A MATRIX-VALUED COUNTING PROCESS WITH FIRST-ORDER INTERACTIVE INTENSITIES

by

A. C. PEDROSO de LIMA
and
PRANAB K. SEN

Department of Biostatistics
University of North Carolina

Institute of Statistics
Mimeo Series No. 2132

June 1994

A Matrix-Valued Counting Process with First-Order Interactive Intensities

A.C.Pedroso de Lima*   P.K.Sen†

Abstract

A matrix-valued counting process is presented that allows the modelling of multivariate failure-time data. The inclusion of covariates in a Cox-type model is considered and asymptotic properties for the estimates of the parameters involved in the model are studied.

Key words: matrix-valued counting process, multivariate failure-time, martingale, predictable process, Cox-type model.

1 Introduction

The modern theory of counting process and martingales as developed by, e.g., Brémaud (1981) has provided the necessary theoretical background for the development of rigorous and general theory of the regression models adapted to censored data. The seed of such approach seems to reside in the work of Aalen (1975). Much has been done since then and more recently two books have been published in the subject [Fleming and Harrington (1991) and Andersen, Borgan, Gill and Keiding (1993)], that take into account a very broad spectrum of application for the methodology. In this work we propose a model to handle multivariate failure-time data. Our main goal is to consider situations where one is interested in the effect of covariates in more than one event of interest, and, hence, the ultimate objective of the

*Universidade de São Paulo, Brazil and University of North Carolina at Chapel Hill. Supported by CAPES-Brazil.

†University of North Carolina at Chapel Hill.

model is the estimation and hypothesis testing of the parameters involved in the model.

In the next section we introduce the matrix-valued model in the bivariate setup. The development is based on the heuristic interpretation given to the *multiplicative intensity* model presented by Aalen. Then, we consider such model in some parametric models where the corresponding intensity models result in nice interpretable expressions. Following that, we consider the inclusion of covariates in the model and derive some asymptotic results for the parameters involved. Finally, although the results are true in a general setting, we discuss the case of two-sample data with time-independent covariate.

2 The matrix-valued counting process model

In order to develop the model and asymptotic properties we will consider a bivariate model. The extension of the results to the k-variate situation is discussed later. Let $(T_1,\ T_2)$ be non-negative random vector defined in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. In principle we assume that the elements of such vector are not independent, having joint survival function given by $\mathbb{S}_{12}(t_1, t_2)$. The marginal survival functions are represented by $\mathbb{S}_1(t_1)$ and $\mathbb{S}_2(t_2)$. We also assume the presence of censoring, represented by the non-negative random vector $(C_1,\ C_2)$ independent of $(T_1;\ T_2)$, defined in the same probability space of $(T_1,\ T_2)$. Typically in real data one does not observe necessarily T_h or C_h but the minimum between them, represented by $Z_h = T_h \wedge C_h$ and $\delta_h = \mathbb{I}\{Z_h = T_h\}$, where $\mathbb{I}\{A\}$ represents the indicator or characteristic function. Looking more closely to this problem one may note that what is actually being observed are random events occurring in time and, hence, the use of stochastic process to study the situation becomes natural. Therefore, we define the counting processes

$$(2.1) \qquad N_h(t) = \mathbb{I}\{Z_h \le t;\ \delta_h = 1\}, \qquad t \ge 0, \quad h = 1, 2,$$

representing a right-continuous function that assumes value zero, jumping to one when the particular event associated to T_h occurs. Since the quantities in (2.1) are defined on dependent random variables, it makes sense to consider also the random-vector

$$N(t) = \begin{pmatrix} N_1(t) \\ N_2(t) \end{pmatrix}. \tag{2.2}$$

In order to better define the quantities above, we consider a sequence of sub-σ-fields defined by $\{\mathcal{N}_t^{\#},\ t \geq 0\}$ (i.e., a *filtration*), that is the *self-exciting* or *natural* filtration $\sigma\{N^{\#}(s),\ 0 \leq s \leq t\}$, defined by the the vector-valued counting process $N^{\#}(t)$ with elements $N_h^{\#}(t) = \mathbb{I}\{Z_h \leq t\}$. Such filtration can also be made complete in order to satisfy the so-called *les conditions habituelles* [1]. Note that $\mathcal{N}_t^{\#}$ also contains information on the processes $N_h(t)$ as well as on their dependence.

In order to characterize the counting processes above, let us define the $\mathcal{N}_t^{\#}$-*predictable* processes

$$Y_h(t) = \mathbb{I}\{Z_h \geq t\}, \qquad t \geq 0, \quad h = 1, 2, \tag{2.3}$$

that corresponds to the information whether or not the component h is still at risk (i.e., uncensored and alive or working.) Such process is assumed to have its value at instant t known just before t, and this property plays a fundamental role in the martingale property, as we will see later. If we pretend for a moment that the components of N are independent, then the multiplicative intensity model of Aalen (1978) would apply, i.e., the associate intensity process of N_h would be given by

$$\lambda_h(t) = \alpha_h(t) Y_h(t), \qquad h = 1, 2. \tag{2.4}$$

where $\alpha_h(t)$ is the marginal hazard function, defined by

$$\alpha_h(t) = \lim_{\Delta t \to 0} \frac{\mathbb{P}\{T_h \in (t,\ t + \Delta t] \mid T_h > t\}}{\Delta t} \tag{2.5}$$

If we collect the intensity processes defined in (2.4) in a vector $\boldsymbol{\lambda}$, then we could write (under the assumption of independence)

$$\boldsymbol{\lambda}(t) = \begin{pmatrix} \lambda_1(t) \\ \lambda_2(t) \end{pmatrix} = \begin{pmatrix} \alpha_1(t) & 0 \\ 0 & \alpha_2(t) \end{pmatrix} \begin{pmatrix} Y_1(t) \\ Y_2(t) \end{pmatrix} = \boldsymbol{\alpha}(t) \boldsymbol{Y}(t) \tag{2.6}$$

[1] Specifically, a complete, increasing and right continuous sequence of sub-σ-fields

This intensity fully specify the counting process defined in (2.2) when the independence is true. It is our goal now to modify (2.6) in order to get a model for a more general situation where the independence is not feasible. In such case it is expected that the interpretation for the unknown deterministic functions α_h should change and, also, the off-diagonal elements should be different than zero. Let us approach this situation considering a generalization in the heuristic argumentation given for (2.4) [see, e.g., Andersen et al. (1993)) for the univariate case.] In this case one may write

$$\lambda_h(t) = \mathbb{E}\{\ \mathrm{d}N_h(t) \mid \mathcal{N}^{\#}_{t-}\}, \tag{2.7}$$

i.e., the average of jumps for component h given the information available just before t. We may note that in this case $\mathcal{N}^{\#}_{t-}$ contains information whether or not one (or both) component(s) have failed just before t. Since the processes Y_h are predictable, this means we know the value of Y_h at the instant t. If the component h have failed before t, then expression (2.7) equals zero. In other words, we need to consider the situations (i) no component has failed at instant t, i.e., $Y_1(t) = Y_2(t) = 1$; (ii) the first component has failed before t but the second has not, i.e., $Y_1(t) = 0$ and $Y_2(t) = 1$; (iii) only second component has failed before t, that is, $Y_1(t) = 1$ and $Y_2(t) = 0$; and (iv) both components failed before t, in which case $Y_1(t) = Y_2(t) = 0$. If we want to consider the intensity process for the first component, then we only consider cases where $Y_1(t) = 1$. This together with expression (2.7) allow us to write

$$\begin{aligned} \lambda_1(t) &= \mathbb{E}\{\ \mathrm{d}N_1(t) \mid \mathcal{N}^{\#}_{t-}\} \\ &= p_1^{(1)}(t)Y_1(t)[1 - Y_2(t)] + p_2^{(1)}(t)Y_1(t)Y_2(t) \end{aligned} \tag{2.8}$$

where $p_1^{(1)}(t) = \lim_{\Delta t \to 0}(\Delta t)^{-1}\mathbb{P}\{T_1 \in (t, t + \Delta t] \mid T_1 > t; T_2 \leq t\}$ and $p_2^{(1)}(t) = \lim_{\Delta t \to 0}(\Delta t)^{-1}\mathbb{P}\{T_1 \in (t, t + \Delta t] \mid T_1 > t; T_2 > t\}$ may be interpreted as conditional hazard functions, given what happened with the other component. Similarly, for component 2,

$$\begin{aligned} \lambda_2(t) &= \mathbb{E}\{\ \mathrm{d}N_2(t) \mid \mathcal{N}^{\#}_{t-}\} \\ &= p_2^{(2)}(t)Y_2(t)[1 - Y_1(t)] + p_1^{(2)}(t)Y_1(t)Y_2(t) \end{aligned} \tag{2.9}$$

for $p_2^{(2)}(t) = \lim_{\Delta t \to 0}(\Delta t)^{-1}\mathbb{P}\{T_2 \in (t, t + \Delta t] \mid T_1 \leq t; T_2 > t\}$ and $p_1^{(2)}(t) = \lim_{\Delta t \to 0}(\Delta t)^{-1}\mathbb{P}\{T_2 \in (t, t + \Delta t] \mid T_1 > t; T_2 > t\}$.

Based on (2.8) and (2.9) we can represent the intensity process by the product of matrices

$$\boldsymbol{\lambda}(t) = \begin{pmatrix} \lambda_1(t) \\ \lambda_2(t) \end{pmatrix} = \begin{pmatrix} Y_1(t) & 0 \\ 0 & Y_2(t) \end{pmatrix} \begin{pmatrix} \alpha_{11}(t) & \alpha_{12}(t) \\ \alpha_{21}(t) & \alpha_{22}(t) \end{pmatrix} \begin{pmatrix} Y_1(t) \\ Y_2(t) \end{pmatrix}$$

$$= \text{Diag}(\boldsymbol{Y}(t))\boldsymbol{\alpha}(t)\boldsymbol{Y}(t) \tag{2.10}$$

where the elements of $\boldsymbol{\alpha}$ are given by

$$\begin{aligned} \alpha_{11}(t) &= p_1^{(1)}(t) \\ \alpha_{12}(t) &= p_2^{(1)}(t) - p_1^{(1)}(t) \\ \alpha_{21}(t) &= p_1^{(2)}(t) - p_2(2)(t) \\ \alpha_{22}(t) &= p_2^{(2)}(t) \end{aligned}$$

The *matrix-valued counting process model* is defined in the following way. Suppose that $\boldsymbol{N}_1, \ldots, \boldsymbol{N}_n$ are n copies of the process $\boldsymbol{N}$ defined on (2.2). Then the matrix-valued counting process is given by

$$\boldsymbol{N}(t) = (\boldsymbol{N}_1(t), \ldots, \boldsymbol{N}_n(t)), \tag{2.11}$$

with an associated intensity process given by (2.10). Note that the columns of $\boldsymbol{N}$ are independent and each column, in this case, is constituted by 2 dependent elements.

In order to illustrate the bivariate model, let us consider a parametric model in the following example.

EXAMPLE 1 Sarkar (1987) considers an absolutely continuous bivariate exponential distribution where the joint survival function for the vector $(T_1,\ T_2)$ is given by

$$\mathbb{P}\{T_1 \geq t_1; T_2 \geq t_2\}$$

$$= \begin{cases} e^{-(\beta_2+\beta_{12})t_2}\{1 - A(\beta_1 t_2)\}^{-\gamma} A(\beta_1 t_1)^{1+\gamma}, & 0 < t_1 \leq t_2 \\ e^{-(\beta_1+\beta_{12})t_1}\{1 - A(\beta_2 t_1)\}^{-\gamma} A(\beta_2 t_2)^{1+\gamma}, & 0 < t_2 \leq t_1, \end{cases} \tag{2.12}$$

where $\beta_1 > 0$, $\beta_2 > 0$, $\beta_{12} > 0$, $\gamma = \beta_{12}/(\beta_1 + \beta_2)$ and $A(z) = 1 - e^{-z}$, $z > 0$. The model is based on modifications in the characterization property of bivariate exponential distributions that states the following three results are true: (i) T_1 and T_2 are marginally exponential, (ii) $\min(T_1,\ T_2)$ is exponential and (iii) $\min(T_1,\ T_2)$ and $T_1 - T_2$ are independent. Note that by (2.12) if $\beta_{12} = 0$ the joint distribution factorizes in two exponential distribution and then T_1 and T_2 are independent.

Assuming that the survival times are given by (2.12) when no censoring is present, define the bivariate counting process with elements $N_1(t) = \mathbb{I}\{T_1 \leq t\}$, $N_2(t) = \mathbb{I}\{T_2 \leq t\}$ and the predictable processes $Y_1(t) = \mathbb{I}\{T_1 \geq t\}$ and $Y_2(t) = \mathbb{I}\{T_2 \geq t_2\}$. Then, after some long algebraic manipulations we obtain, with $\beta = \beta_1 + \beta_2 + \beta_{12}$,

$$\begin{aligned}
p_1^{(1)}(t) &= \frac{\beta_1 \beta A(\beta_2 t) + \beta_2 \beta_{12}}{(\beta_1 + \beta_2) A(\beta_2 t)}, \\
p_2^{(1)}(t) &= \frac{\beta_1 \beta}{\beta_1 + \beta_2}, \\
p_1^{(2)}(t) &= \frac{\beta_2 \beta A(\beta_1 t) + \beta_1 \beta_{12}}{(\beta_1 + \beta_2) A(\beta_1 t)}, \\
p_2^{(2)}(t) &= \frac{\beta_2 \beta}{\beta_1 + \beta_2}.
\end{aligned}$$

Therefore, the elements of the matrix $\boldsymbol{\alpha}(t)$ may be obtained by taking linear combinations of the quantities $p_i^{(h)}$ as shown after expression (2.10) and, hence, the intensity process can be expressed as

$$\begin{pmatrix} \lambda_1(t) \\ \lambda_2(t) \end{pmatrix}
= \begin{pmatrix} Y_1(t) & 0 \\ 0 & Y_2(t) \end{pmatrix}
\begin{pmatrix} \frac{\beta_1 \beta A(\beta_2 t) + \beta_2 \beta_{12}}{(\beta_1 + \beta_2) A(\beta_2 t)} & -\frac{\beta_2 \beta_{12}}{(\beta_1 + \beta_2) A(\beta_2 t)} \\ -\frac{\beta_1 \beta_{12}}{(\beta_1 + \beta_2) A(\beta_1 t)} & \frac{\beta_2 \beta A(\beta_1 t) + \beta_1 \beta_{12}}{(\beta_1 + \beta_2) A(\beta_2 t)} \end{pmatrix}
\begin{pmatrix} Y_1(t) \\ Y_2(t) \end{pmatrix}.$$

This particular model can also be rewritten in a more interpretable way that takes into account, explicitly, the dependence parameter γ. After few manipulations we obtain

$$\begin{aligned}
\begin{pmatrix} \lambda_1(t) \\ \lambda_2(t) \end{pmatrix} &= \begin{pmatrix} \beta_1 & 0 \\ 0 & \beta_2 \end{pmatrix} \begin{pmatrix} Y_1(t) \\ Y_2(t) \end{pmatrix} \\
&\quad + \gamma \begin{pmatrix} \beta_1 + \frac{\beta_2}{A(\beta_2 t)} & -\frac{\beta_2}{A(\beta_2 t)} Y_1(t) \\ -\frac{\beta_1}{A(\beta_1 t)} Y_2(t) & \beta_2 + \frac{\beta_1}{A(\beta_1 t)} \end{pmatrix} \begin{pmatrix} Y_1(t) \\ Y_2(t) \end{pmatrix}.
\end{aligned}$$

In this expression it is clear the dependence structure in the sense that if $\gamma = 0$ then the resulting expression for the intensity process vector will be the same one would obtain when working with two independent exponential random variables, with parameters β_1 and β_2. □

The matrix-valued model can be thought of when there exists more than 2 components in the model. In this case the intensity process is somewhat more complicate since higher order of combinations of the predictable processes must be taken into account. To illustrate this point, let us consider the case of three components. Therefore, consider the nonnegative random vector $\boldsymbol{T} = (T_1,\ T_2,\ T_3)'$, where each element represents the time up to the occurrence of events of interest. Similarly to the bivariate case, define the matrix-valued counting process (2.11) where each column now is given by a 3×1 vector of counting processes $\boldsymbol{N}_i = (N_{1i},\ N_{2i},\ N_{3i})'$, based on n copies $\boldsymbol{T}_i$ of $\boldsymbol{T}$. Also, the predictable vector is given by $\boldsymbol{Y}_i = (Y_{1i},\ Y_{2i},\ Y_{3i})'$, where $Y_{hi}(t) = \mathbb{I}\{Z_{hi} \geq t\}$.

In order to compute the intensity processes we need to consider the $2^3 = 8$ possibilities represented by the combinations of 0's and 1's of the elements of the vector $\boldsymbol{Y}_i(t)$. Since only makes sense to consider the intensity for a component which has not failed yet, only four combinations are considered when computing the intensity for each component (those for which the corresponding predictable process is not zero at time t). For example, let us consider the first component. We consider only the cases where $Y_{1i}(t) = 1$ because when this is not true, the component has already failed and the conditional hazard function will be zero. Therefore, the first element of $\boldsymbol{Y}$ will be fixed and there are $2^2 = 4$ possibilities to be considered, represented by the failure or not of the other two components. The following notations are then defined for the conditional hazard functions (dropping out the subscript i to simplify the notation)

- When $Y_1(t) = Y_2(t) = Y_3(t) = 1$ no component has failed at time t and the conditional hazard is given by

$$p_{123}^{(1)}(t) = \lim_{\Delta t \to 0} \frac{\mathbb{P}\{T_1 \in (t,\ t + \Delta t] \mid T_1 > t,\ T_2 > t,\ T_3 > t\}}{\Delta t};$$

- when $Y_1(t) = Y_2(t) = 1$ and $Y_3(t) = 0$, components 1 and 2 have not failed and component 3 failed before t, so that the conditional hazard is

$$p_{12}^{(1)}(t) = \lim_{\Delta t \to 0} \frac{\mathbb{P}\{T_1 \in (t,\ t+\Delta t] \mid T_1 > t,\ T_2 > t,\ T_3 \le t\}}{\Delta t};$$

- when $Y_1(t) = Y_3(t) = 1$ and $Y_2(t) = 0$, only component 2 has failed before t and so,

$$p_{13}^{(1)}(t) = \lim_{\Delta t \to 0} \frac{\mathbb{P}\{T_1 \in (t,\ t+\Delta t] \mid T_1 > t,\ T_2 \le t,\ T_3 > t\}}{\Delta t};$$

- when $Y_1(t) = 1$ and $Y_2(t) = Y_3(t) = 0$, only component 1 has not failed and in this situation the conditional hazard function will be denoted by

$$p_{1}^{(1)}(t) = \lim_{\Delta t \to 0} \frac{\mathbb{P}\{T_1 \in (t,\ t+\Delta t] \mid T_1 > t,\ T_2 \le t,\ T_3 \le t\}}{\Delta t},$$

so that the intensity process will be given by one of the four expressions above, that, depending on the value of $\boldsymbol{Y}$, can be written as,

(2.13)

$$\begin{aligned}
\lambda_1(t) =& p_1^{(1)}(t)Y_1(t)[1-Y_2(t)][1-Y_3(t)] \\
&+ p_{12}^{(1)}(t)Y_1(t)Y_2(t)[1-Y_3(t)] \\
&+ p_{13}^{(1)}(t)Y_1(t)[1-Y_2(t)][1-Y_3(t)] \\
&+ p_{123}^{(1)}(t)Y_1(t)Y_2(t)Y_3(t) \\
=& p_1^{(1)}(t)Y_1(t) + (p_{13}^{(1)} - p_1^{(1)})Y_1(t)Y_3(t) + (p_{12}^{(1)}(t) - p_1^{(1)}(t))Y_1(t)Y_2(t) \\
&+ (p_{123}^{(1)}(t) - p_{12}^{(1)}(t) - p_{13}^{(1)}(t) + p_1^{(1)}(t))Y_1(t)Y_2(t)Y_3(t) \\
=& \alpha_1^{(1)}(t)Y_1(t) + \alpha_{13}^{(1)}(t)Y_1(t)Y_3(t) + \alpha_{12}^{(1)}(t)Y_1(t)Y_2(t) \\
&+ \alpha_{123}^{(1)}(t)Y_1(t)Y_2(t)Y_3(t).
\end{aligned}$$

The same scheme applies for the second and third components, with only changes in notation, such that, for the second component,

(2.14)

$$\begin{aligned}
\lambda_2(t) =& p_2^{(2)}(t)Y_2(t) + (p_{23}^{(2)} - p_2^{(2)})Y_2(t)Y_3(t) + (p_{12}^{(2)}(t) - p_2^{(2)}(t))Y_1(t)Y_2(t) \\
&+ (p_{123}^{(2)}(t) - p_{12}^{(2)}(t) - p_{23}^{(2)}(t) + p_2^{(2)}(t))Y_1(t)Y_2(t)Y_3(t) \\
=& \alpha_2^{(2)}(t)Y_2(t) + \alpha_{23}^{(2)}(t)Y_2(t)Y_3(t) + \alpha_{12}^{(2)}(t)Y_1(t)Y_2(t) \\
&+ \alpha_{123}^{(2)}(t)Y_1(t)Y_2(t)Y_3(t),
\end{aligned}$$

and for the third component the intensity process will be given by

(2.15)
$$\begin{aligned}\lambda_3(t) =& p_3^{(3)}(t)Y_3(t) + (p_{23}^{(3)} - p_3^{(3)})Y_2(t)Y_3(t) + (p_{13}^{(3)}(t) - p_3^{(3)}(t))Y_1(t)Y_3(t) \\ &+ (p_{123}^{(3)}(t) - p_{13}^{(3)}(t) - p_{23}^{(3)}(t) + p_3^{(3)}(t))Y_1(t)Y_2(t)Y_3(t) \\ =& \alpha_3^{(3)}(t)Y_3(t) + \alpha_{23}^{(3)}(t)Y_2(t)Y_3(t) + \alpha_{13}^{(3)}(t)Y_1(t)Y_3(t) \\ &+ \alpha_{123}^{(3)}(t)Y_1(t)Y_2(t)Y_3(t).\end{aligned}$$

Based on expressions (2.13)–(2.15) we may note that each expression has a term involving the predictable process for the corresponding component, $\binom{3}{2}$ terms involving the product of two predictable processes and one term involving the product of the three processes $Y_h,\ h = 1, 2, 3$. This structure can resemble the models used in analysis of variance or categorical data, where usually one considers models involving the main effects and first or higher order interactions. When collecting all three quantities defined above in a vector of intensity process, one may write the model trying to emphasize this,

(2.16)
$$\begin{aligned}\boldsymbol{\lambda}(t) = \begin{pmatrix}\lambda_1(t)\\ \lambda_2(t)\\ \lambda_3(t)\end{pmatrix} &= \begin{pmatrix}\alpha_1^{(1)}(t)\\ 0\\ 0\end{pmatrix} Y_1(t) + \begin{pmatrix}0\\ \alpha_2^{(2)}(t)\\ 0\end{pmatrix} Y_2(t) + \begin{pmatrix}0\\ 0\\ \alpha_3^{(3)}(t)\end{pmatrix} Y_3(t) \\ &+ \begin{pmatrix}\alpha_{12}^{(1)}(t)\\ \alpha_{12}^{(2)}(t)\\ 0\end{pmatrix} Y_1(t)Y_2(t) + \begin{pmatrix}\alpha_{13}^{(1)}(t)\\ 0\\ \alpha_{13}^{(3)}(t)\end{pmatrix} Y_1(t)Y_3(t) + \begin{pmatrix}0\\ \alpha_{23}^{(2)}(t)\\ \alpha_{23}^{(3)}(t)\end{pmatrix} Y_2(t)Y_3(t) \\ &+ \begin{pmatrix}\alpha_{123}^{(1)}(t)\\ \alpha_{123}^{(2)}(t)\\ \alpha_{123}^{(3)}(t)\end{pmatrix} Y_1(t)Y_2(t)Y_3(t)\end{aligned}$$

where the first three terms in the r.h.s. of expression (2.16) represent the main effects, the following three terms the first order interaction and the last term the second order interaction.

An alternative way of expressing the model is to write (2.16) as a product of matrix similarly the one in (2.10), given by

$$\boldsymbol{\lambda}(t) = \begin{pmatrix} Y_1(t) & 0 & 0 \\ 0 & Y_2(t) & 0 \\ 0 & 0 & Y_3(t) \end{pmatrix} \begin{pmatrix} \alpha_{11}(t) & \alpha_{12}(t) & \alpha_{13}(t) \\ \alpha_{21}(t) & \alpha_{22}(t) & \alpha_{23}(t) \\ \alpha_{31}(t) & \alpha_{32}(t) & \alpha_{33}(t) \end{pmatrix} \begin{pmatrix} Y_1(t) \\ Y_2(t) \\ Y_3(t) \end{pmatrix}$$

$$+ \begin{pmatrix} \alpha_{123}^{(1)}(t) \\ \alpha_{123}^{(2)}(t) \\ \alpha_{123}^{(3)}(t) \end{pmatrix} Y_1(t)Y_2(t)Y_3(t), \tag{2.17}$$

where the elements α_{ij} are defined by the equality

$$\begin{pmatrix} \alpha_1^{(1)}(t) & \alpha_{12}^{(1)}(t) & \alpha_{13}^{(1)}(t) \\ \alpha_{12}^{(2)}(t) & \alpha_2^{(2)}(t) & \alpha_{23}^{(2)}(t) \\ \alpha_{13}^{(3)}(t) & \alpha_{23}^{(3)}(t) & \alpha_3^{(3)}(t) \end{pmatrix} = \begin{pmatrix} \alpha_{11}(t) & \alpha_{12}(t) & \alpha_{13}(t) \\ \alpha_{21}(t) & \alpha_{22}(t) & \alpha_{23}(t) \\ \alpha_{31}(t) & \alpha_{32}(t) & \alpha_{33}(t) \end{pmatrix}.$$

Assuming that the second order interaction is negligible, expression (2.14) turns out to be

$$\boldsymbol{\lambda}(t) = \text{Diag}(\boldsymbol{Y}(t))\boldsymbol{\alpha}(t)\boldsymbol{Y}(t), \tag{2.18}$$

that is similar to (2.10).

The same reasoning can be considered for higher dimension problems, with the additional complication that one has to deal with higher order interactions. For example, for a K component problem, the intensity process will involve up to the $(K-1)$th order interaction. In fact, in this situation, each component will have intensity process that can be written as

$$\lambda_j(t) = \beta_j(t)Y_j(t)(1 + \sum_{\substack{l=1 \\ l \neq j}}^{K} \gamma_{jl}(t)Y_l(t) + \sum_{\substack{l=1 \\ l \neq l' \\ l \neq j}}^{K} \delta_{jll'}(t)Y_l(t)Y_{l'}(t) + \dots)$$

where $\beta_j = \alpha_{jj}^{(j)}$, $\gamma_{jl} = \alpha_{jl}^{(j)}/\alpha_{jj}^{(j)}$, $\delta_{jll'}^{(j)} = \alpha_{jll'}^{(j)}/\alpha_{jj}^{(j)}$ depend on the conditional hazard functions as in the case $K = 3$. In this representation the first term inside parenthesis is related to the independent situation, the second term with the first order interaction, and so on. If it is reasonable to assume that the second and higher order interactions are null, then, this model can be rewritten as in (2.10) and (2.18), i.e., in general,

$$(2.19)\ \boldsymbol{\lambda}(t) = \begin{pmatrix} Y_1(t) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & Y_K(t) \end{pmatrix} \begin{pmatrix} \alpha_{11}(t) & \dots & \alpha_{1K}(t) \\ \vdots & \ddots & \vdots \\ \alpha_{K1}(t) & \dots & \alpha_{KK}(t) \end{pmatrix} \begin{pmatrix} Y_1(t) \\ \vdots \\ Y_K(t) \end{pmatrix}.$$

We finish this section with a brief remark about the assumption that the second or higher order interactions are null. It may be noted that the model represented by expression (2.19) has K^2 infinite-dimensional parameters represented by the functions α_{ij}. As we will see later, additional assumptions are imposed on such model in order to reduce the dimensionality of the parameter space and allow us to estimate as well as develop asymptotic properties for the corresponding estimators. If the higher order interactions are allowed in the model, then the problem becomes much more complex in the sense that additional assumptions will have to be made. Assuming that interactions are null is a common practice in some fields of Statistics and we will also consider this approach since we believe the simplifications are considerable; however, further investigation on the implications of such assumption is needed. A more careful examination shows that this assumption implies in assuming that the failure times are *conditionally independent*, e.g., when $K = 3$, assuming that there is no second order interaction is equivalent to say that, given one of the failure times, the other two are independent. Finally we note that in the case where $K = 2$ no assumption is needed since the model will involve only first order interactions that are being taken into account in model (2.19).

3 The bivariate model with covariates

In this section we consider the bivariate model specified by (2.10) and assume a Cox-type of model in order to include covariates. Throughout this section we will assume that $t \in [0, \tau]$, $\tau > 0$ and, in addition to the quantities defined earlier, we also have a set of time-dependent covariates $X_1(t), \dots, X_q(t)$. The covariates are assumed to be observed for all individuals. In order to simplify the notation we consider the case $q = 1$ and in some expressions we will omit t from the notation for the processes involved. Also, we consider the censoring variables are the same for all components, i.e., $C_1 = C_2 = C$ such that the observable variable is given by $Z_i = T_i \wedge C$, $i = 1, 2$.

If n represents the number of individuals, let $\boldsymbol{N}_1, \dots, \boldsymbol{N}_n$ be copies of $\boldsymbol{N}$ defined in (2.2) with corresponding intensity processes given by $\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_n$, where

$$\boldsymbol{\lambda}_i(t) = \begin{pmatrix} \lambda_{1i}(t) \\ \lambda_{2i}(t) \end{pmatrix} = \begin{pmatrix} \alpha_{11}^i(t)Y_{1i}(t) + \alpha_{12}^i(t)Y_{1i}(t)Y_{2i}(t) \\ \alpha_{22}^i(t)Y_{2i}(t) + \alpha_{21}^i(t)Y_{1i}(t)Y_{2i}(t) \end{pmatrix}. \tag{3.1}$$

Since for each individual i one observes the covariate X_i, $i = 1, \dots, n$, we assume that each element of $\boldsymbol{\alpha}$ can be expressed through a multiplicative form such that

$$\begin{array}{l} p_{1i}^{(1)}(t) = \gamma_{11}(t)e^{\beta_1 X_i} \\ p_{2i}^{(1)}(t) = \gamma_{12}(t)e^{\beta_1 X_i} \\ p_{1i}^{(2)}(t) = \gamma_{21}(t)e^{\beta_2 X_i} \\ p_{2i}^{(2)}(t) = \gamma_{22}(t)e^{\beta_2 X_i} \end{array} \Rightarrow \begin{cases} \alpha_{11}^i = \gamma_{11}(t)e^{\beta_1 X_i} = \alpha_{11}^0(t)e^{\beta_1 X_i} \\ \alpha_{12}^i = (\gamma_{12}(t) - \gamma_{11}(t))e^{\beta_1 X_i} = \alpha_{12}^0(t)e^{\beta_1 X_i} \\ \alpha_{21}^i = (\gamma_{21}(t) - \gamma_{22}(t))e^{\beta_2 X_i} = \alpha_{21}^0(t)e^{\beta_2 X_i} \\ \alpha_{22}^i = \gamma_{22}(t)e^{\beta_2 X_i} = \alpha_{22}^0(t)e^{\beta_2 X_i} \end{cases}$$

In addition, we simplify further the model with the (strong) assumption that $\alpha_{11}^0(t) = \theta_1^{-1}\alpha_{12}^0(t)$ and $\alpha_{22}^0(t) = \theta_2^{-1}\alpha_{21}^0(t)$, for $\theta_h > -1$, $h = 1, 2$. Then, the intensity process vector can be written as

$$\boldsymbol{\lambda}_i(t) = \begin{pmatrix} \alpha_{11}^0(t)(Y_{1i}(t) + \theta_1 Y_{1i}(t)Y_{2i}(t))e^{\beta_1 X_i} \\ \alpha_{22}^0(t)(Y_{2i}(t) + \theta_2 Y_{1i}(t)Y_{2i}(t))e^{\beta_2 X_i} \end{pmatrix} \tag{3.2}$$

Based on that, the problem at hand consists in finding estimates for β_j and θ_j, $j = 1, 2$. Since both failure times (for the two components) are assumed to be observed at the exact instant they occur, we have that no two components can jump at the same instant t for the same subject and, hence, when estimating β we will consider a likelihood whose contribution of individual i at time t, if any, will be restricted to $\lambda_{1i}(t)/\sum_{j=1}^n \lambda_{1j}(t)$ when N_{1i} jumps or $\lambda_{2i}(t)/\sum_{j=1}^n \lambda_{2j}(t)$ when N_{2i} jumps.

The likelihood can then be written as the product of the two ratios above and considering the proportionality assumption for the off-diagonal terms in $\boldsymbol{\alpha}(t)$ we are able to cancel the unknown baseline functions $\alpha_{ij}^0(t)$. Let

$\boldsymbol{\delta} = (\boldsymbol{\delta}_1,\ \boldsymbol{\delta}_2)$, with $\boldsymbol{\delta}_k = (\beta_k,\ \theta_k)$ and suppose the true parameter value is represented by $\boldsymbol{\delta}^0$. Then, the likelihood can be expressed as

$$L(\boldsymbol{\delta}) = \prod_{t\geq 0}\prod_{i=1}^{n}\prod_{h=1}^{2}\left(\frac{\lambda_{hi}(t)}{\sum_{j=1}^{n}\lambda_{hj}(t)}\right)^{\mathrm{d}N_{hi}(t)}$$

$$= \prod_{t\geq 0}\prod_{i=1}^{n}\left(\frac{\alpha_{11}^{0}(t)e^{\beta_1 X_i}(Y_{1i}(t)+\theta_1 Y_{1i}(t)Y_{2i}(t))}{\sum_{j=1}^{n}\alpha_{11}^{0}(t)e^{\beta_1 X_j}(Y_{1j}(t)+\theta_1 Y_{1j}(t)Y_{2j}(t))}\right)^{\mathrm{d}N_{1i}(t)}$$

$$\times\left(\frac{\alpha_{22}^{0}(t)e^{\beta_2 X_i}(Y_{2i}(t)+\theta_2 Y_{1i}(t)Y_{2i}(t))}{\sum_{j=1}^{n}\alpha_{22}^{0}(t)e^{\beta_2 X_j}(Y_{2j}(t)+\theta_2 Y_{1j}(t)Y_{2j}(t))}\right)^{\mathrm{d}N_{2i}(t)}$$

$$= \prod_{t\geq 0}\prod_{i=1}^{n}\prod_{h=1}^{2}\left(\frac{e^{\beta_h X_i}(Y_{hi}(t)+\theta_h Y_{1i}(t)Y_{2i}(t))}{\sum_{j=1}^{n}e^{\beta_h X_j}(Y_{hj}(t)+\theta_h Y_{1j}(t)Y_{2j}(t))}\right)^{\mathrm{d}N_{hi}(t)} \tag{3.3}$$

Expression (3.3) is in fact a partial likelihood and it can be thought of as a multinomial type of likelihood where at a given time t there a certain probability that one of the components will fail. Another justification for this expression can be given using the concept of *profile* likelihood along the same lines as presented by Andersen et al. (1993), pages 481-482. The log-likelihood is given by

$$\log L(\boldsymbol{\delta}) = \int_0^{\tau}\sum_{i=0}^{n}\sum_{h=1}^{2}\beta_h X_i(t) + \log\{Y_{hi}(t)+\theta_h Y_{1i}(t)Y_{2i}(t)\}$$
$$-\log\{\sum_{j=1}^{n}e^{\beta_h X_j(t)}(Y_{hj}(t)+\theta_h Y_{1j}(t)Y_{2j}(t))\}\ \mathrm{d}N_{hi}(t)$$

Computation of the score vector takes place for a pair of parameter for each component. Therefore it is convenient to consider a partitioned vector where the first element is a 2×1 vector containing the derivatives of the log-likelihood with respect to the parameters related to the first component and the same quantities for the second element with information related to the second component. It should be noted that the score vector is also a stochastic process in $[0,\tau]$. For $t=\tau$, we write

$$\boldsymbol{U}(\tau;\boldsymbol{\delta}) = \begin{pmatrix}\boldsymbol{U}^{(1)}(\tau;\boldsymbol{\delta})\\ \boldsymbol{U}^{(2)}(\tau;\boldsymbol{\delta})\end{pmatrix},$$

where the first element of $\boldsymbol{U}^{(h)} = (U_1^{(h)}, U_2^{(h)})'$, $h = 1,2$ is given by

$$U_1^{(h)}(\tau;\boldsymbol{\delta}) = \frac{\partial \log L(\boldsymbol{\delta})}{\partial \beta_h}$$

$$(3.4) \qquad = \int_0^\tau \sum_{i=1}^n \left(X_i - \frac{\sum_{j=1}^n X_j w_j(\boldsymbol{\delta}_h)}{\sum_{j=1}^n w_j(\boldsymbol{\delta}_h)} \right) \, \mathrm{d}N_{hi},$$

where $w_j(\boldsymbol{\delta}_h) = e^{\beta_h X_j}(Y_{hj} + \theta_h Y_{1j} Y_{2j})$. Note that (3.4) is similar to the expression one obtains when considering the univariate case. The basic difference relies on the *weights* w_j that are taking into account the predictable processes associated with both components. The second element for the score vector is given by

$$U_2^{(h)}(\tau;\boldsymbol{\delta}) = \frac{\partial \log L(\boldsymbol{\delta})}{\partial \theta_h}$$

$$(3.5) \qquad = \int_0^\tau \sum_{i=1}^n \left(\frac{Y_{1i}Y_{2i}}{Y_{hi} + \theta_h Y_{1i} Y_{2i}} - \frac{\sum_{j=1}^n e^{\beta_h X_j} Y_{1j} Y_{2j}}{\sum_{j=1}^n w_j(\boldsymbol{\delta}_h)} \right) \, \mathrm{d}N_{hi}.$$

Maximum partial likelihood estimators (MPLE) can be obtained by solving the equations

$$(3.6) \qquad \boldsymbol{U}(\tau;\boldsymbol{\delta}) = \mathbf{0},$$

which need to be computed iteratively since no analytical expression for the estimators can be derived. Let us denote the PMLE by $\hat{\boldsymbol{\delta}}$. Asymptotic properties for such estimator are studied using the standard martingale theory, that implies in computing the information matrix, derive the martingale property for some of the quantities involved and making use of Taylor's expansions. Hence, we first note that expressions (3.4) and (3.5) can be written as martingales when $\boldsymbol{\delta} = \boldsymbol{\delta}^0$. This is done in a similar manner as in the univariate case, developed in Andersen and Gill (1982). First we note that $\mathrm{d}N_{hi}(t) = N_{hi}(t) - N_{hi}(t^-) = \mathrm{d}M_{hi}(t) - \lambda_{hi}(t)\,\mathrm{d}t$, where $M_{hi}(t)$ is a local square integrable martingale. Plugging this quantity into expressions (3.4) and (3.5) we obtain a difference of two integrals, one involving the martingale M_{hi} and the other involving the intensity process. It turns out that the latter is given by

$$\int_0^\tau \sum_{i=1}^n \left(X_i - \frac{\sum_{j=1}^n X_j e^{\beta_h^0 X_j}(Y_{hj} + \theta_h^0 Y_{1j} Y_{2j})}{\sum_{j=1}^n e^{\beta_h^0 X_j}(Y_{hj} + \theta_h^0 Y_{1j} Y_{2j})} \right) w_i(\boldsymbol{\delta}_h^0)\alpha_{hh}^0 \, \mathrm{d}t = 0.$$

Therefore, it follows that the first element of the score function can be written as

$$(3.7) \qquad U_1^{(h)}(\tau;\boldsymbol{\delta}^0) = \int_0^\tau \sum_{i=1}^n \left(X_i - \frac{\sum_{j=1}^n X_j w_j(\boldsymbol{\delta}_h^0)}{\sum_{j=1}^n w_j(\boldsymbol{\delta}_h^0)} \right) \mathrm{d}M_{hi},$$

that is a linear combination of integrals depending on predictable processes and in the square integrable martingales, and, hence, (3.7) is also a square integrable martingale. A similar result follows for the second element of the score vectors, such that

$$(3.8) \quad U_2^{(h)}(\tau;\boldsymbol{\delta}^0) = \int_0^\tau \sum_{i=1}^n \left(\frac{Y_{1i}Y_{2i}}{Y_{hi} + \theta_h^0 Y_{1i}Y_{2i}} - \frac{\sum_{j=1}^n e^{\beta_h^0 X_j} Y_{1j}Y_{2j}}{\sum_{j=1}^n w_j(\boldsymbol{\delta}_h^0)} \right) \mathrm{d}M_{hi},$$

also a square integrable martingale. Therefore, the vector-valued score function can be thought of a square integrable martingale. This fact is used to derive the asymptotic distribution for the score function, based on the Rebolledo's central limit theorem for martingales. Such theorem assumes that the *predictable variation processes* satisfy certain conditions. In order to compute such processes denoted by $\langle \cdot \ , \ \cdot \rangle$, the following processes are defined (extending those quantities usually considered in the literature), .

$$\begin{aligned}
S_h^{(j)}(\boldsymbol{\delta},x) &= (1/n) \sum_{i=1}^n X_i^j(x)(Y_{hi}(x) + \theta_h Y_{1i}(x)Y_{2i}(x)) e^{\beta_h X_i(x)}, \quad j = 0,1,2, \\
S_h^{(3)}(\boldsymbol{\delta},x) &= (1/n) \sum_{i=1}^n e^{\beta_h X_i(x)} Y_{1i}(x)Y_{2i}(x), \\
S_h^{(4)}(\boldsymbol{\delta},x) &= (1/n) \sum_{i=1}^n \frac{Y_{1i}(x)Y_{2i}(x)}{Y_{hi}(x) + \theta_h Y_{1i}(x)Y_{2i}(x)} e^{\beta_h X_i(x)}, \\
S_h^{(5)}(\boldsymbol{\delta},x) &= (1/n) \sum_{i=1}^n X_i(x)Y_{1i}(x)Y_{2i}(x) e^{\beta_h X_i(x)}.
\end{aligned}$$

Defining also $U_j^{(h,n)} = n^{-1/2} U_j^{(h)}$, $j = 1,2$ and using well-known properties of the predictable processes involved, the predictable process of $U_1^{(h,n)}$ will be given by

$$(3.9) \qquad \langle U_1^{(h,n)}, U_1^{(h,n)} \rangle(t) = \int_0^t \left(S_h^{(2)}(\boldsymbol{\delta}^0,s) - \frac{(S_h^{(1)}(\boldsymbol{\delta}^0,s))^2}{S_h^{(0)}(\boldsymbol{\delta}^0,s)} \right) \alpha_{hh}^0 \, \mathrm{d}s.$$

Similarly, the predictable variation processes for $U_2^{(h,n)}$ and predictable co-variation processes between $U_1^{(h,n)}$ and $U_2^{(h,n)}$ are given by

$$(3.10)\qquad \langle U_2^{(h,n)}, U_2^{(h,n)}\rangle(t) = \int_0^t \left(S_h^{(4)}(\boldsymbol{\delta}^0) - \frac{(S_h^{(3)}(\boldsymbol{\delta}^0))^2}{S_h^{(0)}(\boldsymbol{\delta}^0)} \right) \alpha_{hh}^0 \, \mathrm{d}s$$

and

$$(3.11)\quad \langle U_1^{(h,n)}, U_2^{(h,n)}\rangle(t) = \int_0^t \left\{ S_h^{(5)}(\boldsymbol{\delta}^0) - \frac{S_h^{(1)}(\boldsymbol{\delta}^0) S_h^{(3)}(\boldsymbol{\delta}^0)}{S_h^{(0)}(\boldsymbol{\delta}^0)} \right\} \alpha_{hh}^0 \, \mathrm{d}s.$$

Given the processes above, we state now a list of conditions that will be used in the proofs for the theorems concerning the asymptotic properties of quantities of interest. Such conditions are based on those assumed in the univariate case.

CONDITIONS:

C.1. $\int_0^\tau \alpha_{hh}(s)\, \mathrm{d}s < \infty$,

C.2. For a neighborhood $\mathcal{D}$ around the true value for the parameter vector, $\sup_{\{t\in[0,\tau];\, \delta\in\mathcal{D}\}} \|S_h^{(j)}(\boldsymbol{\delta}, t) - s_h^{(j)}(\boldsymbol{\delta}, t)\| \xrightarrow{\mathbb{P}} 0$, with $s_h^{(0)}$ bounded away from zero on $\mathcal{D} \times [0, \tau]$ and $s_h^{(j)}$ continuous functions of $\boldsymbol{\delta}$ on $\mathcal{D}$, $j = 1\ldots, 5$,

C.3. for all $\gamma > 0$, $n^{-1/2} \sup_{\{1\le i\le n;\, t\in[0,\tau]\}} |X_i| Y_{hi} \mathbb{I}\{\beta_h^0 X_i > -\gamma |X_i|\} \xrightarrow{\mathbb{P}} 0$, and

C.4. The matrix Σ is positive definite.

Now we state the following theorem, describing the asymptotic distribution for the score function.

THEOREM 1 *For a bivariate counting process with intensity process defined by (3.2) assume that conditions* C.1–C.3 *are true. Then, the stochastic process* $n^{-1/2}\boldsymbol{U}$*, with* $\boldsymbol{U}$ *defined in (3.4)-(3.5) converges in distribution to a continuous Gaussian martingale* $\boldsymbol{W}$ *with covariance function given by*

$$\Sigma(t) = \begin{pmatrix} \Sigma_1(t) & \mathbf{0} \\ \mathbf{0} & \Sigma_2(t) \end{pmatrix}$$

where $\Sigma_h(t)$ *is a matrix with elements*

$$(\Sigma_h(t))_{11} = \int_0^t \left(s_h^{(2)}(\boldsymbol{\delta}^0, s) - \frac{(s_h^{(1)}(\boldsymbol{\delta}^0, s))^2}{s_h^{(0)}(\boldsymbol{\delta}^0, s)} \right) \alpha_{hh}^0(s)\ \mathrm{d}s$$

$$(\Sigma_h(t))_{22} = \int_0^t \left(s_h^{(4)}(\boldsymbol{\delta}^0, s) - \frac{(s_h^{(3)}(\boldsymbol{\delta}^0, s))^2}{s_h^{(0)}(\boldsymbol{\delta}^0, s)} \right) \alpha_{hh}^0(s)\ \mathrm{d}s$$

$$(\Sigma_h(t))_{12} = \int_0^t \left(s_h^{(5)}(\boldsymbol{\delta}^0, s) - \frac{s_h^{(1)}(\boldsymbol{\delta}^0, s) s_h^{(3)}(\boldsymbol{\delta}^0, s)}{s_h^{(0)}(\boldsymbol{\delta}^0, s)} \right) \alpha_{hh}^0(s)\ \mathrm{d}s$$

Remark: Since $\boldsymbol{W}$ is a continuous Gaussian martingale, the cross-covariance function $\mathbb{E}\{\boldsymbol{W}(s)[\boldsymbol{W}(t)]'\}$ is given by $\Sigma(s \wedge t)$.

Proof: The proof is based on the paper by Andersen and Gill (1982) with some modifications to include our more general setup. It is based on the Rebolledo's central limit theorem and basically we have to show that (i) the predictable processes in (3.9)–(3.11) converge to deterministic functions and (ii) the predictable variation processes converge to continuous functions as $n \to \infty$.

First we note that since it is assumed that no two components can fail at the same time,

$$\langle U_k^{(h,n)}, U_l^{(j,n)} \rangle(t) = 0, \qquad h \neq j, \quad h, j = 1, 2.$$

Then, based on assumptions (C.1) and (C.2) one may interchange the limits and integrals of the quantities involved and, hence,

$$\langle U_1^{(h,n)}, U_1^{(h,n)} \rangle(t) \xrightarrow{\mathbb{P}} \int_0^t \left(s_h^{(2)}(\boldsymbol{\delta}^0) - \frac{(s_h^{(1)}(\boldsymbol{\delta}^0))^2}{s_h^{(0)}(\boldsymbol{\delta}^0)} \right) \alpha_{hh}^0\ \mathrm{d}s,$$

$$\langle U_2^{(h,n)}, U_2^{(h,n)} \rangle(t) \xrightarrow{\mathbb{P}} \int_0^t \left(s_h^{(4)}(\boldsymbol{\delta}^0) - \frac{(s_h^{(3)}(\boldsymbol{\delta}^0))^2}{s_h^{(0)}(\boldsymbol{\delta}^0)} \right) \alpha_{hh}^0\ \mathrm{d}s,$$

and

$$\langle U_1^{(h,n)}, U_2^{(h,n)} \rangle(t) \xrightarrow{\mathbb{P}} \int_0^t \left(s_h^{(5)}(\boldsymbol{\delta}^0) - \frac{s_h^{(1)}(\boldsymbol{\delta}^0) s_h^{(3)}(\boldsymbol{\delta}^0)}{s_h^{(0)}(\boldsymbol{\delta}^0)} \right) \alpha_{hh}^0\ \mathrm{d}s,$$

taking care of (i).

With respect to (ii), we write for the predictable processes in the elements of the score vector,

$$H_{1i}(t,h) = X_i(t) - \frac{S_h^{(1)}(\delta^0,t)}{S_h^{(0)}(\delta^0,t)}$$

and

$$H_{2i}(t,h) = \frac{Y_{1i}(t)Y_{2i}(t)}{Y_{hi}(t)+\theta_h^0 Y_{1i}(t)Y_{2i}(t)} - \frac{S_h^{(3)}(\delta^0,t)}{S_h^{(0)}(\delta^0,t)}.$$

Then, let $H_{ki}^{(n)}(t,h) = n^{-1/2}H_{ki}(t,h)$, $k = 1,2$, and let the martingales containing the jumps of (3.7) and (3.8) be defined by

$$(3.12)\qquad {}_\epsilon U_1^{(h,n)}(t;\delta^0) = \sum_{i=1}^{n}\int_0^t H_{1i}^{(n)}(s,h)\mathbb{I}\{|H_{1i}^{(n)}(s,h)| > \epsilon\}\,\mathrm{d}M_{hi}(s)$$

and

$$(3.13)\qquad {}_\epsilon U_2^{(h,n)}(t;\delta^0) = \sum_{i=1}^{n}\int_0^t H_{2i}^{(n)}(s,h)\mathbb{I}\{|H_{2i}^{(n)}(s,h)| > \epsilon\}\,\mathrm{d}M_{hi}(s).$$

So, to prove that the predictable variation processes converge to continuous functions is equivalent to prove that $\langle {}_\epsilon U_k^{(h,n)}, {}_\epsilon U_k^{(h,n)}\rangle \xrightarrow{\mathbb{P}} 0$, as $n \to \infty$. For the scores related to the parameters β_h we have the inequality

(3.14)

$$\begin{aligned}
&\langle {}_\epsilon U_1^{(h,n)}, {}_\epsilon U_1^{(h,n)}\rangle \\
&\leq \sum_{i=1}^{n}\int 4\frac{1}{n}X_i^2\mathbb{I}\{n^{-1/2}|X_i| > \epsilon\}w_i(\delta_h^0)\alpha_{hh}^0\,\mathrm{d}s \\
&+\sum_{i=1}^{n}\int 4\frac{1}{n}\left(\frac{S_h^{(1)}(\delta^0)}{S_h^{(0)}(\delta^0)}\right)^2\mathbb{I}\left\{n^{-1/2}\left|\frac{S_h^{(1)}(\delta^0)}{S_h^{(0)}(\delta^0)}\right| > \epsilon\right\}w_i(\delta_h^0)\alpha_{hh}^0\,\mathrm{d}s.
\end{aligned}$$

As a consequence of assumptions (C.1) and (C.2), the second term in the r.h.s. of (3.14) converges to zero in probability, since $S_h^{(0)} \xrightarrow{\mathbb{P}} s_h^{(0)}$, $S_h^{(1)} \xrightarrow{\mathbb{P}}$

$s_h^{(1)}$, and $s_h^{(0)}$ is bounded away from zero, what implies that the indicator function in the integral will converge to zero as $n \to \infty$.

For the first term in the r.h.s. of (3.14), we have, for all $\gamma > 0$,

$$
\begin{aligned}
&\sum_{i=1}^{n} \int \frac{1}{n} X_i^2 \mathbb{I}\{n^{-1/2}|X_i| > \epsilon\} w_i(\delta_h^0) \alpha_{hh}^0 \, \mathrm{d}s \\
=&\sum_{i=1}^{n} \int \frac{1}{n} X_i^2 \mathbb{I}\{n^{-1/2}|X_i| > \epsilon;\ \beta_h^0 X_i \le -\gamma|X_i|\} e^{\beta_h^0 X_i}(Y_{hi} + \theta_h^0 Y_{1i} Y_{2i}) \alpha_{hh}^0 \, \mathrm{d}s \\
&+\sum_{i=1}^{n} \int \frac{1}{n} X_i^2 \mathbb{I}\{n^{-1/2}|X_i| > \epsilon;\ \beta_h^0 X_i > -\gamma|X_i|\} e^{\beta_h^0 X_i}(Y_{hi} + \theta_h^0 Y_{1i} Y_{2i}) \alpha_{hh}^0 \, \mathrm{d}s \\
=&I_1 + I_2.
\end{aligned}
$$

The term I_1 is bounded from above by

$$
\sum_{i=1}^{n} \int \frac{1}{n} X_i^2 \mathbb{I}\{n^{-1/2}|X_i| > \epsilon\} e^{-\gamma|X_i|}(Y_{hi} + \theta_h^0 Y_{1i} Y_{2i}) \alpha_{hh}^0 \, \mathrm{d}s,
$$

and since $Y_{hi} + \theta_h^0 Y_{1i} Y_{2i} \le Y_{hi} + |\theta_h^0| Y_{1i} Y_{2i} \le 1 + |\theta_h^0|$, such a quantity is bounded by

$$
\begin{aligned}
&(1 + |\theta_h^0|) \sum_{i=1}^{n} \int \frac{1}{n} X_i^2 \mathbb{I}\{n^{-1/2}|X_i| > \epsilon\} e^{-\gamma|X_i|} \alpha_{hh}^0 \, \mathrm{d}s \\
&\le (1 + |\theta_h^0|) \eta \int \alpha_{hh}^0 \, \mathrm{d}s
\end{aligned}
$$

where the last inequality is a consequence from the fact that, since $\gamma > 0$, $\lim_{x \to \infty} x^2 e^{-\gamma x} = 0$, and, hence, for all $\eta > 0$ there exists x sufficiently large such that $x^2 e^{-\gamma x} < \eta$. Therefore, taking η arbitrarily small, we may conclude that I_1 converges to zero in probability.

In virtue of assumption (C.3), the same conclusion is true for expression I_2. In order to make this clear, note that such expression is smaller or equal than

$$
\begin{aligned}
&4 \sum_{i=1}^{n} \int \frac{1}{n} X_i^2 \mathbb{I}\{\beta_h^0 X_i > -\gamma|X_i|\} e^{\beta_h^0 X_i} Y_{hi} \alpha_{hh}^0 \, \mathrm{d}s \\
&+ 4|\theta_h^0| \sum_{i=1}^{n} \int \frac{1}{n} X_i^2 \mathbb{I}\{\beta_h^0 X_i > -\gamma|X_i|\} e^{\beta_h^0 X_i} Y_{1i} Y_{2i} \alpha_{hh}^0 \, \mathrm{d}s \\
\le&4(1 + |\theta_h^0|) \sum_{i=1}^{n} \int \frac{1}{n} X_i^2 \mathbb{I}\{\beta_h^0 X_i > -\gamma|X_i|\} e^{\beta_h^0 X_i} Y_{hi} \alpha_{hh}^0 \, \mathrm{d}s
\end{aligned}
$$

and, hence, the last expression converges to zero in probability.

For the process ${}_\epsilon U_2^{(h,n)}$ we have that,

$$
\begin{aligned}
&\langle {}_\epsilon U_2^{(h,n)}, {}_\epsilon U_2^{(h,n)} \rangle(t) \\
=&\sum_{i=1}^{n} \int [H_{2i}^{(n)}(s,h)]^2 \mathbb{I}\{|H_{2i}^{(n)}(s,h)| > \epsilon\} \, \mathrm{d}\lambda_{hi}(s) \\
=&\sum_{i=1}^{n} \int \frac{1}{n} \left(\frac{Y_{1i}Y_{2i}}{Y_{hi} + \theta_h^0 Y_{1i}Y_{2i}} - \frac{S_h^{(3)}(\delta^0)}{S_h^{(0)}(\delta^0)} \right)^2 \mathbb{I}\{|H_{2i}^{(n)}(h)| > \epsilon\} w_i(\delta_h^0) \alpha_{hh}^0 \, \mathrm{d}s \\
\leq& 4 \sum_{i=1}^{n} \int \frac{1}{n} \left(\frac{Y_{1i}Y_{2i}}{Y_{hi} + \theta_h^0 Y_{1i}Y_{2i}} \right)^2 \mathbb{I}\left\{ \left| \frac{Y_{1i}Y_{2i}}{Y_{hi} + \theta_h^0 Y_{1i}Y_{2i}} \right| > \epsilon \right\} w_i(\delta_h^0) \alpha_{hh}^0 \, \mathrm{d}s \\
&+ 4 \int \left(\frac{S_h^{(3)}(\delta^0)}{S_h^{(0)}(\delta^0)} \right)^2 \mathbb{I}\left\{ \left| \frac{S_h^{(3)}(\delta^0)}{S_h^{(0)}(\delta^0)} \right| > \epsilon \right\} S_h^{(0)}(\delta^0) \alpha_{hh}^0 \, \mathrm{d}s \\
=& I_3 + I_4
\end{aligned}
$$

By the same reasons as pointed out earlier, we have, by assumptions (C.1) and (C.2) that expression I_4 converges to zero in probability. For expression I_3, we make use of the fact that since $\theta_h^0 > -1$,

$$
\frac{Y_{1i}Y_{2i}}{Y_{hi} + \theta_h^0 Y_{1i}Y_{2i}} \leq \frac{1}{1 + \theta_h^0}
$$

and, hence,

$$
\begin{aligned}
& n^{-1/2} \left| \frac{Y_{1i}Y_{2i}}{Y_{hi} + \theta_h^0 Y_{1i}Y_{2i}} \right| > \epsilon \Rightarrow n^{1/2} \left| \frac{1}{1 + \theta_h^0} \right| > \epsilon \\
& \Rightarrow \mathbb{I}\{\frac{1}{1 + \theta_h^0} > n^{1/2}\epsilon\} \geq \mathbb{I}\left\{ \left| \frac{Y_{1i}Y_{2i}}{Y_{hi} + \theta_h^0 Y_{1i}Y_{2i}} \right| > n^{1/2}\epsilon \right\}.
\end{aligned}
$$

Therefore,

$$\sum_{i=1}^{n}\int \frac{1}{n}\left(\frac{Y_{1i}Y_{2i}}{Y_{hi}+\theta_h^0 Y_{1i}Y_{2i}}\right)^2 \mathbb{I}\left\{\left|\frac{Y_{1i}Y_{2i}}{Y_{hi}+\theta_h^0 Y_{1i}Y_{2i}}\right| > \epsilon\right\} w_i(\boldsymbol{\delta}_h^0)\alpha_{hh}^0 \, \mathrm{d}s$$

$$=\sum_{i=1}^{n}\int \frac{1}{n}\frac{Y_{1i}Y_{2i}}{Y_{hi}+\theta_h^0 Y_{1i}Y_{2i}} \mathbb{I}\left\{\left|\frac{Y_{1i}Y_{2i}}{Y_{hi}+\theta_h^0 Y_{1i}Y_{2i}}\right| > \epsilon\right\} e^{\beta_h^0 X_i}\alpha_{hh}^0 \, \mathrm{d}s$$

$$\leq \frac{1}{1+\theta_h^0}\sum_{i=1}^{n}\int \frac{1}{n}\mathbb{I}\left\{\left|\frac{Y_{1i}Y_{2i}}{Y_{hi}+\theta_h^0 Y_{1i}Y_{2i}}\right| > \epsilon\right\} \alpha_{hh}^0 \, \mathrm{d}s.$$

$$\leq \frac{1}{1+\theta_h^0}\sum_{i=1}^{n}\int \frac{1}{n}e^{\beta_h^0 X_i}\mathbb{I}\left\{\frac{1}{1+\theta_h^0} > n^{1/2}\epsilon\right\} \alpha_{hh}^0 \, \mathrm{d}s.$$

Since θ_h^0 is a fixed value, it is always possible to get n sufficiently large such that the expression above is zero.

In conclusion, the conditions of the Rebolledo's central limit theorem are satisfied and, hence, $n^{-1/2}\boldsymbol{U}$ converges to a Gaussian process, with covariance matrix given by the limit of the predictable (co)variation processes. □

Theorem 1 will be considered when proving the asymptotic distribution for the MPLE below. Also, we will need to estimate the covariance matrix Σ. In this case, it will be of interest to work with the observed information matrix with elements given by (minus) the second derivatives of the log-likelihood, i.e., we define

$$\mathcal{I}(\boldsymbol{\delta}) = \begin{pmatrix} \mathcal{I}_1(\boldsymbol{\delta}) & \mathbf{0} \\ \mathbf{0} & \mathcal{I}_2(\boldsymbol{\delta}) \end{pmatrix} \tag{3.15}$$

where each symmetric matrix $\mathcal{I}_h$ as elements in the main diagonal given by

$$-\frac{\partial U_1^{(h)}}{\partial \beta_h}(\tau;\boldsymbol{\delta}) = -\int_0^{\tau}\sum_{i=1}^{n}\left\{\frac{\sum_j X_j^2 w_j(\boldsymbol{\delta}_h)}{\sum_j w_j(\boldsymbol{\delta}_h)} - \left(\frac{\sum_j X_j w_j(\boldsymbol{\delta}_h)}{\sum_j w_j(\boldsymbol{\delta}_h)}\right)^2\right\} \mathrm{d}N_{hi},$$

$$-\frac{\partial U_2^{(h)}}{\partial \theta_h}(\tau;\boldsymbol{\delta}) = -\int_0^{\tau}\sum_{i=1}^{n}\left\{\left(\frac{Y_{1i}Y_{2i}}{Y_{hi}+\theta_h Y_{1i}Y_{2i}}\right)^2 - \left(\frac{\sum_j e^{\beta_h X_j} Y_{1j}Y_{2j}}{\sum_j w_j(\boldsymbol{\delta}_h)}\right)^2\right\} \mathrm{d}N_{hi},$$

and the remaining elements are given by

$$-\frac{\partial U_1^{(h)}}{\partial \theta_h} = -\frac{\partial U_2^{(h)}}{\partial \beta_h}$$
$$= -\int_0^{\tau} \sum_{i=1}^{n} \left(\frac{\sum_j X_j e^{\beta_h X_j} Y_{1j} Y_{2j}}{\sum_j w_j(\boldsymbol{\delta}_h)} - \frac{\sum_j X_j w_j(\boldsymbol{\delta}_h) \sum_j e^{\beta_h X_j} Y_{1j} Y_{2j}}{(\sum_j w_j(\boldsymbol{\delta}_h))^2} \right) \mathrm{d}N_{hi}$$

The asymptotic distribution of the maximum partial likelihood is given by he following theorem. The approach used for the proof is based on the proof for the maximum likelihood estimator presented in Sen and Singer (1993), due to LeCam (1956)

THEOREM 2 *Let $\widehat{\boldsymbol{\delta}}$ be a value that maximizes the partial likelihood (3.3) and suppose that conditions* C.1–C.4 *hold. Then, if $\boldsymbol{\delta}^0$ is the true value for the parameter $\boldsymbol{\delta}$,*

$$n^{1/2}(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}^0) \xrightarrow{\mathbb{D}} \mathcal{N}(\mathbf{0}, \Sigma^{-1}).$$

Proof: For $\|\mathbf{u}\| \leq K$, $0 < K < \infty$ and remembering that $\boldsymbol{\delta}^0$ represents the true value for the vector of parameters, define

$$\lambda_n(\mathbf{u}) = \log L(\boldsymbol{\delta}^0 + n^{-1/2}\mathbf{u}) - \log L(\boldsymbol{\delta}^0)$$
$$= \sum_{i=1}^{n} \{\log L_i(\boldsymbol{\delta}^0 + n^{-1/2}\mathbf{u}) - \log L_i(\boldsymbol{\delta}^0)\} \tag{3.16}$$

Expanding $\log L_i(\boldsymbol{\delta}^0 + n^{-1/2}\mathbf{u})$ around $\boldsymbol{\delta}^0$, we get

$$\log L_i(\boldsymbol{\delta}^0 + n^{-1/2}\mathbf{u}) = \log L_i(\boldsymbol{\delta}^0) + n^{-1/2} \left[\frac{\partial \log L_i(\boldsymbol{\delta}^0)}{\partial \boldsymbol{\delta}}\right]' \mathbf{u}$$
$$+ \frac{1}{2n}\mathbf{u}' \frac{\partial^2 \log L_i(\boldsymbol{\delta}^*)}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}'} \mathbf{u} \tag{3.17}$$

for $\boldsymbol{\delta}^*$ in the line segment formed by $\boldsymbol{\delta}^0$ and $\boldsymbol{\delta}^0 + n^{-1/2}\mathbf{u}$. Hence,

(3.18)
$$\log L_i(\boldsymbol{\delta}^0 + n^{-1/2}\mathbf{u}) - \log L_i(\boldsymbol{\delta}^0)$$
$$= n^{-1/2} \left[\frac{\partial \log L_i(\boldsymbol{\delta}^0)}{\partial \boldsymbol{\delta}}\right]' \mathbf{u} + \frac{1}{2n}\mathbf{u}' \frac{\partial^2 \log L_i(\boldsymbol{\delta}^*)}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}'} \mathbf{u}.$$

Considering

$$Z_n(\mathbf{u}) = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\partial^2 \log L_i(\boldsymbol{\delta}^*)}{\partial\boldsymbol{\delta}\partial\boldsymbol{\delta}'} - \frac{\partial^2 \log L_i(\boldsymbol{\delta}^0)}{\partial\boldsymbol{\delta}\partial\boldsymbol{\delta}'}\right\},$$

we write

$$\lambda_n(\mathbf{u}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\frac{\partial \log L_i(\boldsymbol{\delta}^0)}{\partial\boldsymbol{\delta}}\right]'\mathbf{u} + \frac{1}{2n}\mathbf{u}'\sum_{i=1}^{n}\frac{\partial^2 \log L_i(\boldsymbol{\delta}^0)}{\partial\boldsymbol{\delta}\partial\boldsymbol{\delta}'}\mathbf{u} + \frac{1}{2}\mathbf{u}'Z_n\mathbf{u}$$

(3.13)

$$= \frac{1}{\sqrt{n}}[U(\boldsymbol{\delta}^0)]'\mathbf{u} - \frac{1}{2n}\mathbf{u}'\mathcal{I}(\boldsymbol{\delta}^0)\mathbf{u} + \frac{1}{2}\mathbf{u}'Z_n\mathbf{u}.$$

If $\|Z_n(\mathbf{u})\|$ converges in probability to zero, uniformly in $\mathbf{u}$, then, maximizing (3.13) with respect to $\mathbf{u}$ corresponds closely to obtaining a MPLE for $\boldsymbol{\delta}^0$, and using the asymptotic distribution for the score function, we will be able to find the asymptotic distribution of the MPLE. In order to proceed, let us consider the following representation for the elements of $\mathcal{I}$ (remembering that such matrix may be expressed as a partitioned matrix), for $h = 1, 2$,

$$(\mathcal{I}^{(h)})_{11} = \sum_{i=1}^{n}\int_0^{\tau}\left\{\left(\frac{S_h^{(1)}(\boldsymbol{\delta},x)}{S_h^{(0)}(\boldsymbol{\delta},x)}\right)^2 - \frac{S_h^{(2)}(\boldsymbol{\delta},x)}{S_h^{(0)}(\boldsymbol{\delta},x)}\right\}\,\mathrm{d}N_{hi}(x)$$

$$= \sum_{i=1}^{n}\int_0^{\tau} V_{11}^{(h)}(\boldsymbol{\delta},x)\,\mathrm{d}N_{hi}(x),$$

$$(\mathcal{I}^{(h)})_{12} = (\mathcal{I}^{(h)})_{21}$$

$$= \sum_{i=1}^{n}\int_0^{\tau}\left\{\frac{S_h^{(3)}(\boldsymbol{\delta},x)S_h^{(1)}(\boldsymbol{\delta},x)}{(S_h^{(0)}(\boldsymbol{\delta},x))^2} - \frac{S_h^{(5)}(\boldsymbol{\delta},x)}{S_h^{(0)}(\boldsymbol{\delta},x)}\right\}\,\mathrm{d}N_{hi}(x)$$

$$= \sum_{i=1}^{n}\int_0^{\tau} V_{12}^{(h)}(\boldsymbol{\delta},x)\,\mathrm{d}N_{hi}(x),$$

$$(\mathcal{I}^{(h)})_{22} = \sum_{i=1}^{n}\int_0^{\tau}\left\{\left(\frac{S_h^{(3)}(\boldsymbol{\delta},x)}{S_h^{(0)}(\boldsymbol{\delta},x)}\right)^2 - \left(\frac{Y_{1i}Y_{2i}}{Y_{hi}+\theta_h Y_{1i}Y_{2i}}\right)^2\right\}\,\mathrm{d}N_{hi}(x).$$

Also, define functions $v_{11}^{(h)}(\cdot)$ and $v_{12}^{(h)}(\cdot)$ similarly to $V_{ij}^{(h)}$, but with $S_h^{(i)}$ replaced by $s_h^{(i)}$, as defined earlier. Then, in order to examine the convergence (in probability) of $\|Z_n^{(h)}\|$, let us study each particular element.

$$
\begin{aligned}
|(Z_n^{(h)}(\mathrm{u}))_{11}| =& \left|\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} V_{11}^{(h)}(\delta^*, x) - V_{11}^{(h)}(\delta^0, x)\, \mathrm{d}N_{hi}(x)\right| \\
\leq& \left|\sup_{\{l:\, \|l\|<\|\mathrm{u}\|/\sqrt{n}\}} \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}(V_{11}^{(h)}(\delta^0 + l, x) - V_{11}^{(h)}(\delta^0, x))\, \mathrm{d}N_{hi}(x)\right| \\
\leq& \left|\sup_{\{l:\, \|l\|<\|\mathrm{u}\|/\sqrt{n}\}} \int_0^{\tau}(V_{11}^{(h)}(\delta^0 + l, x) - v_{11}^{(h)}(\delta^0 + l, x))\, \frac{\mathrm{d}N_{h\cdot}(x)}{n}\right| \\
&+ \left|\sup_{\{l:\, \|l\|<\|\mathrm{u}\|/\sqrt{n}\}} \int_0^{\tau}(V_{11}^{(h)}(\delta^0, x) - v_{11}^{(h)}(\delta^0, x))\, \frac{\mathrm{d}N_{h\cdot}(x)}{n}\right| \\
&+ \left|\sup_{\{l:\, \|l\|<\|\mathrm{u}\|/\sqrt{n}\}} \int_0^{\tau}(v_{11}^{(h)}(\delta^0 + l, x) - v_{11}^{(h)}(\delta^0, x))\, \frac{\mathrm{d}N_{h\cdot}(x)}{n}\right| \\
=& I_5 + I_6 + I_7.
\end{aligned}
$$

Now we note that, by the Lenglart's inequality, for all ρ, η

$$
\begin{aligned}
\mathbb{P}\left\{\frac{N_{h\cdot}(\tau)}{n} \geq \eta\right\} \leq& \frac{\rho}{\eta} + \mathbb{P}\left\{\frac{1}{n}\int_0^{\tau} \lambda_{h\cdot}(t)\, \mathrm{d}t \geq \rho\right\} \\
=& \frac{\rho}{\eta} + \mathbb{P}\left\{\frac{1}{n}\int_0^{\tau}\sum_{j=1}^{n}(Y_{hj} + \theta_h^0 Y_{1j}Y_{2j})e^{\beta_h^0 X_j}\alpha_{hh}^0(t)\, \mathrm{d}t \geq \rho\right\} \\
=& \frac{\rho}{\eta} + \mathbb{P}\left\{\int_0^{\tau} S_h^{(0)}(\delta^0, t)\alpha_{hh}^0(t)\, \mathrm{d}t \geq \rho\right\}.
\end{aligned}
$$

Taking $\rho > \int_0^{\tau} s_h^{(0)}(\delta^0)\alpha_{hh}^0\, \mathrm{d}t$ and considering assumption (C.2) then, as $n \uparrow \infty$ it follows that $\mathbb{P}\{\int_0^{\tau} S_h^{(0)}(\delta^0)\alpha_{hh}^0\, \mathrm{d}t > \rho\} \to 0$ and, hence,

$$
\lim_{\eta\to\infty}\lim_{n\to\infty}\mathbb{P}\left[\frac{N_{h\cdot}(\tau)}{n} > \eta\right] = 0.
$$

By assumption (C.2) expression I_5 converges to zero. If in addition we consider n sufficiently large such that $\delta^0 + l \in \mathcal{D}$, then it follows that expression I_6 also converges in probability to zero, as $n \to \infty$. The continuity on $s^{(j)}$ in condition (C.2) guarantees that, as $n \to \infty$, expression I_7 goes to zero in probability, and, hence,

$$
\sup_{\{\mathrm{u}:\, \|\mathrm{u}\|\leq K\}} |(Z_n^{(h)}(\bar{\mathrm{u}}))_{11}| \xrightarrow{\mathbb{P}} 0.
$$

Similar argumentation lead us to conclude that

$$|(\boldsymbol{Z}_n^{(h)}(\mathbf{u}))_{12}| = |(\boldsymbol{Z}_n^{(h)}(\mathbf{u}))_{21}| \xrightarrow{\mathbb{P}} 0, \qquad \textit{uniformly in } \mathbf{u}.$$

To deal with $(\boldsymbol{Z}_n^{(h)})_{22}$, define $M_h(\boldsymbol{\delta}, x) = S_h^{(3)}(\boldsymbol{\delta}, x)/S_h^{(0)}(\boldsymbol{\delta}, x)$. Then one may write

$$\begin{aligned}(\boldsymbol{Z}_n^{(h)}(\mathbf{u}))_{22} =& \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\{[M_h(\boldsymbol{\delta}^*, x)]^2 - (\frac{Y_{1i}Y_{2i}}{Y_{hi} + \theta_h^* Y_{1i}Y_{2i}})^2\}\, \mathrm{d}N_{hi} \\ &- \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\{[M_h(\boldsymbol{\delta}^0, x)]^2 - (\frac{Y_{1i}Y_{2i}}{Y_{hi} + \theta_h^0 Y_{1i}Y_{2i}})^2\}\, \mathrm{d}N_{hi},\end{aligned}$$

and, hence,

(3.20)

$$\begin{aligned}&|(\boldsymbol{Z}_n^{(h)}(\mathbf{u}))_{22}| \\ \leq& \left|\sup_{\{l:\, \|l\| \leq \|\mathbf{u}\|/\sqrt{n}\}} \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}(M_h(\boldsymbol{\delta}^0 + l, x))^2 - (M_h(\boldsymbol{\delta}^0, x))^2\, \mathrm{d}N_{hi}\right| \\ &+ \left|\sup_{\{l:\, \|l\| \leq \|\mathbf{u}\|/\sqrt{n}\}} \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}\{(\frac{Y_{1i}Y_{2i}}{Y_{hi} + (\theta_h^0 + l)Y_{1i}Y_{2i}})^2 - (\frac{Y_{1i}Y_{2i}}{Y_{hi} + \theta_h^0 Y_{1i}Y_{2i}})^2\}\, \mathrm{d}N_{hi}\right|.\end{aligned}$$

The first term in the r.h.s. of (3.20) is bounded from above by

$$\begin{aligned}&\left|\sup_{\{l:\, \|l\| \leq \|\mathbf{u}\|/\sqrt{n}\}} \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}(M_h(\boldsymbol{\delta}^0 + l, x))^2 - (m_h(\boldsymbol{\delta}^0 + l, x))^2\, \mathrm{d}N_{hi}\right| \\ &+ \left|\sup_{\{l:\, \|l\| \leq \|\mathbf{u}\|/\sqrt{n}\}} \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}(M_h(\boldsymbol{\delta}^0, x))^2 - (m_h(\boldsymbol{\delta}^0, x))^2\, \mathrm{d}N_{hi}\right| \\ &+ \left|\sup_{\{l:\, \|l\| \leq \|\mathbf{u}\|/\sqrt{n}\}} \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}(m_h(\boldsymbol{\delta}^0 + l, x))^2 - (m_h(\boldsymbol{\delta}^0, x))^2\, \mathrm{d}N_{hi}\right| \\ =& I_8 + I_9 + I_{10}.\end{aligned}$$

where $m_h = s_h^{(3)}/s_h^{(0)}$. By assumption (C.2) and the boundedness conditions imposed on the quantities involved, I_8 and I_9 converge to zero in probability. Also, if we consider that $m_h(\boldsymbol{\delta}, x)$ is a continuous function in a neighborhood of the true value $\boldsymbol{\delta}^0$, I_{10} converges to zero as $n \to \infty$.

Remains to show that the second term in the r.h.s. of (3.20) converges to zero. To do so, we note that such expression is not larger than

$$\sup_{\{l:\,|l|\le\|\mathbf{u}\|/\sqrt{n}\}}\left|\frac{1}{(1+\theta_h^0+l)^2}-\frac{1}{(1+\theta_h^0)^2}\right|\int_0^\tau\frac{\mathrm{d}N_{h\cdot}}{n}.$$

that converges to zero as $n\to\infty$. Hence it follows that expression (3.20) converge in probability to zero.

Hence, we have shown that

$$\sup_{\{\|\mathbf{u}\|\in[-K,K]\}}|\boldsymbol{Z}_n(\mathbf{u})|\xrightarrow{\mathbb{P}}0, \tag{3.21}$$

i.e., $\boldsymbol{Z}_n$ converges in probability to zero, uniformly in $\mathbf{u}$.

Also, it is a direct consequence of the results presented earlier that

$$n^{-1}\mathcal{I}(\boldsymbol{\delta}^0)\xrightarrow{\mathbb{P}}\Sigma(\tau), \tag{3.22}$$

for Σ defined in theorem 1.

Using (3.21) we may rewrite expression (3.16) as

$$\begin{aligned}\lambda_n(\mathbf{u})=&\frac{1}{\sqrt{n}}[\boldsymbol{U}(\boldsymbol{\delta}^0)]'\mathbf{u}-\frac{1}{2n}\mathbf{u}'\mathcal{I}(\boldsymbol{\delta}^0)\mathbf{u}+o_p(1)\\=&\frac{1}{\sqrt{n}}[\boldsymbol{U}(\boldsymbol{\delta}^0)]'\mathbf{u}-\frac{1}{2}\mathbf{u}'\Sigma\mathbf{u}+\frac{1}{2}\mathbf{u}'(\Sigma-\frac{1}{n}\mathcal{I}(\boldsymbol{\delta}^0))\mathbf{u}+o_p(1)\end{aligned}$$

that, in virtue of (3.22), may be written, uniformly in the set $\{\mathbf{u}\colon\|\mathbf{u}\|\le K\}$,

$$\lambda_n(\mathbf{u})=\frac{1}{\sqrt{n}}[\boldsymbol{U}(\boldsymbol{\delta}^0)]'\mathbf{u}-\frac{1}{2}\mathbf{u}'\Sigma\mathbf{u}+o_p(1) \tag{3.23}$$

Maximization of (3.23) with respect to $\mathbf{u}$ (ignoring the negligible part for a moment) will give

$$\hat{\mathbf{u}}=\frac{1}{\sqrt{n}}\Sigma^{-1}\boldsymbol{U}(\boldsymbol{\delta}^0),$$

or more precisely,

$$\hat{\mathbf{u}}=\frac{1}{\sqrt{n}}\Sigma^{-1}\boldsymbol{U}(\boldsymbol{\delta}^0)+o_p(1).$$

Noting that a maximum on λ_n corresponds to a maximum on the (log) partial likelihood, if $\widehat{\boldsymbol{\delta}}$ is a point of maximum in (3.3), it follows that

$$\widehat{\boldsymbol{\delta}} = \boldsymbol{\delta}^0 + n^{-1/2}\widehat{\mathbf{u}} = \boldsymbol{\delta}^0 + n^{-1}\Sigma^{-1}\boldsymbol{U}(\boldsymbol{\delta}^0) + o_p(1),$$

such that

$$\sqrt{n}(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}^0) = n^{-1/2}\Sigma^{-1}\boldsymbol{U}(\boldsymbol{\delta}^0) + o_p(1).$$

Therefore, by Theorem 1, $n^{-1/2}\boldsymbol{U}$ computed at $t = \tau$ converges to a multivariate normal distribution with covariance matrix given by Σ, and, hence, applying the Slutsky's theorem and assumption (C.4),

$$n^{1/2}(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}^0) \xrightarrow{\mathbb{D}} \mathcal{N}(\mathbf{0}, \Sigma^{-1}),$$

concluding the proof. □

Based on theorem 2 one needs an estimate for the covariance matrix such that one may compute confidence intervals and perform hypotheses tests for the parameters of interest. Then, to complete this section we state the following corollaries for the theorem.

COROLLARY 3 *The estimator $\widehat{\boldsymbol{\delta}}$ is a consistent estimator for $\boldsymbol{\delta}^0$.*

This corollary is a direct consequence of the theorem, i.e., one can immediately verify that $\|\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}^0\| = o_p(1)$

COROLLARY 4 *The covariance matrix Σ can be consistently estimated by $n^{-1}\mathcal{I}(\widehat{\boldsymbol{\delta}})$.*

Corollary 4 follows from the assumptions made in theorem 2 and from corollary 3. In fact, since $\widehat{\boldsymbol{\delta}}$ is a consistent estimator, there exists a value n_0 such that, for all $n \geq n_0$ it will be in the neighborhood $\mathcal{D}$, such that condition (B.4) will be true, and, hence, each element of the matrix $n^{-1}\mathcal{I}(\widehat{\boldsymbol{\delta}})$ will converge to the respective element of Σ.

4 Application to the two-sample problem

In this section we briefly illustrate the results discussed in the previous sections considering the two-sample problem.

Consider that in a clinical trial one is interested in studying the occurrence of two events, possibly censored. Also, each one of the patients is randomized in one of two groups, *placebo* and *treatment* and the interest resides in studying the efficacy of treatment with respect to prolonging the time for the occurrence of one (or both) events. Here, $K = 2$ and we define a time-independent covariate X_i assuming values zero or one, depending whether a particular individual i is assigned to the placebo or treatment groups. $N_{ki}(t)$ will represent the counting process as defined earlier, indicating if the kth event has occurred for individual i at time t.

In this particular setting some of the quantities can be rewritten in a more interpretable way. To do so, we define the sets of indexes $\mathrm{Tr} = \{i\colon X_i = 1\}$ and $\mathrm{Pl} = \{i\colon X_i = 0\}$ containing the indexes for the individuals in the treatment and placebo groups, respectively. The log-likelihood can be written as

$$\begin{aligned}
&\log L(\boldsymbol{\delta}) \\
&= \sum_{h=1}^{2} \beta_h \sum_{i \in \mathrm{Tr}} \int_0^{\tau} \mathrm{d}N_{hi} + \sum_{h=1}^{2} \sum_{i=1}^{n} \int_0^{\tau} \log\left(Y_{hi} + \theta_h Y_{1i} Y_{2i}\right) \mathrm{d}N_{hi} \\
&\quad - \sum_{h=1}^{2} \sum_{i=1}^{n} \int_0^{\tau} \log\left\{ \sum_{j \in \mathrm{Pl}} (Y_{hj} + \theta_h Y_{1j} Y_{2j}) + e^{\beta_h} \sum_{j \in \mathrm{Tr}} (Y_{hj} + \theta_h Y_{1j} Y_{2j}) \right\} \mathrm{d}N_{hi}.
\end{aligned}$$

Based on such log-likelihood one obtains iteratively the MPLE $\widehat{\boldsymbol{\delta}}$ for the parameter $\boldsymbol{\delta}$. Asymptotic properties for such estimator are given by theorems 1 and 2 and we now analyse the assumptions described there. Initially, note that for fixed t, each one of the $S_k^{(j)}(\cdot, t)$ may be thought of as an average of independent and identically distributed random variables. Therefore, one may apply the Khintchine law of large numbers and show the point-wise (for each t) convergence (in probability) to a deterministic function. If this function is monotone, then considering lemma 3.1 presented in Heiller and Willers (1988) we will have that the point-wise convergence is equivalent to the uniform convergence in assumption C.2.

Thus, for $S_h^{(0)}(\boldsymbol{\delta},t) = (1/n)\sum_{i=1}^n w_i(\boldsymbol{\delta}_h,t)$ we note that

$$
\begin{aligned}
\mathbb{E}\{w_i(\boldsymbol{\delta}_h,t)\} =& \mathbb{E}\left[(Y_{hi}+\theta_h Y_{1i}Y_{2i})e^{\beta_h X_i}\right] \\
=& \mathbb{E}\left[\mathbb{E}\left((Y_{hi}+\theta_h Y_{1i}Y_{2i})e^{\beta_h X_i}|X_i\right)\right] \\
=& \mathbb{E}\left[e^{\beta_h X_i}(\mathbb{S}_h(t)+\theta_h\mathbb{S}_{12}(t))\right] \\
=& (\pi_0+\pi_1 e^{\beta_h})(\mathbb{S}_h(t)+\theta_h\mathbb{S}_{12}(t)) \qquad (4.1)
\end{aligned}
$$

where $\pi_0 = 1-\pi_1 = \mathbb{P}(X_i = 0)$ is the probability that a particular individual will be assigned to placebo, $\mathbb{S}_h(t) = \mathbb{P}(T_{hi} \geq t)$ is the marginal survival function and $\mathbb{S}_{12}(t) = \mathbb{P}(T_{1i} \geq t;\ T_{2i} \geq t)$ is the joint survival function. Since $\mathbb{S}_h$ and $\mathbb{S}_{12}$ are non-increasing functions, the uniform convergence C.2 follows.

Similarly, for $j = 1, 2$,

$$
\begin{aligned}
\mathbb{E}[X_i^j(Y_{hi}+\theta_h Y_{1i}Y_{2i})e^{\beta_h X_i}] =& \mathbb{E}[X_i^j e^{\beta_h X_i}(\mathbb{S}_h(t)+\theta_h\mathbb{S}_{12}(t))] \\
=& \pi_1 e^{\beta_h}(\mathbb{S}_h(t)+\theta_h\mathbb{S}_{12}(t)) \qquad (4.2)
\end{aligned}
$$

that takes care of $S_h^{(1)}$ and $S_h^{(2)}$.

For $S_h^{(3)}$ we have

$$
\mathbb{E}[Y_{1i}Y_{2i}e^{\beta_h X_i}] = (\pi_0+\pi_1 e^{\beta_h})\mathbb{S}_{12}(t) \qquad (4.3)
$$

that is also a monotone function, and, hence, the uniform convergence is true.

For $S_h^{(4)}$ we also have the same result since

$$
\mathbb{E}\left[\frac{Y_{1i}Y_{2i}}{Y_{hi}+\theta_h Y_{1i}Y_{2i}}e^{\beta_h X_i}\right] = (\pi_0+\pi_1 e^{\beta_h})\frac{\mathbb{S}_{12}(t)}{1+\theta_h} \qquad (4.4)
$$

that, as a function of t is also a monotone function.

For $S_h^{(5)}$,

$$
\mathbb{E}[X_i Y_{1i}Y_{2i}e^{\beta_h X_i}] = \pi_1 e^{\beta_h}\mathbb{S}_{12}(t) \qquad (4.5)
$$

a monotone function. It follows then, for $j = 1, \ldots, 5$ the uniform condition C.2 is true, and, as $n \to \infty$,

$$
\begin{aligned}
\langle U_1^{(h,n)}, U_1^{(h,n)}\rangle(t) &\xrightarrow{\mathbb{P}} \frac{\pi_0\pi_1 e^{\beta_h}}{\pi_0+\pi_1 e^{\beta_h}}\int_0^t(\mathbb{S}_h(s)-\theta_h\mathbb{S}_{12}(s))\alpha_{hh}^0(s)\,\mathrm{d}s, \\
\langle U_2^{(h,n)}, U_2^{(h,n)}\rangle(t) &\xrightarrow{\mathbb{P}} \frac{\pi_0+\pi_1 e^{\beta_h}}{1+\theta_h}\int_0^t\frac{\mathbb{S}_h(s)-\mathbb{S}_{12}(s)}{\mathbb{S}_h(s)+\theta_h\mathbb{S}_{12}(s)}\alpha_{hh}^0(s)\,\mathrm{d}s, \\
\langle U_1^{(h,n)}, U_2^{(h,n)}\rangle(t) &\xrightarrow{\mathbb{P}} 0.
\end{aligned}
$$

Since in this case the continuity in condition C.2 and condition C.3 are trivial, then the asymptotic convergence of $n^{-1/2}(\widehat{\boldsymbol{\delta}}-\boldsymbol{\delta}_0)$ is given by a normal distribution with mean zero and covariance matrix Σ whose elements can be estimated by

$$(\widehat{\Sigma}_h)_{11} = \int_0^{\tau} \left(\frac{(Y_{h\cdot}^{\mathrm{Tr}} + \widehat{\theta}_h Y_{12\cdot}^{\mathrm{Tr}})e^{\widehat{\beta}_h}}{(Y_{h\cdot}^{\mathrm{Pl}} + \widehat{\theta}_h Y_{12\cdot}^{\mathrm{Pl}}) + (Y_{h\cdot}^{\mathrm{Tr}} + \widehat{\theta}_h Y_{12\cdot}^{\mathrm{Tr}})e^{\widehat{\beta}_h}} \right)^2 - \frac{(Y_{h\cdot}^{\mathrm{Tr}} + \widehat{\theta}_h Y_{12\cdot}^{\mathrm{Tr}})e^{\widehat{\beta}_h}}{(Y_{h\cdot}^{\mathrm{Pl}} + \widehat{\theta}_h Y_{12\cdot}^{\mathrm{Pl}}) + (Y_{h\cdot}^{\mathrm{Tr}} + \widehat{\theta}_h Y_{12\cdot}^{\mathrm{Tr}})e^{\widehat{\beta}_h}} \, \mathrm{d}\bar{N}_{h\cdot}$$

$$(\widehat{\Sigma}_h)_{22} = \int_0^{\tau} \left(\frac{Y_{12\cdot}^{\mathrm{Pl}} + e^{\widehat{\beta}_h} Y_{12\cdot}^{\mathrm{Tr}}}{(Y_{h\cdot}^{\mathrm{Pl}} + \widehat{\theta}_h Y_{12\cdot}^{\mathrm{Pl}}) + (Y_{h\cdot}^{\mathrm{Tr}} + \widehat{\theta}_h Y_{12\cdot}^{\mathrm{Tr}})e^{\widehat{\beta}_h}} \right)^2 \mathrm{d}\bar{N}_{h\cdot} - \frac{1}{(1+\widehat{\theta}_h)^2} \int_0^{\tau} \mathrm{d}\bar{N}_{h\cdot}^{11}$$

$$(\widehat{\Sigma}_h)_{12} = (\widehat{\Sigma}_h)_{21} = \int_0^{\tau} \frac{(Y_{12\cdot}^{\mathrm{Pl}} + Y_{12\cdot}^{\mathrm{Tr}} e^{\widehat{\beta}_h})(Y_{h\cdot}^{\mathrm{Tr}} + \widehat{\theta}_h Y_{12\cdot}^{\mathrm{Tr}})e^{\widehat{\beta}_h}}{[(Y_{h\cdot}^{\mathrm{Pl}} + \widehat{\theta}_h Y_{12\cdot}^{\mathrm{Pl}}) + (Y_{h\cdot}^{\mathrm{Tr}} + \widehat{\theta}_h Y_{12\cdot}^{\mathrm{Tr}})e^{\widehat{\beta}_h}]^2} - \frac{Y_{12\cdot}^{\mathrm{Tr}} e^{\widehat{\beta}_h}}{(Y_{h\cdot}^{\mathrm{Pl}} + \widehat{\theta}_h Y_{12\cdot}^{\mathrm{Pl}}) + (Y_{h\cdot}^{\mathrm{Tr}} + \widehat{\theta}_h Y_{12\cdot}^{\mathrm{Tr}})e^{\widehat{\beta}_h}} \, \mathrm{d}\bar{N}_{h\cdot}$$

where Y_h^{Tr} and Y_h^{Pl} represent the number of individuals with component h at risk, at time t, for treatment and placebo groups respectively, $Y_{12\cdot}^{\mathrm{Tr}}$ and $Y_{12\cdot}^{\mathrm{Pl}}$ represent the number of individuals with both components at risk, at time t, for treatment and placebo groups, and $\bar{N}_{h\cdot}^{11}$ represents the number of failures, at time t, divided by n, for those individuals with no failure in any component.

References

Aalen, O. O. (1975). *Statistical Inference for a family of Counting Processes*, PhD thesis, Department of Statisitcs, University of California, Berkeley.

Aalen, O. O. (1978). Nonparametric inference for a family of counting processes, *The Annals of Statistics* **6**(4): 701–726.

Andersen, P. K. and Gill, R. (1982). Cox's regression model for counting processes: A large sample study, *The Annals of Statistics* **10**(4): 1100–1120.

Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*, Springer-Verlag.

Brémaud, P. (1981). *Point Processes and Queues*, Springer-Verlag.

Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis*, New York: Wiley.

Heiller, S. and Willers, R. (1988). Asymptotic normality of R-estimates in the linear model, *Statistics* **19**: 173–184.

LeCam, L. (1956). On the asymptotic theory of estimation and testing hypotheses, *Proceedings of the third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, pp. 129–156. Berkeley: University of California Press.

Sarkar, S. K. (1987). A continuous bivariate exponential distribution, *Journal of the American Statistical Association* pp. 667–675.

Sen, P. K. and Singer, J. M. (1993). *Large Sample Methods in Statistics: An Introduction with Applications*, New York:Chapman-Hall.