t + dt | E^{(\rho)}(1) < \ldots < E^{(\rho)}(s-1) < t \leqslant E^{(\rho)}(s), X(t), D \geqslant t] \\
&\quad = \mathbb{P}[t \leqslant E^{(1)}(s) < t + dt | E^{(1)}(1) < \ldots < E^{(1)}(s-1) < t \leqslant E^{(1)}(s), X(t), D \geqslant t].
\end{aligned}
$$

Dividing each quantity by dt and taking the limit when dt tends to 0 shows that the $E^{(\rho)}(s), s = 1 \ldots, B_{max}$ and the $E^{(1)}(s), s = 1 \ldots, B_{max}$, have the same rate function which is defined in the main paper, by (2.1) for the Cox model and by (2.2) for the Aalen model.

Finally note that in the Weibull case, there is also a direct way to simulate the recurrent events. In the multiplicative case, simulate the first recurrent event, $E_i(1) = E_i^{(1)}(1)$ as before, i.e. as a Weibull distribution with shape parameter $a_{\mathcal{W}}$ and scale parameter $\exp(X_i\beta_0(1))$. Then, for $s = 2, \ldots, B$, simulate the sth recurrent event $E_i(s)$ conditionally on the previous recurrent event as a left truncated Weibull distribution with shape parameter $a_{\mathcal{W}}$ and scale parameter $\exp(X_i\beta_0(s))$, truncated at $E_i(s-1)$. In the additive case, simulate $E_i(1) = E_i^{(1)}(1)$ as before and for $s = 2, \ldots, B$, simulate $E_i(s)$ as the law of $\mathcal{E}(X\beta_0(s)) \wedge \mathcal{LTW}(a_{\mathcal{W}}, 1)$ where $\mathcal{LTW}(a_{\mathcal{W}}, 1)$ represents the left truncated Weibull distribution with shape parameter $a_{\mathcal{W}}$ and scale parameter

1, truncated at $E_i(s-1)$. This simulation scheme is identical to the acceptation-reject simulation method in the Weibull case.

4.2 *Results for the Weibull distribution with* $a_{\mathcal{W}} = 1.5$

The results of the simulation study for the additive and multiplicative models with the baseline function following a Weibull distribution of parameter $a_{\mathcal{W}} = 1.5$ with $\mathrm{P_{obs}} = 28\%$ are presented in Tables 1-2.

4.3 *Results for the Weibull distribution with* $a_{\mathcal{W}} = 0.5$

The results of the simulation study for the additive and multiplicative models with the baseline function following a Weibull distribution of parameter $a_{\mathcal{W}} = 0.5$ are presented in Tables 3-6.

4.4 *Results for the Gompertz distribution with* $a_{\mathcal{G}} = 1.5$

The results of the simulation study for the additive and multiplicative models with the baseline function following a Gompertz distribution of parameter $a_{\mathcal{G}} = 1.5$ are presented in Tables 7-8.

4.5 *Results for the Gompertz distribution with* $a_{\mathcal{G}} = 0.5$

The results of the simulation study for the additive and multiplicative models with the baseline function following a Gompertz distribution of parameter $a_{\mathcal{G}} = 2.5$ are presented in Tables 9-10.

5. Dataset

In the bladder tumour data analysis, a small comparison of each estimator in the multiplicative model is provided using the Akaike and Bayesian information criteria. For a given β, these criteria are defined by:

$$\text{AIC}(\beta) = L(\beta) + 2k, \qquad \text{BIC}(\beta) = L(\beta) + \log(n)k,$$

where $k = 4 + \sum_j \mathbf{1}(\text{TV}(\beta^j) \neq 0)$ represents the complexity of the model and $L(\beta)$ is equal to $2nL_n^{PL}(\beta)$. Note that no such criteria have been found in the existing literature for the additive model. This is due to the additive structure of the model which makes the likelihood function difficult to work with.

From the results of Table 11, the smallest AIC is obtained from the two-step estimator which indicates a better fit from this estimator to the data set. Consequently, this shows that our estimation procedure can be applied to a given dataset to perform an automatic selection of the optimal (or near optimal) estimator in terms of AIC or BIC.

In addition, the constant, unconstrained, total variation and reweighted total-variation estimators in the additive model are displayed in Tables 12, 13 and Figure 1. It is of great interest to notice that the same number of change of variations at the same locations are obtained for the two-steps estimators in each model.

References

Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. (1993). *Statistical models based on counting processes*. Springer Series in Statistics. Springer-Verlag, New York.

Donoho, D. L. and Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.* **90,** 1200–1224.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes.* Springer Series in Statistics. Springer-Verlag, New York. With applications to statistics.

Tables

Table 1. *Simulation results in the multiplicative model for* $P_{obs} = 28\%$ *with* $a_{\mathcal{W}} = 1.5$

n	Unconstrained			Constant			TV			two-steps TV		
	MSE	SPEC	SENS	MSE	SPEC	SENS	MSE	SPEC	SENS	MSE	SPEC	SENS
50	81.440	0	1	207.515	1	0	54.730	0.269	0.803	49.269	0.584	0.672
100	30.837	0	1	199.984	1	0	32.572	0.240	0.848	26.525	0.644	0.671
500	10.466	0	1	197.787	1	0	15.744	0.225	0.894	17.381	0.979	0.515
1000	8.671	0	1	196.964	1	0	14.125	0.265	0.856	17.181	0.998	0.500

MSE: mean square error, SPEC: specificity, SENS: sensitivity.

Table 2. *Simulation results in the additive model for* $P_{obs} = 28\%$ *with* $a_{\mathcal{W}} = 1.5$

n	Unconstrained			Constant			TV			two-steps TV		
	MSE	SPEC	SENS	MSE	SPEC	SENS	MSE	SPEC	SENS	MSE	SPEC	SENS
50	923.026	0	1	429.533	1	0	267.314	0.312	1	382.825	0.605	1
100	436.022	0	1	411.630	1	0	181.608	0.231	1	242.721	0.576	1
500	195.300	0	1	408.446	1	0	138.549	0.125	1	170.339	0.508	1
1000	165.560	0	1	408.302	1	0	133.361	0.108	1	157.367	0.414	1

MSE: mean square error, SPEC: specificity, SENS: sensitivity.

Table 3. *Simulation results in the multiplicative model for* $P_{obs} = 28\%$ *with* $a_{\mathcal{W}} = 0.5$

n	Unconstrained			Constant			TV			two-steps TV		
	MSE	SPEC	SENS	MSE	SPEC	SENS	MSE	SPEC	SENS	MSE	SPEC	SENS
50	228.245	0	1	189.510	1	0	58.861	0.237	0.826	85.976	0.693	0.653
100	121.202	0	1	177.853	1	0	48.403	0.217	0.837	61.130	0.743	0.608
500	68.863	0	1	170.375	1	0	44.150	0.176	0.879	38.864	0.992	0.501
1000	64.008	0	1	169.397	1	0	41.324	0.224	0.808	37.714	1	0.500

MSE: mean squared error, FP: false positives, FN: false negatives.

Table 4. *Simulation results in the multiplicative model for* $P_{obs} = 14\%$ *with* $a_{\mathcal{W}} = 0.5$

n	Unconstrained			Constant			TV			two-steps TV		
	MSE	SPEC	SENS	MSE	SPEC	SENS	MSE	SPEC	SENS	MSE	SPEC	SENS
50	1110.171	0	1	191.820	1	0	58.199	0.250	0.813	81.072	0.658	0.633
100	145.322	0	1	180.574	1	0	47.125	0.203	0.830	63.250	0.681	0.636
500	72.530	0	1	173.950	1	0	42.608	0.191	0.889	38.608	0.984	0.505
1000	65.570	0	1	172.620	1	0	39.702	0.197	0.825	36.945	0.998	0.500

MSE: mean squared error, FP: false positives, FN: false negatives.

Table 5. *Simulation results in the additive model for* $P_{obs} = 28\%$ *with* $a_{\mathcal{W}} = 0.5$

n	Unconstrained			Constant			TV			two-steps TV		
	MSE	SPEC	SENS	MSE	SPEC	SENS	MSE	SPEC	SENS	MSE	SPEC	SENS
50	2195.405	0	1	463.149	1	0	704.072	0.246	1	980.529	0.539	0.998
100	1057.826	0	1	435.446	1	0	467.772	0.181	1	659.561	0.508	1
500	463.173	0	1	425.972	1	0	365.889	0.065	1	449.020	0.355	1
1000	405.147	0	1	425.604	1	0	355.974	0.028	1	416.129	0.217	1

MSE: mean squared error, FP: false positives, FN: false negatives.

Table 6. *Simulation results in the additive model for* $\mathrm{P}_{obs} = 14\%$ *with* $a_{\mathcal{W}} = 0.5$

n	Unconstrained			Constant			TV			two-steps TV		
	MSE	SPEC	SENS	MSE	SPEC	SENS	MSE	SPEC	SENS	MSE	SPEC	SENS
50	2420.518	0	1	464.267	1	0	731.579	0.273	1	1007.030	0.553	0.984
100	1162.455	0	1	406.450	1	0	520.417	0.172	1	703.507	0.525	1
500	472.533	0	1	375.098	1	0	366.240	0.067	1	445.434	0.356	1
1000	403.280	0	1	371.508	1	0	351.461	0.035	1	409.135	0.234	1

MSE: mean squared error, FP: false positives, FN: false negatives.

Table 7. *Simulation results in the multiplicative model for* $\mathrm{P}_{obs} = 14\%$ *with* $a_{\mathcal{G}} = 1.5$

n	Unconstrained			Constant			TV			two-steps TV		
	MSE	SPEC	SENS	MSE	SPEC	SENS	MSE	SPEC	SENS	MSE	SPEC	SENS
50	2458.389	0	1	413.077	1	0	398.089	0.344	0.656	391.726	0.500	0.500
100	446.965	0	1	399.164	1	0	384.661	0.228	0.732	381.807	0.378	0.560
500	404.385	0	1	403.240	1	0	388.731	0.140	0.863	389.800	0.247	0.778
1000	400.621	0	1	401.898	1	0	388.649	0.095	0.924	388.454	0.222	0.831

MSE: mean squared error, FP: false positives, FN: false negatives.

Table 8. *Simulation results in the additive model for* $\mathrm{P}_{obs} = 14\%$ *with* $a_{\mathcal{G}} = 1.5$

n	Unconstrained			Constant			TV			two-steps TV		
	MSE	SPEC	SENS	MSE	SPEC	SENS	MSE	SPEC	SENS	MSE	SPEC	SENS
50	1348.887	0	1	404.295	1	0	397.596	0.323	1	525.807	0.610	0.998
100	631.320	0	1	364.233	1	0	275.558	0.215	1	347.038	0.555	1
500	230.762	0	1	344.619	1	0	159.434	0.114	1	196.384	0.463	1
1000	190.031	0	1	342.986	1	0	150.256	0.088	1	177.867	0.399	1

MSE: mean squared error, FP: false positives, FN: false negatives.

Table 9. *Simulation results in the multiplicative model for* $\mathrm{P}_{obs} = 14\%$ *with* $a_{\mathcal{G}} = 0.5$

n	Unconstrained			Constant			TV			two-steps TV		
	MSE	SPEC	SENS	MSE	SPEC	SENS	MSE	SPEC	SENS	MSE	SPEC	SENS
50	46565.252	0	1	342.382	1	0	305.771	0.240	0.775	293.672	0.414	0.639
100	325.573	0	1	337.498	1	0	290.698	0.165	0.811	283.154	0.316	0.698
500	280.627	0	1	335.447	1	0	281.469	0.087	0.871	279.878	0.262	0.723
1000	278.590	0	1	336.101	1	0	280.443	0.076	0.886	282.248	0.309	0.705

MSE: mean squared error, FP: false positives, FN: false negatives.

Table 10. *Simulation results in the additive model for* $\mathrm{P}_{obs} = 14\%$ *with* $a_{\mathcal{G}} = 0.5$

n	Unconstrained			Constant			TV			two-steps TV		
	MSE	SPEC	SENS	MSE	SPEC	SENS	MSE	SPEC	SENS	MSE	SPEC	SENS
50	1417.616	0	1	409.587	1	0	400.878	0.313	1	527.478	0.597	0.990
100	601.365	0	1	361.610	1	0	250.447	0.236	1	320.110	0.568	1
500	220.894	0	1	334.459	1	0	152.116	0.120	1	187.303	0.463	1
1000	184.552	0	1	332.156	1	0	146.099	0.093	1	173.086	0.397	1

MSE: mean squared error, FP: false positives, FN: false negatives.

Table 11. *AIC and BIC values in the multiplicative model*

Estimators	K	$L(\beta)$	AIC(β)	BIC(β)
Constant	4	1109.90	1117.90	1128.91
Unconstrained	20	1097.25	1137.25	1192.32
TV	7	1106.50	1120.50	1139.77
Two-steps TV	6	1104.76	1116.76	1133.28

Table 12. *Unconstrained and constant parameters estimates for the bladder data in the additive model*

s	Unconstrained				Constant			
	PYRIDOXINE	THIOTEPA	SIZE	NUMBER	PYRIDOXINE	THIOTEPA	SIZE	NUMBER
1	-0.0141	-0.0206	-0.0011	0.0105	-0.0014	-0.0158	0.0009	0.0096
2	0.3051	0.0029	0.0027	0.0008	-0.0014	-0.0158	0.0009	0.0096
3	-0.0196	0.0005	0.0095	0.0219	-0.0014	-0.0158	0.0009	0.0096
4	0.1227	-0.0102	0.0012	0.0286	-0.0014	-0.0158	0.0009	0.0096
5	0.0929	-0.0231	0.0071	0.0165	-0.0014	-0.0158	0.0009	0.0096

Table 13. *Total variation and two-steps total variation parameters estimates for the bladder data in the additive model*

s	TV				two-steps TV			
	PYRIDOXINE	THIOTEPA	SIZE	NUMBER	PYRIDOXINE	THIOTEPA	SIZE	NUMBER
1	-0.0057	-0.0144	0.0007	0.0070	-0.0069	-0.0149	0.0011	0.0064
2	-0.0005	-0.0144	0.0010	0.0070	-0.0069	-0.0149	0.0011	0.0064
3	-0.0005	-0.0144	0.0038	0.0193	-0.0069	-0.0149	0.0011	0.0218
4	0.0616	-0.0144	0.0038	0.0193	0.1080	-0.0149	0.0011	0.0218
5	0.0616	-0.0144	0.0038	0.0193	0.1080	-0.0149	0.0011	0.0218

Figure Captions

Fig. 1. Estimates for the bladder data in the additive model. The crosses represent the constant estimator, the filled circles the unconstrained estimator, the circles the total variation estimator and the squares the two steps total variation estimator.

Figures

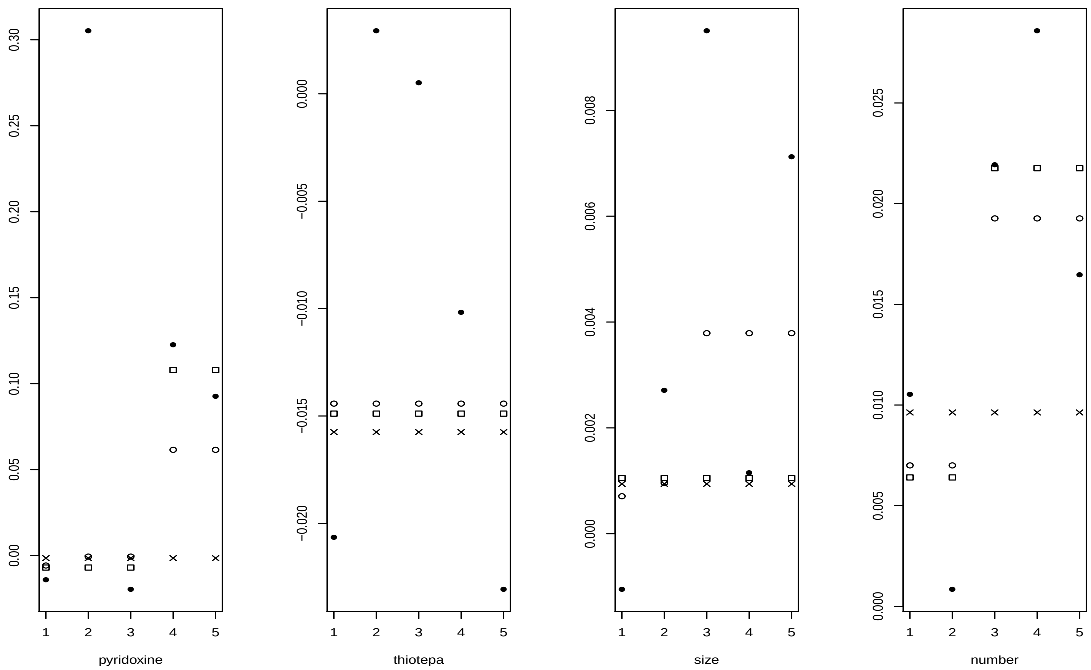

Fig. 1. Estimates for the bladder data in the additive model. The crosses represent the constant estimator, the filled circles the unconstrained estimator, the circles the total variation estimator and the squares the two steps total variation estimator.