A SEMI-PARAMETRIC MATRIX-VALUED COUNTING PROCESS MODEL FOR SURVIVAL ANALYSIS

by

Antonio Carlos Pedroso de Lima
Department of Biostatistics
University of North Carolina

Institute of Statistics
Mimeo Series No. 2145

May 1995
A Semi-Parametric Matrix-Valued Counting Process Model for Survival Analysis

by

Antonio Carlos Pedroso de Lima

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics.

Chapel Hill, 1995

Approved by:

[Signatures]
Advisor
Reader
Reader
Reader
ANTONIO CARLOS PEDROSO DE LIMA. A Semi-Parametric Matrix-Valued Counting Process Model for Survival Analysis. (Under the direction of Pranab Kumar Sen.)

ABSTRACT

In this work we consider the situation where \( n \) individuals are followed up with respect to the occurrence of several events of interest. For a particular individual we associate a vector-valued counting process where each element assumes value zero at \( t = 0 \) and jumps to one when a particular event is observed. The vectors are combined in a matrix giving origin to a matrix-valued counting process.

In the univariate problem, the counting process is usually assumed to satisfy the so-called Aalen multiplicative intensity model and in survival analysis this model can be heuristically justified based on the hazard function and on the information whether or not a particular individual is at risk (i.e., alive and uncensored.) We propose a modification in the multiplicative intensity process in order to take into account the multivariate nature of our setup. Additional assumptions allow us to obtain a general expression for any number of events observed per individual.

In order to include the effect of explanatory variables we consider a Cox-type model, that in our case will involve conditional hazard functions. Assuming that the ratio of such functions are time-independent, we are able to obtain asymptotic properties for the maximum partial likelihood estimators using the classical approach. Relaxation of such assumption will lead us to work on an infinite-dimensional parameter space. In addition, we assume that the coefficients associated with the covariates are also time-dependent. The corresponding estimators are considered
using the method of sieves and asymptotic properties such as consistency and weak convergence are proved taking into account the martingale representation for the score functions.

The results obtained are discussed in some situations of practical interest. Such a discussion is based on sup-norm statistics in the infinite-dimensional case. For the finite-dimensional parameter space the asymptotic distribution for the score function and for the maximum partial likelihood estimators are indicative that the usual methods considered in the univariate case can be applied.
Acknowledgments

This work represents a learning process. In this process, the most important factor was the direction I have received from my advisor, Professor Pranab Kumar Sen. He taught me how to deal with the intricacies of research activity and I hope to be able to make the best possible use of his teachings. I am deeply grateful to Professor Sen for his constant support and incentive. I would also like to thank the members of my committee, Dr. Chuanshu Ji, Dr. Alan Karr, Dr. Bahjat Qaqshi, and Dr. Chirayath Suchindran, whose comments helped me to have a better understanding of practical as well as theoretical issues arising from my work.

This project would not be possible without financial support from the University of São Paulo and Capes, both institutions from Brazil. In the final stages of this work I received financial support from the Biometric Consulting Laboratory, where I had the privilege to interact with Dr. Gary Koch.

From all points of view I would not be able to have overcome all the difficulties without the support, love and attention of my wife Beti. All these years here at Chapel Hill strengthened the respect and admiration I feel toward her.

Although several thousand miles apart, I have always felt the presence of my friends Coelho, Lisa and Koba. Solange Andreoni helped me to deal with several cultural adjustments and Shubhabrata Das reminded me that friendship is a universal concept.

Finally, I would like to dedicate this work to my parents Neusa and Manoel, whose constant encouragement and support I will never be able to return.
# TABLE OF CONTENTS

Chapter 1  Introduction and Literature Review  
1  Introduction .................................................. 1  
2  The counting process approach  
   2.1  Introduction ............................................... 4  
   2.2  The Nelson-Aalen estimator ............................... 9  
   2.3  The Kaplan-Meier estimator ............................. 13  
   2.4  Cox’s regression model ................................. 16  
   2.5  Further results ........................................... 22  
   2.6  Some examples .......................................... 25  
   2.7  Two important results .................................. 28  
3  Outline of the research ...................................... 29

Chapter 2  A Matrix-Valued Counting Process Model  
1  Introduction .................................................. 32  
2  The matrix-valued counting process model .................... 32  
3  Some examples ............................................... 42  
   3.1  Gumbel bivariate exponential distribution .............. 43  
   3.2  Sarkar bivariate exponential distribution .............. 46
### Chapter 3  The Matrix-Valued Model With Finite-Dimensional Parameters

1  Introduction ................................................. 50
2  Bivariate set-up ............................................. 51
   2.1  The model ............................................... 51
   2.2  The likelihood function ............................... 53
3  The score function, information matrix and the predictable variation processes ............................................. 55
   3.1  The score function and its martingale representation .... 56
   3.2  The information matrix and its conditional expectation .... 58
   3.3  The predictable variation and covariation processes ....... 61
4  Asymptotic properties ...................................... 62

### Chapter 4  The Matrix-Valued Model with Time Dependent Parameters

1  Introduction ................................................. 79
2  Model specification ......................................... 80
3  The sieve estimator ......................................... 81
4  Asymptotic properties for the sieve estimator ................ 88
5  Proofs ................................................................ 109
   5.1  Proof for Lemma 4.1 ................................. 109
   5.2  Proof for Lemma 4.2 ................................. 121
   5.3  Proof for Lemma 4.3 ................................. 122
Chapter 5  Possible Applications and Plans for Further Research  125

1  Introduction .................................................. 125

2  Testing the proportionality assumption  .................................. 126
   2.1  Linear covariance structure ........................................ 126
   2.2  The two-sample case ............................................. 130
   2.3  A general covariance matrix .................................... 134

3  Tests for the finite-dimensional model  .............................. 135

4  Topics for further research ........................................... 139

Bibliography  ................................. 141
CHAPTER 1

Introduction and Literature Review

1 Introduction

In several practical situations where statistical methods need to be applied, the information that will be collected may be represented by nonnegative random variables. For example, often in engineering applications we are interested in evaluating the performance of a new equipment and to do so, several units are kept working until failure and we record the time each unit takes before it fails; in biomedical applications, it is common the situation where researchers have as a main objective to test if a new treatment is more efficient than a standard one, and patients (random selected to receive one of the treatments) are followed until an event of interest (e.g., recurrence of a disease) takes place. In this case, each patient has the time (from randomization) up to the event recorded. Several parametric and non-parametric methods have been developed to deal with such situations and usually they are grouped under the survival analysis or life history analysis denomination.¹

Under the parametric approach, we usually obtain more efficient analyses if the

¹Although some authors distinguish between these two names, we shall consider here survival analysis as a general denomination for situations dealing with nonnegative random variables in the sense we will discuss later.
underlying model is correctly selected. For this reason, usually quality control applications make use of parametric models for statistical analyses. On the other hand, in biological or biomedical settings, the verification of distributional assumptions is more difficult to be accomplished and, hence, the nonparametric approach is an important alternative given the robustness properties of such procedures.

In general survival data are incomplete given that, for example, some of the patients may be lost to follow up or the event of interest may not be observable prior to the termination of the trial. In such situations, we have to consider the presence of censored observations. Under the nonparametric approach several methods previously used in the uncensored situations have been generalized to the censored case. For example, Kaplan and Meier (1958) considered the estimation of the survival distribution function (generalizing the empirical distribution function) by studying an idea developed by Böhmer in 1912 and generalizations of the Wilcoxon, Kruskal-Wallis and Savage tests to censored data were developed. One of the most important generalization is the so-called semi-parametric regression model, due to Cox (1972), that allows for covariates to be included in the analysis.

We consider here a different approach that has been recently developed to study survival analysis. It is based in the fact that in a typical problem, we are observing not only the random times but random events occurring in time. Therefore, it becomes natural to think of stochastic processes in order to study such situations. In most cases, we are interested in a collection of individuals moving in a finite number of states and, based on the data collected, usually it is possible to count the number of patients or units in a state at a given time. Therefore, we may make use of counting processes and the associated probabilistic theory developed for such processes. Based on this idea, Aalen (1978) is often cited as the starting point for the unification of the treatment given in survival analysis methodologies based on counting processes. He studied basic nonparametric statistical problems for censored
data using the *intensity process* associated with the counting process recording the events in function of the time. More specifically, he showed that in most situations the intensity process can be written as a product of two functions (the *multiplicative intensity process*). The counting process can then be decomposed in two parts, one being a *martingale* and the other part involving the intensity process. Based on that, several situations can be treated in a unified way. The basic idea is to write estimators and test statistics as stochastic integrals involving martingales and using martingale-related results we may prove small- and large-sample properties.

Although the formal treatment is quite elaborated, it yields considerable generality for the results and this is one of the most important characteristics of the use of stochastic processes. For example, proofs for the distributional properties of estimators and test statistics under very general censoring patterns can be seen in Aalen (1978), Aalen and Johansen (1978) and Gill (1980). In the mathematical background, often the proofs of consistency make use of the *Lenglart’s inequality*, discussed in Andersen and Gill (1982). Another frequently used result is the *martingale central limit theorem*, that was proved by several authors. In this context, the version given in Andersen and Gill (1982) (also known as the *Rebolledo’s martingale central limit theorem*) is usually considered.

Some papers giving heuristic formulation of the various uses of counting processes can be found. For example, the martingale approach on the Cox’s regression model has been considered in a very intuitive way in Gill (1984), although quite formally in Andersen and Gill (1982); also, a heuristic introduction to counting process models for life history data is given in Andersen and Borgan (1985). Recently, two books have been published on the subject. Fleming and Harrington (1991) consider the application of counting process in the most known models used in survival analysis whereas Andersen, Borgan, Gill and Keiding (1993) describes the development of the theory on more general models.
It is our intention here to review the counting process approach that has been developed to study survival analysis. In section 2 we present a basic literature review about the subject, presenting in an informal way the basic mathematical concepts and their applications in some known models. Also, in order to illustrate the applicability of the results, two practical examples are presented. For the sake of completeness the Lenglart's inequality and the Rebolledo's theorem are stated since both results are often cited in this chapter and used several times in the development of this work; other results could also have been included but we decided to include only those two so that this chapter will not be too technical. Finally in Section 3 we briefly describe the contents of the remaining of this dissertation outlining the distribution of the material and summarizing the findings of this work.

2 The counting process approach

In the next subsection we introduce some quantities that are going to be used throughout this work. We present the initial results as intuitive as possible, but at the same time we try to keep in perspective the formal underlying structure. Measurability questions are not addressed here because, as discussed in Andersen et al. (1993), the associated technical points do not arise in the applications considered. Since we are studying the occurrence of random events in time, all results are presented based on a continuous-time interval $\mathcal{T} = [0, \tau)$, for a given terminal time $\tau$, $0 < \tau \leq \infty$, unless otherwise mentioned.

2.1 Introduction

Survival analysis deals with (discrete) random events occurring on time. Therefore, it is natural to consider stochastic processes to model such situation. Under this
perspective, Aalen (1975) is often cited as one of the first to study the use of modern theory of stochastic process in survival analysis. The revised main results of his Ph.D. thesis, published in Aalen (1978), deal with the study of nonparametric inference based on counting processes, specifically, the empirical cumulative intensity estimation [Nelson (1969), Altshuler (1970)] and nonparametric two-sample tests. Based on the counting process approach given by Brémaud (1981) and making use of the theory of time-continuous martingales and stochastic integrals, such studies started an unified way to deal with several methods usually considered in survival analysis.

As a starting point, suppose that \( n \) items (or individuals) are observed up to the occurrence of a particular event of interest \( \mathcal{E} \) (e.g., death caused by a disease under study, failure of electronic equipments, etc.) Let \( T_1, \ldots, T_n \) be the times of occurrence of \( \mathcal{E} \) for each item and assume that sometimes \( T_j \) may not be observable, in which case we would instead observe \( C_j \), another quantity possibly random. Then we may define, for \( i = 1, \ldots, n \),

\[
Z_i = T_i \wedge C_i,
\]

\[
\delta_i = \mathbb{I}\{T_i \leq C_i\},
\]

where \( \mathbb{I} \) is the characteristic or indicator function for the set \( A \). The pairs \((T_i, \delta_i)\) are assumed to be independent. Also, \( \{T_i\}_{1 \leq i \leq n} \) is a sequence of nonnegative random variables with distribution function \( F_i(\cdot) \) and hazard function defined by

\[
\alpha_i(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \mathbb{P}\{T_i \in [t, t + \Delta t) \mid T_i \geq t\}, \tag{1.1}
\]

that can be heuristically written as \( \alpha_i(t) \, dt = \mathbb{P}\{T_i \in [t, t + dt) \mid T_i \geq t\} \). The cumulative (integrated) hazard function is given by

\[
A_i(t) = \int_0^t \alpha_i(x) \, dx.
\]
In order to use the counting process approach, consider a right-continuous multivariate counting process \( \mathbf{N}(t) = \{N_1(t), \ldots, N_n(t)\} \), \( t \geq 0 \), such that each element is defined as

\[
N_i(t) = \mathbb{I}\{Z_i \leq t; \delta_i = 1\}, \quad i = 1, \ldots, n.
\]  

(1.2)

In other words, (1.2) is zero before \( Z_i = T_i \) and jumps to 1 if the event \( \mathcal{E} \) occurs at \( Z_i \). To identify the corresponding intensity process, consider the left-continuous process defined by

\[
Y_i(t) = \mathbb{I}\{Z_i \geq t\}.
\]

(1.3)

If we observe the interval \( I_{dt} = [t, t + \ dt) \), given the information available up to time \( t^- \) (i.e., just prior to \( t \)), the probability that \( N_i(t) \) will jump on \( I_{dt} \) (i.e., \( dN_i(t) = N_i((t + \ dt)^-) - N_i(t^-) = 1 \)) will be zero if \( \mathcal{E} \) has already occurred or if a censoring took place prior to \( t \) [in which case (1.3) equals zero] and will be \( \alpha(t) \ dt \) if the item is still being observed, i.e., neither \( \mathcal{E} \) nor censoring were observed prior to \( t \) [in which case (1.3) equals 1]. Therefore, we may write

\[
\mathbb{P}\{ dN_i(t) = 1 \mid \mathcal{F}_{t^-}^{(i)} \} = \alpha_i(t)Y_i(t) \ dt,
\]

where \( \mathcal{F}_{t^-}^{(i)} \) represents all information available about the item \( i \) prior to time \( t \); formally, it is a pre-\( t \) \( \sigma \)-algebra in the sense that is the smallest \( \sigma \)-algebra containing all \( \mathcal{F}_s^{(i)} \), \( s < t \), with \( \{\mathcal{F}_t^{(i)}\} \) being an increasing right-continuous family of sub-\( \sigma \)-algebras (a filtration) of \( \mathcal{F}^{(i)} \), the \( \sigma \)-algebra pertaining to the probability space where all quantities above are defined.

The jump \( dN_i(t) \) is a 0-1 random variable and, hence, we may write

\[
\mathbb{E}\{ dN_i(t) \mid \mathcal{F}_{t^-}^{(i)} \} = \alpha_i(t)Y_i(t) \ dt.
\]

(1.4)

The quantity (1.4) is random since the expectation is computed conditioning on the random variables generating \( \mathcal{F}_{t^-}^{(i)} \). Also, the process \( Y_i(t) \) is a predictable process, i.e.,
its value is known just prior to \( t \). Formally, this process is defined on the \( \sigma \)-algebra generated by predictable rectangles (for details, see Fleming and Harrington (1991), page 32.) Based on the predictability of \( Y_i(t) \), we may define a martingale from (1.4). To do so, consider the integrated intensity process
\[
\Lambda_i(t) = \int_0^t \alpha_i(s) Y_i(s) \, ds \\
= \int_0^t \lambda_i(s) \, ds.
\] (1.5)

Informally, we may define \( M_i(t) \) as
\[
M_i(t) = N_i(t) - \Lambda_i(t)
\] (1.6)
\[
\Rightarrow dM_i(t) = dN_i(t) - d\Lambda_i(t),
\]
where \( \Lambda_i(\cdot) \) is called the compensator of the counting process \( N_i(\cdot) \). Using (1.6), it follows that
\[
\mathbb{E}\{ dM_i(t) \mid \mathcal{F}_{t-}^{(i)} \} = \mathbb{E}\{ dN_i(t) \mid \mathcal{F}_{t-}^{(i)} \} - \mathbb{E}\{ d\Lambda_i(t) \mid \mathcal{F}_{t-}^{(i)} \}
= \alpha_i(t) Y_i(t) \, dt - \alpha_i(t) Y_i(t) \, dt
= 0,
\] (1.7)
where the predictability of \( Y_i(t) \) (and, hence, of \( \Lambda_i(t) \)) was considered when writing \( \mathbb{E}\{ \alpha_i(t) Y_i(t) \mid \mathcal{F}_{t-}^{(i)} \} = \alpha_i(t) Y_i(t) \, dt \), because given the information in \( \mathcal{F}_{t-}^{(i)} \), \( Y_i(t) \) is known.

Since \( M_i(0) = 0 \), the result in (1.7) implies that \( M_i(t) \) is a zero-mean martingale, what is of fundamental importance in the development of the counting process approach. A formal presentation of the derivation above is given in Fleming and Harrington (1991) and Andersen et al. (1993). The expression (1.6) may also be interpreted as a decomposition of the counting process \( N_i(\cdot) \) into a systematic part [the compensator \( \Lambda_i(\cdot) \)] and a pure noise process [the martingale \( M_i(\cdot) \)]. This decomposition is often used to represent estimators and test statistics as martingales.
Then, using certain inequalities and theorems (e.g., Lenglart's Inequality and Rebolloedo's central limit theorem for martingales), consistency as well as large sample distribution may be derived.

Under this setup, suppose that \( \alpha_i(t) = \alpha(t), \ i = 1, \ldots, n \), and define \( N(t) = \sum_{i=1}^{n} N_i(t) \). Also, since \( N_i, \ i = 1, \ldots, n \) are independent [recall that \((T_i, C_i), \ i = 1, \ldots, n\), are assumed independent], we may consider the product filtration \( \mathcal{F}_i \) based on \( \mathcal{F}_{i}^{(i)}, \ i = 1, \ldots, n \), and hence

\[
\mathbb{E}\{ \mathrm{d} N(t) \mid \mathcal{F}_{t^-}\} = \sum_{i=1}^{n} \mathbb{E}\{ \mathrm{d} N_i(t) \mid \mathcal{F}_{t^-}^{(i)}\} = \alpha(t)Y(t) \, \mathrm{d} t
\]  

(1.8)

where \( Y(t) \) may be interpreted as the risk set (at time \( t \)). Hence, all results described above are also applicable to the counting process \( N(t) \).

Situations where the intensity process (1.4) arises can be found in a more general setup. For example, we may consider a time-continuous Markov chain, with a finite state space \( S = \{1, 2, \ldots, s\} \), where the infinitesimal transition probabilities from state \( i \) to state \( j \) may be represented by \( \alpha_{ij}(t) \). Then, as done before, the counting process \( N_{ij}(t) \) may be defined, representing the number of transitions from state \( i \) to state \( j (\neq i) \) at time \( t \). In this case we may also define \( Y_i(t) \) as the number of individuals in state \( i \) at time \( t \); here, also, \( Y_i(t) \) has left-continuous paths. Hence, \( N(t) = \{N_{ij}(t)\} \) is a multivariate counting process. The situation with one transient state is studied by Aalen (1978), where it is mentioned that the intensity process for \( N_{ij}(t) \) is given by \( \alpha_{ij}(t)Y_i(t) \). A general Markov chain (i.e., considering more than one transient state) is considered in Fleming (1978) and in Aalen and Johansen (1978) but the former studies only uncensored observations. The connection between the Markov chain and the counting process with similar approach to that considered here is also addressed in Jacobsen (1982).

In the following we will review how counting process and martingales have been used in deriving properties for estimators in some situations. We begin with an
informal description based on Aalen (1978), where the estimation of the cumulative hazard function is considered. Similar results can be applied in getting properties for the Kaplan-Meier estimator, what is described in subsection (2.3). Following that, we review some of the results obtained for the Cox's regression model [subsection (2.4)]. In the subsection (2.5) we present a very brief review of results obtained in some other settings. Finally two examples are included in subsection (2.6) in order to illustrate some practical applications.

2.2 The Nelson-Aalen estimator

Using the results and definitions presented above, we may now describe the estimator derived heuristically by Aalen (1978) for the cumulative hazard function

\[ A_i(t) = \int_0^t \alpha_i(x) \, dx, \]  

(1.9)

considering our simple situation where only one event \( E \) is of interest. The estimator is defined assuming that the intensity process for \( N_i(\cdot) \), denoted by \( \lambda_i(t) \), is of the special form given in (1.4), i.e.,

\[ \lambda_i(t) = \alpha_i(t)Y_i(t), \]  

(1.10)

called a multiplicative intensity model. Here, as before, \( \alpha_i(\cdot) \) is a (nonnegative) deterministic function and \( Y_i(\cdot) \) is a predictable process.

The proposed estimator, called initially empirical \( \beta \)-function by Aalen (1978) became known as the Nelson-Aalen estimator because it is a generalization of the empirical cumulative intensity estimator proposed by Nelson (1969). It was proposed mainly to be used for plotting purposes and we consider it here in order to review how counting processes were first used to study small and large sample properties. At this point, we borrow the sequence given in Andersen et al. (1993) and Andersen and Borgan (1985) to introduce the estimator and its properties.
Initially we may note that a natural estimator for (1.9) would be given by
\[
\int_0^t \frac{dN_i(x)}{Y_i(x)}.
\]
However, when \(Y_i(x) = 0\), such an estimator is undefined. Thus, Aalen (1978) proposed the equivalent estimator
\[
\hat{A}_i(t) = \int_0^t J_i(x) \frac{dN_i(x)}{Y_i(x)},
\]
where \(J_i(x) = \mathbb{I}\{Y_i(x) > 0\}\) and \(J_i(x)/Y_i(x) = 0\) whenever \(Y_i(x) = 0\). It is noted that \(\hat{A}_i(\cdot)\) is a simple sum of nonnegative terms and, hence, is an increasing right-continuous step-function.

Now if we consider the quantity
\[
A^*_i(t) = \int_0^t \alpha_i(x)J_i(x) \, dx,
\]
we may notice first that \(A^*_i(t)\) is almost the same as \(A(t)\) (the function we want to estimate) when there is a small probability that \(J_i(x)\) [and, hence, \(Y_i(x)\)] equals zero. Second, taking the difference between (1.11) and (1.12), we get
\[
\hat{A}_i(t) - A^*_i(t) = \int_0^t J_i(x) \frac{dN_i(x)}{Y_i(x)} - \int_0^t J_i(x)\alpha_i(x) \, dx
\]
\[
= \int_0^t J_i(x) \left[ \frac{dN_i(x)}{Y_i(x)} - Y_i(x)\alpha_i(x) \right] \, dx
\]
\[
= \int_0^t J_i(x) \frac{dM_i(x)}{Y_i(x)},
\]
where \(M_i(t) = N_i(t) - \int_0^t Y_i(x)\alpha_i(x) \, dx\) is the martingale (1.6). The integral given by (1.13) is a stochastic integral. It is by itself a stochastic process and in the situations considered here, its paths coincide with the corresponding Lebesgue-Stieltjes integral (conditions for this are discussed in Aalen (1978).) It follows that (Andersen et al. (1993), theorem II.3.1) (1.13) are also zero-mean martingales and hence, \(A^*_i(t)\) can be interpreted as the compensator for \(\hat{A}_i(t)\).
Since (1.13) is a zero-mean martingale and assuming that the expectations exist, we have

\[
E\{\hat{A}_i(t)\} = E\{A_i^*(t)\} = \int_0^t \alpha_i(x)E\{J_i(x)\} \, dx
\]

\[
= \int_0^t \alpha_i(x)P\{Y_i(x) > 0\} \, dx.
\]

(1.14)

Therefore, \(\hat{A}_i(t)\) is an approximately unbiased estimator for \(A_i(t)\), in the sense that when \(P\{Y_i(x) = 0\}\) is small, the bias will be negligible.

In order to get information about the variability of (1.13), we consider its predictable variation process, denoted by \(\langle \hat{A}_i - A_i^* \rangle(t)\). So, we have

\[
d\langle \hat{A}_i - A_i^* \rangle(t) = \text{Var}\{d(\hat{A}_i(t) - A_i^*(t)) \mid \mathcal{F}_{i-}\}
\]

\[
= \text{Var}\{\frac{J_i(t)}{Y_i(t)} \, dM_i(t) \mid \mathcal{F}_{i-}\} = \left\{\frac{J_i(t)}{Y_i(t)}\right\}^2 \text{Var}\{dM_i(t) \mid \mathcal{F}_{i-}\}
\]

\[
= \left\{\frac{J_i(t)}{Y_i(t)}\right\}^2 \text{Var}\{dN_i(t) \mid \mathcal{F}_{i-}\} = \left\{\frac{J_i(t)}{Y_i(t)}\right\}^2 \alpha_i(t)Y_i(t) \, dt.
\]

Therefore, the predictable variation process is

\[
\langle \hat{A}_i - A_i^* \rangle(t) = \int_0^t \frac{J_i(x)}{Y_i(x)} \alpha_i(x) \, dx.
\]

The predictable covariation process can be defined in a similar fashion, but the variance of the increments \(d(\hat{A}_i(t) - A_i^*(t))\) (given the past), is replaced by the covariance between increments. As discussed in Andersen et al. (1993), when the intensity process exists and it is continuous, we have that

\[
\langle M_i, M_j \rangle(t) = 0, \quad i \neq j,
\]

and the martingales are said orthogonal. Therefore, since

\[
\langle \hat{A}_i - A_i^*; \hat{A}_j - A_j^* \rangle(t) = \int_0^t \frac{J_i(x)}{Y_i(x)} \frac{J_j(x)}{Y_j(x)} \, d\langle M_i; M_j \rangle(x), \quad i \neq j,
\]

11
it follows that

$$\langle \hat{A}_i - A_i^*; \hat{A}_j - A_j^* \rangle (t) = 0, \quad i \neq j,$$

and hence, $\hat{A}_i - A_i^*$ are also orthogonal. As discussed in Andersen and Borgan (1985), this has some practical implications, for example, plots for the Nelson-Aalen estimator for each $i$ can be analyzed separately.

When $Y_i(x)$ is positive with high probability [i.e., when $A_i^*(t)$ equals $A_i(t)$], in order to study the variability of $\hat{A}(t)$, Aalen (1978) defines the mean square error function for $\hat{A}_i(t)$ as

$$\tilde{\sigma}_i(t) = \mathbb{E}[\hat{A}_i(t) - A_i^*(t)]^2.$$

Since $\hat{A}_i(t) - A_i^*(t)$ is a martingale, $\{\hat{A}_i(t) - A_i^*(t)\}^2$ is a sub-martingale and hence, applying the Doob-Meier decomposition [Fleming and Harrington (1991), p.37], we have that its compensator is given by $\langle \hat{A}_i - A_i^* \rangle (t)$. Therefore, we have

$$\mathbb{E}\{(\hat{A}_i(t) - A_i^*(t))^2 - \langle \hat{A}_i - A_i^* \rangle (t)\} = 0$$

$$\Rightarrow \mathbb{E}\{\hat{A}_i(t) - A_i^*(t)\}^2 = \mathbb{E}\{\langle \hat{A}_i - A_i^* \rangle (t)\} = \int_0^t \mathbb{E}\left[\frac{J_i(x)}{Y_i(x)}\right] \alpha_i(x) \, dx.$$

(1.15)

Noting that

$$\int_0^t \frac{J_i(x)}{Y_i(x)} \alpha_i(x) \, dx = \int_0^t \frac{J_i(x)}{[Y_i(x)]^2} Y_i(x) \alpha_i(x) \, dx$$

$$= \int_0^t \frac{J_i(x)}{[Y_i(x)]^2} \, dN_i(x) - \int_0^t \frac{J_i(x)}{[Y_i(x)]^2} \, dM_i(x),$$

it follows that

$$\mathbb{E}\{\hat{A}_i(t) - A_i^*(t)\}^2 = \mathbb{E}\left\{\int_0^t \frac{J_i(x)}{Y_i(x)} \alpha_i(x) \, dx\right\}$$

$$= \mathbb{E}\left\{\int_0^t \frac{J_i(x)}{[Y_i(x)]^2} \, dN_i(x)\right\} - \mathbb{E}\left\{\int_0^t \frac{J_i(x)}{[Y_i(x)]^2} \, dM_i(x)\right\}$$

$$= \mathbb{E}\left\{\int_0^t \frac{J_i(x)}{[Y_i(x)]^2} \, dN_i(x)\right\}.$$
Therefore, Aalen (1978) proposed to estimate (1.15) using
\[ \hat{\sigma}_i(t) = \int_0^t \frac{J_i(x)}{[Y_i(x)]} \, dN_i(x). \] (1.16)

The process above is also called the optional variation process [Andersen et al. (1993), Section II.3.2] and is denoted by \([\hat{A}_i - \hat{A}_i^0](t)\).

### 2.3 The Kaplan-Meier estimator

In this subsection we review the use of counting process in deriving properties for the Kaplan-Meier estimator. This review is based mainly in Gill (1980) where the estimator is studied for arbitrary survival distributions. However, we restrict ourselves here to absolutely continuous distributions and we do not detail formal aspects such as the product-integral approach to the estimator, but emphasize the application of similar results as those developed in the previous subsection. However we should mention that a formal treatment is given in Andersen et al. (1993).

We consider the same setup as that in the previous subsection but in addition here we assume that \(F_1 = F_2 = \cdots = F_n = F\). Also, define \(N(t) = \sum_{i=1}^n N_i(t)\) as the number of failures or deaths up to time \(t\); \(Y(t) = \sum_{i=1}^n Y_i(t)\) as the number of individuals at risk at time \(t\) (still a predictable process); and \(M(t) = \sum_{i=1}^n M_i(t)\) as the martingale based on \(N(t)\) and \(\int_0^t Y(x) \alpha(x) \, dx\). Also, let \(J(t) = \mathbb{I}\{Y(t) > 0\}\).

The survival function is defined by \(S(t) = 1 - F(t)\) and by standard results it is known that
\[ S(t) = \exp\left\{-\int_0^t \alpha(x) \, dx \right\} = \exp\{-A(t)\}. \]

Therefore, we may estimate \(S(t)\) by replacing \(A(t)\) with the Nelson-Aalen estimator to get \(\hat{S}(t) = \exp\{-\hat{A}(t)\} = \exp\{-\int_0^t d\hat{A}(x)\}\). Since \(\hat{A}(t)\) is a step-function, we may write
\[ \hat{S}(t) = \prod_{x \leq t} \left\{1 - \frac{\Delta N(x)}{Y(x)} \right\}, \] (1.17)
the so-called *Kaplan-Meier estimator* or *product-limit estimator*. Analogously to the previous section, we may also define

$$S^*(t) = \exp\{-A^*(t)\},$$

that is equal to $S(t)$ except when $Y(t) = 0$.

Based on $\hat{S}(t)$ and $S^*(t)$, we may apply the *Duhamel equation* [Andersen et al. (1993), theorem II.6.2] to get

$$\frac{\hat{S}(t) - S^*(t)}{S^*(t)} = -\int_0^t \frac{\hat{S}(x^-) J(x)}{S^*(x) Y(x)} dM(x). \tag{1.18}$$

Here, as before, $M$ is a (local) square integrable martingale. Also, $\hat{S}(x^-)$ is a left-continuous (adapted) function and hence predictable; $J(u)$ and $Y(u)$ are also predictable, and so is $S^*$ since it depends on a deterministic process $[\alpha(t)]$ and a predictable function. In addition, the integrand is bounded, what implies that (1.18) is also a (local) square integrable martingale. Obviously $(\hat{S}(0) - S^*(0))/S^*(0) = 0$, and hence,

$$1 = \mathbb{E}\{\frac{\hat{S}(t)}{S^*(t)}\} \leq \frac{\mathbb{E}\{\hat{S}(t)\}}{S(t)} \Rightarrow \mathbb{E}\{\hat{S}(t)\} \geq S(t).$$

Therefore, $\hat{S}(t)$ is biased but when $J(t) = 0$ with small probability, the bias becomes negligible. However, when considering the weak convergence, the influence of such bias may be considerable (since we have the term $\sqrt{n}$ involved) and, $S^*(t)$ [instead of $S(t)$] is considered below.

An estimate for the variance of the Kaplan-Meier estimator can be obtained by noticing that $\text{Var}\{(\hat{S}(t) - S^*(t))/S^*(t)\} = \mathbb{E}\{(\hat{S}(t) - S^*(t))/S^*(t)\}^2$. Since the squared quantity is a submartingale, we may apply the Doob-Meier decomposition and then,

$$\mathbb{E}\left\{\frac{\hat{S}(t) - S^*(t)}{S^*(t)}\right\}^2 = \mathbb{E}\left\{\frac{\hat{S}(t) - S^*(t)}{S^*(t)}\right\}^2.$$ 

---

2 More recent works have considered the idea of *localization* in order to relax boundedness conditions.
where the term within \( \langle \cdot \rangle \) is the compensator for \( ((\tilde{S}(t) - S^*(t))/S^*(t))^2 \). If \( \mathcal{F}_t \) is the filtration of the counting process, we have that

\[
d\left( \frac{\tilde{S}(t) - S^*(t)}{S^*(t)} \right) = \text{Var}\left\{ \frac{d(\tilde{S}(t) - S^*(t))}{S^*(t)} \bigg| \mathcal{F}_{t-} \right\}
= \text{Var}\left\{ \frac{\tilde{S}(t^-) J(t)}{S^*(t) Y(t)} dM(t) \bigg| \mathcal{F}_{t-} \right\} = \left( \frac{\tilde{S}(t^-) J(t)}{S^*(t) Y(t)} \right)^2 \text{Var}\left\{ dM(t) \bigg| \mathcal{F}_{t-} \right\}
\approx \left( \frac{\tilde{S}(t^-) J(t)}{S^*(t) Y(t)} \right)^2 \alpha(t) \, dt = \left( \frac{\tilde{S}(t^-) J(t)}{S^*(t) Y(t)} \right)^2 \frac{J(t)}{Y(t)} \, dA(t).
\]

Therefore,

\[
\text{Var}\left\{ \frac{\tilde{S}(t) - S^*(t)}{S^*(t)} \right\} = \mathbb{E}\left\{ \int_0^t \left( \frac{\tilde{S}(x^-)}{S^*(x)} \right)^2 \frac{J(x)}{Y(x)} \, dA(x) \right\}
= \int_0^t \mathbb{E}\left\{ \left( \frac{\tilde{S}(x^-)}{S^*(x)} \right)^2 \frac{J(x)}{Y(x)} \right\} \, dA(x).
\]

If we replace \( \tilde{S}(t^-) \) and \( S^*(t) \) by \( S(t) \) [what, in some sense, may be justified by the fact that \( S(t) \) is assumed continuous and assuming a high probability for the event \( \{ Y(t) > 0 \} \)], then an estimate for the variance of the Kaplan-Meier estimator would be given by

\[
\text{Var}\{\tilde{S}(t)\} = (\tilde{S}(t))^2 \int_0^t \frac{dN(x)}{[Y(x)]^2}.
\]

Alternatively, Gill (1980) considers a general function \( A \) which is not necessarily continuous and ends up with an estimator for the variance of \( \tilde{S}(t) \) equals to the Greenwood’s formula.

Large sample properties for the Kaplan-Meier estimator using counting processes also relies heavily on (1.18). The consistency for the estimator can be proved using the Lenglart’s inequality, i.e., it is shown that on intervals \([0, t]\) for which \( S(t) > 0 \),

\[
\sup_{x \in [0, t]} |\tilde{S}(x) - S(x)| \xrightarrow{p} 0,
\]

15
what implies the uniform consistency.

The weak convergence was also proved in Gill (1980). However, a direct martingale-based proof is presented in Andersen et al. (1993) and basically makes use of the Rebolledo’s martingale central limit theorem to show that, satisfied certain conditions,

$$\sqrt{n} \frac{\hat{S} - S^*}{S^*} \overset{D}{\to} -U,$$

as $n \to \infty$, where $U$ is a Gaussian martingale with $U(0) = 0$ and covariance given by $\text{Cov}\{U(x_i), U(x_j)\} = \sigma^2(x_i \wedge x_j)$. Further, it is shown that

$$\sup_{x \in [0, t]} \sqrt{n} \left( \frac{S(x) - S^*(x)}{S^*(x)} \right) \overset{p}{\to} 0,$$

so that

$$\sqrt{n} \left( \frac{\hat{S} - S}{S^*} \right) = \sqrt{n} \left( \frac{\hat{S} - S^*}{S^*} \right) + \sqrt{n} \left( \frac{S^* - S}{S^*} \right) \overset{D}{\to} -U.$$

A further extension was presented by Gill (1983) where the weak convergence as well as some other related results are presented considering the whole support of $\hat{S}$, making use of stopped processes and considering a continuous distribution $F$ in a random censorship model.

2.4 Cox’s regression model

Assume now that the failure times $\{T_i; i = 1, \ldots, n\}$ have the hazard function (1.1) given by

$$\alpha_i(t) = \alpha_0(t) \exp\{\beta_0^T x_i\}, \quad t \geq 0,$$

(1.19)

where $\beta_0$ is a $p \times 1$ vector of unknown coefficients that parameterizes the regression of survival times $T_i$ on the $p \times 1$ vector of covariates $x_i$ and $\alpha_0(t)$ is an arbitrary
and unspecified baseline hazard function. The model (1.19) proposed by Cox (1972) is a consequence of the assumption that individuals with different covariates have proportional hazard functions (for all $t$) and it is known as the proportional hazards model.

Estimates of $\beta_0$ can be derived considering the partial likelihood function [Cox (1975)], given by

$$L(\beta) = \prod_{i \in D} \left\{ \frac{\exp\{\beta^T x_i\}}{\sum_{j \in R_i} \exp\{\beta^T x_j\}} \right\},$$

where $D$ denotes the set of indices $i = 1, \ldots, n$ corresponding to individuals who died and $R_i$ is the risk set, i.e., the set of indices corresponding to individuals alive and uncensored at time $t_i$. Based on (1.20) we can obtain the maximum partial likelihood estimator $\hat{\beta}$. Asymptotic properties for such estimator in the traditional approach was considered by Tsiatis (1981). He expressed the log-likelihood using empirical quantities and proved strong consistency. He also proved that $\sqrt{n}(\hat{\beta} - \beta_0)$ converges to a normal distribution, approximating it by a sum of random variables which converges jointly to a multivariate normal distribution. Similar techniques as those used by Breslow and Crowley (1974) were considered. Estimates of the cumulative hazard function were also studied using stochastic integrals, providing the asymptotic normality.

Using discrete-time martingales, Sen (1981a) developed the asymptotic distribution theory for the score function test statistic based on the Cox's likelihood. These results are based on induced order statistics and permutational arguments (valid under the null hypothesis $H_0: \beta = 0$.)

Counting processes may be used to derive properties for the Cox's regression model as shown by Gill (1984), based on formal derivations developed in Andersen and Gill (1982). In order to describe their work, we will consider the same notation and framework as in the previous section, but in addition, we assume that $\{T_i; i =$

17
1, \ldots, n} have a more general hazard function than (1.19), given by
\[ \alpha_i(t) = \alpha_0(t) \exp\{\beta_0^T X_i(t)\}, \quad t \geq 0, \tag{1.21} \]
where \( X_i(t) \) is a \( p \times 1 \) vector of (possibly) time-dependent covariates for the \( i \)-th individual.

Now we consider that the \( i \)-th element of \( N(t) \) has intensity (1.3) given by
\[ \lambda_i(t) = Y_i(t) \alpha_0(t) \exp\{\beta_0^T X_i(t)\}, \tag{1.22} \]
where \( Y_i(t) \) and \( X_i(t) \) are assumed to be predictable and locally bounded random processes. Using this set up, we may rewrite (1.20) as
\[ L(\beta) = \prod_{s \geq 0} \prod_{i=1}^n \left\{ \frac{Y_i(s) \exp\{\beta^T X_i(s)\}}{\sum_{j=1}^n Y_j(s) \exp\{\beta^T X_j(s)\}} \right\} \ dN_i(s). \tag{1.23} \]
Treating (1.23) as an ordinary likelihood function in \( \beta \), we can make inferences on \( \beta_0 \) (e.g., obtaining confidence intervals and significance tests). As usual, we take the logarithm of the Cox's likelihood evaluated at time \( t \),
\[ C(\beta, t) \propto \int_0^t \sum_{i=1}^n \beta^T X_i(s) \ dN_i(s) - \int_0^t \log \sum_{j=1}^n Y_j(s) \exp\{\beta^T X_j(s)\} \ dN(s), \]
where \( N(t) = \sum_{i=1}^n N_i(t) \). Taking the derivative in relation to \( \beta \), it follows that
\[ U(\beta, t) = \frac{\partial C(\beta, t)}{\partial \beta} \]
\[ = \int_0^t \sum_{i=1}^n X_i(s) \ dN_i(s) - \int_0^t \frac{\sum_{j=1}^n Y_j(s) X_j(s) \exp\{\beta^T X_j(s)\}}{\sum_{j=1}^n Y_j(s) \exp\{\beta^T X_j(s)\}} \ dN(s). \tag{1.24} \]

If \( \beta = \beta_0 \) and using (1.6) with \( \overline{M}(t) = \sum_{i=1}^n M_i(t) \), we have
\[ U(\beta_0, t) = \int_0^t \sum_{i=1}^n X_i(s) \ dM_i(s) - \int_0^t \frac{\sum_{j=1}^n X_j(s) \lambda_j(s)}{\sum_{j=1}^n \lambda_j(s)} \ d\overline{M}(s). \tag{1.25} \]
It may be noted here that \( \lambda_j \) depends only on \( X_j \) and \( Y_j \) as random components and those are predictable processes. Thus, \( \lambda_j \) is also a predictable process and hence, (1.24) can be seen as the sum of \( n \) stochastic integrals. The same may be observed for (1.25) but the integrals involve also the martingales \( M_i \), what implies that \( U(\beta_0, t) \) is also a martingale.

Considering the Taylor expansion of (1.24) around \( \beta_0 \) for a fixed \( t \), we may write

\[
U(\beta, t) = U(\beta_0, t) + \frac{\partial U(\beta, t)}{\partial \beta} \bigg|_{\beta = \beta^*} (\beta - \beta_0),
\]

where \( \beta^* \) is on the line segment between \( \beta \) and \( \beta_0 \). Here

\[
\frac{\partial U(\beta, t)}{\partial \beta} = -\int_0^t \sum_{j=1}^n \frac{Y_j(s)X_j(s)\exp{\{\beta^T X_j(s)\}}}{\sum_{j=1}^n Y_j(s)\exp{\{\beta^T X_j(s)\}}} dN(s)
\]

\[
+ \int_0^t \left( \frac{\sum_{j=1}^n Y_j(s)X_j(s)\exp{\{\beta^T X_j(s)\}}}{\sum_{j=1}^n Y_j(s)\exp{\{\beta^T X_j(s)\}}} \right)^2 dN(s),
\]

where, using the notation given in Andersen and Gill (1982), \( a^\otimes^2 = aa^T \) for a column vector \( a \).

Assuming that the total interval \( \mathcal{T} \) in which we are working on is \([0, 1]\), we may define the estimator \( \hat{\beta} \) as the solution to the likelihood equation \((\partial/\partial \beta)C(\beta, 1) = U(\beta, 1) = 0\). Inserting \( \hat{\beta} \) in (1.26), and considering \( \mathcal{J}(\beta, t) = -\frac{\partial U}{\partial \beta}U(\beta, t) \), we get

\[
n^{-1/2}(U(\beta_0, 1) - U(\hat{\beta}, 1)) = n^{-1/2}\mathcal{J}(\beta^*, 1)(\hat{\beta} - \beta_0)
\]

\[
\Rightarrow n^{-1/2}U(\beta_0, 1) = \{n^{-1}\mathcal{J}(\beta^*, 1)\}n^{1/2}(\hat{\beta} - \beta_0).
\]

Andersen and Gill (1982) proved the consistency of \( \hat{\beta} \) noting that the process \( X(\beta, 1) = n^{-1}(C(\beta, 1) - C(\beta_0, 1)) \) has unique maximum at \( \beta = \hat{\beta} \) (with probability tending to 1.) Applying the Lenglart’s inequality they proved that \( X(\beta, 1) \) converges (in probability) to a concave function of \( \beta \), which has a unique maximum at \( \beta_0 \). It

19
is then proved that \( \hat{\beta} \xrightarrow{p} \beta_0 \). Boundary problems as those encountered by Tsiatis (1981) are avoided by using the concavity of \( X(\beta, 1) \), what implies that it has a unique root \( \hat{\beta} \), and it converges to a deterministic function which is also proved to be concave. Hence, using some convex function theory, the pointwise convergence (in a compact set) of such function is proved to be a uniform convergence, what proves the desired result about the convergence of \( \hat{\beta} \).

The asymptotic normality of \( n^{1/2}(\hat{\beta} - \beta) \) is proved considering (1.27). Basically, we have to prove the weak convergence of \( U(\beta_0, \cdot) \) to a Gaussian process and the convergence in probability of \( n^{-1}g(\beta^*, 1) \) to a nonsingular matrix. Since, as noted earlier, \( U(\beta_0, t) \) is a martingale, Andersen and Gill (1982) make use of the Rebolledo's central limit theorem for martingales. Using the intuitive approach described in Gill (1984), this implies in showing that, as \( n \to \infty \), \( n^{-1/2}U(\beta_0, t) \) becomes nearly continuous and also that its predictable variation process becomes deterministic. In order to have an idea about how this can be proved, the (simpler) following situation is considered. It is assumed that \( \beta = \beta \) and \( X_i = X_1 \) are scalar and that \( |X_i(t)| \leq C/2 < \infty \) for all \( i \) and \( t \). Thus, we may write

\[
|d(n^{-1/2}U(\beta_0, t))| \leq n^{-1/2} \sum_{i=1}^{n} \left( |X_i(t)| + \frac{\sum_{j=1}^{n} |X_j(t)| \lambda_j(t)}{\sum_{j=1}^{n} \lambda_j(t)} \right) dM_i(t)
\]

\[
\leq n^{-1/2} \sum_{i=1}^{n} \left( \frac{C}{2} + \frac{\sum_{j=1}^{n} (C/2) \lambda_j(t)}{\sum_{j=1}^{n} \lambda_j(t)} \right) dM_i(t)
\]

\[
= n^{-1/2} \sum_{i=1}^{n} C \, dM_i(t) \leq n^{-1/2} C,
\]

where the last inequality follows from the fact that the jumps of \( M_i \) have size +1 (coinciding with the jumps of \( N_i \)) and by the assumption that they do not jump at the same time. By (1.28) we see that \( |d(n^{-1/2}U(\beta, t))| \to 0 \) as \( n \to \infty \) and, hence, \( n^{-1/2}U(\beta, t) \) tends to be continuous as \( n \to \infty \).

In order to verify the second condition of the Rebolledo's theorem, we need to compute the predictable variation process for the martingale \( n^{-1/2}U(\beta_0, t) \). In
other words, we compute $\text{Var}\{ d(n^{-1/2}U(\beta_0, t)) \mid \mathcal{F}_t \}$ and working on (1.25) we end up with an expression based on $\text{Var}\{ dM_i(t) \mid \mathcal{F}_t \}$ (using again the predictability of $X_i$ and $\lambda_i$.) By expression (1.6) this will imply in computing $\text{Var}\{ dN_i(t) \mid \mathcal{F}_t \}$ that is equal to $\lambda_i(t) \, dt$ since $dN_i(t)$ is a 0–1 random variable. Working in the resulting expression, we get the quantity presented by Gill (1984),

$$
(n^{-1/2}U(\beta_0, t)) = \int_0^t \frac{1}{n} \sum_{i=1}^n X_i^2(s)Y_i(s) \exp\{\beta_0X_i(s)\} - \frac{((1/n) \sum_{i=1}^n X_i(s)Y_i(s) \exp\{\beta_0X_i(s)\})^2}{(1/n) \sum_{j=1}^n Y_j(s) \exp\{\beta_0X_j(s)\}} \alpha_0 \, ds.
$$

(1.29)

If the averages $(1/n) \sum_{i=1}^n (X_i(s))^rY_i(s) \exp\{\beta_0X_i\}$, $r = 0, 1, 2$, converge to a constant, then the desired result holds, i.e., the predictable process converge to a deterministic function, as $n \to \infty$. Hence, the two conditions of the Rebolloedo's Theorem are satisfied and the asymptotic normality for the score statistic is proved.

The convergence in probability of $n^{-1}J(\beta^*, 1)$ is proved based on the fact that it converges in probability to, say, $\sigma^2$ for any $\beta^* \xrightarrow{p} \beta_0$ and taking into account the consistency of $\hat{\beta}$. Hence, by the Slutsky theorem, we have $n^{1/2}(\hat{\beta} - \beta_0) \xrightarrow{D} N(0, \sigma^{-2})$.

The results described above were extended to a $nk$-dimensional counting process by Andersen and Borgan (1985) considering a multivariate counting process with elements $N_{ij}(t)$ representing the number of type $i$ events on $[0, t]$ for individual $j$. In such case, some additional regularity conditions on the limiting functions for the averages of the covariates, as well as boundedness condition for the averages, are needed.

Risk functions others than the exponential type used above were considered by Prentice and Self (1983) giving similar results. However, additional conditions were needed since the concavity of $X(\beta, 1)$ holds only inside a neighborhood of $\beta_0$.

It is noted that the model (1.22) assumes time-dependent covariates gen-
eralizing, in this sense, the results obtained by Tsiatis (1981). A further generalization is studied by Murphy and Sen (1991) where time-dependent coefficient $\beta(t)$ is considered. The approach used in this case is based on the method of sieves in order to deal with an infinite-dimensional parameter. The basic idea is to partition the time domain into a large number ($K_n$) of sub-intervals, say $I^n_i$, $i = 1, \ldots, K_n$, such that the ratio of the maximum to the minimum length be finite as the number of elements in the partition approaches infinity. In each sub-interval we consider a fixed value for the parameter $\beta_i$, in such way that, as the number of elements in the partition increases, we can approximate the original function $\beta(t)$. In other words, the parameter space is discretized and can be written as \( \{\beta: \beta(u) = \sum_{i=1}^{K_n} \beta_i I\{u \in I^n_i\}; (\beta_1, \ldots, \beta_{K_n}) \in \mathbb{R}^{K_n}\} \). The Cox's log-likelihood may be rewritten accordingly and used to compute the maximum likelihood estimator. One important aspect is the number of sub-intervals in the partition, as a too small number will increase the bias whereas a too large number will increase the variance of the estimator. Murphy and Sen (1991) also discussed the generalization of the results when $\beta(t)$ is a $p \times 1$ vector.

### 2.5 Further results

Johansen (1983) derived the Nelson-Aalen estimator as the maximum likelihood estimator in an extended model, where the compensators for the counting processes are not assumed absolutely continuous. Other MLE's approaches were derived in Jacobsen (1982) and Jacobsen (1984) and although the resulting estimators are different, they are asymptotically equivalent to the Nelson-Aalen estimator. At this point we should mention that the estimation of the hazard function (rather than the cumulative hazard function) has also been considered by several authors. We may mention, for example, Karr (1987) that considered the maximum likelihood

The study of product-limit type estimators for the survival distribution and transition probabilities of Markov chain models has also been considered under the counting process perspective. Considering the situation and notation briefly described at the end of subsection (2.1), we may recall that the intensity process for the counting process $N_{ij}(t)$ (counting the number of transitions from $i$ to $j$, $i, j \in S$) has the multiplicative form $\alpha_{ij}(t)Y_j(t)$. The approach given by Aalen and Johansen (1978) is based on the fact that the transition probability matrix $P(s, t) = (P_{ij}(t))$ is given by [cf. Theorem II.6.7 in Andersen et al. (1993)]

$$P(s, t) = \prod_{(s,t)} (I + dA(u)), \quad s \leq t < \tau,$$

where $I$ is the identity matrix, $A$ is a matrix whose $(i, j)$th element is given by $A_{ij} = \int \alpha_{ij}(x) \, dx$, (with $\alpha_{ii} = -\sum_{i \neq j} \alpha_{ij}$) and, $\tau = \sup\{u: \int_0^u \alpha_{ij}(x) \, dx < \infty, \ i \neq j\}$. Also, the expression above involves the product-integral representation that becomes a finite product over the jump times of $A$ when it is a step-function. As was the case with the Kaplan-Meier estimator, we may replace $A_{ij}$ by the Nelson-Aalen estimator $\hat{A}_{ij} = \int_{Y_i \leq \tau} dN_{ij}$, $I_{i} = I\{Y_i > 0\}$, and then define the Aalen-Johansen estimator

$$\bar{P}(s, t) = \prod_{(s,t)} (I + d\hat{A}(x)), \quad \text{(1.30)}$$

where $\hat{A}_{ii} = -\sum_{i \neq j} \hat{A}_{ij}$. Based on those quantities, we may again obtain a martingale representation for (1.30) and then the uniform consistency and the weak convergence for $\bar{P}(s, \cdot)$.

With respect to the issue of nonparametric testing for the hazard function, Andersen, Borgan, Gill and Keiding (1982) showed how the counting process representation of test statistics can actually be considered as a generalization of the one-sample log-rank test. They developed the idea that for testing $H_0: \alpha = \alpha_0$ we
can compare the increments of the Nelson-Aalen estimator $\tilde{A}(t)$ with $\alpha_0(t) \, dt$. Thus, using the quantities $N(\cdot)$ and $Y(\cdot)$ as defined in (1.8) and $J(t) = \mathbb{I}\{Y(t) > 0\}$, we may write, using (1.11) and $A_0^\ast(t) = \int_0^t J(x) \alpha_0(x) \, dx$,

$$Z(t) = \int_0^t K(x) \, d\{\tilde{A}(x) - A_0^\ast(x)\}, \quad (1.31)$$

where $K$ is a locally bounded nonnegative predictable process. Thus, under $H_0$, $Z$ will be a (local) square integrable martingale with predictable variation process given by $\langle Z \rangle(t) = \int_0^t K^2(x) \, d\langle \tilde{A} - A_0^\ast \rangle(x)$. Thus, it can be shown that $Z_0(t)\langle\langle Z \rangle(t)\rangle^{-1/2}$ is approximately standard normally distributed (the proof of this fact can be seen in Andersen et al. (1993) and is a direct application of Rebolledo's theorem.) By choosing $K = Y$ we end up with the one-sample log-rank statistic.

The case of the two-sample problem of testing $H_0: \alpha_1 = \alpha_2$, was also considered in Aalen (1978) and following such work, Andersen et al. (1982) introduced a class of statistics for testing $H_0: \alpha_1 = \alpha_2 = \ldots = \alpha_k$. In that paper it is also derived the large sample properties for weighted stochastic processes involving the Nelson-Aalen estimator and an estimator of the hypothesized common value similar to (1.31), applying the martingale central limit theorem. This general test statistic may be derived in a somewhat similar manner as above. Now we consider $N_i(t)$ with intensity process $\alpha_i(t)Y_i(t)$, $i = 1, \ldots, k$. For each group we compute the Nelson-Aalen estimator $\tilde{A}_i(t)$ and the idea is to compare them with the hypothesized common value $A(t) = \int_0^t \alpha(s) \, ds$, with $\alpha$ denoting the common value of $\alpha_1, \ldots, \alpha_k$. $A(t)$ can be estimated by $\tilde{A}(t) = \int_0^t J(x)[Y_+(x)]^{-1} \, dN_+(x)$, where $Y_+ = \sum_{i=1}^k Y_i$, $N_+ = \sum_{i=1}^k N_i$, and, $J(t) = \mathbb{I}\{Y_+(t) > 0\}$. The comparison makes sense only for those $t$ such that $Y_i(t) > 0$, so, we may define $\tilde{A}_i(t) = \int_0^t J_i(x) \, d\tilde{A}(x)$. Then, similarly to (1.31) we define the stochastic process $Z_i(t) = \int_0^t K_i(x) \, d(\tilde{A} - \tilde{A})(x)$, where $K_i$ again is assumed to be a locally bounded predictable weight process. Quadratic forms based on $Z^T(t) = (Z_1(t), \ldots, Z_k(t))$ may be defined, and, for suitable choices
of $K_i$, used to test $H_0$. For example, if $K(t) = I\{Y_i(t) > 0\}$, we obtain the log-rank test for censored failure time data and $K(t) = \sum_{i=1}^{k} Y_i(t)$ will give the Gehan-Breslow generalization (with respect to right-censored data) for the Wilcoxon and Kruskal-Wallis tests. It can be shown that several other test statistics proposed in the literature are particular cases of the quantities described.

Although we are not considering the parametric approach, it should be mentioned that when the underlying hazard function is given via a parametric specification in the multiplicative intensity model, maximum likelihood methods can be used. This situation has been considered by Borgan (1984). Considering some commonly used expressions for the underlying hazard functions [cf. described in, for example, Kalbfleisch and Prentice (1980)], we can express the corresponding likelihood in terms of counting processes. Borgan (1984) showed that there is, with probability one, exactly one consistent solution to the parameters involved and that it is asymptotically multinormal. The Lenglart’s inequality is used to prove the consistency, based on the law of large numbers in the classical set-up. The asymptotic normality is derived by the martingale central limit theorem. It follows that the likelihood ratio test statistics can be considered to test hypothesis using the chi-square distribution.

Finally, other results where counting processes have been applied may be found in the literature but we do not consider them here. However, we may refer to Andersen et al. (1993) for such material, where detailed formulations may be found.

2.6 Some examples

In this subsection we describe two examples that can be found in the literature where the counting process approach has been applied. We have collected them from Andersen et al. (1993), where more complete analyses as well as bibliographic
references may be found.

We first consider the study of survival among 225 patients with skin cancer (melanoma) which had the tumor removed at a hospital in Denmark, in the period of 1962 to 1973. Deaths (when observed) of patients were recorded until the end of 1977. In this prospective clinical experiment, the main objective was to assess how significant were age at operation and sex, among other risk factors, on the survival time since operation. Also, characteristics related with the tumor (such as width, location on the body, type of malignant cells, etc.) were considered. The status of these characteristics was assessed only at time of entry. Here, patients who were still alive at the study termination were considered right-censored, as well as those who died of causes unrelated to cancer. We may define, for each individual, the quantities $T_i$ as the time from operation until death from the disease, $C_i$ as the time from operation until the end of the study or death by another (unrelated) disease, $N(t)$ as the number of deaths from cancer at time $t$ (here also we can define $N_{\text{male}}$ and $N_{\text{female}}$ as those quantities discriminated by sex), and $Y(t)$ as the number of individuals alive and uncensored at time $t$ (with corresponding definitions for males and females.) Thus, the Nelson-Aalen estimator for the cumulative hazard function as well as the Kaplan-Meier estimator for the survival function may be defined. Comparisons between males and females may be considered using the log-rank test (with $K(t) = I\{Y(t) > 0\}$ in the previous section). Considering the covariates $Z_{1i} = I\{\text{patient } i \text{ is male}\}$ and $Z_{2i}$ as the centered thickness in millimeters of the tumor for patient $i$, the results derived for the Cox's regression model may be used to study the effect of those factors on the survival time.

As a second example, we consider a randomized clinical trial where 438 patients, with histologically verified liver cirrhosis at several hospitals in Copenhagen, were randomly divided in two groups. The first one (with 251 patients) received the hormone prednisone whereas the second group (237 patients) received placebo. The
The basic objective of this trial was to assess the effect of prednisone in prolonging the survival of patients with cirrhosis, but was also of interest to study the effect of the hormone on the level of prothrombin. Thus, each individual had the prothrombin index measured in each scheduled visit to the hospital. If the index was less than 70% of the normal value, the patient was considered to be in the low state, otherwise he or she was considered to be in the normal state. Variables such as sex and age were also recorded at the time of entry and others such as alcohol consumption and nutritional status were recorded at the time of entry and in each follow-up visit. Patients entered in the study between 1962 and 1969 but were followed until September 1974. Those alive at that date or lost to follow-up alive before were considered censored.

In this example, individuals with low levels of prothrombin could change to normal levels and vice-versa. The situation is illustrated in Figure 1. Here we may consider a representation based on a three-state Markov chain with two transient states (normal and low) and one absorbing state (death). Thus, a multivariate counting process may be defined and represented by \((N_{01}, N_{02}, N_{10}, N_{12})\) where \(N_{ij}(t)\) counts the number of transitions from state \(i\) to state \(j\) up to time \(t\) and, also, we may define \(Y_i(t), i = 0, 1,\) as the number of patients at risk in state \(i\) at time \(t\). Supposing \(s < T_1 < T_2 < \cdots < T_m \leq t\) as the times of the observed transitions between any two states, then we may estimate the transition matrix using the Aalen-Johansen estimator (1.30) to get \(\hat{P}(s, t) = \prod_{i=1}^{m}(I + \Delta \hat{A}(T_i))\), where

\[
I + \Delta \hat{A}(T_i) = \begin{pmatrix}
1 - \Delta N_0(T_i)/Y_0(T_i) & \Delta N_{01}(T_i)/Y_0(T_i) & \Delta N_{02}(T_i)/Y_0(T_i) \\
\Delta N_{01}(T_i)/Y_0(T_i) & 1 - \Delta N_1(T_i)/Y_1(T_i) & \Delta N_{12}(T_i)/Y_1(T_i) \\
0 & 0 & 1
\end{pmatrix},
\]

with \(N_0 = N_{01} + N_{02}\) and \(N_1 = N_{10} + N_{12}\). The behavior of the prednisone in relation to placebo has been analyzed considering the Nelson-Aalen estimates \(\hat{A}_{02}\) and \(\hat{A}_{12}\) (with respect to survival experience), and \(\hat{A}_{01}\) and \(\hat{A}_{10}\) with respect to
transitions between low and normal levels of the prothrombin index. Also, using the Nelson-Johansen estimates the groups were compared with respect to the survival probabilities $1 - P_{h2}(0, t)$, $h = 0$ or 1. The groups were statistically compared using the log-rank and the Gehan-Breslow tests [for suitable choices of $K_i(t)$, cf. section (2.5)]. The comparisons took place for each possible transition separately.

2.7 Two important results

In the counting process approach two central results dealing with square integrable martingales are frequently used. Since in this work they are considered in several occasions, we state them in this subsection.

The first result usually called the Lenglart’s Inequality, is often considered when proving consistency. It allows us to work with the predictable variation process to draw conclusions about quantities involving martingales. It can be stated as follows. Let $N$ be a univariate counting process with continuous compensator $\Lambda$, such that $M = N - \Lambda$. In addition, let $H$ be a bounded and predictable process. Then, for all $\delta, \rho$ and any $t \geq 0$,

$$\mathbb{P}\left\{ \sup_{0 \leq y \leq t} \left| \int_0^y H(x) \, dM(x) \right| \geq \rho \right\} \leq \frac{\delta}{\rho^2} + \mathbb{P}\left\{ \int_0^t H^2(x) \, d\Lambda(x) \geq \delta \right\}.$$  

The second result is a version of the Central Limit Theorem for martingales,
known as the Rebolledo’s Theorem and it can be stated as follows. If $M^{(n)} = (M_1^{(n)}, \ldots, M_k^{(n)})$ is a vector of $k$ square integrable martingales, for $n = 1, 2, \ldots$ and defining for each $\epsilon > 0$, the jump process $M_\epsilon^{(n)}$ (a $k$-vector of square integrable martingales containing the jumps of the components of $M^{(n)}$ that are larger than $\epsilon$), then considering the conditions (for $t \in \mathcal{T}$)

$$\langle M^{(n)} \rangle \xrightarrow{P} V(t),$$  \hspace{1cm} (1.32)

$$[M^{(n)}](t) \xrightarrow{P} V(t),$$  \hspace{1cm} (1.33)

$$\langle M^{(n)}_{ch} \rangle(t) \xrightarrow{P} V(t),$$  \hspace{1cm} (1.34)

as $n \to \infty$, it follows that either of (1.32) and (1.33), together with (1.34), will imply

$$(M^{(n)}(t_1), \ldots, M^{(n)}(t_i)) \xrightarrow{D} (W(t_1), \ldots, W(t_i)) \quad \text{as } n \to \infty$$

for all $t_1, \ldots, t_i \in \mathcal{T}$, where $W$ is a continuous Gaussian vector martingale with $\langle W \rangle = [W] = V$, a continuous deterministic $k \times k$ positive definite matrix-valued function with positive semi-definite increments, zero at time zero.

If in addition $\mathcal{T}$ is dense, then the same conditions imply

$$M^{(n)} \xrightarrow{D} W \quad \text{in } (D(\mathcal{T}))^k \text{ as } n \to \infty$$

and $\langle M^{(n)} \rangle$ and $[M^{(n)}]$ converge uniformly on compact subsets of $\mathcal{T}$, in probability, to $V$. [Here $(D(\mathcal{T}))^k$ is the space of $\mathbb{R}^k$-valued cadlag functions on $\mathcal{T}$ endowed with the Skorohod topology].

3 Outline of the research

In this work we study a model based on a matrix-valued counting process in a one-dimensional time scale. The idea is that $n$ individuals are followed up with respect to the occurrence of a certain number of events, assumed to be correlated. The
information related to each individual is then represented by a vector of counting process, and the collection of all \( n \) vectors gives origin to the matrix-valued process. The associated intensity process, for each individual, is assumed to follow a Cox-type of model, allowing the presence of covariates.

In Chapter 2 the model is presented, initially in the case where for each individual there are two components (in other words, we will observe the occurrence of the event of interest in each component). Some heuristic interpretation for the quantities in the model is based on conditional hazard functions. The basic idea is some sort of generalization of the multiplicative intensity model discussed earlier, so that the use of predictable process associated to each component introduces the possibility that individuals can be right-censored. When more than two components are considered, the intensity becomes somewhat complicated as factors we denote high-order interactions between components take place. If we assume that such interactions are negligible then we are able to obtain a general expression for the intensity process. Considering that no censoring is present, two bivariate parametric models are briefly considered to explore the behavior of our model.

In Chapter 3 we consider the presence of covariates, and making some extra assumptions on the ratio of the conditional hazard functions for each component, we introduce parameters that will, in some sense, indicate the dependence between the components. Such parameter is assumed to be time-independent and hence, we are able to work in a finite-dimensional parameter space. A likelihood based on the concept of product-integral is considered and, modifying the approach considered for the univariate case described earlier we obtain asymptotic properties (such as consistency and asymptotic distribution) for quantities of interest. The two-sample case is considered as an illustration of the regularity conditions.

In Chapter 4 we relax the assumption on the dependence parameter, allowing it to vary with time. In addition, we introduce time-dependent coefficients that
should allow the use of this model when the proportionality assumption is not true. Based on Murphy and Sen (1991) we obtain asymptotic properties for parameter estimators, considering the method of *histogram sieves*.

Finally, in Chapter 5 we discuss possible applications for the model, first considering some approaches to verify if in a particular situation it is possible to work with the model studied in Chapter 3 rather than the time-dependent parameter model considered in Chapter 4. Such tests are based on naive estimators and the discussion is quite informal. A test to verify the treatment effect in the two-sample case for a two-component situation is also discussed. An empirical approach for general covariance matrices is suggested. For the finite-dimensional model discussed in Chapter 3 we propose the use of the traditional test statistics usually considered in the univariate Cox model and their computation is outlined based on some hypothesis of interest.

Several topics need to be considered before effective application of the models and procedures discussed in this work can be considered in practical situations. Those points, that shall be further investigated, are delineated at the end of Chapter 5.
CHAPTER 2

A Matrix-Valued Counting Process Model

1 Introduction

In this chapter we introduce a model based on a matrix-valued counting process, in a one-dimensional time scale. The main idea is to consider a situation where we are interested on the occurrence of more than one event of interest for each individual or item. In the next section we introduce the model, based on the idea that each individual will have a vector of counting processes giving, at time $t$, the status of each component for a particular individual. We start with the bivariate case and then comment on the situation where more components may be observed. In Section 3 two parametric bivariate models are considered in order to illustrate the model in terms of the parameters involved.

2 The matrix-valued counting process model

In order to introduce and develop the model initially we consider a bivariate setting. The extension of the results to the $k$-variate situation is discussed later. Let $(T_1, T_2)$ be a non-negative random vector defined in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. In
principle we assume that the elements of such a vector are not independent, having joint survival function given by $S_{12}(t_1, t_2)$. The marginal survival functions are represented by $S_1(t_1)$ and $S_2(t_2)$. We also assume the presence of censoring, represented by the non-negative random variable $C$ independent of the failure times and defined in the same probability space of $(T_1, T_2)$. The minimum of $T_k$ and $C$ is represented by $Z_k = T_k \wedge C$, with $\delta_k = \mathbb{I}\{Z_k = T_k\}$. We define the counting processes

$$N_k(t) = \mathbb{I}\{Z_k \leq t; \delta_k = 1\}, \quad t \geq 0, \quad k = 1, 2,$$  \hspace{1cm} (2.1)

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

$$N(t) = (N_1(t), N_2(t))^T.$$  \hspace{1cm} (2.2)

We consider the self-exciting filtration \(\{N^*_t, t \geq 0\} = \sigma\{N^*_s, 0 \leq s \leq t\}\), defined by the vector-valued counting process $N^*_t(s)$ with elements $N^*_k(t) = \mathbb{I}\{T_k \leq t\}$ and to take into account the extra random variation introduced by the censoring, we define $\mathcal{G}_t = \sigma\{\mathbb{I}\{s \leq C\}, s \leq t\}$ and consider the enlarged filtration $\mathcal{F}_t = N^*_t \vee \mathcal{G}_t$. In the derivation that follows we will work with $\mathcal{F}_t$ but technically we could consider a more convenient filtration in terms of the available information, i.e., we could work with a filtration generated by the observable part of $N$, defined as $N^c_k = \int \mathbb{I}\{s \leq C\} \, dN_k(s)$. We refer to Andersen et al. (1993), page 136, for a more detailed discussion.

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

$$Y_k(t) = \mathbb{I}\{Z_k \geq t\}, \quad t \geq 0, \quad k = 1, 2,$$  \hspace{1cm} (2.3)
that corresponds to the information whether or not the component $k$ is still at risk (i.e., uncensored and alive or working.) Such a process is assumed to have its value at instant $t$ known just before $t$. 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 intensity process of $N_k$ would be given by

$$\lambda_k(t) = \alpha_k(t)Y_k(t), \quad k = 1, 2,$$  \hspace{1cm} (2.4)

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

$$\alpha_k(t) = \lim_{\Delta t \to 0} \frac{P\{T_k \in [t, t + \Delta t) \mid T_k \geq t\}}{\Delta t}$$  \hspace{1cm} (2.5)

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

$$\lambda(t) = \begin{pmatrix} \lambda_1(t) \\ \lambda_2(t) \end{pmatrix} = \begin{pmatrix} \alpha_1(t) & 0 \\ 0 & \alpha_2(t) \end{pmatrix} \begin{pmatrix} Y_1(t) \\ Y_2(t) \end{pmatrix} = \alpha(t)Y(t)$$  \hspace{1cm} (2.6)

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 of the unknown deterministic functions $\alpha_k$ should change and, also, the off-diagonal elements should be different than zero. Let us approach this situation by considering a generalization of the heuristic arguments given for (2.4) [see, e.g., Andersen et al. (1993)] for the univariate case.] In this case we may write

$$\lambda_k(t) = E\{ dN_k(t) \mid \mathcal{F}_{t-}\},$$  \hspace{1cm} (2.7)

i.e., the average of jumps for component $k$ given the information available just before $t$. We may note that in this case $\mathcal{F}_{t-}$ contains information whether or not one (or
both) component(s) have failed just before \( t \). Since the processes \( Y_k \) are predictable, this means we know the value of \( Y_k \) at the instant \( t \). If the component \( k \) have failed before \( t \), then expression (2.7) equals zero. In other words, we need to consider the situations (i) no component has failed before time \( t \), 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

\[
\lambda_1(t) = \mathbb{E}\{ dN_1(t) \mid \mathcal{F}_t \} = p_1^{(1)}(t)Y_1(t)[1 - Y_2(t)] + p_2^{(1)}(t)Y_1(t)Y_2(t)
\]

(2.8)

where \( p_1^{(1)}(t) = \lim_{\Delta t \to 0}(\Delta t)^{-1}\mathbb{P}\{T_1 \in [t, t + \Delta t) \mid T_1 \geq t; T_2 < t\} \) and \( p_2^{(1)}(t) = \lim_{\Delta t \to 0}(\Delta t)^{-1}\mathbb{P}\{T_1 \in [t, t + \Delta t) \mid T_1 \geq t; T_2 \geq t\} \) may be interpreted as conditional hazard functions, given the outcome for the other component. Similarly, for component 2,

\[
\lambda_2(t) = \mathbb{E}\{ dN_2(t) \mid \mathcal{F}_t \} = p_2^{(2)}(t)Y_2(t)[1 - Y_1(t)] + p_1^{(2)}(t)Y_1(t)Y_2(t)
\]

(2.9)

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

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

\[
\lambda(t) = \begin{pmatrix}
\lambda_1(t) \\
\lambda_2(t)
\end{pmatrix} = \begin{pmatrix}
Y_1(t) & 0 \\
0 & Y_2(t)
\end{pmatrix} \begin{pmatrix}
\alpha_{11}(t) & \alpha_{12}(t) \\
\alpha_{21}(t) & \alpha_{22}(t)
\end{pmatrix} \begin{pmatrix}
Y_1(t) \\
Y_2(t)
\end{pmatrix} = \text{Diag}(Y(t))\alpha(t)Y(t)
\]

(2.10)
where the elements of $\alpha$ are given by

\[
\begin{align*}
\alpha_{11}(t) &= p_1^{(1)}(t) \\
\alpha_{12}(t) &= p_2^{(1)}(t) - p_1^{(1)}(t) \\
\alpha_{21}(t) &= p_1^{(2)}(t) - p_2^{(2)}(t) \\
\alpha_{22}(t) &= p_2^{(2)}(t)
\end{align*}
\]

The matrix-valued counting process model is defined in the following way. Suppose that $N_1, \ldots, N_n$ are $n$ copies of the process $N$ defined on (2.2). Then the matrix-valued counting process is given by

\[N(t) = (N_1(t), \ldots, N_n(t)), \quad (2.11)\]

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

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 complicated 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 $T = (T_1, T_2, T_3)'$, 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-vector of counting processes $N_i = (N_{i1}, N_{i2}, N_{i3})'$, based on $n$ copies $T_i$ of $T$. Also, the predictable vector is given by $Y_i = (Y_{i1}, Y_{i2}, Y_{i3})'$, where $Y_{ki}(t) = 1\{Z_{ki} \geq t\}$.

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 $Y_i(t)$. 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 $\mathbf{Y}$ 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

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

- 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

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

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

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

- 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

$$p_{1}^{(1)}(t) = \lim_{\Delta t \to 0} \frac{\mathbb{P}\{T_1 \in [t, t + \Delta t] \mid T_1 \geq t, T_2 < t, T_3 < t\}}{\Delta t},$$
so that the intensity process will be given by one of the four expressions above, that, depending on the value of $Y$, can be written as,

\[
\lambda_1(t) = p_1^{(1)}(t)Y_1(t)[1 - Y_2(t)][1 - Y_3(t)] \\
+ p_{12}^{(1)}(t)Y_1(t)Y_2(t)[1 - Y_3(t)] \\
+ p_{13}^{(1)}(t)Y_1(t)[1 - Y_2(t)][1 - Y_3(t)] \\
+ p_{123}^{(1)}(t)Y_1(t)Y_2(t)Y_3(t)
\]

\[
= p_1^{(1)}(t)Y_1(t) + (p_{13}^{(1)} - p_1^{(1)})Y_1(t)Y_3(t) + (p_{12}^{(1)} - p_1^{(1)}Y_1(t)Y_2(t) \\
+ (p_{123}^{(1)} - p_{12}^{(1)} - p_{13}^{(1)} + p_1^{(1)}Y_1(t)Y_2(t)Y_3(t) \\
= \alpha_1^{(1)}(t)Y_1(t) + \alpha_1^{(1)}(t)Y_1(t)Y_3(t) + \alpha_1^{(1)}(t)Y_1(t)Y_2(t) + \alpha_1^{(1)}(t)Y_1(t)Y_2(t)Y_3(t).
\]

(2.12)

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

\[
\lambda_2(t) = p_2^{(1)}(t)Y_2(t) + (p_{23}^{(1)} - p_2^{(1)}Y_2(t)Y_3(t) + (p_{12}^{(1)} - p_2^{(1)}Y_1(t)Y_2(t) \\
+ (p_{123}^{(1)} - p_{12}^{(1)} - p_{23}^{(1)} + p_2^{(1)}Y_1(t)Y_2(t)Y_3(t) \\
= \alpha_2^{(1)}(t)Y_2(t) + \alpha_2^{(1)}(t)Y_2(t)Y_3(t) + \alpha_2^{(1)}(t)Y_1(t)Y_2(t) + \alpha_2^{(1)}(t)Y_1(t)Y_2(t)Y_3(t),
\]

(2.13)

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

\[
\lambda_3(t) = p_3^{(1)}(t)Y_3(t) + (p_{32}^{(1)} - p_3^{(1)}Y_2(t)Y_3(t) + (p_{13}^{(1)} - p_3^{(1)}Y_1(t)Y_3(t) \\
+ (p_{123}^{(1)} - p_{13}^{(1)} - p_{32}^{(1)} + p_3^{(1)}Y_1(t)Y_2(t)Y_3(t) \\
= \alpha_3^{(1)}(t)Y_3(t) + \alpha_3^{(1)}(t)Y_2(t)Y_3(t) + \alpha_3^{(1)}(t)Y_1(t)Y_3(t) + \alpha_3^{(1)}(t)Y_1(t)Y_2(t)Y_3(t).
\]

(2.14)

Based on expressions (2.12)–(2.14) we may note that each expression has a term involving the predictable process for the corresponding component, $(3)^{(2)}$ terms.
involving the product of two predictable processes and one term involving the product of the three processes \( Y_k, \ k = 1, 2, 3 \). 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, we may write the model trying to emphasize this,

\[
\lambda(t) = \begin{pmatrix}
\lambda_1(t) \\
\lambda_2(t) \\
\lambda_3(t)
\end{pmatrix} = \begin{pmatrix}
\alpha_1^{(1)}(t) \\
\alpha_2^{(2)}(t) \\
\alpha_3^{(3)}(t)
\end{pmatrix} Y_1(t) + \begin{pmatrix}
0 \\
0 \\
0
\end{pmatrix} Y_2(t) + \begin{pmatrix}
0 \\
0 \\
\alpha_3^{(3)}(t)
\end{pmatrix} Y_3(t)
\]

\[
+ \begin{pmatrix}
\alpha_1^{(1)}(t) \\
\alpha_2^{(2)}(t) \\
0
\end{pmatrix} Y_1(t) Y_2(t) + \begin{pmatrix}
\alpha_1^{(1)}(t) \\
0 \\
\alpha_3^{(3)}(t)
\end{pmatrix} Y_1(t) Y_3(t) + \begin{pmatrix}
0 \\
\alpha_2^{(2)}(t) \\
\alpha_3^{(3)}(t)
\end{pmatrix} Y_2(t) Y_3(t)
\]

\[
+ \begin{pmatrix}
\alpha_1^{(1)}(t) \\
\alpha_2^{(2)}(t) \\
\alpha_3^{(3)}(t)
\end{pmatrix} Y_1(t) Y_2(t) Y_3(t)
\]

(2.15)

where the first three terms in the r.h.s. of expression (2.15) 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 (2.15) as a product of
matrices similarly to the one in (2.10), given by

\[
\lambda(t) = \begin{pmatrix} Y_1(t) & 0 & 0 \\ 0 & Y_2(t) & 0 \\ 0 & 0 & Y_3(t) \end{pmatrix} \begin{pmatrix} \alpha_{11}(t) & \alpha_{12}(t) & \alpha_{13}(t) \\ \alpha_{21}(t) & \alpha_{22}(t) & \alpha_{23}(t) \\ \alpha_{31}(t) & \alpha_{32}(t) & \alpha_{33}(t) \end{pmatrix} \begin{pmatrix} Y_1(t) \\ Y_2(t) \\ Y_3(t) \end{pmatrix} \\
+ \begin{pmatrix} \alpha_{123}^{(1)}(t) \\ \alpha_{123}^{(2)}(t) \\ \alpha_{123}^{(3)}(t) \end{pmatrix} Y_1(t)Y_2(t)Y_3(t),
\]

(2.16)

where the elements \(\alpha_{ij}\) are defined by the equality

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

Assuming that the second order interaction is negligible, the above expression turns out to be

\[
\lambda(t) = \text{Diag}(Y(t))\alpha(t)Y(t),
\]

(2.17)

that is similar to (2.10).

The same reasoning can be considered for higher dimension problems, with the additional complication that we have 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

\[
\lambda_j(t) = \beta_j(t)Y_j(t)(1 + \sum_{i=1, i\neq j}^K \gamma_{ij}(t)Y_i(t) + \sum_{i,j=1, i\neq j}^K \delta_{ij}(t)Y_i(t)Y_j(t) + \ldots)
\]

where \(\beta_j = \alpha_{jj}^{(j)}\), \(\gamma_{ij} = \alpha_{ij}^{(j)} / \alpha_{jj}^{(j)}\), \(\delta_{ij}^{(j)} = \alpha_{ij}^{(j)} / \alpha_{jj}^{(j)}\) 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 (2.17), i.e., in general,

\[
\lambda(t) = \begin{pmatrix}
Y_1(t) & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & Y_K(t)
\end{pmatrix}
\begin{pmatrix}
\alpha_{11}(t) & \cdots & \alpha_{1K}(t) \\
\vdots & \ddots & \vdots \\
\alpha_{K1}(t) & \cdots & \alpha_{KK}(t)
\end{pmatrix}
\begin{pmatrix}
Y_1(t) \\
\vdots \\
Y_K(t)
\end{pmatrix}
\]  

(2.18)

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 (2.18) has \(K^2\) infinite-dimensional parameters represented by the functions \(\alpha_{ij}\). As we will see later, additional assumptions are imposed to such a 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 (2.18).

In what follows we illustrate the bivariate model considering two parametric models.
3 Some examples

There are several definitions of bivariate exponential models and a exposition of some of them can be seen in Johnson and Kotz (1972). Here we will consider two models, one described by Gumbel (1960) that is a special application of the general formula

$$F_{X,Y}(x, y) = F_X(x)F_Y(y)[1 + \alpha(1 - F_X(x))(1 - F_Y(y))], \quad -1 \leq \alpha \leq 1.$$ 

Note that $X$ and $Y$ are independent if and only if $\alpha = 0$. The other model is based on the work by Sarkar (1987) who proposed an absolutely continuous model.

The basic goal is to derive an expression for the 2-vector of intensity process $\lambda(t)$. The basic setup is as follows.

We consider a non-censored situation where the lifetimes for two components in a given unit are represented by $T_1$ and $T_2$, satisfying a given bivariate exponential distribution. Based on these nonnegative random variables we define the processes

$$N_i(t) = \mathbb{I}\{T_i \leq t\} \quad \text{and} \quad Y_i(t) = \mathbb{I}\{T_i \geq t\}, \quad i = 1, 2,$$

for $t \in \mathcal{T}$. Our goal is to compute $\mathbb{E}\{\, dN_i(t) \mid \mathcal{F}_t^-\}$, where the history $\mathcal{F}_t^- = \sigma\{(N_1(s), N_2(s)) : s < t\}$ incorporates all the information up to (but not including) $t$ for the random variables $T_i$ as well as the relationship between them.

Given the dependence between $T_1$ and $T_2$ (and, hence, between $N_1$ and $N_2$), at a given time $t \in \mathcal{T}$ we may have the following situations:

i. $T_1 < t$ and $T_2 < t$;

ii. $T_1 < t$ and $T_2 \geq t$;

iii. $T_1 \geq t$ and $T_2 < t$;

iv. $T_1 \geq t$ and $T_2 \geq t$.
The intensity process for $N_1$ ($N_2$) will be zero in the first (last) two situations so that, by (2.8) and (2.9) we have

$$\lambda_i(t) = \mathbb{E}\{ \, dN_i(t) \mid \mathcal{F}_t \} = \mathbb{P}\{ \, dN_i(t) = 1 \mid \mathcal{F}_t \}$$

$$= \lim_{\Delta t \to 0} \frac{\Delta t}{\mathbb{P}\{ T_i \in [t, t + \Delta t) \mid T_i \geq t; T_j < t \} Y_i(t)(1 - Y_j(t))}$$

$$+ \lim_{\Delta t \to 0} \frac{\Delta t}{\mathbb{P}\{ T_i \in [t, t + \Delta t) \mid T_i \geq t; T_j \geq t \} Y_i(t)Y_j(t)}$$

$$= \lim_{\Delta t \to 0} \frac{\Delta t}{\mathbb{P}\{ T_i \in [t, t + \Delta t); T_j < t \} Y_i(t)(1 - Y_j(t))}$$

$$+ \lim_{\Delta t \to 0} \frac{\Delta t}{\mathbb{P}\{ T_i \in [t, t + \Delta t); T_j \geq t \} Y_i(t)Y_j(t)}$$

(2.19)

for $i, j = 1, 2$, $i \neq j$.

In the next two subsections we consider the two bivariate models, compute the joint probabilities specified in (2.19) and write the bivariate intensity process in a matrix form trying to emphasize the underlying dependence structure.

### 3.1 Gumbel bivariate exponential distribution

In this subsection we consider the Gumbel model described in Johnson and Kotz (1972), but we assume that the marginal exponential distributions have means different than one. In other word, the joint distribution function for $(T_1, T_2)$, depending on the parameters $\delta = (\theta_1, \theta_2, \alpha)$, is given by

$$F(u, v) = (1 - e^{-\theta_1 u})(1 - e^{-\theta_2 v})(1 + \alpha e^{-(\theta_1 u + \theta_2 v)}), \quad \theta_1, \theta_2 > 0,$$

where $\alpha \in [-1, 1]$. As noted earlier, when $\alpha = 0$, $T_1$ and $T_2$ are independent, and $F(u, v)$ reduces to a product of two exponential distributions with parameters $\theta_1$ and $\theta_2$.

The density function is given by

$$f(u, v) = \theta_1 \theta_2 e^{-(\theta_1 u + \theta_2 v)} \left(1 + \alpha \theta_1 \theta_2 (2e^{-\theta_1 u} - 1)(2e^{-\theta_2 v} - 1)\right).$$
Based on expression (2.19), we derive the intensity process for \(N_1(t)\), given by \(p_1^{(1)}[Y_1(1-Y_2)] + p_2^{(1)}Y_1Y_2\). That implies in computing the conditional hazard functions, given by ratios of quantities involving the joint density. Hence, the numerator of \(p_1^{(1)}\) is given by

\[
\lim_{\Delta t \to 0} (\Delta t)^{-1} \mathbb{P}\{T_1 \in [t, t + \Delta t); T_2 < t\} = \int_0^t f(t, v) \, dv = \\
= \int_0^t \theta_1 \theta_2 e^{-(\theta_1 \xi + \theta_2 \xi)} \{1 + \alpha \theta_1 \theta_2 (2e^{-\theta_1 \xi} - 1)(2e^{-\theta_2 \xi} - 1)\} \, dv \\
= \theta_1 e^{-\theta_1 t}(1 - e^{-\theta_2 t})\{1 + \alpha \theta_1 \theta_2 (2e^{-\theta_1 t} - 1)e^{-\theta_2 t}\}.
\]

The denominator is given by

\[
\mathbb{P}\{T_1 \geq t; T_2 < t\} = \int_t^\infty \int_0^t f(u, v) \, dv \\
= \int_t^\infty \int_t^v \theta_1 \theta_2 e^{-(\theta_1 u + \theta_2 \xi)} \{1 + \alpha \theta_1 \theta_2 (1 - 2e^{-\theta_1 u})(1 - 2e^{-\theta_2 \xi})\} \, du \, dv \\
= e^{-\theta_1 t}(1 - e^{-\theta_2 t})\{1 + \alpha \theta_1 \theta_2 (e^{-\theta_1 t} - 1)e^{-\theta_2 t}\}.
\]

Hence, the conditional hazard function for the first component when \(Y_1(t) = 1\) and \(Y_2(t) = 0\) is given by

\[
p_1^{(1)}(t) = \frac{\theta_1 \{1 + \alpha \theta_1 \theta_2 (2e^{-\theta_1 t} - 1)e^{-\theta_2 t}\}}{1 + \alpha \theta_1 \theta_2 (e^{-\theta_1 t} - 1)e^{-\theta_2 t}} = \theta_1 + \alpha \frac{\theta_1 \theta_2 e^{-(\theta_1 + \theta_2) t}}{1 + \alpha \theta_1 \theta_2 (e^{-\theta_1 t} - 1)e^{-\theta_2 t}}.
\]

As for \(p_2^{(1)}\), the numerator is given by

\[
\lim_{\Delta t \to 0} (\Delta t)^{-1} \mathbb{P}\{T_1 \in [t, t + \Delta t); T_2 \geq t\} = \int_t^\infty f(u, v) \, dv \\
= \theta_1 e^{-(\theta_1 + \theta_2) t}\{1 + \alpha \theta_1 \theta_2 (2e^{-\theta_1 t} - 1)(e^{-\theta_2} - 1)\}.
\]

Similarly, the denominator is

\[
\mathbb{P}\{T_1 \geq t; T_2 \geq t\} = \int_t^\infty \int_t^\infty f(u, v) \, dv \\
= e^{-(\theta_1 + \theta_2) t}\{1 + \alpha \theta_1 \theta_2 (e^{-\theta_1 t} - 1)(e^{-\theta_2 t} - 1)\}.
\]
Therefore, the conditional hazard function when \( Y_1(t) = Y_2(t) = 1 \) is given by
\[
p_2^{(1)}(t) = \frac{\theta_1 \{ 1 + \alpha \theta_1 \theta_2 (2e^{-\theta_1 t} - 1)(e^{-\theta_2 t} - 1) \}}{1 + \alpha \theta_1 \theta_2 (e^{-\theta_1 t} - 1)(e^{-\theta_2 t} - 1)} = \theta_1 + \alpha g_2^{(1)}(t, \delta).
\]

Combining both the conditional hazard functions above, if follows that the intensity process for the first component of \( \mathbf{N} \) is given by
\[
\lambda_1(t) = [\theta_1 + \alpha g_1^{(1)}(t, \delta)]Y_1(t)[1 - Y_2(t)] + [\theta_1 + \alpha g_2^{(1)}(t, \delta)]Y_1(t)Y_2(t)
\]
\[
= [\theta_1 + \alpha g_1^{(1)}(t, \delta)]Y_1(t) + \alpha [g_2^{(1)}(t, \delta) - g_1^{(1)}(t, \delta)]Y_1(t)Y_2(t).
\]

The computations for the intensity process associated to the second component follow the same steps above so that, defining \( g_1^{(2)} \) and \( g_2^{(2)} \) conveniently, we have
\[
\lambda_2(t) = [\theta_2 + \alpha g_1^{(2)}(t, \delta)]Y_2(t) + \alpha [g_2^{(2)}(t, \delta) - g_1^{(2)}(t, \delta)]Y_1(t)Y_2(t),
\]
and hence, the vector of intensity processes is given by
\[
\mathbf{\lambda}(t) = \begin{pmatrix}
[\theta_1 + \alpha g_1^{(1)}(t, \delta)]Y_1(t) + \alpha [g_2^{(1)}(t, \delta) - g_1^{(1)}(t, \delta)]Y_1(t)Y_2(t) \\
[\theta_2 + \alpha g_1^{(2)}(t, \delta)]Y_2(t) + \alpha [g_2^{(2)}(t, \delta) - g_1^{(2)}(t, \delta)]Y_1(t)Y_2(t)
\end{pmatrix}
\]
\[
= \begin{pmatrix}
\theta_1 & 0 \\
0 & \theta_2
\end{pmatrix}
\begin{pmatrix}
Y_1(t) \\
Y_2(t)
\end{pmatrix}
\]
\[
+ \alpha
\begin{pmatrix}
Y_1(t) & 0 \\
0 & Y_2(t)
\end{pmatrix}
\begin{pmatrix}
g_1^{(1)}(t, \delta) & g_2^{(1)}(t, \delta) - g_1^{(1)}(t, \delta) \\
g_2^{(2)}(t, \delta) - g_1^{(2)}(t, \delta) & g_2^{(2)}(t, \delta)
\end{pmatrix}
\begin{pmatrix}
Y_1(t, \delta) \\
Y_2(t, \delta)
\end{pmatrix}.
\]

In the above expression we see that the relationship between the two components is quite complex (given the functional form of the functions \( g_i^{(j)} \)), however note that, as expected, when \( T_1 \) and \( T_2 \) are independent, \( \mathbf{\lambda} \) reduces to (2.6) for two independent exponential variables.
3.2 Sarkar bivariate exponential distribution

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

\[
P\{T_1 \geq u; T_2 \geq v\} = \begin{cases} 
  e^{-(\beta_2 + \beta_{12})v(1 - [A(\beta_1 v)]^{-\gamma}[A(\beta_1 u)]^{1+\gamma})}, & 0 < u \leq v \\
  e^{-(\beta_1 + \beta_{12})u(1 - [A(\beta_2 u)]^{-\gamma}[A(\beta_2 v)]^{1+\gamma})}, & 0 < v \leq u
\end{cases}
\]

where \(\beta_1 > 0, \beta_2 > 0, \beta_{12} \geq 0, \gamma = \beta_{12}/(\beta_1 + \beta_2)\) and \(A(z) = 1 - e^{-z}, z > 0\). This model is based on modifications on the characterization property for bivariate exponential distributions, stating that the following 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 (2.20) if \(\beta_{12} = 0\) the joint distribution can be expressed as the product of two exponential distributions and then \(T_1\) and \(T_2\) are independent.

The corresponding density function can be expressed as

\[
f(u, v) = \begin{cases} 
  \beta_1(1 + \gamma)e^{-\beta_1 u - (\beta_2 + \beta_{12})v}\{\beta_1 \gamma + \beta_2(1 + \gamma)A(\beta_1 v)\} \\
  \quad \times [A(\beta_1 u)]^{-\gamma}[A(\beta_1 v)]^{-(1+\gamma)}, & 0 < u \leq v, \\
  \beta_2(1 + \gamma)e^{-\beta_2 v - (\beta_1 + \beta_{12})u}\{\beta_2 \gamma + \beta_1(1 + \gamma)A(\beta_2 u)\} \\
  \quad \times [A(\beta_2 v)]^{-\gamma}[A(\beta_2 u)]^{-(1+\gamma)}, & 0 < v \leq u
\end{cases}
\]

where \(\beta = \beta_1 + \beta_2 + \beta_{12}\). Based on this model we define the bivariate counting process as well as the corresponding predictable processes as discussed earlier. Hence, for \(t \in \mathcal{T}\), the numerator of the conditional hazard function \(p_1^{(1)}(t)\) is given by

\[
\lim_{\Delta t \to 0} (\Delta t)^{-1}P\{T_1 \in [t, t + \Delta t); T_2 < t\} = \int_0^t f(t, v) \, dv \\
= (1 + \gamma)e^{-(\beta_1 + \beta_{12})t}[A(\beta_2 t)]^{-(1+\gamma)}\{\beta_2 \gamma + \beta_1(1 + \gamma)A(\beta_2 t)\} \int_0^t \beta_2 e^{-\beta_2 v}[A(\beta_2 v)]^{-\gamma} \, dv \\
= e^{-(\beta_1 + \beta_{12})t}\{\beta_2 \gamma + \beta_1(1 + \gamma)A(\beta_2 t)\}.
\]
As for the denominator a more elaborated manipulation is needed. In this case,

\[ \mathbb{P}\{T_1 \geq t; \ T_2 < t\} = \int_t^\infty \int_0^t f(u,v) \, dv \]

\[ = (1 + \gamma) \int_t^\infty e^{\beta_1 + \beta_{12}} u \{ \beta_2 \gamma + \beta_1 (1 + \gamma) A(\beta_2 u)\} [A(\beta_2 u)]^{-(1 + \gamma)} \, du \]

\[ \times \int_0^t \beta_2 e^{-\beta_2 v}[A(\beta_2 v)]^\gamma \, dv. \]

The second integral in the expression above is equal to \([A(\beta_2 t)]^{\gamma+1}/(\gamma + 1)\). The first integral can be expressed as

\[ \beta_2 \gamma \int_t^\infty e^{-(\beta_1 + \beta_{12}) u}[A(\beta_2 u)]^{-(1+\gamma)} \, du + \beta_1 (1 + \gamma) \int_t^\infty e^{-(\beta_1 + \beta_{12}) u}[A(\beta_2 u)]^{-\gamma} \, du. \]

(2.21)

Using the relation \((1 - t)^a = \sum_{k \geq 0} \binom{a}{k} (-t)^k\), for any number \(a\) and \(t \in (-1, 1)\) [see, e.g., Feller (1968), page 51], then by the dominated convergence theorem we can write for the first factor in (2.21)

\[ \int_t^\infty e^{-(\beta_1 + \beta_{12}) u} \sum_{k \geq 0} (-1)^k \binom{-(1 + \gamma)}{k} (e^{-\beta_2 u})^k \, du \]

\[ = \sum_{k \geq 0} (-1)^k \binom{-(1 + \gamma)}{k} \int_t^\infty e^{-(\beta_1 + \beta_{12} + k \beta_2) u} \, du \]

\[ = e^{-(\beta_1 + \beta_{12}) t} \sum_{k \geq 0} (-1)^k \binom{-(1+\gamma)}{k} e^{k \beta_2 t} \frac{1}{\beta_1 + k \beta_2 + \beta_{12}}. \]

Similarly, the second integral in (2.21) is

\[ \int_t^\infty e^{-(\beta_1 + \beta_{12}) u}[A(\beta_2 u)]^{-\gamma} = e^{-(\beta_1 + \beta_{12}) t} \sum_{k \geq 0} (-1)^k \frac{(-\gamma)}{k} e^{k \beta_2 t} \frac{1}{\beta_1 + k \beta_2 + \beta_{12}}. \]

Combining both terms and using that for \(a > 0\), \((-a)_k = (-1)^k \binom{a + k - 1}{k}\) [Feller (1968),
expression (2.21) is equal to
\[
e^{-\left(\beta_1 + \beta_{12}\right)t} \sum_{k \geq 0} \left(-1\right)^k \frac{e^{-k\beta_2 t}}{\beta_1 + k\beta_2 + \beta_{12}} \left\{ \beta_2 \gamma \binom{-\left(1 + \gamma\right)}{k} + \beta_1 \left(1 + \gamma\right) \binom{-\gamma}{k} \right\}
\]
\[
= e^{-\left(\beta_1 + \beta_{12}\right)t} \sum_{k \geq 0} \frac{e^{-k\beta_2 t}}{\beta_1 + k\beta_2 + \beta_{12}} \left\{ \beta_2 \gamma \binom{\gamma + k}{k} + \beta_1 \left(1 + \gamma\right) \binom{\gamma + k - 1}{k} \right\}
\]
\[
= e^{-\left(\beta_1 + \beta_{12}\right)t} \sum_{k \geq 0} \left(\frac{\gamma + k - 1}{k}\right) e^{-k\beta_2 t} = e^{-\left(\beta_1 + \beta_{12}\right)t} \sum_{k \geq 0} \left(-1\right)^k \left(\frac{-\gamma}{k}\right) \left(e^{-\beta_2 t}\right)^k
\]
\[
= e^{-\left(\beta_1 + \beta_{12}\right)t} \left(1 - e^{-\beta_2 t}\right)^{-\gamma}.
\]

Hence,
\[
\mathbb{P}\{T_1 \geq t; T_2 < t\} = e^{-\left(\beta_1 + \beta_{12}\right)t} A(\beta_2 t),
\]

so that
\[
p_1(1)(t) = \beta_1 \left(1 + \gamma\right) + \gamma \frac{\beta_2}{A(\beta_2 t)}.
\]

The same type of manipulations applies to determining $p_2(1)(t)$. In this case the numerator will be given by
\[
\lim_{\Delta t \to 0} (\Delta t)^{-1} \mathbb{P}\{T_1 \in [t, t + \Delta t); T_2 \geq t\} = \int_t^\infty f(t, v) \, dv = \beta_1 \left(1 + \gamma\right) e^{-\beta t},
\]
and the denominator is given by the survival function (2.20) in the case $u \leq v$, so that
\[
p_2(1)(t) = \frac{\beta_1 \left(1 + \gamma\right) e^{-\beta t}}{e^{-\beta t}} = \beta_1 \left(1 + \gamma\right).
\]

Thus,
\[
\lambda_1(t) = p_1(1)(t)Y_1(t) + \left[p_2(1)(t) - p_1(1)(t)\right]Y_1(t)Y_2(t)
\]
\[
= \left\{\beta_1 \left(1 + \gamma\right) + \gamma \frac{\beta_2}{A(\beta_2 t)}\right\}Y_1(t) - \gamma \frac{\beta_2}{A(\beta_2 t)} Y_1(t)Y_2(t).
\]
Similar derivation can be employed to find the intensity process for the second component, so that,

\[
\lambda_2(t) = p_1^{(2)}(t)Y_2(t) + [p_2^{(2)}(t) - p_1^{(2)}(t)]Y_1(t)Y_2(t)
\]

\[
= \left\{ \beta_2(1 + \gamma) + \gamma \frac{\beta_1}{A(\beta_1 t)} \right\} Y_2(t) - \gamma \frac{\beta_1}{A(\beta_1 t)} Y_1(t)Y_2(t),
\]

and hence, the intensity process for the vector \( \mathbf{N}(t) \) is given by

\[
\lambda(t) = \begin{pmatrix} Y_1(t) & 0 \\ 0 & Y_2(t) \end{pmatrix} \begin{pmatrix} \beta_1(1 + \gamma) + \gamma \frac{\beta_2}{A(\beta_2 t)} & -\gamma \frac{\beta_2}{A(\beta_2 t)} \\ -\gamma \frac{\beta_1}{A(\beta_1 t)} & \beta_2(1 + \gamma) + \gamma \frac{\beta_1}{A(\beta_1 t)} \end{pmatrix} \begin{pmatrix} Y_1(t) \\ Y_2(t) \end{pmatrix}.
\]

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

\[
\lambda(t) = \begin{pmatrix} \beta_1 & 0 \\ 0 & \beta_2 \end{pmatrix} \begin{pmatrix} Y_1(t) \\ Y_2(t) \end{pmatrix} + \gamma \begin{pmatrix} \beta_1 + \frac{\beta_2}{A(\beta_2 t)} & -\frac{\beta_2}{A(\beta_2 t)} Y_1(t) \\ -\frac{\beta_1}{A(\beta_1 t)} Y_2(t) & \beta_2 + \frac{\beta_1}{A(\beta_1 t)} \end{pmatrix} \begin{pmatrix} Y_1(t) \\ Y_2(t) \end{pmatrix}.
\]

In this expression it is clear the dependence structure in the sense that when \( \gamma = 0 \) the resulting expression for the intensity process vector is the same we would obtain when working with two independent exponential random variables with parameters \( \beta_1 \) and \( \beta_2 \).
CHAPTER 3

The Matrix-Valued Model With Finite-Dimensional Parameters

1 Introduction

In this chapter we generalise the intensity process discussed in Chapter 2 considering that, in addition to the failure times we also observe (time-dependent) covariates. This information is incorporated into the model by considering a semi-parametric approach similar to the one used for the Cox-model in the univariate case. It is assumed that the ratio of the conditional hazard functions (for each component) is constant. Such an assumption will be relaxed in the next chapter but here it will allow us to work with a finite-dimensional parameter space.

The derivation that follows is based on the situation where we have two events of interest for each item or individual. The advantage of considering this case is that we do not need any further assumption with respect to higher-order interactions; however, in the case where we have more than two events, the results can be easily extended (except for a more cumbersome notation) if we assume that the effect of those high-order interactions can be neglected.

In Section 2 we define the bivariate model and corresponding intensity process and discuss the likelihood function; in Section 3 we derive the quantities that
are needed to prove the asymptotic properties for the maximum partial likelihood estimator (MPLE), what is considered in Section 4, making use of the theory of martingales developed for counting processes presented in Chapter 1.

2 Bivariate set-up

In this section we consider the situation where \( n \) individuals are followed-up and for each one we are interested in the lifetime of two components. In other words, we observe the occurrence of up to two events for each individual. The censoring mechanism is related to the individual rather than to the component. We begin with the description of the model.

2.1 The model

Keeping in mind the ideas discussed in Chapter 2, suppose that \((T_1, T_2)\) are two dependent nonnegative random variables representing failure times and \(C\) is a non-negative censoring variable. Based on \((T_1, T_2)\) and \(C\), let \((N_1(t), N_2(t))^{T}\) be a bivariate counting process observed over the time interval \(\mathcal{T} = [0, \tau]\) and defined on a base \((\Omega, \{\mathcal{F}_t : t \in \mathcal{T}\}, \mathbb{P})\). Recall that \(Z_k = T_k \wedge C\) and \(Y_k(t) = 1\{Z_k \geq t\}\) is assumed \(\mathcal{F}_t\)-predictable. In addition now we assume that a set of \(q\) time-dependent covariates is observed for each individual and, without loss of generality we consider \(q = 1\). This covariate, represented by \(X(t)\) is supposed to be \(\mathcal{F}_t\)-predictable and bounded.

The intensity process for such model is represented by the product of matrices

\[
\lambda(t) = \begin{pmatrix}
\lambda_1(t) \\
\lambda_2(t)
\end{pmatrix} = \begin{pmatrix}
Y_1(t) & 0 \\
0 & Y_2(t)
\end{pmatrix} \begin{pmatrix}
\alpha_{11}(t) & \alpha_{12}(t) \\
\alpha_{21}(t) & \alpha_{22}(t)
\end{pmatrix} \begin{pmatrix}
Y_1(t) \\
Y_2(t)
\end{pmatrix}
= \text{Diag}(Y(t))\alpha(t)Y(t),
\]

51
where, for \( t \in \mathcal{T} \) the (unknown) elements in \( \alpha(t) \) are

\[
\alpha_{11}(t) = \lim_{h \to 0} h^{-1} \mathbb{P}\{T_1 \in [t, t+h) \mid T_1 \geq t; T_2 < t\} = p_1^{(1)}(t),
\]

\[
\alpha_{12}(t) = \lim_{h \to 0} h^{-1} \mathbb{P}\{T_1 \in [t, t+h) \mid T_1 \geq t; T_2 \geq t\}
- \lim_{h \to 0} h^{-1} \mathbb{P}\{T_1 \in [t, t+h) \mid T_1 \geq t; T_2 < t\}
= p_2^{(1)}(t) - p_1^{(1)}(t),
\]

\[
\alpha_{21}(t) = \lim_{h \to 0} h^{-1} \mathbb{P}\{T_2 \in [t, t+h) \mid T_1 \geq t; T_2 \geq t\}
- \lim_{h \to 0} h^{-1} \mathbb{P}\{T_2 \in [t, t+h) \mid T_1 < t; T_2 \geq t\}
= p_1^{(2)}(t) - p_2^{(2)}(t),
\]

\[
\alpha_{22}(t) = \lim_{h \to 0} h^{-1} \mathbb{P}\{T_2 \in [t, t+h) \mid T_1 < t; T_2 \geq t\} = p_2^{(2)}(t). \tag{3.1}
\]

Consider now \( n \) copies \( N_1, \ldots, N_n \) of \( N \) with corresponding intensity processes given by \( \lambda_1, \ldots, \lambda_n \), where

\[
\lambda_i(t) = \begin{pmatrix} \lambda_{1i}(t) \\ \lambda_{2i}(t) \end{pmatrix} = \begin{pmatrix} \alpha_{11}(t)Y_{1i}(t) + \alpha_{12}(t)Y_{1i}(t)Y_{2i}(t) \\ \alpha_{22}(t)Y_{2i}(t) + \alpha_{21}(t)Y_{1i}(t)Y_{2i}(t) \end{pmatrix}. \tag{3.2}
\]

Since for each individual \( i \) we observe \( X_i(t), \ i = 1, \ldots, n \), let us follow the univariate setting, assuming that each element of \( \alpha(t) \) can be expressed through a multiplicative form. In other words, we assume that each limit described in (3.1) may be represented as

\[
\begin{align*}
    p_1^{(1)}(t) &= \gamma_{11}(t)e^{\beta_1 X_i(t)} \\
    p_2^{(1)}(t) &= \gamma_{12}(t)e^{\beta_1 X_i(t)} \\
    p_1^{(2)}(t) &= \gamma_{21}(t)e^{\beta_2 X_i(t)} \\
    p_2^{(2)}(t) &= \gamma_{22}(t)e^{\beta_2 X_i(t)}
\end{align*}
\Rightarrow
\begin{align*}
    \alpha_{11}^i &= \gamma_{11}(t)e^{\beta_1 X_i(t)} = \alpha_{11}^2(t) e^{\beta_1 X_i(t)} \\
    \alpha_{12}^i &= (\gamma_{12}(t) - \gamma_{11}(t))e^{\beta_1 X_i(t)} = \alpha_{12}^2(t) e^{\beta_1 X_i(t)} \\
    \alpha_{21}^i &= (\gamma_{21}(t) - \gamma_{22}(t))e^{\beta_2 X_i(t)} = \alpha_{21}^2(t) e^{\beta_2 X_i(t)} \\
    \alpha_{22}^i &= \gamma_{22}(t)e^{\beta_2 X_i(t)} = \alpha_{22}^2(t) e^{\beta_2 X_i(t)} \tag{3.3}
\end{align*}
\]

In addition, we simplify further the model with the assumption that \( \alpha_{12}(t) = \theta_1 \alpha_{11}(t) \) and \( \alpha_{21}(t) = \theta_2 \alpha_{22}(t) \). Note that this assumption implies that, for each
component the conditional hazard functions are proportional, e.g., for component 1, 
$p_{2i}^{(1)}(t) = (1 + \theta_1)p_{1i}^{(1)}(t)$ and in order to avoid the degenerated cases we shall assume
$\theta_k > -1$, $k = 1, 2$. Therefore, using the notation $\alpha_{kk} = \alpha_k$, the intensity process
vector can be written as

$$\lambda_i(t) = \left( \begin{array}{c}
\alpha_1(t)(Y_{1i}(t) + \theta_1 Y_{1i}(t) Y_{2i}(t))e^{\beta_1 X_i(t)} \\
\alpha_2(t)(Y_{2i}(t) + \theta_2 Y_{1i}(t) Y_{2i}(t))e^{\beta_2 X_i(t)}
\end{array} \right),$$

for $i = 1, \ldots, n$.

In order to illustrate future developments and assumptions, let us consider a simple situation described in the following example.

**EXAMPLE 3.1** 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 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 $k$th event has occurred for individual $i$ at time $t$. \(\Box\)

### 2.2 The likelihood function

The problem at hand consists in finding estimates for the parameters $\beta_k$ and $\theta_k$, $k = 1, 2$ and in order to determine an appropriated likelihood we proceed as follows. 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 $\beta_k$ and $\theta_k$ 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 \( \alpha(t) \) we are able to cancel the unknown baseline functions \( \alpha_{ij}^\circ(t) \). Let \( \delta = (\delta_1, \delta_2) \), with \( \delta_k = (\beta_k, \theta_k) \in \mathbb{R} \times (-1, +\infty) \) and suppose the true parameter value is represented by \( \delta^\circ \). Then, by the discussion above we write the likelihood as

\[
L_n(\delta) = \prod_{t \geq 0} \prod_{i=1}^{n} \left( \frac{\lambda_{1i}(t)}{\sum_{j=1}^{n} \lambda_{1j}(t)} \right) \frac{dN_{1i}(t)}{dN_{2i}(t)} \left( \frac{\lambda_{2i}(t)}{\sum_{j=1}^{n} \lambda_{2j}(t)} \right) \frac{dN_{2i}(t)}{dN_{2i}(t)}
\]

\[
= \prod_{t \geq 0} \prod_{i=1}^{n} \left( \frac{\alpha_{11}^\delta(t)e^{\beta_1 X_{1i}(t)} + \alpha_{12}^\delta(t)e^{\beta_2 X_{1i}(t)}Y_{1i}(t)Y_{2i}(t)}{\sum_{j=1}^{n} \alpha_{1j}^\delta(t)e^{\beta_1 X_{1j}(t)} + \alpha_{12}^\delta(t)e^{\beta_1 X_{1i}(t)}Y_{1j}(t)Y_{2j}(t)} \right) \frac{dN_{1i}(t)}{dN_{1i}(t)}
	imes \left( \frac{\alpha_{21}^\delta(t)e^{\beta_2 X_{2i}(t)} + \alpha_{22}^\delta(t)e^{\beta_2 X_{2i}(t)}Y_{1i}(t)Y_{2i}(t)}{\sum_{j=1}^{n} \alpha_{2j}^\delta(t)e^{\beta_2 X_{2j}(t)} + \alpha_{22}^\delta(t)e^{\beta_2 X_{2i}(t)}Y_{1j}(t)Y_{2j}(t)} \right) \frac{dN_{2i}(t)}{dN_{2i}(t)}
\]

\[
= \prod_{t \geq 0} \prod_{i=1}^{n} \prod_{k=1}^{2} \left( \frac{\alpha_k^\delta(t)e^{\beta_k X_{ki}(t)}(Y_{ki}(t) + \theta_k Y_{1i}(t)Y_{2i}(t))}{\sum_{j=1}^{n} \alpha_k^\delta(t)e^{\beta_k X_{kj}(t)}(Y_{kj}(t) + \theta_k Y_{1j}(t)Y_{2j}(t))} \right) \frac{dN_{ki}(t)}{dN_{ki}(t)}
\]

(3.5)

Expression (3.5) is in fact a partial likelihood and it can be thought of as a multinomial type of likelihood. Considering the concept of profile likelihood we can justify its form in the following way.\(^1\) We write the partial likelihood for \( N \) as

\[
\prod_{t \geq 0} \prod_{i=1}^{n} \left( d\Lambda_{1i}(t) \right) dN_{1i}(t) \left( d\Lambda_{2i}(t) \right) dN_{2i}(t) \left( 1 - d\Lambda_{..}(t) \right)^{N_{..}(t)}
\]

(3.6)

where \( \Lambda_{ki} = \int \lambda_{ki}, \Lambda_{..} = \sum_{k,i} \Lambda_{ki} \) and \( N_{..} = \sum_{k,i} N_{ki} \). Under the assumption \( \Lambda(\cdot) \) is absolutely continuous and using the properties inherent to the product-integral notation, (3.6) may be rewritten as, with \( w_i(t; \theta_k) = (Y_{ki}(t) + \theta_k Y_{1i}(t)Y_{2i}(t)) \),

\[
\prod_{t \geq 0} \prod_{i=1}^{n} \prod_{k=1}^{2} \left[ d\Lambda_{ki}(t) \right] dN_{ki}(t) \exp \left\{ - \int_0^T d\Lambda_{..}(s) \right\}
\]

\[
= \prod_{t \geq 0} \prod_{i=1}^{n} \prod_{k=1}^{2} \left[ dA_k(t)w_i(t; \theta_k) e^{\beta_k X_i(t)} \right] dN_{ki}(t) e^{-\sum_k \int_0^T dA_k(t)W(t; \delta_k) dt}
\]

(3.7)

\(^1\) Following the same approach given by Andersen et al. (1993), pages 481–482.
where \( A_k(t) = \int_0^t \alpha_k(s) \, ds \) and \( W(t; \delta_k) = \sum_{i=1}^n w_i(t; \theta_k)e^{\delta_k X_i(t)} \).

We may get a value for \( dA_k(t) \) that (point-wise) maximizes (3.7), for \( \beta_k \) and \( \theta_k \) fixed. Initially, consider the log-likelihood

\[
\sum_{t,k,i} dN_{ki}(t) \log( dA_k(t)w_i(t; \theta_k)e^{\delta_k X_i(t)}) - \sum_k \int_0^\tau W(t; \delta_k) \, dA_k(t),
\]

so that, for \( N_k(t) = \sum_{i=1}^n N_{ki}(t) \), taking the derivative with respect to \( dA_k(t) \) we get

\[
N_k(t)/dA_k(t) - W(t; \delta_k) = 0,
\]

what implies

\[
d\hat{A}_k(t) = dN_k(t)/W(t; \delta_k).
\]  

(3.8)

Considering the profile likelihood with (3.8) plugged into (3.7) we get

\[
\prod_{t,k,i} \left[ d\hat{A}_k(t)w_i(t; \theta_k)e^{\delta_k X_i(t)} \right] \frac{dN_{ki}(t)}{d\hat{A}_k(t)} \exp\left\{ -\sum_k \int_0^\tau d\hat{A}_k(t)W(t; \delta_k) \right\}
= \prod_{t,k,i} \left( \frac{dN_k(t)}{W(t; \delta_k)}w_i(t; \theta_k)e^{\delta_k X_i(t)} \right) \frac{dN_{ki}(t)}{dN_k(t)} \exp\left\{ -\sum_k \int_0^\tau \frac{dN_k(t)}{W(t; \delta_k)}W(t; \delta_k) \right\}
= \prod_{t,k,i} \left( \frac{w_i(t; \theta_k)e^{\delta_k X_i(t)}}{W(t; \delta_k)} \right) \times \prod_{t,k,i} \left( \frac{dN_k(t)}{dN_k(t)} \right) e^{-N_k(t)},
\]  

(3.9)

where the first component of (3.9) agrees with (3.5).

Based on (3.5), we derive the score function and information matrix that are used in the computation of estimates as commented below.

3 The score function, information matrix and the predictable variation processes

In this section we compute the functions that are needed in the derivations of the asymptotic properties for the MPLE. Since we will make use of the martingale
theory, we also discuss the martingale representation for the score function and derive the predictable (co)variation processes.

3.1 The score function and its martingale representation

Computation of the score vector takes place for a pair of parameters for each component. Therefore it is convenient to consider a partitioned vector where the first element is a 2-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 \( \mathcal{T} \). For \( t \in \mathcal{T} \), we have

\[
U(t; \delta) = \begin{pmatrix} U^{(1)}(t; \delta) \\ U^{(2)}(t; \delta) \end{pmatrix},
\]

where \( U^{(1)} = (U_1^{(1)}, U_2^{(1)})^T \) and \( U^{(2)} = (U_1^{(2)}, U_2^{(2)})^T \). The first element in these vectors corresponds to the parameter \( \beta_k \) and the second element to the parameter \( \theta_k \).

In order to compute such quantities we have that, by (3.5), the log-likelihood is

\[
\log L_n(\delta) = \int_0^r \sum_{i=1}^n \sum_{k=1}^2 \beta_k X_i(t) + \log \left\{ Y_{ki}(t) + \theta_k Y_{1i}(t)Y_{ki}(t) \right\} \\
- \log \left\{ \sum_{j=1}^n e^{\beta_k X_j(t)}(Y_{kj}(t) + \theta_k Y_{1j}(t)Y_{kj}(t)) \right\} \, \text{d}N_{ki}(t) \tag{3.10}
\]

The vector-valued score function may then be computed by simple differentiation, such that we have, with respect to \( \beta_k \) (neglecting the argument \( t \) to simplify the
notation)

\[ U^{(k)}_1(\tau, \delta) = \frac{\partial \log L_n(\delta)}{\partial \beta_k} \]

\[ = \int_0^\tau \sum_{i=1}^n \left( X_i - \frac{\sum_{j=1}^n X_j e^{\beta_k X_j} (Y_{kj} + \theta_k Y_{k_j} Y_{2j})}{\sum_{j=1}^n e^{\beta_k X_j} (Y_{kj} + \theta_k Y_{k_j} Y_{2j})} \right) dN_{ki} \]

\[ = \int_0^\tau \sum_{i=1}^n \left( X_i - \frac{\sum_{j=1}^n X_j w_j(\theta_k) e^{\beta_k X_j}}{\sum_{j=1}^n w_j(\theta_k) e^{\beta_k X_j}} \right) dN_{ki}, \tag{3.11} \]

Note that (3.11) is similar to the expression we obtain when considering the univariate case. The basic difference relies on the weights \( w_j \) that are taking into account the predictable processes related to both components.

As for the element associated with \( \theta_k \) we have

\[ U^{(k)}_2(\delta) = \frac{\partial \log L_n(\delta)}{\partial \theta_k} \]

\[ = \int_0^\tau \sum_{i=1}^n \left( \frac{Y_{k_i} Y_{2i}}{Y_k + \theta_k Y_{k_i} Y_{2i}} - \frac{\sum_{j=1}^n e^{\beta_k X_j} Y_{kj} Y_{2j}}{\sum_{j=1}^n e^{\beta_k X_j} (Y_{kj} + \theta_k Y_{k_j} Y_{2j})} \right) dN_{ki} \]

\[ = \int_0^\tau \sum_{i=1}^n \left( \frac{Y_{k_i} Y_{2i}}{Y_k + \theta_k Y_{k_i} Y_{2i}} - \frac{\sum_{j=1}^n e^{\beta_k X_j} Y_{kj} Y_{2j}}{\sum_{j=1}^n w_j(\theta_k) e^{\beta_k X_j}} \right) dN_{ki}, \tag{3.12} \]

for \( k = 1, 2 \).

Maximum partial likelihood estimators can be obtained by solving the equations

\[ U(\tau, \delta) = 0, \tag{3.13} \]

what need to be computed iteratively as no analytical expression for the estimators can be derived. Let us denote the resulting MPLE by \( \hat{\delta} \). Asymptotic properties for such an estimator are studied using the standard martingale theory, that implies in computing the information matrix, deriving the martingale property for some of the quantities involved and making use of Taylor expansions. Hence, we first note that the elements in the score vector can be represented by a stochastic integral with respect to a martingale when \( \delta = \delta^0 \). Recalling that by the Doob-Meier
decomposition of the counting process, \( dN_{ki}(t) = dM_{ki}(t) + \lambda_{ki}(t) \, dt \) where \( M_{ki}(t) \) is a square integrable martingale, and noting that for \( k = 1, 2 \)

\[
\int_0^\tau \sum_{i=1}^n \left( X_i - \frac{\sum_{j=1}^n X_j(t) w_j(t; \theta_k^o) e^{\beta_k X_j(t)}}{\sum_{j=1}^n w_j(t; \theta_k^o) e^{\beta_k X_j(t)}} \right) w_i(t; \theta_k^o) e^{\beta_k X_i(t)} \alpha_k^o(t) \, dt = 0,
\]

it follows that

\[
U_1^{(k)}(\tau, \delta^o) = \int_0^\tau \sum_{i=1}^n \left( X_i(t) - \frac{\sum_{j=1}^n X_j(t) w_j(t; \theta_k^o) e^{\beta_k^o X_j(t)}}{\sum_{j=1}^n w_j(t; \theta_k^o) e^{\beta_k^o X_j(t)}} \right) dM_{ki}(t),
\]

(3.14)

for \( k = 1, 2 \). Similarly one can easily show that

\[
\int_0^\tau \sum_{i=1}^n \left( \frac{Y_{1i}(t) Y_{2i}(t)}{w_i(t; \theta_k^o)} - \frac{\sum_{j=1}^n e^{\beta_k X_i(t)} Y_{1j}(t) Y_{2j}(t)}{\sum_{j=1}^n w_j(t; \theta_k^o) e^{\beta_k^o X_j(t)}} \right) \lambda_{ki}(t) \, dt = 0,
\]

what implies

\[
U_2^{(k)}(\tau, \delta^o) = \int_0^\tau \sum_{i=1}^n \left( \frac{Y_{1i}(t) Y_{2i}(t)}{w_i(t; \theta_k^o)} - \frac{\sum_{j=1}^n e^{\beta_k X_i(t)} Y_{1j}(t) Y_{2j}(t)}{\sum_{j=1}^n w_j(t; \theta_k^o) e^{\beta_k^o X_j(t)}} \right) dM_{ki}(t),
\]

(3.15)

for \( k = 1, 2 \). Given the predictability and boundedness of \( Y_{ki} \) and \( X_i \) it follows that \( U \) is a vector of square integrable martingales. This fact is used to derive the asymptotic distribution for the score function.

### 3.2 The information matrix and its conditional expectation

After proving the asymptotic normality for the score function we will use a Taylor’s expansion to obtain the weak convergence for the MPLE. At that point we will have to deal with the information matrix. Therefore, we compute the second derivatives of the log-likelihood and then compute their (conditional) expectation in order to obtain interpretation for those quantities.

The second derivatives based on (3.11) and (3.12) are given by (again here we neglect the argument \( t \))

\[
\frac{\partial U_1^{(k)}}{\partial \beta_k}(\tau, \delta) = - \int_0^\tau \sum_{i=1}^n \left\{ \frac{\sum_j X_j^2 w_j(\theta_k) e^{\beta_k X_j}}{\sum_j w_j(\theta_k) e^{\beta_k X_j}} - \left( \frac{\sum_j X_j w_j(\theta_k) e^{\beta_k X_j}}{\sum_j w_j(\theta_k) e^{\beta_k X_j}} \right)^2 \right\} dN_{ki},
\]

(3.16)
\[ \frac{\partial U^{(k)}_1}{\partial \beta_k}(\tau, \delta) = \frac{\partial U^{(k)}_2}{\partial \beta_k}(\tau, \delta) \]
\[ = - \int_0^\tau \sum_{i=1}^n \left( \frac{X_i e^{\beta_k X_i} Y_{i1} Y_{i2}}{\sum_j w_j(\theta_k) e^{\beta_k X_j}} \right) \, dN_{ki} \]

\[ = - \sum_{i=1}^n \left( \frac{X_i e^{\beta_k X_i}}{\sum_j w_j(\theta_k) e^{\beta_k X_j}} \right) \left( \frac{Y_{i1} Y_{i2}}{\sum_j w_j(\theta_k) e^{\beta_k X_j}} \right) \, dN_{ki} \]  
(3.17) 
and

\[ \frac{\partial U^{(k)}_2}{\partial \beta_k}(\tau, \delta) = - \int_0^\tau \sum_{i=1}^n \left\{ \left( \frac{Y_{i1} Y_{i2}}{w_i(\theta_k) e^{\beta_k X_i}} \right)^2 - \left( \frac{\sum_j e^{\beta_k X_j} Y_{i1} Y_{i2}}{\sum_j w_j(\theta_k) e^{\beta_k X_j}} \right)^2 \right\} \, dN_{ki} \]  
(3.18) 

Note that for each \( k \) all other derivatives are zero. This means that the information matrix is block diagonal, with symmetric matrices with main diagonal given by minus the expressions (3.16) and (3.18) and the off-diagonal elements given by minus expression (3.17), i.e.,

\[ J(\delta) = \begin{pmatrix} J_1(\delta) & 0 \\ 0 & J_2(\delta) \end{pmatrix} \]  
(3.19) 

where

\[ J_k(\delta) = \begin{pmatrix} -(\partial/\partial \beta_k)U^{(k)}_1 & -(\partial/\partial \beta_k)U^{(k)}_1 \\ -(\partial/\partial \beta_k)U^{(k)}_2 & -(\partial/\partial \beta_k)U^{(k)}_2 \end{pmatrix} \]

Looking more closely to the elements in the main diagonal of \( J_k(\delta) \) we have, from (3.16), and with \( E_k(t) = \left( \sum_j X_j w_j e^{\beta_k X_j} \right) / \left( \sum_j w_j e^{\beta_k X_j} \right), \)

\[ -\frac{\partial U^{(k)}_1}{\partial \beta_k} = \int_0^\tau \sum_{i=1}^n \frac{X_i^2 w_j(\theta_k) e^{\beta_k X_i} - \sum_{j=1}^n w_j(\theta_k) e^{\beta_k X_j} E_k^2}{\sum_{j=1}^n w_j(\theta_k) e^{\beta_k X_j}} \, dN_{ki} \]
\[ = \int_0^\tau \sum_{i=1}^n \left[ \left( \sum_{j=1}^n X_j^2 w_j(\theta_k) e^{\beta_k X_j} - 2 \sum_{j=1}^n w_j(\theta_k) e^{\beta_k X_j} E_k^2 \right) + \sum_{j=1}^n w_j(\theta_k) e^{\beta_k X_j} E_k^2 \right] \, dN_{ki} \]
\[ \int_0^r \sum_{i=1}^{n} \left[ \left( \sum_{j=1}^{n} X_j^2 w_j(\theta_k) e^{\beta_k X_j} \right) - 2 \sum_{j=1}^{n} X_j w_j(\theta_k) e^{\beta_k X_j} \right] \left( \sum_{j=1}^{n} E_k^2 w_j(\theta_k) e^{\beta_k X_j} \right) \right] \ \text{d}N_{ki} \]

\[ = \int_0^r \sum_{i=1}^{n} \sum_{j=1}^{n} \left( \frac{n_j^2 - 2X_j E_k + E_k^2}{\sum_{j=1}^{n} w_j(\theta_k) e^{\beta_k X_j}} \right) w_j(\theta_k) e^{\beta_k X_j} \ \text{d}N_{ki} \]

\[ = \int_0^r \sum_{i=1}^{n} \sum_{j=1}^{n} \left( \frac{n_j^2 - 2X_j E_k + E_k^2}{\sum_{j=1}^{n} w_j(\theta_k) e^{\beta_k X_j}} \right) \frac{n_j}{\sum_{j=1}^{n} w_j(\theta_k) e^{\beta_k X_j}} \ \text{d}N_{ki}, \]

so that, since the processes \( X_j, E_k(t) \), and \( w_j \) are predictable, when \( \delta = \delta^0 \) the conditional expected value (given \( \mathcal{F}_t \)) is given by

\[ \int_0^r \sum_{i=1}^{n} (X_i(t) - E_k(t))^2 w_i(t; \theta_k^0) e^{\beta_k X_i(t)} \alpha_k^0(t) \ \text{d}t. \]  

(3.20)

Similarly, the conditional expected value of (minus) the expression (3.18) is given by

\[ \int_0^r \sum_{i=1}^{n} \left\{ \left( \frac{Y_{2i}}{w_i(\theta_k^0)} \right)^2 - \left( \frac{\sum_{j=1}^{n} Y_{1j} Y_{2j} e^{\beta_k X_j}}{\sum_{j=1}^{n} w_j(\theta_k^0) e^{\beta_k X_j}} \right)^2 \right\} w_i(t; \theta_k^0) e^{\beta_k X_i(t)} \alpha_k^0(t) \ \text{d}t \]

\[ = \int_0^r \sum_{i=1}^{n} \left\{ \left( \frac{Y_{1i}(t) Y_{2i}(t)}{w_i(t; \theta_k^0)} \right)^2 - \left( \frac{\sum_{j=1}^{n} Y_{1j}(t) Y_{2j}(t) e^{\beta_k X_j(t)}}{\sum_{j=1}^{n} w_j(t; \theta_k^0) e^{\beta_k X_j(t)}} \right)^2 \right\} w_i(t; \theta_k^0) e^{\beta_k X_i(t)} \alpha_k^0(t) \ \text{d}t \]  

(3.21)

Note that expression (3.21) can be rewritten as

\[ \int_0^r \sum_{i=1}^{n} \left( Z_{ki} - \sum_{j=1}^{n} Z_{kj} \frac{w_j(t; \theta_k^0) e^{\beta_k X_j(t)}}{\sum_{i=1}^{n} w_i(t; \theta_k^0) e^{\beta_k X_i(t)}} \right)^2 w_i(t; \theta_k^0) e^{\beta_k X_i(t)} \alpha_k^0(t) \ \text{d}t \]  

(3.22)

where \( Z_{ki} = Y_{1i} Y_{2i} / (Y_{1i} + Y_{2i} \theta_k Y_{1i} Y_{2i}) \), and, hence, expressions (3.20) and (3.22) may be interpreted as weighted measures of variability, the former related to the covariate \( X \) and the latter to the variates \( Z_{kj} \).

For the remaining elements the conditional expected value will result

\[ \int_0^r \sum_{i=1}^{n} \left( \sum_{j=1}^{n} X_{ij} Y_{1j} e^{\beta_k X_j} - \sum_{j=1}^{n} X_{ij} w_j(\theta_k^0) e^{\beta_k X_j} \sum_{j=1}^{n} Y_{1j} Y_{2j} e^{\beta_k X_j} \right) \alpha_k^0(t) \ \text{d}t \]

\[ = \int_0^r \sum_{i=1}^{n} \left( X_i(t) - E_k(t) \right) \left( Z_{ki} - \sum_{j=1}^{n} Z_{kj} \frac{w_j(t; \theta_k^0) e^{\beta_k X_j(t)}}{\sum_{i=1}^{n} w_i(t; \theta_k^0) e^{\beta_k X_i(t)}} \right) w_i(t; \theta_k^0) e^{\beta_k X_i(t)} \alpha_k^0(t) \ \text{d}t \]
3.3 The predictable variation and covariation processes

In order to use the Rebolledo’s theorem for the quantities expressed earlier, we need to consider the predictable process associated with the martingales (3.14) and (3.15). Initially we define the following functions (extending those quantities usually considered in the literature for the univariate case)

\[ S_k^{(j)}(\theta, x) = \left( \frac{1}{n} \right) \sum_{i=1}^{n} [X_i(x)]^j w_i(x; \theta_k) e^{\theta_k X_i(x)}, \quad j = 0, 1, 2, \]

\[ S_k^{(3)}(\theta, x) = \left( \frac{1}{n} \right) \sum_{i=1}^{n} Y_{1i}(x) Y_{2i}(x) e^{\theta_k X_i(x)}, \]

\[ S_k^{(4)}(\theta, x) = \left( \frac{1}{n} \right) \sum_{i=1}^{n} \frac{Y_{1i}(x) Y_{2i}(x)}{w_i(x; \theta_k)} e^{\theta_k X_i(x)}, \]

\[ S_k^{(5)}(\theta, x) = \left( \frac{1}{n} \right) \sum_{i=1}^{n} X_i(x) Y_{1i}(x) Y_{2i}(x) e^{\theta_k X_i(x)}. \]

Defining also \( U_j^{(k,n)} = n^{-1/2} U_j^{(k)} \), \( j = 1, 2 \) and using well-known properties of the predictable processes involved, the predictable variation process for \( U_1^{(k,n)} \) is given by

\[
\langle U_1^{(k,n)}, U_1^{(k,n)} \rangle(t) = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{t} \left( X_i(s) - \frac{\sum_{j=1}^{n} X_j w_j(s; \theta_k) e^{\theta_k X_j(s)}}{\sum_{j=1}^{n} w_j(s; \theta_k) e^{\theta_k X_j(s)}} \right)^2 \lambda_k(s) \, ds
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{t} \left( X_i(s) - \frac{S_k^{(1)}(\theta, s)}{S_k^{(0)}(\theta, s)} \right)^2 \lambda_k(s) \, ds
\]

\[
= \int_{0}^{t} \left( S_k^{(2)}(\theta, s) - \frac{(S_k^{(1)}(\theta, s))^2}{S_k^{(0)}(\theta, s)} \right) \alpha_k(s) \, ds. \quad (3.23)
\]

Similarly, the predictable variation process for \( U_2^{(k,n)} \) will be given by

\[
\langle U_2^{(k,n)}, U_2^{(k,n)} \rangle(t) = \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{t} \left( \frac{Y_{1i}(s) Y_{2i}(s)}{w_i(s; \theta_k)} - \frac{\sum_{j=1}^{n} e^{\theta_k X_j(s)} Y_{1j}(s) Y_{2j}(s)}{\sum_{j=1}^{n} e^{\theta_k X_j(s)}} \right)^2 \lambda_k(s) \, ds
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{t} \left( \frac{Y_{1i} Y_{2i}}{w_i(s; \theta_k)} - \frac{S_k^{(3)}(\theta, s)}{S_k^{(0)}(\theta, s)} \right)^2 \lambda_k(s) \, ds
\]
\[ = \int_0^t \left( S_k^{(4)}(\delta^o, s) - \frac{\left(S_k^{(3)}(\delta^o, s)\right)^2}{S_k^{(0)}(\delta^o, s)} \right) \alpha_k^2(s) \, ds. \]  

(3.24)

Finally, we need to compute the predictable covariation process between \( U_1^{(k,n)} \) and \( U_2^{(k,n)} \). This is given by

\[
\langle U_1^{(k,n)}, U_2^{(k,n)} \rangle(t) = \frac{1}{n} \sum_{i=1}^n \int_0^t \left( X_i(s) - \frac{S_k^{(1)}(\delta^o, s)}{S_k^{(0)}(\delta^o, s)} \right) \left( \frac{Y_{2i}(s)}{\omega_i(s; \theta_k^o)} - \frac{S_k^{(3)}(\delta^o, s)}{S_k^{(0)}(\delta^o, s)} \right) \alpha_k(s) \, ds
\]

\[
= \int_0^t \left( \frac{S_k^{(5)}(\delta^o, s)}{S_k^{(0)}(\delta^o, s)} \right) \alpha_k^2(s) \, ds
\]

(3.25)

4 Asymptotic properties

In this section, making use of the Rebolloso's central limit theorem, we show that the score function converges weakly to a Gaussian process. Then we use this fact to prove the weak convergence for the MPLE. An estimator for the covariance matrix as well as its consistency is derived. We begin stating a set of conditions that are based on those assumed in the univariate case.

CONDITIONS:

C.1. \( \int_0^\tau \alpha_k^2(s) \, ds < \infty \),

C.2. For a neighborhood \( \mathcal{D} \) around the true value for the parameter vector,

\[
\sup_{\delta \in \mathcal{D}} \sup_{t \in [0, \tau]} |S_k^{(j)}(\delta, t) - s_k^{(j)}(\delta, t)| \xrightarrow{\mathcal{P}} 0,
\]

for \( j = 1, \ldots, 5 \), with \( s_k^{(0)} \) bounded away from zero on \( \mathcal{D} \times [0, \tau] \) and \( s_k^{(j)} \) continuous functions of \( \delta \) on \( \mathcal{D} \), \( j = 1, \ldots, 5 \),

C.3. \( \exists \gamma > 0 \) such that \( n^{-1/2} \max_{1 \leq i \leq n, \, t \in [0, \tau]} |X_i|Y_{ki} I\{\beta_k^2 X_i > -\gamma |X_i|\} \xrightarrow{\mathcal{P}} 0 \), and
C.4. The symmetric matrix $\Sigma_k(t)$ with elements given by

$$(\Sigma_k(t))_{11} = \int_0^t \left( s_k^{(2)}(\delta^\circ, u) - \frac{s_k^{(1)}(\delta^\circ, u)^2}{s_k^{(0)}(\delta^\circ, u)} \right) \alpha_k^2(u) \, du$$

$$(\Sigma_k(t))_{22} = \int_0^t \left( s_k^{(4)}(\delta^\circ, u) - \frac{s_k^{(3)}(\delta^\circ, u)^2}{s_k^{(0)}(\delta^\circ, u)} \right) \alpha_k^2(u) \, du$$

$$(\Sigma_k(t))_{12} = \int_0^t \left( s_k^{(5)}(\delta^\circ, u) - \frac{s_k^{(1)}(\delta^\circ, u)s_k^{(3)}(\delta^\circ, u)}{s_k^{(0)}(\delta^\circ, u)} \right) \alpha_k^2(u) \, du$$

is positive definite when $t = \tau$.

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

**Theorem 3.1** For a bivariate counting process with intensity process defined by (3.4) such that conditions C.1–C.3 are true, the stochastic process $n^{-1/2}U$, with $U$ defined in (3.11)-(3.12) converges in distribution to a continuous Gaussian martingale $W$ with covariance function given by

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

where the matrices $\Sigma_k(t)$ are defined in condition C.4.

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

**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 (3.23)-(3.25) converge to deterministic functions and (ii) the predictable variation processes converge to continuous functions as $n \to \infty$. 

63
Considering the predictable processes (3.23)-(3.25), and conditions (C.1)-(C.2) we can interchange the limits and integrals, and, hence,

\[ \langle U_{1}^{(k,n)}, U_{1}^{(k,n)} \rangle (t) \overset{P}{\rightarrow} \int_{0}^{t} \left( s_{k}^{(2)}(\delta^{\circ}, u) - \frac{s_{k}^{(1)}(\delta^{\circ}, u)}{s_{k}^{(0)}(\delta^{\circ}, u)} \right) \alpha_{k}^{\circ}(u) \, du \]

\[ \langle U_{2}^{(k,n)}, U_{2}^{(k,n)} \rangle (t) \overset{P}{\rightarrow} \int_{0}^{t} \left( s_{k}^{(4)}(\delta^{\circ}, u) - \frac{s_{k}^{(3)}(\delta^{\circ}, u)}{s_{k}^{(0)}(\delta^{\circ}, u)} \right) \alpha_{k}^{\circ}(u) \, du \]

\[ \langle U_{i}^{(k,n)}, U_{2}^{(k,n)} \rangle (t) \overset{P}{\rightarrow} \int_{0}^{t} \left( s_{k}^{(5)}(\delta^{\circ}, u) - \frac{s_{k}^{(1)}(\delta^{\circ}, u) s_{k}^{(3)}(\delta^{\circ}, u)}{s_{k}^{(0)}(\delta^{\circ}, u)} \right) \alpha_{k}^{\circ}(u) \, du \]

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

\[ H_{1i}(t, k) = X_i - \frac{S_{k}^{(1)}(\delta^{\circ}, t)}{S_{k}^{(0)}(\delta^{\circ}, t)} \]

(3.26)

and

\[ H_{2i}(t, k) = \frac{Y_{1i} Y_{2i}}{w_{i}(\theta_{k}^{\circ}, t)} - \frac{S_{k}^{(3)}(\delta^{\circ}, t)}{S_{k}^{(0)}(\delta^{\circ}, t)}. \]

(3.27)

Then, let \( H_{hi}^{(n)}(t, k) = n^{-1/2} H_{hi}(t, k) \) for \( h = 1, 2 \) and let the martingales containing the jumps of (3.14) and (3.15) be defined by

\[ \epsilon U_{1}^{(k,n)}(t, \delta^{\circ}) = \sum_{i=1}^{n} \int_{0}^{t} H_{1i}^{(n)}(u, k) \mathbb{I}\{|H_{1i}^{(n)}(u, k)| > \epsilon\} \, dM_{ki}(u), \]

(3.28)

and

\[ \epsilon U_{2}^{(k,n)}(t, \delta^{\circ}) = \sum_{i=1}^{n} \int_{0}^{t} H_{2i}^{(n)}(u, k) \mathbb{I}\{|H_{2i}^{(n)}(u, k)| > \epsilon\} \, dM_{ki}(u). \]

(3.29)

So, to prove that the predictable variation processes converge to continuous functions is equivalent to prove that \( \langle \epsilon U_{1}^{(k,n)}, \epsilon U_{2}^{(k,n)} \rangle \overset{P}{\rightarrow} 0 \), as \( n \to \infty \).

Considering \( \epsilon U_{1}^{(k,n)} \) we have

\[ \langle \epsilon U_{1}^{(k,n)}, \epsilon U_{1}^{(k,n)} \rangle (t) = \sum_{i=1}^{n} \int [H_{1i}^{(n)}(u, k)]^2 \mathbb{I}\{|H_{1i}^{(n)}(u, k)| > \epsilon\} \, d\lambda_{ki}(u) \]

\[ = \sum_{i=1}^{n} \int \frac{1}{n} \left( X_i(u) - \frac{S_{k}^{(1)}(\delta^{\circ}, u)}{S_{k}^{(0)}(\delta^{\circ}, u)} \right)^2 \mathbb{I}\{n^{-1/2}|X_i(u) - \frac{S_{k}^{(1)}(\delta^{\circ}, u)}{S_{k}^{(0)}(\delta^{\circ}, u)}| > \epsilon\} \]

\[ \times w_{i}(u; \theta_{k}^{\circ}) e^{\theta_{k}^{\circ} X_i(u)} \alpha_{k}^{\circ}(u) \, du \]

(3.30)
Considering the inequality (as presented in Andersen and Gill (1982))

\[ |a - b|^2 \mathbb{I}\{|a - b| > \epsilon\} \leq 4|a|^2 \mathbb{I}\{|a| > \epsilon/2\} + 4|b|^2 \mathbb{I}\{|b| > \epsilon/2\}, \]

we have

\[
\langle \epsilon U_1^{(k,n)}, \epsilon U_1^{(k,n)} \rangle \\
\leq \sum_{i=1}^{n} \int \frac{1}{n} X_i(u) o \mathbb{I}\{n^{-1/2} |X_i(u)| > \epsilon\} w_i(u; \theta_k^o) \alpha_k^o(u) \, du \\
+ \sum_{i=1}^{n} \int \frac{1}{n} \left( \frac{S_k^{(1)}(\delta^o, u)}{S_k^{(0)}(\delta^o, u)} \right)^2 \mathbb{I}\{n^{-1/2} \frac{S_k^{(1)}(\delta^o, u)}{S_k^{(0)}(\delta^o, u)} > \epsilon\} w_i(u; \delta_k^o) \alpha_k^o(u) \, du. \tag{3.31}
\]

Note that (3.32) is equal to

\[
4 \int \left( \frac{S_k^{(1)}(\delta^o, u)}{S_k^{(0)}(\delta^o, u)} \right)^2 \mathbb{I}\{n^{-1/2} \frac{S_k^{(1)}(\delta^o, u)}{S_k^{(0)}(\delta^o, u)} > \epsilon\} S_k^{(0)}(\delta^o, u) \alpha_k^o(u) \, du. \tag{3.32}
\]

As a consequence of assumptions (C.1) and (C.2), we have that \( S_k^{(1)}(\delta^o, t) \overset{P}{\rightarrow} s_k^{(1)}(\delta^o, t) \) and since \( s_k^{(1)}(\delta^o, t) \) is bounded and \( s_k^{(0)}(\delta^o, t) \) is bounded away from zero, it follows that, \( \forall \epsilon' > 0 \), fixed, there exists \( n \) sufficiently large such that

\[
P\{w: n^{-1/2} \frac{S_k^{(1)}(\delta^o, t)}{S_k^{(0)}(\delta^o, t)} > \epsilon\} \leq \epsilon'.
\]

Since \( \epsilon' \) is arbitrary we may conclude that (3.32) converges to zero in probability.

Regarding (3.31) we have, for a given \( \gamma > 0 \),

\[
\sum_{i=1}^{n} \int \frac{1}{n} X_i^2(u) \mathbb{I}\{n^{-1/2} |X_i(u)| > \epsilon\} w_i(u; \theta_k^o) \alpha_k^o(u) \, du \\
= \sum_{i=1}^{n} \int \frac{1}{n} X_i^2(u) \mathbb{I}\{n^{-1/2} |X_i(u)| > \epsilon; \beta_k^o X_i(u) \leq -\gamma |X_i(u)|\} e^{\beta_k^o X_i(u)} w_i(u; \theta_k^o) \alpha_k^o(u) \, du \tag{3.33}
\]

\[
+ \sum_{i=1}^{n} \int \frac{1}{n} X_i^2(u) \mathbb{I}\{n^{-1/2} |X_i(u)| > \epsilon; \beta_k^o X_i(u) > -\gamma |X_i(u)|\} e^{\beta_k^o X_i(u)} w_i(u; \theta_k^o) \alpha_k^o(u) \, du \tag{3.34}
\]
The term (3.33) is bounded from above by
\begin{align*}
\sum_{i=1}^{n} \int \frac{1}{n} X_i(u)^2 \mathbb{I}\{|X_i(u)| > \varepsilon; \beta_k^2 X_i(u) \leq -\gamma|X_i(u)|\} e^{-\gamma|X_i(u)|} w_i(u; \theta_k^0) \alpha_k^0(u) \, du \\
\leq \sum_{i=1}^{n} \int \frac{1}{n} X_i^2(u) \mathbb{I}\{n^{-1/2}|X_i(u)| > \varepsilon\} e^{-\gamma|X_i(u)|} w_i(u; \theta_k^0) \alpha_k^0(u) \, du.
\end{align*}
(3.35)

Since \( Y_{k1} + \theta_k^0 Y_{11} Y_{21} \leq Y_{k1} + |\theta_k^0| Y_{11} Y_{21} \leq 1 + |\theta_k^0| \), it follows the expression in (3.35) is bounded above by
\begin{align*}
(1 + |\theta_k^0|) \sum_{i=1}^{n} \int \frac{1}{n} X_i^2(u) \mathbb{I}\{n^{-1/2}|X_i(u)| > \varepsilon\} e^{-\gamma|X_i(u)|} \alpha_k^0(u) \, du \\
\leq (1 + |\theta_k^0|) \eta \int \alpha_k^0(u) \, du
\end{align*}
(3.36)

where the last inequality results 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 \( \eta \) arbitrarily small, we may conclude that (3.33) converges to zero in probability.

In virtue of assumption (C.3), the same conclusion is true for expression (3.34). In order to make this clear we note that such an expression is smaller or equal than
\begin{align*}
4 \sum_{i=1}^{n} \int \frac{1}{n} X_i^2(u) \mathbb{I}\{\beta_k^0 X_i(u) > -\gamma|X_i(u)|\} e^{\beta_k^0 X_i(u)} Y_{k1}(u) \alpha_k^0(u) \, du \\
+ 4|\theta_k^0| \sum_{i=1}^{n} \int \frac{1}{n} X_i^2(u) \mathbb{I}\{\beta_k^0 X_i(u) > -\gamma|X_i(u)|\} e^{\beta_k^0 X_i(u)} Y_{11}(u) Y_{21}(u) \alpha_k^0(u) \, du \\
\leq 4(1 + |\theta_k^0|) \sum_{i=1}^{n} \int \frac{1}{n} X_i^2(u) \mathbb{I}\{\beta_k^0 X_i(u) > -\gamma|X_i(u)|\} e^{\beta_k^0 X_i(u)} Y_{k1}(u) \alpha_k^0(u) \, du,
\end{align*}
(3.37)

and hence the last expression converges to zero in probability.
As for process \( U^{(k,n)}_2 \), we have

\[
\langle \epsilon U^{(k,n)}_2, \epsilon U^{(k,n)}_2 \rangle (t) \\
= \sum_{i=1}^{n} \int [H_{2i}^{(n)}(u, k)]^2 \mathbb{I}\{|H_{2i}^{(n)}(u, k)| > \epsilon\} \, d\lambda_k(u) \\
= \sum_{i=1}^{n} \int \frac{1}{n} \left( \frac{Y_{1i}(u)Y_{2i}(u)}{w_i(u; \theta_k)} - \frac{S^{(3)}_k(\delta^o, u)}{S^{(0)}_k(\delta^o, u)} \right)^2 \mathbb{I}\{|H_{2i}^{(n)}(u, k)| > \epsilon\} \, w_i(u; \theta_k^o) \alpha_k^o(u) \, du \\
\leq 4 \sum_{i=1}^{n} \int \frac{1}{n} \left( \frac{Y_{1i}(u)Y_{2i}(u)}{w_i(u; \theta_k^o)} \right)^2 \mathbb{I}\{n^{-1/2} \frac{Y_{1i}(u)Y_{2i}(u)}{w_i(u; \theta_k^o)} > \epsilon\} \, w_i(u; \theta_k^o) e^{\beta_k X_i(u)} \alpha_k^o(u) \, du \\
+ 4 \int \left( \frac{S^{(3)}_k(\delta^o, u)}{S^{(0)}_k(\delta^o, u)} \right)^2 \mathbb{I}\{n^{-1/2} \frac{S^{(3)}_k(\delta^o, u)}{S^{(0)}_k(\delta^o, u)} > \epsilon\} \, S^{(0)}_k(\delta^o, u) \alpha_k^o(u) \, du
\]  

(3.38)

Using the same arguments as in the previous case we recall that by assumption (C.2) \( S^{(3)}_k \) converges in probability to \( s^{(3)}_k \) and \( S^{(0)}_k \) converges in probability to \( s^{(0)}_k \). Based on this fact in addition to assumption (C.1) we claim that, for all \( \epsilon' > 0 \) there exists a \( n_0 \) sufficiently large such that, \( \forall n \geq n_0 \),

\[
P\{w: n^{-1/2} \frac{S^{(3)}_k(\delta^o, t)}{S^{(0)}_k(\delta^o, u)} > \epsilon\} \leq \epsilon',
\]

and, hence, expression (3.39) converges in probability to zero.

Finally, with respect to (3.38) we note that

\[
\sum_{i=1}^{n} \int \frac{1}{n} \left( \frac{Y_{1i}(u)Y_{2i}(u)}{w_i(u; \theta_k^o)} \right)^2 \mathbb{I}\{n^{-1/2} \frac{Y_{1i}(u)Y_{2i}(u)}{w_i(u; \theta_k^o)} > \epsilon\} \, w_i(u; \theta_k^o) e^{\beta_k X_i(u)} \alpha_k^o(u) \, du \\
= \sum_{i=1}^{n} \int \frac{1}{n} \frac{Y_{1i}(u)Y_{2i}(u)}{w_i(u; \theta_k^o)} \mathbb{I}\{n^{-1/2} \frac{Y_{1i}(u)Y_{2i}(u)}{w_i(u; \theta_k^o)} > \epsilon\} e^{\beta_k X_i(u)} \alpha_k^o(u) \, du \\
\leq \frac{1}{1 + \theta_k^o} \sum_{i=1}^{n} \int \frac{1}{n} \mathbb{I}\{n^{-1/2} \frac{Y_{1i}(u)Y_{2i}(u)}{w_i(u; \theta_k^o)} > \epsilon\} e^{\beta_k X_i(u)} \alpha_k^o(u) \, du.
\]  

(3.40)

Also, we note that, for \( i = 1, \ldots, n \),

\[
\frac{Y_{1i}Y_{2i}}{Y_{ki} + \theta_k^o Y_{1i}Y_{2i}} = \begin{cases} 
1/(1 + \theta_k^o), & \text{if } Y_{1i} = Y_{2i} = 1; \\
0, & \text{otherwise,}
\end{cases}
\]

67
so that, since $\theta_k^o > -1$,

$$\frac{Y_{1i} Y_{2i}}{Y_{ki} + \theta_k^o Y_{1i} Y_{2i}} \leq \frac{1}{1 + \theta_k^o}$$

and, hence,

$$n^{-1/2} \left| \frac{Y_{1i} Y_{2i}}{Y_{ki} + \theta_k^o Y_{1i} Y_{2i}} \right| > \epsilon \rightarrow n^{-1/2} \left| \frac{1}{1 + \theta_k^o} \right| > \epsilon$$

$$\Rightarrow \mathbb{I}\left\{(1 + \theta_k^o)^{-1} > n^{1/2} \epsilon\right\} \geq \mathbb{I}\left\{\left| \frac{Y_{1i} Y_{2i}}{Y_{ki} + \theta_k^o Y_{1i} Y_{2i}} \right| > n^{1/2} \epsilon\right\}.$$

Therefore, expression (3.40) is bounded from above by

$$\frac{1}{1 + \theta_k^o} \sum_{i=1}^{n} \int \frac{1}{n} e^{\theta_k^o X_i(u)} \mathbb{I}\left\{\frac{1}{1 + \theta_k^o} > n^{1/2} \epsilon\right\} \alpha_k^o(u) \, du.$$

Since $\theta_k^o$ is a fixed value, it is always possible to get $n$ sufficiently large such that the expression above is zero, what implies that (3.40) converges to zero in probability.

The results above complete the conditions of the Rebolledo’s central limit theorem and, hence, one can claim that $n^{-1/2} U$ converges to a multivariate Gaussian process, with covariance matrix given by the limit of the predictable (co)variation processes.

In order to illustrate the conditions imposed to the functions $S_k^{(j)}$, let us consider the following example.

**Example 3.2 [Example (3.1) continued]:** In the case of binary covariates, the existence of the functions $s_k^{(j)}$ is easier to verify. Initially we note that for a 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 condition (C.2).
Thus, for $S^{(0)}_k(\delta^o, t) = (1/n) \sum_{i=1}^n w_i(t; \theta^o_k) e^{\beta_k X_i}$ we note that

$$E\{w_i(t; \theta^o_k) e^{\beta_k X_i}\} = E\left[(Y_{ki} + \theta^o_k Y_{1i} Y_{2i}) e^{\beta_k X_i}\right] = E\left[E\left((Y_{ki} + \theta^o_k Y_{1i} Y_{2i}) e^{\beta_k X_i} \mid X_i\right)\right] = \pi_0 \left(S^{pl}_k(t) + \theta^o_k S^{tr}_k(t)\right) + \pi_1 e^{\beta_k} \left(S^{pl}_k(t) + \theta^o_k S^{tr}_k(t)\right)$$

(3.41)

where $\pi_0 = 1 - \pi_1 = P(X_i = 0)$ is the probability that a particular individual will be assigned to placebo, $S^{pl}_k(t) = P(T_{ki} \geq t \mid X_i = 0)$ is the marginal survival function related to component $k$, $S^{tr}_k(t) = P(T_{1i} \geq t; T_{2i} \geq t \mid X_i = 0)$ is the joint survival function for the placebo group and $S^{tr}_k(t), S^{pl}_k(t)$ are similarly defined for the treatment group. Since the survival functions in the above expression are non-increasing functions, the uniform convergence (C.2) follows. Analogously, for $j = 1, 2$,

$$E\left[X_i^j (Y_{ki} + \theta^o_k Y_{1i} Y_{2i}) e^{\beta_k X_i}\right] = \pi_1 e^{\beta_k} \left(S^{pl}_k(t) + \theta^o_k S^{tr}_k(t)\right)$$

(3.42)

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

For $S^{(3)}_k$ we have

$$E\{Y_{1i} Y_{2i} e^{\beta_k X_i}\} = \pi_0 S^{pl}_{12}(t) + \pi_1 e^{\beta_k} S^{tr}_{12}(t)$$

(3.43)

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

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

$$E\left[\frac{Y_{1i} Y_{2i}}{Y_{ki} + \theta^o_k Y_{1i} Y_{2i}} e^{\beta_k X_i}\right] = \frac{1}{1 + \theta^o_k} \left(\pi_0 S^{pl}_{12}(t) + \pi_1 e^{\beta_k} S^{tr}_{12}(t)\right)$$

(3.44)

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

For $S^{(5)}_k$,

$$E[X_i Y_{1i} Y_{2i} e^{\beta_k X_i}] = \pi_1 e^{\beta_k} S^{tr}_{12}(t)$$

(3.45)
a monotone function. It follows then, for \( j = 1, \ldots, 5 \) the uniform condition (C.2) is true, and, as \( n \to \infty \),

\[
\langle U_1^{(k,n)}, U_1^{(k,n)} \rangle(t) \overset{P}{\to} \\
\pi_0 \pi_1 e^{\beta^*_k} \int_0^t \frac{[S_k^{tr}(u) + \theta_k^0 S_{12}^{tr}(u)][S_k^{pl}(u) + \theta_k^0 S_{12}^{pl}(u)]}{\pi_0 \pi_1 e^{\beta^*_k} [S_k^{tr}(u) + \theta_k^0 S_{12}^{tr}(u)] + \pi_1 e^{\beta^*_k} [S_k^{pl}(u) + \theta_k^0 S_{12}^{pl}(u)]} \alpha_k(u) \, du,
\]

\[
\langle U_2^{(k,n)}, U_2^{(k,n)} \rangle(t) \overset{P}{\to} \\
\int_0^t \frac{[\pi_0 S_{12}^{pl}(u) + e^{\beta^*_k} \pi_1 S_{12}^{tr}(u)][\pi_0 (S_k^{pl}(u) - S_k^{pl}(u)) + \pi_1 e^{\beta^*_k} (S_k^{tr}(u) - S_k^{tr}(u))]}{\pi_0 S_k^{tr}(u) + \pi_1 e^{\beta^*_k} [S_k^{tr}(u) + \theta_k^0 S_{12}^{tr}(u)]} \alpha_k(u) \, du,
\]

\[
\langle U_1^{(k,n)}, U_2^{(k,n)} \rangle(t) \overset{P}{\to} \\
e^{\beta^*_k} \pi_0 \pi_1 \int_0^t \frac{S_k^{tr}(u) S_k^{pl}(u) - S_k^{tr}(u) S_{12}^{pl}(u)}{\pi_0 S_k^{pl}(u) + \pi_1 e^{\beta^*_k} [S_k^{tr}(u) + \theta_k^0 S_{12}^{tr}(u)]} \alpha_k(u) \, du.
\]

\[
\square
\]

Theorem 3.1 will be considered when proving the asymptotic distribution for the MPLE below. We also will make use of the information matrix computed on subsection 3.2.

The next theorem refers to the asymptotic distribution for the MPLE and it is based on the approach used for the proof of asymptotic distribution for the usual maximum likelihood estimator due to LeCam (1956). [For more details we may refer to the presentation given in theorem 5.2.1 of Sen and Singer (1993).]

**Theorem 3.2** Let \( \hat{\delta} \) be a value that maximizes the partial likelihood (3.10) and suppose that conditions C.1–C.4 hold. Then, if \( \delta^0 \) is the true value for the parameter \( \delta \),

\[
n^{1/2}(\hat{\delta} - \delta^0) \overset{D}{\to} \mathcal{N}(0, \Sigma^{-1}).
\]

70
Proof: Let $u$ be a 4-vector such that $\|u\| \leq K$, $0 < K < \infty$ and define

$$
\lambda_n(u) = \log L_n(\delta^o + n^{-1/2}u) - \log L_n(\delta^o) \\
= \sum_{i=1}^{n} \{ \log L_i(\delta^o + n^{-1/2}u) - \log L_i(\delta^o) \}.
$$

(3.46)

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

$$
\log L_i(\delta^o + n^{-1/2}u) = \log L_i(\delta^o) + n^{-1/2} \left[ \frac{\partial \log L_i(\delta^o)}{\partial \delta} \right]^T u \\
+ \frac{1}{2n} u^T \frac{\partial^2 \log L_i(\delta^o)}{\partial \delta \partial \delta^T} u,
$$

(3.47)

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

$$
\log L_i(\delta^o + n^{-1/2}u) - \log L_i(\delta^o) = n^{-1/2} \left[ \frac{\partial \log L_i(\delta^o)}{\partial \delta} \right]^T u \\
+ \frac{1}{2n} u^T \frac{\partial^2 \log L_i(\delta^o)}{\partial \delta \partial \delta^T} u.
$$

(3.48)

Considering

$$
Z_n(u) = \frac{1}{n} \sum_{i=1}^{n} \left[ \frac{\partial^2 \log L_i(\delta^*)}{\partial \delta \partial \delta^T} \right] - \frac{\partial^2 \log L_i(\delta^o)}{\partial \delta \partial \delta^T},
$$

we write

$$
\lambda_n(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ \frac{\partial \log L_i(\delta^o)}{\partial \delta} \right]^T u + \frac{1}{2n} u^T \sum_{i=1}^{n} \frac{\partial^2 \log L_i(\delta^o)}{\partial \delta \partial \delta^T} u + \frac{1}{2} u^T Z_n u \\
= \frac{1}{\sqrt{n}} [U(\delta^o)]^T u - \frac{1}{2n} u^T J(\delta^o) u + \frac{1}{2} u^T Z_n u,
$$

(3.49)

where $J(\cdot)$ was defined in (3.19)

If $\|Z_n(u)\|$ converges in probability to zero, uniformly in $u$, then, maximizing (3.49) with respect to $u$ will correspond closely to obtain a MPLE for $\delta^o$, 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 $J$ (remembering that such matrix may
be expressed as a partitioned matrix, for \( k = 1, 2, \)

\[
(J^{(k)})_{11} = -\sum_{i=1}^{n} \frac{\partial^2 \log L_i(\delta)}{\partial \beta_k^2} = \sum_{i=1}^{n} \int_{0}^{\tau} \left\{ \frac{S_k^{(2)}(\delta, x)}{S_k^{(0)}(\delta, x)} - \left( \frac{S_k^{(1)}(\delta, x)}{S_k^{(0)}(\delta, x)} \right)^2 \right\} dN_{ki}(x)
\]

\[
= \sum_{i=1}^{n} \int_{0}^{\tau} V_{11}^{(k)}(\delta, x) \, dN_{ki}(x),
\]

\[
(J^{(k)})_{12} = (J^{(k)})_{21} = -\sum_{i=1}^{n} \frac{\partial^2 \log L_i(\delta)}{\partial \beta_k \partial \theta_k} = \sum_{i=1}^{n} \int_{0}^{\tau} \left\{ \frac{S_k^{(3)}(\delta, x)}{S_k^{(0)}(\delta, x)} - \frac{S_k^{(3)}(\delta, x)S_k^{(1)}(\delta, x)}{(S_k^{(0)}(\delta, x))^2} \right\} dN_{ki}(x)
\]

\[
= \sum_{i=1}^{n} \int_{0}^{\tau} V_{12}^{(k)}(\delta, x) \, dN_{ki}(x),
\]

\[
(J^{(k)})_{22} = -\sum_{i=1}^{n} \frac{\partial^2 \log L_i(\delta)}{\partial \theta_k^2} = \sum_{i=1}^{n} \int_{0}^{\tau} \left\{ \left( \frac{Y_{1i}(x)Y_{2i}(x)}{w_i(x; \theta_k)} \right)^2 - \left( \frac{S_k^{(3)}(\delta, x)}{S_k^{(0)}(\delta, x)} \right)^2 \right\} dN_{ki}(x).
\]

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

\[
|(Z_n^{(k)}(u))_{11}| = \left| \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\tau} V_{11}^{(k)}(\delta^*, x) \, dN_{ki}(x) - \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\tau} V_{11}^{(k)}(\delta^0, x) \, dN_{ki}(x) \right|
\]

\[
\leq \left| \sup_{\{h: \|h\| \leq \|u\|/\sqrt{n}\}} \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\tau} V_{11}^{(k)}(\delta^0 + h, x) - V_{11}^{(k)}(\delta^0, x) \, dN_{ki}(x) \right|
\]

\[
+ \left| \sup_{\{h: \|h\| \leq \|u\|/\sqrt{n}\}} \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\tau} V_{11}^{(k)}(\delta^0, x) - v_{11}^{(k)}(\delta^0, x) \, dN_{ki}(x) \right|
\]

\[
+ \left| \sup_{\{h: \|h\| \leq \|u\|/\sqrt{n}\}} \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\tau} v_{11}^{(k)}(\delta^0 + h, x) - v_{11}^{(k)}(\delta^0, x) \, dN_{ki}(x) \right|
\]

72
\[ \begin{align*}
&= \left| \sup_{\{|h| < \|x\|/\sqrt{n}\}} \int_0^r v_{11}^{(k)}(\delta^o + h, x) - v_{11}^{(k)}(\delta^o + h, x) \, \frac{dN_k(x)}{n} \right| \\
&\quad + \left| \int_0^r v_{11}^{(k)}(\delta^o, x) - v_{11}^{(k)}(\delta^o, x) \, \frac{dN_k(x)}{n} \right| \\
&\quad + \left| \sup_{\{|h| < \|x\|/\sqrt{n}\}} \int_0^r v_{11}^{(k)}(\delta^o + h, x) - v_{11}^{(k)}(\delta^o, x) \, \frac{dN_k(x)}{n} \right|. \\
&\quad \quad \quad \quad \quad \quad \quad \text{(3.50)} \\
&\quad \quad \quad \quad \quad \quad \quad \text{(3.51)} \\
&\quad \quad \quad \quad \quad \quad \quad \text{(3.52)}
\end{align*} \]

Using the Lenglart's inequality, we have that, for all \( \rho > 0 \) and \( \eta > 0 \),
\[
\begin{align*}
\mathbb{P}\left\{ \frac{N_k(\tau)}{n} \geq \eta \right\} &\leq \frac{\rho}{\eta} + \mathbb{P}\left\{ \frac{1}{n} \int_0^r \lambda_k(t) \, dt \geq \rho \right\} \\
&= \frac{\rho}{\eta} + \mathbb{P}\left\{ \frac{1}{n} \int_0^r \sum_{j=1}^n w_j(t; \theta_k^o) e^{\theta_k^o X_j(t)} \alpha_k^o(t) \, dt \geq \rho \right\} \\
&= \frac{\rho}{\eta} + \mathbb{P}\left\{ \int_0^r \left[ \frac{1}{n} \sum_{j=1}^n w_j(t; \theta_k^o) e^{\theta_k^o X_j(t)} \right] \alpha_k^o(t) \, dt \geq \rho \right\} \\
&= \frac{\rho}{\eta} + \mathbb{P}\left\{ \int_0^r S_k^{(0)}(\delta^o, t) \alpha_k^o(t) \, dt \geq \rho \right\}.
\end{align*}
\]

Taking \( \rho > \int_0^r S_k^{(0)}(\delta^o, t) \alpha_k^o(t) \, dt \) and considering assumption (C.2) then, as \( n \uparrow \infty \) it follows that \( \mathbb{P}\{\int_0^r S_k^{(0)}(\delta^o, t) \alpha_k^o(t) \, dt > \rho \} \to 0 \) and, hence,
\[
\lim_{\eta \to 0} \lim_{n \to \infty} \mathbb{P}\left[ \frac{N_k(\tau)}{n} > \eta \right] = 0. \quad (3.53)
\]

Under condition (C.2) it follows that, by the boundedness conditions on \( s_k^{(j)} \),
\[
\sup_{t \in [0, \tau]} |v_{11}^{(k)}(\delta, t) - v_{11}^{(k)}(\delta, t)| \xrightarrow{p} 0. \quad (3.54)
\]

If we consider \( n \) sufficiently large such that \( \delta^o + h \in \mathcal{D} \), it follows that expressions (3.50) and (3.51) will converge in probability to zero, as \( n \to \infty \), in virtue of (3.54) and Condition (C.2). Also, as a consequence of (3.54), as \( n \to \infty \), expression (3.52) will converge to zero, in probability. By the continuity of the elements of \( v_{11}^{(k)}(\delta, t) \) in a neighborhood of \( \delta^o \) we also conclude that (3.52) converges to zero in probability.
Hence,

\[ \sup_{\{u: \|u\| \leq K\}} |(Z_n^{(k)}(u))_{11}| \xrightarrow{P} 0. \]

Similar argumentation lead us to conclude that

\[ |(Z_n^{(k)}(u))_{12}| = |(Z_n^{(k)}(u))_{21}| \xrightarrow{P} 0, \quad \text{uniformly in } u. \]

To deal with \((Z_n^{(k)})_{22}\), define \(M_k(\delta, x) = S_k^{(3)}(\delta, x)/S_k^{(0)}(\delta, x)\). Then we may write

\[
(Z_n^{(k)}(u))_{22} = \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau \left\{ [M_k(\delta^*, x)]^2 - \left( \frac{Y_{1i}(x)Y_{2i}(x)}{w_i(x; \theta_k^*)} \right)^2 \right\} dN_{ki}(x)
- \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau \left\{ [M_k(\delta^o, x)]^2 - \left( \frac{Y_{1i}(x)Y_{2i}(x)}{w_i(x; \theta_k^o)} \right)^2 \right\} dN_{ki}(x),
\]

and, hence,

\[
|(Z_n^{(k)}(u))_{22}|
\leq \sup_{\{h: |h| \leq \|u\|/\sqrt{n}\}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau \left\{ [M_k(\delta^o + h, x)]^2 - \left( \frac{Y_{1i}Y_{2i}}{w_i(x; \theta_k^o + h)} \right)^2 \right\} dN_{ki}(x) \right\}
- \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau \left\{ [M_k(\delta^o, x)]^2 - \left( \frac{Y_{1i}Y_{2i}}{w_i(x; \theta_k^o)} \right)^2 \right\} dN_{ki}(x) \right\}
\leq \sup_{\{h: |h| \leq \|u\|/\sqrt{n}\}} \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau \left\{ (M_k(\delta^o + h, x))^2 - (M_k(\delta^o, x))^2 \right\} dN_{ki}(x) \right\]
+ \sup_{\{h: |h| \leq \|u\|/\sqrt{n}\}} \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau \left\{ \left( \frac{Y_{1i}Y_{2i}}{w_i(x; \theta_k^o + h)} \right)^2 - \left( \frac{Y_{1i}Y_{2i}}{w_i(x; \theta_k^o)} \right)^2 \right\} dN_{ki}(x). \quad (3.55)
\]

The first term in the r.h.s. of (3.55) is bounded from above by

\[
\left| \sup_{\{h: |h| \leq \|u\|/\sqrt{n}\}} \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau \left\{ (M_k(\delta^o + h, x))^2 - (m_k(\delta^o + h, x))^2 \right\} dN_{ki}(x) \right| \quad (3.56)
\]

\[
+ \left| \sup_{\{h: |h| \leq \|u\|/\sqrt{n}\}} \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau \left\{ (M_k(\delta^o, x))^2 - (m_k(\delta^o, x))^2 \right\} dN_{ki}(x) \right| \quad (3.57)
\]

\[
+ \left| \sup_{\{h: |h| \leq \|u\|/\sqrt{n}\}} \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau \left\{ (m_k(\delta^o + h, x))^2 - (m_k(\delta^o, x))^2 \right\} dN_{ki}(x) \right| \quad (3.58)
\]
where \( m_k = s_k^{(3)}/s_k^{(0)} \). By condition (C.2) and the boundedness conditions imposed on the quantities involved, (3.56), (3.57), and (3.58) converge to zero in probability as \( n \to \infty \).

Remains to show that the second term in the r.h.s. of (3.55) converges to zero. To do so, we note that such expression is not larger than

\[
\sup_{\{h: |h| \leq ||u||/\sqrt{n}\}} \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{\tau} \left| \frac{Y_{1i}(x)Y_{2i}(x)}{w_i(x; \theta_k^0 + h)} - \frac{Y_{1i}(x)Y_{2i}(x)}{w_i(x; \theta_k^0)} \right|^2 dN_{ki}(x)
\]

\[
\leq \sup_{\{h: |h| \leq ||u||/\sqrt{n}\}} \left| \frac{1}{(1 + \theta_k^0 + h)^2} - \frac{1}{(1 + \theta_k^0)^2} \right| \int_{0}^{\tau} \frac{dN_{ki}(x)}{n}.
\]

Now we note that, for all \( \eta > 0 \) it is possible to get \( \varepsilon > 0 \) such that, for all \( h < \varepsilon \),

\[
\left| \frac{1}{(1 + \theta_k^0 + h)^2} - \frac{1}{(1 + \theta_k^0)^2} \right| < \eta.
\]

Taking \( \eta \downarrow 0 \) in addition to (3.53) it follows that (3.55) will converge in probability to zero.

Hence, we have shown that

\[
\sup_{\{||u|| \in [-K,K]\}} \|Z_n(u)\| \xrightarrow{P} 0,
\]

(3.59)

i.e., \( Z_n \) converges in probability to zero, uniformly in \( u \).

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

\[
n^{-1}J(\delta^o) \xrightarrow{P} \Sigma(\tau),
\]

(3.60)

for \( \Sigma \) defined in theorem 3.1.

Using (3.59) we may rewrite expression (4.49) as

\[
\lambda_n(u) = \frac{1}{\sqrt{n}}[U(\delta^o)]^T u - \frac{1}{2n} u^T J(\delta^o) u + o_p(1)
\]

\[
= \frac{1}{\sqrt{n}}[U(\delta^o)]^T u - \frac{1}{2} u^T \Sigma u + \frac{1}{2} u^T (\Sigma - \frac{1}{n} J(\delta^o)) u + o_p(1)
\]

75
that, in virtue of (3.60), may be written, uniformly in the set \( \{ u : \| u \| \leq K \} \),

\[
\lambda_n(u) = \frac{1}{\sqrt{n}}[U(\delta^o)]^T u - \frac{1}{2} u^T \Sigma u + o_p(1)
\]

(3.61)

Maximization of (3.61) with respect to \( u \) (ignoring the negligible part for a moment) will give

\[
\hat{u} = \frac{1}{\sqrt{n}} \Sigma^{-1} U(\delta^o),
\]

or more precisely,

\[
\hat{u} = \frac{1}{\sqrt{n}} \Sigma^{-1} U(\delta^o) + o_p(1).
\]

Noting that a maximum on \( \lambda_n \) corresponds to a maximum on the (log) partial likelihood, if \( \hat{\delta} \) is a point of maximum in (3.10), it follows that

\[
\hat{\delta} = \delta^o + n^{-1/2} \hat{u} = \delta^o + n^{-1} \Sigma^{-1} U(\delta^o) + o_p(1),
\]

such that

\[
\sqrt{n}(\hat{\delta} - \delta^o) = n^{-1/2} \Sigma^{-1} U(\delta^o) + o_p(1).
\]

Therefore, by Theorem 3.1, \( n^{-1/2} U \) computed at \( t = \tau \) converges to a multivariate normal distribution with covariance matrix given by \( \Sigma \), and, by the Slutsky's theorem one may conclude that \( \sqrt{n}(\hat{\delta} - \delta^o) \) converges to a multivariate normal distribution, as specified in the theorem, concluding the proof.

Two important consequences arise from the above development, and we enunciate them in the following corollaries.

**Corollary 3.1** *The estimator \( \hat{\delta} \) is a consistent estimator for \( \delta^o \).*

This corollary is a direct consequence of the theorem, i.e., one can immediately verify that \( \| \hat{\delta} - \delta^o \| = o_p(1) \)
Corollary 3.2 The covariance matrix \( \Sigma \) can be consistently estimated by \( n^{-1}J(\hat{\delta}) \).

Corollary 3.2 follows from the assumptions made in theorem 3.2 and from corollary 3.1. In fact, since \( \hat{\delta} \) 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 (C.2) will be true, and, hence, each element of the matrix \( n^{-1}J(\hat{\delta}) \) will converge to the respective element of \( \Sigma \).

Example 3.3 [Example (3.2) continued]: Since in the case of the example 3.1 the continuity in condition C.2 and condition C.3 are trivial, the asymptotic convergence of \( n^{1/2}(\hat{\delta} - \delta^0) \) is given by a normal distribution with mean zero and covariance matrix given by the inverse of \( \Sigma \) whose elements can be consistently estimated by

\[
(\hat{\Sigma}_k)_{11} = \int_0^T \left\{ \frac{(Y_k^{\text{Tr}} + \hat{\theta}_k Y_{12}^{\text{Tr}}) e^{\hat{\beta}_k}}{(Y_k^{\text{Pl}} + \hat{\theta}_k Y_{12}^{\text{Pl}}) + (Y_k^{\text{Tr}} + \hat{\theta}_k Y_{12}^{\text{Tr}}) e^{\hat{\beta}_k}} \right. \\
 \left. - \frac{(Y_k^{\text{Tr}} + \hat{\theta}_k Y_{12}^{\text{Tr}}) e^{\hat{\beta}_k}}{(Y_k^{\text{Pl}} + \hat{\theta}_k Y_{12}^{\text{Pl}}) + (Y_k^{\text{Tr}} + \hat{\theta}_k Y_{12}^{\text{Tr}}) e^{\hat{\beta}_k}} \right\}^2 d\hat{N}_k.
\]

\[
(\hat{\Sigma}_k)_{22} = \frac{1}{(1 + \hat{\theta}_k)^2} \int_0^T d\hat{N}_{k_1} - \int_0^T \frac{Y_{12}^{\text{Pl}} + e^{\hat{\beta}_k} Y_{12}^{\text{Tr}}}{(Y_k^{\text{Pl}} + \hat{\theta}_k Y_{12}^{\text{Pl}}) + (Y_k^{\text{Tr}} + \hat{\theta}_k Y_{12}^{\text{Tr}}) e^{\hat{\beta}_k}}^2 d\hat{N}_k.
\]

\[
(\hat{\Sigma}_k)_{12} = (\hat{\Sigma}_k)_{21} = \int_0^T \left\{ \frac{Y_{12}^{\text{Tr}} e^{\hat{\beta}_k}}{(Y_k^{\text{Pl}} + \hat{\theta}_k Y_{12}^{\text{Pl}}) + (Y_k^{\text{Tr}} + \hat{\theta}_k Y_{12}^{\text{Tr}}) e^{\hat{\beta}_k}} \\
\right. \\
\left. - \frac{(Y_k^{\text{Pl}} + Y_{12}^{\text{Tr}} e^{\hat{\beta}_k})(Y_k^{\text{Tr}} + \hat{\theta}_k Y_{12}^{\text{Tr}}) e^{\hat{\beta}_k}}{[(Y_k^{\text{Pl}} + \hat{\theta}_k Y_{12}^{\text{Pl}}) + (Y_k^{\text{Tr}} + \hat{\theta}_k Y_{12}^{\text{Tr}}) e^{\hat{\beta}_k}]^2} \right\} d\hat{N}_k.
\]

where \( Y_k^{\text{Tr}} \) and \( Y_k^{\text{Pl}} \) represent the number of individuals with component \( k \) at risk, at time \( t \), for treatment and placebo groups respectively, \( Y_{12}^{\text{Tr}} \) and \( Y_{12}^{\text{Pl}} \) represent the number of individuals with both components at risk, at time \( t \), for treatment and
placebo groups, and $\tilde{N}_k^{11}$ represents the number of failures of component $k$, at time $t$, divided by $n$, for those individuals with no failure in any component.
CHAPTER 4

The Matrix-Valued Model with Time Dependent Parameters

1 Introduction

In the previous chapter we studied the matrix-valued model assuming that the parameters in the intensity processes were finite-dimensional. In particular, the dependence parameter \( \theta_k \) was assumed to be time independent in order to simplify the theoretical derivations. In this chapter we relax that assumption, allowing \( \theta_k \) to depend on the time \( t \). In addition, we also consider that the coefficients associated with the covariates are time-dependent, extending in this sense the work by Murphy and Sen (1991), based on the method of sieves; therefore this chapter is strongly based on that paper with some modifications in the proofs in order to take into account the larger number of parameters present in our model. Other approaches dealing with time-dependent coefficients have also been discussed in the literature (e.g., Zucker and Karr (1990) using the penalized likelihood) but at this time we will concentrate in the histogram sieves.

The treatment here is more technical than in the previous chapters, therefore we present the material in the following way. In the next section the time-dependent coefficient model is discussed, in the same lines of the Chapter 3. Then, in Section 3
we describe the method of sieves. Following that, the asymptotic properties for
the sieves estimator derived for our model are stated and proved, dealing with the
consistency and weak convergence; several lemmas are needed for the main theorems
and they are proved in details in Section 5.

2 Model specification

Here again we consider the bivariate setting. However, extensions to higher dimen-
sions should not be very different, as long as we assume that higher order interactions
are negligible and accept a more cumbersome notation. As defined in Chapter 2,
we consider that \(n\) individuals are followed and for each one we associate a vector-
valued counting process \(N_i(t) = (N_{1i}(t), N_{2i}(t))^\top, t \in \mathcal{T} = [0, \tau], \tau < \infty\), with
intensity process given by

\[
\lambda_i(t) = \begin{pmatrix}
\alpha_{11}^i(t)Y_{1i}(t) + \alpha_{12}^i(t)Y_{1i}(t)Y_{2i}(t) \\
\alpha_{22}^i(t)Y_{2i}(t) + \alpha_{21}^i(t)Y_{1i}(t)Y_{2i}(t)
\end{pmatrix},
\]

for \(i = 1, \ldots, n\). However, now the functions \(\alpha_{ij}^i(\cdot)\) include time-dependent param-
eters, so that for the \(i\)th individual,

\[
\alpha_{11}^i(t) = \alpha_{11}^o(t)e^{\beta_1(t)X_i(t)} \\
\alpha_{12}^i(t) = \alpha_{12}^o(t)e^{\beta_1(t)X_i(t)} \\
\alpha_{21}^i(t) = \alpha_{21}^o(t)e^{\beta_2(t)X_i(t)} \\
\alpha_{22}^i(t) = \alpha_{22}^o(t)e^{\beta_2(t)X_i(t)}.
\]

We write, as before, \(\alpha_{12}^o(t) = \alpha_{11}^o(t)\theta_1(t)\) and \(\alpha_{21}^o(t) = \alpha_{22}^o(t)\theta_2(t)\), but now the
dependence parameter \(\theta_k\) is also assumed to be a function of \(t\) so that the vector of
intensity processes is written as (with the notation $\alpha_{kk}^o = \alpha_k^o$)

$$
\lambda_i(t) = \begin{pmatrix} 
\alpha_i^1(t)(Y_{1i}(t) + \theta_1(t)Y_{1i}(t)Y_{2i}(t))e^{\beta_1(t)X_i(t)} \\
\alpha_i^2(t)(Y_{2i}(t) + \theta_2(t)Y_{1i}(t)Y_{2i}(t))e^{\beta_2(t)X_i(t)}
\end{pmatrix}
$$

$$
= \begin{pmatrix} 
\alpha_i^1(t)w_i(t; \theta_1)e^{\beta_1(t)X_i(t)} \\
\alpha_i^2(t)w_i(t; \theta_2)e^{\beta_2(t)X_i(t)}
\end{pmatrix}
$$

The likelihood is derived using the same multinomial approach as before, so that

$$
L(\delta) = \prod_{t \geq 0} \prod_{i=1}^{n} \prod_{k=1}^{2} \left( \frac{\lambda_{ki}(t)}{\sum_{j=1}^{n}\lambda_{kj}(t)} \right)^{dN_{ki}(t)}
$$

$$
= \prod_{t \geq 0} \prod_{i=1}^{n} \prod_{k=1}^{2} \left( \frac{\alpha_k^o(t)w_i(t; \theta_k)e^{\beta_k(t)X_i(t)}}{\sum_{j=1}^{n}\alpha_j^o(t)e^{\beta_j(t)X_j(t)}w_j(t; \theta_j)} \right)^{dN_{ki}(t)}
$$

$$
= \prod_{t \geq 0} \prod_{i=1}^{n} \prod_{k=1}^{2} \left( \frac{e^{\beta_k(t)X_i(t)}w_i(t; \theta_k)}{\sum_{j=1}^{n}e^{\beta_j(t)X_j(t)}w_j(t; \theta_j)} \right)^{dN_{ki}(t)}. \tag{4.1}
$$

Hence, the log-likelihood will be given by

$$
\log L(\delta) = \sum_{i=1}^{n} \sum_{k=1}^{2} \int_{0}^{T} \left\{ \beta_k(t)X_i(t) + \log \frac{w_i(t; \theta_k)}{\sum_{j=1}^{n}w_j(t; \theta_j)e^{\beta_j(t)X_j(t)}} \right\} dN_{ki}(t). \tag{4.2}
$$

### 3 The sieve estimator

In the infinite-dimensional case, the (functional) parameter space is given by $\Delta = \{(\beta(t), \theta(t)) \in \mathbb{R}^2 \times (-1, \infty)^2, \ t \in \mathcal{T}\}$ and to be able to apply the maximum likelihood method to such parameter space, we consider the method of sieves [Grenander (1981)] using a similar approach to the one considered by Murphy and Sen (1991). More specifically, we consider the histogram sieves that consists in obtaining MPLE for piecewise constant functions defined on a partition $\Pi$, based on subintervals $I_j^\nu$ of length $l_j, j = 1, \ldots, \nu$ of the time interval $\mathcal{T}$. Therefore, we define a family of
subsets of $\Delta$ given by

$$S_\nu = \left\{ (\beta(t), \theta(t)) : \beta_k(t) = \sum_{j=1}^\nu \beta_{kj} I\{t \in I_j^\nu\} \text{ for } (\beta_{k1}, \ldots, \beta_{k\nu}) \in \mathbb{R}^\nu \right\}$$

$$\theta_k(t) = \sum_{j=1}^\nu \theta_{kj} I\{t \in I_j^\nu\} \text{ for } (\theta_{k1}, \ldots, \theta_{k\nu}) \in (-1, \infty)^\nu \}.$$  

Note that although not explicit in the notation, $\nu = \nu(n)$. Therefore, the issue here is to allow $\nu$ to increase with $n$ in such a way that the consistency as well as the efficiency of the MPLEs are assured.

When working on $S_\nu$, the corresponding log-likelihood will be given by

$$\log L(\delta^\nu) = \sum_{i=1}^n \sum_{k=1}^2 \int_0^T \left\{ \beta_k^\nu(t) X_i(t) + \log \frac{w_i(t; \theta_k^\nu)}{\sum_{j=1}^n w_j(t; \theta_k^\nu) e^{\beta_k^\nu(t) X_j(t)}} \right\} dN_{ki}(t).$$

(4.3)

Therefore, based on (4.3) we compute a MPLE estimator for $\delta^\nu$, designated $\hat{\delta}^\nu$, for a fixed $\nu$. Allowing $\nu$ to increase with $n$, we then obtain a sequence of estimators that hopefully will converge to the true (infinite-dimensional) parameter $\delta$. As pointed out in Murphy and Sen (1991), the sequence $(S_\nu)$ of parameter spaces must satisfy certain properties. For example, such a sequence must be increasing, each $S_\nu$ must contain a MPLE and $\cup_\nu S_\nu$ must be dense in the parameter space $S_\infty$. The last assumption will assure that there actually exists a sequence $(\hat{\delta}_\nu)_{\nu \geq 1}$ converging to the true value $\delta^o \in S_\infty$. We note that the histogram sieves can be seen as simple functions (as they are linear combinations of indicators) and, hence, one has the denseness condition.

Direct comparison between $\hat{\delta}^\nu$ and $\delta^o$ is not straightforward, since they belong to different spaces. Therefore, we need to consider a quantity in $S_\nu$, say $\bar{\delta}^\nu$ that is close enough to $\delta^o$ and compute its distance to $\hat{\delta}^\nu$. Note that as $\nu \rightarrow \infty$, we will require $\bar{\delta}^\nu \rightarrow \delta^o$, uniformly in $t$. Therefore, the choice of $\bar{\delta}^\nu$ is crucial. One possible approach is given in Murphy and Sen (1991). The basic idea is to use, within each
bin, the mean value of the function \( \delta(u) \). To obtain such value, consider a first order Taylor approximation for the partial likelihood defined above. Based on (4.2), we define

\[
 f_{ki}(u; \delta) = \beta_k(u) X_i(u) + \log w_i(u; \theta_k) - \log \sum_{j=1}^{n} w_j(u; \theta_k) e^{\theta_k(u) X_j(u)},
\]

so that

\[
 \log L_n(\delta) = \sum_{i=1}^{n} \sum_{k=1}^{2} \int_0^\tau f_{ki}(u; \delta) \, dN_{ki}(u).
\]

Expansion of (4.4) around \( \delta^o = (\delta_1^o, \delta_2^o)^T \) will give (dropping the variable \( u \) for a moment)

\[
 f_{ki}(\delta) = f_{ki}(\delta^o) + (\delta_k - \delta_k^o)^T \frac{\partial f_{ki}(\delta^o)}{\partial \delta_k} + \frac{1}{2} (\delta_k - \delta_k^o)^T \frac{\partial^2 f_{ki}(\delta^o)}{\partial \delta_k \partial \delta_k^T} (\delta_k - \delta_k^o).
\]

Therefore we may write

\[
 \frac{1}{n} \{ \log L_n(\delta) - \log L_n(\delta^o) \} = \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{2} \int_0^\tau (f_{ki}(u; \delta) - f_{ki}(u; \delta^o)) \, dN_{ki}(u)
\]

\[
 = \frac{1}{n} \sum_{i=1}^{n} \int_0^\tau (\delta_k(u) - \delta_k^o(u))^T \frac{\partial f_{ki}(u; \delta^o)}{\partial \delta_k} \, dN_{ki}(u)
\]

\[
 + \frac{1}{2n} \sum_{i=1}^{n} \sum_{k=1}^{2} \int_0^\tau (\delta_k(u) - \delta_k^o(u))^T \frac{\partial^2 f_{ki}(u; \delta^o)}{\partial \delta_k \partial \delta_k^T} (\delta_k(u) - \delta_k^o(u)) \, dN_{ki},
\]

Heuristically we show that, as \( n \to \infty \), the term (4.5) will converge to zero and then we find the quantity \( \delta^o \) that will maximize the term (4.6). The idea with respect to the linear term is to write the expression as a zero mean martingale, what is accomplished by remembering that \( dM_{ki} = dN_{ki} - d\Lambda_{ki} \) is a zero-mean martingale so that, for (4.5),

\[
 \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{2} \int_0^\tau (\delta_k(u) - \delta_k^o(u))^T \frac{\partial f_{ki}(u; \delta^o)}{\partial \delta_k} \, dN_{ki}
\]

\[
 = \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{2} \int_0^\tau (\delta_k(u) - \delta_k^o(u))^T \frac{\partial f_{ki}(u; \delta^o)}{\partial \delta_k} \, dM_{ki}
\]

\[
 + \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{2} \int_0^\tau (\delta_k(u) - \delta_k^o(u))^T \frac{\partial f_{ki}(u; \delta^o)}{\partial \delta_k} \, d\Lambda_{ki}
\]
The first term in the r.h.s. of the above expression must converge to zero for \( n \to \infty \), since it is an average of zero-mean martingales. As for the second term, we define the following quantities generalizing the definitions given in Chapter 3:

\[
S_k^{(j)}(\delta, u) = n^{-1} \sum_{i=1}^{n} w_i(u; \theta_k) X_i(u)^j e^{\beta_k(u)} X_i(u), \quad j = 0, \ldots, 4,
\]

\[
S_k^{(5)}(\delta, u) = n^{-1} \sum_{i=1}^{n} X_i(u) Y_{i1}(u) Y_{2i}(u) e^{\beta_k(u)} X_i(u),
\]

\[
S_k^{(6)}(\delta, u) = n^{-1} \sum_{i=1}^{n} X_i(u)^2 Y_{i1}(u) Y_{2i}(u) e^{\beta_k(u)} X_i(u),
\]

\[
S_k^{(7)}(\delta, u) = n^{-1} \sum_{i=1}^{n} e^{\beta_k(u)} X_i(u) Y_{i1}(u) Y_{2i}(u),
\]

\[
S_k^{(8)}(\delta, u) = n^{-1} \sum_{i=1}^{n} Y_{i1}(u) Y_{2i}(u) \frac{e^{\beta_k(u)} X_i(u)}{w_i(u; \theta_k)} = \frac{1}{1 + \theta_k(u)} S_k^{(7)}(\delta, u).
\]

Also, we define the following quantities that will be useful in the next sections,

\[
E_\beta(\delta_k, t) = \frac{S_k^{(1)}(\delta, t)}{S_k^{(0)}(\delta, t)} \quad \text{and} \quad E_\theta(\delta_k, t) = \frac{S_k^{(7)}(\delta, t)}{S_k^{(0)}(\delta, t)}.
\]

Since \( d\Lambda_k(u) = w_i(u; \theta_k^O) e^{\beta_k(u)} X_i(u) \alpha_k^O(u) \) \( du \),

\[
\frac{1}{n} \sum_{k=1}^{2} \frac{1}{2} \int_{0}^{T} (\delta_k(u) - \delta_k^O(u))^2 \sum_{i=1}^{n} \left( \frac{\partial f_{k1}(w; \delta^O)}{\partial \theta_k} \right) w_i(u; \theta_k^O) e^{\beta_k(u)} X_i(u) \alpha_k^O(u) \ du
\]

\[
= \frac{1}{n} \sum_{k=1}^{2} \frac{1}{2} \int_{0}^{T} (\delta_k(u) - \delta_k^O(u))^2 \sum_{i=1}^{n} \left( \frac{\partial f_{k1}(w; \delta^O)}{\partial \theta_k} \right) w_i(u; \theta_k^O) e^{\beta_k(u)} X_i(u) \alpha_k^O(u) \ du
\]

\[
= \frac{1}{n} \sum_{k=1}^{2} \frac{1}{2} \int_{0}^{T} (\delta_k(u) - \delta_k^O(u))^2 \left( \frac{S_k^{(1)}(u; \delta^O) - S_k^{(1)}(u; \delta^O)}{S_k^{(0)}(u; \delta^O)} \right) w_i(u; \theta_k^O) e^{\beta_k(u)} X_i(u) \alpha_k^O(u) \ du = 0.
\]

Therefore, it is to be expected that the contribution of the linear term tends to be negligible as the sample size becomes large and, hence, we turn our attention to obtain a value \( \delta^O \) that maximizes the quadratic term (4.6). Using the same
approach considered for the first term,
\[
\frac{1}{2n} \sum_{i=1}^{n} \sum_{k=1}^{2} \int_{0}^{\tau} (\delta_k - \delta_k^0)^T \frac{\partial^2 f_{ki}(\delta^0)}{\partial \delta_k \partial \delta_k^T} (\delta_k - \delta_k^0) \, dN_{ki} \\
= \frac{1}{2n} \sum_{i=1}^{n} \sum_{k=1}^{2} \int_{0}^{\tau} (\delta_k - \delta_k^0)^T \frac{\partial^2 f_{ki}(\delta^0)}{\partial \delta_k \partial \delta_k^T} (\delta_k - \delta_k^0) \, dM_{ki} \\
+ \frac{1}{2n} \sum_{i=1}^{n} \sum_{k=1}^{2} \int_{0}^{\tau} (\delta_k - \delta_k^0)^T \frac{\partial^2 f_{ki}(\delta^0)}{\partial \delta_k \partial \delta_k^T} (\delta_k - \delta_k^0) w_i(\theta_k^0) e^{\delta_k^0 X_i} \, du
\]
(4.7)

Here also the first term is a zero-mean martingale and, hence, should behave as zero as $n$ increases. As for (4.7), the second derivatives are given by
\[
\frac{\partial f_{ki}}{\partial \beta_k} = - \frac{S_k^{(1)}}{S_k^{(0)}} + \frac{S_k^{(2)}}{S_k^{(0)}}^2 \\
\frac{\partial f_{ki}}{\partial \beta_k \partial \theta_k} = - \sum_{i=1}^{n} Y_{1i} Y_{2i} e^{\beta_k X_i} + \frac{S_k^{(1)} \sum_{i=1}^{n} Y_{1i} Y_{2i} e^{\beta_k X_i}}{(S_k^{(0)})^2} \\
\frac{\partial f_{ki}}{\partial \theta_k^2} = \frac{(Y_{1i} Y_{2i})^2}{w_i^2(\theta_k^0)} + \frac{S_k^{(7)} \sum_{i=1}^{n} Y_{1i} Y_{2i} e^{\beta_k X_i}}{(S_k^{(0)})^2} \\
= - \frac{(Y_{1i} Y_{2i})^2}{w_i(\theta_k^0)} + \frac{S_k^{(7)}}{S_k^{(0)}}
\]

So that we define $V_{\beta_k}$, $V_{\theta_k}$, and $C_{\delta_k}$ such that
\[
\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 f_{ki}}{\partial \beta_k^2} w_i(\theta_k^0) e^{\beta_k X_i} = - \left\{ \frac{S_k^{(2)}}{S_k^{(0)}} - \left( \frac{S_k^{(1)}}{S_k^{(0)}} \right)^2 \right\} S_k^{(0)} = - V_{\beta_k} S_k^{(0)}, \\
\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 f_{ki}}{\partial \beta_k \partial \theta_k} w_i(\theta_k^0) e^{\beta_k X_i} = - \left\{ \frac{S_k^{(5)}}{S_k^{(0)}} - \frac{S_k^{(1)} S_k^{(7)}}{(S_k^{(0)})^2} \right\} S_k^{(0)} = - C_{\delta_k} S_k^{(0)}
\]

and
\[
\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 f_{ki}}{\partial \theta_k^2} w_i(\theta_k^0) e^{\beta_k X_i} = \sum_{i=1}^{n} \left\{ - \frac{(Y_{1i} Y_{2i})^2}{w_i(\theta_k^0)} + \frac{(S_k^{(7)})^2}{S_k^{(0)}} \right\} \frac{w_i(\theta_k^0) e^{\beta_k X_i}}{n}
\]

85
\[-\frac{1}{2} \sum_{k=1}^{n} \int_{0}^{\tau} (\delta_{k} - \delta_{k}^{0})^T v_k (\delta_{k} - \delta_{k}^{0}) s_k^{(0)} \alpha_k^{2} \, du.\]

In $S_{\nu}$, the vector-valued function $\delta_{k}(\cdot)$ is of the form $\delta_{k}(u) = \sum_{h=1}^{\nu} \delta_{kh}^{\nu} I_h(u)$, so that

\[
-\frac{1}{n} \{ \log L_n(\delta^{\nu}) - \log L_n(\delta^{0}) \} 
\approx -\frac{1}{2} \sum_{k=1}^{n} \int_{0}^{\tau} \left( \sum_{h=1}^{\nu} \delta_{kh}^{\nu} I_h^{\nu} - \delta_{k}^{0} \right)^T v_k \left( \sum_{h=1}^{\nu} \delta_{kh}^{\nu} I_h^{\nu} - \delta_{k}^{0} \right) s_k^{(0)} \alpha_k^{2} \, du. \tag{4.8}
\]

Taking the derivative w.r.t. $\delta_{kh}^{\nu}$,

\[
-\int_{0}^{\tau} I_{h'}^{\nu} v_k \left( \sum_{h=1}^{\nu} \delta_{kh}^{\nu} I_h^{\nu} - \delta_{k}^{0} \right) s_k^{(0)} \alpha_k^{2} \, du \\
= -\int_{0}^{\tau} I_{h'}^{\nu} v_k \left( \sum_{h=1}^{\nu} \delta_{kh}^{\nu} I_h^{\nu} \right) s_k^{(0)} \alpha_k^{2} \, du + \int_{0}^{\tau} I_{h'}^{\nu} v_k \delta_{k}^{0} s_k^{(0)} \alpha_k^{2} \, du \\
= -\int_{0}^{\tau} I_{h'}^{\nu} v_k \delta_{kh}^{\nu} s_k^{(0)} \alpha_k^{2} \, du + \int_{0}^{\tau} I_{h'}^{\nu} v_k \delta_{k}^{0} s_k^{(0)} \alpha_k^{2} \, du = 0.
\]

Solving this equation for $\delta_{kh}^{\nu}$, we obtain

\[
\delta_{kh}^{\nu} = \sum_{h'=-1}^{1} \int_{0}^{\tau} I_{h'}^{\nu}(u) v_k(u) \delta_{k}^{0}(u) s_k^{(0)}(\delta^{0}; u) \alpha_k^{2}(u) \, du, \tag{4.9}
\]

86
where
\[ \Sigma_{kh'} = \int_0^T \Pi_k(u) \nu_k(u) \alpha_k^2(u) \, du \]
is a 2 \times 2 matrix, assumed to be positive definite.

The quantity in expression (4.9) can be interpreted as a mean value for the
function vector \( \delta_k^o(u) \) in a particular bin. Note that if we define the quantity \( \eta_{kh'} = \int_0^T \Pi_k(u) \alpha_k^2(u) \, du \) so that \( \Sigma_{kh'}^o = (\eta_{kh'})^{-1} \Sigma_{kh'} \), then (4.9) can be rewritten as
\[ \tilde{\delta}_{kh'}^o = (\Sigma_{kh'}^o)^{-1} \left( \int_0^T \Pi_k(u) \nu_k(u) \delta_k^o(u) s_k^{(0)}(u) \alpha_k^2(u) \, du / \eta_{kh'} \right) \]
and \( \Sigma_{kh'}^o \) may be interpreted as an average variation matrix for the variance on
the \( h' \)th bin. Also, we will need to have \( \eta_{kh'} \sim K_{kh'} l_{h'} \), for \( K_{kh'} > 0 \). This last assumption is needed in order to guarantee that \( \tilde{\delta}_{kh'}^o \) will have a limit as the mesh
size tends to zero.

Another consideration is related to the fact that since the quantity given by
\( (1/n) \{ \log L_n(\delta^o) - \log L_n(\delta^o) \} \) is an average, it is expected that it will converge to
\( \mathbb{E}\{ \log L_1(\delta^o) - \log L_1(\delta^o) \} \). This last quantity is the (negative) Kullback-Leibler
information [cf.Kullback (1978)] and as discussed earlier, it is approximated (for
large \( n \)) by expression (4.8), similarly to the case discussed in Murphy and Sen
(1991). Also, as mentioned in condition C2(a) in Geman and Hwang (1982), such
an expression indicates the natural norm to be considered in proving consistency.

In our case we can write then
\[ \mathbb{E}\{ \log \frac{L_1(\delta^o)}{L_1(\delta^o)} \} = \frac{1}{2} \sum_{k=1}^2 \int_0^T (\delta_k^o(u) - \delta_k^o(u))^T v_k(u) (\delta_k^o(u) - \delta_k^o(u)) s_k^{(0)}(u) \alpha_k^2(u) \, du \]
\[ = \frac{1}{2} \sum_{k=1}^2 \int_0^T (\delta_k^o(u) - \delta_k^o(u))^T v_k(u) (\delta_k^o(u) - \delta_k^o(u)) \, d\mu_k, \]
where \( d\mu_k = s_k^{(0)}(u) \alpha_k^2(u) \, du \).
4 Asymptotic properties for the sieve estimator

We follow now by considering the consistency of the sieve estimator. To do so, we will need the following set of conditions (generalizations of the ones usually required for the Cox model).

Conditions

A. For $k = 1, 2$,

(1) there exist $s_k^{(i)}(\delta, t), i = 0, 1, \ldots, 7$ such that, for all $\gamma > 0$,

$$\sup_{t \in \mathcal{T}} \sup_{\|d - \delta^{(i)}(t)\| \leq \gamma} |S_k^{(i)}(d, t) - s_k^{(i)}(d, t)| = o_p(1).$$

Also, we define $s_k^{(8)}(\delta, t) = (1 + \theta_k^{(i)}(u))^{-1}s_k^{(7)}(\delta, t), \ t \in \mathcal{T}$. In addition, we assume that

(2)

$$n \int_0^\tau (S_k^{(i)}(\delta, t) - s_k^{(i)}(\delta, t))^2 \ dt = O_p(1).$$

(3) For $i = 1, 2, \ldots, 6$, and $\gamma > 0$,

$$\sup_{t \in \mathcal{T}} \sup_{d \in \mathbb{R} \times (-1, \infty)} \sup_{\|d - \delta^{(i)}_k\| < \gamma} \frac{|S_k^{(i)}(d, t)|}{S_k^{(0)}(d, t)} = O_p(1).$$

B. For $k=1,2$, there exists a constant $\psi > 0$ such that

$$n^{-1/2} \sup_{u \in \mathcal{T}} \max_{1 \leq i \leq n} |X_i(u)| |Y_k(u)| \mathbb{I}\{ \beta_k(u)X_i(u) > -\psi |X_i(u)| \} = o_p(1).$$

C. For $k = 1, 2$,

(1) The functions $s_k^{(i)}(\cdot, t)$ are continuous in $\delta$, uniformly in $t \in \mathcal{T}$, and there exist positive constants $U_k^{(1)}$ and $U_k^{(2)}$, such that $\alpha_k^2(t) \leq \alpha_k(t) \leq U_k^{(1)}$, $s_k^{(i)}(\delta, t) \leq U_k^{(1)}$, $i = 0, \ldots, 8$, and $s_k^{(0)}(\delta, t) \geq U_k^{(2)}$. 

88
(2) Define \( c_\beta = s^{(1)}_k / s^{(0)}_k \) and \( e_{\delta^o_k} = s^{(7)}_k / s^{(0)}_k \) and consider the symmetric matrix \( v_k(t) \) with elements

\[
v_{\beta^o_k}(t) = \frac{s^{(2)}_k(\delta^o, t)}{s^{(0)}_k(\delta^o, t)} - e^2_{\beta}(\delta^o_k, t),
\]

\[
v_{\delta^o_k}(t) = \frac{s^{(8)}_k(\delta^o, t)}{s^{(0)}_k(\delta^o, t)} - e^2_{\delta}(\delta^o_k, t)
\]

in the main diagonal and the other elements given by

\[
c_{\delta^o_k}(t) = \frac{s^{(5)}_k(\delta^o, t)}{s^{(0)}_k(\delta^o, t)} - e_{\beta}(\delta^o_k, t)e_{\delta}(\delta^o_k, t).
\]

We suppose that there exists a constant \( L > 0 \) such that the smallest eigenvalue of the matrix \( v_k(t)s^{(0)}_k(\delta^o, t)\alpha^o_k(t) > L \), uniformly in \( t \).

D. For \( k = 1, 2, \)

(1) For \( k=1,2, \) \( \delta^o_k(u) \) is component-wise first-order Lipschitz.

(2) For \( k = 1, 2, \) the dependence parameter \( \theta^o_k \) is bounded away from -1, i.e.,

\[
\inf_{t \in \tau} \theta^o_k(t) \geq -1 + \eta, \text{ for some } \eta > 0.
\]

(3) The second derivatives of the elements of \( \delta^o_k(u) \) are bounded.

Remarks: Condition A is usually referred to as asymptotic stability. Note that condition A(3) is not required for \( S^{(7)}_k \) and \( S^{(8)}_k \), as those quantities (when divided by \( S^{(0)}_k \)) are bounded by one. Condition in B is related to the Lindeberg condition and is considered when proving weak convergence. Condition C relates to asymptotic regularity and will allow exchange of limits in several occasions. Conditions in D(1) relates to the proofs involving the bias and D(2) guarantees that the conditional hazards functions (defined when we discussed the intensity process for the matrix-valued model) are not degenerated.

Several results are needed for proving the consistency and they are presented in the following lemmas.
Lemma 4.1 Assuming that conditions A, C(1) and D(1) are true, in addition to \( \limsup_{n \to \infty} (l_{(\nu)}/l_{(1)}) < \infty \), then we have that for \( k = 1, 2, \)

\[
\sum_{h=1}^{\nu} \left\{ (nl_h)^{-1} \left\| \frac{\partial \log L_n(\delta^\nu)}{\partial \delta_{kh}} \right\|^2 \right\} - \leq 2(\|l\|^4 n)^{-1} \left\{ \|l\|^4 \int_0^T \sum_{h=1}^{\nu} l_h^{-2} (v_{\rho_k}(t) + v_{\varphi_k}(t)) s_k^{(0)}(\delta^\nu, t) \alpha_k^2(t) \, dt + o_P(1) \right\} 
+ O_p(1/n) + o_p(\|l\|^2).
\]

(4.10)

The proof consists in showing that the quantity in the l.h.s. of 4.10 can be bounded by a submartingale plus a predictable variation process. The submartingale is shown to converge to the first quantity in the r.h.s. of 4.10 using the Doob-Meier decomposition to express it as a zero-mean martingale and then working on the corresponding compensator with the Lenglart’s inequality. Using first-order Taylor expansion, the predictable variation process is shown to be of the mentioned order.

Details are presented in Section 5.

Lemma 4.2 Let \( M_h^+ \) be a matrix with elements taken equal to the absolute value of the elements of

\[
(nl_h)^{-1} \frac{\partial^2 \log L_n(\delta^\nu)}{\partial \delta_{kh} \partial \delta_{kh}^i} + l_h^{-1} \Sigma_{kh}.
\]

Then, under conditions A(1), A(3), C(1), and D(2),

\[
\max_{1 \leq h \leq \nu} 1^T M_h^+ 1 = O_p(n^{-1/2}\|l\|^{-2}) + o_p(1).
\]

In order to obtain the result stated in the lemma we work with each element of the matrix \( M_h^+ \). For each term the triangular inequality is used to obtain quantities depending on the counting processes \( N_{k_t} \). By the Doob-Meier decomposition those quantities can be expressed as martingales, that can be shown to have the desired rate by the Lenglart’s inequality. For details we refer to Section 5.
**Lemma 4.3** Let $M^*_h$ be a matrix with elements taken equal to the absolute value of the elements of

$$(nl^{-1}_h)\left\{ \frac{\partial^2 \log L_n(\delta^*_{kh})}{\partial \delta^*_{kh} \partial \delta^*_{kh}} - \frac{\partial^2 \log L_n(\bar{\delta}^*_{kh})}{\partial \delta^*_{kh} \partial \bar{\delta}^*_{kh}} \right\},$$

for $||\delta^*_k - \bar{\delta}^*_k|| = o(1)$, $\delta^*_k = (\delta^*_{k1}, \ldots, \delta^*_{k\nu})$ and $\bar{\delta}^*_\nu = (\bar{\delta}^*_{k1}, \ldots, \bar{\delta}^*_{k\nu})$. Then, under conditions A(1), A(3), and C(1),

$$\max_{1 \leq k \leq \nu} 1^T M^*_h 1 = o_p((n^{1/2}||l||^2)^{-1}) + o_p(1).$$

Similar to the previous lemma, the proof here consists in showing that each element of $M^*_h$ can be expressed in terms of the counting processes $N_{ki}$. Direct application of conditions A(1), A(3), and C(1) and Lemma 1 of Murphy and Sen (1991) completes the proof that is detailed in Section 5.

Now we prove the consistency for the sieve estimator, expressed in terms of an integrated process. The basic idea is to show that the distance between the estimator and $\bar{\delta}^*_{\nu}$ and the distance between the latter and $\delta^*$ goes to zero as $n \to \infty$. Then the result follows by the triangular inequality.

**Theorem 4.1** Assume that together with conditions A, C, D(1), we have

1. $\lim_{n \to \infty} n||l||^6 < \infty$,

2. $\lim_{n \to \infty} n||l||^4 = \infty$,

3. $\limsup_{n \to \infty} 1_{(\nu)} / 1_{(1)} < \infty$.

Then, for $\hat{\delta}$ maximizing $\log L_n(\delta)$ in $S_{\nu},$

$$\int_0^T ||\hat{\delta}(u) - \delta^*(u)||^2 \, du = O_p((n||l||^2)^{-1}) + O(||l||^4), \quad \text{as} \ n \to \infty.$$
Remarks: We note that assumption (1) is stronger here than the original one in Murphy and Sen (1991). This is due to the fact that we do not work with the third derivatives in the Taylor expansions. However, it is possible to obtain the same rate as Murphy and Sen (1991) but with increasing complication on the algebraic manipulations. Also, it will be desirable in practice to work with partitions satisfying (2) and such that $n\|l\|^8$ has bounded limit, and that implies our assumption (1) automatically.

Proof: Given the orthogonality between $\delta_1$ and $\delta_2$, we consider a generic $\delta_k$. We can then write
\[
\int_0^r \|\delta_k(u) - \delta_k^*(u)\|^2 \, du \leq 2 \int_0^r \|\delta_k(u) - \delta_k^*(u)\|^2 \, du + 2 \int_0^r \|\delta_k^*(u) - \delta_k^0(u)\|^2 \, du. \tag{4.11}
\]

The proof then consists in showing that (4.11) is $O((n\|l\|^2)^{-1})$ and that (4.12) is $O(||l||^4)$. Regarding the latter, we have that
\[
\delta_k^*(u) - \delta_k^0(u) = \sum_{h=1}^{\nu} \delta_k^{(h)}(u) \beta_h(u) - \delta_k^0(u)
\]
\[
= \sum_{h=1}^{\nu} \left\{ \sum_{k=1}^{\Sigma} \int_0^r I_h^k(s) \delta_k^0(s) \delta_k^{(0)}(s) \alpha_h^0(s) \, ds - \sum_{h=1}^{\Sigma} \sum_{k=1}^{\Sigma} \delta_k^0(u) I_h^k(u) \right\} I_h^k(u)
\]
\[
= \sum_{h=1}^{\nu} \sum_{k=1}^{\Sigma} \left\{ \int_0^r I_h^k(s) \delta_k^0(s) \delta_k^{(0)}(s) \alpha_h^0(s) \, ds - \int_0^r I_h^k(s) \delta_k^0(u) \delta_k^{(0)}(s) \alpha_h^0(s) \, ds \right\} I_h^k(u)
\]
\[
= \sum_{h=1}^{\nu} \sum_{k=1}^{\Sigma} \int_0^r I_h^k(s) \delta_k^0(s) \delta_k^{(0)}(s) \alpha_h^0(s) \, ds - \int_0^r I_h^k(s) \delta_k^0(u) \delta_k^{(0)}(s) \alpha_h^0(s) \, ds I_h^k(u).
\]

Recall that for a symmetric matrix $A$ and a column vector $x$ [cf. Rao (1973)],
\[
\sup_x \frac{x^T A x}{x^T x} = \lambda_1 \quad \text{and} \quad \inf_x \frac{x^T A x}{x^T x} = \lambda_2, \tag{4.13}
\]
where $\lambda_1 (\lambda_2)$ is the largest (smallest) eigenvalue of the matrix $A$. Therefore, denoting $x_h = \int_0^r I_h^k(s) \delta_k^0(s) \delta_k^{(0)}(s) \alpha_h^0(s) \, ds$, using the first expression
in (4.13) we write
\[
\|\delta_k(u) - \delta_k^*(u)\|^2 \leq \sum_{h=1}^\nu \left[\mathbf{x}_h^\top \left(\Sigma_{kh}^{-1} \mathbf{x}_h\right) \Sigma_{kh}^{-1} \mathbf{x}_h\right] \mathbb{I}_h(u) = \sum_{h=1}^\nu \mathbf{x}_h^\top \Sigma_{kh}^{-2} \mathbf{x}_h \mathbb{I}_h(u)
\]
\[
\leq \sum_{h=1}^\nu \|\mathbf{x}_h\|^2 \lambda_1(\Sigma_{kh}^{-2}) \mathbb{I}_h(u) \leq \sum_{h=1}^\nu \|\mathbf{x}_h\|^2 \text{trace}(\Sigma_{kh}^{-2}) \mathbb{I}_h(u).
\]

Denoting the elements of \(\Sigma_{kh}\) by
\[
\Sigma_{kh} = \begin{pmatrix}
\sigma_{h1} & \sigma_{h12} \\
\sigma_{h12} & \sigma_{h2}
\end{pmatrix},
\]
condition C results \(\sigma_{h1} = \int_0^\tau \mathbb{I}_h(u) v_{\beta_k}(u) s_k^{(0)}(0, u) \alpha_k^2(u) \, du \leq O(1) \int_0^\tau \mathbb{I}_h(u) \, du = O(l_h)\). Similarly, \(\sigma_{h12} = O(l_h)\) and \(\sigma_{h2} = O(l_h)\). Consequently, since \(v_k\) is assumed positive definite, we conclude that \(\text{det}(\Sigma_{kh}) = |\Sigma_{kh}| = O(l_h)\) and hence,
\[
\text{trace}(\Sigma_{kh}^{-2}) = |\Sigma_{kh}|^{-1}(\sigma_{h2}^2 + \sigma_{h12}^2) = O(l_h^{-4}) O(l_h^2) = O(l_h^{-2}).
\]

(4.14)

On the other hand, using conditions C and D(1) we have that
\[
\|\mathbf{x}_h\|^2 = \int_0^\tau \mathbb{I}_h(s) [\delta_k^*(s) - \delta_k^*(u)] \mathbf{v}_k(s) s_k^{(0)}(0, s) \alpha_k^2(s) \, ds
\]
\[
\times \int_0^\tau \mathbb{I}_h(s) \mathbf{v}_k(s) [\delta_k^*(s) - \delta_k^*(u)] s_k^{(0)}(0, s) \alpha_k^2(s) \, ds
\]
\[
= \left(\int_0^\tau \mathbb{I}_h(s) \left[\beta_k^2(s) - \beta_k^2(u)\right] \mathbf{v}_k(s) s_k^{(0)}(0, s) \alpha_k^2(s) \, ds\right)^2
\]
\[
+ \left(\int_0^\tau \mathbb{I}_h(s) \left[\theta_k^2(s) - \theta_k^2(u)\right] c_{\delta_k}(s) + \left[\theta_k^2(s) - \theta_k^2(u)\right] v_{\delta_k}(s) \right) s_k^{(0)}(0, s) \alpha_k^2(s) \, ds\right)^2
\]
\[
\leq \left(\int_0^\tau \mathbb{I}_h(s) \left[\beta_k^2(s) - \beta_k^2(u)\right] \mathbf{v}_k(s) s_k^{(0)}(0, s) \alpha_k^2(s) \, ds\right)^2
\]
\[
+ \left(\int_0^\tau \mathbb{I}_h(s) \left[\theta_k^2(s) - \theta_k^2(u)\right] c_{\delta_k}(s) + \left[\theta_k^2(s) - \theta_k^2(u)\right] v_{\delta_k}(s) \right) s_k^{(0)}(0, s) \alpha_k^2(s) \, ds\right)^2
\]
\[
\leq O(1) \left(\int_0^\tau \mathbb{I}_h(s) \left[s - u\right] \mathbf{v}_k(s) + c_{\delta_k}(s) \right) s_k^{(0)}(0, s) \alpha_k^2(s) \, ds\right)^2
\]
\[
+ O(1) \left(\int_0^\tau \mathbb{I}_h(s) \left[s - u\right] \mathbf{v}_k(s) + v_{\delta_k}(s) \right) s_k^{(0)}(0, s) \alpha_k^2(s) \, ds\right)^2
\]
\[
\leq O(1) l_h^2 \left(\int_0^\tau \mathbb{I}_h(s) \, ds\right)^2 = O(l_h^2).\]

(4.15)
Hence, by expressions (4.14) and (4.15),
\[ \|\tilde{\delta}^\nu_k(u) - \delta^\nu_k(u)\|^2 = \sum_{h=1}^\nu O(l^2_h)O(l^2_h)l^2_h\Pi^\nu_k(u) = O(1) \sum_{h=1}^\nu l^2_h\Pi^\nu_k(u). \]

Consequently,
\[ \int_0^\tau \|\tilde{\delta}^\nu_k(u) - \delta^\nu_k(u)\|^2 \, du = O(1) \int_0^\tau \sum_{h=1}^\nu l^2_h\Pi^\nu_k(u) \, du = O(1) \sum_{h=1}^\nu l^2_h l_h \]
\[ \leq O(1) \sum_{h=1}^\nu l^2_h l_{(\nu)} \leq O(1)\|l\|^2 l_{(\nu)} \leq O(1)\|l\|^2\|l\|^2 = O(\|l\|^4). \]

In order to deal with (4.11) we first note that
\[ \int_0^\tau \|\tilde{\delta}_k(u) - \tilde{\delta}^\nu_k(u)\|^2 \, du = \int_0^\tau (\tilde{\beta}_k(u) - \tilde{\beta}^\nu_k(u))^2 \, du + \int_0^\tau (\tilde{\theta}_k(u) - \tilde{\theta}^\nu_k(u))^2 \, du \]
\[ = \int_0^\tau (\sum_{h=1}^\nu (\tilde{\beta}_{kh} - \tilde{\beta}^\nu_{kh})\Pi^\nu_k(u))^2 \, du + \int_0^\tau (\sum_{h=1}^\nu (\tilde{\theta}_{kh} - \tilde{\theta}^\nu_{kh})\Pi^\nu_k(u))^2 \, du \]
\[ = \int_0^\tau \sum_{h=1}^\nu (\tilde{\beta}_{kh} - \tilde{\beta}^\nu_{kh})^2\Pi^\nu_k(u) \, du + \int_0^\tau \sum_{h=1}^\nu (\tilde{\theta}_{kh} - \tilde{\theta}^\nu_{kh})^2\Pi^\nu_k(u) \, du \]
\[ \leq 2\|l\|^2\|\tilde{\delta}_k - \tilde{\delta}^\nu_k\|^2, \]
where the norm in the last expression is the $l_2$-norm. Thus, in order to prove that
\[ \int_0^\tau \|\tilde{\delta}_k - \tilde{\delta}^\nu_k\|^2 \, du = O_p((n\|l\|^2)^{-1}) \]
it is sufficient to show that
\[ \|\tilde{\delta}_k - \tilde{\delta}^\nu_k\|^2 = O_p((n\|l\|^4)^{-1}). \]

To accomplish this we adapt the fixed point theorem in Murphy and Sen (1991), i.e., we have to show that for $\delta_k \in S_\nu$,

\[ \sup_{A} \sum_{h=1}^\nu (nl_h)^{-1}(\delta_{kh} - \tilde{\delta}^\nu_{kh})^T \frac{\partial \log L_n(\delta)}{\partial \delta_{kh}} < 0 \]

where $A$ is conveniently taken as the set of all $\delta_k \in S_\nu$ such that $\|\delta_k - \tilde{\delta}_k\| = (n\|l\|^4)^{-1/2}\psi_\nu$, for

\[ \psi_\nu^2 = 8L^{-2}\|l\|^4 \int_0^\tau \sum_{h=1}^\nu \Pi^\nu_k(t)l^{-2}_h(v_{\beta^o_k}(t) + v_{\theta^o_k}(t))s^{(0)}_k(\delta^o, t)\alpha^o_k(t) \, dt. \]

94
If (4.18) is true, then it follows by lemma 2 of Aitchison and Silvey (1958) that 
\[ \exists \delta^{\nu} \in \delta^{*} \text{ such that } \| (\delta^{\nu} - \delta^{\nu}) / (\|l\|^4n)^{-1/2} \psi_{\nu} \| \leq 1 \rightarrow \| \delta^{\nu} - \delta^{\nu} \| \leq (\|l\|^4n)^{-1/2} \psi_{\nu}. \]
As \( \psi_{\nu} \) is bounded (and bounded away from zero), then the result for (4.11) will follow.

We rewrite the l.h.s. of (4.18) by considering the Taylor expansion
\[
\frac{\partial \log L_n(\delta)}{\partial \delta_{kh}} = \frac{\partial \log L_n(\delta^{\nu})}{\partial \delta_{kh}} + \frac{\partial^2 \log L_n(\delta^{\nu})}{\partial \delta_{kh} \partial \delta^{\nu}_{kh}} (\delta_{kh} - \delta^{\nu}_{kh}),
\]
for \( \| \delta_{kh} - \delta^{\nu}_{kh} \| \leq \| \delta_{kh} - \delta^{\nu}_{kh} \| \), so that
\[
\sum_{h=1}^{\nu} (n_{lh})^{-1} (\delta_{kh} - \delta^{\nu}_{kh})^T \frac{\partial \log L_n(\delta)}{\partial \delta_{kh}} = \\
= \sum_{h=1}^{\nu} (n_{lh})^{-1} (\delta_{kh} - \delta^{\nu}_{kh})^T \frac{\partial \log L_n(\delta^{\nu})}{\partial \delta_{kh}} \]
\[
+ \sum_{h=1}^{\nu} (n_{lh})^{-1} (\delta_{kh} - \delta^{\nu}_{kh})^T \left[ \frac{\partial^2 \log L_n(\delta^{\nu})}{\partial \delta_{kh} \partial \delta^{\nu}_{kh}} - \frac{\partial \log L_n(\delta^{\nu})}{\partial \delta_{kh} \partial \delta^{\nu}_{kh}} \right] (\delta_{kh} - \delta^{\nu}_{kh}). \tag{4.19}
\]
\[
+ \sum_{h=1}^{\nu} (n_{lh})^{-1} (\delta_{kh} - \delta^{\nu}_{kh})^T \left[ \frac{\partial^2 \log L_n(\delta^{\nu})}{\partial \delta_{kh} \partial \delta^{\nu}_{kh}} - \frac{\partial \log L_n(\delta^{\nu})}{\partial \delta_{kh} \partial \delta^{\nu}_{kh}} \right] (\delta_{kh} - \delta^{\nu}_{kh}). \tag{4.20}
\]
For expression (4.19) we use Lemma 4.1 so that such a quantity is bounded above by
\[
\sum_{h=1}^{\nu} (n_{lh})^{-1} \| (\delta_{kh} - \delta^{\nu}_{kh})^T \frac{\partial \log L_n(\delta)}{\partial \delta_{kh}} \| \leq \sum_{h=1}^{\nu} (n_{lh})^{-1} \| \delta_{kh} - \delta^{\nu}_{kh} \| \frac{\partial \log L_n(\delta)}{\partial \delta_{kh}} \|
\leq [\sum_{h=1}^{\nu} (n_{lh})^{-2} \| \frac{\partial \log L_n(\delta^{\nu})}{\partial \delta_{kh}} \|^2]^{1/2} \left[ \sum_{h=1}^{\nu} \| \delta_{kh} - \delta^{\nu}_{kh} \|^2 \right]^{1/2}
\leq [\sum_{h=1}^{\nu} \| (n_{lh})^{-2} \frac{\partial \log L_n(\delta^{\nu})}{\partial \delta_{kh}} \|^2]^{1/2} \| \delta_{k} - \delta^{\nu}_{k} \|
\leq \left( n \|l\|^4 \right)^{-1} \left\{ \frac{L^2 \psi_{\nu}^2}{4} + o_p(1) \right\} + o_p(\|l\|^2) \right\}^{1/2} \| \delta_{k} - \delta^{\nu}_{k} \| \tag{4.22}
For expression (4.20), considering $\Sigma_{kh}$ defined in (4.9), we have that

$$
\sum_{h=1}^{\nu} (nl_h)^{-1} (\delta_{kh} - \bar{\delta}_{kh})^T \partial^2 \log L_n(\bar{\delta}^\nu) \frac{\partial}{\partial \delta_{kh} \partial \delta_{kh}^T} (\delta_{kh} - \bar{\delta}_{kh})
= \sum_{h=1}^{\nu} (\delta_{kh} - \bar{\delta}_{kh})^T \left[(nl_h)^{-1} \frac{\partial^2 \log L_n(\bar{\delta}^\nu)}{\partial \delta_{kh} \partial \delta_{kh}^T} + l_h^{-1} \Sigma_{kh} \right] (\delta_{kh} - \bar{\delta}_{kh})
- \sum_{h=1}^{\nu} (\delta_{kh} - \bar{\delta}_{kh})^T [l_h^{-1} \Sigma_{kh}] (\delta_{kh} - \bar{\delta}_{kh}).
$$

(4.23)

The second term in the r.h.s. above can be written as

$$
\sum_{h=1}^{\nu} l_h^{-1} \int_0^T \Pi_h(u) (\delta_{kh} - \bar{\delta}_{kh})^T v_k(u) s_k^{(0)}(\delta^\circ, u) \alpha_k^e(u) (\delta_{kh} - \bar{\delta}_{kh}) \, du
\geq \sum_{h=1}^{\nu} l_h^{-1} \int_0^T \Pi_h(u) \lambda_2(v_k(u) s_k^{(0)}(\delta^\circ, u) \alpha_k^e(u))) ||\delta_{kh} - \bar{\delta}_{kh}||^2 \, du \quad \text{[by (4.13)]}
\geq \sum_{h=1}^{\nu} l_h^{-1} ||\delta_{kh} - \bar{\delta}_{kh}||^2 L \int_0^T \Pi_h(u) \, du \quad \text{[by C(2)]}
= L||\delta_{\hat{k}} - \bar{\delta}_{\hat{k}}||^2.
$$

On the other hand, if for the first term on (4.23) we denote $x_h = (\delta_{kh} - \bar{\delta}_{kh})$ and $M_h = [(nl_h)^{-1} (\partial^2 / \partial \delta_{kh} \partial \delta_{kh}^T) \log L_n(\bar{\delta}^\nu) + l_h^{-1} \Sigma_{kh}]$, then we write (for $m_{ij}$ representing the elements of $M_h$)

$$
\sum_{h=1}^{\nu} x_h^T M_h x_h = \sum_{h=1}^{\nu} \sum_{i,j=1}^{2} x_i m_{ij} x_j \leq \sum_{i,j=1}^{2} |m_{ij}| x_i^T x_j
\leq \sum_{h=1}^{\nu} \max_{i=1,2} |x_i|^2 \sum_{i,j=1}^{2} |m_{ij}| \leq \sum_{h=1}^{\nu} \|x_h\|^2 1^T M_h^+ 1
\leq \max_{1 \leq h \leq \nu} 1^T M_h^+ 1 \|x\|^2 = \max_{1 \leq h \leq \nu} 1^T M_h^+ 1 \|\delta_{\hat{k}} - \bar{\delta}_{\hat{k}}\|^2,
$$

where the elements of $M_h^+$ are equal to the absolute value of the elements of $M_h$. It then follows by Lemma 4.2 that the last term above is of order $O_p(n^{-1/2} \|l\|^2) + o_p(1)$,
and hence,
\[
\sum_{h=1}^{\nu} (nl_h)^{-1} (\delta_{kh} - \delta_{kh}^\nu)^T \frac{\partial^2 \log L_n(\delta^\nu)}{\partial \delta_{kh} \delta_{kh}^\nu} (\delta_{kh} - \delta_{kh}^\nu) \\
\leq \{ O_p(n^{-1/2} ||l||^{-2}) + o_p(1) - L \} ||\delta_k - \delta_k^\nu||^2.
\] (4.24)

We can follow a similar approach to deal with expression (4.21), i.e., defining
\( M_k^* \) as a 2 \times 2 matrix with elements given by the absolute value of the elements of
the matrix \((nl_h)^{-1} (\partial^2 / \partial \delta_{kh} \partial \delta_{kh}^\nu) (\log L_n(\delta^*) - \log L_n(\delta^\nu))\), then expression (4.21) is
less than or equal to
\[
\max_{1 \leq h \leq \nu} 1^T M_k^* 1 ||\delta_k - \delta_k^\nu||^2 = \left[ o_p((n^{1/2} ||l||^2)^{-1}) + o_p(1) \right] ||\delta_k - \delta_k^\nu||^2,
\] (4.25)
by Lemma 4.3.

Taking \( ||\delta_k - \delta_k^\nu||^2 = (n||l||^4)^{-1} \psi_\nu^2 \) and using (4.22), (4.24), and (4.25),
\[
\sum_{h=1}^{\nu} (nl_h)^{-1} (\delta_{kh} - \delta_{kh}^\nu)^T \frac{\partial \log L_n(\delta)}{\partial \delta_{kh}} \\
\leq \left[ \frac{(n||l||^4)^{-1} \{ (L^2 \psi_\nu^2)/4 + o_p(1) \} + O_p(n^{-1}) + o_p(||l||^2)^{-1/2} ||\delta_k - \delta_k^\nu||^2}{\psi_\nu^2} \right] + \left[ \{ O_p(n^{-1/2} ||l||^{-2}) + o_p(1) - L \} + \{ o_p((n^{1/2} ||l||^2)^{-1}) + o_p(1) \} \right] ||\delta_k - \delta_k^\nu||^2 \\
= \left[ \frac{L}{2} \left( 1 + \frac{o_p(1) + o_p(||l||^4) + o_p(n||l||^6)}{\psi_\nu^2} \right) \right]^{1/2} \left( \frac{L}{2} \left( 1 + \frac{o_p(1) + o_p(||l||^4) + o_p(n||l||^6)}{\psi_\nu^2} \right) \right)^{-1/2} \\
+ O_p(n^{-1/2} ||l||^{-2}) + o_p(1) - L + o_p(n^{-1/2} ||l||^{-2}) + o_p(1) ||\delta_k - \delta_k^\nu||^2
\]
Hence, taking the supremum in \( A \), it follows that as \( n \) becomes large, the above
expression will have increasing probability of assuming the sign of \( L/2 - L < 0 \), and
consequently condition (4.18) is verified. This completes the proof for (4.11) and
for the theorem.

The weak convergence is considered with the same underlying idea as seem
earlier. In other words, we work with an integrated process, express it in terms of two

97
terms involving the function \( \delta^\nu \) and show that one term converges to a Gaussian process and the other converges in probability to 0. The result is given by the following theorem.

**Theorem 4.2** Assume that together with conditions A(1), A(3), B, C(1), and D, we have

1. \( \lim n \| l \|^6 < \infty \),
2. \( \lim n \| l \|^4 = \infty \),
3. \( \limsup l(\nu)/l(1) < \infty \).

Then, the integrated process \( n^{1/2} \int_0^t (\delta_h(u) - \delta^\nu_h(u)) \, du \) converges in distribution to a continuous multivariate Gaussian martingale \( W \), such that \( W(0) = 0 \) and

\[
\text{Cov}(W(t), W(r)) = \int_0^{t \wedge r} \left[ v_h(u) s_k^0(\delta^\nu, u) \sigma_k(u) \right]^{-1} \, du, \quad r, t \in \mathcal{T}.
\]

**Proof:** Initially we write

\[
n^{1/2} \int_0^t \left( \delta(u) - \delta^\nu(u) \right) \, du = n^{1/2} \int_0^t \left( \delta(u) - \bar{\delta}^\nu(u) \right) \, du
\]

\[
+ n^{1/2} \int_0^t \left( \bar{\delta}^\nu(u) - \delta^\nu(u) \right) \, du \tag{4.26}
\]

For expression (4.27) we have that

\[
n^{1/2} \int_0^t \left( \bar{\delta}^\nu(u) - \delta^\nu(u) \right) \, du = n^{1/2} \int_0^t \sum_{h=1}^r \left( \delta_h^\nu - \delta^\nu(u) \right) \Pi_h^\nu(u) \, du.
\]

Let \( u_h \) be such that, for a constant \( c \), \( \delta^\nu_h = \delta^\nu(u_h) + c l_h^2 \). Using condition D(3) we Taylor expand \( \delta^\nu(u) \) around \( u_h \), such that, for \( |u - u^*| \leq |u - u_h| \),

\[
\delta^\nu(u) = \delta^\nu(u_h) + \frac{d\delta^\nu(u_h)}{du}(u - u_h) + \frac{1}{2} \frac{d^2\delta^\nu(u^*)}{du^2}(u - u_h)^2.
\]
Hence, using conditions in D, we can write
\[
\left\| n^{1/2} \int_0^t \left( \delta^v(u) - \delta^o(u) \right) \, du \right\|^2 \leq n \int_0^t \left\| \delta^v(u) - \delta^o(u) \right\|^2 \, du \\
= n \int_0^t \sum_{h=1}^\nu \left\| \delta_h^v - \delta_h^o \right\|^2 I_h^v(u) \, du \leq n \int_0^t \sum_{h=1}^\nu \left( c l_h^o + O(1) l_h^o + O(1) l_h^o \right) I_h^v(u) \, du \\
\leq n\|l\|^8 O(1) \int_0^\nu \sum_{h=1}^\nu I_h^v(u) \, du = O(n\|l\|^8).
\]

Therefore we conclude that \( \sup_{u \in T} \left\| n^{1/2} \int_0^t (\delta^v(u) - \delta^o(u)) \, du \right\| = O(n^{1/2}\|l\|^4) \).

As for expression (4.27) we first consider, for each interval \( I_h^v, h = 1, \ldots, \nu \), the function
\[
\lambda(\xi_h) = \log L_n(\delta^v + (nl_h)^{-1/2} \xi_h) - \log L_n(\delta^v), \quad \|\xi_h\| \leq K \in [0, \infty).
\]

For \( u \in I_h^v \), consider the Taylor expansion of \( \log L_n(\delta^v + (nl_h)^{-1/2} \xi_h) \) around \( \delta^v(u) \)
\[
\log L_n(\delta^v + (nl_h)^{-1/2} \xi_h) = \log L_n(\delta^v) + (nl_h)^{-1/2} \xi_h^T \frac{\partial \log L_n(\delta^v)}{\partial \delta_h} \\
+ \frac{1}{2} (nl_h)^{-1} \xi_h^T \frac{\partial^2 \log L_n(\delta^v)}{\partial \delta_h \partial \delta_h} \xi_h,
\]
for \( \|\delta^* - \delta^v(u)\| \leq \|(nl_h)^{-1/2} \xi_h\| \), so that, using Lemma 4.2 and assumption 2,
\[
\lambda(\xi_h) = (nl_h)^{-1/2} \xi_h^T \frac{\partial \log L_n(\delta^v)}{\partial \delta_h} + \frac{1}{2n l_h} \xi_h^T \frac{\partial^2 \log L_n(\delta^v)}{\partial \delta_h \partial \delta_h} \xi_h \\
= (nl_h)^{-1/2} \xi_h^T \left[ (nl_h)^{-1} \frac{\partial^2 \log L_n(\delta^v)}{\partial \delta_h \partial \delta_h} + l_h^{-1} \Sigma_h \right] \xi_h \\
+ (1/2) \xi_h^T \left( (nl_h)^{-1} \frac{\partial^2 \log L_n(\delta^v)}{\partial \delta_h \partial \delta_h} + l_h^{-1} \Sigma_h \right) \xi_h,
\]
where \( \Sigma_h \) is a block diagonal matrix of dimension \( 4 \times 4 \), with matrices in the main diagonal given by \( \Sigma_{kh}, k = 1, 2 \), as defined on page 93. Now define \( U_h(\cdot) = (\partial/\partial \delta_h) \log L_n(\cdot) \). Then we can write
\[
\lambda(\xi_h) = (nl_h)^{-1/2} \xi_h^T U_h(\delta^v) - (1/2) l_h^{-1} \Sigma_h \xi_h + o_p(1).
\]
The value \( \hat{\xi}_h \) that maximizes \( \lambda(\xi_h) \) will be given by \( \hat{\xi}_h = (n^{-1} l_h)^{1/2} \Sigma_h^{-1} U_h(\delta^\nu) + o_p(1) \), so that, by the definition of \( \lambda(\cdot) \), such a value also corresponds to the value that maximizes \( \log L_n(\delta^\nu + (n l_h)^{-1/2} \xi_h) \) in \( I_h^\nu \), given by \( \hat{\delta}_h \), the maximum partial likelihood estimator of \( \delta(u) \in S_\nu \). Therefore,

\[
\hat{\delta}_h = \delta^\nu_h + (n^{-1} l_h)^{1/2}(n l_h)^{-1/2} \Sigma_h^{-1} U_h(\delta^\nu) + o_p(1)
\]

\[
= \delta^\nu_h + n^{-1} \Sigma_h^{-1} U_h(\delta^\nu) + o_p(1),
\]

and hence, we write the integrated process as

\[
n^{1/2} \int_0^t \left( \tilde{\delta}(u) - \delta^\nu(u) \right) \, du = n^{1/2} \sum_{h=1}^{\nu} \int_0^t \left( \hat{\delta}_h - \delta^\nu_h \right) \Pi_h^\nu(u) \, du + o_p(1)
\]

\[
= n^{1/2} \sum_{h=1}^{\nu} \int_0^t n^{-1} \Sigma_h^{-1} U_h(\delta^\nu) \Pi_h^\nu(u) \, du + o_p(1)
\]

\[
= n^{-1/2} \int_0^t \sum_{h=1}^{\nu} \Pi_h^\nu(u) \Sigma_h^{-1} U_h(\delta^\nu) \, du + o_p(1)
\]

As we describe below, the first term in the last expression above can be thought of as a **piecewise linear approximation.** Such approach was considered by Murphy and Sen (1991) and is based on Lemma 4.1 of McKeague (1988). The basic idea is that if we are able to show that a given function converges weakly to a Gaussian process, then its piecewise linear approximation will also (weakly) converge to a Gaussian process, as long as \( \max(l_1, \ldots, l_\nu) \to 0 \), what is the case here. In order to make this clear, we note that for \( t \in I_r^\nu = [a_{r-1}, a_r) \),

\[
n^{-1/2} \int_0^t \sum_{h=1}^{r-1} \Pi_h^\nu(u) \Sigma_h^{-1} U_h(\delta^\nu) \, du
\]

\[
= n^{-1/2} \sum_{h=1}^{r-1} \int_0^t \Pi_h^\nu(u) \, du \Sigma_h^{-1} U_h(\delta^\nu) + \int_{a_{r-1}}^t \, du \Sigma_r^{-1} U_r(\delta^\nu)
\]

\[
= n^{-1/2} \sum_{h=1}^{r-1} l_h \Sigma_h^{-1} U_h(\delta^\nu) + \frac{t - a_{r-1}}{l_r} \Sigma_r^{-1} U_r(\delta^\nu)
\]

\[
= X(a_{r-1}) + \left( \frac{t - a_{r-1}}{l_r} \right) (X(a_r) - X(a_{r-1}))
\]

(4.28)
where $X(t) = (X_1(t), X_2(t))^T$ is a 4-vector with [recall the definition of $f_{ki}$ in (4.4)]

$$X_k(t) = n^{-1/2} \sum_{i=1}^n \int_0^t \sum_{h=1}^\nu l_h \Sigma^{-1}_{kh} \frac{\partial f_{ki}(u, \tilde{\delta}^\nu)}{\partial \delta_{kh}} \, dN_{ki}(u). \tag{4.29}$$

Hence, using the terminology considered in McKeague (1988), the expression given in (4.28) is the piecewise linear approximation for (4.29), and hence, the weak convergence of the latter will imply the weak convergence of the former to a Gaussian process. Therefore, we concentrate now in determining the asymptotic behavior of (4.29). Hence, using the Doob-Meier decomposition we have that

$$X_k(t) = n^{-1/2} \sum_{i=1}^n \int_0^t \sum_{h=1}^\nu l_h \Sigma^{-1}_{kh} \frac{\partial f_{ki}(u, \tilde{\delta}^\nu)}{\partial \delta_{kh}} \, dM_{ki}(u) \tag{4.30}$$

$$+ n^{-1/2} \sum_{i=1}^n \int_0^t \sum_{h=1}^\nu l_h \Sigma^{-1}_{kh} \frac{\partial f_{ki}(u, \tilde{\delta}^\nu)}{\partial \delta_{kh}} \, d\Lambda_{ki}(u).$$

For the second factor in the r.h.s. of (4.30) we have that, for $\lambda_1(\cdot)$ representing the largest eigenvalue of the matrix taken as its argument,

$$\left\| n^{-1/2} \sum_{i=1}^n \int_0^t \sum_{h=1}^\nu l_h \Sigma^{-1}_{kh} \frac{\partial f_{ki}(u, \tilde{\delta}^\nu)}{\partial \delta_{kh}} \, d\Lambda_{ki}(u) \right\|^2$$

$$\leq n^{-1} \sum_{h=1}^\nu \left\| l_h^2 \Sigma^{-1}_{kh} \sum_{i=1}^n \int_0^t \frac{\partial f_{ki}(u, \tilde{\delta}^\nu)}{\partial \delta_{kh}} \, d\Lambda_{ki}(u) \right\|^2 \tag{by (4.13)}$$

$$\leq n^{-1} \sum_{h=1}^\nu l_h^2 \lambda_1(\Sigma_{kh}) \left\| \sum_{i=1}^n \int_0^t \frac{\partial f_{ki}(u, \tilde{\delta}^\nu)}{\partial \delta_{kh}} \, d\Lambda_{ki}(u) \right\|^2 \tag{by (4.14)}$$

$$\leq n^{-1} \sum_{h=1}^\nu l_h^2 O(l_h^2) \left\| \sum_{i=1}^n \int_0^t \frac{\partial f_{ki}(u, \tilde{\delta}^\nu)}{\partial \delta_{kh}} \, d\Lambda_{ki}(u) \right\|^2$$

$$\leq O(n^{-1} \sum_{h=1}^\nu (nl_h)^2 \left\| \sum_{i=1}^n \int_0^t \frac{\partial f_{ki}(u, \tilde{\delta}^\nu)}{\partial \delta_{kh}} \, d\Lambda_{ki}(u) \right\|^2) \leq O(n\|l\|^4) \sum_{h=1}^\nu \left[ (nl_h)^{-1} \left\| \sum_{i=1}^n \int_0^t \frac{\partial f_{ki}(u, \tilde{\delta}^\nu)}{\partial \delta_{kh}} \, d\Lambda_{ki}(u) \right\|^2. \tag{using (3))}$$

Note that the last term above is equal to the quantity $X(t)$ defined in expression (4.38), on page 109; that expression was shown to be of order $O_p(1/n) + o_p(\|l\|^2)$. Therefore, we may conclude that

$$\left\| n^{-1/2} \sum_{i=1}^n \int_0^t \sum_{h=1}^\nu l_h \Sigma^{-1}_{kh} \frac{\partial f_{ki}(u, \tilde{\delta}^\nu)}{\partial \delta_{kh}} \, d\Lambda_{ki}(u) \right\|^2 = O_p(\|l\|^4) + o_p(n\|l\|^6).$$

101
As for the first term in (4.30), denote its predictable variation process by \( \langle Z(t) \rangle \). If we can show that (a) \( \langle Z(t) \rangle \) converges to a continuous function, and (b) the predictable variation related to the jump process converges to zero in probability then, by the Rebolledo’s central limit theorem, that factor converges to a Gaussian process. The predictable variation process is given by

\[
\langle Z(t) \rangle = n^{-1} \sum_{h=1}^{n} \sum_{i=1}^{\nu} \int_{0}^{t} l_{k}^{2} \left[ \Sigma^{-1}_{kh} \frac{\partial f_{ki}(u, \bar{\delta}^{\nu})}{\partial \delta_{kh}} \right] \phi_{i}(u, \theta_{k}) e^{\theta_{k}(u)X_{i}(u)} \alpha_{k}(u) \, du \\
= \sum_{h=1}^{\nu} \int_{0}^{t} I_{h}^{2}(u) l_{h}^{2} \Sigma^{-1}_{kh} A(u) \Sigma^{-1}_{kh} \alpha_{k}(u) \, du,
\]

where the matrix \( A = (a_{ij}) \) has elements given by

\[
a_{11}(u) = V_{\theta_{k}}(u)S_{k}^{(0)}(\delta^{\circ}, u) + [E_{\theta}(\delta_{k}, u) - E_{\theta}(\bar{\delta}^{\nu}_{k}, u)]^{2} S_{k}^{(0)}(\delta^{\circ}, u) \\
a_{12}(u) = a_{21}(u) = C_{\delta_{k}}(u)S_{k}^{(0)}(\delta^{\circ}, u) + \left\{ \left( \frac{1 + \theta_{k}(u)}{1 + \theta_{k}^{\nu}(u)} - 1 \right) C_{\delta_{k}}(u) \\
+ \left( \frac{1 + \theta_{k}^{\nu}(u)}{1 + \theta_{k}(u)} \right) [E_{\theta}(\delta_{k}, u) - E_{\theta}(\bar{\delta}^{\nu}_{k}, u)][E_{\theta}(\bar{\delta}^{\nu}_{k}, u) - E_{\theta}(\delta_{k}, u)] \right\} S_{k}^{(0)}(\delta^{\circ}, u) \\
a_{22}(u) = V_{\theta_{k}}(u)S_{k}^{(0)}(\delta^{\circ}, u) + \left\{ \left( \frac{1 + \theta_{k}(u)}{1 + \theta_{k}^{\nu}(u)} - 1 \right) V_{\theta_{k}}(u) \\
+ \left( \frac{1 + \theta_{k}^{\nu}(u)}{1 + \theta_{k}(u)} \right) E_{\theta}(\delta_{k}, u) - E_{\theta}(\bar{\delta}^{\nu}_{k}, u) \right\}^{2} S_{k}^{(0)}(\delta^{\circ}, u).
\]

Therefore, using conditions A(1), A(3), C(1), and D(2),

\[
A(u) = V_{k}(u)S_{k}^{(0)}(\delta^{\circ}, u) + o_{p}(1),
\]

and hence,

\[
\langle Z(t) \rangle \\
= \sum_{h=1}^{\nu} \int_{0}^{t} I_{h}^{2}(u) l_{h}^{2} \Sigma^{-1}_{kh} V_{k}(u)S_{k}^{(0)}(\delta^{\circ}, u) \Sigma^{-1}_{kh} \alpha_{k}(u) \, du + \sum_{h=1}^{\nu} \int_{0}^{t} I_{h}^{2}(u) l_{h}^{2} \Sigma^{-2}_{kh} \, du o_{p}(1).
\]

As seen in (4.14), each element of \( \Sigma^{-2}_{kh} \) is \( O(l_{h}^{-2}) \), so that the second term in the r.h.s. above is \( o_{p}(1) \). As for the first term we use the continuity of the elements in
\( v_k(u)S_k^{(0)}(\delta^o, u)\alpha_k^o(u) \) to claim that

\[
\max_{1 \leq h \leq \nu} \sup_{u \in I_h} |l_h^{-1} (\Sigma_{kh})_{ij} - (v_k(u)S_k^{(0)}(\delta^o, u)\alpha_k^o(u))_{ij}| \to 0,
\]

so that the matrix \( l_h^{-1} \Sigma_{kh} - v_k(u)S_k^{(0)}(\delta^o, u)\alpha_k^o(u) \) converges to a null matrix as \( n \to \infty \). Therefore,

\[
\sum_{h=1}^{\nu} \int_0^t I_h^r(u) l_h \Sigma_{kh}^{-1} V_k(u) S_k^{(0)}(\delta^o, u) \Sigma_{kh}^{-1} l_h \alpha_k^o(u) \, du = \sum_{h=1}^{\nu} \int_0^t I_h^r(u) l_h \Sigma_{kh}^{-1} \left( V_k(u) S_k^{(0)}(\delta^o, u) - v_k(u)S_k^{(0)}(\delta^o, u) \right) \Sigma_{kh}^{-1} l_h \alpha_k^o(u) \, du + \sum_{h=1}^{\nu} \int_0^t I_h^r(u) l_h \Sigma_{kh}^{-1} v_k(u)S_k^{(0)}(\delta^o, u) \alpha_k^o(u) \Sigma_{kh}^{-1} l_h \, du
\]

\[
= o_p(1) \sum_{h=1}^{\nu} \int_0^t I_h^r(u) l_h^2 \Sigma_{kh}^{-2} \, du + \sum_{h=1}^{\nu} \int_0^t I_h^r(u) l_h \Sigma_{kh}^{-1} \left( v_k(u)S_k^{(0)}(\delta^o, u)\alpha_k^o(u) - l_h^{-1} \Sigma_{kh} \right) \Sigma_{kh}^{-1} l_h \, du + \sum_{h=1}^{\nu} \int_0^t I_h^r(u) l_h \Sigma_{kh}^{-1} l_h^{-1} \Sigma_{kh} \Sigma_{kh}^{-1} l_h \, du
\]

\[
= o_p(1) + o_p(1) \sum_{h=1}^{\nu} \int_0^t I_h^r(u) l_h^2 \Sigma_{kh}^{-2} \, du + \sum_{h=1}^{\nu} \int_0^t I_h^r(u) \left( l_h \Sigma_{kh}^{-1} - \left[ v_k(u)S_k^{(0)}(\delta^o, u)\alpha_k^o(u) \right]^{-1} \right) \, du + \sum_{h=1}^{\nu} \int_0^t I_h^r(u) \left[ v_k(u)S_k^{(0)}(\delta^o, u)\alpha_k^o(u) \right]^{-1} \, du
\]

\[
= \int_0^t \left[ v_k(u)S_k^{(0)}(\delta^o, u)\alpha_k^o(u) \right]^{-1} \, du
\]

In other words, we have that

\[
\langle Z(t) \rangle \overset{P}{\to} \int_0^t \left[ v_k(u)S_k^{(0)}(\delta^o, u)\alpha_k^o(u) \right]^{-1} \, du,
\]

a continuous function. That takes care of (a). With respect to (b), the manipulation is more complicated as we have to consider each element of the vector of jump
process associated with the second factor in the r.h.s. of (4.30) and show that their predictable variation process converge to zero in probability. In order to proceed, we define first the process $H_i$, a 2-vector, such that $H_i = (H_i^{(1)}(u), H_i^{(2)}(u))^\tau = \sum_{h=1}^n \Sigma^{-1}_{kh} \partial \delta u}_h f_{ki}(u, \delta^\nu)$, where

$$H_i^{(1)}(u) = \sum_{h=1}^n n^{-1/2} l_h |\Sigma_{kh}|^{-1} [\sigma_{h2} f'_{\delta^\nu}(u) - \sigma_{h12} f'_i(u)]$$

$$H_i^{(2)}(u) = \sum_{h=1}^n n^{-1/2} l_h |\Sigma_{kh}|^{-1} [\sigma_{h1} f'_{\delta^\nu}(u) - \sigma_{h12} f'_i(u)],$$

and the functions $f'_i$ correspond to the partial derivatives of $f_{ki}$ w.r.t. the elements of $\delta$. Based on those quantities we define the jump processes $Z_j(t)$ as the 2-vector with elements

$$Z_j(t) = \sum_{i=1}^n \int_0^t H_i^{(j)}(u) I\{H_i^{(j)}(u) > \epsilon\} dM_{ki}(u), \quad j = 1, 2.$$

The corresponding predictable variation process is then given by

$$\langle \epsilon Z_j(t) \rangle = \sum_{i=1}^n \int_0^t (H_i^{(j)}(u))^2 I\{H_i^{(j)}(u) > \epsilon\} w_i(u, \theta_k^\nu) e^{\delta^\nu X_i(u)} \alpha_k(u) du, \quad j = 1, 2.$$

Using the inequality $|a - b|^2 I\{|a - b| > \epsilon\} \leq 4|a|^2 I\{|a| > \epsilon/2\} + 4|b|^2 I\{|b| > \epsilon/2\}$, we have that for the first element of the predictable variation (jump) process,

$$\langle \epsilon Z_1(t) \rangle \leq 4 \sum_{i=1}^n \sum_{h=1}^n \int_0^t n^{-1/2} l_h^2 |\Sigma_{kh}|^{-2} [\sigma_{h2} f'_{\delta^\nu}(u)]^2 w_i(u, \theta_k^\nu) e^{\delta^\nu X_i(u)}$$

$$\times I\{\sum_{h=1}^n n^{-1/2} l_h |\Sigma_{kh}|^{-1} \sigma_{h2} f'_{\delta^\nu}(u) > \epsilon\} \alpha_k^\nu(u) du$$

$$+ 4 \sum_{i=1}^n \sum_{h=1}^n \int_0^t n^{-1/2} l_h^2 |\Sigma_{kh}|^{-2} [\sigma_{h12} f'_i(u)]^2 w_i(u, \theta_k^\nu) e^{\delta^\nu X_i(u)}$$

$$\times I\{\sum_{h=1}^n n^{-1/2} l_h |\Sigma_{kh}|^{-1} \sigma_{h12} f'_i(u) > \epsilon\} \alpha_k^\nu(u) du.$$

(4.31)
The first term in the r.h.s. above equals

\[ \int_0^\tau \sum_{h=1}^\nu \int_h(u)l_h^2|\Sigma_{kh}|^{-2}\sigma_{kh}^2n^{-1}\sum_{i=1}^n(X_i(u) - E_\beta(\delta_k, u))^2 w_i(u, \theta_k^0)e^{\beta_k^0(u)X_i(u)} \times \mathbb{I}\{|X_i(u) - E_\beta(\delta_k, u)| \sum_{h=1}^\nu \int_h(u)l_h|\Sigma_{kh}|^{-1}\sigma_{kh} > \epsilon n^{1/2}\} \alpha_k^0(u) \, du. \]

But we have seen that \(|\Sigma_{kh}|^{-2}\sigma_{kh}^2 = O(\iota_h^{-2})\), so that \(\sum_{h=1}^\nu \int_h(u)l_h^2|\Sigma_{kh}|^{-2}\sigma_{kh}^2 = O(1)\).

Thus, there exists a constant \(C > 0\) such that the above expression is less than or equal to

\[ C \int_0^\tau n^{-1}\sum_{i=1}^n(X_i(u) - E_\beta(\delta_k, u))^2 w_i(u, \theta_k^0)e^{\beta_k^0(u)X_i(u)} \times \mathbb{I}\{|X_i(u) - E_\beta(\delta_k, u)| > \frac{\epsilon n^{1/2}}{C}\} \alpha_k^0(u) \, du \]

\[ \leq 4C \int_0^\tau n^{-1}\sum_{i=1}^n X_i^2(u) w_i(u, \theta_k^0)e^{\beta_k^0(u)X_i(u)} \mathbb{I}\{|X_i(u)| > \frac{\epsilon n^{1/2}}{2C}\} \alpha_k^0(u) \, du \]

\[ + \int_0^\tau n^{-1} E_\beta^2(\delta_k, u) \mathbb{I}\{|E_\beta(\delta_k, u)| > \frac{\epsilon n^{1/2}}{2C}\} S_k^{(0)}(\delta_k, u) \alpha_k^0(u) \, du \]

\[ = 4C \int_0^\tau n^{-1}\sum_{i=1}^n X_i^2(u) w_i(u, \theta_k^0)e^{\beta_k^0(u)X_i(u)} \mathbb{I}\{|X_i(u)| > \frac{\epsilon n^{1/2}}{2C}\} \alpha_k^0(u) \, du + o_p(1), \]

using condition A(1) and A(3) in the last step. Showing that the first term in the r.h.s. above is also \(o_p(1)\) will follow by considering the standard derivation presented in, e.g., Andersen and Gill (1982). That is accomplished by writing such an expression as

\[ \int_0^\tau n^{-1}\sum_{i=1}^n X_i^2(u) w_i(u, \theta_k^0)e^{\beta_k^0(u)X_i(u)} \times \mathbb{I}\{|X_i(u)| > \epsilon n^{1/2}, \beta_k^0(u)X_i(u) \leq -\psi |X_i(u)|\} \alpha_k^0(u) \, du \]

\[ + \int_0^\tau n^{-1}\sum_{i=1}^n X_i^2(u) w_i(u, \theta_k^0)e^{\beta_k^0(u)X_i(u)} \times \mathbb{I}\{|X_i(u)| > \epsilon n^{1/2}, \beta_k^0(u)X_i(u) > -\psi |X_i(u)|\} \alpha_k^0(u) \, du \]

105
\begin{align}
&\leq \int_0^\tau (1 + \theta^o_k(u))n^{-1}\sum_{i=1}^n X_i^2(u)e^{-\psi|X_i(u)|}\mathbb{I}\{|X_i(u)| > \epsilon n^{1/2}\} \alpha^o_k(u) \, du \\
&\quad + \int_0^\tau n^{-1}\sum_{i=1}^n X_i^2(u)\mathbb{I}\{|X_i(u)| > \epsilon n^{1/2}, \beta^o_k(u)X_i(u) > -\psi|X_i(u)|\} Y_{ki}(u) \times w_i(u, \theta^o_k) e^{\beta^o_k(u)X_i(u)} \alpha^o_k(u) \, du
\end{align}

(4.33)

Since $X^2_i e^{-\psi X_i} \to 0$ as $X_i \to \infty$, than for $n$ large enough, there exists $\eta > 0$ such that expression (4.32) is bounded by $\eta \int_0^\tau (1 + \theta^o_k) \alpha^o_k \, du$. Since $\eta$ is arbitrary, it follows that (4.32) is $o_p(1)$. On the other hand, by condition B we have that, as $n \to \infty$, 

$$
P\{\exists i, u: n^{-1/2}|X_i(u)| > \epsilon, \beta^o_k(u)X_i(u) > -\psi|X_i(u)|, Y_{ki}(u) = 1\} \to 0.$$

Therefore, we may conclude that (4.33) is also $o_p(1)$.

As for the second term in (4.31), we follow the same steps above to write

$$
\sum_{i=1}^n \sum_{h=1}^\nu \int_0^\tau n^{-1} l^2_h |\Sigma_{kh}|^{-2} [\sigma_{h12} f_i' \psi(u)]^2 w_i(u, \theta^o_k) e^{\beta^o_k(u)X_i(u)}
$$

$$
\times \mathbb{I}\{|\sum_{h=1}^\nu n^{-1/2} l_h |\Sigma_{kh}|^{-1} \sigma_{h12} |f_i' \psi(u)| > \epsilon\} \alpha^o_k(u) \, du
$$

$$
\leq C \int_0^\tau n^{-1} \sum_{i=1}^n \left(\frac{Y_{1i}(u)Y_{2i}(u)}{w_i(u, \theta^o_k)} - E_0(\tilde{\xi}^\nu, u)\right)^2 w_i(u, \theta^o_k) e^{\beta^o_k(u)X_i(u)}
$$

$$
\times \mathbb{I}\{|\frac{Y_{1i}(u)Y_{2i}(u)}{w_i(u, \theta^o_k)} - E_0(\tilde{\xi}^\nu, u)| > \frac{\epsilon n^{1/2}}{C}\} \alpha^o_k(u) \, du
$$

$$
\leq 4C \int_0^\tau n^{-1} \sum_{i=1}^n \left(\frac{Y_{1i}(u)Y_{2i}(u)}{w_i(u, \theta^o_k)}\right)^2 w_i(u, \theta^o_k) e^{\beta^o_k(u)X_i(u)} \mathbb{I}\{|\frac{Y_{1i}(u)Y_{2i}(u)}{w_i(u, \theta^o_k)}| > \frac{\epsilon n^{1/2}}{2C}\} \alpha^o_k(u) \, du
$$

$$
+ o_p(1)
$$

$$
\leq O(1) \int_0^\tau \mathbb{I}\{\frac{1}{1 + \theta^o_k(u)} O(1) > \frac{\epsilon n^{1/2}}{2C}\} \alpha^o_k(u) \, du + o_p(1)
$$

$$
= o_p(1),
$$

where the last step follows by condition D(2). Hence, we conclude that $\langle e Z_1(t) \rangle = o_p(1)$.

106
Given the similarity between the components $\langle \epsilon Z_1(t) \rangle$ and $\langle \epsilon Z_2(t) \rangle$, the same manipulation can be considered for the latter so that we also conclude that $\langle \epsilon Z_2(t) \rangle = o_p(1)$. Therefore,

$$\langle \epsilon Z(t) \rangle = o_p(1),$$

and hence, (b) is also verified. As a consequence, the Rebolledo’s central limit theorem applies and this concludes the proof of the theorem. □

Using the result in Theorem 4.2 we will be able to develop test statistics and confidence intervals as long as we estimate the covariance matrix. Based on the classical development for maximum likelihood estimators, a natural choice would be to consider the second derivative and use the expression

$$\overline{\text{Cov}}_k \{W(s), W(t)\} = \int_0^{\tau} \left[ -\sum_{h=1}^\nu \frac{1}{n} \sum_{i=1}^n \frac{1}{(n l_h)^{-1}} \frac{\partial^2 \log L_n(\hat{\delta})}{\partial \delta_{kh} \partial \delta_{kh}^t} \right]^{-1} du.$$

(4.34)

The consistency for (4.34) is considered in the following theorem.

**Theorem 4.3** Assume that conditions A(1), A(3), C, D(1), and D(2) are true, as well as

1. $\lim_{n \to \infty} n||l||^4 = \infty$, and

2. $\lim_{n \to \infty} l(\nu)/l(1) < \infty$.

Then,

$$\int_0^t \left[ -\sum_{h=1}^\nu \frac{1}{n} \sum_{i=1}^n \frac{1}{(n l_h)^{-1}} \frac{\partial^2 \log L_n(\hat{\delta})}{\partial \delta_{kh} \partial \delta_{kh}^t} \right]^{-1} du \overset{p}{\longrightarrow} \int_0^t \left[ v_k(u) s_k^{(0)}(\delta^0, u) \alpha_k^2(u) \right]^{-1} du,$$

(4.35)

uniformly in $t$.  

107
Proof: Let $D_{2h}(\delta)$ be (minus) the second derivative of the log-likelihood w.r.t. $\delta_k$, computed at $\delta \in S_\nu$, and let $\Sigma_k$ be the matrix defined on page 93. Then, we write

$$
\int_0^t \left[ \sum_{h=1}^\nu I_k^\nu(u)(nl_h)^{-1}D_{2h}(\delta) - v_k(u)s_k^{(0)}(\delta, u)\alpha_k^\nu(u) \right] du
$$

$$
= \int_0^t \left[ \sum_{h=1}^\nu I_k^\nu(u)(nl_h)^{-1}[D_{2h}(\delta) - D_{2h}(\delta^\nu)] \right] du
$$

$$
+ \int_0^t \sum_{h=1}^\nu I_k^\nu(u) \left[ (nl_h)^{-1}D_{2h}(\delta^\nu) - l_h^{-2}\Sigma_k \right] du
$$

$$
+ \int_0^t \sum_{h=1}^\nu I_k^\nu(u) \left[ l_h^{-2}\Sigma_k - v_k(u)s_k^{(0)}(\delta, u)\alpha_k^\nu(u) \right] du
$$

$$
= \int_0^t \sum_{h=1}^\nu I_k^\nu(u)(nl_h)^{-1}M_{1h}(\delta^\nu, \delta^\nu) du + \int_0^t \sum_{h=1}^\nu I_k^\nu(u)M_{2h}(\delta^\nu, \delta^\nu) du + o_p(1),
$$

where the third term is a consequence of the continuity of $v_k(\delta, u)s_k^{(0)}(\delta, u)\alpha_k^\nu(u)$ in $u$. Using the notation $M_{1h}^+$ to indicate the matrix with elements equal to the absolute value of the elements of $M_{1h}$, $i = 1, 2$, we have that

$$
\sup_{t \in \mathbb{T}} \left| I^\nu \int_0^t \left[ \sum_{h=1}^\nu I_k^\nu(u)(nl_h)^{-1}D_{2h}(\delta) - v_k(u)s_k^{(0)}(\delta, u)\alpha_k^\nu(u) \right] du \right|
$$

$$
\leq \int_0^t \sum_{h=1}^\nu I_k^\nu(u)(nl_h)^{-1}I^\nu M_{1h}^+(\delta, \delta^\nu)1 du + \int_0^t \sum_{h=1}^\nu I_k^\nu(u)I^\nu M_{2h}^+(\delta^\nu, \delta^\nu)1 du + o_p(1)
$$

$$
\leq \tau \max_{1 \leq h \leq \nu} I^\nu (nl_h)^{-1}M_{1h}^+(\delta, \delta^\nu)1 du + \tau \max_{1 \leq h \leq \nu} I^\nu M_{2h}^+(\delta^\nu, \delta^\nu)1 du + o_p(1)
$$

$$
= O_p(n^{-1/2}\|l\|^{-2}) + o_p(1),
$$

by Lemma 4.2 (applied to the factor involving $M_{2h}^+$) and Lemma 4.3 (for the factor with $M_{1h}^+$). By assumption 1 it follows that the term above is $o_p(1)$. Therefore, since by condition C(2) $v_k$ is positive definite uniformly in $u$, the result given in the theorem follows. \qed
5 Proofs

In this section we present the technical details of the proofs for the lemmas used in Section 4. Since some of the proofs are long, we consider each lemma as a subsection.

5.1 Proof for Lemma 4.1

The proof follows quite similar to the one presented by Murphy and Sen (1991).

Using the function $f_{ki}$ defined in (4.4), we have that for $t \in J$,

$$
\sum_{h=1}^{\nu} \left\{ (n_l)^{-1} \left\| \frac{\partial \log L_n(t; \delta')}{\partial \delta_{kh}} \right\| \right\}^2 = \sum_{h=1}^{\nu} \left\{ (n_l)^{-1} \left\| \sum_{j=1}^n \int_0^t \frac{\partial f_{kj}(\delta', u)}{\partial \delta_{kh}} \, dN_{kj}(u) \right\| \right\}^2 
$$

$$
= \sum_{h=1}^{\nu} \left\{ (n_l)^{-1} \left\| \sum_{j=1}^n \int_0^t \frac{\partial f_{kj}(\delta', u)}{\partial \delta_{kh}} \, d(M_{kj}(u) + d\Lambda_{kj}(u)) \right\| \right\}^2 
$$

$$
\leq 2 \sum_{h=1}^{\nu} \left\{ (n_l)^{-1} \left\| \sum_{j=1}^n \int_0^t \frac{\partial f_{kj}(\delta', u)}{\partial \delta_{kh}} \, dM_{kj}(u) \right\| \right\}^2 
$$

$$
+ 2 \sum_{h=1}^{\nu} \left\{ (n_l)^{-1} \left\| \sum_{j=1}^n \int_0^t \frac{\partial f_{kj}(\delta', u)}{\partial \delta_{kh}} \, d\Lambda_{kj}(u) \right\| \right\}^2 
$$

(4.36)

The proof now is divided in two parts: (i) show that (4.37) is $O_p(1/n) + o_p(\|l\|^2)$ and (ii) show that (4.36) is less than or equal to the first quantity in the r.h.s. of (4.10).

For part (i), let us consider first expression (4.37). Denoting its value divided by 2 as $X(t)$, we have

$$
X(t) = \sum_{h=1}^{\nu} \left\{ (n_l)^{-1} \left\| \sum_{j=1}^n \int_0^t \frac{\partial f_{kj}(\delta', u)}{\partial \delta_{kh}} \, d\Lambda_{kj}(u) \right\| \right\}^2 
$$

$$
= \sum_{h=1}^{\nu} \left\{ (n_l)^{-1} \left\| \sum_{j=1}^n \int_0^t \frac{\partial f_{kj}(\delta', u)}{\partial \beta_{kh}} \, d\Lambda_{kj}(u) \right\| \right\}^2 
$$

$$
+ \sum_{h=1}^{\nu} \left\{ (n_l)^{-1} \left\| \sum_{j=1}^n \int_0^t \frac{\partial f_{kj}(\delta', u)}{\partial \theta_{kh}} \, d\Lambda_{kj}(u) \right\| \right\}^2 
$$

$$
= X_\delta(t) + X_\theta(t). 
$$

(4.38)
For $X_\beta$, we have that such an expression equals

$$X_\beta(t) = \sum_{h=1}^{\nu} \left\{ l_h^{-1} \int_0^t \mathbb{I}_h^\nu(u)[E_\beta(\delta_k^\nu(u) - \delta_k^\nu(u)] S_k^{(0)}(\delta^\nu, u) \alpha_k^\nu(u) \, du \right\}^2.$$  

The term within square brackets can be replaced by the terms in the Taylor series below, considered for a fixed $u$ and $||\delta_k^\nu(u) - \delta_k^\nu(u)|| \leq ||\delta_k^\nu(u) - \delta_k^\nu(u)||$,

$$E_\beta(\delta_k^\nu(u) - \delta_k^\nu(u)) = E_\beta(\delta_k^\nu(u) - \delta_k^\nu(u)) + \delta_k^\nu(u) \frac{\partial E_\beta(\delta_k^\nu(u) - \delta_k^\nu(u))}{\partial \delta_k^\nu(u)}$$

$$= E_\beta(\delta_k^\nu(u) - \delta_k^\nu(u)) + \delta_k^\nu(u) \frac{\partial E_\beta(\delta_k^\nu(u) - \delta_k^\nu(u))}{\partial \delta_k^\nu(u)}$$

$$+ \delta_k^\nu(u) \left[ \frac{\partial E_\beta(\delta_k^\nu(u) - \delta_k^\nu(u))}{\partial \delta_k^\nu(u)} - \frac{\partial E_\beta(\delta_k^\nu(u) - \delta_k^\nu(u))}{\partial \delta_k^\nu(u)} \right].$$

Therefore,

$$X_\beta(t) \leq 2 \sum_{h=1}^{\nu} \left\{ l_h^{-1} \int_0^t \mathbb{I}_h^\nu(u)[(\delta_k^\nu(u) - \delta_k^\nu(u))^T \frac{\partial E_\beta(\delta_k^\nu(u) - \delta_k^\nu(u))}{\partial \delta_k^\nu(u)}] S_k^{(0)}(\delta^\nu, u) \alpha_k^\nu(u) \, du \right\}^2$$

$$+ 2 \sum_{h=1}^{\nu} \left\{ l_h^{-1} \int_0^t \mathbb{I}_h^\nu(u)(\delta_k^\nu(u) - \delta_k^\nu(u))^T \left[ \frac{\partial E_\beta(\delta_k^\nu(u) - \delta_k^\nu(u))}{\partial \delta_k^\nu(u)} - \frac{\partial E_\beta(\delta_k^\nu(u) - \delta_k^\nu(u))}{\partial \delta_k^\nu(u)} \right] \right. \times S_k^{(0)}(\delta^\nu, u) \alpha_k^\nu(u) \, du \right\}$$

(4.39)  

$$= 0 \quad \text{so that (4.39) equals}$$

$$\sum_{h=1}^{\nu} \left\{ l_h^{-1} \int_0^t \mathbb{I}_h^\nu(u)[(\delta_k^\nu(u) - \delta_k^\nu(u))^T \left[ (V_{\delta_k}^{(0)}) S_k^{(0)}(\delta^\nu) \right. \left. (V_{\delta_k}^{(0)}) S_k^{(0)}(\delta^\nu) - (V_{\delta_k}^{(0)}) S_k^{(0)}(\delta^\nu) \right] \alpha_k^\nu(u) \, du \right\}$$

$$\leq \sum_{h=1}^{\nu} \left\{ l_h^{-1} \int_0^t \mathbb{I}_h^\nu(u)[\delta_k^\nu(u) - \delta_k^\nu(u))^T \left[ (V_{\delta_k}^{(0)}) S_k^{(0)}(\delta^\nu) - (V_{\delta_k}^{(0)}) S_k^{(0)}(\delta^\nu) \right] \alpha_k^\nu(u) \, du \right\}$$

$$\leq \sum_{h=1}^{\nu} \left\{ l_h^{-1} \int_0^t \mathbb{I}_h^\nu(u)[\delta_k^\nu(u) - \delta_k^\nu(u))^T \left[ (V_{\delta_k}^{(0)}) S_k^{(0)}(\delta^\nu) - (V_{\delta_k}^{(0)}) S_k^{(0)}(\delta^\nu) \right] \alpha_k^\nu(u) \, du \right\}$$

$$\leq \sum_{h=1}^{\nu} l_h^{-2} \int_0^t \mathbb{I}_h^\nu(u)[\delta_k^\nu(u) - \delta_k^\nu(u))^2 \int_0^t \left[ (V_{\delta_k}^{(0)}) S_k^{(0)}(\delta^\nu) - (V_{\delta_k}^{(0)}) S_k^{(0)}(\delta^\nu) \right]^2 \alpha_k^\nu(u) \, du \right\}$$

$$\leq O(1) \sum_{h=1}^{\nu} l_h^{-2} \int_0^t \mathbb{I}_h^\nu(u) \sum_{i=1}^{\nu} l_i^2 \mathbb{I}_i^\nu(u) \left[ (V_{\delta_k}^{(0)}) S_k^{(0)}(\delta^\nu) - (V_{\delta_k}^{(0)}) S_k^{(0)}(\delta^\nu) \right]^2 \alpha_k^\nu(u) \, du \right\}$$

$$= O(1) \sum_{h=1}^{\nu} \int_0^t \mathbb{I}_h^\nu(u) \int_0^t \left[ (V_{\delta_k}^{(0)}) S_k^{(0)}(\delta^\nu) - (V_{\delta_k}^{(0)}) S_k^{(0)}(\delta^\nu) \right]^2 \alpha_k^\nu(u) \, du \right\}$$

110
\[ = O(1) \int_0^T \left( V_{\beta \delta} S_k^{(0)}(\delta) - \nu_{\beta \delta} s_k^{(0)}(\delta) \right)^2 + O(1) \int_0^T \left( C_{\delta \delta} S_k^{(0)}(\delta) - c_{\delta \delta} s_k^{(0)}(\delta) \right)^2 \, du. \] 

(4.41)

For the first expression in the last line above we note that \( V_{\beta \delta} s_k^{(0)} = S_k^{(2)} - s_k^{(2)} - (S_k^{(1)} E_\beta - s_k^{(1)} e_\beta) \); and also we have \( S_k^{(1)} E_\beta - s_k^{(1)} e_\beta = (S_k^{(1)} - s_k^{(1)}) E_\beta + (E_\beta - e_\beta) s_k^{(1)} \) so that

\[
\int_0^T \left( V_{\beta \delta} s_k^{(0)}(\delta) - \nu_{\beta \delta} s_k^{(0)}(\delta) \right)^2 \, du \leq 2 \int_0^T \left( S_k^{(2)}(\delta) - s_k^{(2)}(\delta) \right)^2 \, du \\
+ 4 \int_0^T \left( S_k^{(1)}(\delta) - s_k^{(1)}(\delta) \right)^2 E_\beta^2 \, du + 4 \int_0^T \left( E_\beta - e_\beta \right)^2 (s_k^{(1)}(\delta))^2 \, du \\
= O_p(n^{-1}) + O_p(1) \int_0^T \left( S_k^{(1)}(\delta) - s_k^{(1)}(\delta) \right)^2 \, du \\
+ O(1) \int_0^T \left( E_\beta - e_\beta \right)^2 \, du = O_p(n^{-1}) + O_p(n^{-1}) + O_p(n^{-1}),
\]

where we have made use of conditions A(2), A(3), C(1) and the fact that \( (E_\beta - e_\beta) = (S_k^{(1)} - s_k^{(1)})/s_k^{(0)} - E_\beta (S_k^{(0)} - s_k^{(0)})/s_k^{(0)}. \)

The same steps can be followed for the second term in (4.41) so that we have

\[
\sum_{h=1}^\nu \left( l_h^{-1} \int_0^t \mathbb{I}_h(u) \left[ (\delta_k^*(u) - \delta_k^o(u))^2 \frac{\partial E_\beta(\delta_k^o, u)}{\partial \delta_k^o(u)} \right] S_k^{(0)}(\delta^o, u) \alpha_k^o(u) \, du \right)^2 = O_p(1/n).
\]

(4.42)

With respect to expression (4.40), using Cauchy-Schwarz such expression is less than or equal to

\[
2 \sum_{h=1}^\nu \left( l_h^{-1} \int_0^t \mathbb{I}_h(u) \left[ (\delta_k^*(u) - \delta_k^o(u))^2 \frac{\partial E_\beta(\delta_k^o, u)}{\partial \delta_k^o(u)} - \frac{\partial E_\beta(\delta_k^o, u)}{\partial \delta_k^o(u)} \right] S_k^{(0)}(\delta^o, u) \alpha_k^o(u) \, du \right)^2 \\
\leq 2 \sum_{h=1}^\nu \left\{ \sup_{u \in T} \left| \delta_k^o(u) - \delta_k^o(u) \right| \left| \frac{\partial E_\beta(\delta_k^o, u)}{\partial \delta_k^o(u)} - \frac{\partial E_\beta(\delta_k^o, u)}{\partial \delta_k^o(u)} \right| \right\} \\
\times l_h^{-1} \int_0^t \mathbb{I}_h(u) \left| (\delta_k^*(u) - \delta_k^o(u))^2 \right| S_k^{(0)}(\delta^o, u) \alpha_k^o(u) \, du \right)^2 \\
\leq 2 \left\{ \sup_{u \in T} \left| \delta_k^o(u) - \delta_k^o(u) \right| \left( V_{\beta \delta} - V_{\beta \delta} \right)^2 + \sup_{u \in T} \left| \delta_k^o(u) - \delta_k^o(u) \right| \left( C_{\delta \delta} - C_{\delta \delta} \right)^2 \right\} \\
\times \sum_{h=1}^\nu \left( l_h^{-1} \int_0^t \mathbb{I}_h(u) \left| (\delta_k^*(u) - \delta_k^o(u))^2 \right| S_k^{(0)}(\delta^o, u) \alpha_k^o(u) \, du \right)^2
\]

111
\[ o_p(1) \sum_{h=1}^{\nu} \int_0^t I_h^*(u) \| \tilde{\delta}^v(u) - \delta^v_k(u) \| O_p(1) \ du \quad \text{[by Condition A(1)]} \]
\[ = o_p(1) \sum_{h=1}^{\nu} \int_0^t I_h^*(u) l_h^2 \ du \leq o_p(1) \| l \|^{-2} \sum_{h=1}^{\nu} l_h^2 I_h = o_p(\| l \|^2). \]

The result above together with (4.42) implies that

\[ X_{\theta}(t) = O_p(1/n) + o_p(\| l \|^2). \]

As for \( X_{\theta} \), we have that

\[ X_{\theta}(t) \]
\[ = \sum_{h=1}^{\nu} \left( \frac{1}{n} \sum_{j=1}^{n} \int_0^t I_h^*(u) \left\{ \frac{Y_{1j}(u) Y_{2j}(u)}{w_j(u, \theta_k^v)} - \frac{S_k^{(7)}(\tilde{\delta}^v, u)}{S_k^{(0)}(\tilde{\delta}^v, u)} w_j(u, \theta_k^v) e^{\beta_k(u)} X_j(u) \alpha_k^v(u) \ du \right\} \right)^2 \]
\[ = \sum_{h=1}^{\nu} \left( \frac{1}{n} \sum_{j=1}^{n} \frac{Y_{1j}(u) Y_{2j}(u)}{w_j(u, \theta_k^v)} w_j(u, \theta_k^v) e^{\beta_k(u)} X_j(u) - E_{\theta}(\tilde{\delta}^v, u) S_k^{(0)}(\delta^v, u) \alpha_k^v(u) \ du \right)^2 \]

Noting that

\[ \sum_{h=1}^{\nu} \frac{Y_{1j} Y_{2j}}{w_j(\theta_k^v)} w_j(\theta_k^v) e^{\beta_k(u)} X_j = \frac{1 + \theta_k^v}{1 + \theta_k^v} \sum_{h=1}^{\nu} Y_{1j} Y_{2j} e^{\beta_k(u)} X_j = \frac{1 + \theta_k^v}{1 + \theta_k^v} \bar{c}^{(7)}(\delta^v), \]

we have that

\[ X_{\theta} = \sum_{h=1}^{\nu} \left[ \frac{1}{l_h^2} \int_0^t I_h^*(u) \left( \frac{1 + \theta_k^v(u)}{1 + \theta_k^v(u)} \right) S_k^{(7)}(\delta^v, u) - E_{\theta}(\tilde{\delta}^v, u) S_k^{(0)}(\delta^v, u) \alpha_k^v(u) \ du \right] \]
\[ = \sum_{h=1}^{\nu} \left[ \frac{1}{l_h^2} \int_0^t I_h^*(u) \left( \frac{1 + \theta_k^v(u)}{1 + \theta_k^v(u)} \right) \left( (1 + \theta_k^v(u)) E_{\theta}(\delta^v, u) - (1 + \theta_k^v(u)) E_{\theta}(\tilde{\delta}^v, u) \right) \right] \]
\[ \times S_k^{(0)}(\delta^v, u) \alpha_k^v(u) \ du \right]^2. \]
Taylor expanding \((1 + \hat{\theta}_k^\nu)E_\theta(\bar{\delta}_k^\nu)\) around \(\delta_k^\nu\), we obtain

\[
X_\theta(t) \leq 2 \sum_{h=1}^{\nu} \left[ l_h^{-1} \int_0^t \mathbb{I}_h^\nu(u) \frac{1 + \theta_k^\nu(u)}{1 + \hat{\theta}_k^\nu(u)} (\bar{\delta}_k^\nu(u) - \delta_k^\nu(u))^T \left( \frac{C_{\theta_k}^{\nu}(u)}{V_{\theta_k}^{\nu}(u)} \right) S_k^{(0)}(\delta^\nu, u) \alpha_k^\nu(u) \, du \right]^2 \\
+ 2 \sum_{h=1}^{\nu} \left[ l_h^{-1} \int_0^t \mathbb{I}_h^\nu(u) \frac{1 + \theta_k^\nu(u)}{1 + \hat{\theta}_k^\nu(u)} (\bar{\delta}_k^\nu(u) - \delta_k^\nu(u))^T \left[ \left( \frac{C_{\theta_k}^{\nu}(u)}{V_{\theta_k}^{\nu}(u)} \right) - \left( \frac{C_{\theta_k}^{\nu}(u)}{V_{\theta_k}^{\nu}(u)} \right) \right] \\
\times S_k^{(0)}(\delta^\nu, u) \alpha_k^\nu(u) \, du \right]^2 \\
\leq O(1) \sum_{h=1}^{\nu} \left[ l_h^{-1} \int_0^t \mathbb{I}_h^\nu(u) (\bar{\delta}_k^\nu(u) - \delta_k^\nu(u))^T \left( \frac{C_{\theta_k}^{\nu}(u)}{V_{\theta_k}^{\nu}(u)} \right) S_k^{(0)}(\delta^\nu, u) \alpha_k^\nu(u) \, du \right]^2 \\
+ O(1) \sum_{h=1}^{\nu} \left[ l_h^{-1} \int_0^t \mathbb{I}_h^\nu(u) (\bar{\delta}_k^\nu(u) - \delta_k^\nu(u))^T \left[ \left( \frac{C_{\theta_k}^{\nu}(u)}{V_{\theta_k}^{\nu}(u)} \right) - \left( \frac{C_{\theta_k}^{\nu}(u)}{V_{\theta_k}^{\nu}(u)} \right) \right] \\
\times S_k^{(0)}(\delta^\nu, u) \alpha_k^\nu(u) \, du \right]^2, \\
\tag{4.43}
\end{align*}

using the fact that, under condition D(2), \((1 + \theta_k^\nu)/(1 + \hat{\theta}_k^\nu) = O(1)\).

Therefore, looking closely to expressions (4.43)--(4.44) and (4.39)--(4.40) we note that the same approach considered for \(X_\beta\) can be used here, such that we also conclude that

\[
X_\theta(t) = O_p(1/n) + o_p(||I||^2),
\]

concluding part (i) of the proof.

For part (ii), note that (4.36) can be written as

\[
\sum_{h=1}^{\nu} (nl_h)^{-2} \left\{ \left( \sum_{j=1}^{\nu} \int_0^t \frac{\partial f_{kj}(\bar{\delta}^\nu, u)}{\partial \beta_{kh}} \, dM_{kj}(u) \right)^2 + \left( \sum_{j=1}^{\nu} \int_0^t \frac{\partial f_{kj}(\bar{\delta}^\nu, u)}{\partial \theta_{kh}} \, dM_{kj}(u) \right)^2 \right\}
\]

\[
= \sum_{h=1}^{\nu} \left\{ (nl_h)^{-1} \sum_{j=1}^{\nu} \int_0^t \mathbb{I}_h^{\nu}(u) \left( X_j(u) - E_\theta(\bar{\delta}^\nu, u) \right) \, dM_{kj}(u) \right\}^2 \\
+ \sum_{h=1}^{\nu} \left\{ (nl_h)^{-1} \sum_{j=1}^{\nu} \int_0^t \mathbb{I}_h^{\nu}(u) \left( \frac{Y_{1j}(u)Y_{2j}(u)}{w_j(u, \theta_k^\nu)} - E_\theta(\bar{\delta}^\nu, u) \right) \, dM_{kj}(u) \right\}^2 \\
= Z_\beta(t) + Z_\theta(t).
\]

113
Considering the terms above separately, we first note that since (4.45) is a sub-
martingale, it follows by the Doob-Meier decomposition that there exists a process
$P_\beta$ such that

$$Z_\beta(t) - P_\beta(t) = \sum_{h=1}^{\nu} \{ (n l_h)^{-1} \sum_{j=1}^{n} \int_{0}^{t} \mathbb{I}_h(u) \{ X_j(u) - E_{\beta}(\delta_k^u; u) \} \, dM_{kj}(u) \}^2$$

$$- \sum_{h=1}^{\nu} n^{-1} l_h^{-2} \sum_{j=1}^{n} \int_{0}^{t} \mathbb{I}_h(u) \frac{1}{n} (X_j(u) - E_{\beta}(\delta_k^u; u))^2 \, d\Lambda_{kj}(u)$$

is a martingale. If we can show that such a quantity converges to zero (in probability)
as $n \to \infty$, then we can work with the limit of $P_\beta(t)$ to deal with $Z_\beta(t)$, as $n \to \infty$. A convenient approach is to consider the Lenglart's inequality, that will consist in showing that the predictable variation process of $Z_\beta(t) - P_\beta(t)$ goes to zero as
$n \to \infty$. First, however, we note that since $Z_\beta(t) - P_\beta(t) \leq \sum_{h=1}^{\nu} n^{-1} l_h^{-2} A_h \leq n^{-1} ||l||^{-4} \sum_{h=1}^{\nu} A_h O(1)$, it suffices to show that $\langle n ||l||^4(Z_\beta(t) - P_\beta(t)) \rangle \overset{P}{\longrightarrow} 0$. We can then follow the same steps in Murphy and Sen (1991) to conclude that

$$\sup_{t \in \mathcal{T}} ||l||^4 n |Z_\beta(t) - \int_{0}^{t} \sum_{h=1}^{\nu} \mathbb{I}_h(u) l_h^{-2} n^{-1} v_{\beta_\xi}(u) s_{k}^{(0)}(\delta^u, u) \alpha_k(u) \, du| = o_p(1).$$

(4.47)

As for (4.46) a similar idea is considered so that the compensated process

$$Z_\theta(t) - P_\theta(t)$$

$$= \sum_{h=1}^{\nu} \{ (n l_h)^{-1} \sum_{j=1}^{n} \int_{0}^{t} \mathbb{I}_h(u) \left( \frac{Y_{1j}(u) Y_{2j}(u)}{w_j(u, \delta_k^u)} - E_\theta(\delta_k^u, u) \right) \, dM_{kj}(u) \}^2$$

$$- \sum_{h=1}^{\nu} (n l_h)^{-2} \sum_{j=1}^{n} \int_{0}^{t} \mathbb{I}_h(u) \left( \frac{Y_{1j}(u) Y_{2j}(u)}{w_j(u, \delta_k^u)} - E_\theta(\delta_k^u, u) \right)^2 \, d\Lambda_{kj}(u)$$

is a martingale. We then turn our attention to determine the predictable variation process of $n ||l||^4(Z_\theta(t) - P_\theta(t))$ and try to show that such a process converges to zero in probability. To do so we consider its optional variation process $[n ||l||^4(Z_\theta(t) - P_\theta(t))] = \sum_{u \leq t} (n ||l||^4 \Delta(Z_\theta(u) - P_\theta(u)))^2$. Using the same idea as in Murphy and
Sen (1991), we have

\[ \sum_{u \leq t} (n \|l\|^4 \Delta(Z_\theta(u) - P_\theta(u)))^2 \]

\[ = (n \|l\|^4)^2 \sum_{j=1}^{n} \int_{0}^{t} \sum_{h=1}^{\nu} \Pi_{h}^{j}(u) \left\{ M_{h}^{j}(u)^2 (n l_{h})^{-2} \left( \frac{Y_{1j}(u) Y_{2j}(u)}{w_{j}(u, \theta_{k}^{u})} - E_{\theta}(\delta_{k}^{u}, u) \right)^2 \right. \]

\[ \left. + (n l_{h})^{-4} \left( \frac{Y_{1j}(u) Y_{2j}(u)}{w_{j}(u, \theta_{k}^{u})} - E_{\theta}(\delta_{k}^{u}, u) \right)^4 \right. \]

\[ + 2 M_{h}^{j}(u) (n l_{h})^{-3} \left( \frac{Y_{1j}(u) Y_{2j}(u)}{w_{j}(u, \theta_{k}^{u})} - E_{\theta}(\delta_{k}^{u}, u) \right)^3 \right\} \ dN_{kj}(u), \]

(4.48)

where, for \( a_h \) being the lower limit of the \( h \)th interval in the partition \( \Pi \),

\[ M_{h}^{j}(u) = 2 n^{-1} \sum_{j=1}^{n} \int_{a_h}^{u} l_{h}^{-1} \left( \frac{Y_{1j}(s) Y_{2j}(s)}{w_{j}(s, \theta_{k}^{u})} - E_{\theta}(\delta_{k}^{u}, s) \right) dM_{kj}(s). \]

(4.49)

In order to get (4.48) it is used the fact that \( \Delta(Z_\theta(t) - P_\theta(t)) = \Delta Z_\theta(t) \) (since \( P_\theta(t) \) is continuous) and that \( \Delta M_{kj}(t) = \Delta N_{kj}(t) \).

As well known, for a martingale \( M \), \( M^2 - \langle M \rangle \) and \( [M] - \langle M \rangle \) are also martingales, therefore, the predictable variation process of \( n \|l\|^4(Z_\theta - P_\theta) \) is the same as the compensator of (4.48). Since \( N_{kj} - \Lambda_{kj} \) is a martingale and \( M_{h}^{j}(u) \) is predictable, it follows the predictable variation process we want is given by

\[ [n \|l\|^4(Z_\theta(t) - P_\theta(t))] \]

\[ = (n \|l\|^4)^2 \sum_{j=1}^{n} \int_{0}^{t} \sum_{h=1}^{\nu} \Pi_{h}^{j}(u) \left\{ M_{h}^{j}(u)^2 (n l_{h})^{-2} \left( \frac{Y_{1j}(u) Y_{2j}(u)}{w_{j}(u, \theta_{k}^{u})} - E_{\theta}(\delta_{k}^{u}, u) \right)^2 \right. \]

\[ \left. + (n l_{h})^{-4} \left( \frac{Y_{1j}(u) Y_{2j}(u)}{w_{j}(u, \theta_{k}^{u})} - E_{\theta}(\delta_{k}^{u}, u) \right)^4 \right. \]

\[ + 2 M_{h}^{j}(u) (n l_{h})^{-3} \left( \frac{Y_{1j}(u) Y_{2j}(u)}{w_{j}(u, \theta_{k}^{u})} - E_{\theta}(\delta_{k}^{u}, u) \right)^3 \right\} \ d\Lambda_{kj}(u) \]
\[(n||l||^4)^2 \sum_{h=1}^\nu \int_0^t \Pi_h^\nu(u) \left\{ M_h^\nu(u)^2 n^{-1} I_h^{-2} \sum_{j=1}^n \left( \frac{Y_{ij}(u)Y_{2j}(u)}{w_j(u, \theta^\nu_k)} - E_{\theta}(\delta^\nu_k, u) \right)^2 \right\}
\times w_j(u, \theta^\nu_k)e^{\delta^\nu_k(u)X_j(u)} du.\]

\[+ n^{-3} I_h^{-4} \sum_{j=1}^n \left( \frac{Y_{ij}(u)Y_{2j}(u)}{w_j(u, \theta^\nu_k)} - E_{\theta}(\delta^\nu_k, u) \right)^4 w_j(u, \theta^\nu_k)e^{\delta^\nu_k(u)X_j(u)}\]

\[+ 2M_h^\nu(u)n^{-2} I_h^{-3} \sum_{j=1}^n \left( \frac{Y_{ij}(u)Y_{2j}(u)}{w_j(u, \theta^\nu_k)} - E_{\theta}(\delta^\nu_k, u) \right)^3 w_j(u, \theta^\nu_k)e^{\delta^\nu_k(u)X_j(u)} \alpha_k^\nu(u) du.\]

In (4.50), we have that

\[
\frac{1}{n} \sum_{j=1}^n \left( \frac{Y_{ij}(u)Y_{2j}(u)}{w_j(u, \theta^\nu_k)} - E_{\theta}(\delta^\nu_k, u) \right)^2 w_j(u, \theta^\nu_k)e^{\delta^\nu_k(u)X_j(u)} \leq 2 \left\{ \frac{1}{n} \sum_{j=1}^n \left( \frac{Y_{ij}(u)Y_{2j}(u)}{w_j(u, \theta^\nu_k)} \right)^2 w_j(u, \theta^\nu_k)e^{\delta^\nu_k(u)X_j(u)} + \left( E_{\theta}(\delta^\nu_k, u) \right)^2 S_k^{(0)}(\delta^\nu, u) \right\}
\]

\[= 2 \left\{ \left( \frac{1}{1 + \theta^\nu_k(u)} \right)^2 S_k^{(8)}(\delta^\nu, u) + \left( E_{\theta}(\delta^\nu_k, u) \right)^2 S_k^{(0)}(\delta^\nu, u) \right\}
\]

\[\leq O_p(1) \left\{ \sup_{u \in \mathcal{I}} |S_k^{(8)}(\delta^\nu, u) - S_k^{(8)}(\delta^\nu, u)| + s_k^{(8)}(\delta^\nu, u) \right\}
\]

\[\leq O_p(1) \left\{ \sup_{u \in \mathcal{I}} |S_k^{(8)}(\delta^\nu, u) - s_k^{(8)}(\delta^\nu, u)| + s_k^{(8)}(\delta^\nu, u) \right\}
\]

\[= O_p(1), \quad (4.53)\]

by conditions A(1) and C(1). Also, in the intermediate steps we made use of conditions A(3) and D(2). Similarly, we can write

\[n^{-1} \sum_{j=1}^n (Y_{ij}Y_{2j}/w_j(\theta^\nu_k) - E_{\theta}(\delta^\nu_k)) = O_p(1), \quad i = 3, 4 \text{ in (4.52) and (4.51) respectively, so that using C(1)},\]

\[\langle n||l||^4(Z_\theta(t) - P_\theta(t)) \rangle = (n||l||^4)^2 \sum_{h=1}^\nu \int_0^t \Pi_h^\nu(u) \left\{ M_h^\nu(u)^2 n^{-1} I_h^{-2} duO_p(1) \right\}
\]

\[+ (n||l||^4)^2 \sum_{h=1}^\nu \int_0^t \Pi_h^\nu(u)n^{-3} I_h^{-4} duO_p(1) \]

\[+ (n||l||^4)^2 \sum_{h=1}^\nu \int_0^t \Pi_h^\nu(u)|M_h^\nu(u)|n^{-2} I_h^{-3} duO_p(1) \]

116
\begin{align*}
&\leq n^2 ||l||^8 n^{-2} \sum_{h=1}^\nu \int_0^t \mathbb{I}_h^\nu(u) \left( M_h^\nu(u) \right)^2 \, du O_p(1) + n^2 ||l||^9 n^{-3} \sum_{h=1}^\nu l_h O_p(1) \\
&\quad + n^2 ||l||^8 n^{-2} \sum_{h=1}^\nu \int_0^t \mathbb{I}_h^\nu(u) |M_h^\nu(u)| \, du O_p(1) \\
&\leq n ||l||^4 \sum_{h=1}^\nu \int_0^t \mathbb{I}_h^\nu(u) \left( M_h^\nu(u) \right)^2 \, du + O_p(n^{-1}) \\
&\quad + ||l||^2 \sum_{h=1}^\nu \int_0^t \mathbb{I}_h^\nu(u) |M_h^\nu(u)| \, du O_p(1) \\
\end{align*}
(4.54)

At this point we have to deal with $M_h^\nu$. For the third term in the r.h.s. of (4.54) we have (recall that $\mathbb{I}_h^\nu$ represents the $h$th interval in the partition $\Pi$ with lower end-point equals to $a_h$)

\[ ||l||^2 \sum_{h=1}^\nu \int_0^t \mathbb{I}_h^\nu(u) |M_h^\nu(u)| \, du \leq ||l||^2 \sum_{h=1}^\nu \sup_{u \in I_h^\nu} |M_h^\nu(u)| l_h. \]

Denoting $f_j^\nu = Y_{1j}Y_{2j}/w_j(\hat{\theta}_k^\nu) - E_\theta(\hat{\theta}_k^\nu)$ and using the definition of $M_h^\nu$ given in (4.49),

\[
\sup_{u \in I_h^\nu} |M_h^\nu| = \sup_{u \in I_h^\nu} \left| 2n^{-1} \sum_{j=1}^n \int_{a_h}^u l_h^{-1} f_j^\nu(s) \, dM_kj(s) \right|
\]
\[
= 2n^{-1} \sup_{u \in I_h^\nu} \left| \sum_{j=1}^n \int_{a_h}^u \mathbb{I}_h^\nu(s) l_h^{-1} f_j^\nu(s) \, dM_kj(s) \right|
\]
\[
\leq 2 \sup_{u \in I_h^\nu} \left| \sum_{j=1}^n \int_{a_h}^u \mathbb{I}_h^\nu(s) l_h^{-1} f_j^\nu(s) \, dM_kj(s) \right|
\]

The quantity within $| \cdot |$ is a martingale with predictable variation process given by

\[
n^{-2} \sum_{j=1}^n \int_0^u \mathbb{I}_h^\nu(s) l_h^{-2} \left( f_j^\nu(s) \right)^2 w_j(s, \theta_k^\nu) e^{\theta_k^\nu(s)X_j(s)} \alpha_k^\nu(s) \, ds
\]
\[
= n^{-1} \int_0^u \mathbb{I}_h^\nu(s) l_h^{-2} \frac{1}{n} \sum_{j=1}^n \left( f_j^\nu(s) \right)^2 w_j(s, \theta_k^\nu) e^{\theta_k^\nu(s)X_j(s)} \alpha_k^\nu(s) \, ds
\]
\[
= O_p(n^{-1}) l_h^{-2} \int_0^u \mathbb{I}_h^\nu(s) \, ds \leq O_p(n^{-1}) l_h^{-1},
\]

where in the last line we have used the result obtained in (4.53). Therefore, using
Lenglart’s inequality, for \( K > 0 \) we have that for all \( \epsilon > 0 \),

\[
\mathbb{P}\left\{ \sup_{u \in \mathcal{I}_h} \left| n^{-1} \sum_{j=1}^{n} \int_0^u \mathbb{I}_h(s) l_{h}^{-1} \left( \frac{Y_{ij}(s) Y_{2j}(s)}{w_j(s, \theta^*_k)} - E_{\theta}(\theta^*_k, s) \right) \, dM_{kj}(s) \right| > \left( \frac{2}{\epsilon} Kn^{-1} l_{h}^{-1} \right)^{1/2} \right\} \\
\leq \frac{\epsilon}{2} + \mathbb{P}\left\{ \frac{1}{n} \sum_{j=1}^{n} \int_0^u \mathbb{I}_h(s) l_{h}^{-2} \left( \frac{Y_{ij}(s) Y_{2j}(s)}{w_j(s, \theta^*_k)} - E_{\theta}(\theta^*_k, s) \right)^2 \, d\Lambda_{kj}(s) > n^{-1} l_{h}^{-1} K \right\} \leq \epsilon.
\]  

(4.55)

In other words, we conclude that \( \sup_{u \in \mathcal{I}_h} |M_h^*(u)| = O_p(n^{-1/2} l_{h}^{-1/2}) \), so that

\[
||l||^2 \sum_{h=1}^{\mathcal{V}} \int_0^t \mathbb{I}_h(u) |M_h^*(u)| \, du \leq ||l||^2 O_p(n^{-1/2}) \sum_{h=1}^{\mathcal{V}} l_{h}^{-1/2} l_{h} = O_p(n^{-1/2} ||l||).
\]

A similar approach can be used for the first term in the r.h.s. of (4.54), i.e.,

\[
n ||l||^4 \sum_{h=1}^{\mathcal{V}} \int_0^t \mathbb{I}_h(u) (M_h^*(u))^2 \, du \leq n ||l||^4 \sum_{h=1}^{\mathcal{V}} \sup_{u \in \mathcal{I}_h^*} (M_h^*(u))^2 l_{h},
\]

and \( \sup_{u \in \mathcal{I}_h^*} (M_h^*(u))^2 \leq 4n^{-2} \sup_{u \in \mathcal{I}_h^*} \left( \sum_{j=1}^{n} \int_0^u \mathbb{I}_h(s) l_{h}^{-1} f_j^*(s) \, dM_{kj}(s) \right)^2 \), then it follows by (4.55) that

\[
n ||l||^4 \sum_{h=1}^{\mathcal{V}} \sup_{u \in \mathcal{I}_h^*} (M_h^*(u))^2 l_{h} \leq n ||l||^4 \sum_{h=1}^{\mathcal{V}} l_{h}^{-1} O_p(n^{-1}) l_{h} \leq ||l||^4 O_p(1) ||l||^{-2},
\]

and hence we conclude that

\[
n ||l||^4 \sum_{h=1}^{\mathcal{V}} \int_0^t \mathbb{I}_h(u) (M_h^*(u))^2 \, du = O_p(||l||^2).
\]

All the intermediate steps developed above allow us to conclude that

\[
\langle n ||l||^4 (Z_\theta(t) - P_\theta(t)) \rangle = O_p(||l||^2) + O_p(n^{-1}) + O_p(n^{-1/2} ||l||).
\]

Therefore we conclude that \( \langle n ||l||^4 (Z_\theta(t) - P_\theta(t)) \rangle \xrightarrow{p} 0 \), and hence by the Lenglart’s inequality we conclude that the limit of \( Z_\theta \) and \( P_\theta \) coincide and to compute
the limit of the latter we note that

\[ P_{\theta}(t) = \sum_{h=1}^{n} (n l_h)^{-2} \sum_{j=1}^{n} \int_{0}^{t} \mathbb{P}_{h} \left( \frac{Y_{ij} Y_{2j}}{w_j(\theta_k)} - E_{\theta}(\tilde{\delta}_k^{(\nu)}) \right)^2 w_j(\theta_k^0) e^{\beta k x_j} \alpha_k^{(\nu)} \, du \]

\[ \leq \sum_{h=1}^{n} l_h^{-2} n^{-1} \int_{0}^{t} \mathbb{P}_{h} \left| \sum_{j=1}^{n} \frac{1}{n} \left( \frac{Y_{ij} Y_{2j}}{w_j(\theta_k)} - E_{\theta}(\tilde{\delta}_k^{(\nu)}) \right)^2 w_j(\theta_k^0) e^{\beta k x_j} - v_{\beta k} s_k^{(0)}(\delta) \right| \alpha_k^{(\nu)} \, du \]

\[ + \sum_{h=1}^{n} l_h^{-2} n^{-1} \int_{0}^{t} \mathbb{P}_{h} v_{\beta k} s_k^{(0)}(\delta) \alpha_k^{(\nu)} \, du. \]

Considering only the term within \(| \cdot |\), we write

\[ \frac{1}{n} \sum_{j=1}^{n} \left( \frac{Y_{ij} Y_{2j}}{w_j(\theta_k)} - E_{\theta}(\tilde{\delta}_k^{(\nu)}) \right)^2 w_j(\theta_k^0) e^{\beta k x_j} - v_{\beta k} s_k^{(0)}(\delta) \]

\[ = \left\{ \left( \frac{1 + \theta_k^0}{1 + \theta_k^{(\nu)}} \right)^2 S_k^{(8)}(\delta) - s_k^{(8)}(\delta) \right\} + \left\{ c_{\theta}^{2}(\delta_k^{(\nu)}) s_k^{(0)}(\delta) - 2 E_{\theta}(\tilde{\delta}_k^{(\nu)}) S_k^{(7)}(\delta) \left( \frac{1 + \theta_k^0}{1 + \theta_k^{(\nu)}} \right) \right\} \]

\[ + \left[ E_{\theta}(\tilde{\delta}_k^{(\nu)}) \right]^{2} S_k^{(0)}(\delta). \] (4.56)

For the first term above we recall that \((1 + \theta_k^0)/(1 + \theta_k^{(\nu)}) = O(1)\), so that such an expression is bounded above by

\[ O(1) \sup_{\theta \in \mathcal{F}} |S_k^{(8)}(\delta) - s_k^{(8)}(\delta)| + \left[ \left( \frac{1 + \theta_k^0}{1 + \theta_k^{(\nu)}} \right)^2 - 1 \right] s_k^{(8)}(\delta) \]

\[ \leq o_p(1) + \left( \frac{\theta_k^0 - \theta_k^{(\nu)}(2 + |\theta_k^0 - \theta_k^{(\nu)}|)}{(1 + \theta_k^{(\nu)})^2} \right) s_k^{(8)}(\delta) = o_p(1). \]

The second term equals

\[ c_{\theta}^{2}(\delta_k^{(\nu)}) s_k^{(0)}(\delta) - 2 E_{\theta}(\tilde{\delta}_k^{(\nu)}) S_k^{(7)}(\delta) \left( \frac{1 + \theta_k^0}{1 + \theta_k^{(\nu)}} - 1 \right) - 2 E_{\theta}(\tilde{\delta}_k^{(\nu)}) S_k^{(7)}(\delta) \]

\[ - \left[ E_{\theta}(\tilde{\delta}_k^{(\nu)}) \right]^{2} S_k^{(0)}(\delta) + 2 \left[ E_{\theta}(\tilde{\delta}_k^{(\nu)}) \right]^{2} S_k^{(0)}(\delta). \]

Noting that \(|(1 + \theta_k^0)/(1 + \theta_k^{(\nu)}) - 1| = o(1)\) and combining the other terms we have
that the above expression is less than or equal to

\[
\begin{align*}
&\alpha_p(1) + 2E_\theta(\tilde{e}_k^\nu)S_k^{(0)}(\delta^\nu) |E_\theta(S_k^\nu) - E_\theta(\tilde{e}_k^\nu)| + \frac{1}{2}|E_\theta(\tilde{e}_k^\nu)|^2 |S_k^{(0)}(\delta^\nu) - s_k^{(0)}(\delta^\nu)| \\
&+ |e_\theta^0(\delta^\nu) - E_\theta(\tilde{e}_k^\nu)|s_k^{(0)}(\delta^\nu) \\
\leq &\alpha_p(1) + \alpha_p(1) + \alpha_p(1) + (E_\theta(\tilde{e}_k^\nu) + e_\theta(\delta^\nu))s_k^{(0)}(\delta^\nu) |E_\theta(\tilde{e}_k^\nu) - e_\theta(\delta^\nu)| \\
\leq &\alpha_p(1) + O_p(1) |E_\theta(\tilde{e}_k^\nu) - E_\theta(\delta^\nu)| + O_p(1) |E_\theta(\tilde{e}_k^\nu) - e_\theta(\delta^\nu)| \\
\leq &\alpha_p(1) + O_p(1) |E_\theta(\tilde{e}_k^\nu) - e_\theta(\delta^\nu)|,
\end{align*}
\]

making use of conditions A(1), A(3) and C(1). Finally, we note that

\[
\sup_{u \in \mathcal{F}} |E_\theta(\delta_k^\nu) - e_\theta(\delta_k^\nu)| \\
\leq O_p(1) \sup_{u \in \mathcal{F}} |S_k^{(0)}(\delta^\nu) - s_k^{(0)}(\delta^\nu)| + O(1) \sup_{u \in \mathcal{F}} |S_k^{(7)}(\delta_k^\nu) - s_k^{(7)}(\delta_k^\nu)| = \alpha_p(1),
\]

by conditions A(1), A(3), and C(1). Hence, we conclude that (4.56) is \( \alpha_p(1) \) so that

\[
P_\theta(t) = \sum_{h=1}^\nu l_h^{-2}n^{-1} \int_0^t I_h^\nu v_\beta^\nu s_k^{(0)}(\delta^\nu) \alpha_k^\nu du + n^{-1} \sum_{h=1}^\nu l_h^{-1} \alpha_p(1),
\]

and hence,

\[
(n\|l\|^4) |Z_\theta(t) - \sum_{h=1}^\nu l_h^{-2}n^{-1} \int_0^t I_h^\nu v_\beta^\nu s_k^{(0)}(\delta^\nu) \alpha_k^\nu du| \\
\leq n\|l\|^4 |Z_\theta(t) - P_\theta(t)| + n\|l\|^4 \left| \sum_{h=1}^\nu l_h^{-2}n^{-1} \int_0^t I_h^\nu v_\beta^\nu s_k^{(0)}(\delta^\nu) \alpha_k^\nu du \right| \\
= \alpha_p(1) + n\|l\|^4 n^{-1} \sum_{h=1}^\nu l_h^{-1} \alpha_p(1) = \alpha_p(1).
\]

Thus,

\[
Z_\theta(t) = (n\|l\|^4)^{-1} \left\{ \|l\|^4 \sum_{h=1}^\nu l_h^{-2} \int_0^t I_h^\nu v_\beta^\nu s_k^{(0)}(\delta^\nu, u) \alpha_k^\nu(u) du + \alpha_p(1) \right\},
\]

and the proof for the lemma is complete.
5.2 Proof for Lemma 4.2

The proof consists in showing that all elements of $M_h^+$ have the same order, what implies the result since $1^T M_h^+ 1$ corresponds to take the sum of the elements of $M_h^+$.

For elements $m_{11}$ and $m_{12} (= m_{21})$, the proof follows exactly as in Lemma 2, part 2, of Murphy and Sen (1991). For $m_{22}$, some extra work needs to be done. First we have that

\[
\left| (l_h n)^{-1} \frac{\partial^2 \log L_n(\delta^\nu)}{\partial \theta^2_k} + l_h^2 \sigma_{h2} \right| = \left| (l_h n)^{-1} \sum_{i=1}^n \int_0^r \mathbb{I}_k^i \left[ \frac{Y_{i1} Y_{2i}}{w_i(\theta_k^\nu)} - E_{\theta}(\delta^\nu) \right] dN_{ki} + l_h^2 \sigma_{h2} \right|
\]

\[
\leq \left| (l_h n)^{-1} \sum_{i=1}^n \int_0^r \mathbb{I}_k^i \left( \frac{Y_{i1} Y_{2i}}{w_i(\theta_k^\nu)} - \frac{S_k^{(8)}(\delta^\nu)}{S_k^{(0)}(\delta^\nu)} \right) dN_{ki} \right|
\]

\[
+ \left| (l_h n)^{-1} \sum_{i=1}^n \int_0^r \mathbb{I}_k^i V_{\theta_k^\nu} dN_{ki} + l_h^2 \sigma_{h2} \right|
\]

Second term in the r.h.s. above can receive the same treatment as in Lemma 2, part 2 of Murphy and Sen (1991) so that it is $O_p(n^{-1/2} ||l||^{-2}) + o_p(1)$. As for the first term, we use the Doob-Meyer decomposition so that it is bounded above by

\[
\left| (l_h n)^{-1} \sum_{i=1}^n \int_0^r \mathbb{I}_k^i \left( \frac{Y_{i1} Y_{2i}}{w_i(\theta_k^\nu)} - \frac{S_k^{(8)}(\delta^\nu)}{S_k^{(0)}(\delta^\nu)} \right) dM_{ki} \right|
\]

\[
+ \left| (l_h n)^{-1} \sum_{i=1}^n \int_0^r \mathbb{I}_k^i \left( \frac{Y_{i1} Y_{2i}}{w_i(\theta_k^\nu)} - \frac{S_k^{(8)}(\delta^\nu)}{S_k^{(0)}(\delta^\nu)} \right) w_i(\theta_k^2) e^{\theta_k^2 X_i} d\theta_k^2 \right|
\]

\[
= \left| (l_h n)^{-1} \sum_{i=1}^n \int_0^r \mathbb{I}_k^i \left( \frac{Y_{i1} Y_{2i}}{w_i(\theta_k^\nu)} - E_{\theta}(\delta^\nu) \right) dM_{ki} \right|
\]

\[
+ \left| (l_h n)^{-1} \int_0^r \mathbb{I}_k^i \left( \frac{Y_{i1} Y_{2i}}{w_i(\theta_k^\nu)} - E_{\theta}(\delta^\nu) \right) \left( \frac{1 + \theta_k^2}{1 + \theta_k^2} \right) \left( \frac{S_k^{(8)}(\delta^\nu)}{S_k^{(8)}(\delta^\nu)} - \frac{S_k^{(0)}(\delta^\nu)}{S_k^{(0)}(\delta^\nu)} \right) \right| \]

\[
(4.57)
\]

\[
(4.58)
\]

For (4.58) we use condition C(2) together with (4.16) and condition A(1) to conclude that such an expression is $o_p(1)$. For (4.57) we consider the Lenglart’s inequality, i.e., denoting

\[
M_k^*(t) = (l_h n)^{-1} \sum_{i=1}^n \int_0^t \mathbb{I}_k^i \left( \frac{Y_{i1} Y_{2i}}{w_i(\theta_k^\nu)} - E_{\theta}(\delta^\nu) \right) dM_{ki},
\]

121
the corresponding predictable variation process is given by

\[
\langle M_k^\nu(t) \rangle = (l_h n)^{-2} \sum_{i=1}^n \int_0^t \mathbb{I}_h^\nu \left( \frac{1}{(1 + \theta_k^\nu)^2} \left( \frac{Y_{i1} Y_{2i}}{w_i(\theta_k^\nu)} - E_\theta(\delta_k^\nu) \right) \right)^2 \, d\lambda_{ki}
\]

\[
= l_h^{-2} n^{-1} \int_0^t \mathbb{I}_h^\nu O_p(1) \, du \quad \text{[by condition C(2) and (4.53)]}
\]

\[
= l_h^{-1} n^{-1} O_p(1) \leq l_h^{-1} n^{-1} O_p(1) \leq \|l\|^{-2} n^{-1} O_p(1).
\]

So, using the Lenglart's Inequality, it follows that for \( K > 0 \), we have \( \forall \epsilon > 0 \),

\[
\mathbb{P}\left\{ \sup_{t \in \tau} |M_k^\nu(t)| > \left( \frac{2}{\epsilon} K n \|l\|^{-2} \right)^{-1/2} \right\} \leq \frac{\epsilon}{2} + \mathbb{P}\left\{ \langle M_k^\nu(\tau) \rangle > \|l\|^{-2} n^{-1} K \right\}
\]

\[
\leq \epsilon.
\]

Hence, (4.57) is \( O_p((n \|l\|^{-2})^{-1/2}) \), so that we can conclude

\[
\max_{1 \leq h \leq \nu} \left| (l_h n)^{-1} \frac{\partial^2 \log L_n(\delta^\nu)}{\partial \theta_{ki}^2} + l_h^{-2} \sigma_{h_2} \right| = O_p(n^{-1/2} \|l\|^{-2}) + o_p(1),
\]

completing the proof for the lemma. \( \square \)

### 5.3 Proof for Lemma 4.3

Following the same idea of the previous subsection, we have to show that all elements of \( M_k^\nu \) are of the given order, what implies the result. For the element \((1,1)\), considering \( N_k = \sum_{i=1}^n N_{ki} \),

\[
\max_{1 \leq h \leq \nu} \left| (l_h n)^{-1} \int_0^T \mathbb{I}_h^\nu(u) \left( V_{\alpha^\nu}(u) - V_{\beta^\nu}(u) \right) \, dN_k(u) \right|
\]

\[
\leq \sup_{u \in \mathcal{T}} |V_{\alpha^\nu}(u) - V_{\beta^\nu}(u)| \max_{1 \leq h \leq \nu} \int_0^T (n l_h)^{-1} \mathbb{I}_h^\nu(u) \, dN_k(u).
\] (4.59)
For the first term in the r.h.s. above we consider that, since $\delta^\nu(u) \to \delta^\circ(u)$, for all $u \in \mathcal{J}$, then

$$\sup_{u \in \mathcal{J}} |V_{\beta^*}(u) - V_{\beta^\nu}(u)| \leq \sup_{u \in \mathcal{J}} |V_{\beta^*}(u) - v_{\beta^*}(u)|$$

$$+ \sup_{u \in \mathcal{J}} |V_{\beta^\nu}(u) - v_{\beta^\nu}(u)| + \sup_{u \in \mathcal{J}} |v_{\beta^*}(u) - v_{\beta^\nu}(u)|$$

$$\leq \sup_{u \in \mathcal{J}} \sup_{\|\delta^\nu(u) - \delta^\circ(u)\| < \gamma} |V_{\beta^*}(u) - v_{\beta^*}(u)| + \sup_{u \in \mathcal{J}} \sup_{\|\delta^\nu(u) - \delta^\circ(u)\| < \gamma} |V_{\beta^\nu}(u) - v_{\beta^\nu}(u)|$$

$$+ \sup_{u \in \mathcal{J}} |v_{\beta^*}(u) - v_{\beta^\nu}(u)|.$$

Using conditions A(1) and A(3) we conclude that the first and second terms in the r.h.s. above are $o_p(1)$ and by condition C(1) together with the fact that we can make $\delta^\nu$ arbitrarily close to $\delta^\circ$, the third term is $o(1)$.

For the second term in the r.h.s. of (4.59), we first note that

$$\max_{1 \leq h \leq \nu} \int_0^\tau (n l_h)^{-1} \Pi^\nu_k(u) \, dN_k(u) \leq l_h^{-1} \max_{1 \leq h \leq \nu} \int_0^\tau n^{-1} \Pi^\nu_k(u) \, dN_k(u)$$

$$\leq O(1)\|l\|^{-2} \max_{1 \leq h \leq \nu} \int_0^\tau n^{-1} \Pi^\nu_k(u) \, dN_k(u).$$

But using $N_k = M_k + \Lambda_k$ and Lemma 1, part 1 of Murphy and Sen (1991), we have that

$$\int_0^\tau n^{-1} \Pi^\nu_k(u) \, dN_k(u) = n^{-1} \int_0^\tau \Pi^\nu_k(u) \, dM_k(u) + \int_0^\tau \Pi^\nu_k(u) \mathcal{S}^{(0)}(\delta^\circ, u) \, d\mathcal{L}(u)$$

$$= O_p(n^{-1/2}) + O_p(1)l_h = O_p(n^{-1/2}) + O_p(\|l\|^2).$$

Hence,

$$\max_{1 \leq h \leq \nu} \int_0^\tau (n l_h)^{-1} \Pi^\nu_k(u) \, dN_k(u) = O_p(n^{-1/2}\|l\|^{-2}) + o_p(1).$$

Combining this result with the previous one we have that

$$\max_{1 \leq h \leq \nu} (n l_h)^{-1} \int_0^\tau \Pi^\nu_k(u) \left(V_{\beta^*}(u) - V_{\beta^\nu}(u)\right) \, dN_k(u) = o_p(n^{-1/2}\|l\|^{-2}) + o_p(1).$$
For the off diagonal elements of $M^*_h$ the exact same proof follows with $V_\beta$ replaced by $C_\delta$. However, for element $(2,2)$ an extra step needs to be considered. First we note that such term can be written as
\[
\max_{1 \leq h \leq \nu} (nl_h)^{-1} \left| \sum_{j=1}^n \int_0^\tau I^*_h(u) \left\{ \left( \frac{Y_{1i}(u)Y_{2i}(u)}{[w_i(u, \theta^*_h)]^2} - E^2_\theta(u, \delta^*_h) \right) - \left( \frac{Y_{1i}(u)Y_{2i}(u)}{[w_i(u, \theta^*_h)]^2} - E^2_\theta(u, \delta^*_h) \right) \right\} dN_{ki}(u) \right| \leq \max_{1 \leq h \leq \nu} (nl_h)^{-1} \left| \sum_{j=1}^n \int_0^\tau I^*_h(u) \left( \frac{Y_{1i}(u)Y_{2i}(u)}{[w_i(u, \theta^*_h)]^2} - \frac{Y_{1i}(u)Y_{2i}(u)}{[w_i(u, \theta^*_h)]^2} \right) dN_{ki}(u) \right| + \max_{1 \leq h \leq \nu} (nl_h)^{-1} \left| \sum_{j=1}^n \int_0^\tau \left( E^2_\theta(u, \delta^*_h) - E^2_\theta(u, \delta^*_h) \right) dN_{ki}(u) \right|.
\]  
(4.60)

Following the same sequence we considered for the element $(1,1)$, we can show that (4.61) is $o_p(n^{-1/2}\|l\|^{-2}) + o_p(1)$. As for expression (4.60), using condition D(2) and the fact that $||\delta^*_h - \delta^*_h|| = o(1)$,
\[
\max_{1 \leq h \leq \nu} (nl_h)^{-1} \left| \sum_{j=1}^n \int_0^\tau I^*_h(u) \left( \frac{Y_{1i}(u)Y_{2i}(u)}{[w_i(u, \theta^*_h)]^2} - \frac{Y_{1i}(u)Y_{2i}(u)}{[w_i(u, \theta^*_h)]^2} \right) dN_{ki}(u) \right| \leq \max_{1 \leq h \leq \nu} (nl_h)^{-1} \left| \sum_{j=1}^n \int_0^\tau I^*_h(u)Y_{1i}(u)Y_{2i}(u) \left( \frac{1}{[1 + \theta^*_h(u)]^2} - \frac{1}{[1 + \theta^*_h(u)]^2} \right) dN_{ki}(u) \right| \leq o(1) \max_{1 \leq h \leq \nu} (nl_h)^{-1} \int_0^\tau I^*_h(u) dN_{ki}(u) = o_p(n^{-1/2}\|l\|^{-2}) + o_p(1),
\]
and hence, the result for the lemma follows. \(\square\)
CHAPTER 5

Possible Applications and Plans for Further Research

1 Introduction

In this chapter we discuss some possible applications of the results obtained in the previous chapters and comment on topics which are intended to in depth future investigation.

In Chapter 3 we consider a model based on the assumption that the ratio of the conditional hazard function for each component, denoted by $\theta_k$ does not depend on time (viz. Section 2.1). Such assumption is relaxed in Chapter 4 together with the inclusion of a time-dependent coefficient $\beta(t)$ requiring then a more elaborated framework to obtain estimates and their properties. In practice it may be clear that the latter approach, while more general, will be costlier in the sense that algorithms to determine the estimators will be more complex and more demanding in terms of computing time. Therefore, it makes sense to derive a procedure that would allow us to determine if it is possible to work with a model defined on a time-independent parameter space.

Keeping this in mind, in the next section we discuss an approach that can be considered for verifying the assumptions that $\delta(t) = \delta$, for all $t \in \mathcal{T}$ made in
Chapter 3. Initially we consider that the asymptotic covariance matrix for quantities of interest satisfies a given structure. Following that we consider a slightly different approach that can be considered in the two-sample setting. In Section 3 we indicate tests for the parameters involved in the finite-dimensional case when they are time-independent and in Section 4 we discuss some open questions for future research.

2 Testing the proportionality assumption

In this section we consider the situation where the main interest is to verify if for a given problem, the parameters in the intensity process for the matrix-valued model considered in Chapter 4 can be considered as quantities belonging to a finite-dimensional space (i.e., they do not depend on \( t \in T \)). If this is the case and the covariates are also time-independent then the ratio of the hazards related to a given component, for any two individuals with same type of components at risk will not depend on \( t \). Hence, in this sense we are testing if the model is (component-wise) of proportional hazards. Consequently, we will denominate this test as a test for the proportionality assumption.

In what follows \( \delta \) is assumed to be a \( p \)-vector. Note that when there are \( K \) components, \( p = 2K \) as we are considering only one independent variable (therefore a dependence parameter \( \theta_k \) and a coefficient \( \beta_k \) is associated to each component.)

2.1 Linear covariance structure

We consider the following framework. As we have shown in Chapter 4, the MPLE for the parameter \( \delta(t) \), when properly normalized will converge to a multivariate Gaussian martingale with a given covariance matrix which we rewrite here as \( \Gamma_t \).
In other words, we saw in Theorem 4.2 that
\[ \sqrt{n} \int_0^t (\hat{\delta}(s) - \delta^o(s)) \, ds \xrightarrow{D} W(t), \tag{5.1} \]
where Cov\{W(s), W(t)\} = \Gamma_{s\wedge t}. Note that the martingale structure implies that \( \Gamma_t - \Gamma_s \) is p.s.d. for all \( s \leq t \). Therefore we may consider the situation where the asymptotic covariance matrix can be expressed as \( \Gamma_t = c_t \Gamma \), where \( \Gamma \) is a positive definite matrix not depending on time \( t \) and \( c_t \) is an increasing function. In addition, at a certain point we will consider the case where \( c_t = t \) for all \( t \in \mathcal{T} \).

The main interest is to examine the hypothesis \( H_0: \delta(t) \) is constant, for all \( t \in \mathcal{T} \), against the alternative \( H_a \) that \( \delta(t) \) is not constant. We introduce the following quantity
\[ \hat{\delta} = \tau^{-1} \int_0^\tau \hat{\delta}(s) \, ds, \]
representing an average value for the MPLE on \( \mathcal{T} \), and then we rewrite the l.h.s. of 5.1 as
\[ \sqrt{n} \int_0^t (\hat{\delta}(s) - \delta^o(s)) \, ds = \sqrt{n} \int_0^t (\hat{\delta}(s) - \hat{\delta}) \, ds + \sqrt{n} \int_0^t (\hat{\delta} - \delta^o(s)) \, ds. \]
Note that under \( H_0 \), \( \delta(s) \) is constant \( \forall s \), equal to \( \delta^o \), say, so that the integrand in the second term of the r.h.s. above does not depend on \( s \). Hence, working with the definition of \( \hat{\delta} \) we can write this term as \( (t/\tau) \sqrt{n} \int_0^\tau (\delta(u) - \delta^o) \, du \). Denoting
\[ Z^*(t) = \Gamma^{-1/2} \sqrt{n} \int_0^t (\hat{\delta}(u) - \delta^o) \, du \quad \text{and} \quad Z(t) = \Gamma^{-1/2} \sqrt{n} \int_0^t (\delta(u) - \hat{\delta}) \, du, \]
we have
\[ Z(t) = Z^*(t) - (t/\tau)Z^*(\tau). \tag{5.2} \]
Note that since \( Z(0) = Z(\tau) = 0 \), the process \( \{Z(t), t \in \mathcal{T}\} \) converges to a tied-down Gaussian process. Also, its form resembles a Brownian bridge; however,
if we denote the asymptotic covariance by $\text{Cov}_A(\cdot, \cdot)$, then for $s \leq t$,

$$
\text{Cov}_A\{Z(s), Z(t)\} = \text{Cov}_A\{Z^*(s), Z^*(t)\} - (t/\tau)\text{Cov}_A\{Z^*(s), Z^*(\tau)\} - (s/\tau)\text{Cov}_A\{Z^*(t), Z^*(\tau)\} + (st/\tau^2)\text{Cov}_A\{Z^*(s), Z^*(\tau)\}
$$

$$
= \Gamma^{-1/2}_\tau \{ \Gamma_s - \frac{t}{\tau} \Gamma_s - \frac{s}{\tau} \Gamma_t + \frac{st}{\tau^2} \Gamma_{\tau} \} \Gamma^{-1/2}_\tau
$$

$$
= c^{-1}_\tau \{ c_s - \frac{t}{\tau} c_s - \frac{s}{\tau} c_t + \frac{st}{\tau^2} c_s \} I
$$

Therefore, $Z(s)$ is not in general a Brownian bridge. It turns out that if we make use of the assumption $c_t = t$ mentioned earlier we can write for $s, t \in \mathcal{I}$,

$$
\text{Cov}_A\{Z(s), Z(t)\} = \frac{s}{\tau} \left( 1 - \frac{t}{\tau} \right) I, \quad s \leq t,
$$

and in this particular situation, $Z$ will weakly converge to $W^\circ(t)$, a vector consisting of $p$ independent Brownian bridges on $[0, \tau]$. Assuming that the sieves estimator is computed based on a partition $0 = t_0 \leq t_1 \leq \cdots \leq t_\nu = \tau$, we define the statistic

$$
\max_{1 \leq j \leq \nu} \|Z(t_j)\|^2,
$$

with $\Gamma_\tau$ replaced by $\tau \hat{\Gamma}$. This estimator is given by a block-diagonal matrix

$$
\hat{\Gamma} = \begin{pmatrix}
\hat{\Gamma}^{(1)} & 0 \\
0 & \hat{\Gamma}^{(2)}
\end{pmatrix}
$$

where under $H_o$, each $\hat{\Gamma}^{(k)}$ can be computed using the expression derived for $n^{-1}J(\hat{\delta})$, with $J$ given in Chapter 3. An issue here is which estimator for $\delta$ should be considered when evaluating $J(\hat{\delta})$. Since the statistic is computed under $H_o$, a possible choice could be the MPLE defined in Chapter 3 for the finite dimensional model; however, given that the sieve estimator is consistent, it could also be used in the computations, as well as $\hat{\delta}$ defined earlier. In practice we will have to compute the sieve estimator in any case, therefore we adopt it here. Considerations related to the power and
robustness of the test also favors the sieve estimator, given that it does not assume
that \( H_0 \) is true.

Therefore, the estimator for the covariance matrix is given by

\[ \hat{\Sigma}^{(k)} = -\left[ \frac{\partial^2 \log L_n(\delta)}{\partial \delta_k \partial \delta_k'} \right]^{-1}. \] (5.4)

Assuming that a sieve estimator \( \hat{\delta}(u) = \sum_{h=1}^{\nu} \mathbb{I}_{h}^{\nu}(u) \hat{\delta}_h \) is computed based on a
partition with subintervals of same length \( l_h = \tau / \nu \), \( h = 1, \ldots, \nu \), then the element
(1,1) based on (minus) the second derivative of the log-likelihood is given by

\[
- \frac{1}{n} \frac{\partial^2 \log L_n(\hat{\delta})}{\partial \beta_k \partial \beta_k'} = \frac{1}{n} \int_0^\tau \left\{ \frac{S_k^{(2)}(\hat{\delta}, t)}{S_k^{(0)}(\hat{\delta}, t)} - \left( \frac{S_k^{(1)}(\hat{\delta}, t)}{S_k^{(0)}(\hat{\delta}, t)} \right)^2 \right\} dN_k(t)
= \frac{1}{n} \sum_{h=1}^{\nu} \int_{\frac{h-1}{\nu}}^{\frac{h}{\nu}} \left\{ \frac{S_k^{(2)}(\hat{\delta}_h, t_j^{(h)})}{S_k^{(0)}(\hat{\delta}_h, t_j^{(h)})} - \left( \frac{S_k^{(1)}(\hat{\delta}_h, t_j^{(h)})}{S_k^{(0)}(\hat{\delta}_h, t_j^{(h)})} \right)^2 \right\} dN_k(t)
= \frac{1}{n} \sum_{h=1}^{\nu} \sum_{j \in D_h^{(k)}} \left\{ \frac{S_k^{(2)}(\hat{\delta}_h, t_j^{(h)})}{S_k^{(0)}(\hat{\delta}_h, t_j^{(h)})} - \left( \frac{S_k^{(1)}(\hat{\delta}_h, t_j^{(h)})}{S_k^{(0)}(\hat{\delta}_h, t_j^{(h)})} \right)^2 \right\}
\]

where \( D_h^{(k)} \) contains the indexes of individuals for which a failure for component
\( k \) occurs at time \( t_j^{(h)} \) in the subinterval \([ (h-1) \tau / \nu; h \tau / \nu ] \). Similarly, the other
elements will be given by

\[
- \frac{1}{n} \frac{\partial^2 \log L_n(\hat{\delta})}{\partial \beta_k \partial \theta_k} = \frac{1}{n} \sum_{h=1}^{\nu} \sum_{j \in D_h^{(k)}} \left\{ \frac{S_k^{(5)}(\hat{\delta}_h, t_j^{(h)})}{S_k^{(0)}(\hat{\delta}_h, t_j^{(h)})} - \frac{S_k^{(1)}(\hat{\delta}_h, t_j^{(h)}) S_k^{(3)}(\hat{\delta}_h, t_j^{(h)})}{[S_k^{(0)}(\hat{\delta}_h, t_j^{(h)})]^2} \right\}
\]

and

\[
- \frac{1}{n} \frac{\partial^2 \log L_n(\hat{\delta})}{\partial \theta_k^2} = \frac{1}{n} \sum_{h=1}^{\nu} \sum_{j \in D_h^{(k)}} \left\{ \frac{\sum_{i=1}^{n} Y_{1i}(t_j^{(h)}) Y_{2i}(t_j^{(h)})}{(1 + \theta_{kh})^2} - \left( \frac{S_k^{(3)}(\hat{\delta}_h, t_j^{(h)})}{S_k^{(0)}(\hat{\delta}_h, t_j^{(h)})} \right)^2 \right\}
\]

Since the sieve estimator was shown to be consistent in Theorem 4.1, it follows by
Corollary 3.2 that the estimator in (5.4) is a consistent estimator for the covariance
matrix.
As a consequence of the weak convergence for the processes in (5.3), such statistic converges in distribution to \( \sup_{t \in \mathcal{T}} \| W(t) \|_2^2 \overset{D}{=} \tau \sup_{u \in [0,1]} \| B(u) \|_2^2 \), where \( B \) is a vector of independent Brownian bridges defined on \([0,1]\). These two processes have the same distribution because if we consider \( t = \tau u \), for \( u \in [0,1] \), then we can write \( \mathbb{P}\{W_k(\tau u) \geq \lambda\} = \mathbb{P}\{\tau^{-1/2}W_k(u) \geq \lambda\} \). Therefore,

\[
W_k^*(\tau u) = W_k(\tau u) - (\tau u/\tau)W_k(\tau) \overset{D}{=} \tau^{1/2}W_k(u) - u\tau^{1/2}W_k(1)
= \tau^{1/2}(W_k(u) - uW_k(1)) = \tau^{1/2}B_k^*(u).
\]

Critical values for the supremum involving \( B \) for some figures of \( p \) and \( \alpha \) have been computed numerically and can be found in Sinha (1979) or DeLong (1980). If for a given \( \alpha \) we have that \( \max_{1 \leq i \leq n} \| Z(t_i) \|_2^2 \) is larger than such critical value (times \( \tau \)), then we reject the hypothesis \( H_0: \delta(t) \) is constant \( \forall t \in \mathcal{T} \); otherwise we have evidences in favor of the time-independent model studied in Chapter 3 and hence, we can proceed with such model as described in Section 3 below.

### 2.2 The two-sample case

The two-sample setting described in Examples 3.1–3.3 in Chapter 3 can be considered as a special case of the situation considered in the previous section. Assume \( K = 2 \) (two components per individual.)

Recall that in this case the coefficients \( \beta_k(t) \), \( k = 1, 2 \), represent the treatment effect. When the interest is in testing if there is a treatment effect, the dependence parameter \( \theta_k(t) \) may be considered as a nuisance parameter and the main goal will be to test not only that \( \beta_k(t) \) is time-independent but also that is is equal to zero.

An interesting feature of the two-sample problem is that, under the hypothesis that the parameters do not depend on time, the asymptotic covariance between
\( \hat{\beta}_k \) and \( \hat{\theta}_k \) can be expressed as (cf. Example 3.2, page 70)

\[
e^{2\pi_0 \pi_1 t} \int_0^t \frac{S_{12}^{Tr}(u)S_{1}^{Pl}(u) - S_{1}^{Tr}(u)S_{12}^{Pl}(u)}{\pi_0 [S_{1}^{Pl}(u) + \theta_k S_{12}^{Pl}(u)] + \pi_1 e^{2\pi_0 \pi_1 [S_{1}^{Tr}(u) + \theta_k S_{12}^{Tr}(u)]}} \alpha_k(u) \, du,
\]

where the functions in the integrand are survival functions related to the individuals in the treatment (superscript Tr) and placebo (superscript Pl) groups. Under the hypothesis of no treatment effect, the survival functions for treatment and placebo will coincide, and the expression above will be equal to zero; in other words, \( \hat{\theta}_k \) and \( \hat{\beta}_k \) will be asymptotically independent. Based on this fact, we may approach this problem considering the marginal hypotheses

\[ H_{0k}: \beta_k(t) = 0, \quad \forall t \in \mathcal{T}, \]

for \( k = 1, 2 \), so that the hypothesis of no treatment effect is given by the intersection \( H_{01} \cap H_{02} \). In this case, the alternative hypothesis will be that there is a treatment effect, this effect may be different for each component and also that it can vary as a function of the time \( t \in \mathcal{T} \).

We associate to each marginal hypothesis the process

\[ Z_k^*(t) = \sqrt{n}(\gamma_k(\tau))^{-1} \int_0^t (\hat{\beta}_k(s) - \beta_k(s)) \, ds, \]

where \( \gamma_k^2(\tau) \) is the asymptotic variance of \( \sqrt{n} \int_0^t (\hat{\beta}_k - \beta_k) \, ds \), evaluated at \( t = \tau \). Under \( H_0 \), \( \beta_k(t) = 0 \) for all \( t \); if in this case \( Z_k^* \) is large we would have evidences against the null hypothesis. In virtue of Theorem 4.2, \( Z_k^* \) converges to a Gaussian process. Moreover, from the above discussion, \( Z_1^* \) and \( Z_2^* \) are asymptotically independent.

In order to use results usually found in the literature, would be desirable to have the processes converging to standard Brownian motions. In order to overcome this problem we define

\[ W_k^{(n)}(u) = Z_k^*(\xi(u)), \quad 0 \leq u \leq 1, \]

131
where \( \xi(u) = \sup\{t: \gamma_k^2(t)/\gamma_k^2(\tau) \leq u\} \). This definition is based on a result obtained by Scott (1973) for discrete-time martingales stating that \( W_k^{(n)} \) converges weakly to a standard Brownian motion \( B_k \) on \([0, 1]\). Sen (1981b), page 31 elaborates on that and then gives a derivation for the continuous-time setup. Note that \( W_k^{(n)}(u) \) corresponds to a transformation in the time scale, that is expanded (or compressed) by the function \( \xi(u) \). However, the values assumed by such a process will be in the same range as the values assumed by \( Z_k^*(t) \) on \([0, \tau]\). Consequently,

\[
\sup_{t \in \tau} |Z_k^*(t)| = \sup_{u \in [0, 1]} |W_k^{(n)}(u)| \xrightarrow{D} \sup_{u \in [0, 1]} |B_k(u)|.
\]

The limiting process given above has a known expression for its probability, given by

\[
P\{ \sup_{t \in [0, 1]} |B(u)| \leq \lambda \} = \sum_{m=-\infty}^{+\infty} (-1)^k \left( \Phi((2m + 1)\lambda) - \Phi((2m - 1)\lambda) \right)
\approx \Phi(\lambda) - \Phi(-\lambda) = 2\Phi(\lambda) - 1,
\]

where \( \Phi \) is the standard normal distribution function.

The test statistic is defined first considering \( T_k = \sup_{t \in \tau} |Z_k^*(t)| \) and then defining

\[
T = \max_{k=1,2} T_k.
\]

Large values of \( T \) are indicative of treatment effect and hence we should reject \( H_0 \) if \( T \) is larger than a critical value \( c_\alpha \), for a given \( \alpha \). The computation \( c_\alpha \) is based on the convergence of \( T_1 \) and \( T_2 \), i.e., since \( B_1 \) and \( B_2 \) are independent standard Brownian motions,

\[
P\{ \max_{k=1,2} \sup_{u \in [0, 1]} |B_k(u)| \leq c_\alpha \} = \left[ P\{ \sup_{u \in [0, 1]} |B_1(u)| \leq c_\alpha \} \right]^2 \approx \left[ 2\Phi(c_\alpha) - 1 \right]^2,
\]

and equating the last expression to \( 1 - \alpha \) we obtain the value \( c_\alpha \).
The procedure described above is feasible as long as we replace the unknown variance function \( \gamma_k^2(t) \) by some suitable consistent estimate. From the result derived in Theorem 4.3 in Chapter 4, we may consider the estimator given in expression (4.34). In order to derive that particular expression to the two-sample case, let us assume that the sieve estimator \( \hat{\delta}_k(u) \) is computed based on a partition \( 0 = t_0 < t_1 < \cdots < t_\nu = \tau \). Then,

\[
\hat{\gamma}_k(\tau) = \int_0^\tau \left[ - \sum_{h=1}^{\nu} \frac{\partial^2 \log L_n(\hat{\delta})}{\partial \beta_{kh}^2} \right]^{-1} \, du = \sum_{h=1}^{\nu} n_{lh}^2 \left[ - \frac{\partial^2 \log L_n(\hat{\delta})}{\partial \beta_{kh}^2} \right]^{-1},
\]

where

\[
- \frac{\partial^2 \log L_n(\hat{\delta})}{\partial \beta_{kh}^2} = \int_0^\tau \text{I}_h(u) \left\{ \frac{S_k^{(2)}(\hat{\delta}_h, u)}{S_k^{(0)}(\hat{\delta}_h, u)} - \left( \frac{S_k^{(1)}(\hat{\delta}_h, u)}{S_k^{(0)}(\hat{\delta}_h, u)} \right)^2 \right\} \, dN_k(u)
\]

\[
= \sum_{l=1}^{m_h^{(k)}} \left\{ \frac{(n_{kl}^{\text{T}} + \hat{\delta}_{kh} n_{kl}^{\text{P}}) e^{\beta_{kh}}}{(n_{kl}^{\text{T}} + \hat{\delta}_{kh} n_{kl}^{\text{P}}) + (n_{kl}^{\text{P}} + \hat{\delta}_{kh} n_{kl}^{\text{T}}) e^{\beta_{kh}}} \right\}.
\]

In the above expression, \( m_h^{(k)} \) is the total number of failures related to component \( k \) at times \( t_l^{(k)} \in I_h^k \), \( l = 1, \ldots, m_h^{(k)} \), \( n_{kl}^{\text{T}} \) and \( n_{kl}^{\text{P}} \) represent the number of individuals with, respectively, the \( k \)th and both components at risk at time \( t_l^{(k)} \) for the treatment group and \( n_{kl}^{\text{T}} \) and \( n_{kl}^{\text{T}} \) are the same quantities for the placebo group. The quantities \( \hat{\beta}_{kh} \) and \( \hat{\delta}_{kh} \) are the sieve estimators associated to the interval \( I_h^k \).

Finally, we note that the estimation of \( \gamma_k^2 \) considered is based on the whole parameter space rather than under the restriction imposed by \( H_0 \). Under this circumstance, it is expected that the test will have some robustness when the null hypothesis is not true; this should not imply in extra computations, as both sieve estimators will have to be computed in any case.
2.3 A general covariance matrix

The covariance structure considered in the previous subsections may not be, in
general, a reasonable assumption. Therefore, in this subsection we consider an
empirical approach to help in deciding whether a reduction in the parameter space
is possible. Again we consider the case $K = 2$.

Since the integrated process in (5.1) converges to a Gaussian process, its
finite-dimensional distributions will be multivariate normal. We use this fact to
compute replicates of such distribution (with a proper choice for the covariance
matrix) and based on those we compute an empirical critical value for the statistic
(5.3).

As before, we assume that the time interval $\mathcal{T}$ is divided in $\nu$ subintervals,
defined by $0 < t_1 < t_2 < \cdots < t_\nu = \tau$. Based on such partition define the $4\nu$-vector

$$ x_r^T = (x_r^T(t_1), x_r^T(t_2), \ldots, x_r^T(t_\nu)), $$

where the subscript $r$ represents a particular replicate. This vector should corre-
spond to observations generated from a multivariate normal distribution with mean
0 and covariance matrix given by

$$
\Gamma^* = \begin{pmatrix}
\Gamma_{t_1} & \Gamma_{t_1 \wedge t_2} & \cdots & \Gamma_{t_1 \wedge t_\nu} \\
\Gamma_{t_2 \wedge t_1} & \Gamma_{t_2} & \cdots & \Gamma_{t_2 \wedge t_\nu} \\
\vdots & \vdots & \ddots & \vdots \\
\Gamma_{t_\nu \wedge t_1} & \Gamma_{t_\nu \wedge t_2} & \cdots & \Gamma_{t_\nu}
\end{pmatrix}
= \begin{pmatrix}
\Gamma_{t_1} & \Gamma_{t_1} & \cdots & \Gamma_{t_1} \\
\Gamma_{t_2} & \Gamma_{t_2} & \cdots & \Gamma_{t_2} \\
\vdots & \vdots & \ddots & \vdots \\
\Gamma_{t_\nu} & \Gamma_{t_\nu} & \cdots & \Gamma_{t_\nu}
\end{pmatrix}.
$$

Using the Cholesky decomposition it is possible to find a $4\nu \times 4\nu$ matrix $B$ such
that $\Gamma^* = BB^T$. Then, given the available data, we estimate $\Gamma^*$ using the estimator
considered in (4.34) and obtain the corresponding $\hat{B}$. Considering some standard
procedure we then generate a $4\nu$-vector $y_r$ from a multivariate normal
distribution with mean zero and covariance given by the identity matrix, and then we obtain

134
\( x_r = \hat{B} y_r \). Therefore, for each component of \( x_r \) we compute \( z^*_{r}(t_j) = [\hat{\Gamma}_r]^{-1/2} x_r(t_j) \) and then

\[
z_r(t_j) = z^*_{r}(t_j) - (t_j/\tau) z^*_{\nu}(t_{\nu}), \quad j = 1, \ldots, \nu.
\]

Based on these values we can compute values for the statistic

\[
T = \sup_{t \in \mathcal{T}} \| Z(t) \|^2
\]

given by \( \zeta_r = \max_{1 \leq j \leq \nu} \| z_r(t_j) \|^2 \).

Repeating the procedure described a certain number of times, say \( R \), we obtain the empirical distribution function for \( T \) based on \( \zeta_r, r = 1, \ldots, R \), and from that an empirical critical value \( c_\alpha \).

Therefore, based on the sample at hand we compute

\[
T = \sqrt{n} \hat{\Gamma}_r \sup_{t \in \mathcal{T}} \| \int_0^t (\delta(s) - \hat{\delta}) \, ds \|^2,
\]

and if \( T > c_\alpha \) we reject the hypothesis of proportionality; otherwise we conclude that there are evidences in favor of the finite-dimensional model of Chapter 3.

### 3 Tests for the finite-dimensional model

In the previous section we discussed some procedures proposed to help in deciding if a reduction in the model (in terms of the dimensionality of the parameter space) is possible. When that is the case, the problem is greatly simplified with respect to the computation of estimates and test statistics. In this section we discuss some possible approaches that could be considered to verify some hypotheses of interest.

For simplicity we consider the bivariate case, i.e., let us consider \( \delta^T = (\delta_1, \delta_2) \), where \( \delta_k = (\beta_k, \theta_k) \). Some hypotheses could be formulated as follow.
If, for example, the interest resides only in assessing treatment effect, then the hypothesis would be given by

\[ H_0 : \beta_k = 0, \quad k = 1, 2, \]  
(5.5)

and in this case \( \theta_1 \) and \( \theta_2 \) are nuisance parameters. Another situation, more related to the modelling itself, refers to testing if the components can be treated separately, in which case a traditional approach based on the univariate Cox model could be considered for each component. That would imply in verifying

\[ H_0 : \theta_k = 0, \quad k = 1, 2. \]  
(5.6)

Other hypotheses of interest could be stated as

\[ H_0 : \beta_1 = \beta_2, \]  
(5.7)

to assess if treatment effect is the same for both components and

\[ H_0 : \theta_1 = \theta_2, \]  
(5.8)

meaning the dependence structure is \emph{symmetric}, in the sense that the behavior of the ratios of the conditional hazard functions for both components is similar.

We note that in the case of any one of the hypotheses above we are testing if further reductions in the parameter space are possible and given the asymptotic properties developed in Chapter 3 for the score statistic, the traditional approaches related to the score function and MPLE can be used in those cases. Therefore, we may consider the use of the score statistic as well as simpler methods such as the Wald and likelihood ratio test statistics.

For the score statistic, we consider the result proved in Theorem 3.1, where it was shown that the process \( n^{-1/2}U(t, \delta^0) \) converges to a Gaussian process with
covariance function given by $\Sigma(t)$, $t \in T$. Consequently, when $t = \tau$, the asymptotic distribution of $n^{-1/2}U(\delta^o)$ is multivariate normal with mean zero and covariance matrix $\Sigma$. Here we are dropping $\tau$ from the expressions to simplify the notation. In order to discuss this method, let us consider the hypothesis given in 5.5. Consider a partition for the score vector such that it can be expressed as $U(\delta^o) = (U_\beta(\delta^o), U_\theta(\delta^o))$ with corresponding covariance matrix partitioned as

$$
\Sigma = \begin{pmatrix}
\Sigma_\beta & \Sigma_\delta \\
\Sigma_\delta & \Sigma_\theta 
\end{pmatrix}
$$

(5.9)

In this notation, the subscript $\beta$ (or $\theta$) refers to derivatives w.r.t. those parameters whereas the subscript $\delta$ corresponds to quantities with partial derivatives w.r.t. $\beta$ and $\theta$. Also, we may note that all sub-matrices above are diagonal.

From the multivariate normal theory we conclude that the asymptotic conditional distribution of $n^{-1/2}U_\beta(\delta)$ given $U_\theta(\delta)$ is also normal with mean zero and covariance given by [see e.g., Mardia, Kent and Bibby (1979), page 63]

$$
\Sigma^*_\beta = \Sigma_\theta - \Sigma_\delta \Sigma^{-1}_\beta \Sigma_\delta.
$$

(5.10)

Under the hypothesis $H_0$: $\beta_k = \beta^o_k = 0$, $k = 1, 2$, we compute the restricted MPLE for $\theta_k$, say, $\hat{\theta}_k(\beta^o_k)$ and estimate 5.10 by

$$
n^{-1}(J_\beta - J_\delta J^{-1}_\beta J_\delta),
$$

where $J_\beta$, $J_\delta$, and $J_\delta$ are sub-matrices of $J$ [partitioned similarly to (5.9)] given in Chapter 3, page 59 and evaluated at $\beta_k = 0$ and $\theta_k = \hat{\theta}_k(\beta^o_k)$. Given those definitions, the score statistic can be defined as

$$
Q_S = [U_\beta(\beta^o, \hat{\theta})]^T[J_\delta - J_\delta J^{-1}_\beta J_\delta]^{-1}U_\beta(\beta^o, \hat{\theta}),
$$

(5.11)

and asymptotically this statistic will follow a chi-square distribution with 2 degrees of freedom. Note that a similar approach can be considered for the hypothesis defined in 5.6.
The Wald statistic will also follow a chi-square distribution. This result is based on Theorem 3.2 where we have shown that $\sqrt{n}(\hat{\delta} - \delta^o) \Rightarrow N(0, \Sigma^{-1})$. To illustrate this case, let us consider the hypothesis given in 5.7. Denoting $c^T = [1 \ 0 \ -1 \ 0]$, it will follow that

$$\sqrt{n}(c^T\hat{\delta} - c^T\delta^o) \Rightarrow N(0, c^T\Sigma^{-1}c).$$

In this case the null hypothesis can also be stated as $H_0: c^T\delta = 0$. The Wald statistic is given by

$$Q_W = n(c^T\hat{\delta} - c^T\delta^o)^T[c^T\Sigma^{-1}(\delta)c]^{-1}(c^T\hat{\delta} - c^T\delta^o).$$

It follows from Corollary 3.2 and the Slutsky theorem that $Q_W \Rightarrow \chi^2_1$, where $\chi^2_1$ denotes the chi-square distribution with one degree of freedom. Given the results proved in Chapter 3 this asymptotic distribution can be proved similarly to the classical theory as presented, e.g., in Theorem 5.6.1 of Sen and Singer (1993). Note that $Q_W$ can also be used to test the other hypotheses presented, with corresponding modifications in the degrees of freedom for the asymptotic distribution.

Finally, the likelihood ratio statistic is also an important option in our setting. Let us consider the hypotheses in (5.5)-(5.8) formulated as $H_0: \delta \in \Omega_o$, for $\Omega_o$ representing the whole parameter space $\Omega = [\mathbb{R} \times (-1, \infty)]^2$, constrained by the restriction(s) given in $H_o$. Then, the likelihood ratio statistic is defined as

$$Q_L = -2\log\sup_{\delta \in \Omega_o} L_n(\delta) - \log\sup_{\delta \in \Omega} L_n(\delta).$$

As an example, in the case of the hypothesis (5.5), we set the value of $\beta_k$, $k = 1, 2$ to zero and obtain the MPLE $\hat{\theta}_k(\beta_k = 0)$ for $\theta_k$. Replacing this value in the log-likelihood given in (3.10) we obtain the value for $\log\sup_{\delta \in \Omega_o} L_n(\delta)$. The same procedure follows for the other hypotheses. Note that the other quantity in $Q_L$ is computed with the unrestricted MPLE for $\delta$. 

138
The statistic $Q_L$ will also follow a chi-square distribution with degrees of freedom given by the dimension of $\Omega_0$. This can be proved in our case because of the weak convergence of the MPLE to a normal distribution given in Theorem 3.2 and by the consistency of $n^{-1}J(\hat{\delta})$ given in Corollary 3.2. Then, the proof follows similarly to Theorem 5.6.1 in Sen and Singer (1993), and is based on a Taylor expansion of the log-likelihood (evaluated at $\hat{\delta}$) around $\delta^0$.

We finalize this section by noting that the statistics considered will provide a consistent test when fixed alternative hypotheses are considered. This can be clearly seen from the expression for $Q_W$, where we have a quadratic form. If $H_0$ is not true, then as $n \to \infty$ the difference between $c^T\hat{\delta}$ and $c^T\delta^0$ will be different than zero and given the factor $n$ in the expression, $Q_W \to \infty$. However, when other types of alternatives are considered, it is not clear how those statistics will behave under our model.

4 Topics for further research

In this work we consider mainly the methodology to be used in the proposed matrix-valued model. In terms of applicability of the model, it is necessary to perform numerical studies in order to assess the viability for the proposed procedures; simulation studies would be a starting point. Routines for the estimation in the finite-dimensional case can be based on the Newton-Raphson algorithm; however, more sophisticated techniques can be considered. In the infinite-dimensional case, one possible approach would be to extend the procedure presented in the appendix of Kalbfleisch and Prentice (1980) formulated for the time-dependent covariates. Those issues were not considered here but they are fundamental before the practical use of the model can be considered and they will be addressed in a near future, if possible with evaluations based also on a real dataset.
On the methodological side, would be interesting to develop a procedure to assess if the assumption that interactions of second or higher order are negligible is reasonable or not. In this work such assumption is considered without a formal justification. Another point of interest is how another approach to the infinite-dimensional situation could be considered in order to simplify and improve the results when deriving the asymptotic properties for the integrated processes. As mentioned in Chapter 4, the penalized likelihood approach could be considered.

Perhaps the simplest approach to the problem of time-dependent coefficients relies on the use of resampling techniques for deriving approximations for the asymptotic distribution of the test statistics discussed in section 2. As we saw, the assumptions on the covariance matrix may not be reasonable and in this situation bootstrap as well as jackknife techniques as discussed in Sen (1994) for the change-point problem in the univariate case could be considered and such approaches must be further investigated.

Regarding the tests described for the finite-dimensional model, would be interesting to study the behavior of the three tests presented under some types of alternative hypotheses, e.g., local alternatives.

A more difficult issue refers to the estimation for the joint survival function. In the way our model is presented in this work, its use for such an estimation would not be a reasonable choice, as the time is taken in the one-dimensional scale. Therefore, would be very valuable to extend the model to higher dimensions (in terms of the time scale). However, it has been pointed out in the literature that the counting process approach becomes surprisingly complicated in such situations. One of the reasons is that the martingale structure is more complex to define. A possible solution would be to consider weak martingales. It turns out that no central limit theorem has been derived for this case.
BIBLIOGRAPHY


141


