



HAL
open science

A two-tier proportional hazards regression model with change-points in the baseline function

Abdullah Oueslati, Olivier Lopez

► **To cite this version:**

Abdullah Oueslati, Olivier Lopez. A two-tier proportional hazards regression model with change-points in the baseline function. 2012. hal-00655891

HAL Id: hal-00655891

<https://hal.science/hal-00655891>

Preprint submitted on 3 Jan 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A two-tier proportional hazards regression model with change-points in the baseline function

Abdullah OUESLATI¹, Olivier LOPEZ²

January 3, 2012

Abstract

In this article, we consider a new regression model for counting processes under a proportional hazards assumption. This model is motivated by the need of understanding the evolution of the booking process of a railway company. The main novelty of the approach consists in assuming that the baseline hazard function is piecewise constant. Hence, the parameters of the model can be separated into two different types: parameters that measure the influence of the covariates, and parameters from a (multiple) change-point model for the baseline. Cox's semiparametric regression can be seen as a limit case of our model. We develop an iterative procedure to estimate the different parameters, and a test procedure that allows to perform change-point detection in the baseline. Our technique is supported by simulation studies and a real data analysis, which show that our model can be a reasonable alternative to Cox's regression model, particularly in the presence of tied event times.

Key words: proportional hazards regression, change-point detection, iterative procedures, dynamic programming, revenue-management, survival analysis.

Short title: A two-tier proportional hazards model

¹ SNCF, Direction de l'Innovation et de la Recherche & Laboratoire de Statistique Théorique et Appliquée, Université Pierre et Marie Curie Paris VI, 4 place Jussieu, 75005 Paris, France, E-mail: abdullah.oueslati@gmail.com, abdullah.oueslati@sncf.fr.

² Laboratoire de Statistique Théorique et Appliquée, Université Pierre et Marie Curie Paris VI, 4 place Jussieu, 75005 Paris, France, E-mail: olivier.lopez0@upmc.fr.

1 Introduction

In Yield-Management, a critical issue is to infer on the stochastic behaviour of a counting process (representing, for example, the number of reservations for a transportation company, or the number of reservations in an hotel) in order to be able to develop strategies that optimize the profit. In this paper, we develop a statistical model that has been firstly designed to apply to the study of the amount of reservations in a public transport company. The key feature of our new approach stands in the fact that we try to identify two different factors that influence the general dynamics of reservations. The first one is a classical capacity constraint: when some types of seats become unavailable, the consumers change their behaviour, by either preferring another category of seats, or by not buying any ticket. The second one is a time evolution, caused by the proximity of the date of departure. In the model that we consider, the number of reservations is considered as a counting process that evolves through a multiplicative hazard model. Multiplicative hazard regression models are a very common tool when one wants to model the influence of some set of covariates, which, in the particular situation that we will latter consider, will consist of informations on the availability of the different types of seats. In these models, a baseline hazard function determines some general behaviour of the process, while the influence of the covariates appears only through a multiplicative term. Back to our example, the baseline hazard function will represent some general evolution of the behaviour of the consumers, that is not caused by the lack of availability, but only by a general time evolution.

If such a multiplicative hazard model holds, it becomes very easy to determine the influence of a specific covariate, identifying its contribution to the increase of the hazard rate. In the literature of multiplicative hazard models, one must distinguish between two major types of approaches. The fully parametric approach consists of specifying the baseline hazard function. See for example Martinussen and Sheike (2006) for a review of the different techniques in this framework. On the other hand, the semiparametric approach initially developed by Cox (1972) (see also Andersen and Gill 1982) does not require the specification of the baseline hazard rate, whose estimation relies on nonparametric approaches. The most classical of these approaches is developed by Breslow (1972), which introduces a nonparametric estimator of the cumulative baseline hazard rate, but other approaches have been developed, as the one proposed by Dabrowska (1997) using kernel smoothing, or the one based on regression splines in Sleeper and Harrington (1990). One of the drawbacks of such techniques stands in the difficulty of interpretation of the

resulting estimator of the baseline hazard function.

The approach that we develop in this paper is parametric, and is close to Cox's regression model in the sense that Cox's regression model can be seen as a limit case. We propose to model the baseline hazard function as a piecewise constant function, without prior knowledge of the times at which the jumps occur. This methodology is inspired by multiple change-point techniques (see e.g. Csörgo and Hórvath 1997; Basseville and Nikiforov 1993; Lebarbier 2005), which are particularly adapted to our problem, since they will allow to estimate some time of change of behaviour of the consumers. Moreover, this framework allows us to develop test procedures that consists of detecting the presence or absence of change-points. From a practical viewpoint, such type of test procedure will allow to distinguish between changes of behaviours that are only caused by the covariates (that is, the changes in the availability), and changes which represent a general evolution of the dynamics. Moreover, compared to Cox's regression model, our approach can be seen as a reasonable compromise between additional assumptions and computational efficiency. Indeed, our estimation procedure is expected to outperform Cox's partial likelihood maximization in situations where the number of tied event times can be high, which is shown through simulations.

The rest of this paper is organized as follows. In section 2, we introduce our hazard regression model. Section 3 is devoted to the estimation of the parameters, which relies on maximum likelihood estimation. We especially focus on the practical implementation, developing a dynamic programming strategy for the estimation of the change-points, and an iterative algorithm to perform the joint estimation of the change-points and of the regression parameters. We also propose a test procedure to perform the change-point detection, and discuss how to compute the critical values through a simulation approach. In section 4, we investigate the practical behaviour of our technique. Through a simulation study, we first compare the behaviour of our estimation procedure to the use of Cox's partial likelihood when we only focus on the influence of covariates, before investigating on the change-point estimation. Next, we apply our technique to a real case provided by the French national railway company SNCF.

2 Proportional hazards model with change-point detection and parametric regression

In this section, we present the statistical model that will be used throughout this paper, and the procedure that we propose to estimate its parameters. Section 2.1 details notations and assumptions that will be used in the following. As announced in the introduction, our model combines a proportional hazards regression model, with a change-point approach in order to estimate the baseline intensity. A small discussion to compare our model to Cox's regression model is provided in section 4.2.

2.1 Model

We consider observations made of m independent counting processes, observed at discrete dates $t_0 < \dots < t_n$, that is we observe $(N_j(t_i))_{\substack{0 \leq i \leq n \\ 1 \leq j \leq m}}$, where $t_n = \tau$ is the total time of observation. Since all processes may not be observed during the same amount of time, we also observe indicators $Y_{ij} \triangleq Y_j(t_i)$ taking the value 1 if N_j is at risk at time t_i , and $Y_j(t_i) = 0$ otherwise. In addition, one observes a time-dependent vector of covariates $(X_{ij})_{\substack{0 \leq i \leq n \\ 0 \leq j \leq m}} \triangleq (X_j(t_i))_{\substack{0 \leq i \leq n \\ 0 \leq j \leq m}}$.

Data are modeled through a proportional regression model, that is the hazard intensity of the process $(N_j(t))_{0 \leq t \leq \tau}$ is

$$\lambda_j(t) = Y_j(t) \lambda_0(t) e^{\beta X_j(t)}, \quad (2.1)$$

where $Y_j(t)$ is the ‘‘at risk’’ process, $\lambda_0(t)$ is the baseline hazard function, β is a some unknown regression parameter to estimate and $X_j(t)$ is a time-dependent variable.

Such type of regression models have been widely used for the study of recurrent events processes (see Ghosh and Lin 2003). An important version is the semiparametric Cox regression model (Cox 1972; Andersen and Gill 1982). In Cox's model, the baseline hazard function λ_0 is not specified. Compared to a fully parametric model, Cox's approach provides additional flexibility to the model. Nevertheless, one of the disadvantages stands in the difficulty of interpreting the estimation of the nonparametric part. In our approach, we therefore chose to rely on a fully parametric model, which is adapted to some specific situations such as the one we consider in our application in section 4.3 below. We assume that the function λ_0 is piecewise constant, with jumps only at l unknown times $(t_{k_s})_{1 \leq s \leq l}$,

that is

$$\lambda_0(t) = \sum_{s=0}^l \mu_s \mathbb{1}\{t_{k_s} \leq t < t_{k_{s+1}}\}, \quad (2.2)$$

which can be seen as a multiple change-point model on the baseline hazard function. The values μ_s of the intensity on $[t_{k_s}, t_{k_{s+1}})$ are also unknown in our framework. An interesting feature of this model stands in the fact that the time of change can be interpreted as structural changes in the hazard rate that are not caused by the covariates, but which come from a change in the global dynamics. In the example we provide below in section 4.3, this change is interpreted by a structural and global increase (or decrease) in the number of bookings pro period of time. In a Cox-regression model, the classical way to estimate the nonparametric part consists of using the estimator of Breslow (1972). In this case, the resulting estimator is a piecewise constant hazard function with n jumps (at each time of observation), hence it can be seen as a limit case of model (2.2).

3 Estimation procedure

We now present our methodology to estimate the different parameters of the model considered in section 2. Our procedure consists of applying maximum likelihood inference (see section 3.1). Since this approach leads to numerical optimization problems when it comes to implementation, we provide an iterative algorithm, detailed in section 3.2, to perform this maximization. The principle of this strategy consists of distinguishing between the two different structures present in our model. This is done by separating the estimation of the regression parameters from the estimation of the change-points. Section 3.3 proposes a numerical method involving dynamic programming in order to perform change-point estimation. Finally, a test procedure for change-point detection is provided in section 3.4.

3.1 Maximum likelihood estimation

From a statistical point of view, we have to estimate the following parameters $\theta = (\beta, k_1, \dots, k_l, \mu_0, \dots, \mu_l)$. In this type of parametric problems, it is natural to rely on a maximum likelihood approach. Following Martinussen and Scheike (2006), Aalen et al.

(2008), the log-likelihood in our model can be expressed as,

$$L_l(\theta) = - \sum_{s=0}^l \mu_s \sum_{i=k_s}^{k_{s+1}-1} d_i \sum_{j=1}^m Y_{ij} e^{\beta X_{ij}} + \sum_{s=0}^l \log(\mu_s) \sum_{\substack{k_s \leq i \leq k_{s+1}-1 \\ 1 \leq j \leq m}} Z_{ij} + \beta \sum_{\substack{1 \leq i \leq n \\ 1 \leq j \leq m}} X_{ij} Z_{ij},$$

where $Z_{ij} \triangleq N_j(t_i) - N_j(t_{i-1})$ is the variation of the process N_j on $[t_{i-1}, t_i)$, and $(d_1, \dots, d_n) \triangleq (t_1 - t_0, \dots, t_n - t_{n-1})$ denotes the length of the different time intervals (which may not be equal in practice). The maximum likelihood estimator is then defined as

$$\hat{\theta}_l = \underset{\theta}{\operatorname{argmax}} L_l(\theta).$$

Unfortunately, direct maximization of a functional of the type of the log-likelihood leads to technical problems, due to the complexity of the numerical techniques that would be involved if one wishes to achieve exact maximization of this functional. Therefore, we provide in the following section an iterative procedure that allows to fill this gap.

3.2 Estimation algorithm

The algorithm detailed in this section is based on the separation of the parameter θ into two parts: the regression parameter β and the change-point parameters $(\underline{k}, \underline{\mu})$ with $\underline{k} = (k_1, \dots, k_l)$ and $\underline{\mu} = (\mu_0, \dots, \mu_l)$. Similar approaches are classically developed in semiparametric regression (see e.g. Xia et al. 2002). In our framework, the model is fully parametric. Nevertheless, in our approach, regarding the estimation of the baseline, we are close to a nonparametric problem, in the sense that the number of parameters to estimate may be high if we increase the number l of change-points.

Therefore, to optimize $L_l(\beta, (\underline{k}, \underline{\mu}))$, we develop the following iterative algorithm, containing two steps in each iteration:

Algorithm 1 Find θ which maximizes $L_l(\theta)$

```

 $\beta^{(0)} \leftarrow \beta_0$ 
 $(\underline{k}^{(0)}, \underline{\mu}^{(0)}) \leftarrow \operatorname{argmax}_{(\underline{k}, \underline{\mu})} L_l(\beta^{(0)}, (\underline{k}, \underline{\mu}))$ 
 $\theta^{(0)} \leftarrow (\beta^{(0)}, (\underline{k}^{(0)}, \underline{\mu}^{(0)}))$ 
 $\beta^{(1)} \leftarrow \operatorname{argmax}_{\beta} L_l(\beta, (\underline{k}^{(0)}, \underline{\mu}^{(0)}))$ 
 $(\underline{k}^{(1)}, \underline{\mu}^{(1)}) \leftarrow \operatorname{argmax}_{(\underline{k}, \underline{\mu})} L_l(\beta^{(1)}, (\underline{k}, \underline{\mu}))$ 
 $\theta^{(1)} \leftarrow (\beta^{(1)}, (\underline{k}^{(1)}, \underline{\mu}^{(1)}))$ 
 $p \leftarrow 1$ 
while  $|L_l(\theta^{(p)}) - L_l(\theta^{(p-1)})| > \varepsilon$  do
   $p \leftarrow p + 1$ 
  Change-point step :  $(\underline{k}^{(p)}, \underline{\mu}^{(p)}) \leftarrow \operatorname{argmax}_{(\underline{k}, \underline{\mu})} L_l(\beta^{(p-1)}, (\underline{k}, \underline{\mu}))$ 
  Regression step :  $\beta^{(p)} \leftarrow \operatorname{argmax}_{\beta} L_l(\beta, (\underline{k}^{(p)}, \underline{\mu}^{(p)}))$ 
   $\theta^{(p)} \leftarrow (\beta^{(p)}, (\underline{k}^{(p)}, \underline{\mu}^{(p)}))$ 
end while

```

At the regression step, the problem

$$\operatorname{argmax}_{\beta} L_l(\beta, (\underline{k}, \underline{\mu})),$$

can be solved by a quasi-Newton algorithm (available on the stat package of R software), while, at the change-point step, the maximization problem

$$\operatorname{argmax}_{(\underline{k}, \underline{\mu})} L_l(\beta, (\underline{k}, \underline{\mu})), \quad (3.1)$$

requires high computational time to be solved. Indeed, the likelihood has to be maximized for all combinations of vector (k_1, \dots, k_l) . Dynamic programming has been used by Bellman (1961) and Jackson et al. (2005) to fix some similar kind of problems. The algorithm developed in the next section is a direct extension of these approaches, but with a different contrast to minimize, which is adapted to our particular framework.

3.3 Dynamic programming for change-point detection

Solving the problem (3.1) is equivalent to solving

$$\operatorname{argmax}_{\underline{k}} L_l(\beta, (\underline{k}, \widehat{\underline{\mu}}(\underline{k})))$$

with $\widehat{\underline{\mu}}(\underline{k}) = (\widehat{\mu}_0(k_0, k_1), \dots, \widehat{\mu}_l(k_l, k_{l+1}))$ and $\widehat{\mu}_s(k_s, k_{s+1}) = \frac{\sum_{j=1}^m \sum_{i=k_s}^{k_{s+1}-1} Z_{ij}}{\sum_{i=k_s}^{k_{s+1}-1} d_i \sum_{j=1}^m Y_{ij} e^{\beta X_{ij}}}$. Indeed, $\widehat{\underline{\mu}}(\underline{k})$ is the maximum likelihood estimator of the vector of intensities $\underline{\mu}$ if we assume that

the changes occur at times t_{k_1}, \dots, t_{k_l} for a fixed vector $\underline{k} = (k_1, \dots, k_l)$.

Therefore, solving (3.1) consists of solving an optimization problem of the following type:

$$\operatorname{argmax}_{\underline{k}_l \in \mathcal{C}_{n-1}^l} f(\underline{k}_l),$$

where \mathcal{C}_{n-1}^l is the set of $\binom{n-1}{l}$ vector combinations $\underline{k}_l = (1, k_1, \dots, k_l, n+1)$ such that $(1 < k_1 < k_2 < \dots < k_l < n+1)$, with

$$f(\underline{k}_l) = \sum_{s=0}^l g(k_s, k_{s+1}),$$

where g is defined as,

$$g(k_s, k_{s+1}) = \log \left(\frac{\sum_{j=1}^m \sum_{i=k_s}^{k_{s+1}-1} Z_{ij}}{\sum_{i=k_s}^{k_{s+1}-1} d_i \sum_{j=1}^m Y_{ij} e^{\beta X_{ij}}} \right) \sum_{j=1}^m \sum_{i=k_s}^{k_{s+1}-1} Z_{ij}.$$

To define our dynamic programming procedure, we also need to introduce,

$$V(b, i) \triangleq \max_{\underline{k}_i \in \mathcal{C}^i(1, b)} f(\underline{k}_i),$$

$$\widehat{V}(b, 1) \triangleq \operatorname{argmax}_{\underline{k}_i \in \mathcal{C}^i(1, b)} f(\underline{k}_i),$$

where $\mathcal{C}^i(1, b)$ is the set of possible vectors $\underline{k}_i = (1, k_1, \dots, k_i, b)$ such that $1 < k_1 < \dots < k_i < b$.

The following $l+1$ steps algorithm provides the vector of right combination $\widehat{\underline{k}}_l$ which maximizes the objective function f . The main idea of the algorithm can be summarized in the following way : assuming that we know, for each possible possible location of the i^{th} change-point, the optimal combination of $i-1$ first change-points, we just have to choose the i^{th} change-point that maximizes the function f . We then apply iteratively this procedure to get the optimal position of $i=1$ change-point, then $i=2, \dots$, until $i=l$ change-points.

against the following alternative,

$$H_1 : \exists(k, \mu_0, \mu_1) \text{ s.a. } \lambda_0(t) = \mu_0 \mathbb{1}\{t < t_k\} + \mu_1 \mathbb{1}\{t \geq t_k\}.$$

A natural way to proceed consists of using a likelihood ratio test. Let $\hat{\theta}_1$ (resp. $\hat{\theta}_2$) denote the maximum likelihood estimators developed in the previous sections in the case where one assumes that there is no change-point (resp. one change-point), that is under H_0 (resp. under H_1). The likelihood ratio test procedure consists of rejecting H_0 when $T_n = L_2(\hat{\theta}_2) - L_1(\hat{\theta}_1)$ is larger than some critical value s_α that permits to achieve the proper level α . Rather than rely on an asymptotic approximation of T_n , we suggest to use a bootstrap procedure to compute this critical value.

The test procedure described below can be extended to test the presence of k -change-points, against the model with $(k - 1)$ -change-points. Then, it can be used as a forward procedure to decide what number of change-points should be considered (see application in section 4.4).

Simulation technique to compute the critical value

1. **Estimation** : first, we estimate $\hat{\theta}_1$ from the data. If H_0 is true, the probability law defined by $\hat{\theta}_1$ should be close to the true probability distribution.
2. **Simulation** : for $b = 1, \dots, B$ simulate a multivariate process according to the law defined by $\hat{\theta}_1$ and $(X_{ij}, Y_{ij})_{\substack{0 \leq i \leq n \\ 1 \leq j \leq m}}$, where we used the matrix X and Y observed on the data-set.
3. **Computation of the critical value** : for each of the simulated trajectory, compute $\hat{\theta}_1^b$ (the estimator based on the b -th simulated trajectory, with the assumption that there is no change-point), and $\hat{\theta}_2^b$ (the estimator based on the b -th simulated trajectory, with the assumption that there is one change-point), and compute the corresponding test statistic T_n^b . Take s_α such as a proportion α of the $(T_n^b)_{1 \leq b \leq B}$ is larger than s_α .

4 Application : simulations and SNCF study

In this section, we investigate the empirical behaviour of our procedure. In section 4.1, we consider a fixed value of the regression parameter β and we observe how the change-point modeling influences the quality of estimation of this parameter. Moreover, we

investigate the quality of our final estimators of the intensities λ_j . In section 4.2, we develop a comparison with Cox's regression model, which suffers, in our framework, from the presence of tied event times. Section 4.3 presents the application of our model to the study of bookings in the French company SNCF.

4.1 Estimation

The simulation setting that we propose in order to enlighten the practical behaviour of our procedure is inspired by the situation that is considered in the study developed in section 4.3. We consider $n = 20$, $m = 42$, $\beta = 2$, and at-risk processes $Y_j \equiv 1$. The lines of the matrix X are made of vectors X_i , with components equal to zero up to some index $j_0(i)$, and equal to 1 for the remaining components. The choice of such a structure for the matrix X is motivated by our example in the case of the SNCF study, in which a value 1 corresponds to a class of seats which is unavailable. For all our simulation study, we considered the same matrix X , which has been randomly generated in order to match the structure observed in the real data set, that is a proportion of 31.07% of 1 in the coefficients. The interval between t_i and t_{i+1} is assumed to be constant and equal to one.

For different values of λ_0 (that is different values of $(\underline{k}, \underline{\mu})$), we simulate $q = 1000$ samples and estimate the parameters from our estimation procedure for each simulated sample. We compare our estimator with the one obtained by maximizing Cox's partial likelihood. Since there is a significant number of tied event times in the simulated datasets, one requires to use approximations of this partial likelihood. In our simulations, we consider the case of Efron's approximation (1977) and of Breslow's approximation (1974). The performance of each technique is measured through different error criterions. We focus first on the quality of estimation for each separated parameter, and then look at the global estimator of the intensities. For a parameter $x \in \mathbb{R}$, and an estimate $x^{(p)} \in \mathbb{R}$ of this quantity, we consider the following criterions:

$$\begin{aligned}
 e_1(x) &= \frac{1}{q} \sum_{p=1}^q |\hat{x}^{(p)} - x|, \\
 e_2(x) &= \frac{1}{q} \sum_{p=1}^q (\hat{x}^{(p)} - x)^2, \\
 e_\infty(x) &= \max_{1 \leq p \leq q} |\hat{x}^{(p)} - x|.
 \end{aligned}$$

Similarly, we define errors for an estimator $\hat{\lambda}$ of the matrix $\lambda = (\lambda_j(t_i, \theta))_{\substack{1 \leq i \leq n \\ 1 \leq j \leq m}}$ as

$$\begin{aligned} e_1(\hat{\lambda}) &= \frac{1}{q} \sum_{p=1}^q \sum_{j=1}^m \sum_{i=1}^n |\hat{\lambda}_j(t_i, \hat{\theta}^{(p)}) - \lambda_j(t_i, \theta)|, \\ e_2(\hat{\lambda}) &= \frac{1}{q} \sum_{p=1}^q \sum_{j=1}^m \sum_{i=1}^n (\hat{\lambda}_j(t_i, \hat{\theta}^{(p)}) - \lambda_j(t_i, \theta))^2, \\ e_\infty(\hat{\lambda}) &= \frac{1}{q} \sum_{p=1}^q \max_{\substack{1 \leq i \leq n \\ 1 \leq j \leq m}} |\hat{\lambda}_j(t_i, \hat{\theta}^{(p)}) - \lambda_j(t_i, \theta)|. \end{aligned}$$

Compared to the separate study of the estimation of each parameter, evaluating some distance between our estimated $\hat{\lambda}$ and the true intensity matrix λ presents the interest to have a global view on the performance of the estimation procedure in our model. More precisely, we will compare the estimated intensity to the one that would be obtained in a model with no change-point, to try to understand if the improvement obtained through our approach is significant or not.

We consider different settings, first in the case where there is one change-point, and the two change-points situation.

First case: one change-point.

In this case, we need three additional parameters to characterize the processes, that is the intensity μ_0 before the “jump”, the intensity $\mu = 1$ after the jump, and the rank k of the time of the change t_k . In all our simulations, we take $\mu_0 = 1$. Clearly, the performance of our approach will depend on our faculty to properly identify the jump, which depends on the size of the jump $|\mu_1 - \mu_0|$ and of the time at which the jump occurs. For this reason, we consider intensity jumps of 20 %, 10 % and 5 % between μ_0 and μ_1 . We consider, for each case, the situation of a “late” (resp. “early”) change-point date, that is $k = 17$ (resp. $k = 7$) over a total number $n = 20$ of dates, and an intermediate situation ($k = 11$).

Table 1 below presents the results for the different settings.

	$k = 17$			$k = 11$			$k = 7$			
	e_1	e_2	e_∞	e_1	e_2	e_∞	e_1	e_2	e_∞	
Jump 20 % $\mu_0 = 1$, $\mu_1 = 1.2$	β	0.050092941	0.003963936	0.248468853	0.055721109	0.004622622	0.214266014	0.054186728	0.004407378	0.257698624
	β (Ef.)	1.186774	1.412986	1.388577	1.297066	1.686361	1.492439	1.349833	1.825504	1.521232
	β (Br.)	1.690324	2.858266	1.781873	1.746075	3.049542	1.822275	1.772479	3.142387	1.844868
	μ_0	0.037107611	0.002300499	0.345238095	0.042708980	0.003366053	0.571428571	0.067267145	0.008067782	0.714285714
	μ_1	0.069407934	0.007896021	0.428938236	0.069889660	0.007453335	0.299418533	0.067273797	0.006925657	0.303741091
	λ^{TRUE}	86.98218	32.77981	451.85678	86.86101	29.98575	666.81994	88.12037	22.82999	939.84461
	λ^0	209.0466	145.9577	774.6417	130.82063	62.31422	1126.11419	95.30184	17.81868	1239.30885
Jump 10 % $\mu_0 = 1$, $\mu_1 = 1.1$	β	0.051709285	0.004133502	0.225043017	0.059351228	0.005332092	0.212276779	0.060646875	0.005410152	0.223756055
	β (Ef.)	1.187414	1.414494	1.413568	1.242805	1.548736	1.421431	1.267657	1.611097	1.451633
	β (Br.)	1.687982	2.850353	1.768820	1.716789	2.948368	1.791424	1.730969	2.997096	1.802645
	μ_0	0.046509656	0.004809497	0.452380952	0.053101563	0.006792678	0.619047619	0.06930684	0.00884062	0.45238095
	μ_1	0.066134710	0.006946465	0.298221301	0.072659852	0.007958981	0.262544256	0.076805279	0.008392841	0.281927396
	λ^{TRUE}	99.09034	34.35615	473.10180	93.59773	30.35525	631.62636	90.54560	24.32775	752.46746
	λ^0	114.72577	43.13866	436.34106	84.91746	23.39024	570.30491	70.39245	11.59017	624.25144
Jump 5 % $\mu_0 = 1$, $\mu_1 = 1.05$	β	0.059356508	0.005394952	0.231662288	0.06163810	0.00564299	0.21691081	0.063459137	0.005919686	0.244816285
	β (Ef.)	1.182138	1.401861	1.371664	1.210940	1.470665	1.394149	1.225724	1.506598	1.429457
	β (Br.)	1.685886	2.843369	1.780099	1.698342	2.885402	1.791498	1.710376	2.926386	1.811282
	μ_0	0.058231361	0.007360461	0.404761905	0.065319651	0.009823075	0.476190476	0.07511100	0.01123867	0.59523810
	μ_1	0.075571031	0.008409454	0.263930833	0.079995921	0.009078988	0.309142683	0.081847629	0.009490729	0.318434829
	λ^{TRUE}	95.65260	29.37298	466.53510	90.72911	26.85267	553.14071	86.79729	23.61415	580.05336
	λ^0	74.68713	17.41288	272.06133	64.17256	12.46306	316.74985	59.091348	8.686976	325.976527

Table 1: Estimation performance in several situations in the one-change-point case, β (Ef.) (resp. β (Br.)) denotes estimator of β using Efron's (resp. Breslow's) approximation of Cox's partial likelihood, λ^{TRUE} denotes the estimated intensity obtained by our procedure, and λ^0 denotes the estimated intensity without considering any change-point.

In this study, it clearly appears that, when we increase the size of the jump, the behaviour of our method is significantly improved for each parameter of the model and for each criterion. Compared to Cox’s partial likelihood approximations, the estimation of β obtained by our method leads to far better estimation results. Besides, estimates of the intensities μ_0 and μ_1 are more accurate when the change-point occurs later (i.e. when k increases). Concerning the global estimation of the hazard function λ , our change-point model (λ^{TRUE}) leads to significant improvements compared to the approach with no change-point (λ^0), except for small sizes of jumps and early date of change-point. In this latter case, we observe slightly poorer results. This mainly comes from the fact that the error in estimating μ_0 , μ_1 and k is higher. However, the estimation of β seems still relevant even in this extreme case.

Second case: two change-points.

In the two change-points case, the process is characterized by $\theta = (\beta, k_1, k_2, \mu_0, \mu_1, \mu_2)$, where (k_1, k_2) are the ranks of the times of change (t_{k_1}, t_{k_2}) and (μ_0, μ_1, μ_2) are the intensities in the three periods defined by the change-points. As in the previous case, we considered different values of these parameters in order to assess the quality of the estimations. We have first simulated datasets where the ranks of the two change-points are quite close (separated by 3 ranks). The change-points are assumed to occur “late” $((k_1, k_2) = (14, 17))$ or at an intermediate date $((k_1, k_2) = (8, 11))$. We also consider the case where the two change-points divide the time in three approximately equal periods $((k_1, k_2) = (7, 11))$. The same jump of 10% is applied at the first change point and jumps of size 0.05, 0.1 or 0.2 from the value 1.1 are used at the second change-point.

Table 2 below presents the results for the different settings.

	$(k_1, k_2) = (14, 17)$				$(k_1, k_2) = (8, 11)$				$(k_1, k_2) = (7, 15)$			
	e_1	e_2	e_∞		e_1	e_2	e_∞		e_1	e_2	e_∞	
$\mu_0 = 1,$ $\mu_1 = 1.1,$ $\mu_2 = 1.3$	β	0.064817713	0.006190545	0.219135497	β	0.057574912	0.005021856	0.208565731	β	0.062534508	0.005973307	0.237327399
	β (Ef.)	1.229069	1.514848	1.397504	β (Ef.)	1.370795	1.882647	1.563088	β (Ef.)	1.311279	1.723167	1.477761
	β (Br.)	1.707914	2.917884	1.790513	β (Br.)	1.779649	3.167793	1.856412	β (Br.)	1.753362	3.075044	1.832443
	μ_0	0.06612649	0.01004802	0.61904762	μ_0	0.07724306	0.01202679	0.47619048	μ_0	0.072414961	0.009407029	0.500000000
	μ_1	0.08908864	0.01414965	0.46696570	μ_1	0.13692390	0.02955524	0.59181152	μ_1	0.10711075	0.01917136	0.53349325
μ_2	0.10179238	0.01535896	0.38695007	μ_2	0.09173299	0.01279416	0.34997984	μ_2	0.10035438	0.01467417	0.33134670	
λ^{TRUE}	120.06754	52.40541	716.98462	λ^{TRUE}	118.39842	50.53192	936.53635	λ^{TRUE}	117.93811	49.19202	919.02605	
λ^0	269.0137	248.2904	1217.5827	λ^0	157.05333	79.28889	1735.06326	λ^0	202.1427	131.6136	1522.3824	
$\mu_0 = 1,$ $\mu_1 = 1.1,$ $\mu_2 = 1.2$	β	0.064260237	0.005961951	0.223232609	β	0.059986231	0.005482251	0.242598120	β	0.064078998	0.006029081	0.232295462
	β (Ef.)	1.219214	1.490809	1.417622	β (Ef.)	1.321616	1.750359	1.509318	β (Ef.)	1.291754	1.672550	1.462749
	β (Br.)	1.703353	2.902389	1.787137	β (Br.)	1.756059	3.084520	1.832974	β (Br.)	1.742895	3.038508	1.818627
	μ_0	0.07482665	0.01169863	0.52380952	μ_0	0.08362933	0.01356685	0.52380952	μ_0	0.08234189	0.01205407	0.47619048
	μ_1	0.08821589	0.01389789	0.54378136	μ_1	0.13430696	0.02873167	0.52514781	μ_1	0.10808110	0.02115315	0.60000000
μ_2	0.09488646	0.01298862	0.30831266	μ_2	0.0919037	0.0126769	0.3263178	μ_2	0.09388448	0.01306089	0.35813276	
λ^{TRUE}	121.72250	51.53993	735.02953	λ^{TRUE}	117.22853	46.53641	947.02413	λ^{TRUE}	120.11474	47.19206	889.99899	
λ^0	174.2237	107.9067	863.8115	λ^0	107.33235	32.49791	1177.55438	λ^0	122.17310	42.38263	1062.88695	
$\mu_0 = 1,$ $\mu_1 = 1.1,$ $\mu_2 = 1.15$	β	0.065146666	0.006358655	0.267650610	β	0.063798466	0.006068773	0.237228090	β	0.061757848	0.005693481	0.220565021
	β (Ef.)	1.213334	1.476361	1.415002	β (Ef.)	1.291749	1.672146	1.447051	β (Ef.)	1.279627	1.641281	1.460097
	β (Br.)	1.704309	2.905729	1.791570	β (Br.)	1.742821	3.038196	1.820148	β (Br.)	1.738093	3.021759	1.809625
	μ_0	0.08330602	0.01432253	0.57142857	μ_0	0.08784710	0.01502091	0.52380952	μ_0	0.09013876	0.01472177	0.500000000
	μ_1	0.0981028	0.0173706	0.5854899	μ_1	0.13949539	0.03164465	0.60000000	μ_1	0.11863639	0.02557955	0.57619048
μ_2	0.09547936	0.01377333	0.36256315	μ_2	0.09287966	0.01340663	0.44032470	μ_2	0.0933072	0.0130673	0.3777881	
λ^{TRUE}	119.08308	48.22871	753.01141	λ^{TRUE}	114.43641	43.64743	946.78327	λ^{TRUE}	117.51992	44.05131	915.34084	
λ^0	130.93558	62.17867	685.58948	λ^0	87.85115	19.70437	891.63818	λ^0	87.96409	20.06260	846.32478	

Table 2: Estimation performance in several situations in the two-change-points case, β (Ef.) (resp. β (Br.)) denotes estimator of β using Efron's (resp. Breslow's) approximation of Cox's partial likelihood, λ^{TRUE} denotes the estimated intensity obtained by our procedure, and λ^0 denotes the estimated intensity without considering any change-point.

Comparing the estimations for $(k_1, k_2) = (14, 17)$ and $(k_1, k_2) = (8, 11)$, we observe, as in previous case, that the estimates of μ_0, μ_1 are more accurate when the change-points occur later. According to the $e_i(\lambda^{\text{TRUE}})$ and $e_i(\lambda^0)$ criterions, as in the previous case, the improvement of global estimation of λ is stronger when change-points occur later ($(k_1, k_2) = (14, 17)$) and weaker or non-existent when changes are smaller and occur earlier ($(k_1, k_2) = (8, 11)$).

However, in the case where the change-points are uniformly located ($(k_1, k_2) = (7, 15)$) our model with the change-points (λ^{TRUE}) provides the best improvement to the no-change-point model (λ^0).

4.2 Comparison with Cox's proportional hazards model

The continuous one-jump counting process that we modelize by (2.1) is only observed at $n + 1$ dates $(t_i)_{0 \leq i \leq n}$. The event times are thus grouped into intervals and it implies many tied event times, especially in the case defined in the application of section 4.3. In this context, as explained by Therneau and Grambsch (2000), two options are possible to apply Cox maximum partial likelihood estimation. The first one is to consider the process as a discrete time process. In addition to this incorrect assumption, the maximization of the exact partial likelihood requires a much important computation time. The second one is to keep the continuous time scale and maximize an approximation of the partial likelihood either from Breslow's or Efron's method (Breslow 1974; Efron 1977). Computationally, these two methods require also an increasing time when the number of ties becomes high. Hertz-Picciotto and Rockhill (1997) compared the two approximations, showing that Efron's method is more accurate especially for a high number of ties.

Our model is neither based on the assumption of discrete time scale nor on an approximate likelihood. Moreover, the algorithm described in section 3.2 does not require a high computational time.

Besides, the specification of the form of the baseline hazard allows a more accurate estimation than in classical semi-parametric context and an easier interpretation of this part of the model. The baseline function represents the intrinsic dynamics of the process while the regression part stands for the influence of the variable on the dynamics. Assumption 2.2 means that we assume $l + 1$ homogeneous periods in the internal dynamics of the process.

In our model, if we increase β , we mechanically increase the number of tied event times. Therefore, we illustrate in Table 3 below how the quality of estimation of the

regression parameter evolves when we increase β . The errors are normalized in order to be able to compare each case. We consider a one change-point situation, with $k = 11$, $\mu_0 = 1$ and $\mu_1 = 1.2$.

	β	-2	-1	1	2
$\left \frac{e_1(\beta)}{\beta} \right $	Cox (Breslow)	0.188381	0.3097856	0.7664887	0.8727165
	Cox (Efron)	0.1061763	0.13179192	0.4283994	0.6483565
	Developed model	0.0790222	0.11637283	0.058707430	0.02751048
$\left \frac{e_2(\beta)}{\beta^2} \right $	Cox (Breslow)	0.04335568	0.1080335	0.5883579	0.7618253
	Cox (Efron)	0.01741894	0.02716132	0.1875493	0.4212818
	Developed model	0.00997855	0.02248492	0.005314939	0.001160346
$\left \frac{e_\infty(\beta)}{\beta} \right $	Cox (Breslow)	0.4422733	0.5956260	0.8544077	0.910485
	Cox (Efron)	0.4889666	0.62851338	0.6369225	0.7309385
	Developed model	0.3361397	0.50231980	0.303411172	0.08143495

Table 3: Empirical behaviour of the estimators of β for different values of β in a one change-point model, $k = 11$, $\mu_0 = 1$, $\mu_1 = 1.2$.

In this table, the error criterions are normalized with respect to β . The approximations of Efron (1977) and Breslow (1974) seem to lead to relative errors that increase with β , which is not the case using our procedure. Moreover, the computation time increases considerably for Cox’s approximations when β increases, while the computation is not impacted in our approach.

4.3 Real data set: reservations to train (SNCF)

The model presented in this paper has been applied in the practical case of reservations data in the French railway company SNCF. Booking process is observed, for a determined tariff class, on $n+1 = 21$ dates $(t_i)_{0 \leq i \leq n}$. $N_j(t)$ denotes the cumulated number of bookings in the train j (for $j = 1, \dots, m = 42$) between 0 and time t . In our particular dataset, we have $Y_j(t) \equiv 1$, although the techniques applies for more general situations. The aim of the study is to identify the impact of “rival” tariff classes on the booking process. We considered the case of a single rival class, corresponding to a less expensive tariff. To measure the impact of this class, we consider the process $X(t)$, taking value 0 if the rival class is still open at time t , and taking value 1 otherwise. In the situation we considered, rival classes do not re-open after being closed, although this situation could be considered through our model.

Estimation

We fix a model by determining the number l of change-points in the baseline hazard function and then estimate the parameter $\theta = (\beta, (\underline{k}, \underline{\mu}))$. The regression parameter β represents the influence of “rival” tariff class or train while the change-point parameters $(\underline{k}, \underline{\mu})$ explain the structural increase (or decrease) of the reservations in the considered tariff class or train.

We considered different models with a fixed number l of change-points. The table 4 below presents the results for $l = 1, \dots, 8$. We explain later how we selected the model that seemed to us the most appropriate.

We observe that, the more change-points are estimated, the less important is the influence of the “rival” class. If we assume 3 change-points (i.e. 4 homogeneous reservation periods), the risk of reservation is multiplied by $e^{0.660} = 1.93$ as soon as the “rival” class is closed to reservation.

l	$\hat{\beta}$	\hat{k}	$\hat{\mu}$
0	3.3426051	X	0.47974
1	0.8344358	11	0.19299
2	0.7206932	5 12	0.69372 8.71990
3	0.6598940	4 8 12	0.03273 0.28189 1.40280
4	0.4292370	4 8 12 19	0.020408 0.2835231 1.487212 8.86125
5	0.4289637	4 8 12 19 20	0.02040816 2.835253 1.487334 8.863116 23.80212
6	0.2814054	4 8 12 17 19 20	0.02040816 0.2843890 1.535951 8.469844 12.19908
7	0.2784701	4 5 8 12 17 19 20	0.02040816 0.1190476 0.3626156 1.536894 8.479895 12.23217
8	0.2673407	4 5 8 11 12 17 19 20	0.02040816 0.1190476 0.3627254 1.334798 3.133703 8.516242 12.35266 27.89254 10.39531

Table 4: Estimates of $\theta = (\beta, k_1, \dots, k_l, \mu_0, \dots, \mu_l)$ for $l = 1, \dots, 8$ for the real data set.

Model selection

In the following figure, we plotted the values of the maximized log-likelihood, with respect to the number of change-points l that are considered in the model. Since the models are nested, this log-likelihood increases with l . After a fast increase between 1 and 5, the increase slows down, and the curve becomes almost linear. A possible model choice would consist of taking $l = 5$ or 6, since after that, the improvement of the log-likelihood does not seem particularly important. Alternatively to this graphical method, one can rely on techniques such as “slope heuristic”, see e.g. Baudry, Maugis and Michel (2012). The technique consists of determining the slope p of the linear part of the log-likelihood curve (here, the part of the curve between $l = 5$ and $l = 16$), and use this slope to determine a penalized criterion. The penalized criterion which is recommended is obtained as follows,

$$\hat{l} = \operatorname{argmax}_l \max_{\theta} L_l(\theta) - 2pl.$$

In our application, we estimate the following value $p = 10.662$ of the slope. In the following table, the penalized criterion $\max_{\theta} L_l(\theta) - 2pl$ is shown for different values of l .

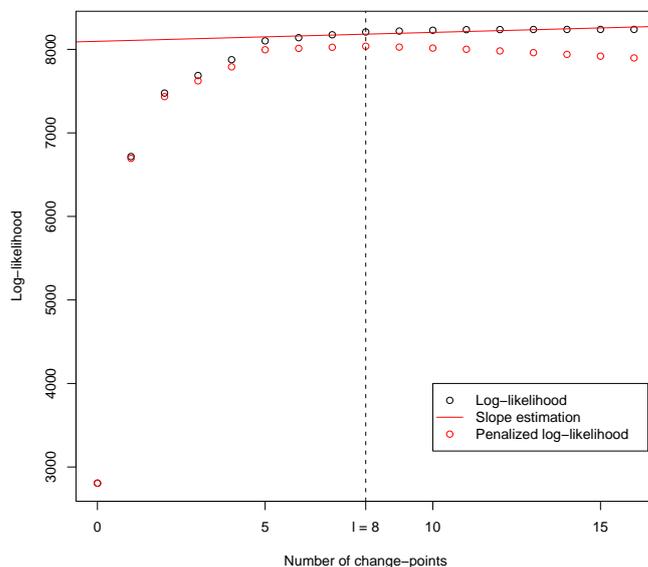


Figure 1: Log-likelihood, penalized likelihood and estimation of linear regression coefficients (slope: 10.66172, intercept: 8096.505).

l	Log-likelihood L_l	$L_l - 2pl$
0	2806.6728312	2806.6728312
1	6716.8506334	6695.527
2	7477.5020833	7434.855
3	7687.344	7623.374
4	7876.631	7791.337
5	8102.203	7995.586
6	8140.190	8012.249
7	8174.818	8025.554
8	8208.273	8037.685
9	8219.261	8027.35
10	8228.928	8015.694
11	8235.892	8001.334
12	8237.372	7981.491
13	8238.118	7960.913
14	8238.532	7940.004
15	8238.85936440	7919.008
16	8238.99457638	7897.82

Table 5: Log-likelihood and penalized log-likelihood for $l = 0, \dots, 16$.

It appears that, according to this criterion, the selected model would be $\hat{l} = 8$. However, since the penalized criterion does not increase strongly between $l = 6$ and $l = 8$, it may suggest that the models with 6 or 7 change-points are still reasonable candidates, if one wishes to promote models with fewer parameters.

4.4 Testing procedure results

Simulations and one-change-point detection

In order to assess the validity of the test procedures that we used on the real data-set, we can try to determine through simulation the power of our test procedure. In table 4.4 below, we present the results of this evaluation in the most simple case, that is when we consider the simplest situation mentioned in section 3.4, where the null hypothesis corresponds to the no change-point situation, which is tested against an alternative corresponding to the one-change point situation. To simulate the law under the null hypothesis, we considered the value of β obtained on the sample, and we took $\mu_0 = 1$. For each testing case (each compartment of the following table), we simulate 1000 samples either from hypothesis H_0 ($\mu_0 = 1$; column 1) or from hypothesis H_1 ($\mu_0 = 1$ and various jumps to μ_1 ; see columns 3-5). In the latter case, we test different positions for the change-point:

late ($k = 17$), middle ($k = 11$) or early ($k = 5$). For every sample, we simulate $B = 100$ observations $(T_n^j)_{1 \leq j \leq B}$ of the test statistic and compare T_n to the obtained threshold s_α ($\alpha = 5\%$). We report in the table the proportion of rejection of H_0 over the 1000 simulated samples.

No change point	One change-point			
$k = 21$		$k = 17$	$k = 11$	$k = 5$
5.2%	$\mu_1 = 1.1$	53%	56.1%	67%
	$\mu_1 = 1.2$	98%	98.4%	99.9%
	$\mu_1 = 1.3$	100%	100%	100%

Table 6: Tests results under hypothesis H_0 and H_1 .

We observe that, under the null hypothesis (which corresponds to the case $k = 21$), the simulation approach almost achieves the expected level which was 5%. The earlier change-point occurs, easier is the detection. Indeed, the results show an increase of the power of the test when the alternative is $k = 11$. Nevertheless, the size of the change, that is the difference between μ_0 and μ_1 , has a more significant impact. In the case where $\mu_1 = 1.2$, change-point is almost systematically detected no matter the time of occurrence.

Forward procedure and multiple-change-point detection

The testing procedure described in section 3.4 is implemented in order to perform the change-point detection in our real data set case. Successive tests for model with l change-points against model with $l + 1$ change-points are performed to select the right number of change-points.

l	Test statistic	Threshold s_α	Rejection of H_0
0	3910.17780	3.79950	Yes
1	760.65144	4.01493	Yes
2	209.84210	4.35410	Yes
3	189.2871	4.36418	Yes
4	225.5715	3.99669	Yes
5	37.98701	4.184904	Yes
6	34.62845	4.128622	Yes
7	33.45466	4.014498	Yes
8	10.98819	4.050255	Yes
9	9.666715	3.830744	Yes
10	6.964444	3.595357	Yes
11	1.479882	3.321706	No
12	0.7458188	2.506621	No
13	0.4144982	1.873998	No
14	0.3268795	1.269713	No
15	0.1352120	0.9744056	No
16	0.3290045	0.6506319	No

Table 7: Tests results for hypothesis H_l : “There are l change-points” against H_{l+1} : “There are $l + 1$ change-points” ($l = 0, \dots, 16$) at the level $\alpha = 5\%$.

Compared to the model selection defined in section 4.3 above, the forward procedure leads to $l = 10$ change-points at the level of $\alpha = 5\%$.

5 Conclusion

In this paper, we proposed a statistical model that is adapted to the study of a counting process which is influenced by some covariates and structural changes in the baseline hazard function. Estimation is performed through an iterative algorithm that combines dynamic programming for the change-point part, and a regular maximum likelihood estimation phase in a multiplicative hazard regression model. Numerical investigation show that our procedures seems to behave correctly, even for small sample sizes. In the particular case of the SNCF data that we studied, they allowed us to detect some change in the dynamics of the reservation which is not completely caused by the fluctuations of the availability for the different categories of seats. They provide a simple modelization of the characteristics of the reservation process, which then can be used in a simulation phase, in order to simulate the impact of different commercial strategies, making this approach quite valuable for Yield-Management.

The proportional hazards model presented in this paper has been developed to describe reservation process but can also be used in recurrent event data analysis as well as survival data analysis as an alternative for parametric and Cox proportional hazards model.

References

- [1] O. Aalen (1978), Nonparametric inference for a family of counting processes, *The Annals of Statistics* **6**, 701–726.
- [2] O. Aalen, O. Borgan, H. K. Gjessing (2008), *Survival and Event History Analysis*, Springer.
- [3] P. K. Andersen, R. D. Gill (1982), Cox’s regression model for counting processes: a large sample study, *The Annals of Statistics* **10**, 1100–1120.
- [4] M. Basseville, I. V. Nikiforov (1993), *Detection of Abrupt Changes: Theory and Application*, Prentice-Hall, Inc.
- [5] J.-P. Baudry, C. Maugis, B. Michel (2012), Slope heuristics: overview and implementation, *Statistics and Computing* **22**, 455–470.
- [6] R. Bellman (1961), On the approximation of curves by line segments using dynamic programming, *Communications of the ACM* **4**, 284.
- [7] N. E. Breslow (1972), Contribution to the discussion of the paper by Dr. Cox, *Journal of the Royal Statistical Society, Series B (Methodological)* **34**, 216–217.
- [8] N. E. Breslow (1974), Covariance analysis of censored survival data, *Biometrics* **30**, 89–99.
- [9] D. R. Cox (1972), Regression models and life-tables, *Journal of the Royal Statistical Society, Series B (Methodological)* **34**, 187–220.
- [10] M. Csörgö, L. Hórvath (1997), *Limit Theorems in Change-Point Analysis*, Wiley, Chichester.
- [11] D. Dabrowska (1997), Smoothed Cox regression, *The Annals of Statistics* **25** (4), 1510–1540.

- [12] B. Efron (1977), The efficiency of Cox's likelihood function for censored data, *Journal of the American Statistical Association* **72**, 557–565.
- [13] D. Ghosh and D. Lin (2003), Semiparametric analysis of recurrent events in the presence of dependent censoring, *Biometrics* **59**, 877–885.
- [14] I. Hertz-Picciotto and B. Rockhill (1997), Validity and efficiency of approximation methods for tied survival times in Cox regression, *Biometrics* **53**, 1151–1156.
- [15] P. Hougaard (2003), Analysis of multivariate survival data, *Springer*.
- [16] B. Jackson et al. (2005), An algorithm for optimal partitioning of data on an interval, *IEEE Signal Processing Letters* **12 (2)**, 105–108.
- [17] E. Lebarbier (2005), Detecting multiple change-points in the mean of Gaussian process by model selection, *Signal Processing* **85**, 717–736.
- [18] D. Y. Lin (2007), On the Breslow estimator, *Lifetime Data Analysis* **13**, 471–480.
- [19] T. Martinussen, T. H. Scheike (2006), Dynamic Regression Models for Survival Data, *Springer*.
- [20] L. A. Sleeper and D.P. Harrington (1990), Regression splines in the Cox model with application to covariate effects in liver disease, *Journal of the American Statistical Association* **85**, 941–949.
- [21] K. T. Talluri, G. J. Van Ryzin (2004), The theory and practice of revenue management. *Kluwer Academic Publishers*.
- [22] T. M. Therneau, P. M. Grambsch (2000), Modeling Survival Data, Extending the Cox Model, *Springer*.
- [23] Y. Xia, H. Tong, W. K. Li, L.-X. Zhu (2002), An adaptive estimation of dimension reduction space, *Journal of the Royal Statistical Society, Series B* **64**, 363–410.
- [24] R. H. Zeni (2001), Improved forecast accuracy in airline Revenue Management by unconstraining demand estimates from censored data, *Dissertation*.