

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

## 1. ALGORITHMS AND IMPLEMENTATION

### 1.1. Link to the lasso

The minimizers of the partial log-likelihood (respectively the partial least-squares) penalized with a covariate specific total variation are defined as:

$$\hat{\beta}_{TV/mult} \in \operatorname{argmin}_{\beta \in \mathbb{R}^{p \times B}} \left\{ L_n^{PL}(\beta) + \frac{\lambda_n}{n} \sum_{j=1}^p \operatorname{TV}(\beta^j) \right\} \quad (1)$$

and

$$\hat{\beta}_{TV/add} \in \operatorname{argmin}_{\beta \in \mathbb{R}^{p \times B}} \left\{ L_n^{PLS}(\beta) + \frac{\lambda_n}{n} \sum_{j=1}^p \operatorname{TV}(\beta^j) \right\}. \quad (2)$$

They can be seen as lasso estimators by introducing a block matrix  $D$  of size  $(pB \times pB)$  with  $p$  diagonal blocks being equal to a lower triangular matrix with nonzero elements equal to 1 and  $p^2 - p$  off-diagonal blocks being zeros matrices:

$$D = \begin{pmatrix} T_B & O_B & \cdots & O_B \\ O_B & T_B & \cdots & O_B \\ \cdots & \cdots & \cdots & \cdots \\ O_B & O_B & \cdots & T_B \end{pmatrix} \text{ with } T_B = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & 1 & \cdots & 1 \end{pmatrix} \text{ and } O_B = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}. \quad (3)$$

The minimization problems of Equations (1) and (2) can then be rewritten as

$$\begin{aligned} \hat{\beta}_{TV/mult} &= D\hat{\gamma}_{TV/mult} \text{ with} \\ \hat{\gamma}_{TV/mult} &\in \operatorname{argmin}_{\gamma \in \mathbb{R}^{p \times B}} \left\{ L_n^{PL}(D\gamma) + \frac{\lambda_n}{n} \sum_{j=1}^p \sum_{s=2}^B |\gamma^j(s)| \right\} = \operatorname{argmin}_{\gamma \in \mathbb{R}^{p \times B}} \left\{ L_n^{PL}(D\gamma) + \frac{\lambda_n}{n} \|\gamma\|_1 \right\} \end{aligned} \quad (4)$$

and

$$\begin{aligned} \hat{\beta}_{TV/add} &= D\hat{\gamma}_{TV/add} \text{ with} \\ \hat{\gamma}_{TV/add} &\in \operatorname{argmin}_{\gamma \in \mathbb{R}^{p \times B}} \left\{ L_n^{PLS}(D\gamma) + \frac{\lambda_n}{n} \sum_{j=1}^p \sum_{s=2}^B |\gamma^j(s)| \right\} = \operatorname{argmin}_{\gamma \in \mathbb{R}^{p \times B}} \left\{ L_n^{PLS}(D\gamma) + \frac{\lambda_n}{n} \|\gamma\|_1 \right\}, \end{aligned} \quad (5)$$

where, in both cases,  $\hat{\gamma}_{TV} = (\hat{\beta}_{TV}^1(1), \Delta\hat{\beta}_{TV}^1(2), \dots, \Delta\hat{\beta}_{TV}^p(B))^\top$ . The related R functions can be found at <http://www.lsta.upmc.fr/guilloux.php?main=publications>.

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

n	Unconstrained			Constant			TV			two-steps TV		
	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN
50	0.164	2	0	0.402	0	2	0.090	1.59	0.05	0.081	0.98	0.05
100	0.046	2	0	0.403	0	2	0.042	1.96	0	0.034	1.02	0
500	0.018	2	0	0.404	0	2	0.019	1.96	0	0.016	1.47	0
1000	0.017	2	0	0.405	0	2	0.017	1.99	0	0.015	1.42	0

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

### 1·2. Reweighted lasso

To enhance the sparsity in  $\hat{\gamma}_{TV/mult}$   $\hat{\gamma}_{TV/add}$  (or equivalently to force  $\hat{\beta}_{TV/mult}$  and  $\hat{\beta}_{TV/add}$  to be more constant), we consider a reweighed lasso in Sections 5 and 6. In both models, the two steps (or reweighted lasso) estimators are defined as

25       $\tilde{\beta}_{TV/mult} = D\tilde{\gamma}_{TV/mult}$  with

$$\tilde{\gamma}_{TV/mult} \in \underset{\gamma \in \mathbb{R}^{p \times B}}{\operatorname{argmin}} \left\{ L^{PL}(D\gamma) + \frac{\lambda_n}{n} \sum_{j=1}^p \sum_{s=2}^B \frac{1}{|\hat{\gamma}_{TV/mult}^j(s)| + |\hat{\gamma}_0|} |\gamma^j(s)| \right\} \quad (6)$$

and

$$\begin{aligned} \tilde{\beta}_{TV/add} &= D\tilde{\gamma}_{TV/add} \text{ with} \\ \tilde{\gamma}_{TV/add} &\in \underset{\gamma \in \mathbb{R}^{p \times B}}{\operatorname{argmin}} \left\{ L^{PLS}(D\gamma) + \frac{\lambda_n}{n} \sum_{j=1}^p \sum_{s=2}^B \frac{1}{|\hat{\gamma}_{TV/add}^j(s)| + |\hat{\gamma}_0|} |\gamma^j(s)| \right\} \end{aligned} \quad (7)$$

30      where  $\hat{\gamma}_{TV/mult}$  and  $\hat{\gamma}_{TV/add}$  are defined in Equation (4) and (5) and, in both cases,  $|\hat{\gamma}_0| = \min\{|\hat{\gamma}_{TV}^j(s)|, j = 1, \dots, p, s = 1, \dots, B, |\hat{\gamma}_{TV}^j(s)| \neq 0\}$ . We refer the reader to Candès et al. (2008) for a similar algorithm in a different situation.

### 1·3. Implementation

The constant and event specific estimators are calculated via the functions `coxph` (in the R package `survival`, see Therneau (2011)) and `ahaz` (in the R package `ahaz`, see Gorst-Rasmussen (2011)). We develop ad hoc functions to calculate the penalized estimators, they are based on the functions `coxnet` (in the R package `glmnet`, see Simon et al. (2011)) and `ahazpen` (in the R package `ahaz`). In particular, we choose the regularization parameters via cross-validation, as proposed in the latter packages.

## 40      2. SIMULATION STUDY (CONTINUED)

### 2·1. Results for the Weibull distribution with $a_W = 0.5$

### 2·2. Results for the Gompertz distribution with $a_G = 2 \cdot 5$

### 2·3. Results for the Gompertz distribution with $a_G = 0 \cdot 5$

## 3. DATASET

45      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

Table 2. *Simulation results in the multiplicative model for  $P_{obs} = 14\%$* 

n	Unconstrained			Constant			TV			two-steps TV		
	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN
50	NA	NA	NA	0.433	0	2	0.222	1.38	0.25	0.545	0.98	0.29
100	4.694	2	0	0.397	0	2	0.096	1.68	0.01	0.201	1.15	0
500	0.032	2	0	0.393	0	2	0.033	1.90	0	0.030	1.46	0
1000	0.027	2	0	0.394	0	2	0.029	1.93	0	0.026	1.33	0

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

Table 3. *Simulation results in the additive model for  $P_{obs} = 28\%$* 

n	Unconstrained			Constant			TV			two-steps TV		
	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN
50	8.794	2	0	0.494	0	2	0.583	0.90	0.79	0.923	0.66	0.89
100	1.410	2	0	0.371	0	2	0.340	1.34	0.22	0.568	1.09	0.41
500	0.216	2	0	0.307	0	2	0.113	1.94	0	0.133	1.67	0.09
1000	0.096	2	0	0.295	0	2	0.064	1.99	0	0.082	1.80	0.02

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

Table 4. *Simulation results in the additive model for  $P_{obs} = 14\%$* 

n	Unconstrained			Constant			TV			two-steps TV		
	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN
50	NA	NA	NA	0.626	0	2	NA	NA	NA	1.618	0.72	1.1
100	6.592	2	0	0.436	0	2	0.987	1.56	0.29	1.036	1.17	0.47
500	0.429	2	0	0.326	0	2	0.188	1.96	0	0.248	1.69	0.09
1000	0.223	2	0	0.311	0	2	0.138	1.99	0.01	0.155	1.87	0.01

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

Table 5. *Simulation results in the multiplicative model for  $P_{obs} = 28\%$* 

n	Unconstrained			Constant			TV			two-steps TV		
	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN
50	0.498	2	0	0.409	0	2	0.079	1.55	0.045	0.069	0.89	0.05
100	0.039	2	0	0.407	0	2	0.037	1.62	0	0.030	0.96	0
500	0.0119	2	0	0.412	0	2	0.014	1.91	0	0.012	1.28	0
1000	0.010	2	0	0.413	0	2	0.012	1.98	0	0.011	1.29	0

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

a given  $\beta$ , these criteria are defined by:

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

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 13, the smallest AIC is obtained from the two-steps 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.

Table 6. Simulation results in the multiplicative model for  $P_{obs} = 14\%$ 

n	Unconstrained			Constant			TV			two-steps TV		
	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN
50	NA	NA	NA	0.433	0	2	0.244	1.36	0.26	0.290	0.90	0.36
100	1.790	2	0	0.412	0	2	0.087	1.59	0.03	0.07	1.10	0.05
500	0.026	2	0	0.410	0	2	0.027	1.92	0	0.024	1.45	0
1000	0.021	2	0	0.411	0	2	0.023	1.96	0	0.022	1.38-	0

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

Table 7. Simulation results in the additive model for  $P_{obs} = 28\%$ 

n	Unconstrained			Constant			TV			two-steps TV		
	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN
50	2.398	2	0	0.409	0	2	0.389	0.90	0.475	0.520	0.54	0.79
100	1.200	2	0	0.350	0	2	0.297	1.41	0.23	0.396	0.95	0.43
500	0.139	2	0	0.305	0	2	0.082	1.94	0.01	0.104	1.47	0.05
1000	0.079	2	0	0.300	0	2	0.054	1.99	0	0.058	1.69	0.01

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

Table 8. Simulation results in the additive model for  $P_{obs} = 14\%$ 

n	Unconstrained			Constant			TV			two-steps TV		
	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN
50	NA	NA	NA	0.537	0	2	5.318	0.92	0.83	1.362	0.73	0.91
100	4.071	2	0	0.384	0	2	0.577	1.53	0.22	0.935	1.11	0.48
500	0.317	2	0	0.329	0	2	0.153	1.96	0.01	0.184	1.67	0.12
1000	0.162	2	0	0.315	0	2	0.101	1.99	0	0.124	1.87	0.02

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

Table 9. Simulation results in the multiplicative model for  $P_{obs} = 28\%$ 

n	Unconstrained			Constant			TV			two-steps TV		
	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN
50	2.964	2	0	0.411	0	2	0.091	1.5	0.06	0.071	0.81	0.07
100	0.043	2	0	0.413	0	2	0.036	1.64	0	0.032	1.01	0
500	0.013	2	0	0.409	0	2	0.013	1.90	0	0.013	1.37	0
1000	0.012	2	0	0.413	0	2	0.011	1.94	0	0.011	1.35	0

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

In addition, the total variation estimator in the multiplicative model, the total variation and reweighted total-variation estimators in the additive model are respectively displayed in Figures 1-3. 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.

## APPENDIX: PROOFS

### Proof of Lemma A1

By its definition,  $dN_i(t) = \mathbf{1}(C_i(t) \geq t)dN_i^*(t)$  and as a consequence

$$\begin{aligned} & \mathbb{E}[dN_i(t) \mid X_i(t), D_i \wedge C_i \geq t, N_i(t) = s-1] = \mathbb{E}[dN_i(t) \mid X_i(t), D_i \wedge C_i \geq t, N_i^*(t) = s-1] \\ & = \mathbf{1}(C_i \geq t)E[dN_i^*(t) \mid X_i(t), D_i \wedge C_i \geq t, N_i^*(t) = s-1] = \mathbf{1}(C_i \wedge D_i \geq t)\rho_0(t, s, X_i(t))dt, \end{aligned}$$

Table 10. *Simulation results in the multiplicative model for  $P_{obs} = 14\%$* 

n	Unconstrained			Constant			TV			two-steps TV		
	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN
50	NA	NA	NA	0.432	0	2	1.532	1.35	0.27	0.201	0.86	0.37
100	0.292	2	0	0.417	0	2	0.080	1.66	0.02	0.068	1.11	0.03
500	0.026	2	0	0.409	0	2	0.027	1.91	0	0.024	1.37	0
1000	0.021	2	0	0.411	0	2	0.023	1.93	0	0.021	1.27	0

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

Table 11. *Simulation results in the additive model for  $P_{obs} = 28\%$* 

n	Unconstrained			Constant			TV			two-steps TV		
	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN
50	5.061	2	0	0.444	0	2	0.611	1.05	0.60	0.686	0.64	0.91
100	1.311	2	0	0.350	0	2	0.286	1.39	0.22	0.366	0.91	0.44
500	0.153	2	0	0.307	0	2	0.089	1.95	0	0.109	1.45	0.07
1000	0.083	2	0	0.300	0	2	0.057	1.98	0	0.062	1.71	0.02

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

Table 12. *Simulation results in the additive model for  $P_{obs} = 14\%$* 

n	Unconstrained			Constant			TV			two-steps TV		
	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN	MSE	FP	FN
50	NA	NA	NA	0.530	0	2	0.993	1.03	0.89	1.433	0.83	0.96
100	4.590	2	0	0.391	0	2	0.654	1.40	0.27	0.815	1.08	0.45
500	0.308	2	0	0.329	0	2	0.142	1.97	0	0.174	1.66	0.07
1000	0.155	2	0	0.316	0	2	0.093	1.99	0	0.117	1.84	0.02

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

Table 13. *AIC and BIC values in the multiplicative model*

Estimators	K	$L(\beta)$	AIC( $\beta$ )	BIC( $\beta$ )
Constant	4	1109 · 9	1117 · 895	1128 · 909
Unconstrained	20	1097 · 25	1137 · 246	1192 · 317
TV	7	1106 · 5	1120 · 496	1139 · 772
Two-steps TV	6	1104 · 76	1116 · 759	1133 · 28

where the last equality comes from Assumption 1.

#### 4. PROOFS

##### *Proof of Lemma A2*

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

$$\begin{aligned}
 L_n^{PLS}(\beta) &= \frac{1}{n} \sum_{i=1}^n \sum_{s=1}^B \int \{X_i(t)\beta(s) - \bar{X}^s(t)\beta(s)\}^2 Y_i^s(t) dt \\
 &\quad - \frac{2}{n} \sum_{i=1}^n \sum_{s=1}^B \int \{X_i(t)\beta(s) - \bar{X}^s(t)\beta(s)\} \mathbf{1}(N_i(t) = s) dN_i(t).
 \end{aligned} \tag{8} \quad 70$$

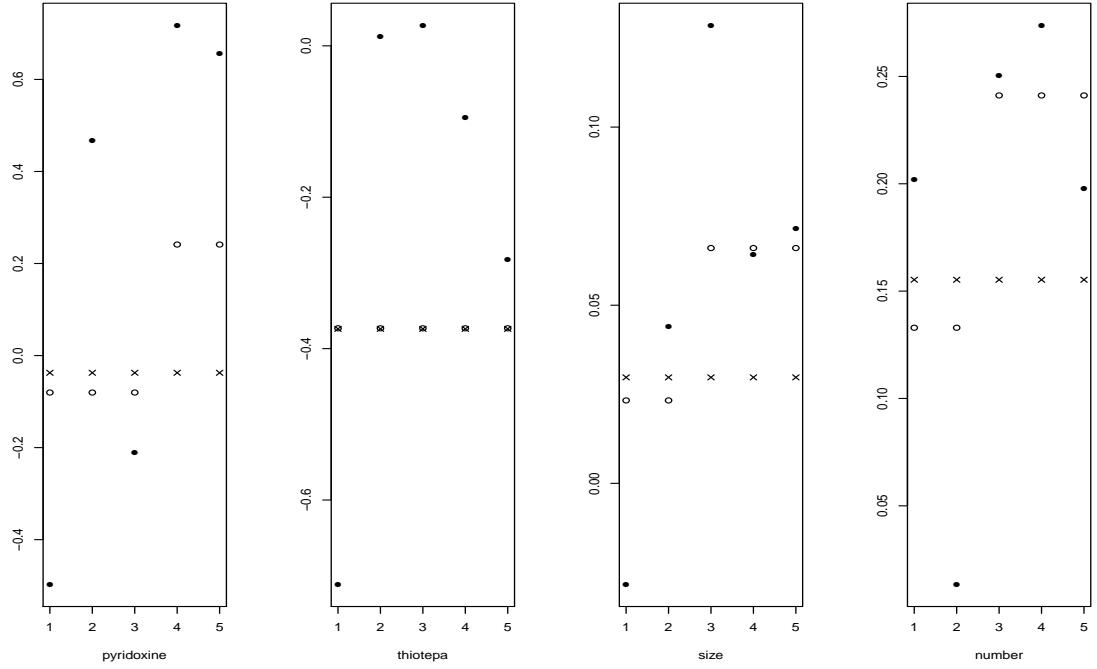


Fig. 1. Parameter estimates for the bladder data in the multiplicative model. The triangles represent the constant estimator, the filled circles the unconstrained estimator and the circles the total variation estimator.

Apply Lemma A1 in the equation  $dM_i(t) = dN_i(t) - \mathbb{E}(dN_i(t) | X_i(t), D_i \wedge C_i \geq t, N_i(t) = s)$  and conclude the proof using the relation

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

75

#### *Proof of Lemma A3*

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

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

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

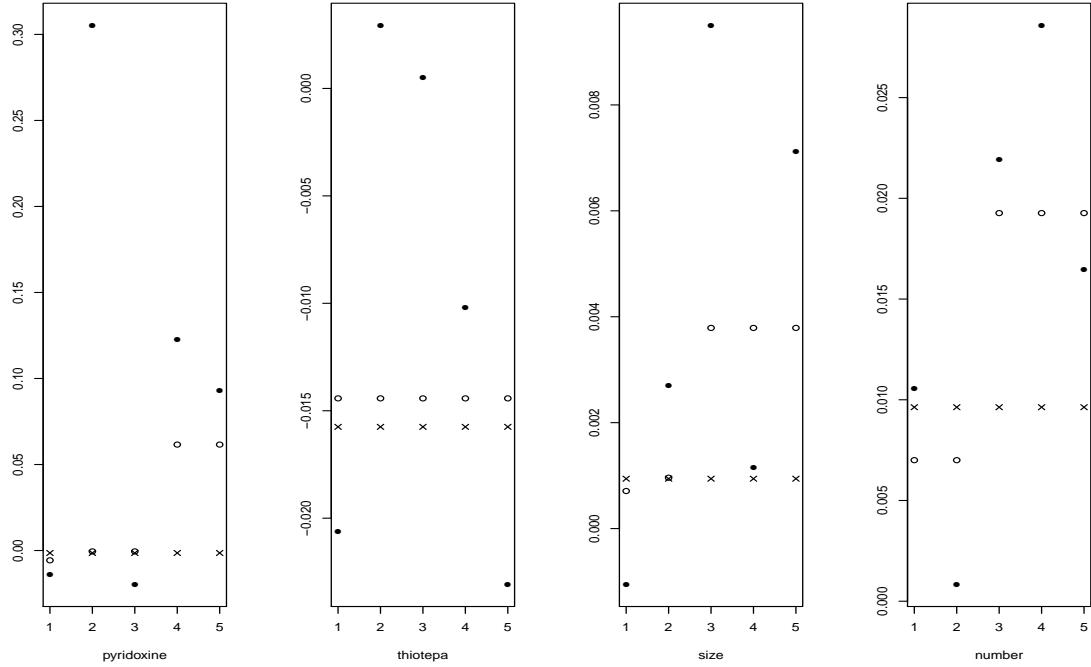


Fig. 2. Parameter estimates for the bladder data in the additive model. The triangles represent the constant estimator, the filled circles the unconstrained estimator and the circles the total variation estimator.

## REFERENCES

- CANDÈS, E. J., WAKIN, M. B. & BOYD, S. P. (2008). Enhancing sparsity by reweighted  $l_1$  minimization. *J. Fourier Anal. Appl.* **14**, 877–905.  
 GORST-RASMUSSEN, A. (2011). *ahaz: Regularization for semiparametric additive hazards regression*. R package version 1.1.  
 SIMON, N., FRIEDMAN, J., HASTIE, T. & TIBSHIRANI, R. (2011). Regularization paths for cox's proportional hazards model via coordinate descent. *Journal of Statistical Software* **39**, 1–13.  
 THERNEAU, T. (2011). *survival: Survival analysis, including penalised likelihood*. R package version 2.36-10, Splus->R port by Thomas Lumley.  
 VAN DER VAART, A. W. & WELLNER, J. A. (1996). *Weak convergence and empirical processes*. Springer Series in Statistics. New York: Springer-Verlag. With applications to statistics.

85

90

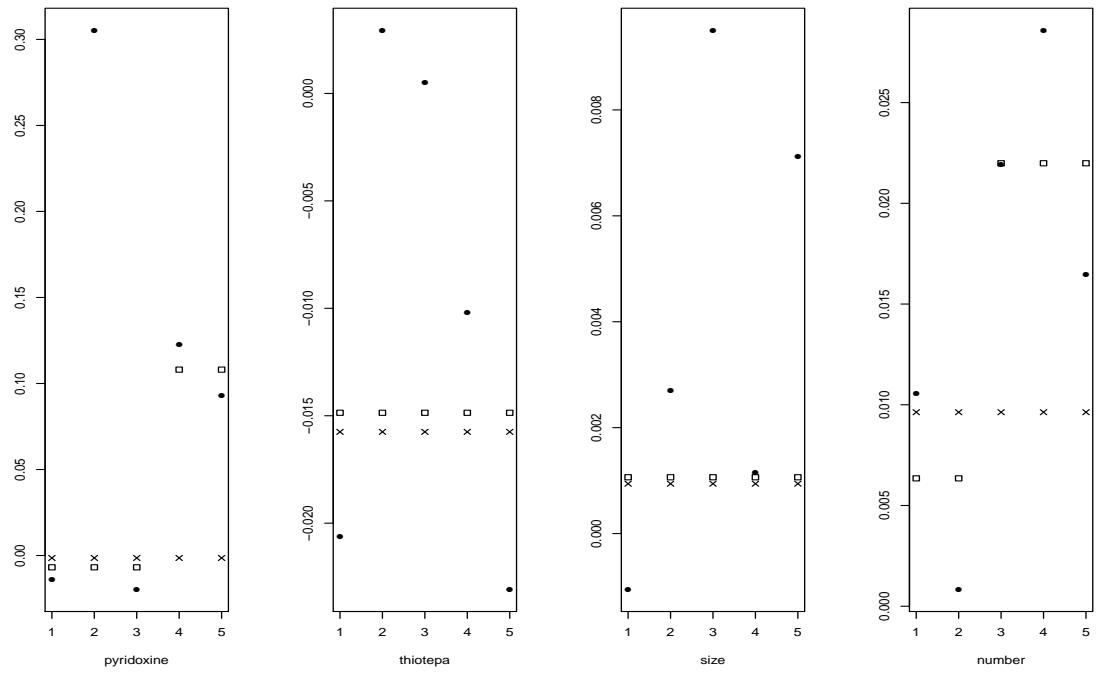


Fig. 3. Parameter estimates for the bladder data in the additive model. The triangles represent the constant estimator, the circles the unconstrained estimator and the squares the two-steps total variation estimator.