UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Departamento de Estatística e Investigación Operativa

MULTI-STATE MODELS FOR BIOMEDICAL RESEARCH: NEW CONTRIBUTIONS IN STATISTICAL MODELLING, SOFTWARE DEVELOPMENT, AND APPLICATIONS

Luís Filipe Meira Machado

July 2005

D. Carmen Cadarso Suárez, Profesora Titular del Departamento de Estatística e Investigación Operativa de la Universidade de Santiago de Compostela, y D. Jacobo de Uña Álvarez, Profesor Titular del Departamento de Estatística e Investigación Operativa de la Universidade de Vigo,

INFORMAN que la memoria titulada *"Multi-state models for biomedical research: new contributions in statistical modelling, software development, and applications"* ha sido realizada bajo su dirección por Luís Filipe Meira Machado, estimando que el interesado se encuentra en condiciones de optar al grado de doctor en Ciencias Matemáticas, por lo que solicitan que ésta sea admitida a trámite para su lectura y defensa pública.

Santiago de Compostela, 12 de Julio de 2005.

D. Luís Filipe Meira Machado, profesor Assistente de la Universidade do Minho (Portugal),

DECLARA que ha realizado la Tesis Doctoral titulada *"Multi-state models for biomedical research: new contributions in statistical modelling, software development, and applications"* y solicita su admisión a tramite en el Departamento de Estatística e Investigación Operativa de la Universidade de Santiago de Compostela.

To my family

Summary

This thesis contains some contributions for statistical models studying the disease progression. Methods developed in this thesis are largely motivated by the applications to the medical sciences.

Disease progression can be well described using multi-state models. These models may be considered as a generalization of the survival process where several (intermediate) events occur successively over time. Multi-state models offer a better understanding of the process of the illness, leading to a better knowledge of the evolution of the disease over time. Issues of interest include the estimation of progression rates, assessing the effects of individual risk factors, survival rates or prognostic forecasting. The influence of these intermediate events in survival is often analyzed using the Cox time-dependent regression model. This thesis contains a comprehensive review of several multi-state models for studying disease progression. Differences between these approaches and the time-dependent Cox regression model are discussed, focusing on possible advantages and disadvantages for each method. Software implementation of these models has been developed in the form of an R library.

Traditionally, statistical methods for analyzing such models depend on the Markov assumption, for which future evolution only depends on the current state. By ignoring disease history behaviour, these models may carry severe limitations which can make the model inappropriate. One alternative approach is to use a semi-Markov assumption, in which the future of the process does not depend on the current time, but only on the duration in the current state. Finally, we developed a new non-Markov approach for which the future of the process depends on the current time, but also on the

time of transition to the current state. These methodologies can be used for the estimation of transition probabilities, as well as many other interesting quantities. This approach is evaluated using a simulation study and illustrated using data from a clinical study. Some asymptotic results are given for the proposed estimators.

Acknowledgments

I would like to express my sincere thanks to my supervisors, Carmen Cadarso Suárez and Jacobo de Uña Álvarez for their encouragement, support and guidance during these last years. My thanks to Fernandez Sueiro and Willisch Domínguez from the Juan Canalejo University Hospital, and to the Portuguese Oncology Institute of Oporto for providing both the data and a clinician's perspective on the problem. Thanks are also due to the Department of Statistics and Operational Research of the University of Santiago de Compostela for their valuable teaching in the PhD courses. I would like also to thank my wife for her moral support.

I would like to acknowledge the combined financial support provided by University of Minho, Prodep grant 5.3/N/189.015/01 and by Spanish Ministry of Education & Science grant BMF2005-00818 (European FEDER support included).

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Survival analysis

This chapter is concerned with studying the time, *T*, between a well-defined time origin and a subsequent event. In biomedical applications, this is known as survival analysis, and the times may represent the survival time of a living organism or the time until a disease is cured. Although we are mainly interested with the application of these methodologies to data from epidemiological and clinical studies, these methods can also be applied to data from different areas, such as the social sciences, economics and engineering.

It is common in survival analysis for some of the data to be censored. That is, the time elapsed before the occurrence of the event is not known; it is only known that this time exceeds some value (right-censoring) or that it is inferior to some value (left-censoring). Some observations in survival analysis may not be observed for the full time to the event. For example, in a clinical study, some patients will have survived to the end of the study, while for others the survival status at the end of study is unknown because there is either loss to follow-up or they have died due to other causes unrelated to the study. In these studies, complete survival information will be available for some patients, while for others we do not know the exact survival time but rather the time elapsed from the entry in the study until last known survival time. In such cases, we say that the observation is right-censored.

Sometimes the survival time is less than that which is observed. This might occur if subjects are observed at intermittent visits, and only then it is observed that the event had occurred some time before. A typical example is the study concerning the time to recurrence of a tumour. When a patient is found to have a recurrence, the actual

time of recurrence (since the operation) is less than the observed, and the observation is said to be left-censored.

Another form of censoring is interval-censoring, which is commonly present in clinical studies where the observations are made at intermittent visits. In these cases, the exact survival time is not observed; it is only known that the event occurred within an interval of time. The observation is said to be interval-censored.

Censored observations cannot be ignored since they carry important information about the survival, but because of censoring, special methods may be required to apply some model and to analyze the data.

Generic survival data is in the form of (Y,δ), where $Y=\min(T,C)$ and $\delta=\mathbb{I}(T\le C)$ for which C are the censoring times associated with T. With random censoring, we will assume that the time of censoring and the survival time T are independent. This assumption specifies that the probability of an event occurring in the small interval $[t,t+dt[$, given that no event occur until time t, is the same for individuals in general and those whose censoring time is greater than t; that is,

$$\mathbb{P}(t\le T<t+dt\,|\,T\ge t)=\mathbb{P}(t\le T<t+dt\,|\,T\ge t,C\ge t).$$

In addition to censoring, observations may be incomplete because of left-truncation, that is, the observation is only considered if some condition bearing on subject history is met. For example, subjects may not be followed from time 0, but only from a later entry time, conditional on being alive at this entry time.

Basic concepts

Let T be a random non-negative variable representing the individual survival time from a homogeneous population.

Assume T is an absolutely continuous variable with underlying density function $f(t)$. The distribution function of T is given by

$$F(t) = \mathbb{P}(T \le t) = \int_0^t f(u)\,du . \qquad [1.1]$$

The survival function, $S(t)$, is the probability that the survival time is greater than some value t,

$$S(t) = \mathbb{P}(T > t) = 1 - F(t) . \qquad [1.2]$$

The hazard function, $\alpha(t)$, is the probability that an individual "dies" at some time t, conditional that he survived until that time. Thus, the hazard function represents the instantaneous probability that the event will occur at a given time t, and can be written as

$$\alpha(t) = \lim_{dt \to 0} \frac{\mathbb{P}\left[t \le T < t + dt \,|\, T \ge t\right]}{dt} , \qquad [1.3]$$

or $\alpha(t)\,dt \cong \mathbb{P}\left(t \le T < t + dt \,|\, T \ge t\right)$, representing the probability that an event will occur in the small time interval $[t, t + dt[$, given that no event occur until time t. The hazard function is also called risk, the failure rate, the mortality rate, or intensity.

The integrated or cumulative hazard function is defined as

$$A(t) = \int_0^t \alpha(u)\,du . \qquad [1.4]$$

We can now obtain some useful relationships between each one of these functions:

$$\alpha(t) = \frac{f(t)}{S(t)} , \qquad [1.5]$$

$$S(t) = \exp\left(-\int_0^t \alpha(u)\,du \right) , \qquad [1.6]$$

$$S(t) = \exp\left(-A(t)\right), \qquad [1.7]$$

and

$$f(t) = \alpha(t) \times \exp\left(-\int_0^t \alpha(u)du \right). \qquad [1.8]$$

Because of relations [1.1]-[1.8], the distribution of *T* can be univocally specified through any one of the following functions: $f(\cdot)$, $F(\cdot)$, $S(\cdot)$, $\alpha(\cdot)$ and $A(\cdot)$.

The Kaplan-Meier estimator

A basic task in survival analysis is the estimation of survival in the presence of censoring. If there are no censored observations in a sample of dimension *n* with observed survival times, $t_1, t_2, ..., t_n$, the most natural estimator for survival is the empirical estimator, given by

$$\widehat{S_n}(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(t_i \geq t),$$

which is unbiased and is a consistent estimate of $S(t)$.

Whenever observations are censored, alternative estimators are needed.

Kaplan and Meier (1958), obtained a nonparametric estimate of the survival function, called product-limit, which is the generalization of the empirical estimator for censored data.

Suppose there are *n* subjects with observed survival times, $t_1, t_2, ..., t_n$. Some of these observations may be right-censored or there may be tied observations. If we now assume that we have *k* events $(k \leq n)$ and $n-k$ censored observations, then we may write $t_{(1)} < t_{(2)} < ... < t_{(k)}$ for the *k* event times (order statistics) arranged in increased order. Let n_i denote the number of subjects at risk (alive and uncensored) just prior to

the event time $t_{(i)}$, and d_i the number of subjects who experienced the event at that same time.

The Kaplan-Meier estimate of $S(t)$, is defined by,

$$\hat{S}(t) = \prod_{i=1}^{j} \left(1 - \frac{d_i}{n_i}\right), \qquad [1.9]$$

for $t_{(j)} \le t < t_{(j+1)}$ and $j = 1, 2, ..., k$.

The Kaplan-Meier function is a consistent estimator of $S(t)$ which, under general conditions, can be considered a nonparametric maximum likelihood estimator. This estimator is a step function which steps down at each event time (only), with $\hat{S}(t) = 1$ for $t < t_{(1)}$. This estimator can be used in the context of survival data subjected to independent right-censoring and for left-truncation as well. In the presence of left-truncation, the estimation [1.9] is performed as earlier, but now the n_i is the number of subjects who have entered the study before time $t_{(i)}$ and are still in the study just prior to $t_{(i)}$ (Borgan, 1998b).

The variance of the Kaplan-Meier estimator is estimated by

$$\hat{\sigma}^2(t) = \hat{S}(t)^2 \sum_{i=1}^{j} \frac{d_i}{n_i(n_i - d_i)}, \qquad [1.10]$$

for $t_{(j)} \le t < t_{(j+1)}$ and $j = 1, 2, ..., k$. This estimator is known as Greenwood's formula (Greenwood, 1926).

The cumulative hazard may be estimated by the Nelson estimator (Nelson, 1969), originally introduced to check graphically the fit of parametric models. This estimator is also known as Nelson-Aalen estimator and takes the form,

$$\hat{A}(t) = \sum_{i=1}^{j} \frac{d_i}{n_i},$$

for $t_{(j)} \le t < t_{(j+1)}$ and $j = 1,2,...,k$. This estimator leads to an alternative estimator for the survivor function,

$$\hat{S}(t) = \exp\left\{-\hat{A}(t)\right\}.$$

Fleming and Harrington (1984) propose an alternative estimator to the Nelson estimator, which in the presence of ties handles more accurately,

$$\hat{A}(t) = \sum_{i=1}^{j}\left\{\sum_{s=0}^{d_i-1}\frac{1}{n_i - s}\right\}.$$

for $t_{(j)} \le t < t_{(j+1)}$ and $j = 1,2,...,k$.

The Cox proportional hazards model

In many statistical studies, we are facing problems in which the main interest is to study the relationship between variables, or more particularly, to analyze the influence of one or more covariates on the variable of interest. In general, regression models are used to relate the variable of interest with covariates.

A classical model relating the hazard function and a certain number of covariates is given by the proportional hazards model, introduced by Cox (1972). Such a model is a popular choice because it is a semi-parametric model which is readily interpreted and in which censored observations are easily accommodated. Formally, the Cox model assumes that the hazard function of an individual *i* can be written as

$$\alpha_i\left(t|Z_i\right) = \alpha_0(t)\exp\left(\beta^{\mathsf{T}} Z_i\right), \qquad [1.11]$$

where $\alpha_0(t)$ is a non-negative baseline hazard function, $Z_i = \left(Z_{i1}, Z_{i2}, ..., Z_{ip}\right)^{\mathrm{T}}$ a vector of p covariates, and $\beta^{\mathsf{T}} = \left(\beta_1, \beta_2, ..., \beta_p\right)$ the associated vector of unknown regression parameters.

Model [1.11] implies proportionality of the hazards; that is, the hazard ratio for any two individuals with covariate vectors Z_1 and Z_2 is constant over time,

$$\frac{\alpha\left(t|Z_1\right)}{\alpha\left(t|Z_2\right)} = \exp\left\{\beta^{\mathrm{T}}\left(Z_1 - Z_2\right)\right\},$$

which is independent of t, that is, the hazard function for any two patients is proportional over time. The proportional hazards assumption can be checked by plotting a graph of $\log\left(-\log\left(S\left(t\right)\right)\right)$ against t. Displayed curves must be parallel across the distinct groups; if the curves cross, then a proportional hazards model is inappropriate. A more formal way for checking this assumption was suggested by Harrel and Lee (1986).

The form of $\alpha_0\left(t\right)$ can be one known parametric distribution (Exponential, Weibull, Gamma, etc) or a nonparametric function. If we assume a particular form for $\alpha_0\left(t\right)$ then we have a completely parametric model of proportional hazards, such as the exponential model, $\alpha_0\left(t\right) = \alpha$, or the Weibull model $\alpha_0\left(t\right) = r\theta\left(\theta t\right)^{r-1}$. In these cases, the full likelihood (not shown here) can be used to obtain the parameter estimates.

Assuming model [1.11], Cox formulated a *partial likelihood* (*PL*) function that does not depend on $\alpha_0\left(t\right)$ and, therefore, allows the realization of inferences on β without the need for any assumptions on the baseline hazard function $\alpha_0\left(t\right)$. The Partial log-likelihood PL function takes the form

$$PL = \sum_{i=1}^{n} \delta_i \log\left\{\frac{\exp\left(\beta^{\mathrm{T}} Z_i\right)}{\sum_{h\in R\left(t_i\right)} \exp\left(\beta^{\mathrm{T}} Z_h\right)}\right\}, \qquad [1.12]$$

where $t_1, t_2, ..., t_n$ are the survival times for n individuals, $R\left(t_i\right)$ represents the risk set at time t_i, and δ_i is an indicator of censoring, that takes the zero value if the survival time

of the *i*th individual is censored and unity otherwise. This approach assumes no ties in the times at which events occur.

Although survival times can be considered a continuous variable, the collection of survival data often involves an error measure and therefore ties are frequently observed. To include ties, adjustments are necessary.

Suppose $t_{(1)} < t_{(2)} < ... < t_{(k)}$ are the failure times for the n individuals in the sample, and let d_i be the number of subjects observed to fail at $t_{(i)}$. Then the approximation due to Peto (1972) can be used:

$$L = \prod_{i=1}^{k} \frac{\exp\left(\beta^{\mathrm{T}} s_i\right)}{\left[\sum_{h \in R\left(t_{(i)}\right)} \exp\left(\beta^{\mathrm{T}} Z_h\right)\right]^{d_i}}, \qquad [1.13]$$

where s_i is the sum of the covariates of d_i subjects observed to fail at $t_{(i)}$, $s_i = \sum_{j=1}^{d_i} Z_{ij}$.

Other methods for adjustment of ties exist, namely, the proposals of Breslow (1974) and Efron (1977).

Sometimes there are important factors for which the proportionality of the hazards is violated for the different levels. An important generalization that does not make difficult the estimation of β consists in allowing $\alpha_0(t)$ to vary in certain subsets of the data. This is the case when the population is stratified. This generalization is useful if, for example, some covariate does not seem to have a multiplicative effect on the hazard. The *strata* divide the subjects into different groups, each one with a different baseline hazard function but with common values for the vector coefficients.

Suppose that the population is stratified in q stratus, and let $n = \sum_{j=1}^{q} n_j$ be the sample size, where n_j is the number of subjects in the *j*th stratum. We then define the hazard function for an individual in the *j*th stratum of the factor as,

$$\alpha\left(t|Z\right) = \alpha_{0j}\left(t\right)\exp\left(\beta^{\mathsf{T}} Z\right).$$

To obtain the estimates for β, we maximize $L = \prod_{j=1}^{q} L_j$ where L_j is the likelihood function [1.13] for the n_j observations in the *j*th stratum.

In many clinical studies that generate survival data, patients are observed and individual data and covariate information are collected at several occasions through a follow-up period. Covariates whose values change over time, are called *time-dependent covariates*. The introduction of these covariates in the survival process can make the patients risk change from one time point to the next as the values of the covariates change. Time-dependent covariates might represent either a qualitative change in patient's condition (presence of metastases, a change of treatment, etc.) or individual continuous information (tumour size, lymphocyte count, etc.). A classical approach for the analysis of the occurrence of such events is the *time-dependent Cox regression model* (Kalbfleisch and Prentice, 1980).

The Partial log-likelihood [1.12] is now written as,

$$PL = \sum_{i=1}^{n} \delta_i \log\left\{\frac{\exp(\sum_{j=1}^{p} \beta_j Z_{ij}(t_i))}{\sum_{h\in R(t_i)} \exp(\sum_{j=1}^{p} \beta_j Z_{hj}(t_i))}\right\},$$

where the covariates, time-dependent or not, can be updated or fixed.

It should be noticed that the time-dependent Cox model does not require proportional hazards, that is, the hazard ratio for two individuals with fixed covariate vectors may not be constant over time.

Advantages of the proportional hazards model include the ease of interpretation and its availability in the majority of statistical packages. However, this approach provides only constant estimates for the covariate effect over the whole study period, which can be considered a disadvantage of the model, since the parameterization of the effect of covariates is very restricted.

In model [1.11], the effect of the covariate Z_j is assumed to be only through a linear combination of $\beta_j Z_j$. Often there is empirical evidence that the covariate effect may vary with time. One option is to use the Cox model,

$$\alpha_i\left(t|Z_i\right)=\alpha_0\left(t\right)\exp\left(\sum_{j=1}^{p}\beta_j\left(t\right)Z_{ij}\right),$$

where $\beta(t)=\left(\beta_1(t),...,\beta_p(t)\right)^{\mathrm{T}}$ is a vector of smooth unknown functions in t.

Local methods and spline function methods are two non-linear approaches for such problems. There are two main approaches to local fitting methods: using local scoring (Hastie and Tibshirani, 1986; Gray, 1994); or using local likelihood (Hastie and Tibshirani, 1986; Hastie and Tibshirani, 1987; Gentleman and Crowley, 1991). Here we are interested in the spline function methods. Among these, regression spline models replace the $\beta_j Z_j$ by the additive model $\sum_j f_j\left(Z_j\right)$:

$$\alpha_i\left(t|Z_i\right)=\alpha_0\left(t\right)\exp\left(\sum_{j=1}^{p}f_j\left(Z_{ij}\right)\right),$$

where $f_j\left(\cdot\right),\ j=1,...,p$ are unspecified smooth covariate functions.

These models are described in detail by Durrleman and Simon (1989), and Hastie and Tibshirani (1990), using backfitting estimators.

Another possible approach is the smoothing splines (O'Sullivan, 1988; Hastie and Tibshirani, 1990). Among these, cubic smoothing splines are the most used. Smoothing is achieved by imposing a penalty for curve roughness.

Here, we shall use penalized splines, also known as P-splines, as introduced by Eilers and Marx (1996). The unknown functions $f_j,\ j = 1,..., p$ will now be specified as sums of basis functions of the form

$$f_j(x) = \sum_{s=1}^{m} a_s B_s(x),$$

where $B_s(\cdot)$ are basis functions connected to knots and m is the number of basis functions. Obviously, the basis functions and location of knots depends on the function to be fitted. Eilers and Marx (1996) give several reasons for choosing B-splines. Basic references about splines are given by de Boor (1978).

A wide literature on the subjects discussed in this section include books by Kalbfleisch and Prentice (1980), Cox and Oakes (1984) and Fleming and Harrington (1991).

1.2 Multi-state models

The experience of a patient in a survival study may be thought of as a process that involves two states, with one possible transition from the 'alive' state to the 'dead' state. In some studies, however, the state representing those patients 'alive' may be partitioned into two or more intermediate (transient) states, each of which corresponding to a particular stage in the normal progress of the illness. The influence of these events

in survival is often important in the patient's outcome and can be handled using the Cox time-dependent regression model. In such cases, the event of interest is considered as the main event, while intermediate events are often included as time-dependent covariates in a proportional hazards model. The process ends when the patient enters the (absorbing, "dead") state corresponding to the event of interest. No further observations are necessary after this event occurs. Thus, in survival analysis only two states are considered, and the event of interest is the passage from one state to another. As mentioned before, the proportional hazards model has the advantage of being easily interpreted and available in the majority of statistical packages.

Multi-state models may be considered as a generalization of the basic framework dealing with survival data where survival is the ultimate outcome of interest but where intermediate (transient) states are identified. In contrast to survival data, in these models, a sequence of events is observed, leading to more than one observation per individual. In medicine, the states might be based on clinical symptoms (e.g. bleeding episodes), biological markers (e.g. CD4 T-lymphocyte cell counts; serum immunoglobulin levels), some scale of the disease (e.g. stages of cancer or HIV infection) or a non-fatal complication in the course of the illness (e.g. transplantation). A change of state is called a transition, or an event. States can be transient or absorbing, if no transitions emerge from the state (for example, death). The state structure identifies the states and the transitions allowed between states. There is no unique structure for the series of states. Some of the commonly-used state structures are illustrated in Chapter 2.

Multi-state models have several advantages over the Cox proportional hazards model with time-dependent covariates. First, they provide a comprehensive view of the disease process, putting the incomplete information to more efficient use when portions

of the history of an individual's illness are known (efficiently handling heavily censored data). Secondly, in this framework, the so-called transition intensities provide the instantaneous hazard for movement out of one state into another. These intensity functions can be used to determine the mean sojourn time in a given state of illness, the number of individuals in different states at a certain moment, and survival proportions in each state. Finally, covariates in the transition intensities can also explain differences in the course of the illness among the population. In this way, multi-state models can reveal that different covariates affect different transitions, which would not be possible with other models, the Cox regression model for example. In fact, it is very unlikely that the risk of death in patients that received different treatments would be the same. Furthermore, the prognostic factors associated with the risk of death can be different in these groups of patients. In conclusion, multi-state models dynamically evaluate the patient's progress of disease, depending on the occurrence of intermediate events.

Recent years have seen an increasing interest in statistical methods for studying disease progression. Some diseases that have been studied using multi-state models include: HIV infection (Lagakos et al., 1988; Longini et al., 1989 and 1991 ; Gentleman et al., 1994; Satten and Longini, 1996; Aalen et al., 1997), breast cancer (Duffy and Chen, 1995; Chen et al., 1996; Pérez-ócon et al., 2001), cirrhosis of the liver (Kay, 1986; Andersen et al., 2000), leukaemia (Klein et al., 1984 and 1994; Andersen et al., 1999; Keiding et al., 2001; Chevret et al., 2000), asthma (Saint-Pierre et al., 2003), alzheimer's (Commenges et al., 2004), transplantations (Hansen et al., 1994; Klotz and Sharples, 1994; Jackson and Sharples, 2002), diabetes (Andersen, 1988), diabetic retinopathy (Andersen, 1991; Marshall and Jones, 1995), malaria (Gottschau and Hogh, 1995) and multiple sclerosis (Esbjerg et al., 1999). Multi-state models have been also

used in other fields, such as signal processing (Juang and Rabiner, 1991) and reliability analysis (Su et al., 2000).

For example, a search on PUBMED using the terms “multistate model” and/or "multi-state model" takes us to 86 published articles. Figure 1.1 shows the evolution of the use of multi-state models up to September 2005. This figure shows that multi-state models have been increasingly used for modelling survival data in the case where several events occur successively over time.

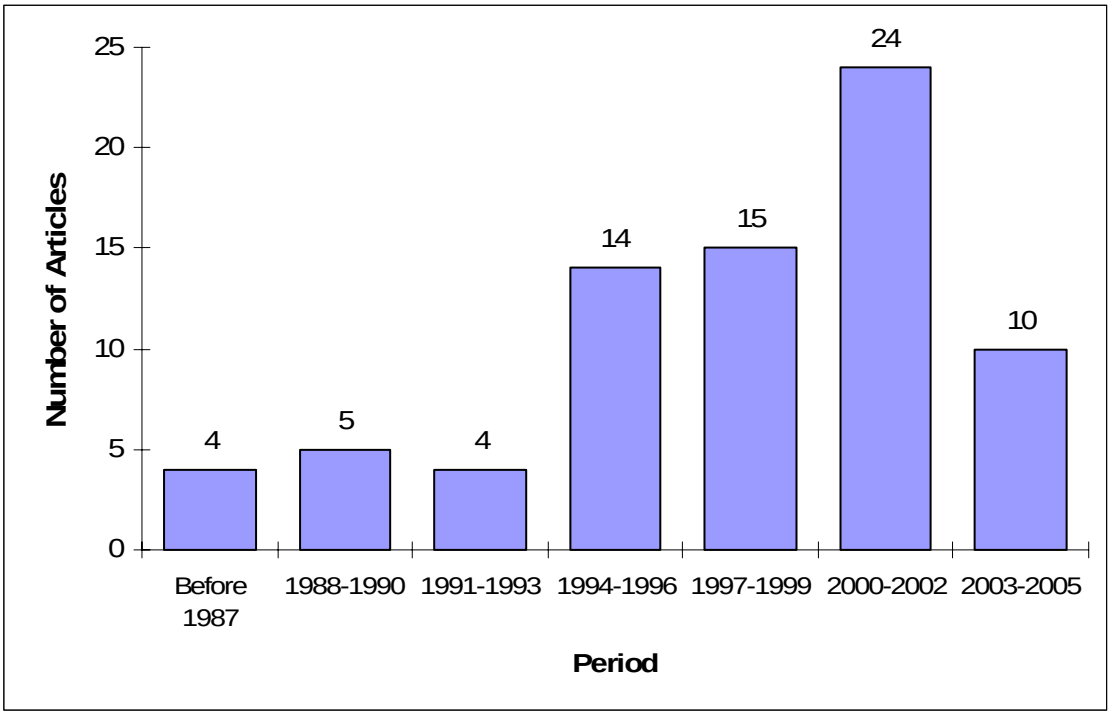


Figure 1.1. Number of published articles in recent years using multi-state models.

The Kaplan-Meier estimate of the survival function is today well-known and extensively used in the medical sciences; the proportional hazards model is also widely used whenever the interest is to study how covariates affect survival. However, in spite of the advantages mentioned previously, there are still few applications in the literature using multi-state models.

The earlier applications of multi-state models have been done with homogeneous Markov models (Kalbfleisch and Lawless, 1985; Kay, 1986). While the research on such models has been increasing in recent years, there are still few applications on non-homogeneous models or non-Markov models. Furthermore, when using multi-state models, different patterns of censoring and the different state structure have to be considered. For these purposes, flexible estimation methods have been proposed along with software. Even so, we feel that the available software for these models is still scarce, difficult to use, and most of them are self-contained. Moreover, although the use of such models has been increasingly suggested by several authors, we feel that the applications of such models are still insufficient in the literature.

In conclusion, and even with the limitations mentioned above, it is our belief that because of the development of more flexible estimation methods, and with the increasing power of numerical methods and computers, that multi-state models will be more and more applied in the medical sciences.

A wide literature on these models includes books by Andersen et al. (1993) and Hougaard (2000).

1.3 Applications

Throughout this thesis we will use some databases to illustrate all presented methodologies. The Stanford Heart Transplant dataset is one of the most widely used examples in survival analysis. An additional advantage of this data is its wide availability to researchers. Because Stanford Heart Transplant data is Markov (when we

consider time since entry in study as the time scale), we may use this database to compare some reviewed methods in Chapter 3.

One of the major aims of this thesis is to study non-Markov multi-state models. To illustrate proposed methodologies, we use data from a clinical trial of cirrhosis of the liver. PROVA clinical trial was conceived to evaluate the effect of propanolol and/or sclerotherapy versus that of no treatment on risk of bleeding and survival. This database was previously used by Andersen et al., (2000), who concluded (among other things) that the Markov assumption is not valid in this case.

In addition to these two databases we shall use our own databases. A sample of 419 subjects suffering from Psoriatic Arthritis from Galicia (Spain), and a database provided by the Portuguese Oncology Institute of Oporto to study the stomach cancer treatment in the north of Portugal.

Stanford Heart Transplantation data

The Stanford Heart Transplant study began in October 1967. The available data in Crowley and Hu's article (Crowley and Hu, 1977) covers the period up to April 1, 1974. Some patients died before an appropriate heart was found. Of the 103 patients, 69 received a heart transplant. The total number of deaths was 75; the remaining 28 patients contributed with censored survival times.

For each individual were recorded the following: an indicator of his final vital status (censored or not), the survival times from the patient's entry into the study (in days), and a covariate vector including age at acceptance, year of acceptance, previous surgery (coded as $1 = \text{yes};\ 0 = \text{no}$), and transplant (coded as $1 = \text{yes}; 0 = \text{no}$). The covariate transplant is the only time-dependent covariate, while the other covariates included are fixed.

In the context of multi-state modelling, we may consider the covariate 'transplant' as an associated state of risk, and then use multi-state models to investigate the effect of the transplantation on survival.

For the Stanford data, it is interesting to compare the hazard of death before and after the transplantation. We may also explore the potential fixed covariate effects in each of the transitions.

Among others, Turnbull, Brown and Hu (1974), Mantel and Byar (1974) and Crowley and Hu (1977) have studied the Stanford Heart data to evaluate the influence of transplantation on hazard.

PROVA clinical trial

PROVA was a Danish multi-centre clinical trial conceived to evaluate the effect of propanolol and/or sclerotherapy versus that of no treatment on risk of bleeding episodes and survival in patients with cirrhosis of the liver. This data cover the period between 1985 and 1989. In this period, 286 patients entered the study. Since their entry into the study, some patients had bleeding episodes, while others did not. The occurrence of these intermediate events may affect the patient outcome and can be included as a transient state in a multi-state model.

Before 1 January 1990, only 50 of the 286 patients developed bleeding episodes and from these, 29 died. The total number of deaths was 75.

Psoriatic Arthritis

Due to large number of people affected by Psoriasis and Psoriatic Arthritis (PsA), there is much demand for information on these chronic diseases. Here, we consider a sample of 419 subjects from Galicia (Spain). All these patients were

diagnosed as suffering from PsA, and then some treatment was considered. This data was obtained retrospectively and conceived to investigate the age of developing PsA. Here, we are interested in investigating the evolution of the disease from birth up to the current state. For this purpose, a series of stages was considered: Psoriatic, Arthritis (inflammatory arthritis), PsA, Remission and Active.

The research goals will determine the time scale to use. For example, in Psoriatic Arthritis data, since our interest focuses on the age of developing of Psoriatic Arthritis, then age is the appropriate choice for a time scale.

Stomach cancer study

We consider a sample of 314 subjects drawn from the database provided by the Portuguese Oncology Institute of Oporto. This study is concerned with stomach cancer treatment in the north of Portugal.

In this study, patient's survival times are recorded from the surgical removal of the tumour. Along with the survival time, the vital status and a covariate vector including age, sex, and clinical staging were recorded for each patient. Following removal of the tumour, 41 patients have a recurrence of stomach cancer while 53 patients fall ill as a result of metastasis to other solid organs. In this context, patients may pass from the initial state (surgical intervention time) through one of two mutually exclusive states ('metastases' and 'recurrence') to an absorbing state (death). In this study, our main goals include: (a) to obtain some informative events such as the mean sojourn time, transition rates and survival rates for each state; (b) to compare the mortality intensities; (c) to explore the potential fixed covariate effects in each of the transitions.

It should be notice that patients only enter our study if the surgical procedure was considered curative, that is to say, if after the surgery there is no evidence of illness. During the 5-year follow-up period, the number of deaths was 68; 38 of which occurred from state 'Metastases' while the remaining 30 were from the recurrent state.

1.4 Outline of the thesis.

Chapter 2 provides a deeper review of multi-state models. We introduce these models as stochastic processes; we present some standard multi-state models, we consider some censoring patterns, and we discuss the most common simplifying assumptions.

In Chapter 3, we review some possible approaches following the methodology of the multi-state Markov models. Specifically, we review the following models: (i) Cox Markov models (Andersen et al., 2000), where the transition intensities are modeled using different Cox models separately; (ii) time-homogeneous Markov models (Kay, 1986); and (iii) Markov models with piecewise constant intensities (Pérez-Ocón et al., 2001). The reviewed methods are illustrated with the data of Stanford heart transplant study, providing some guidance about the use of these methodologies for studying the evolution of the disease. We also use this database to discuss hypothesis testing procedures and methods for model checking, such as the time-homogeneous assumption or the Markov assumption. Furthermore, when examining the data, one is often interested in testing several hypotheses about the model (e.g. hypothesis about the transition intensities, regression coefficients, goodness-of-fit, etc). These checks are also

illustrated (discussed) using Stanford database. Differences between these approaches and the time-dependent Cox regression model are discussed, focusing on possible advantages and disadvantages for each method. Through this illustration, we show that multi-state models can provide new insights while confirming the results obtained when using the Cox regression model.

Special attention must be paid to evaluating the covariate effect on the hazard. As mentioned before, the Cox model provides only constant estimates of the covariate effect over the whole study period. To avoid this problem, we used spline function methods (P-splines) to obtain a dynamic Cox model. Furthermore, we introduce these spline methods in multi-state models to find out possible non-linear covariate effect on the transition intensities. The use of these methods in the scope of multi-state models is new.

While Multi-state Markov models provide non-biased estimates of the importance and ability of covariates to predict the course of the illness, these issues cannot be fully explored using Cox models alone. In order to illustrate possible benefits of using multi-state models, simulation studies were undertaken. Through these simulation studies we provide a useful description of why the Cox model can be misleading (with difficult interpretation) when used in the presence of time-dependent covariates. In this way, we show that a multi-state approach can provide a more conclusive analysis of the covariate effect than the Cox model.

Because, in some cases, Markov models do not fit satisfactorily, alternative models must be considered. These models are considered in Chapter 4. One alternative approach is to use a semi-Markov assumption. In addition to the Cox semi-Markov models, we propose a new approach for illness-death models based on less restrictive assumptions than those based on the Markov property. To date, we believe this to be the

first non-parametric modelling approach completely free from the Markov assumption. Along with theoretical properties, we have presented estimated transition probabilities and cumulative incidence functions for such models. A simulation study is performed to compare both Markovian and non-Markov approaches under a variety of scenarios. For illustration purposes, we have applied our methodology to data from the PROVA clinical trial, described above. In this chapter, we also show that these methods can be extended to more complex multi-state models. As a result of this work, a manuscript is being written and will be submitted for publication.

As previously mentioned, one main difficulty in the implementation of multi-state models is the lack of available software for these models; most of the current ones present some difficulties and limitations in practice. In some clinical studies, a model with the Markov assumption may be appropriate, while for others the semi-Markov model is preferable. In several cases, a homogeneous model will be satisfactory, while in others not. Furthermore, possible comparisons between different multi-state models are rather difficult because each of the current programs requires its own input data structure. In addition, most of the available programs only provide the regression parameters estimates and do not supply graphical output for the survival estimates nor for transition probabilities estimates. To overcome these difficulties, we have developed a user-friendly R library, called **tdc.surv**, which can be used to fit almost all the proposed models (the non-Markov model in Chapter 4 is the exception). Advantages of this software include the same data input for fitting the different models while providing the corresponding numerical and graphical outputs obtained: parameter estimates with standard errors for the covariates; transition rates; survival estimates; transition probabilities estimates; and flexible p-spline hazard ratio estimates for continuous covariates (Eilers and Marx, 1996). Moreover, users are able to include any number of

covariates on the transition intensities. In this way, users may easily analyze the results offered by the various models in order to compare them and make decisions accordingly. A manuscript about this software has been submitted for publication. A detailed description of **tdc.surv** is presented in Chapter 5.

Finally, Chapter 6 summarizes this thesis and features areas of future research.

Chapter 2

Generalities on multi-state processes

2.1. Multi-state processes.

In a general multi-state model, an individual moves from one state to another over time. The next state to which the individual moves, and the time of change, are specified through transition intensities that provide the instantaneous hazard for movement out of one state into another. These models are based on stochastic processes in continuous time allowing individuals to move between a finite number of states.

From now on, we will represent the status of an individual by a stochastic process $\{X_i(t),\ t \in \mathcal{T}\}$ with $X_i(t)$ taking a finite number of values $\{1,...,N\}$, $\mathcal{T} = [0,\tau)$, $\tau \leq \infty$, and fulfilling certain simplification assumptions. Thus, $X_i(t)$ denotes the state occupied by the *i*th subject at time *t* and $S = \{1,...,N\}$ is a finite state space. Therefore, $\{X_i(t),\ t \in \mathcal{T}\}$ have the information of the different transitions that occur to an individual over time, as well as the time at which these transitions take place.

The process starts with the distribution of the initial state given by $p_j(0) = \mathbb{P}[X(0) = j],\ j \in S$. We shall call these probabilities, $p_j(t) = \mathbb{P}[X(t) = j]$, state occupation probabilities. With the evolution of the process over time *t*, a history, or filtration, $\mathcal{F}_{t-}$, will be generated containing all information about the process over the interval $[0,t)$, such as the number of transitions up to time *t*, states previously visited, times of transitions, covariates, etc.

This multi-state process is fully characterized through transition intensities or through transition probabilities between states h and j, that we will express respectively by $\alpha_{hj}(t)$ and $p_{hj}(s,t) = \mathbb{P}(X(t) = j \,|\, X(s) = h)$. Thus, while the

transition probabilities provide important measures to make long-term predictions, each transition intensity, $\alpha_{hj}(t)$, represents the instantaneous hazard of progression to state *j* conditionally on occupying state *h*:

$$\alpha_{hj}(t) = \lim_{dt \to 0} \frac{p_{hj}(t, t+dt)}{dt} = \lim_{dt \to 0} \frac{\mathbb{P}\left(X(t+dt) = j \middle| X(t) = h\right)}{dt}.$$

This expression means that, given its prior history, the conditional probability of making a transition from state h into state j in the small time interval $[t, t+dt)$ is approximately $\alpha_{hj}(t)dt$ for small dt.

The process $X_i(\cdot)$ is then governed by an intensity matrix $Q(t)$ with (h, j) entry $\alpha_{hj}(t)$ or by a transition probability matrix, $P(s,t)$ with (h, j) entry $p_{hj}(s,t)$. For convenience, we define $\alpha_{hh}(t) = \sum_{h \neq j} \alpha_{hj}(t)$, representing the hazard associated with the distribution of the sojourn time in state h.

While the transition probabilities are important for predictive purposes, many other quantities in multi-state analysis are expressed in terms of the intensity functions. For example, the probability of no exit from state *h* at time *t*, given its history up to time *s*, is given by

$$\mathbb{P}(\text{no exit from state } h \text{ by time } t) = \exp\left\{-\int_s^t \alpha_{hh}(u)\,du\right\}.$$

Following Commenges (Commenges, 2002), we then define the stochastic process as

$$X_i(t) \sim MSM\left(\alpha_{hj}(t); h, j = 1, ..., N\right) \qquad [2.1]$$

It should be noted that, by defining $X_i(t)$ as in [2.1], we are assuming that the intensities are the same for all subjects. In practical situations, however, it might be

valuable to relate the individual characteristics with the intensity rates through a covariate vector, *Z*, possibly time-dependent. For a general regression model we can write

$$\alpha_{hji}(\cdot) = \varphi\left(\alpha_{hj0}(\cdot), \beta_{hj}^{\mathrm{T}} Z_i\right)$$

where $\alpha_{hj0}(\cdot)$ is the baseline intensity function between states *h* and *j*, β_{hj} is the vector of regression parameters, and Z_i is the covariate vector for subject *i*.

A popular choice that simplifies the model for inference is the proportional hazards assumption, which is obtained by choosing $\varphi(u(\cdot), v) = u(\cdot)e^{v}$, that is,

$$\alpha_{hji}(t) = \alpha_{hj0}(t)\exp\left(\beta_{hj}^{\mathrm{T}} Z_i\right) \qquad [2.2]$$

An alternative was proposed by Aalen for survival data and later used in multi-state models, which is obtained by choosing $\varphi(u(\cdot), v) = u(\cdot) + v$, that is,

$$\alpha_{hji}(t) = \alpha_{hj0}(t) + \beta_{hj}^{\mathrm{T}} Z_i$$

In this case, the transition intensities in [2.1] will be then represented by $\alpha_{hj}(t|Z)$.

2.2. Commonly-used multi-state models.

The complexity of a multi-state model greatly depends on the number of states defined and by the transitions allowed between these states.

The simplest form of multi-state model is the "two-state" model, or *mortality model*, for survival analysis, illustrated in Figure 2.1. In this model, we have one

transient state, “alive”, and one possible transition from this state to an absorbing state, “dead”, representing some terminal event of interest. Patients enter the study at the “alive” state and then transfer to the “dead” state by some coefficient of transition, $\alpha(t)$, at time *t*. This coefficient of transition, which we shall express as transition intensity, is a generalization of the hazard function of survival analysis, as they provide the instantaneous hazard for movement out of one state into another. Note that for the mortality model, the only relevant information in $\mathcal{F}_{t^-}$ is the individual status (alive or dead) along with the covariate history.

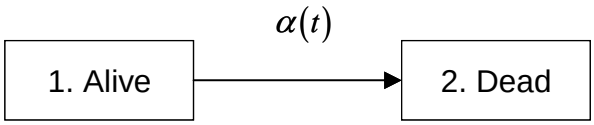


Figure 2.1: Mortality model for survival analysis.

These models have been discussed in more detail in Chapter 1.

Note that the transition (occupation) probabilities are expressed by

$$p_1(t) = p_{11}(0,t) = \exp\left\{-\int_0^t \alpha(u)\,du\right\} = S(t), \qquad [2.3]$$

and

$$p_2(t) = p_{12}(0,t) = 1 - S(t). \qquad [2.4]$$

Splitting the “Alive” state from the simple mortality model for survival data into two transient states, we therefore obtain the simplest *progressive three-state model*, illustrated in Figure 2.2. It has three states and the only possible transitions are $1 \to 2$ and $2 \to 3$. For this model, the entry time into state 2 is a relevant information in $\mathcal{F}_{t^-}$. Note that for the progressive three-state model we assume that the transition intensity

from state 2 into state 3 might depend, in some way, on the entry time in state 2, denoted by t_{12}.

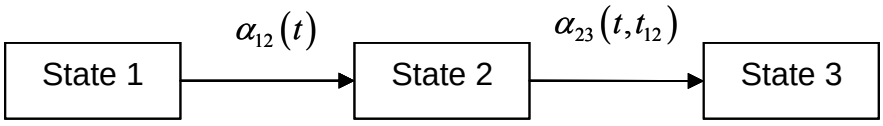


Figure 2.2: Progressive three-state model $(t_{12}$ is the entry time into state $2)$.

When we refer to progressive models, we are referring to all processes for which each state has at most one transition into it, and none into the initial state.

The state occupation probabilities are now given by,

$$p_1(t) = p_{11}(0,t) = \exp\left\{-\int_0^t \alpha_{12}(u)\,du\right\}$$

$$p_2(t) = p_{12}(0,t) = \int_0^t \left[\alpha_{12}(u)\exp\left\{-\int_0^u \alpha_{12}(v)\,dv\right\}\exp\left\{-\int_u^t \alpha_{23}(v,u)\,dv\right\}\right] du$$

$$p_3(t) = 1 - p_1(t) - p_2(t)$$

The mortality model and the progressive three-state model are particular cases of the *k-state progressive model*. This model is illustrated in Figure 2.3.

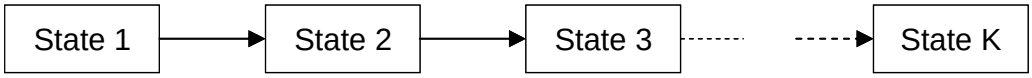


Figure 2.3: *k*-state progressive model.

Another possible and extensively used model in the literature to describe the disease progression is the *illness-death model* (also known as disability model). This model is fully characterized by three transition intensities, each one describing the

instantaneous risk of moving from one state to another, namely: the disease intensity (incidence), the death intensity without disease, and the death intensity after the occurrence of the disease. These models are widely used in the medical literature and can be used to study the incidence of the disease as well as death. One important problem here is to evaluate whether previously diseased subjects have the same risk of death as those who have been healthy all their lives. This model is applied to the Heart Transplant data in Chapter 3, and to the PROVA data in Chapter 4.

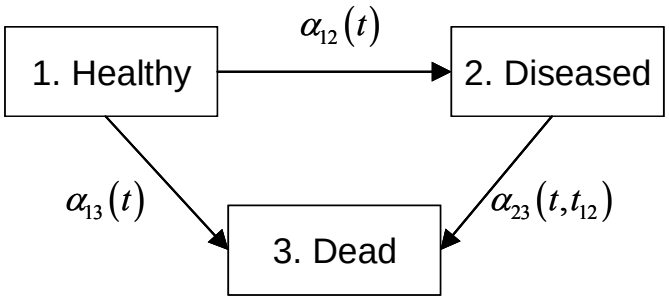


Figure 2.4: Illness-death model.

Again, we assume that the transition intensity from state 2 into state 3 might depend on the entry time in state 2. For such a model, the probability of staying in the healthy state is given by

$$p_1(t) = p_{11}(0,t) = \exp\left\{-\int_0^t \left(\alpha_{12}(u)+\alpha_{13}(u)\right)du\right\}$$

The probability of being in the diseased state is

$$\begin{aligned} p_2(t) &= p_{12}(0,t) \\ &= \int_0^t \left[\alpha_{12}(u)\exp\left\{-\int_0^u \left(\alpha_{12}(v)+\alpha_{13}(v)\right)dv\right\}\exp\left\{-\int_u^t \alpha_{23}(v,u)\,dv\right\}\right]du \end{aligned}$$

and finally, the probability of having died is

$$p_3(t) = 1 - p_1(t) - p_2(t).$$

Note that the illness-death model of Figure 2.4 is not progressive since there are two transitions into the 'death' state. One disadvantage with non-progressive models is the impossibility to know the states visited before and the visiting order just by knowing the current occupied state. Even so, in some cases the state structure can be changed so that the model can become more transparent (Hougaard, 2000). For example, when using an illness-death model, it is not possible to know the previously visited states for a patient found in the 'dead' state. We can, however, extend these models to become progressive by adding an extra state, representing death after disease. This model is depicted in Figure 2.5.

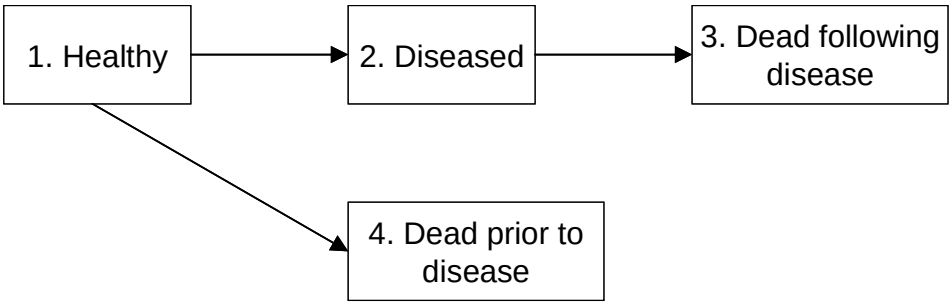


Figure 2.5: Progressive illness-death model.

In this model, if we assume that all patients enter the study in the healthy state, the current state includes the information on the states previously visited, and the order in which they have been visited.

Reversibility between the "healthy" state and the "diseased" state may be considered although this will make the analysis more difficult. These models allow individuals to move back and forth between the two states. Such a model is referred to as the illness-death model with recovery (see Figure 2.6).

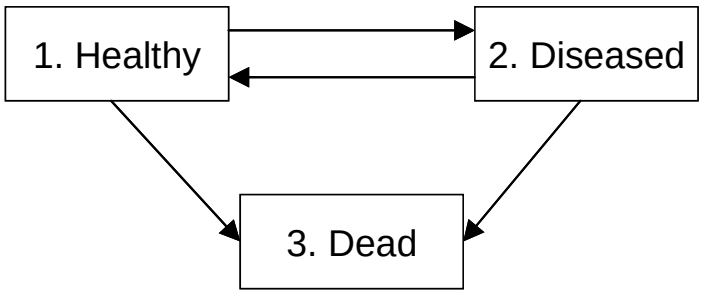


Figure 2.6: Illness-death model with recovery

The competing risks model (Andersen et al., 2002) is another multi-state model which extends the simple mortality model for survival data in which each individual may 'die' due to any of several causes. As shown in Figure 2.7, there is one transient "alive" state and several absorbing "death" states corresponding to different causes of death.

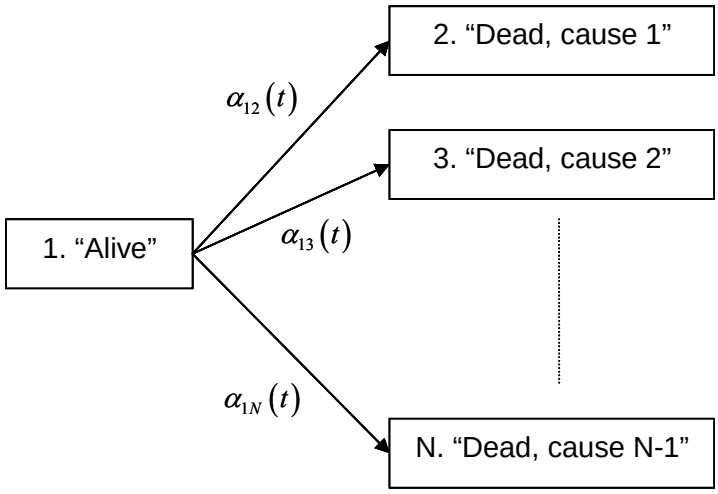


Figure 2.7: Competing risks model.

For the competing risks model, the hazard for the total mortality is given by $\alpha(t) = \sum_{j=2}^{N} \alpha_{1j}(t)$. Thus, the survival function is given by

$$S(t) = \exp\left\{-\int_0^t \alpha(u)\,du\right\}$$

Then, the occupation probabilities are given by

$$p_1(t) = p_{11}(0,t) = S(t)$$

and

$$p_j(t) = p_{1j}(0,t) = \int_0^t \alpha_{1j}(u) \exp\left\{-\int_0^u \sum_{l=2}^{N} \alpha_{1l}(v)\, dv\right\} du$$

The bivariate model (Hougaard, 2000), depicted in Figure 2.8, is the multi-state model for bivariate parallel data, with states 'both alive', 'individual 1 dead', 'individual 2 dead' and 'both dead'.

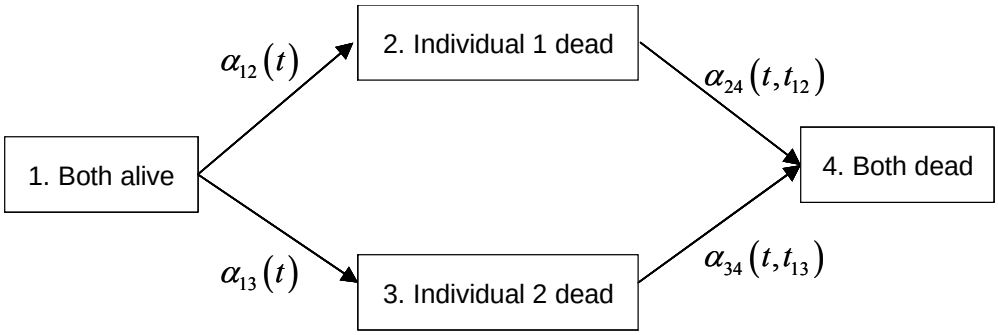


Figure 2.8: The bivariate model.

Although model depicted in Figure 2.8 is non-progressive, we can however, make this model progressive using a similar procedure as those shown above for the illness-death model.

For the bivariate model, the occupation probabilities are given by

$$p_1(t) = p_{11}(0,t) = \exp\left\{-\int_0^t \left(\alpha_{12}(u) + \alpha_{13}(u)\right) du\right\},$$

$$p_2(t) = p_{12}(0,t)$$
$$= \int_0^t \left[\alpha_{12}(u) \exp\left\{-\int_0^u \left(\alpha_{12}(v) + \alpha_{13}(v)\right) dv\right\} \exp\left\{-\int_u^t \alpha_{24}(v,u)\, dv\right\}\right] du,$$

$$p_3(t) = p_{13}(0,t)$$
$$= \int_0^t \left[\alpha_{13}(u) \exp\left\{ -\int_0^u \left(\alpha_{12}(v) + \alpha_{13}(v) \right) dv \right\} \exp\left\{ -\int_u^t \alpha_{34}(v,u)\, dv \right\} \right] du$$

and

$$p_4(t) = 1 - p_1(t) - p_2(t) - p_3(t).$$

The choice of the appropriate state structure depends on the available data as well as the research goal. Goals for the data analysis will be determinant for the choice of the events and the time scale. For example, in the Psoriatic Arthritis data we are interested in studying the evolution of the disease through age. Thus, age is the appropriate choice for the time scale. An overview of different time scale can be found in Keiding (1991).

2.3. Sampling times.

In medical applications, measurements of a disease are often made at several times, providing incomplete observations in some way. Because the process cannot be observed over an infinite time period, at times it will be ended before an absorbing state is reached. In these situations the whole trajectory of the process was not observed, providing in this way right-censored observation times. It can also happen that the process is not observed from its origin, leading to left-censored observations. Usually, patients are observed at intermittent follow-up visits, at which time individual data and covariate information are collected, but the information from the periods between visits

is not available. In such cases the transition times of movement between states are not exactly observed and the states occupied in between visits are not known; these observations are said to be interval-censored.

Figure 2.9, illustrates a possible evolution of a patient in the illness-death model. The process of Figure 2.9 is observed at four occasions during the time period of 9 months, being the states occupied at times 0, 3, 6 and 9 months, the only information available. Following this figure we do not know the exact time of transition between the "Healthy" state and the "Diseased" state. More, if the process has not been observed at time 6, for some reason, the transition between the "Healthy" state and the "Diseased" state was not observed at all.

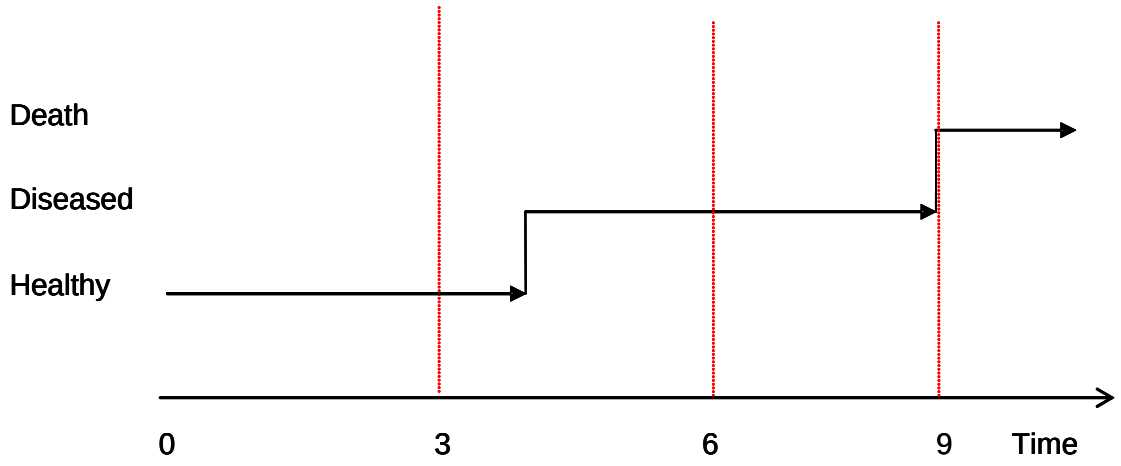


Figure 2.9: Evolution of an illness-death multi-state model.

Also frequent is the presence of left-truncation when dealing with multi-state models. For example, a common occurrence in the illness-death model is that for one patient to be selected, he must be in the "Healthy" state. Note that censoring and truncation are two different concepts. Censoring indicates that the information about the process is only partially observed, whereas truncation represents an exclusion of a part

of the information because it is not observed. In most cases, the truncation and censoring mechanisms are assumed independent from the process.

Ideal data collection should record every event or transition along with transition times and covariates.

2.4. Common simplifying assumptions.

Various possible models for the transition rates between states can be accommodated in expression [2.1]. Transition intensities can depend on the states previously visited, the time since the last event, covariates, etc. Furthermore, they may be constant over time or not. The most common models (see Figure 2.10) are characterized by one of the following assumptions:

1. *Time-Homogeneity*: the intensities are constant over time, that is, transition intensities are independent of t. Therefore we have $\alpha_{hj}\left(t|Z\right)=\alpha_{hj}\left(Z\right)$.

2. *The Markov assumption*: future evolution only depends on the current state and not on the previous history of the individual. That is, transition intensities do not depend on other information prior to time *t*.

3. *The semi-Markov assumption*: future evolution not only depends on the time *t* since origin, but also on the time spent in the current state *h*, that is, $t-t_{hj}$, where t_{hj} is the transition time from *h* to *j*. If we assume, in addition, that the transitions do not depend directly on *t*, we will have intensity functions of the general form $\alpha_{hj}\left(t-t_{hj}|Z\right)$.

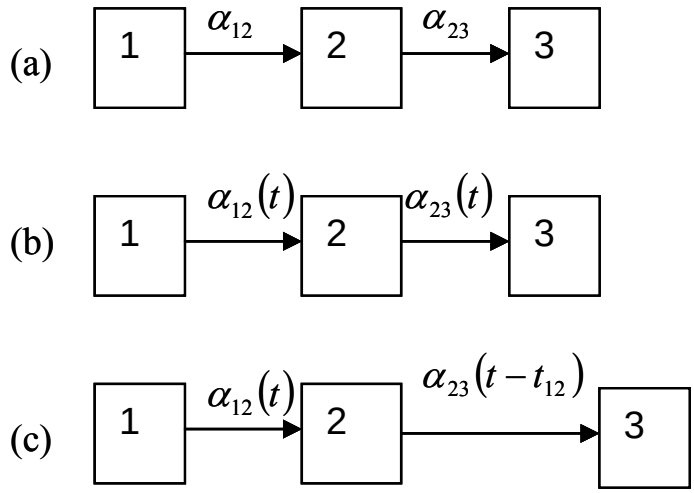


Figure 2.10: Three-state progressive models.
(a) Homogeneous Markov model; (b) Non-homogeneous Markov model;
(c) Semi-Markov model (t_{12} is the transition time from state 1 to state 2).

Because of their simplicity, Markov models are the most used in the literature. In fact, if the state structure has a large number of states, non-Markovian models can rapidly become complicated. Note that $\{X(t),\ t \geq 0\}$ is a Markov process, if for any $s,\ t$ with $0 \leq s < t$, and $h, j, x(u) \in \{1, 2, ..., n\}$ with $h \leq j$, we have

$$\mathbb{P}\left[X(t) = j \middle| X(s) = h, X(u) = x(u), 0 \leq u < s \right] = \mathbb{P}\left[X(t) = j \middle| X(s) = h \right].$$

Thus, the future of the process after time *s* depends only on the state occupied at time *s*.

Traditionally, semi-Markov models are considered when the Markov assumption is violated. For these models the transition intensities depend on $\mathcal{F}_{t-}$ only through the state currently occupied, the elapsed time since the transition and the covariates. In Chapter 4 we consider a different approach to non-Markov modelling, for which, in addition to the current time, the transition intensities are allowed to depend also on the entry time on the current state.

By choosing the best state structure for the data, the assumption about the model can be more transparent, for example, making the process a Markov process. For

example, in some models, the states visited before, along with their order, are implicit just by knowing the currently occupied state. We have seen above that, in some models, the state structure can be changed so that non-progressive models become progressive just by adding extra states.

Finally, note that the mortality model and the competing risks model are trivially Markov.

Chapter 3

Multi-state Markov models

3.1. Introduction.

While in the previous chapter we discussed generalities of the multi-state models, such as the state structure and the most common assumptions, in the present chapter and the next, we will go more deeply into statistical modelling. Here, we consider those models relying on the Markov assumption. The dependence of Markov models only on the current state and not on the previous history of the individual may be seen as advantage for such models. The simplicity of these models can particularly be helpful for evaluating the transition probabilities. In this chapter, we will review some of the commonly-used multi-state Markov models.

Let $\{X_i(t),\ t \in \mathcal{T}\}$ be the stochastic process with finite state space $S = \{1,...,N\}$. Using the theorem of total probability, for $0 \le s < u < t$, we may write,

$$\mathbb{P}\left[X(t) = j \middle| X(s) = h\right] = \sum_{k \in S} \mathbb{P}\left[X(t) = j \middle| X(u) = k, X(s) = h\right] \mathbb{P}\left[X(u) = k \middle| X(s) = h\right].$$

If we assume that the process is Markovian, then the state of the process depends only on the current state. Thus the above equality can be simplified by

$$\mathbb{P}\left[X(t) = j \middle| X(s) = h\right] = \sum_{k \in S} \mathbb{P}\left[X(t) = j \middle| X(u) = k\right] \mathbb{P}\left[X(u) = k \middle| X(s) = h\right].$$

We therefore obtain the Chapman-Kolmogorov equation,

$$p_{hj}(s,t) = \sum_{k \in S} p_{hk}(s,u)\, p_{kj}(u,t), \quad 0 \le s < u < t.$$

We may now derive Kolmogorov's equations. Because

$$p_{hj}(s,t+dt) - p_{hj}(s,t) = \sum_{k \in S} p_{hk}(s,u)\left\{p_{kj}(u,t+dt) - p_{kj}(u,t)\right\},$$

dividing both sides by dt and taking the limit as $dt \to 0$, we obtain

$$\frac{\partial}{\partial t} p_{hj}(s,t) = \sum_{k \in S} p_{hk}(s,u) \frac{\partial}{\partial t} p_{kj}(u,t).$$

Then taking $u \to t$ we obtain the so-called Kolmogorov's forward equation

$$\frac{\partial}{\partial t} p_{hj}(s,t) = \sum_{k \neq j} p_{hk}(s,t)\alpha_{kj}(t) + p_{hj}(s,t)\alpha_{jj}(t).$$

Thus, in Markov models, the transition probability matrix can be calculated from the intensity matrix, Q, by solving the forward Kolmogorov differential equation,

$$\frac{dP(t)}{dt} = P(s,t)Q(t)$$

For the illness-death model, we have the following equations:

$$\frac{dp_{11}}{dt}(s,t) = -p_{11}(s,t)[\alpha_{12}(t) + \alpha_{13}(t)], \qquad [3.1]$$

$$\frac{dp_{12}}{dt}(s,t) = p_{11}(s,t)\alpha_{12}(t) - p_{12}(s,t)\alpha_{23}(t), \qquad [3.2]$$

and

$$\frac{dp_{22}}{dt}(s,t) = -p_{22}(s,t)\alpha_{23}(t). \qquad [3.3]$$

The solutions of these equations are:

$$p_{11}(s,t) = e^{-(A_{12}(s,t)+A_{13}(s,t))},$$

$$p_{22}(s,t) = e^{-A_{23}(s,t)},$$

and

$$p_{12}(s,t) = \int_s^t p_{11}(s,u)\alpha_{12}(u)p_{22}(u,t)du;$$

where $A_{hj}(s,t) = \int_s^t \alpha_{hj}(u)du$ is the cumulative or integrated intensity between *s* and *t*.

3.2. Homogeneous Markov models.

Because homogeneous Markov models ignore both time and history, they are often referred to as the "simplest case" multi-state model. In fact, earlier applications have been done with these models. Such models have been used by various authors to analyse different diseases (breast cancer, liver cancer, leukaemia, diabetes, etc).

In homogeneous Markov models, all transitions are assumed to be constant as functions of time. Therefore, each transition probability $p_{hj}(s,t)$ depends only on the time difference $t-s$, that is, $p_{hj}(s,t) = p_{hj}(0,t-s)$. To simplify notation, we will use a matrix depending on only one argument in time $p_{hj}(t-s) = p_{hj}(0,t-s)$. In other words, $p_{hj}(t) = \mathbb{P}\left[X(t+u) = j \middle| X(u) = h\right]$ for any $u \geq 0$. We can rewrite Kolmogorov's forward equation as

$$\frac{\partial}{\partial t} p_{hj}(t) = \sum_{k \neq j} p_{hk}(t)\alpha_{kj} + p_{hj}(t)\alpha_{jj}.$$

Then, the transition probabilities can be simply expressed in terms of the transition intensities, through the Kolmogorov relation $P(t) = \exp(tQ)$ (Cox and Miller, 1965). For the illness-death model, the intensity matrix is given by

$$Q = \begin{pmatrix} -(\alpha_{12}+\alpha_{13}) & \alpha_{12} & \alpha_{13} \\ 0 & -\alpha_{23} & \alpha_{23} \\ 0 & 0 & 0 \end{pmatrix}$$

Solutions for equations [3.1]-[3.3] are now the following:

$$p_{11}(t) = e^{-(\alpha_{12}+\alpha_{13})t},$$

$$p_{12}(t) = \frac{\alpha_{12}\left(e^{-\alpha_{23}t} - e^{-(\alpha_{12}+\alpha_{13})t}\right)}{\alpha_{12}+\alpha_{13}-\alpha_{23}},$$

$$p_{22}(t) = e^{-\alpha_{23}t},$$

$$p_{13}(t) = 1 - e^{-(\alpha_{12}+\alpha_{13})t} + \frac{\alpha_{12}\left(e^{-(\alpha_{12}+\alpha_{13})t} - e^{-\alpha_{23}t}\right)}{\alpha_{12}+\alpha_{13}-\alpha_{23}},$$

and

$$p_{23}(t) = 1 - e^{-\alpha_{23}t}.$$

Equations for more than three states are more complex and require using special software.

For homogeneous processes, expression [2.1] reduces to $X_i(t) \sim MSM\left(\alpha_{hj}(Z); h, j = 1,...,N\right)$, and then we may use Cox proportional hazards models of type

$$\alpha_{hj}(Z) = \alpha_{hj} \exp\left(\beta_{hj}^{\mathrm{T}} Z\right), \qquad [3.4]$$

to relate the transition intensities α_{hj} with covariates *Z*.

General Likelihood

Under model [3.4], inference is conducted by a general likelihood, which is derived as follows. Assuming that the stochastic process $X_i(\cdot)$ is observed at times $t_{i,0} < t_{i,1} < ... < t_{i,m_i}$, where $i = 1,...,n$ are the indexed individuals, let us consider a general multi-state model, with a pair of states consecutively observed $\left(X_i(t_{i,r}),\ X_i(t_{i,r+1})\right)$. The general likelihood is then the product of all the terms over all the individuals and all the transitions (Kay, 1986),

$$L = \prod_{i=1}^{n} \prod_{r=0}^{m_i-1} l_{i,r},$$

where $l_{i,r} = p_{X_i(t_{i,r}),X_i(t_{i,r+1})}\left(t_{i,r+1} - t_{i,r}\right)$ is the contribution to the likelihood for the *i*th individual for the pair of states $\left(X_i\left(t_{i,r}\right),\ X_i\left(t_{i,r+1}\right)\right)$ observed.

3.3. Non-homogeneous Markov models.

In some applications, the hypothesis of homogeneity may be unrealistic since the illness tends to evolve over time. In these occasions, a non-homogenous model is then recommended. Several non-parametric approaches to non-homogeneous processes have been proposed in the literature (Frydman, 1995; Joly et al., 2002). An alternative (parametric) procedure consists of partitioning the whole study period in two or more intervals and then fitting a piecewise constant intensities model (Pérez-Ocón et al., 2001; Saint-Pierre et al., 2003), leading to transition intensity functions as step functions. Such a model will be described in detail in the next section.

3.3.1. Piecewise homogeneous Markov model.

Here we consider a piecewise process where the transition intensities are defined by stepwise constant functions of type:

$$\alpha_{hj}\left(t\middle|Z\right) = \begin{cases} \alpha_{hj}^{1}\left(Z\right), & 0 \le t \le \theta_1 \\ \alpha_{hj}^{l}\left(Z\right), & \theta_{l-1} < t \le \theta_l \end{cases}, \ l = 2, 3, ..., k$$

where $\theta = (\theta_1,...,\theta_{k-1})$ is the vector of cut-off points $(\theta_1 < \theta_2 < ... < \theta_{k-1} < \theta_k = \infty)$. Transition intensities $\alpha_{hj}^{l}(Z)$, $l = 2,3,...,k$ are often expressed according to the Cox model [3.4].

The likelihood function is built following parametric methodological procedures. Let us assume *k*-1 cut points: $0 = \theta_0, \theta_1, \theta_2, ..., \theta_k = \infty$ and let us define one piecewise intensities matrix:

$$Q(t|Z) = \begin{cases} Q_1(Z) & , \theta_0 \le t < \theta_1 \\ Q_l(Z) & , \theta_{l-1} \le t < \theta_l \end{cases} \qquad l = 2,3,...,k$$

Assuming that the process $X_i(\cdot)$ is observed at times $t_{i,0} < t_{i,1} < ... < t_{i,m_i}$, the likelihood is expressed as $L = \prod_{i=1}^{n} \prod_{r=0}^{m_i - 1} l_{i,r}$, where $l_{i,r} = p_{X_i(t_{i,r}),X_i(t_{i,r+1})}(t_{i,r+1} - t_{i,r})$ is the contribution to the likelihood for the *i*th individual for the pair of states $(X_i(t_{i,r}),\ X_i(t_{i,r+1}))$ observed.

Let us define the following intervals: $B_q = [\theta_q, \theta_{q+1}[$ and $C_s =]\theta_s, \theta_{s+1}]$ with $q, s = 1,2,...,k-1$. Then, each contribution $l_{i,r}$ is constructed as follows:

1. if $t_{i,r} \in B_q$ and $t_{i,r+1} \in C_q$, then

$$l_{i,r} = p^{Q_q(Z_i)}_{X_i(t_{i,r}),X_i(t_{i,r+1})}(t_{i,r+1} - t_{i,r});$$

2. if $t_{i,r} \in B_q$ and $t_{i,r+1} \in C_{q+1}$, then

$$l_{i,r} = p^{Q_q(Z_i)}_{X_i(t_{i,r}),X_i(t_{i,r})}(\theta_{q+1} - t_{i,r}) \times p^{Q_{q+1}(Z_i)}_{X_i(t_{i,r}),X_i(t_{i,r+1})}(t_{i,r+1} - \theta_{q+1});$$

3. if $t_{i,r} \in B_q$ and $t_{i,r+1} \in C_l$ with $l - q \geq 2$, then

$$l_{i,r} = p^{Q_q(Z_i)}_{X_i(t_{i,r}),X_i(t_{i,r})}(\theta_{q+1} - t_{i,r}) \times \left(\prod_{u=q}^{l-2} p^{Q_{u+1}(Z_i)}_{X_i(t_{i,r}),X_i(t_{i,r})}(\theta_{u+1} - \theta_u) \right) \times p^{Q_{l-1}(Z_i)}_{X_i(t_{i,r}),X_i(t_{i,r+1})}(t_{i,r+1} - \theta_{l-1}).$$

Assuming an illness-death model, the survival function, $S(t|Z)$ is expressed as:

1. If $t \in [\theta_0, \theta_1[$, $S(t \mid Z) = p_{11}^{Q_1(Z)}(t \mid Z) + p_{12}^{Q_1(Z)}(t \mid Z)$;

2. If $t \in [\theta_1, \theta_2[$,

$$S(t \mid Z) = p_{11}^{Q_1(Z)}(\theta_1 \mid Z) \times \left\{ p_{11}^{Q_2(Z)}(t - \theta_1 \mid Z) + p_{12}^{Q_2(Z)}(t - \theta_1 \mid Z) \right\} + \\ + p_{12}^{Q_1(Z)}(\theta_1 \mid Z) \times p_{22}^{Q_2(Z)}(t - \theta_1 \mid Z);$$

3. If $t \in [\theta_{q-1}, \theta_q[$ with $q \geq 3$,

$$S(t \mid Z) = \left(\prod_{u=1}^{q-1} p_{11}^{Q_u(Z)}(\theta_u - \theta_{u-1} \mid Z) \right) \times \left\{ p_{11}^{Q_q(Z)}(t - \theta_{q-1} \mid Z) + p_{12}^{Q_q(Z)}(t - \theta_{q-1} \mid Z) \right\} + \\ + p_{12}^{Q_1(Z)}(\theta_1 \mid Z) \times \left(\prod_{u=2}^{q-1} p_{22}^{Q_u(Z)}(\theta_u - \theta_{u-1} \mid Z) \right) \times p_{22}^{Q_q(Z)}(t - \theta_{q-1} \mid Z) + \\ + \sum_{i=2}^{q-1} \prod_{u=1}^{i-1} \left\{ p_{11}^{Q_u(Z)}(\theta_u - \theta_{u-1} \mid Z) \times p_{12}^{Q_i(Z)}(\theta_i - \theta_{i-1} \mid Z) \times \right. \\ \left. \times \prod_{u=i+1}^{q-1} p_{22}^{Q_u(Z)}(\theta_u - \theta_{u-1} \mid Z) \right\} \times p_{22}^{Q_q(Z)}(t - \theta_{q-1} \mid Z).$$

The assumption of time-homogeneity may be assessed using a piecewise model (Kay, 1986). Likelihood ratio tests can be used to compare the piecewise model with the time homogeneous model. Under the null hypothesis of homogeneity, the test statistic has approximately a χ^2_{k-r} distribution, where r is the number of parameters under H_0 and k the number of parameters under H_1. Alternatively, the local score test can be used. A complete description of this method can be found in Kalbfleisch and Lawless (1985).

3.3.2. Cox Markov models.

For simplicity, assume the illness-death model of Figure 2.4. The transition intensities, $\alpha_{hj}(t|Z)$, are modelled using Cox-like models of the form

$$\alpha_{hji}(t|Z_i) = \alpha_{hj0}(t)\exp\left(\beta_{hj}^{\mathrm{T}} Z_i\right), \qquad [3.5]$$

assuming the process to be *Markovian*.

For the hazard of 'death' without disease, $\alpha_{13}(t|Z)$, survival times from diseased patients are taken as censored in disease time. Patients who are healthy also contribute with censored survival times. For the disease intensity, $\alpha_{12}(t|Z)$, the final point is the time of the beginning of the disease. Survival times of patients who did not become diseased are taken as censored, whether they are alive or whether they have died without having been diseased. Finally, to model $\alpha_{23}(t|Z)$, the death intensity after the occurrence of the disease, we only enter the survival times truncated on disease time, censored or not, of the individuals that experienced the disease. Note that patients are at risk only after entering state 2.

Let now $\widehat{A_{hj}}(t|Z) = \widehat{A_{hj0}}(t)\exp\left(\widehat{\beta_{hj}^{\mathrm{T}}} Z\right)$ be the estimate of the cumulative intensity function with $\widehat{A_{hj0}}(\cdot)$ the Breslow estimator for $A_{hj0}(t) = \int_0^t \alpha_{hj0}(u)\,du$.

The estimation of the transition probabilities $p_{hj}(s,t|Z) = \mathbb{P}\left(X(t) = j \middle| X(s) = h, Z\right)$ $s \le t$ and $h \le j$ for a given covariate vector Z, are expressed in the following way:

$$\widehat{p_{11}}(s,t|Z) = \prod_{s<u\le t}\left(1 - \sum_{j=2}^{3} d\widehat{A_{1j}}(u|Z)\right),$$

that is

$$\widehat{p_{11}}(s,t|Z) = \widehat{p_{11}}(s,t-|Z)\left(1-d\widehat{A_{12}}(t|Z)-d\widehat{A_{13}}(t|Z)\right),$$

$$\widehat{p_{12}}(s,t\,|\,Z) = \sum_{u\le t}\widehat{p_{11}}(s,u-|\,Z)\,d\widehat{A_{12}}(u|Z)\,\widehat{p_{22}}(u+,t\,|\,Z),$$

where

$$\widehat{p_{22}}(s,t\,|\,Z) = \prod_{s<u\le t}\left(1-d\widehat{A_{23}}(u|Z)\right),$$

therefore,

$$\widehat{p_{12}}(s,t\,|\,Z) = \widehat{p_{11}}(s,t-|\,Z)\,d\widehat{A_{12}}(t|Z)+\widehat{p_{12}}(s,t-|\,Z)\left(1-d\widehat{A_{23}}(t|Z)\right),$$

$$\widehat{p_{13}}(s,t\,|\,Z) = 1-\widehat{p_{11}}(s,t\,|\,Z)-\widehat{p_{12}}(s,t\,|\,Z),$$

and

$$\widehat{p_{23}}(s,t\,|\,Z) = 1-\widehat{p_{22}}(s,t\,|\,Z).$$

Note, however, that the probability of having died without ever being diseased can be estimated by

$$\widehat{p_{13}^{nd}}(s,t\,|\,Z) = \sum_{s<u\le t}\widehat{p_{11}}(s,u-|\,Z)d\widehat{A_{13}}(u|Z).$$

The conditional survival probability $S(t|Z)$ which is defined as $S(t|Z) = p_{11}(0,t|Z)+p_{12}(0,t|Z)$, can now be estimated by

$$\hat{S}(t|Z) = \widehat{p_{11}}(0,t|Z)+\widehat{p_{12}}(0,t|Z),$$

with changes in its values at observed failure times.

Note that in some cases we may assume some conditions about the baseline hazards. For example, for the illness-death model, one approach that is often used is to assume the baseline hazards for transition $1\to 3$ and for the $2\to 3$ transition to be proportional. In such cases, the model is given by

$$\alpha_{13}(t|Z) = \alpha_{130}(t)\exp\left(\beta_{13}^{\mathrm{T}} Z\right)$$

and

$$\alpha_{23}(t|Z) = \alpha_{130}(t)\exp\left(\beta_{23}^{\mathrm{T}} Z + \delta\right).$$

These models can be fitted using most of the statistical packages as long as we use a counting process notation, representing each patient with several observations. More details on these models, such as the variance of the estimated transition probabilities, were discussed in the book of Andersen et al. (1993).

3.3.3. Non-parametric Markov models.

In this section we will present the simple case of a nonparametric Markov model without covariates. This approach can be thought of as the generalization of the Kaplan-Meier estimate of the simple mortality model. For survival data, the transition probability from the 'alive' state into the 'dead' state may be estimated using the Kaplan-Meier estimator (see [1.9] and [2.4]). In this section, we propose a generalization of this approach to general multi-state models with a finite number of states. Such a generalization was considered by Aalen and Johansen (1978), and is denoted as the Aalen-Johansen estimator. Again, for simplicity, we start by assuming the illness-death model, and later extend such an approach to general models. Note that for the illness-death model, it is enough to consider the transition probabilities $p_{11}(s,t)$, $p_{12}(s,t)$ and $p_{22}(s,t)$; all the others can be obtained from these ones. Because the process is Markov,

$$p_{11}(s,t) = \exp\left\{-\int_s^t \alpha_{12}(u) + \alpha_{13}(u)\, du\right\}$$

$$p_{22}(s,t) = \exp\left\{-\int_s^t \alpha_{23}(u)\, du\right\}$$

$$p_{12}(s,t) = \int_s^t p_{11}(s,u)\,\alpha_{12}(u)\, p_{22}(u,t)\, du$$

Assume that we have a sample of n subjects with observed disease and death times, $t_1, t_2, ..., t_n$. If we now assume that we have k events and $n-k$ censored observations, then we may write $t_{(1)} < t_{(2)} < ... < t_{(k)}$ for the k event times arranged in increased order. Let now n_{1i} and n_{2i} denote the number of healthy and disease subjects, respectively, just prior to the event time $t_{(i)}$. Further, let d_{12i} be the number of subjects who become diseased at time $t_{(i)}$, while d_{13i} and d_{23i} denote, respectively, the numbers of healthy and diseased subjects who die at that same time. Then the above transition probabilities may be estimated by

$$\widehat{p_{11}}(s,t) = \prod_{s<t_{(i)}\le t}\left(1 - \frac{d_{12i} + d_{13i}}{n_{1i}}\right), \qquad [3.6]$$

$$\widehat{p_{22}}(s,t) = \prod_{s<t_{(i)}\le t}\left(1 - \frac{d_{23i}}{n_{2i}}\right), \qquad [3.7]$$

and

$$\widehat{p_{12}}(s,t) = \sum_{s<t_{(i)}\le t} \widehat{p_{11}}\left(s, t_{(i-1)}\right)\frac{d_{12i}}{n_{1i}}\,\widehat{p_{22}}\left(t_{(i)}, t\right). \qquad [3.8]$$

Because [3.6] and [3.7] are Kaplan-Meier estimators, we may use Greenwood's formula [1.10] to achieve a variance estimator for such transition probabilities, whereas,

$$\widehat{\mathrm{var}}\left(\widehat{p_{12}}(s,t)\right) = \sum_{s<t_{(i)}\le t} \widehat{p_{11}}\left(s,t_{(i-1)}\right)^2 \left[\widehat{p_{11}}\left(t_{(i)},t\right) - \widehat{p_{12}}\left(t_{(i)},t\right)\right]^2 \frac{n_{1i}-1}{n_{1i}^{\ 3}} d_{12i}$$
$$+ \sum_{s<t_{(i)}\le t} \left[\widehat{p_{11}}\left(s,t_{(i-1)}\right)\widehat{p_{12}}\left(t_{(i)},t\right)\right]^2 \frac{n_{1i}-1}{n_{1i}^{\ 3}} d_{13i}$$
$$+ \sum_{s<t_{(i)}\le t} \left[\widehat{p_{12}}\left(s,t_{(i-1)}\right)\widehat{p_{22}}\left(t_{(i)},t\right)\right]^2 \frac{n_{2i}-1}{n_{2i}^{\ 3}} d_{23i}$$

For general cases, explicit expressions for the transition probabilities like those mentioned above cannot be given. For such cases, consider d_{hji} the number of transitions from state *h* into state *j* at time $t_{(i)}$, and let n_{hi} be the number of subjects in state *h* just prior to time $t_{(i)}$. For each $i \in \{1,...,k\}$ consider the intensity matrix Q_i with entry (h,j) given by α_{hji}. Consider now $\widehat{Q}_i$, with entry (h,j) given by $\widehat{\alpha_{hji}} = \frac{d_{hji}}{n_{hi}}$ for $h \neq j$; and $\widehat{\alpha_{hhi}} = -\frac{\sum_{j\neq h} d_{hji}}{n_{hi}}$ for (h,h). We may then express the transition probability matrix P in terms of the intensity matrix. The Aalen-Johansen estimator takes the form

$$P(s,t) = \prod_{s<t_{(i)}\le t} \left(I + \widehat{Q}_i\right),$$

where *I* is the identity matrix.

An indispensable assumption here is that censoring is independent, so that censoring times do not carry information on the hazard of transition between states.

Further details can be found in Andersen et al. (1993).

3.4. Simulation study.

In situations where the estimation of the covariate effect is of interest, one question that can arise is whether the inclusion of information on the intermediate events through a multi-state model provides a better knowledge about how the covariate affects the hazard. In this section, we intend to illustrate how the covariate effect can be wrongly interpreted when using simple regression models such as the Cox time-dependent regression model.

For simplicity, we have chosen to generate a homogeneous Markov process. We then compare the homogeneous Markov model and Cox Markov model with the Cox time-dependent regression model.

We consider the illness-death model and the population vector as (T_{12}, T_{13}, T_{23}) where, T_{hj} is the potential sojourn time in state *h* prior to transition to state *j*.

An individual in state 1 is then exposed to two mutually exclusive events, for which only the first (small) of these events is observed. Therefore, patient histories can be decoupled in two groups, $1 \to 2 \to 3$ $(\text{if } T_{12} \le T_{13})$ or $1 \to 3$ $(\text{if } T_{13} < T_{12})$. We assume also that individual's times are at risk of being right-censored by some censoring variable *C*, denoting the potential censoring time. Thus, individuals times can be censored in state 1 $(\text{if } C < \min(T_{12}, T_{13}))$ or in state 2 $\left(\text{if } C < T_{12} + T_{23} \text{ and } T_{12} \le \min(T_{13}, C)\right)$.

Note that the illness-death model leads to the time-dependent Cox model, for which the time-dependent covariate is coded as 1, if a transition from state 2 into state 3 is observed and 0 otherwise.

For each individual, we generate the observations for one covariate, *Z*, from a uniform distribution with a minimum of 0 and a maximum of 10. To achieve a linear covariate effect on transition $1 \to 3$ and $2 \to 3$ (and no effect on transition $1 \to 2$), the sojourn times T_{12}, T_{13}, and T_{23} were independently generated from three exponential distributions, with rate parameters α_{12}, α_{13}, and α_{23} respectively, where:

$$\alpha_{12} = 0.05 \times \exp\left(\beta_{12} \times Z\right),$$

$$\alpha_{13} = 0.04 \times \exp\left(\beta_{13} \times Z\right)$$

and

$$\alpha_{23} = 0.05 \times \exp\left(\beta_{23} \times Z\right),$$

with $\beta_{12} = 0$, $\beta_{13} = 0.08$ and $\beta_{23} = -0.09$.

Censoring times were generated by a uniform distribution with minimum 0 and maximum 80 (yielding a percentage of about 27% of censored observations). Note that since $\left(T_{12}, T_{13}, T_{23}\right)$ are exponentially distributed and mutually independent, we are generating a homogeneous Markov process. Under this setting, when using a multi-state model, a significant covariate effect is expected only on transitions $1 \to 3$ and $2 \to 3$. By choosing a covariate effect on transition $2 \to 3$ opposite to the effect of the same covariate on transition $1 \to 3$, we intend to show that this may eliminate the covariate effect when using a Cox time-dependent regression model.

For this configuration, $K = 500$ data sets were generated with two different sample sizes of $n = 200$ and $n = 500$. In Table 3.1, we present the mean estimate $\left(ME\right)$ (along with the corresponding standard error), $ME = \overline{\widehat{\beta}} = \frac{1}{K}\sum_{i=1}^{K} \widehat{\beta}_i$, and the mean hazard ratio $\left(MHR\right)$ estimate: $MHR = \overline{HR} = \frac{1}{K}\sum_{i=1}^{K} \exp\left(\widehat{\beta}_i\right)$.

In this table, we also present the percentage of observed significant effects $(POSE)$.

As shown in Table 3.1, both multi-state models give rise to estimators which are in any sense better than those obtained from the Cox time-dependent regression model. While estimators for the Cox model can not be interpretable at all, both multi-state models yielded good estimates (compare the mean estimates in Table 3.1 with the real values β_{hj}). Note that for any multi-state model, and especially for the homogeneous Markov model, the mean estimate for the covariate effect on transitions $1 \to 3$ and $2 \to 3$ are very precise, whereas for transition $1 \to 2$ it is near zero (something that would be expected whenever the covariate effect does not exist). The simulation showed that increasing the sample size reduces the variance and increases the *POSE* for all models (but mainly for multi-state models, leading to a more precise covariate effect). For the sample size $n = 500$, when analyzing the homogeneous Markov model we observe that the estimated covariate effect is accurate and statistically significant (i.e., present) on 472 replicates for transition $1 \to 3$ and on 412 for transition $2 \to 3$. Furthermore, for the sample size $n = 1000$ the covariate effect is always present for both transitions (results not shown). For the same $K = 500$ replicates, the covariate effect is statistically significant in 71 replicates (14.2%) when using a Cox time-dependent regression model. As shown in Table 3.1, the obtained mean hazard rate estimate is close to one. Although the generated process is time-homogeneous (thus, favourable to the homogeneous Markov model) the Cox Markov model also achieved good results. Moreover, such a model is the one presenting better results for the *POSE* in all transitions leaving state 1.

Table 3.2 shows the estimates obtained for the homogeneous Markov model, Cox Markov model and Cox time-dependent regression model for a single sample with size

$n = 500$. Through this example, it is shown that the effect of important covariates may vanish when only a model for the hazard rate is analyzed. In contrast, both multi-state models reached a significant and accurate covariate effect on transitions $1 \to 3$ and $2 \to 3$.

Table 3.1. Mean estimate of the covariate effect for the homogeneous Markov model, Cox Markov model and for the Cox time-dependent regression model, according to sample size, over 500 replicates.

	Model	*ME*	*SE*	*MHR*	*POSE*
n = 200	Homogeneous Markov model				
	$1 \to 2$	–0.0015	0.0393	0.9987	4.2%
	$1 \to 3$	0.0798	0.0384	1.0832	58.6%
	$2 \to 3$	–0.0927	0.0550	0.9119	46.0%
	Cox Markov model				
	$1 \to 2$	0.0040	0.0418	1.0049	4.3%
	$1 \to 3$	0.0773	0.0352	1.0811	71,8%
	$2 \to 3$	–0.0950	0.0574	0.9108	35.9%
	Time-dependent Cox model	0.0176	0.0308	1.0183	8.1%
n = 500	Homogeneous Markov model				
	$1 \to 2$	–0.0002	0.0263	0.9997	6.6%
	$1 \to 3$	0.0788	0.0215	1.0820	94.4%
	$2 \to 3$	–0.0916	0.0326	0.9129	82.4%
	Cox Markov model				
	$1 \to 2$	–0.0008	0.0264	0.9995	2.4%
	$1 \to 3$	0.0803	0.0223	1.0839	97.8%
	$2 \to 3$	–0.0910	0.0352	0.9136	68.7%
	Time-dependent Cox model	0.0163	0.0184	1.0166	14.2%

ME = Mean Estimate of the covariate effect; *MHR* = Mean Hazard Ratio;
SE = Standard Error; *POSE* = Percentage of observed significant effects.

Table 3.2. Covariate effect in homogeneous Markov model, Cox Markov model and Cox time-dependent model for a single sample with size $n = 500$.

Model	$\hat{\beta}$	HR	95%CI for HR
Homogeneous Markov model			
$1 \to 2$	-0.005	0.995	0.944-1.048
$1 \to 3$	0.060	1.062	1.019-1.107
$2 \to 3$	-0.095	0.909	0.850-0.972
Cox Markov model			
$1 \to 2$	-0.010	0.990	0.940-1.040
$1 \to 3$	0.055	1.060	1.010-1.100
$2 \to 3$	-0.098	0.906	0.847-0.970
Time-dependent Cox model	0.011	1.010	0.976-1.050

HR = Hazard Ratio; CI=Confidence Interval.

3.5. Application to the Stanford Heart Transplantation study.

For illustration, we have applied reviewed methods of previous sections to data from the Stanford's Heart Transplant study described in detail in Chapter 1.

In the context of multi-state modelling, we may consider the covariate 'transplant' as an associated state of risk, and then use the illness-death model of Figure 3.1. In this model, for a given patient, 'own heart' corresponds to having his/her own heart, whereas 'new heart' represents having had a transplant. Typically, a patient enters the study in the 'own heart' state; after the transplant, he moves to the transplanted population, that is, to the 'new heart' state. With this multi-state formulation of the Stanford data, main goals of this study include: (a) to assess whether or not there exists a beneficial effect of heart transplant on survival, which will be carried out by comparing the transition intensities $\alpha_{13}(t)$ and $\alpha_{23}(t)$, and also (b) to explore the potential fixed covariate effects in each of the transitions.

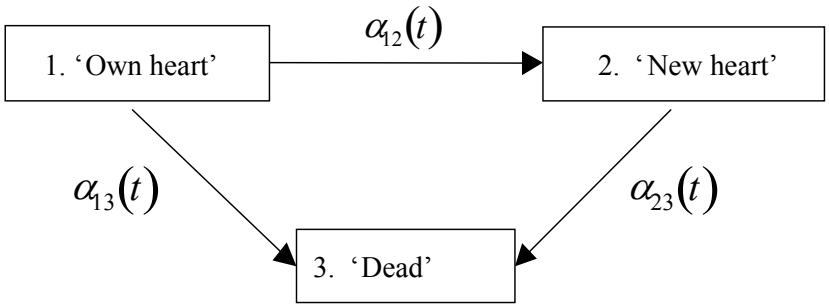


Figure 3.1. Illness-death model for Transplant Heart data.

Time-dependent Cox regression model

We have constructed various time-dependent Cox regression models, all of them including the effect of transplant, among other covariates. Numerical results obtained from fitting eleven different models are presented in Table 3.3. Model comparisons have been done using the Akaike's Information Criterion (AIC), which is defined as

$$AIC = -2 \times \text{log-likelihood} + 2p\,,$$

being p the number of estimated parameters. Smaller values for the AIC suggest preferable models.

For all the fitted models, the influence of age at acceptance on hazard is positive, while effects of year and surgery are both negative. When analyzing the fitted models showing a smaller AIC (i.e. models V, and IX-XI), we see that the effect of the transplant leads to a small reduction in risk, but without reaching statistical significance. For example, for model IX, age (Hazard Ratio, HR:1.027; 95% confidence interval, 95%CI: 1.001-1.055) and year of acceptance $\left(\text{HR:0.864;95\%CI:0.753-0.992}\right)$ are both important factors, while surgery has no significant effect $\left(\text{p-value} > 0.05\right)$. Figure 3.2(a) displays the smoothed hazard ratio for the covariate year, obtained by using penalized

splines together with the 95 percent pointwise confidence bands. This plot suggests that, during the first months a diminishing of the hazard ratio occurs, remaining nearly constant afterwards. However, when we exclude the patients accepted in the first year of the programme, the negative effect of year on hazard ($\widehat{\beta} = -0.0955$) is then negligible (p-value=0.27). As shown in Table 3.3, the effect of age is significant and, by itself, changes the direction of the transplant effect. Figure 3.2 (b) shows the smoothed hazard ratio for this covariate using penalized splines, along with the corresponding 95 percent pointwise confidence bands. According to this figure, it is apparent that the effect of age is non-linear and closer to being quadratic, as it was previously suggested by Crowley and Hu (1977).

Table 3.3. Time-dependent Cox regression models. Effect estimates and corresponding standard errors. Stanford Heart Transplantation data.

Model	Estimate (SE)	transplant	Age	year	surgery	Transplant by age	Transplant by year	AIC
I	$\hat{\beta}$ (SE)	0.127 (0.301)						598.07
	P – value	0.67						
II	$\hat{\beta}$ (SE)	-0.004 (0.312)	0.031 (0.015)					595.07
	P – value	0.99	0.03					
III	$\hat{\beta}$ (SE)	0.123 (0.303)		-0.191 (0.070)				592.55
	P – value	0.68		0.01				
IV	$\hat{\beta}$ (SE)	0.158 (0.297)			-0.749 (0.360)			594.88
	P – value	0.59			0.04			
V	$\hat{\beta}$ (SE)	-0.031 (0.318)	0.027 (0.014)	-0.179 (0.070)				590.58
	P – value	0.92	0.06	0.01				
VI	$\hat{\beta}$ (SE)	0.016 (0.309)	0.031 (0.014)		-0.773 (0.360)			591.52
	P – value	0.96	0.03		0.03			
VII	$\hat{\beta}$ (SE)	0.076 (0.321)	0.012 (0.018)			0.041 (0.028)		594.97
	P – value	0.81	0.51			0.15		
VIII	$\hat{\beta}$ (SE)	-0.282 (0.514)		-0.265 (0.105)			0.137 (0.141)	593.60
	P – value	0.58		0.01			0.33	
IX	$\hat{\beta}$ (SE)	-0.010 (0.314)	0.027 (0.014)	-0.146 (0.070)	-0.637 (0.367)			589.13
	P – value	0.97	0.05	0.04	0.08			
X	$\hat{\beta}$ (SE)	-0.606 (0.540)	0.029 (0.014)	-0.279 (0.106)			0.186 (0.143)	590.84
	P – value	0.26	0.04	0.01			0.19	
XI	$\hat{\beta}$ (SE)	-0.621 (0.531)	0.030 (0.014)	-0.253 (0.105)	-0.664 (0.368)		0.197 (0.140)	589.10
	P – value	0.24	0.03	0.02	0.07		0.16	

SE=Standard error; AIC=Akaike Information Criterion.

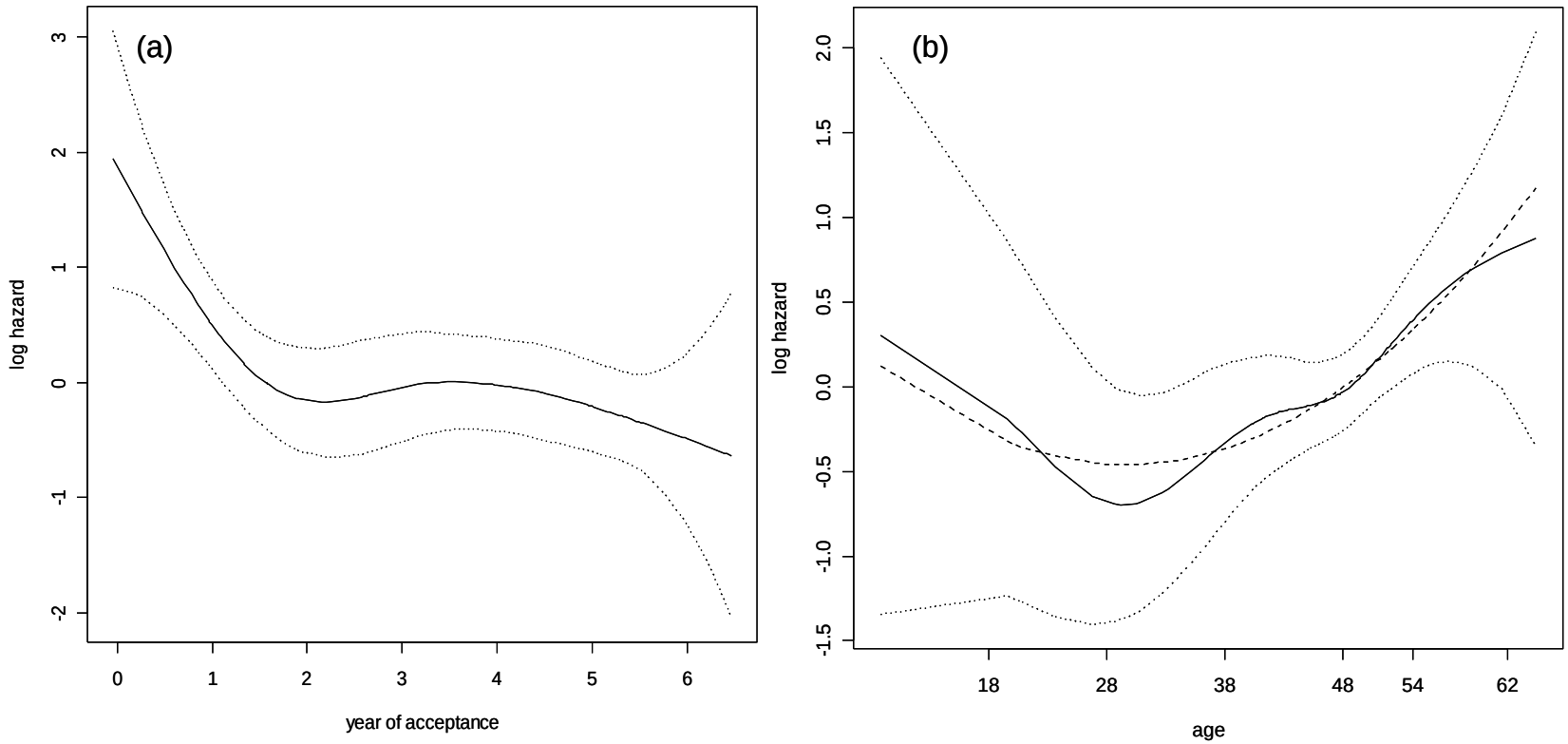


Figure 3.2: (a) Hazard ratio estimation with penalized splines for year of acceptance (with 95% pointwise confidence bands). (b) Hazard ratio estimation with penalized splines for age at acceptance (with 95% pointwise confidence bands), along with quadratic fit (dashed line). Stanford Heart Transplantation data.

Validity of Markov's assumption

The models studied in this chapter rely on the Markov assumption. Before the analyses of the Stanford data, we will check the validity of this assumption, verifying that the transition rate, $\alpha_{23}(t)$, is not affected by the time since transplant. We, therefore, consider $Z_1 = \text{"age"}$ and $Z_2 = \text{"time since transplant"}$ as the covariates, and then we fit the following two hazards: $\alpha_0(t)\exp(\beta_1 Z_1 + \beta_2 + \beta_3 Z_2)$ for patients with transplant, and $\alpha_0(t)\exp(\beta_1 Z_1)$ for patients without. The results obtained (see Table 3.4) show that effect the of time since the transplant is not significant (p-value=0.93), indicating that the transplant does not lead to an increase in mortality in the period immediately after surgery. This allows us to conclude that Markov's model is satisfactory for the Stanford data (with time since entry in study as the time scale). Further details about this issue are discussed in Chapter 6.

Table 3.4. Estimated effects on the mortality intensity in transplanted patients. Stanford Heart Transplantation data.

Parameter	Estimate	SE	P-value
β_1 (age)	0.031	0.015	<0.05
β_2 (indicator of transplant)	-0.015	0.336	0.96
β_3 (Time since transplant)	0.0001	0.0016	0.93

SE=Standard error.

Cox Markov Model

The Cox Markov model [3.5] allows us to observe how the covariate effects behave when transition intensities are modelled separately. Through this multi-state methodology we pretend to show that these methods can provide new biological insights while confirming some of the results obtained using the time-dependent Cox

model. Results obtained from fitting this model are given in Table 3.5. It can be seen that year of acceptance, which revealed a strong effect on survival in the Cox model, under the Cox Markov model only shows a significant effect on $\alpha_{13}(t)$ (HR:0.753; 95%CI: 0.606-0.936), although this covariate ceases to be a significant predictor when the patients admitted in the first year were excluded from the study. Age showed a linear effect on each transition, and is the best predictor for the mortality transition $\alpha_{23}(t)$ of transplanted patients $(\text{HR:1.050;95\%CI:1.008}-1.100)$, as well as for transition $\alpha_{12}(t)$, corresponding to receive a new heart $(\text{HR:1.032;95\%CI:1.004-1.060})$. Interestingly, age has no significant effect on the mortality intensity in patients without transplant $(\text{HR:1.020;95\%CI:0.984}-1.057)$, suggesting that the effect of age enters through the transplant incidence. As shown in Table 3.5, there is no effect of a previous surgery in any transition.

Table 3.5. Final Cox Markov model for all transitions. Stanford Heart Transplantation data.

Transitions	Covariate	$\hat{\beta}$	SE	HR	95% CI for HR	p-value
$1 \rightarrow 2$	Age	0.031	0.014	1.032	1.004 – 1.060	0.03
	Year	0.001	0.070	1.001	0.873 – 1.150	0.99
	Surgery (yes=1,no=0)	0.047	0.315	1.048	0.565 – 1.940	0.88
$1 \rightarrow 3$	Age	0.020	0.018	1.020	0.984 – 1.057	0.27
	Year	-0.283	0.111	0.753	0.606 – 0.936	0.01
	Surgery (yes=1,no=0)	-0.229	0.636	0.796	0.229 – 2.768	0.72
$2 \rightarrow 3$	Age	0.050	0.021	1.050	1.008 – 1.100	0.02
	Year	-0.023	0.097	0.977	0.808 – 1.180	0.81
	Surgery (yes=1,no=0)	-0.817	0.455	0.442	0.181 – 1.080	0.07

SE = Standard Error, HR = Hazard Ratio; CI=Confidence Interval.

Homogeneous Markov Model

In comparison with the preceding studied models, the homogeneous Markov model offers a detailed description of the survival process, making use of all the available information to estimate the transition probabilities and intensity rates. By applying this modelling approach, we refit the Stanford data including the potential effects of age, year and surgery in all transitions. Results obtained from the fitted model (see Table 3.6) are in good agreement with those obtained when using the Cox Markov model. Some exceptions are the surgery effect, which was not a significant factor in any of the previously studied models but now reveals a significant effect on survival for transplanted patients $\left(\text{HR:}0.306;95\%\text{CI:}0.128-0.730\right)$. Results indicate that age is the only covariate showing a significant linear effect in all transitions. This occurs even for the mortality intensity in patients without a transplant, something that did not occur when using the Cox Markov modelling approach. We have also observed that the acceptance time in the study (year) is a significant predictor, though only for mortality intensity in patients without transplant $\left(\text{HR:}0.739;95\%\text{CI:}0.595-0.919\right)$.

We have also observed that the fitted homogeneous Markov model leads to very similar effects of age in the three transitions (see Table 3.6). To assume that these effects are equal, we use the Wald test statistic, yielding a value of 1.163, revealing non significant differences between them (Marshall and Jones, 1995).

In view of the results obtained, we then consider a simplified homogeneous Markov model, including the (same) potential effect of age in all transitions, the effect of year only in the mortality transition of patients without transplant, and the effect of surgery in the mortality transition of transplanted patients. Notice that this new model prevents the over-fitting of data with redundant parameters. In the process, likelihood

ratio tests (LRT) were used to test if the regression parameters are statistically different from zero (Kay,1986) . This statistic has an approximately χ_1^2 distribution under $H_0 : \beta_{hj} = 0$. Estimated parameters for the simplified model (results not shown) did not differ substantially from those obtained by the initial model (shown in Table 3.6). Therefore, hypothesis testing can be carried out using either LRT or Wald tests although opinions regarding which is optimal lack consistency.

Table 3.6. Multi-state homogeneous Markov model. Estimated transition rates and hazard rates. Stanford Heart Transplantation data.

TR (SE)		
	$\widehat{\alpha_{12}}$	$0.0137\ (0.0017)$
	$\widehat{\alpha_{13}}$	$0.0054\ (0.0011)$
	$\widehat{\alpha_{23}}$	$0.0018\ (0.0003)$
HR (95%CI)		
Age		
	$1 \to 2$	$1.068\ (1.039-1.098)$
	$1 \to 3$	$1.056\ (1.020-1.093)$
	$2 \to 3$	$1.076\ (1.030-1.125)$
Year		
	$1 \to 2$	$0.975\ (0.852-1.116)$
	$1 \to 3$	$0.739\ (0.595-0.919)$
	$2 \to 3$	$1.109\ (0.928-1.325)$
Surgery		
	$1 \to 2$	$1.368\ (0.737-2.539)$
	$1 \to 3$	$0.959\ (0.277-3.315)$
	$2 \to 3$	$0.306\ (0.128-0.730)$

TR=Transition rate; SE=Standard error; HR=Hazard ratio; CI=Confidence Interval.

Further, we have used Wald's test to verify whether or not a relation between transplant and survival exists (Kay, 1986). Formally, the hypothesis of no relation is given by $H_0 : \alpha_{13} = \alpha_{23}$, and then Wald's test reduces to $W = \dfrac{\left(\widehat{\alpha_{13}} - \widehat{\alpha_{23}}\right)^2}{v_{11}}$, being

$v_{11} = \text{var}\left(\widehat{\alpha_{13}} - \widehat{\alpha_{23}}\right)$. With our data, under the null hypothesis the W statistic (which follows a χ_1^2) yields a value of 18.5, suggesting that the transplant is significantly associated to a diminishing in mortality risk. In Figure 3.3 we also compare fitted survival probabilities, $\widehat{p_{13}}$ and $\widehat{p_{23}}$, confirming the beneficial effect of the transplant.

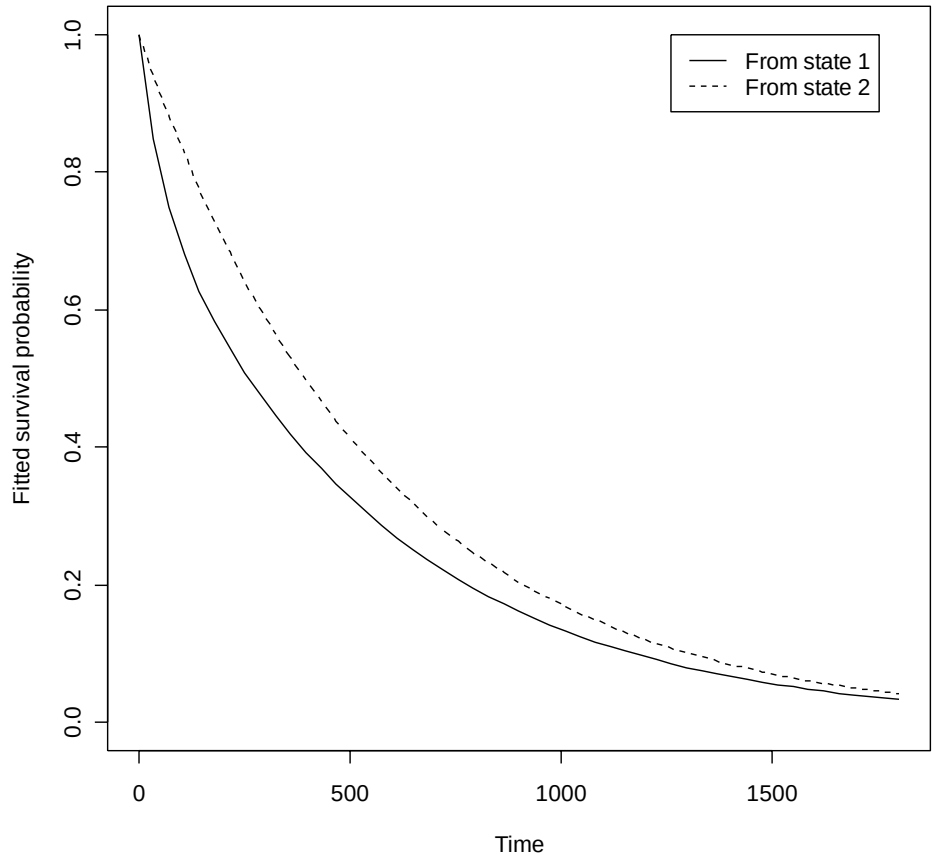


Figure 3.3. Fitted survival probability for the mortality intensity from a multi-state homogeneous continuous-time Markov model. Stanford Heart Transplantation data.

Note however that likelihood ratio tests (fitting unrestricted and restricted models) can also be used for constructing a test of H_0 against the general alternative (under the null hypothesis the test statistic has an approximately χ_1^2 distribution) (Gentleman et al., 1994).

The goodness-of-fit of a multi-state model can be assessed by comparing the observed and predicted number of patients undergoing each transition. Table 3.7 (page 72) reports the observed percentages of patients in states "own heart", "new heart" and "dead", together with the corresponding expected percentages obtained from the fitted homogeneous Markov model. For comparison purposes, we also included the expected percentages obtained from all fitted multi-state models. In this table, we can observe that, for lower survival times, the mortality is underestimated from the fitted homogeneous Markov model. In many cases, these discrepancies can be explained by the failure of the Markov assumption. Another possibility is that the transition rates vary with time, so that the model is non-homogeneous. This is the case for the Stanford data; it is seen that most of the transitions from state 'own heart' to state 'new heart' (approximately 73% from the total) occur up to 51 days of survival. Figure 3.4 compares the resulting transition probabilities $p_{12}(0,t|z)$ from the fitted Cox Markov model and homogeneous Markov model, respectively. This figure seems to point out that the Cox Markov model explains in a better way how transition rates vary with time, indicating a rapid increase of the transition probability up to 51 days, then decreasing quickly afterwards. Taken as a whole, these results suggest that a homogeneous model may be inappropriate. To assess the assumption of time homogeneity, Kay (1986) suggests the use of a piecewise model. Again, likelihood ratio tests can be used to compare the piecewise model with the homogeneous model. For the Stanford Heart Transplantation data the test statistic (which follows a χ^2_4) suggests the use of a non-homogeneous model. Another method for verifying departures from time-homogeneity is via a local score test (Kalbfleisch and Lawless, 1985).

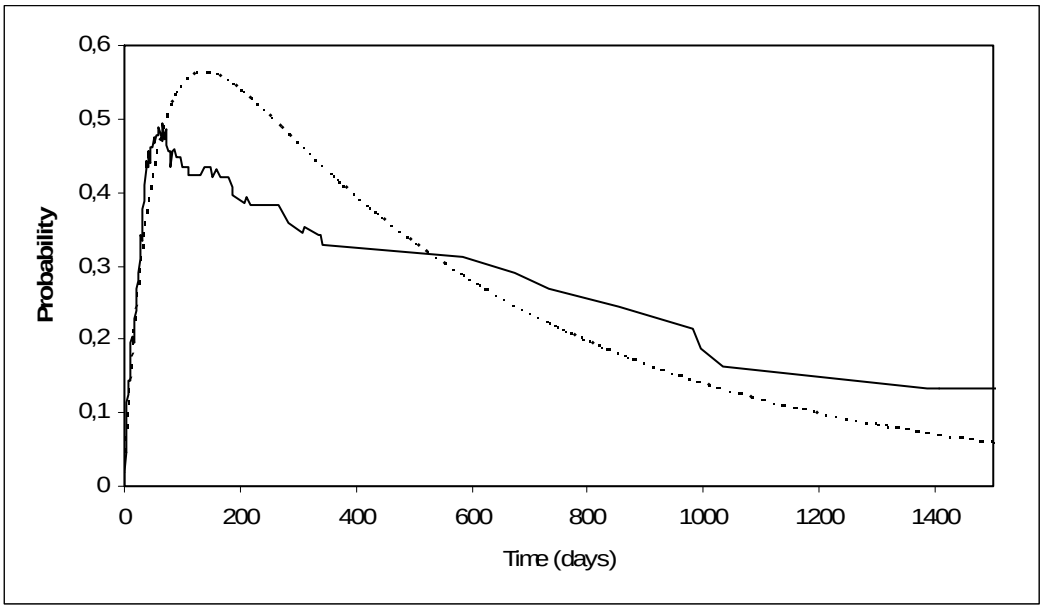


Figure 3.4. Estimated transition probabilities from state 1 ('own heart') to state 2 ('new heart'), obtained from the Cox Markov model (solid line), and the homogeneous Markov model (dashed line). Stanford Heart Transplantation data.

Non-homogeneous model

In this section, we construct a piecewise constant intensities model with one cut-off point, specified from the Stanford data covariates that showed a significant effect when fitting the homogeneous model. After examining the likelihood for several cut-off points θ, a value of $\theta = 90$ days was selected, and two intervals (time ≤ 90 days , time >90 days) were then considered. Focusing mainly on the short-term survival period of 90 days, results in Table 3.7 indicate that, for the non-homogeneous model, the agreement between predicted and observed percentages of patients in each transition is globally satisfactory, being clearly better than that which was obtained previously by using an homogeneous model.

Estimated mortality rates and covariate effects produced from fitting the non-homogeneous model are presented in Table 3.8. It is seen that in both intervals the

resulting estimates for the mortality intensity were lower in transplanted patients, though only in the second interval (time >90 days) a significant difference was found. When examining the fixed covariate effects, we see that, for time ≤ 90 days, age at acceptance is a significant predictor in all transitions $\left(\text{HR:1.032;95\%CI:1.011}-1.053\right)$, while the effect of year is only significant on the mortality intensity in patients without transplant $\left(\text{HR:0.716;95\%CI:0.571}-0.899\right)$ and the effect of surgery only on the mortality intensity in transplanted patients $\left(\text{HR:0.131;95\%CI:0.018}-0.968\right)$. For the second interval (time >90 days), however, the only significant covariate was age at acceptance $\left(\text{HR:1.061;95\%CI:1.008}-1.116\right)$.

Finally, turning to the comparison of the time-dependent Cox model with the four multi-state models described here, Figure 3.5(a) shows the resulting survival functions, $\hat{S}(t)$, obtained from fitting all these models to Stanford data. The range of time has been restricted to 90 days to emphasize the differences between the four functions. In view of results of Table 3.7, it is not surprising that the uppermost survival curve corresponds to the homogeneous Markov model. All the other proposed models produce similar survival curves. Figure 3.5(b) shows the resulting survival function for the entire range of time.

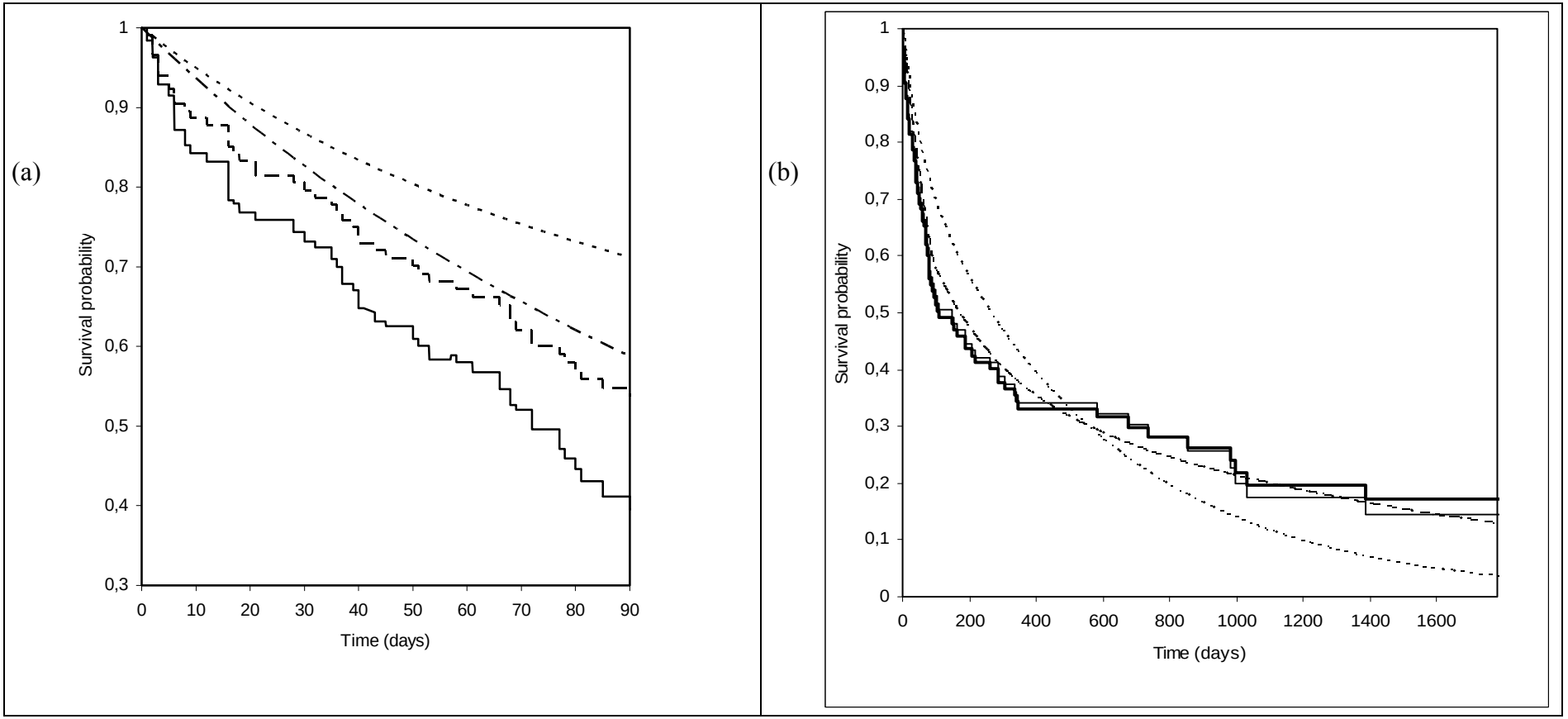


Figure 3.5. (a) Comparison of probability survival curves for time up to 90 days. Homogeneous Markov model (····), Non-homogeneous model (– · –), Time dependent Cox model (– – –), Cox Markov model (—). (b) Survival probabilities curves for the Homogeneous Markov model (····), piecewise Markov model (– – –), Cox Markov model (**—**), and Time-dependent Cox model (—). Stanford Heart Transplantation data.

Table 3.7. Comparison of observed and predicted percentages of patients in each of the three states for Cox Markov Model (CMM), Homogeneous Markov Model (HMM) and Non-Homogeneous Model (NHM). Stanford Heart Transplantation data.

Survival Time	State 1 "Own heart"				State 2 "New heart"				State 3 "Dead"			
(days)	Observed	CMM	HMM	NHM	Observed	CMM	HMM	NHM	Observed	CMM	HMM	NHM
0	100	100	100	100	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
20	57.8	60.1	68.3	59.3	23.5	23.8	22.3	28.5	18.6	16.1	9.4	12.2
50	25.0	24.9	38.6	27.1	43.0	46.2	41.9	46.3	32.0	28.9	19.6	26.6
90	11.0	10.5	18.0	9.6	41.0	44.7	53.2	49.2	48.0	44.8	28.8	41.2
150	7.1	5.9	5.7	5.2	37.8	43.4	56.2	40.4	55.1	50.7	38.1	54.4
250	5.2	3.9	0.9	3.2	33.0	38.3	50.4	38.9	61.9	57.8	48.8	57.9
400	2.2	1.2	0.1	1.5	25.8	32.9	39.3	36.3	72.0	65.9	60.6	62.2
600	1.2	1.2	0.0	0.6	20.7	31.1	27.8	32.5	78.2	67.7	72.2	66.9
900	1.2	1.2	0.0	0.1	14.3	24.4	16.5	27.1	84.5	74.4	83.5	72.8
1200	1.2	1.2	0.0	0.0	7.4	16.2	9.8	22.5	91.4	82.6	90.2	77.5

Table 3.8. Non-homogeneous model. Estimated transition rates and hazard rates on transition intensities. Stanford Heart Transplantation data.

	time ≤ 90 days	**time > 90 days**
Number of parameters	6	4
TR (SE)		
$\widehat{\alpha}_{12}$	$0.0194\ (0.0024)$	$0.0028\ (0.0013)$
$\widehat{\alpha}_{13}$	$0.0068\ (0.0015)$	$0.0022\ (0.0011)$
$\widehat{\alpha}_{23}$	$0.0052\ (0.0012)$	$0.0006\ (0.0002)$
HR (95%CI)		
Age	$1.032\ (1.011-1.053)$	$1.061\ (1.008-1.116)$
Year $1\rightarrow 3$	$0.716\ (0.571-0.899)$	
Surgery $2\rightarrow 3$	$0.131\ (0.018-0.968)$	

TR=Transition rate; SE=Standard error; HR=Hazard ratio; CI=Confidence Interval.

3.6. Application to the Psoriatic Arthritis data.

The purpose of this section is to discuss the application of a multi-state Markov model to a database on Psoriatic Arthritis (described in detail in Chapter 1). Here, the application of a Markov process allows a simplification in the modelling since the transition intensities are assumed to be independent of the history of the process (something that seem reasonable under the clinician perspective).

With the purpose of investigating the evolution of the disease, several stages were considered from birth up to current state: Birth, Psoriatic, Arthritis, PsA and states for the actual stage of the disease, Remission and Active. It is assumed that transitions from one state to the next occur irreversibly. The states and the allowed transitions

between states are shown in Figure 3.6. Table 3.9 summarizes the observed transitions between these states. Since our interest focuses on the age of developing Psoriatic Arthritis (PsA), then age is the appropriate choice for time scale.

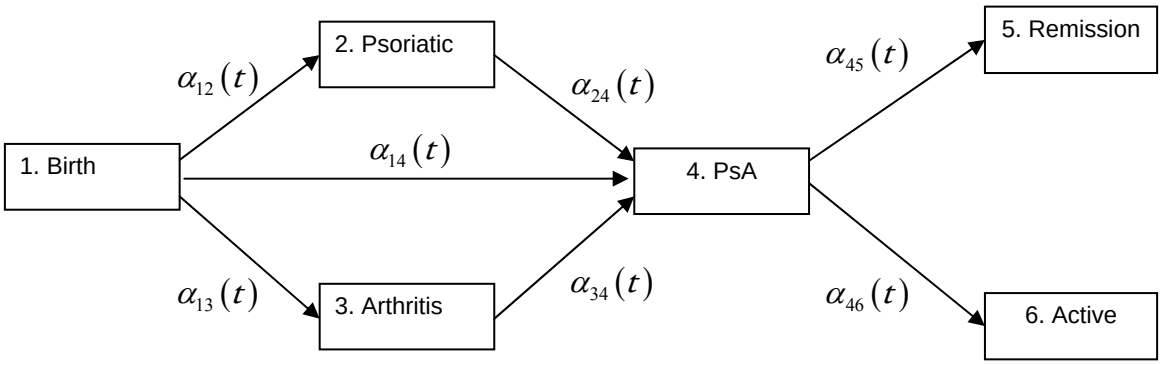


Figure 3.6. multi-state model for Psoriatic Arthritis data.

Table 3.9. Observed transitions in the Psoriatic Arthritis data.

		To				
		2	3	4	5	6
From	1	263	35	32	0	0
	2	0	0	263	0	0
	3	0	0	35	0	0
	4	0	0	2	176	154

In this case several issues must be considered when the interest is to analyse the transition intensities. For example, because this dataset was obtained retrospectively, we have the problem of left-truncation. The analysis of these issues is out of the scope of this thesis.

3.7. Discussion.

Multi-state models have some advantages over the time-dependent Cox regression model. Among others, these models (a) allow estimates for the number of patients in various states; (b) can reveal the effect of each covariate on different transitions; and (c) express the time-dependent covariates in a simpler way. Although multi-state models may be preferable to the Cox regression model, there are some limitations to the use of such models: (a) multi-state methodology requires some assumptions concerning a Markov or semi-Markov structure of the data; (b) most of the existent software assume the process is Markov and time-homogeneous, which can be very restrictive; (c) some of the multi-state models may require large sample sizes in order that accurate estimates may be achieved; (d) when using covariates, the number of parameters increases proportionally to the number of covariates. To overcome some of these difficulties, model assessment techniques can be used.

When analyzing Stanford Heart data using the multi-state methodology, the Markov assumption was satisfactory. Goodness-of-fit was tested for all multi-state models, showing that the Cox Markov model and the piecewise model are more suitable than the homogeneous model. When applying a homogeneous Markov model, we verified that this model underestimates the ‘short-term’ mortality. The reason why this model turned out to fit the Stanford heart data poorly is due to the fact that the survival process is not time-homogeneous.

While the time-dependent Cox regression model suggested a negligible effect of the transplant, the multi-state homogeneous Markov model indicated that the transplant is significantly associated to a diminishing in mortality risk. The application of the

piecewise model only confirmed such an association when the ‘long-term’ survival is analyzed.

Through a simulation study we have shown that the covariate effect on hazard can be deeply influenced by the effect of the same covariate through only one transition. In fact, when analyzing Stanford data we showed that the acceptance time covariate (year), considered the most important predictor in the Cox regression model, only shows a statistically significant effect in the mortality transition for patients without transplant. Some of the multi-state models used here have shown a significant negative influence of previous surgery on hazard in mortality transition for transplanted patients; that is, having a previous surgery enlarges the survival of transplanted patients. Age at acceptance, on the other hand, was revealed to be a significant predictor of survival in any of the studied models, and its positive effect on the hazard indicates that younger patients have a better survival.

In conclusion, the multi-state modelling offers a flexible tool for the study of covariate effects on the various transition rates. These models may bring out important biological insights which may be ignored when using Cox regression models alone. In practice, multi-state models can be used to thoroughly examine and confirm conclusions obtained by applying simpler survival models. Therefore, we should not see the multi-state models as merely an alternative to the time-dependent Cox model but rather as supplements that offer additional information.

Chapter 4

Multi-state non-Markov models

4.1. Introduction.

Traditionally, statistical methods for analyzing multi-state models depend on the Markov assumption. Under the Markov assumption, the transition intensities depend on the current time and the state currently occupied; they do not depend on the patient history (length of stay in the current state; patient characteristics measured before, etc.). By ignoring the disease history behaviour, these models may carry severe limitations which can make the model inappropriate. It is a fact that the future health of recently diseased individuals may be different from those who have been diseased for a long time. One alternative approach is to use a semi-Markov assumption in which future of the process does not depend on the current time but rather on the duration in the current state. Semi-Markov models (Andersen et al., 2000) are also called "clock reset" models, because each time the patient enters a new state time is reset to 0. The aim of the present chapter is to review Cox semi-Markov models and propose a new non-Markov approach, for which the transition intensities are allowed to depend not only on the current time, but also on the time of transition to the current state. For example, when considering the illness-death model (see Figure 2.4), we obtain a non-Markovian model by allowing the mortality intensity after the occurrence of the disease $(\text{transition } 2 \to 3)$ to depend on the current time as well as the time of transition to the diseased state.

4.2. Semi-Markov models.

Transition intensities can depend on the current state, states previously visited, the entry time in each state, the time since the entry into the current state, covariates, etc. As mentioned before, these models can rapidly become very complicated, and therefore simpler models are necessary. If the process is not Markov, then semi-Markov models can be a wise choice.

Semi-Markov models make the assumption that the history of the process depends only on the state currently occupied, the elapsed time since the entry into the current state and the covariates.

4.2.1 Cox semi-Markov models.

Assuming again the illness-death model, the transition intensities, $\alpha_{1j}\left(t|Z\right)$, $j=2,3$, are modelled as in [3.5]. The difference between Markov and semi-Markov models are expressed on transition, $2\to 3$. According to semi-Markov models, future evolution not only depends on the current state, but also on t_{12} $\left(t_{12}\le t\right)$, the time that the individual remains in that same state. Therefore, these transitions intensities will be expressed as

$$\alpha_{23i}\left(t-t_{12}\,|Z_i\right)=\alpha_{230}\left(t-t_{12}\right)\exp\left(\beta_{23}^{\mathrm{T}}Z_i\right)$$

The transition probabilities depend on the history of the process through the time t_{12}, $p_{hj}\left(s,t|Z,t_{12}\right)$, and can be estimated by,

$$\widehat{p_{22}}\left(s,t\,|\,Z,t_{12}\right)=\prod_{s<u\le t}\left(1-d\widehat{A_{23}}\left(u-t_{12}\,|Z\right)\right),$$

$$\widehat{p_{12}}\left(0,t\,|\,Z\right)=\widehat{p_{11}}\left(0,t-\,|\,Z\right)d\widehat{A_{12}}\left(t|Z\right)+\widehat{p_{12}}\left(0,t-\,|\,Z\right)\left(1-d\widehat{A_{23}}\left(t-t_{12}\,|Z\right)\right),$$

and

$$\widehat{p_{23}}\left(s,t\,|\,Z\right)=1-\widehat{p_{22}}\left(s,t\,|\,Z,t_{12}\right).$$

The rest of the transition probabilities are estimated as in the Cox Markov model of section 3.3.2.

The survival probability $S\left(t|Z\right)$ can now be estimated by

$$\hat{S}\left(t|Z\right)=\widehat{p_{11}}\left(0,t|Z\right)+\widehat{p_{12}}\left(0,t|Z\right)$$

with changes on its values at observed failure times.

4.3. Non-Markov models.

There are few references to non-Markov multi-state models in the literature, some exceptions are the works of Strauss and Shavelle (1998), Aalen, Borgan and Fekjaer, (2001), Datta and Satten (2001), and Glidden (2002). Strauss and Shavelle (1998) developed an extension of the Kaplan-Meier estimator for the estimation of transition probabilities, avoiding the Markov assumption. The proposed estimator is constructed by partitioning the survival probability in proportion to the number of live and uncensored patients in each state. Aalen et al. (2001) and Datta and Satten (2001)

studied the performance of the Aalen-Johansen estimator of stage occupancy probabilities when the process is not Markovian. These authors established the consistency of Aalen-Johansen estimators of the prevalence functions in a non-Markov process under independent censoring. Later, Glidden (2002) developed robust confidence bands for those event curves.

Our research on non-Markov models has two main goals. The first goal is to develop an approach based on less restrictive assumptions than those based on the Markov property by completely removing the Markov assumption. The second goal is to compare estimators developed here for the transition probabilities with Aalen-Johansen estimates (derived under the Markov assumption). For this purpose, we initially consider the illness-death model. Later, we consider the extension of this methodology for more complex multi-state models.

4.3.1 The illness-death model.

Assume that we have an illness-death model. Let $\{X(t), t \geq 0, X(0) = 1\}$ denote a non-homogeneous stochastic process. We assume that a finite number of independent histories from the process are observed. We then represent the stochastic behaviour of the process by a random vector (T_{12}, T_{13}, T_{23}) where T_{hj} is the potential sojourn time in state *h* prior to transition to state *j*.

An individual in the healthy state is exposed to two mutually exclusive events, “disease” and “death”, which do not occur simultaneously, and so only the first (small) of these events is observed. We therefore divide patient’s history (“course”) into two

groups according to whether the disease occurred $(1 \to 2 \to 3)$ or not $(1 \to 3)$. The history belongs to the second group if $T_{13} < T_{12}$ and then a transition from state 1 into state 3 is observed. Otherwise $(T_{12} \le T_{13})$ a transition from state 1 to state 2 occurs at time T_{12}, and later at time $T_{12} + T_{23}$ a transition from state 2 into state 3 takes place.

Of course, several issues influence the observation of the variables T_{hj}. Right-censoring may appear due to time limitation in the following-up, lost to follow-up cases, etc. On the other hand, whenever $T_{13} < T_{12}$, one gets a right-censored value of T_{12} $(\text{with censoring time } T_{13})$, and no information on T_{23} is available. In the same way, whenever $T_{12} \le T_{13}$, one gets a right-censored value of T_{13} $(\text{with censoring time } T_{12})$. Throughout this chapter, we assume that complete information for each individual may not be observed due to right-censoring. Therefore, individual's times are at risk of being right-censored by some censoring variable C (denoting the potential censoring time). In this way, individual's times can be censored before the illness $(\text{i.e., } C < \min(T_{12}, T_{13}))$ or after the illness $(\text{i.e., } C < T_{12} + T_{23} \text{ and } T_{12} \le \min(T_{13}, C))$. We will assume that C is independent of (T_{12}, T_{13}, T_{23}).

4.3.1.1. Nonparametric estimation of transition probabilities.

In this section, our main target is the estimation of the transition probability, $p_{hj}(s,t)=\mathbb{P}\left(X(t)=j\middle|X(s)=h\right)$, $s<t$, which gives the probability of being in state *j* at time *t*, conditionally on being in state *h* at time *s*. In multi-state modelling these are one of the most important targets to be estimated. For the illness-death model, it is enough to consider the estimation of the transition probabilities $p_{11}(s,t)$, $p_{12}(s,t)$ and $p_{22}(s,t)$. All the others can be obtained from these since $p_{13}(s,t)=1-p_{11}(s,t)-p_{12}(s,t)$ and $p_{23}(s,t)=1-p_{22}(s,t)$.

Therefore we have,

$$p_{11}(s,t)=\mathbb{P}\left(T_{12}>t,T_{13}>t\middle|T_{12}>s,T_{13}>s\right), \qquad [4.1]$$

$$p_{12}(s,t)=\mathbb{P}\left(T_{12}\le t,T_{12}\le T_{13},T_{12}+T_{23}>t\middle|T_{12}>s,T_{13}>s\right), \qquad [4.2]$$

$$p_{22}(s,t)=\mathbb{P}\left(T_{12}\le t,T_{12}\le T_{13},T_{12}+T_{23}>t\middle|T_{12}\le s,T_{12}\le T_{13},T_{12}+T_{23}>s\right). \qquad [4.3]$$

These quantities [4.1]-[4.3] are determined by the joint distribution of (T_{12},T_{13},T_{23}). Specifically, knowledge of the distribution of $Z=\min(T_{12},T_{13})$ is enough for the recovery of $p_{11}(s,t)$:

$$p_{11}(s,t)=\frac{\mathbb{P}(Z>t)}{\mathbb{P}(Z>s)},$$

while expectations of type $S(\phi)=\mathbb{E}\left[\phi(T_{12},T_{12}+T_{23})\mathbb{I}(T_{12}\le T_{13})\right]$ arise when handling $p_{12}(s,t)$ $\left(\phi(u,v)=\phi_{s,t}(u,v)=\mathbb{I}(s<u\le t,v>t)\right)$,

$$p_{12}(s,t)=\frac{S(\phi_{s,t})}{\mathbb{P}(Z>s)},$$

and $p_{22}(s,t)$ $\left(\phi(u,v)=\tilde{\phi}_{s,t}(u,v)=\mathbb{I}(u\le s, v>t)\right)$,

$$p_{22}(s,t)=\frac{S\left(\tilde{\phi}_{s,t}\right)}{S\left(\tilde{\phi}_{s,s}\right)}.$$

The relevance of the proposed estimation methods comes from the fact that they do not rely on the Markov assumption, typically used in multi-state models. Under the Markov assumption, the conditional probability of moving from one state to another only depends on the state currently occupied. These probabilities are not influenced by the states previously visited and the times of transition among them. In the case of the illness-death model, this assumption implies that the survival prognosis of an individual being in state 2 does not depend on the entry time, T_{12}, a fact that may be unrealistic in some applications. From a mathematical perspective, the Markov assumption restricts the distribution of (T_{12},T_{13},T_{23}) to those distributions satisfying, for any $s<t$, the equation

$$\begin{aligned}&\mathbb{P}\left[T_{12}+T_{23}>t \,\middle|\, T_{12}+T_{23}>s, T_{12}\le s, T_{12}\le T_{13}\right]=\\&\qquad=\mathbb{P}\left[T_{12}+T_{23}>t \,\middle|\, T_{12}+T_{23}>s, T_{12}=s_0, T_{12}\le T_{13}\right]\quad \forall 0<s_0\le s\end{aligned} \qquad [4.4]$$

that is, the conditional independence between $T_{12}+T_{23}$ and T_{12} given that $T_{12}+T_{23}>s$, $T_{12}\le T_{13}$, $T_{12}\le s$.

Proposed non-Markov estimators allow the transition probabilities to depend on the current time and on the time of entry into state 2, T_{12}.

Suppose a sample of *n* individuals under study followed from an entry time to death or censoring. We denote the sample information as

$$\left(U_i,\delta_i,\delta_i.V_i,\delta_i\rho_i,(1-\delta_i)\eta_i\right),\ 1\le i\le n$$

which are assumed to be independent and identically distributed copies of $\left(U,\delta,\delta.V,\delta\rho,(1-\delta)\eta\right)$, where

$U=\min\left(T_{12},T_{13},C\right)$ is the observed sojourn time in state 1;

$\delta=\mathbb{I}\left(T_{12}\le\min\left(T_{13},C\right)\right)$ is an indicator of whether a transition $1\to 2$ occurs;

$V=\min\left(T_{23},C-T_{12}\right)$ is the observed sojourn time in state 2;

$\rho=\mathbb{I}\left(T_{23}\le C-T_{12}\right)$, so that $\delta\rho$ is an indicator of whether a transition $2\to 3$ occurs;

$\eta=\mathbb{I}\left(T_{13}\le C\right)$, so that $(1-\delta)\eta$ is an indicator of whether a transition $1\to 3$ occurs.

Note that, under $\delta=0$, no information on $\left(V,\rho\right)$ is available; while, if $\delta=1$, then the observation of η is not possible.

Therefore, the available data for some individual will be, $\left(T_{13},0,0,0,1\right)$ if in the process a direct transition from state 1 into state 3 occurs; $\left(T_{12},1,T_{23},1,0\right)$ if a transition from state 1 into state 2 occurs and afterwards a transition to state 3; $\left(T_{12},1,C-T_{12},0,0\right)$ if the individuals transfers from state 1 to state 2 at time T_{12}, and afterwards have a censored sojourn time in state 2; and finally $\left(C,0,0,0,0\right)$ if the individual does not leave state 1, providing a censored sojourn time in state 1.

For the estimation of the transition probability [4.1], we need to make inference on Z, whose distribution function we denote by $H(t)=\mathbb{P}\left(Z\le t\right)$. In such cases, since we observe $U=\min\left(Z,C\right)$ and $\gamma=\delta+\left(1-\delta\right)\eta=\mathbb{I}\left(Z\le C\right)$, we consider the Kaplan-Meier product-limit estimator (Kaplan and Meier, 1958) based on the pairs $\left(U_i,\gamma_i\right)$.

Let $U_{(1)} \le U_{(2)} \le ... \le U_{(n)}$ be the ordered *U*-values, and let $\gamma_{[i]}$ be the concomitant of the *i*th order statistic, $U_{(i)}$. The Kaplan-Meier of $H(t)$ is defined as

$$\widehat{H}(t) = \sum_{i=1}^{n} W_{in}^{H} \mathbb{I}\left(U_{(i)} \le t\right), \qquad [4.5]$$

where $W_{in}^{H} = \frac{\gamma_{[i]}}{n-i+1} \prod_{j=1}^{i-1} \left(1 - \frac{\gamma_{[j]}}{n-j+1}\right)$ is the Kaplan-Meier weight attached to $U_{(i)}$.

Note that $W_{in}^{H} = 0$ for any censored observation $U_{(i)}$.

Since *C* and *Z* are assumed to be independent, then $\widehat{H}$ consistently estimates the distribution of *Z*.

Thus, the transition probability [4.1] can be estimated by,

$$\widehat{p_{11}}(s,t) = \frac{1-\widehat{H}(t)}{1-\widehat{H}(s)}. \qquad [4.6]$$

The above estimator [4.6] is equivalent to the Aalen-Johansen estimator [3.6].

Now, in order to introduce an estimator for the expectations of type $S(\phi)$, we need the following two lemmas.

Lemma 1. Let $T = \mathbb{I}(T_{12} \le T_{13})(T_{12} + T_{23}) + \mathbb{I}(T_{12} > T_{13})(T_{13})$ be the survival time of the process. Then: (i) $\mathbb{I}(T \le C) = (1-\delta)\eta + \delta\rho$; and (ii) $U + \delta V = \min(T, C)$.

Proof.

Note first that

$$(1-\delta)\eta + \delta\rho = \mathbb{I}(T_{12} > T_{13})\mathbb{I}(T_{13} \le C) + \mathbb{I}(T_{12} \le T_{13})\mathbb{I}(T_{12} + T_{23} \le C).$$

Now write

$$\mathbb{I}(T \le C) = \mathbb{I}(T_{12} > T_{13})\mathbb{I}(T \le C) + \mathbb{I}(T_{12} \le T_{13})\mathbb{I}(T \le C).$$

Use the definition of T to show (i). For (ii) write

$$U+\delta V=\mathbb{I}\left(T_{12}>T_{13}\right)\left(U+\delta V\right)+\mathbb{I}\left(T_{12}\le T_{13}\right)\left(U+\delta V\right).$$

When $T_{12}\le T_{13}$ we have $T=T_{12}+T_{23}$ and thus,

$$\begin{aligned}U+\delta V&=\min\left(T_{12},C\right)+\mathbb{I}\left(T_{12}\le C\right)\min\left(T_{23},C-T_{12}\right)\\&=\min\left(T_{12},C\right)+\mathbb{I}\left(T_{12}\le C\right)\left(\min\left(T,C\right)-T_{12}\right)\end{aligned}$$

By considering separately cases $T_{12}\le C$ and $T_{12}>C$, it is easily seen that the right-hand side equals $\min\left(T,C\right)$. On the other hand, on $T_{12}>T_{13}$ we have $T=T_{13}$, $\delta=0$, and $U+\delta V=\min\left(T,C\right)$ immediately follows. This shows $\left(ii\right)$. □

Lemma 2. Let *G* denote the distribution function of *C*. For each function ϕ we have

$$S\left(\phi\right)=\mathbb{E}\left[\frac{\phi\left(U,U+\delta V\right)\delta\rho}{1-G\left(\left(U+\delta V\right)^{-}\right)}\right]$$

where $1-G\left(x^{-}\right)=\mathbb{P}\left(C\ge x\right)$ is the left-continuous version of $1-G\left(x\right)$.

Proof.

Write

$$S\left(\phi\right)=\mathbb{E}\left\{\phi\left(T_{12},T_{12}+T_{23}\right)\mathbb{I}\left(T_{12}\le T_{13}\right)\frac{\mathbb{I}\left(T_{12}+T_{23}\le C\right)}{\mathbb{E}\left[\mathbb{I}\left(T_{12}+T_{23}\le C\right)\middle|T_{12},T_{13},T_{23}\right]}\right\}.$$

Now, since *C* is independent of the process,

$$\mathbb{E}\left[\mathbb{I}\left(T_{12}+T_{23}\le C\right)\middle|T_{12},T_{13},T_{23}\right]=1-G\left(\left(T_{12}+T_{23}\right)^{-}\right).$$

On the other hand,

$$\mathbb{I}\left(T_{12}\le T_{13}\right)\mathbb{I}\left(T_{12}+T_{23}\le C\right)=\delta\rho.$$

Since, under $\delta\rho=1$, we have $U=T_{12}$ and $\delta V=T_{23}$, we get

$$S\left(\phi\right)=\mathbb{E}\left[\frac{\phi\left(T_{12},T_{12}+T_{23}\right)\delta\rho}{1-G\left(\left(T_{12}+T_{23}\right)^{-}\right)}\right]=\mathbb{E}\left[\frac{\phi\left(U,U+\delta V\right)\delta\rho}{1-G\left(\left(U+\delta V\right)^{-}\right)}\right].\ \square$$

Let K denote the distribution function of the survival time of the process, T. Since T and C are independent, then, Lemma 1 ensures that the Kaplan-Meier product-limit estimator based on $\left(U_i+\delta_i V_i,\nu_i\right)$, where $\nu_i=\left(1-\delta_i\right)\eta_i+\delta_i\rho_i$, is a consistent estimator for K. This estimator, which we denote by $\widehat{K}$, is important since, for example, it provides an estimator for the occupation probability in state 3, $K\left(t\right)=\mathbb{P}\left(T\le t\right)=p_{13}\left(0,t\right)$.

Furthermore, Lemma 1 guarantees that the Kaplan-Meier product-limit estimator based on the $\left(U_i+\delta_i V_i,1-\nu_i\right)$, say $\widehat{G}$, is a consistent estimator of G. Then, Lemma 2 suggests estimating $S\left(\phi\right)$ by the sample average

$$\widehat{S}\left(\phi\right)=\frac{1}{n}\sum_{i=1}^{n}\frac{\phi\left(U_i,U_i+\delta_i V_i\right)\delta_i\rho_i}{1-\hat{G}\left(\left(U_i+\delta_i V_i\right)^{-}\right)}. \qquad [4.7]$$

As a technical remark, we point out that the result in Lemma 2 is valid for those functions ϕ such that $\phi\left(u,\upsilon\right)=0$ whenever υ is strictly greater than the upper bound τ of the support of $\min\left(T,C\right)$. Otherwise, the result should be read as

$$S^{\tau}\left(\phi\right)=\mathbb{E}\left[\frac{\phi\left(U,U+\delta V\right)\delta\rho}{1-G\left(\left(U+\delta V\right)^{-}\right)}\right],$$

where $S^{\tau}\left(\phi\right)$ is the expectation defined as

$$S^{\tau}\left(\phi\right)=\mathbb{E}\left[\phi\left(T_{12},T_{12}+T_{23}\right)\mathbb{I}\left(T_{12}\le T_{13}\right)\mathbb{I}\left(T_{12}+T_{23}\le\tau\right)\right].$$

Now, consider $\phi=\phi_{s,t}$ or $\phi=\tilde{\phi}_{s,t}$ for some $s\le t$. Since $\phi_{s,t}\left(u,u\right)=\tilde{\phi}_{s,t}\left(u,u\right)=0$, then

$$\widehat{S}\left(\phi_{s,t}\right)=\frac{1}{n}\sum_{i=1}^{n}\frac{\phi_{s,t}\left(U_i,U_i+\delta_i V_i\right)\nu_i}{1-\hat{G}\left(\left(U_i+\delta_i V_i\right)^{-}\right)},$$

and

$$\hat{S}\left(\tilde{\phi}_{s,t}\right)=\frac{1}{n}\sum_{i=1}^{n}\frac{\tilde{\phi}_{s,t}\left(U_i,U_i+\delta_i V_i\right)\nu_i}{1-\hat{G}\left(\left(U_i+\delta_i V_i\right)^{-}\right)}.$$

These expressions show that $\hat{S}\left(\phi_{s,t}\right)$ and $\hat{S}\left(\tilde{\phi}_{s,t}\right)$ are indeed multivariate Kaplan-Meier integrals with respect to the measure associated to $\widehat{K}$ (Stute, 1993). This fact will be very useful when investigating the asymptotic properties of the proposed estimators.

We define then,

$$\begin{aligned}\widehat{p_{12}}(s,t)&=\frac{\hat{S}\left(\phi_{s,t}\right)}{1-\widehat{H}(s)}\\&=\frac{1}{n\left(1-\widehat{H}(s)\right)}\sum_{i=1}^{n}\frac{\mathbb{I}\left(s<U_i\le t\right)\mathbb{I}\left(U_i+\delta_i V_i>t\right)\delta_i\rho_i}{1-\hat{G}\left(\left(U_i+\delta_i V_i\right)^{-}\right)},\end{aligned}\qquad [4.8]$$

and

$$\begin{aligned}\widehat{p_{22}}(s,t)&=\frac{\hat{S}\left(\tilde{\phi}_{s,t}\right)}{\hat{S}\left(\tilde{\phi}_{s,s}\right)}\\&=\frac{\sum_{i=1}^{n}\frac{\mathbb{I}\left(U_i\le s\right)\mathbb{I}\left(U_i+\delta_i V_i>t\right)\delta_i\rho_i}{1-\hat{G}\left(\left(U_i+\delta_i V_i\right)^{-}\right)}}{\sum_{i=1}^{n}\frac{\mathbb{I}\left(U_i\le s\right)\mathbb{I}\left(U_i+\delta_i V_i>s\right)\delta_i\rho_i}{1-\hat{G}\left(\left(U_i+\delta_i V_i\right)^{-}\right)}},\end{aligned}\qquad [4.9]$$

where (as mentioned) the indicator $\delta_i\rho_i$ can be replaced by ν_i.

Estimators proposed here have fewer jump points but with larger steps when compared with Aalen-Johansen estimators. These estimators may step down or up only at event times (disease occurrence or death) and only when the individuals time is not censored. Of course, the number of jump points and the size of the steps will depend on the sample size and on the censoring percentage.

For the uncensored case, estimators [4.6], [4.8] and [4.9] give place to "obvious" estimators:

$$\widehat{p_{11}}(s,t)=\frac{\sum_{i=1}^{n}\mathbb{I}(U_i>t)}{\sum_{i=1}^{n}\mathbb{I}(U_i>s)},$$

$$\widehat{p_{12}}(s,t)=\frac{\sum_{i=1}^{n}\mathbb{I}(s<U_i\le t)\mathbb{I}(U_i+\delta_i V_i>t)\delta_i}{\sum_{i=1}^{n}\mathbb{I}(U_i>s)},$$

and

$$\widehat{p_{22}}(s,t)=\frac{\sum_{i=1}^{n}\mathbb{I}(U_i\le s)\mathbb{I}(U_i+\delta_i V_i>t)\delta_i}{\sum_{i=1}^{n}\mathbb{I}(U_i\le s)\mathbb{I}(U_i+\delta_i V_i>s)\delta_i}.$$

Note that these estimators, $\widehat{p_{hj}}(s,t)$, can be expressed as the proportion of the observations available in state *h* at time *s* which ends up in state *j* at time *t*:

$$\widehat{p_{hj}}(s,t)=\frac{\sum_{i=1}^{n}\mathbb{I}(X_i(s)=h,X_i(t)=j)}{\sum_{i=1}^{n}\mathbb{I}(X_i(s)=h)}.$$

As mentioned before, estimators for the rest of the transition probabilities, $p_{13}(s,t)$ and $p_{23}(s,t)$, can be obtained from the above ones. Certainly,

$$\widehat{p_{23}}(s,t)=1-\widehat{p_{22}}(s,t)$$

and

$$\widehat{p_{13}}(s,t)=1-\widehat{p_{11}}(s,t)-\widehat{p_{12}}(s,t).$$

Alternatively, since

$$\begin{aligned} p_{13}(s,t) &= \mathbb{P}\left[T \le t \middle| Z > s\right] \\ &= \mathbb{P}\left[T_{12} \le T_{13}, T_{12} + T_{23} \le t \middle| Z > s\right] + \mathbb{P}\left[T_{13} \le t, T_{12} > T_{13} \middle| Z > s\right] \\ &= \frac{\mathbb{P}\left[s < T_{12}, T_{12} \le T_{13}, T_{12} + T_{23} \le t\right]}{\mathbb{P}\left[Z > s\right]} + \frac{\mathbb{P}\left[s < T_{13} \le t, T_{12} > T_{13}\right]}{\mathbb{P}\left[Z > s\right]}, \end{aligned} \qquad [4.10]$$

one can certainly provide an alternative estimator for $p_{13}(s,t)$. The first term on the right hand side of [4.10] is the conditional probability of leaving state 1 into state 3 having visited the diseased state (i.e., $1 \to 2 \to 3$), denoted by $p_{13}^{d}(s,t)$. The second term is the conditional probability of leaving state 1 into state 3 without having visited the diseased state (i.e., $1 \to 3$), denoted by $p_{13}^{nd}(s,t)$.

As for the previous transition probabilities, also these two additional quantities, $p_{13}^{d}(s,t)$ and $p_{13}^{nd}(s,t)$, are determined by the joint distribution of (T_{12}, T_{13}, T_{23}). Specifically, the knowledge of the distribution of $Z = \min(T_{12}, T_{13})$ and expectations of type $S(\phi) = \mathbb{E}\left[\phi(T_{12}, T_{12} + T_{23})\mathbb{I}(T_{12} \le T_{13})\right]$ (considered before) arise when handling

$p_{13}^{d}(s,t)$ $\left(\phi(u,v) = \bar{\phi}_{s,t}(u,v) = \mathbb{I}(u > s, v \le t)\right)$,

$$p_{13}^{d}(s,t) = \frac{S\left(\bar{\phi}_{s,t}\right)}{\mathbb{P}(Z > s)},$$

whereas the expectations of type $D(\phi) = \mathbb{E}\left[\phi(T_{13}, T_{13})\mathbb{I}(T_{12} > T_{13})\right]$ arise when handling

$p_{13}^{nd}(s,t)$ $\left(\phi(u,v) = \bar{\phi}_{s,t}(u,v) = \mathbb{I}(u > s, v \le t)\right)$,

$$p_{13}^{nd}(s,t) = \frac{D\left(\bar{\phi}_{s,t}\right)}{\mathbb{P}(Z > s)}.$$

Lemma 2, suggests that $p_{13}^{d}(s,t)$ can be estimated by the sample average,

$$\widehat{p_{13}^{d}}(s,t)=\frac{1}{n\left(1-\hat{H}(s)\right)}\sum_{i=1}^{n}\frac{\mathbb{I}(U_i>s)\mathbb{I}(U_i+\delta_i V_i\le t)\delta_i\rho_i}{1-\hat{G}\left((U_i+\delta_i V_i)^{-}\right)}.$$

For the estimation of the expectation $D(\phi)=\mathbb{E}\left[\phi(T_{13},T_{13})\mathbb{I}(T_{12}>T_{13})\right]$ we need the following Lemma,

Lemma 3. For each function ϕ we have

$$D(\phi)=\mathbb{E}\left[\frac{\phi(U,U)(1-\delta)\eta}{1-G\left(U^{-}\right)}\right].$$

Proof.

Write

$$D(\phi)=\mathbb{E}\left\{\phi(T_{13},T_{13})\mathbb{I}(T_{12}>T_{13})\frac{\mathbb{I}(T_{13}\le C)}{\mathbb{E}\left[\mathbb{I}(T_{13}\le C)\middle|T_{12},T_{13}\right]}\right\}.$$

Now, since C is independent of the process,

$$\mathbb{E}\left[\mathbb{I}(T_{13}\le C)\middle|T_{12},T_{13}\right]=1-G\left((T_{13})^{-}\right).$$

On the other hand,

$$\mathbb{I}(T_{13}<T_{12})\mathbb{I}(T_{13}\le C)=(1-\delta)\eta\,.$$

Since, under $(1-\delta)\eta=1$, we have $U=T_{13}$, we get

$$D(\phi)=\mathbb{E}\left[\frac{\phi(T_{13},T_{13})(1-\delta)\eta}{1-G\left(T_{13}^{-}\right)}\right]=\mathbb{E}\left[\frac{\phi(U,U)(1-\delta)\eta}{1-G\left(U^{-}\right)}\right].\square$$

Then, $D(\phi) = \mathbb{E}\left[\phi(T_{13}, T_{13})\mathbb{I}(T_{12} > T_{13})\right]$ may be estimated by the sample average,

$$\begin{aligned}\widehat{D}(\phi) &= \widehat{\mathbb{E}}\left[\phi(T_{13}, T_{13})\mathbb{I}(T_{12} > T_{13})\right] \\ &= \frac{1}{n}\sum_{i=1}^{n} \frac{\mathbb{I}(s < U_i \le t)(1-\delta_i)\eta_i}{1-\hat{G}(U_i^-)},\end{aligned}$$

and $p_{13}^{nd}(s,t)$ can be estimated by

$$\widehat{p_{13}^{nd}}(s,t) = \frac{1}{n\left(1-\hat{H}(s)\right)}\sum_{i=1}^{n} \frac{\mathbb{I}(s < U_i \le t)(1-\delta_i)\eta_i}{1-\hat{G}(U_i^-)}. \qquad [4.11]$$

Then one possible estimator for $p_{13}(s,t)$ is given by the following expression,

$$\widehat{p_{13}^{\bullet}}(s,t) = \frac{1}{n\left(1-\hat{H}(s)\right)}\sum_{i=1}^{n} \frac{\mathbb{I}(U_i > s, U_i + \delta_i V_i \le t)\nu_i}{1-\hat{G}\left((U_i + \delta_i V_i)^-\right)}. \qquad [4.12]$$

Note that, since

$$p_{13}(s,t) = \mathbb{P}(T \le t | Z > s),$$

then, the above estimator [4.12] can be seen as an estimation problem involving the joint distribution of (Z,T). This problem is discussed below.

4.3.1.2. Some asymptotic results.

In this section our main interest is the asymptotic results for the estimated transition probabilities $\widehat{p_{hj}}(s,t)$. Special attention will be paid to the following issues: (i) consistency; (ii) convergence in distribution to a normal; and (iii) (limit) variance estimation.

We first introduce some results for an estimator of the general expectation

$$\begin{aligned} L(\varphi) &= \mathbb{E}\left[\varphi(Z,T)\right] \\ &= \mathbb{E}\left[\varphi(T_{12},T_{12}+T_{23})\mathbb{I}(T_{12}\le T_{13})\right]+\mathbb{E}\left[\varphi(T_{13},T_{13})\mathbb{I}(T_{12}>T_{13})\right] \end{aligned} \qquad [4.13]$$

Note that $p_{13}(s,t)$ is related to this expectation, through the choice $\varphi(u,v)=\bar{\phi}_{s,t}(u,v)=\mathbb{I}(u>s,v\le t)$. Note that for $s\le t$ we get $L\left(\bar{\phi}_{s,t}\right)=S\left(\bar{\phi}_{s,t}\right)+D\left(\bar{\phi}_{s,t}\right)$. Also, for $\varphi(s,t)=\phi_{s,t}(u,v)=\mathbb{I}(u\le s,v>t)$ and $s\le t$, we get $L\left(\phi_{s,t}\right)=S\left(\phi_{s,t}\right)$ (because $\phi_{s,t}(u,u)=0$, the second expectation on the right-hand side of [4.13] is zero). The same argument shows that $L\left(\tilde{\phi}_{s,t}\right)=S\left(\tilde{\phi}_{s,t}\right)$. Hence, estimators of $L(\varphi)$ are of much practical interest in the scope of the illness-death model.

Now, note that, rather than the pair (Z,T) we observe (U,Y) where $Y=U+\delta V$. Besides, it is known whether the value of Y is an uncensored observation of T $(\nu=1)$ or not $(\nu=0)$. So the problem of estimating $L(\varphi)$ has to do with the evaluation of Kaplan-Meier integrals in the presence of covariates (here, the role of the covariate will be played by the *Z*). In contrast with the existing related literature developed in the nineties (e.g. Stute, 1993), the *Z* may be censored too; however, *U* will offer an uncensored value of *Z* whenever $\nu=1$, and so any Kaplan-Meier integral based

on the weights of $\widehat{K}$ can be computed regardless of the referred issue. This is explained below.

Introduce the Kaplan-Meier integral

$$\hat{L}(\varphi)=\sum_{i=1}^{n} W_{in}^{K}\varphi\left(U_{[i]},Y_{(i)}\right),$$

where $Y_{(1)}\le Y_{(2)}\le\ldots\le Y_{(n)}$ are the ordered values of the Y_i's and $Y_i=U_i+\delta_i V_i$; $U_{[i]}$ is the concomitant of the *i*th order statistic $Y_{(i)}$, and W_{in}^{K} stands for the weight which is attached to $Y_{(i)}$ under $\widehat{K}$. Explicitly,

$$W_{in}^{K}=\frac{\nu_{[i]}}{n-i+1}\prod_{j=1}^{i-1}\left[1-\frac{\nu_{[j]}}{n-j+1}\right],$$

where $\nu_{[i]}$ is the censoring indicator of $Y_{(i)}$. Note that $W_{in}^{K}=0$ for any censored $Y_{(i)}$. Since $U_{[i]}=Z_{[i]}$ whenever $\nu_{[i]}=1$, we may write

$$\hat{L}(\varphi)=\sum_{i=1}^{n} W_{in}^{K}\varphi\left(Z_{[i]},Y_{(i)}\right).$$

A number of results are available for this $\hat{L}(\varphi)$, in the case of complete observation on *Z*. When right-censored values of *Z* correspond to censored observations of *T* (as in our case), the results are easily adapted. The needed identifiably conditions are satisfied in this application (they are consequences of the independence between *C* and (T_{12},T_{13},T_{23}) which of course implies the independence between *C* and (Z,T)), since: (a) *T* and *C* are independent random variables; (b) ν is independent of *Z* conditionally on *T*.

Define,

$$L^{\tau}(\varphi)=\mathbb{E}\left[\varphi(Z,T)\mathbb{I}(T\le\tau)\right],$$

where τ stands for the upper bound of the support of $Y=U+\delta V=\min(T,C)$. Specifically, we have (Stute, 1993):

$$\hat{L}(\varphi)\to L^{\tau}(\varphi) \quad \text{w. p. } 1 \text{ as } n\to\infty,$$

provided that the limit exists.

In the previous section, we obtained the following estimators for the transition probabilities in the illness-death model:

$$\widehat{p_{11}}(s,t)=\frac{1-\widehat{H}(t)}{1-\widehat{H}(s)},$$

$$\widehat{p_{12}}(s,t)=\frac{\hat{L}(\phi_{s,t})}{1-\widehat{H}(s)},$$

$$\widehat{p_{22}}(s,t)=\frac{\hat{L}(\tilde{\phi}_{s,t})}{\hat{L}(\tilde{\phi}_{s,s})},$$

$$\widehat{p_{23}}(s,t)=1-\widehat{p_{22}}(s,t)=1-\frac{\hat{L}(\tilde{\phi}_{s,t})}{\hat{L}(\tilde{\phi}_{s,s})},$$

$$\widehat{p_{13}}(s,t)=1-\widehat{p_{11}}(s,t)-\widehat{p_{12}}(s,t)=1-\frac{1-\widehat{H}(t)}{1-\widehat{H}(s)}-\frac{\hat{L}(\phi_{s,t})}{1-\widehat{H}(s)}.$$

As mentioned before, an alternative estimator for $p_{13}(s,t)$ (see [4.11]) is now expressed as

$$\widehat{p_{13}^{\bullet}}(s,t)=\frac{\hat{L}(\bar{\phi}_{s,t})}{1-\widehat{H}(s)},$$

where $\bar{\phi}_{s,t}(u,v)=\mathbb{I}(u>s,v\le t)$.

Estimators $\widehat{p_{1j}}(s,t)$, as introduced in [4.6], [4.8] and [4.12], do not satisfy the natural restriction

$$\sum_{j=1}^{3}\widehat{p_{1j}}(s,t)=1.$$

This problem can be overcome by considering, when estimating the distribution H of Z, empirical integrals based on the weights associated to $\widehat{K}$, the estimator of the survival time T. Certainly, introduce

$$\widehat{H^*}(t)=\sum_{i=1}^{n}W_{in}^{K}\mathbb{I}\left(U_{[i:n]}\le t\right)=\sum_{i=1}^{n}W_{in}^{K}\mathbb{I}\left(Z_{[i:n]}\le t\right),$$

note now that,

$$\hat{L}(\varphi_t)=\widehat{H^*}(t)\quad\text{where}\quad\varphi_t(u,v)=\mathbb{I}(u\le t).$$

Furthermore, $\hat{L}(\varphi_t)\to H(t)$ follows from the general convergence result above. Introduce now the following estimators:

$$\widehat{p_{11}^*}(s,t)=\frac{1-\hat{L}(\varphi_t)}{1-\hat{L}(\varphi_s)},$$

$$\widehat{p_{12}^*}(s,t)=\frac{\hat{L}(\phi_{s,t})}{1-\hat{L}(\varphi_s)},$$

$$\widehat{p_{13}^*}(s,t)=\frac{\hat{L}(\bar{\phi}_{s,t})}{1-\hat{L}(\varphi_s)};$$

for which

$$\hat{L}(\varphi_t)=\sum_{i=1}^{n}W_{in}^{K}\mathbb{I}\left(Z_{[i]}\le t\right),$$

$$\hat{L}(\phi_{s,t})=\sum_{i=1}^{n}W_{in}^{K}\mathbb{I}\left(s<Z_{[i]}\le t,Y_{(i)}>t\right),$$

$$\hat{L}\left(\bar{\phi}_{s,t}\right)=\sum_{i=1}^{n}W_{in}^{K}\mathbb{I}\left(s<Z_{[i]},Y_{(i)}\le t\right).$$

Because $Z_{[i]}\le Y_{(i)}$ whenever $W_{in}^{K}\neq 0$, then $\hat{L}\left(\bar{\phi}_{s,t}\right)=\sum_{i=1}^{n}W_{in}^{K}\mathbb{I}\left(s<Z_{[i]}\le t,Y_{(i)}\le t\right)$.

Therefore,

$$\begin{aligned}
&\hat{L}\left(\varphi_s\right)+\hat{L}\left(\tilde{\phi}_{s,t}\right)+\hat{L}\left(\bar{\phi}_{s,t}\right)=\\
&\quad=\sum_{i=1}^{n}W_{in}^{K}\mathbb{I}\left(Z_{[i]}\le s\right)+\sum_{i=1}^{n}W_{in}^{K}\mathbb{I}\left(s<Z_{[i]}\le t,Y_{(i)}\le t\right)+\sum_{i=1}^{n}W_{in}^{K}\mathbb{I}\left(s<Z_{[i]}\le t,Y_{(i)}>t\right)\\
&\quad=\sum_{i=1}^{n}W_{in}^{K}\mathbb{I}\left(Z_{[i]}\le s\right)+\sum_{i=1}^{n}W_{in}^{K}\mathbb{I}\left(s<Z_{[i]}\le t\right)\\
&\quad=\sum_{i=1}^{n}W_{in}^{K}\mathbb{I}\left(Z_{[i]}\le t\right)\\
&\quad=\hat{L}\left(\varphi_t\right)
\end{aligned}$$

from which it follows that $\sum_{j=1}^{3}\widehat{p_{1j}^{*}}\left(s,t\right)=1$.

Note that in the case $s=0$, we obtain $\widehat{p_{13}^{*}}\left(0,t\right)=\hat{L}\left(\bar{\phi}_{0,t}\right)=\hat{K}\left(t\right)$, which means that the distribution function of the survival time of the process, *K*, may be estimated by $\hat{L}\left(\bar{\phi}_{0,t}\right)$.

The main goal in the remainder of this section will be establishing the asymptotics for $\widehat{p_{hj}}\left(s,t\right)$ and $\widehat{p_{hj}^{*}}\left(s,t\right)$.

Asymptotics for $\widehat{p_{11}}\left(s,t\right)$

First, consider the empirical weighted average

$$\hat{R}\left(\tilde{\varphi}\right)=\sum_{i=1}^{n}W_{in}^{H}\tilde{\varphi}\left(U_{(i)}\right)$$

which can be regarded as an estimator for $R\left(\tilde{\varphi}\right)=\mathbb{E}\left[\tilde{\varphi}\left(Z\right)\right]$. For simplicity we assume continuity in the following. Introduce $H^{\tau_0}\left(z\right)=\mathbb{P}\left(Z\le\min\left(z,\tau_0\right)\right)$, where τ_0 denotes

the upper bound of the support of U. Of course, H^{τ_0} equals H whenever the support of Z is contained in that of C (otherwise, we can not expect consistency in the right tail of the distribution). Thus if we define

$$R^{\tau_0}\left(\tilde{\varphi}\right)=\int\tilde{\varphi}dH^{\tau_0}=\mathbb{E}\left[\tilde{\varphi}\left(Z\right)\mathbb{I}\left(Z\le\tau_0\right)\right],$$

we have (Stute and Wang, 1993):

$$\hat{R}\left(\tilde{\varphi}\right)\to R^{\tau_0}\left(\tilde{\varphi}\right)\quad\text{w. p. 1,}$$

provided that the limit exists. Let N denote the distribution function of U, and N^j, $j=0,1$, for the subdistributions pertaining to the censored and uncensored subpopulations, respectively; that is, $N^j\left(x\right)=\mathbb{P}\left(U\le x,\gamma=j\right)$, $j=0,1$. We also have (Stute, 1995):

$$\hat{R}\left(\tilde{\varphi}\right)=\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\tilde{\varphi}\left(U_i\right)\gamma_i}{1-G\left(U_i\right)}+\eta_1^{\tilde{\varphi}}\left(U_i\right)\left(1-\gamma_i\right)-\eta_2^{\tilde{\varphi}}\left(U_i\right)\right\}+r_n\left(\tilde{\varphi}\right)$$

where

$$\begin{aligned}\eta_1^{\tilde{\varphi}}\left(x\right)&=\frac{1}{1-N\left(x\right)}\int\frac{\mathbb{I}\left(x<w\right)\tilde{\varphi}\left(w\right)}{1-G\left(w\right)}dN^1\left(w\right)\\&=\frac{1}{1-N\left(x\right)}\mathbb{E}\left[\frac{\mathbb{I}\left(x<U\right)\tilde{\varphi}\left(U\right)\gamma}{1-G\left(U\right)}\right]\\&=\frac{1}{1-N\left(x\right)}\mathbb{E}\left[\mathbb{I}\left(x<Z\right)\tilde{\varphi}\left(Z\right)\mathbb{I}\left(Z<\tau_0\right)\right],\end{aligned}$$

$$\eta_2^{\tilde{\varphi}}\left(x\right)=\int\frac{\mathbb{I}\left(\upsilon<x\right)\eta_1^{\tilde{\varphi}}\left(\upsilon\right)}{1-N\left(\upsilon\right)}dN^0\left(\upsilon\right),$$

and

$$r_n\left(\tilde{\varphi}\right)=o_{\mathbb{P}}\left(n^{-1/2}\right),$$

provided that the following conditions hold:

C1.0 $$\mathbb{E}\left[\frac{\tilde{\varphi}(Z)^2\,\mathbb{I}(Z\le\tau_0)}{1-G(Z)}\right]=\mathbb{E}\left[\frac{\tilde{\varphi}(U)^2\,\gamma}{(1-G(U))^2}\right]=\int\frac{\tilde{\varphi}(z)^2\,dN^1(z)}{(1-G(z))^2}<\infty,$$

C2.0 $$\mathbb{E}\left[\left|\tilde{\varphi}(Z)\right|C_0(Z)^{1/2}\,\mathbb{I}(Z\le\tau_0)\right]=\int\left|\tilde{\varphi}(z)\right|C_0(z)^{1/2}\,dH^{\tau_0}(z)<\infty$$

where $C_0(z)$ is defined as

$$C_0(z)=\int_0^z\frac{dG(x)}{(1-N(x))(1-G(x))}.$$

This result, together with the Central Limit Theorem (CLT), immediately leads to

$$n^{1/2}\left[\hat{R}(\tilde{\varphi})-R^{\tau_0}(\tilde{\varphi})\right]\to N\left(0,\sigma_0(\tilde{\varphi})\right)\quad\text{in law,}$$

where $\sigma_0^2(\tilde{\varphi})=\mathrm{var}\left(\xi_0^{\tilde{\varphi}}\right)$ and

$$\xi_0^{\tilde{\varphi}}=\frac{\tilde{\varphi}(U)\gamma}{1-G(U)}+\eta_1^{\tilde{\varphi}}(U)(1-\gamma)-\eta_2^{\tilde{\varphi}}(U).$$

Moreover, in our applications we need to consider transformations such as $\hat{\theta}(\varphi_1,\varphi_2)=g\left(\hat{R}(\varphi_1),\hat{R}(\varphi_2)\right)$ where typically $g(x,y)=x/y$, $g(x,y)=(1-x)/(1-y)$ or $g(x,y)=x/(1-y)$. Then, the multivariate CLT and the delta method (Serfling, 1980, page 124) ensure

$$n^{1/2}\left[\hat{\theta}(\varphi_1,\varphi_2)-g\left(R^{\tau_0}(\varphi_1),R^{\tau_0}(\varphi_2)\right)\right]\to N\left(0,\sigma_0(\varphi_1,\varphi_2)\right)\quad\text{in law,}$$

where

$$\sigma_0^2(\varphi_1,\varphi_2)=\sigma_{0,1}^2\left[\left.\frac{\partial g}{\partial x}\right|_{(x,y)=\left(R^{\tau_0}(\varphi_1),R^{\tau_0}(\varphi_2)\right)}\right]^2+\sigma_{0,2}^2\left[\left.\frac{\partial g}{\partial y}\right|_{(x,y)=\left(R^{\tau_0}(\varphi_1),R^{\tau_0}(\varphi_2)\right)}\right]^2$$
$$+2\sigma_{0,12}\left.\frac{\partial g}{\partial x}\frac{\partial g}{\partial y}\right|_{(x,y)=\left(R^{\tau_0}(\varphi_1),R^{\tau_0}(\varphi_2)\right)}$$

and where

$\sigma_{0,1}^2 = \sigma_0^2(\varphi_1)$, $\sigma_{0,2}^2 = \sigma_0^2(\varphi_2)$, $\sigma_{0,12} = \operatorname{cov}\left(\xi_0^{\varphi_1}, \xi_0^{\varphi_2}\right)$.

We can easily see that $\sigma_{0,12}$ can be simplified as

$$\sigma_{0,12} = \mathbb{E}\left[\frac{\varphi_1(U)\varphi_2(U)\gamma}{(1-G(U))^2}\right] - \mathbb{E}\left[\eta_1^{\varphi_1}(U)\eta_1^{\varphi_2}(U)(1-\gamma)\right] - R^{\tau_0}(\varphi_1)R^{\tau_0}(\varphi_2),$$

which can be used for computation of the limiting variances and covariances $\sigma_{0,1}^2$, $\sigma_{0,2}^2$, $\sigma_{0,12}$. Besides, the first summand is also written as

$$\mathbb{E}\left[\frac{\varphi_1(U)\varphi_2(U)\gamma}{(1-G(U))^2}\right] = \mathbb{E}\left[\frac{\varphi_1(Z)\varphi_2(Z)\mathbb{I}(Z \le \tau_0)}{1-G(Z)}\right].$$

Now note that $\widehat{p_{11}}(s,t) = \left(1-\hat{R}\left(\tilde{\varphi}_t\right)\right)\Big/\left(1-\hat{R}\left(\tilde{\varphi}_s\right)\right)$ where $\tilde{\varphi}_t(u) = \mathbb{I}(u \le t)$ and $\tilde{\varphi}_s(u) = \mathbb{I}(u \le s)$. The existence of $\mathbb{E}\left[\tilde{\varphi}_t(Z)\mathbb{I}(Z \le \tau_0)\right]$ is guaranteed since the function $\tilde{\varphi}_t$ is bounded (similarly for $\tilde{\varphi}_s$). On the other hand, $\mathbb{E}\left[\tilde{\varphi}_t(Z)\mathbb{I}(Z \le \tau_0)\right] = H(t)$ for each $t \le \tau_0$. Hence, the estimator $\widehat{p_{11}}(s,t)$ converges almost surely to $p_{11}(s,t)$ provided that $t \le \tau_0$. We also obtain, for $t \le \tau_0$,

$$n^{1/2}\left[\widehat{p_{11}}(s,t) - p_{11}(s,t)\right] \to N\left(0, \sigma_0(s,t)\right) \qquad \text{in law,}$$

where

$$\sigma_0^2(s,t) = \sigma_{0,1}^2 \frac{1}{(1-H(s))^2} + \sigma_{0,2}^2 \frac{(1-H(t))^2}{(1-H(s))^4} - 2\sigma_{0,12}\frac{1-H(t)}{(1-H(s))^3},$$

$$\sigma_{0,1}^2 = \mathbb{E}\left[\frac{\mathbb{I}(Z \le t)}{1-G(Z)}\right] - \mathbb{E}\left[\frac{(H(t)-H(U))^2(1-\gamma)}{(1-N(U))^2}\right] - H(t)^2,$$

$$\sigma_{0,2}^2 = \mathbb{E}\left[\frac{\mathbb{I}(Z \le s)}{1-G(Z)}\right] - \mathbb{E}\left[\frac{(H(s)-H(U))^2(1-\gamma)}{(1-N(U))^2}\right] - H(s)^2,$$

and

$$\sigma_{0,12} = \mathbb{E}\left[\frac{\mathbb{I}(Z \le s)}{1-G(Z)}\right] - \mathbb{E}\left[\frac{\big(H(s)-H(U)\big)\big(H(t)-H(U)\big)(1-\gamma)}{\big(1-N(U)\big)^2}\right] - H(s)H(t).$$

For the uncensored case, we simply obtain

$$\sigma_0^2(s,t) = \frac{\big(H(t)-H(s)\big)\big(1-H(t)\big)}{\big(1-H(s)\big)^3}.$$

In order to check that $\tilde{\varphi}_t$ satisfies conditions C1.0 and C2.0 above for $t \le \tau_0$, we write

$$\mathbb{E}\left[\frac{\tilde{\varphi}_t(Z)^2\,\mathbb{I}(Z \le \tau_0)}{1-G(Z)}\right] \le \frac{H(t)}{1-G(t)} < \infty$$

and

$$\mathbb{E}\Big[\Big|\tilde{\varphi}_t(Z)\Big|C_0(Z)^{1/2}\,\mathbb{I}(Z \le \tau_0)\Big] \le C_0(t)^{1/2}\,H(t) \le \frac{H(t)G(t)^{1/2}}{\Big[\big(1-N(t)\big)\big(1-G(t)\big)\Big]^{1/2}} < \infty.$$

We have established that $n^{1/2}\Big[\widehat{p_{11}}(s,t) - p_{11}(s,t)\Big]$ converges in distribution to a normal distribution with mean zero and variance $\sigma_0^2(s,t)$. This asymptotic variance can be consistently estimated in finite samples by plug-in methods with

$$\widehat{\sigma_0^2}(s,t) = \widehat{\sigma_{0,1}^2}\frac{1}{\big(1-\widehat{H}(s)\big)^2} + \widehat{\sigma_{0,2}^2}\frac{\big(1-\widehat{H}(t)\big)^2}{\big(1-\widehat{H}(s)\big)^4} - 2\widehat{\sigma_{0,12}}\frac{1-\widehat{H}(t)}{\big(1-\widehat{H}(s)\big)^3},$$

where

$$\widehat{\sigma_{0,1}^2} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\mathbb{I}(U_i \le t)\gamma_i}{\big(1-\widehat{G}(U_i)\big)^2} - \frac{\big(\widehat{H}(t)-\widehat{H}(U_i)\big)^2(1-\gamma_i)}{\big(1-\widehat{N}(U_i)\big)^2}\right\} - \big(\widehat{H}(t)\big)^2,$$

$$\widehat{\sigma^2_{0,2}} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\mathbb{I}(U_i \le s)\gamma_i}{\left(1-\widehat{G}(U_i)\right)^2} - \frac{\left(\widehat{H}(s)-\widehat{H}(U_i)\right)^2(1-\gamma_i)}{\left(1-\widehat{N}(U_i)\right)^2}\right\} - \left(\widehat{H}(s)\right)^2$$

and

$$\widehat{\sigma_{0,12}} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\mathbb{I}(U_i \le s)\gamma_i}{\left(1-\widehat{G}(U_i)\right)^2} - \frac{\left(\widehat{H}(s)-\widehat{H}(U_i)\right)\left(\widehat{H}(t)-\widehat{H}(U_i)\right)(1-\gamma_i)}{\left(1-\widehat{N}(U_i)\right)^2}\right\} - \left(\widehat{H}(s)\right)\left(\widehat{H}(t)\right);$$

for some estimators $\widehat{H}$, $\widehat{G}$ and $\widehat{N}$.

This variance estimator, $\widehat{\sigma_0}(s,t)$, can be used to construct an asymptotic normal theory confidence interval for $p_{11}(s,t)$. An approximately $1-\alpha$ confidence interval for $p_{11}(s,t)$ is given by

$$\left[\widehat{p_{11}}(s,t) - z_{\alpha/2}\times\frac{\widehat{\sigma_0}(s,t)}{\sqrt{n}}, \widehat{p_{11}}(s,t) + z_{\alpha/2}\times\frac{\widehat{\sigma_0}(s,t)}{\sqrt{n}}\right].$$

It is important to note, that in this case, because our estimator is equivalent to Aalen-Johansen estimator one estimator of $\sigma_0(s,t)$ is given by the generalization of Greenwood's formula to the illness-death model (see section 3.3.3).

Asymptotics for $\widehat{p^*_{11}}(s,t)$

For investigating the asymptotic properties of $\widehat{p^*_{11}}(s,t)$, we first review some existing results for the empirical

$$\hat{L}(\varphi) = \sum_{i=1}^{n} W^K_{in}\varphi\left(Z_{[i]}, Y_{(i)}\right).$$

Let M denote the distribution function of Y, $M^0(y) = \mathbb{P}(Y \le y, \nu = 0)$ and $M^{11}(z,y) = \mathbb{P}(Z \le z, Y \le y, \nu = 1)$. We have (Stute, 1996):

$$\hat{L}(\varphi)=\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\varphi(Z_i,Y_i)\nu_i}{1-G(Y_i)}+\gamma_1^{\varphi}(Y_i)(1-\nu_i)-\gamma_2^{\varphi}(Y_i)\right\}+r_n'(\varphi)$$

where

$$\begin{aligned}\gamma_1^{\varphi}(x)&=\frac{1}{1-M(x)}\int\frac{\mathbb{I}(x<w)\varphi(z,w)}{1-G(w)}M^{11}(dz,dw)\\&=\frac{1}{1-M(x)}\mathbb{E}\left[\frac{\mathbb{I}(x<Y)\varphi(Z,Y)\nu}{1-G(Y)}\right]\\&=\frac{1}{1-M(x)}\mathbb{E}\left[\mathbb{I}(x<T)\varphi(Z,T)\mathbb{I}(T\le\tau)\right],\end{aligned}$$

$$\gamma_2^{\varphi}(x)=\int\frac{\mathbb{I}(\upsilon<x)\gamma_1^{\varphi}(\upsilon)}{1-M(\upsilon)}dM^{0}(\upsilon),$$

and

$$r_n'(\varphi)=o_{\mathbb{P}}\left(n^{-1/2}\right),$$

provided that the following conditions hold:

C1. $$\mathbb{E}\left[\frac{\varphi(Z,T)^2\,\mathbb{I}(T\le\tau)}{1-G(T)}\right]=\mathbb{E}\left[\frac{\varphi(Z,Y)^2\,\nu}{(1-G(Y))^2}\right]<\infty,$$

C2. $$\mathbb{E}\left[\left|\varphi(Z,T)\right|C(T)^{1/2}\,\mathbb{I}(T\le\tau)\right]<\infty$$

where $C(z)$ is defined as

$$C(z)=\int_0^z\frac{dG(x)}{(1-M(x))(1-G(x))}.$$

In the same way as above, an application of the CLT gives

$$n^{1/2}\left[\hat{L}(\varphi)-L^{\tau}(\varphi)\right]\to N(0,\sigma(\varphi))\quad\text{in law,}$$

where $L^{\tau}(\varphi)=\mathbb{E}\left[\varphi(Z,T)\mathbb{I}(T\le\tau)\right]$ and τ is the upper bound of the support of Y. The asymptotic variance is given by $\sigma^2(\varphi)=\mathrm{var}\left(\xi^{\varphi}\right)$ with

$$\xi^{\varphi} = \frac{\varphi(Z,Y)\nu}{1-G(Y)} + \gamma_1^{\varphi}(Y)(1-\nu) - \gamma_2^{\varphi}(Y).$$

On the other hand, for $\hat{\theta}(\varphi_1,\varphi_2) = g\left(\hat{L}(\varphi_1), \hat{L}(\varphi_2)\right)$ we obtain

$$n^{1/2}\left[\hat{\theta}(\varphi_1,\varphi_2) - g\left(L^{\tau}(\varphi_1), L^{\tau}(\varphi_2)\right)\right] \to N\left(0, \sigma(\varphi_1,\varphi_2)\right) \quad \text{in law,}$$

where

$$\sigma^2(\varphi_1,\varphi_2) = \sigma_1^2\left[\left.\frac{\partial g}{\partial x}\right|_{(x,y)=\left(L^{\tau}(\varphi_1),L^{\tau}(\varphi_2)\right)}\right]^2 + \sigma_2^2\left[\left.\frac{\partial g}{\partial y}\right|_{(x,y)=\left(L^{\tau}(\varphi_1),L^{\tau}(\varphi_2)\right)}\right]^2 + 2\sigma_{12}\left.\frac{\partial g}{\partial x}\frac{\partial g}{\partial y}\right|_{(x,y)=\left(L^{\tau}(\varphi_1),L^{\tau}(\varphi_2)\right)}$$

with

$$\sigma_1^2 = \sigma^2(\varphi_1),\ \sigma_2^2 = \sigma^2(\varphi_2),\ \sigma_{12} = \operatorname{cov}\left(\xi^{\varphi_1}, \xi^{\varphi_2}\right).$$

Similarly as above, σ_{12} can be simplified as

$$\sigma_{12} = \mathbb{E}\left[\frac{\varphi_1(Z,Y)\varphi_2(Z,Y)\nu}{\left(1-G(Y)\right)^2}\right] - \mathbb{E}\left[\gamma_1^{\varphi_1}(Y)\gamma_1^{\varphi_2}(Y)(1-\nu)\right] - L^{\tau}(\varphi_1)L^{\tau}(\varphi_2),$$

where the first summand is also written as

$$\mathbb{E}\left[\frac{\varphi_1(Z,Y)\varphi_2(Z,Y)\nu}{\left(1-G(Y)\right)^2}\right] = \mathbb{E}\left[\frac{\varphi_1(Z,T)\varphi_2(Z,T)\mathbb{I}(T\le\tau)}{1-G(T)}\right].$$

Note that $\widehat{p_{11}^*}(s,t) = \left(1-\hat{L}(\varphi_t)\right)\Big/\left(1-\hat{L}(\varphi_s)\right)$ where $\varphi_t(u,v) = \mathbb{I}(u\le t)$ and $\varphi_s(u,v) = \mathbb{I}(u\le s)$. The existence of $\mathbb{E}\left[\varphi_t(Z,T)\mathbb{I}(T\le\tau)\right]$ is guaranteed since the function φ_t is bounded (similarly for φ_s). On the other hand, $\mathbb{E}\left[\varphi_t(Z,T)\mathbb{I}(T\le\tau)\right] = H(t)$ holds if the support of T is contained in that of C. Hence, if $\mathbb{P}(T\le\tau) = 1$, the estimator $\widehat{p_{11}^*}(s,t)$ converges almost surely to $p_{11}(s,t)$.

Considering

$$C_1(t,x) = \mathbb{P}(Z \le t, T > x),$$

we also obtain

$$n^{1/2}\left[\widehat{p_{11}^*}(s,t) - p_{11}(s,t)\right] \to N\left(0, \sigma(s,t)\right) \qquad \text{in law,}$$

where

$$\sigma^2(s,t) = \sigma_1^2 \frac{1}{\left(1-H(s)\right)^2} + \sigma_2^2 \frac{\left(1-H(t)\right)^2}{\left(1-H(s)\right)^4} - 2\sigma_{12} \frac{1-H(t)}{\left(1-H(s)\right)^3},$$

$$\sigma_1^2 = \mathbb{E}\left[\frac{\mathbb{I}(Z \le t)}{1-G(T)}\right] - \mathbb{E}\left[\frac{C_1(t,Y)^2(1-\nu)}{\left(1-M(Y)\right)^2}\right] - H(t)^2,$$

$$\sigma_2^2 = \mathbb{E}\left[\frac{\mathbb{I}(Z \le s)}{1-G(T)}\right] - \mathbb{E}\left[\frac{C_1(s,Y)^2(1-\nu)}{\left(1-M(Y)\right)^2}\right] - H(s)^2,$$

and

$$\sigma_{12} = \mathbb{E}\left[\frac{\mathbb{I}(Z \le s)}{1-G(T)}\right] - \mathbb{E}\left[\frac{C_1(s,Y)C_1(t,Y)(1-\nu)}{\left(1-M(Y)\right)^2}\right] - H(s)H(t).$$

For the uncensored case, we simply obtain

$$\sigma^2(s,t) = \frac{\left(H(t)-H(s)\right)\left(1-H(t)\right)}{\left(1-H(s)\right)^3}.$$

In this case, the estimator $\widehat{H^*}$ is the ordinary empirical distribution function of the Z_i, and both $\widehat{p_{11}^*}(s,t)$ and $\widehat{p_{11}}(s,t)$ coincide.

Note that, unlike in the preceding section, the function φ_t may or may not satisfy conditions C1 and C2 above. Finally, note that it is possible to introduce an estimator for the limiting variance of $\widehat{p_{11}^*}(s,t)$ by plug-in methods with

$$\widehat{\sigma^2}(s,t)=\widehat{\sigma_1^2}\frac{1}{\left(1-\widehat{H}(s)\right)^2}+\widehat{\sigma_2^2}\frac{\left(1-\widehat{H}(t)\right)^2}{\left(1-\widehat{H}(s)\right)^4}-2\widehat{\sigma_{12}}\frac{1-\widehat{H}(t)}{\left(1-\widehat{H}(s)\right)^3},$$

where

$$\widehat{\sigma_1^2}=\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\mathbb{I}(U_i\le t)\nu_i}{\left(1-\widehat{G}(Y_i)\right)^2}-\frac{\left(\widehat{C_1}(t,Y_i)\right)^2(1-\nu_i)}{\left(1-\widehat{M}(Y_i)\right)^2}\right\}-\left(\widehat{H}(t)\right)^2,$$

$$\widehat{\sigma_2^2}=\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\mathbb{I}(U_i\le s)\nu_i}{\left(1-\widehat{G}(Y_i)\right)^2}-\frac{\left(\widehat{C_1}(s,Y_i)\right)^2(1-\nu_i)}{\left(1-\widehat{M}(Y_i)\right)^2}\right\}-\left(\widehat{H}(s)\right)^2,$$

$$\widehat{\sigma_{12}}=\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\mathbb{I}(U_i\le s)\nu_i}{\left(1-\widehat{G}(Y_i)\right)^2}-\frac{\widehat{C_1}(s,Y_i)\widehat{C_1}(t,Y_i)(1-\nu_i)}{\left(1-\widehat{M}(Y_i)\right)^2}\right\}-\left(\widehat{H}(s)\right)\left(\widehat{H}(t)\right);$$

for some estimators $\widehat{H}$, $\widehat{G}$, $\widehat{M}$ and $\widehat{C_1}$.

Asymptotics for $\widehat{p_{22}}(s,t)$

Our goal is to investigate the asymptotic properties for the estimator pertaining to

$$p_{22}(s,t)=\frac{S\left(\tilde{\phi}_{s,t}\right)}{S\left(\tilde{\phi}_{s,s}\right)},\qquad s\le t,$$

where $\tilde{\phi}_{s,t}(u,v)=\mathbb{I}(u\le s,v>t)$. Recall that since $\tilde{\phi}_{s,t}(u,u)=0$ and $\tilde{\phi}_{s,s}(u,u)=0$ we have $L\left(\tilde{\phi}_{s,t}\right)=S\left(\tilde{\phi}_{s,t}\right)$ and $L\left(\tilde{\phi}_{s,s}\right)=S\left(\tilde{\phi}_{s,s}\right)$. In order to avoid problems with censoring in the right tail of the distribution, we can write

$$L\left(\tilde{\phi}_{s,t}\right)=\mathbb{E}\left[\mathbb{I}(Z\le s,T>t)\right]=H(s)-\mathbb{E}\left[\mathbb{I}(Z\le s,T\le t)\right].$$

The distribution H could be estimated by [4.5], whereas the expectation involving (Z,T) is estimated by

$$\hat{L}(\psi_{s,t}) = \sum_{i=1}^{n} W_{in}^{K} \psi_{s,t}(Z_{[i]}, Y_{(i)}),$$

where $\psi_{s,t}(u,v) = \mathbb{I}(u \le s, v \le t)$. This would lead to

$$\widehat{p_{22}^{\triangle}}(s,t) = \frac{\widehat{H}(s) - \hat{L}(\psi_{s,t})}{\widehat{H}(s) - \hat{L}(\psi_{s,s})}.$$

However, here we investigate the asymptotics for

$$\widehat{p_{22}}(s,t) = \frac{\widehat{H^*}(s) - \hat{L}(\psi_{s,t})}{\widehat{H^*}(s) - \hat{L}(\psi_{s,s})} = \frac{\hat{L}(\tilde{\phi}_{s,t})}{\hat{L}(\tilde{\phi}_{s,s})}.$$

For this estimator, the theory reviewed in the preceding section will be used.

First, we have

$$\widehat{p_{22}}(s,t) \to \frac{\mathbb{E}\left[\tilde{\phi}_{s,t}(Z,T)\mathbb{I}(T \le \tau)\right]}{\mathbb{E}\left[\tilde{\phi}_{s,s}(Z,T)\mathbb{I}(T \le \tau)\right]} \quad \text{w. p. 1}$$

where τ denotes the upper bound of the support of *Y*. Hence, if the support of *T* is contained in that of *C*, we get consistency. Assume $\mathbb{P}(T \le \tau) = 1$ in what follows. We have

$$n^{1/2}\left[\widehat{p_{22}}(s,t) - p_{22}(s,t)\right] \to N(0, \pi(s,t)) \qquad \text{in law,}$$

where

$$\pi^2(s,t) = \pi_1^2 \frac{1}{L(\tilde{\phi}_{s,s})^2} + \pi_2^2 \frac{L(\tilde{\phi}_{s,t})^2}{L(\tilde{\phi}_{s,s})^4} - 2\pi_{12} \frac{L(\tilde{\phi}_{s,t})}{L(\tilde{\phi}_{s,s})^3},$$

$$\pi_1^2 = \mathbb{E}\left[\frac{\mathbb{I}(Z \le s, T > t)}{1 - G(T)}\right] - \mathbb{E}\left[\frac{C_1(s, t \vee Y)^2 (1-\nu)}{(1 - M(Y))^2}\right] - L(\tilde{\phi}_{s,t})^2,$$

$$\pi_2^2 = \mathbb{E}\left[\frac{\mathbb{I}\left(Z \le s, T > s\right)}{1-G\left(T\right)}\right] - \mathbb{E}\left[\frac{C_1\left(s, s \vee Y\right)^2\left(1-\nu\right)}{\left(1-M\left(Y\right)\right)^2}\right] - L\left(\tilde{\phi}_{s,s}\right)^2,$$

and

$$\pi_{12} = \mathbb{E}\left[\frac{\mathbb{I}\left(Z \le s, T > t\right)}{1-G\left(T\right)}\right] - \mathbb{E}\left[\frac{C_1\left(s, t \vee Y\right)C_1\left(s, s \vee Y\right)\left(1-\nu\right)}{\left(1-M\left(Y\right)\right)^2}\right] - L\left(\tilde{\phi}_{s,s}\right)L\left(\tilde{\phi}_{s,t}\right).$$

For the uncensored case, we simply obtain

$$\pi^2\left(s,t\right) = \frac{L\left(\tilde{\phi}_{s,t}\right)}{L\left(\tilde{\phi}_{s,s}\right)^3}\left[L\left(\tilde{\phi}_{s,s}\right) - L\left(\tilde{\phi}_{s,t}\right)\right].$$

Again, using analogous procedures it is possible to estimate the limiting variance $\pi^2\left(s,t\right)$ of $\widehat{p_{22}}\left(s,t\right)$ by plug-in methods:

$$\widehat{\pi^2}\left(s,t\right) = \widehat{\pi_1^2}\frac{1}{\hat{L}\left(\tilde{\phi}_{s,s}\right)^2} + \widehat{\pi_2^2}\frac{\hat{L}\left(\tilde{\phi}_{s,t}\right)^2}{\hat{L}\left(\tilde{\phi}_{s,s}\right)^4} - 2\widehat{\pi_{12}}\frac{\hat{L}\left(\tilde{\phi}_{s,t}\right)}{\hat{L}\left(\tilde{\phi}_{s,s}\right)^3},$$

where

$$\widehat{\pi_1^2} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\mathbb{I}\left(U_i \le s\right)\mathbb{I}\left(Y_i > t\right)\nu_i}{\left(1-\widehat{G}\left(Y_i\right)\right)^2} - \frac{\left(\widehat{C_1}\left(s, t \vee Y_i\right)\right)^2\left(1-\nu_i\right)}{\left(1-\widehat{M}\left(Y_i\right)\right)^2}\right\} - \left(\hat{L}\left(\tilde{\phi}_{s,t}\right)\right)^2,$$

$$\widehat{\pi_2^2} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\mathbb{I}\left(U_i \le s\right)\mathbb{I}\left(Y_i > s\right)\nu_i}{\left(1-\widehat{G}\left(Y_i\right)\right)^2} - \frac{\left(\widehat{C_1}\left(s, s \vee Y_i\right)\right)^2\left(1-\nu_i\right)}{\left(1-\widehat{M}\left(Y_i\right)\right)^2}\right\} - \left(\hat{L}\left(\tilde{\phi}_{s,s}\right)\right)^2,$$

and

$$\widehat{\pi_{12}} = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\mathbb{I}\left(U_i \le s\right)\mathbb{I}\left(Y_i > t\right)\nu_i}{\left(1-\widehat{G}\left(Y_i\right)\right)^2} - \frac{\widehat{C_1}\left(s, t \vee Y_i\right)\widehat{C_1}\left(s, s \vee Y_i\right)\left(1-\nu_i\right)}{\left(1-\widehat{M}\left(Y_i\right)\right)^2}\right\} - \hat{L}\left(\tilde{\phi}_{s,s}\right)\hat{L}\left(\tilde{\phi}_{s,t}\right).$$

Regarding technical conditions C1 and C2, we mention that $\tilde{\phi}_{s,t}$ may or may not verify these assumptions. Some care will be needed in applications.

Asymptotics for $\widehat{p_{12}^{*}}(s,t)$

Our goal here is to investigate the asymptotic properties for the estimator pertaining to

$$p_{12}(s,t)=\frac{S(\phi_{s,t})}{1-H(s)}$$

where $\phi_{s,t}(u,v)=\mathbb{I}(s<u\le t,v>t)$. Recall that since $\phi_{s,t}(u,u)=0$ we have $L(\phi_{s,t})=S(\phi_{s,t})$. In order to avoid problems with censoring in the right tail of the distribution, we can write

$$L(\phi_{s,t})=\mathbb{E}\left[\mathbb{I}(s<Z\le t,T>t)\right]=H(t)-H(s)-\mathbb{E}\left[\mathbb{I}(s<Z\le t,T\le t)\right].$$

Here, the expectation involving (Z,T) is estimated by

$$\hat{L}(\bar{\phi}_{s,t})=\sum_{i=1}^{n}W_{in}^{K}\bar{\phi}_{s,t}(Z_{[i]},Y_{(i)})$$

where $\bar{\phi}_{s,t}(u,v)=\mathbb{I}(s<u\le t,v\le t)$. This would lead to

$$\widehat{p_{12}^{\triangle}}(s,t)=\frac{\widehat{H}(t)-\widehat{H}(s)-\hat{L}(\bar{\phi}_{s,t})}{1-\widehat{H}(s)}.$$

However, here we investigate the asymptotics for

$$\widehat{p_{12}^{*}}(s,t)=\frac{\widehat{H^{*}}(t)-\widehat{H^{*}}(s)-\hat{L}(\bar{\phi}_{s,t})}{1-\widehat{H^{*}}(s)}=\frac{\hat{L}(\phi_{s,t})}{1-\hat{L}(\varphi_{s})}.$$

First, we have

$$\widehat{p_{12}^{*}}(s,t)\to\frac{\mathbb{E}\left[\phi_{s,t}(Z,T)\mathbb{I}(T\le\tau)\right]}{1-\mathbb{E}\left[\varphi_{s}(Z,T)\mathbb{I}(T\le\tau)\right]}\quad\text{w. p. 1}$$

where τ denotes the upper bound of the support of *Y*. Hence, if the support of *T* is contained in that of *C*, we get consistency. Assume $\mathbb{P}(T \le \tau) = 1$ in what follows. Introduce

$$C_2(s,t,t \vee x) = \mathbb{P}(s < Z \le t, T > t \vee x).$$

We also have

$$n^{1/2}\left[\widehat{p_{12}^*}(s,t) - p_{12}(s,t)\right] \to N\left(0, \tilde{\pi}(s,t)\right) \qquad \text{in law,}$$

where

$$\tilde{\pi}^2(s,t) = \tilde{\pi}_1^2 \frac{1}{\left(1-L(\varphi_s)\right)^2} + \tilde{\pi}_2^2 \frac{L(\phi_{s,t})^2}{\left(1-L(\varphi_s)\right)^4} - 2\tilde{\pi}_{12} \frac{L(\phi_{s,t})}{\left(1-L(\varphi_s)\right)^3},$$

$$\tilde{\pi}_1^2 = \mathbb{E}\left[\frac{\mathbb{I}(s < Z \le t, T > t)}{1-G(T)}\right] - \mathbb{E}\left[\frac{C_2(s,t,t \vee Y)^2 (1-\nu)}{\left(1-M(Y)\right)^2}\right] - L(\phi_{s,t})^2,$$

$$\tilde{\pi}_2^2 = \mathbb{E}\left[\frac{\mathbb{I}(Z \le s)}{1-G(T)}\right] - \mathbb{E}\left[\frac{C_1(s,Y)^2 (1-\nu)}{\left(1-M(Y)\right)^2}\right] - L(\varphi_s)^2,$$

and

$$\tilde{\pi}_{12} = -\mathbb{E}\left[\frac{C_2(s,t,t \vee Y) C_1(s,Y)(1-\nu)}{\left(1-M(Y)\right)^2}\right] - L(\varphi_s) L(\phi_{s,t}).$$

For the uncensored case, we simply obtain

$$\tilde{\pi}^2(s,t) = \frac{L(\phi_{s,t})}{\left(1-L(\varphi_s)\right)^3}\left[1 - L(\varphi_s) - L(\phi_{s,t}) + 4L(\varphi_s)L(\phi_{s,t})\right].$$

Again, it is possible to estimate the limiting variance $\tilde{\pi}^2(s,t)$ of $\widehat{p_{12}^*}(s,t)$ by plug-in methods:

$$\widehat{\tilde{\pi}^2}(s,t) = \widehat{\tilde{\pi}_1^2} \frac{1}{\left(1-\hat{L}(\varphi_s)\right)^2} + \widehat{\tilde{\pi}_2^2} \frac{\hat{L}(\phi_{s,t})^2}{\left(1-\hat{L}(\varphi_s)\right)^4} - 2\widehat{\tilde{\pi}_{12}} \frac{\hat{L}(\phi_{s,t})}{\left(1-\hat{L}(\varphi_s)\right)^3},$$

where

$$\widehat{\widetilde{\pi}}_1^2 = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\mathbb{I}(s<U_i\le t)\mathbb{I}(Y_i>t)\nu_i}{\left(1-\widehat{G}(Y_i)\right)^2} - \frac{\left(\widehat{C}_2(s,t,t\vee Y_i)\right)^2(1-\nu_i)}{\left(1-\widehat{M}(Y_i)\right)^2}\right\} - \left(\hat{L}(\phi_{s,t})\right)^2,$$

$$\widehat{\widetilde{\pi}}_2^2 = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\mathbb{I}(U_i\le s)\nu_i}{\left(1-\widehat{G}(Y_i)\right)^2} - \frac{\left(\widehat{C}_1(s,Y_i)\right)^2(1-\nu_i)}{\left(1-\widehat{M}(Y_i)\right)^2}\right\} - \left(\hat{L}(\varphi_s)\right)^2,$$

and

$$\widehat{\widetilde{\pi}}_{12} = -\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\widehat{C}_2(s,t,t\vee Y_i)\widehat{C}_1(s,Y_i)(1-\nu_i)}{\left(1-\widehat{M}(Y_i)\right)^2}\right\} - \hat{L}(\varphi_s)\hat{L}(\phi_{s,t}).$$

Note that the above functions $\phi_{s,t}$ and φ_t may or may not verify the assumptions C1 and C2.

Asymptotics for $\widehat{p_{12}}(s,t)$

We now come back to $\widehat{p_{12}}(s,t)$ as defined above. We have then,

$$\widehat{p_{12}}(s,t) = \frac{\widehat{H^*}(t) - \widehat{H^*}(s) - \hat{L}\left(\tilde{\phi}_{s,t}\right)}{1-\widehat{H}(s)}$$
$$= \frac{\hat{L}(\phi_{s,t})}{1-\hat{R}(\tilde{\varphi}_s)}$$

First, we have

$$\widehat{p_{12}}(s,t) \to \frac{\mathbb{E}\left[\phi_{s,t}(Z,T)\mathbb{I}(T\le\tau)\right]}{1-\mathbb{E}\left[\tilde{\varphi}_s(Z)\mathbb{I}(Z\le\tau_0)\right]} \quad \text{w. p. 1}$$

where τ_0 denotes the upper bound of the support of U whereas τ denotes de upper bound of the support of Y. Hence, if the support of T is contained in that of C, we get consistency. Assuming $\mathbb{P}\left(Z \le \tau_0, T \le \tau\right) = 1$ in what follows, we have

$$n^{1/2}\left[\widehat{p_{12}}(s,t) - p_{12}(s,t)\right] \to N\left(0, \widetilde{\sigma}(s,t)\right) \qquad \text{in law,}$$

where

$$\widetilde{\sigma}^2(s,t) = \widetilde{\sigma}_1^2 \frac{1}{\left(1-H(s)\right)^2} + \widetilde{\sigma}_2^2 \frac{L\left(\phi_{s,t}\right)^2}{\left(1-H(s)\right)^4} - 2\widetilde{\sigma}_{12} \frac{L\left(\phi_{s,t}\right)}{\left(1-H(s)\right)^3},$$

$$\widetilde{\sigma}_1^2 = \mathbb{E}\left[\frac{\mathbb{I}\left(s < Z \le t, T > t\right)}{1-G(T)}\right] - \mathbb{E}\left[\frac{C_2\left(s,t,t\vee Y\right)^2\left(1-\nu\right)}{\left(1-M(Y)\right)^2}\right] - L\left(\phi_{s,t}\right)^2,$$

$$\widetilde{\sigma}_2^2 = \mathbb{E}\left[\frac{\mathbb{I}\left(Z \le s\right)}{1-G(Z)}\right] - \mathbb{E}\left[\frac{\left(H(s)-H(U)\right)^2\left(1-\gamma\right)}{\left(1-N(U)\right)^2}\right] - H(s)^2,$$

and

$$\widetilde{\sigma}_{12} = -\mathbb{E}\left[\frac{C_2\left(s,t,t\vee Y\right)\left(H(s)-H(U)\right)\left(1-\nu\right)\left(1-\gamma\right)}{\left(1-N(U)\right)\left(1-M(Y)\right)}\right] - L\left(\phi_{s,t}\right)H(s).$$

For the uncensored case we simply obtain

$$\widetilde{\sigma}^2 = \frac{L\left(\phi_{s,t}\right)}{\left(1-H(s)\right)^3}\left[1-H(s)-L\left(\phi_{s,t}\right)+4H(s)L\left(\phi_{s,t}\right)\right].$$

Again, it is possible to estimate the limiting variance $\widetilde{\sigma}^2(s,t)$ of $\widehat{p_{12}}(s,t)$ by plug-in methods:

$$\widehat{\widetilde{\sigma}^2}(s,t) = \widehat{\widetilde{\sigma}_1^2} \frac{1}{\left(1-\widehat{H}(s)\right)^2} + \widehat{\widetilde{\sigma}_2^2} \frac{\hat{L}\left(\phi_{s,t}\right)^2}{\left(1-\widehat{H}(s)\right)^4} - 2\widehat{\widetilde{\sigma}_{12}} \frac{\hat{L}\left(\phi_{s,t}\right)}{\left(1-\widehat{H}(s)\right)^3},$$

where

$$\widetilde{\sigma}_1^2 = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\mathbb{I}(s<U_i\le t)\mathbb{I}(Y_i>t)\nu_i}{\left(1-\widehat{G}(Y_i)\right)^2} - \frac{\left(\widehat{C}_2(s,t,t\vee Y_i)\right)^2(1-\nu_i)}{\left(1-\widehat{M}(Y_i)\right)^2}\right\} - \left(\hat{L}(\phi_{s,t})\right)^2,$$

$$\widetilde{\sigma}_2^2 = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\mathbb{I}(U_i\le s)\nu_i}{\left(1-\widehat{G}(Y_i)\right)^2} - \frac{\left(\widehat{H}(s)-\widehat{H}(U_i)\right)^2(1-\gamma_i)}{\left(1-\widehat{N}(U_i)\right)^2}\right\} - \left(\widehat{H}(s)\right)^2,$$

and

$$\widetilde{\sigma}_{12} = -\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\widehat{C}_2(s,t,t\vee Y_i)\left(\widehat{H}(s)-\widehat{H}(U_i)\right)(1-\nu_i)(1-\gamma_i)}{\left(1-\widehat{N}(U_i)\right)\left(1-\widehat{M}(Y_i)\right)}\right\} - \hat{L}(\phi_{s,t})\widehat{H}(s).$$

In this section, we have introduced and investigated up to six different estimators for the three different transition probabilities: $p_{11}(s,t)$, $p_{12}(s,t)$ and $p_{22}(s,t)$. As shown before, more different estimators for these parameters could have been introduced, and no general answer regarding their relative efficiency can be given in principal. Two main estimators were considered here: using estimators $\widehat{p_{11}}(s,t)$, $\widehat{p_{12}}(s,t)$ and $\widehat{p_{22}}(s,t)$ (see, [4.6], [4.8] and [4.9]); or using estimators $\widehat{p_{11}^*}(s,t)$, $\widehat{p_{12}^*}(s,t)$. For obtaining the asymptotic results, we have used existing theory devoted to Kaplan-Meier integrals. When the expectation to be estimated depends only on *Z* (but not on *T*), general theory in Stute and Wang (1993) and Stute (1995) is to be used. However, the estimation of expectations involving (Z,T) requires using the extended theory in Stute (1993, 1996). In the following sections, a simulation study and an application are performed to illustrate differences between Aalen-Johansen estimators and non-Markovian estimators. To emphasize the differences between both approaches, we shall use in both cases the transition probability $p_{22}(s,t)$ and the corresponding non-Markov estimator $\widehat{p_{22}}(s,t)$.

4.3.1.3. Simulation study.

In this section, simulation studies were undertaken to compare Markovian (Aalen-Johansen) and non-Markovian estimators for non-homogeneous illness-death model.

Three settings were considered here, differing on transition from state 2 to state 3. In the first setting, times (T_{12}, T_{13}, T_{23}) are exponential distributed and mutually independent, yielding a homogeneous Markov process (see Lemma in the appendix A). For the second setting, times (T_{12}, T_{13}, T_{23}) are mutually independent, T_{12} and T_{13} are exponentially distributed, but for T_{23} times were generated from a Weibull distribution with shape parameter 2 and scale parameter 0.05. In the third setting, independent exponentially distributed times were considered for T_{12} and T_{13}, while $T_{23} = 1.7 \times T_{12}$. For all settings, the rate parameter used here for distributions T_{12} and T_{13} were 0.039 and 0.026 so that $\mathbb{P}(T_{13} \geq T_{12}) = 0.6$. In the first setting we used the rate parameter 0.065 for T_{23}.

Note that settings 2 and 3 do not fulfill the Markov assumption.

The performance of both Markovian (Aalen-Johansen) and non-Markovian estimators was examined under the presence of three levels of censoring (the uncensored case and two different levels of censoring). For each subject, a censoring time was generated from an exponential distribution, with rate parameter 0.013 and 0.035 (determining the heaviness of the censoring). Subjects alive at this censoring time were then censored at that time. The proportion of individuals with censored observations (either in state 1 or state 2) can be found through the following expression:

$$\text{Proportion of censoring} = \mathbb{P}\left[C < \min\left(T_{12}, T_{13}\right)\right] + \\ + \mathbb{P}\left[T_{12} \le \min\left(T_{13}, C\right)\right] \times \mathbb{P}\left[C < T_{12} + T_{23} \,\middle|\, T_{12} \le \min\left(T_{13}, C\right)\right].$$

In this section we compare Markovian and non-Markovian approaches through the estimation of the transition probabilities $p_{22}(s,t)$ for some fixed value of *s*, $(s = 10 \text{ and } s = 25)$. We choose to compare both approaches through this transition probability to emphasize the differences between them. Observe that fixing *s*, $p_{22}(s,t) = 1 - p_{23}(s,t)$ is a function of time *t* alone. For this simulation study, $\widehat{p_{22}}(s,t)$ is the non-Markovian estimator examined.

For the evaluation of the simulations, we shall borrow the work of Couper and Pepe (1997), which contains among other things, a comprehensive simulation study. As Couper and Pepe (1997), we use the integrated absolute bias, integrated variance and the integrated MSE. For each setting we derived the analytic expression of $p_{22}(s,t)$ (given in the appendix A) so that the bias and the MSE of the estimator, could be examined.

For each configuration, $K = 100$ data sets were generated, with two different sample sizes $N = 100$ and $N = 200$.

Let $\widehat{p_{22}^{k}}(s,t)$ denote the estimated transition probability at time *t* for some *s* fixed on the *k*th generated data set. For each time *t* considered, we may obtain the mean for all generated data sets, $\overline{\widehat{p_{22}}}(s,t) = \frac{1}{K}\sum_{k=1}^{K} \widehat{p_{22}^{k}}(s,t)$.

We then define, for some fixed value of *s*, the pointwise estimates of the bias, variance and Mean Square Error (MSE) as:

$$\widehat{\text{bias}}(t) = p_{22}(s,t) - \overline{\widehat{p_{22}}}(s,t),$$

$$\widehat{\text{var}}\left(\widehat{p_{22}}(s,t)\right)=\frac{1}{K-1}\sum_{k=1}^{K}\left(\widehat{p_{22}^{k}}(s,t)-\overline{\widehat{p_{22}}}(s,t)\right)^2,$$

and

$$\widehat{\text{MSE}}\left(\widehat{p_{22}}(s,t)\right)=\frac{1}{K}\sum_{k=1}^{K}\left(\widehat{p_{22}^{k}}(s,t)-p_{22}(s,t)\right)^2.$$

To summarize the results, we also calculate the integrated absolute bias, integrated variance and the integrated MSE, defined in Table 4.1.

Tables 4.2, 4.3 and 4.4, present the summary statistics for the two estimators (Markovian and non-Markovian) for settings 1, 2 and 3, respectively.

Table 4.1. Summary statistics measuring integrated bias, integrated variance and integrated mean square error.

Statistic	**Definition**	**Estimator**
Integrated absolute bias	$\int_{s}^{t_1}\left\|\text{bias}(t)\right\|dt$	$\sum_{t=s}^{t_1}\left\|\widehat{\text{bias}}(t)\right\|\Delta$
Integrated variance	$\int_{s}^{t_1}\text{var}\left(\widehat{p_{22}}(s,t)\right)dt$	$\sum_{t=s}^{t_1}\widehat{\text{var}}\left(\widehat{p_{22}}(s,t)\right)\Delta$
Integrated MSE	$\int_{s}^{t_1}\text{MSE}\left(\widehat{p_{22}}(s,t)\right)dt$	$\sum_{t=s}^{t_1}\widehat{\text{MSE}}\left(\widehat{p_{22}}(s,t)\right)\Delta$

s is a fixed value; $t_1=s+50$; $t=s+l\Delta$ with Δ=0.2 and $l=1,...,250$.

As shown in Tables 4.2-4.4, the non-Markovian estimator revealed itself to be significantly more precise in settings 2 and 3 while the Markovian estimator obtains better results whenever the process is Markov (setting 1).

The non-Markovian estimator obtains in all settings a small bias, while the Markovian estimator is grossly biased in the last two settings. It is clear that the bias clearly dominates the performance of the Markovian estimator in the last two settings (and especially in the last one), leading to a larger MSE. In fact, in some cases (see Tables 4.3 and 4.4) the performance of the Markovian estimator is disappointing despite having obtained a smaller variance than the non-Markovian estimator. The reason for

the bias in Markovian estimator for settings 3 is that the time spent in state 2 is associated with the time of entry into that same state. In this last setting (clearly non-Markov), individuals who enter state 2 early also leaves this state early, so the Markovian estimator will overestimate the time spent in state 2, and therefore $p_{22}(s,t)$. In an analogous way, the individual's time spent in state 2 will be underestimated for individuals with a later entry in state 2.

Although we restrict the integrated bias, variance and mean square error (MSE) to the interval $[s, s+50]$, we verified that, for settings 2 and 3, the enlargement of this interval increases at a faster rate in the Markovian estimator for the bias and MSE (results are not shown). Furthermore, we verified that for other values of s, our estimator still obtains better results whenever the process is not Markov (results are not shown).

The bias for the non-Markovian estimator increases with *s* for setting 1 and 3 but not for setting 2. For the Markovian estimator, the bias increases with *s* only for setting 1. In all cases, the variance increases with *s* and with censoring for both estimators. Furthermore, the variance is smaller for the Markovian estimator in setting 1, but our estimator revealed to be competitive when the process is not Markov (settings 2 and 3).

The simulations showed that increasing the sample size reduces the variance for both estimators. Apart from setting 1, the bias (and consequently the MSE) in the non-Markovian estimator, decreases at a faster rate when increasing the sample size.

In Figure 4.1, we choose to present the curves of MSE, bias and variance obtained by using Markovian and non-Markovian estimators for setting 3 with $s = 25$, $N = 200$ and about 32% of censored observations (see results in Table 4.4). The disparity in the curves for the MSE and the bias clearly reflects the failure of the Markov assumption. This figure shows that the MSE of the Markovian estimate is

increasing as a function of time whereas for the non-Markovian estimator it increases (slowly) until around time $t = 35$ and then slowly decreases. Curves for the bias are clearly different for Markovian and non-Markovian estimators, showing that the Markovian estimator is grossly biased. Note however that the variance for the Markovian estimator is smaller than that of the non-Markovian estimator until around time $t = 67$.

Curves obtained for other scenarios are presented in Appendix A2.

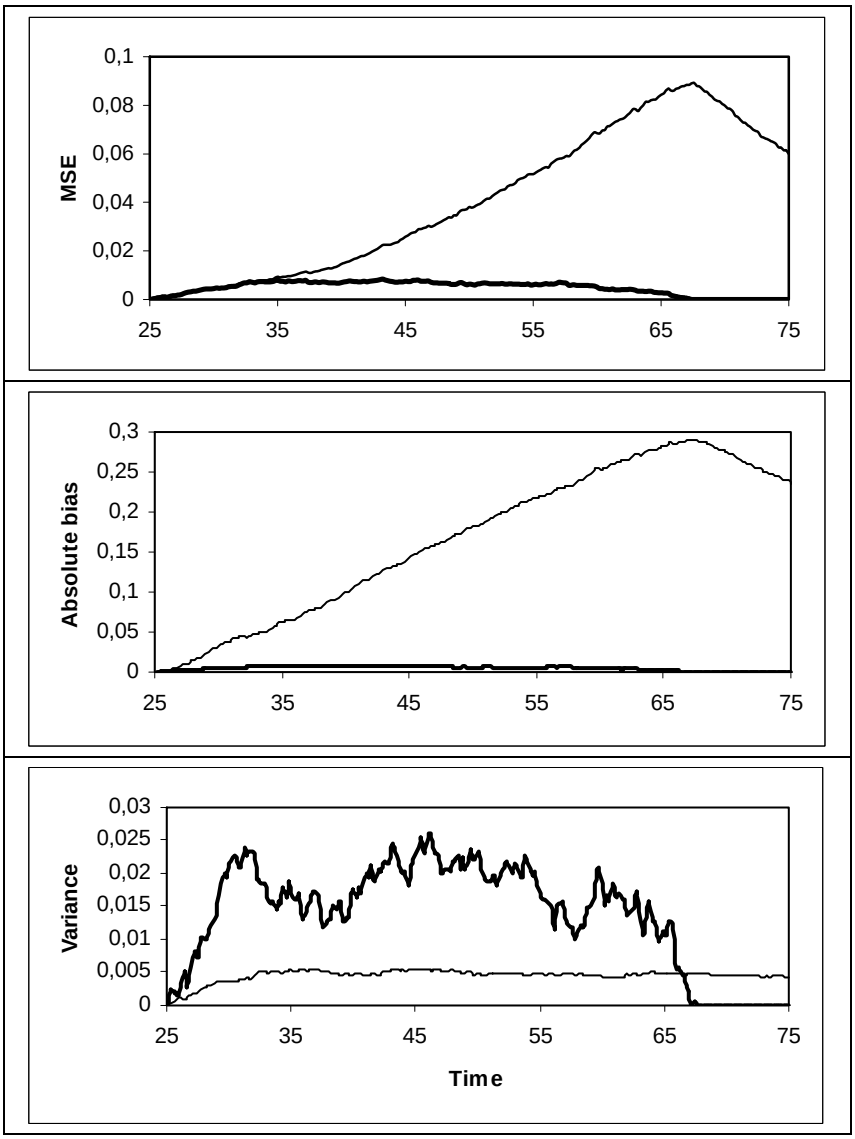


Figure 4.1. Curves obtained for setting 3 with $s = 25$, $N = 200$ and 32% of censored observations for Aalen-Johansen estimator (solid line) and non-Markovian estimator (solid bold line).

Table 4.2. Estimates of integrated absolute bias, integrated variance and integrated mean square error of $\widehat{p_{22}}(s,t)$ for setting 1, according to fixed value *s,* censoring and sample size.

		$s=10$						$s=25$					
	censoring	0 %		25 %		49 %		0 %		25 %		49 %	
	Estimator	Markov	non-Markov	Markov	non-Markov	Markov	non-Markov	Markov	non-Markov	Markov	non-Markov	Markov	non-Markov
N=100	MSE	0.1846	0.3545	0.2194	0.5250	0.4639	1.1462	0.2673	0.3536	0.5414	0.7780	1.2099	1.7156
	Bias	0.1118	0.3185	0.2214	0.6573	0.3160	2.4289	0.3055	0.5372	0.5676	0.8634	0.9749	4.4109
	Variance	0.1860	0.3553	0.2204	0.5174	0.4654	1.0228	0.2673	0.3503	0.5389	0.7660	1.1967	1.4558
N =200	MSE	0.0975	0.1816	0.1213	0.3127	0.2003	0.6057	0.1422	0.1901	0.2200	0.3617	0.4976	0.9535
	Bias	0.1128	0.1964	0.1446	0.2936	0.1363	1.4246	0.1341	0.2722	0.1911	1.0560	0.2020	3.1437
	Variance	0.0981	0.1824	0.1219	0.3135	0.2016	0.5658	0.1431	0.1902	0.2211	0.3393	0.5013	0.7434

Table 4.3. Estimates of integrated absolute bias, integrated variance and integrated mean square error of $\widehat{p_{22}}(s,t)$ for setting 2, according to fixed value *s,* censoring and sample size.

		$s=10$						$s=25$					
	censoring	0 %		27 %		52 %		0 %		27 %		52 %	
	Estimator	Markov	non-Markov	Markov	non-Markov	Markov	non-Markov	Markov	non-Markov	Markov	non-Markov	Markov	non-Markov
N=100	MSE	0.5522	0.2019	0.6284	0.2560	0.6520	0.4986	0.3114	0.1771	0.4597	0.2717	1.0531	0.8721
	Bias	4.0598	0.2611	4.2482	0.2498	3.6002	1.0800	2.3669	0.1599	2.7128	0.2484	2.8612	0.5753
	Variance	0.1435	0.2011	0.1764	0.2566	0.3345	0.4660	0.1543	0.1776	0.2637	0.2721	0.8613	0.8683
N=200	MSE	0.3508	0.0946	0.4309	0.1212	0.5204	0.3020	0.2044	0.0944	0.2517	0.1452	0.4400	0.3341
	Bias	3.3451	0.5010	3.6921	0.1373	3.7938	0.5657	2.2515	0.2641	2.3468	0.1348	2.3717	0.7759
	Variance	0.0625	0.0878	0.0806	0.1219	0.1569	0.2925	0.0829	0.0929	0.1111	0.1459	0.3098	0.3185

Table 4.4. Estimates of integrated absolute bias, integrated variance and integrated mean square error of $\widehat{p_{22}}(s,t)$ for setting 3, according to fixed value *s,* censoring and sample size.

		$s=10$						$s=25$					
	censoring	0 %		32 %		54 %		0 %		32 %		54 %	
	Estimator	Markov	non-Markov	Markov	non-Markov	Markov	non-Markov	Markov	non-Markov	Markov	non-Markov	Markov	non-Markov
N =100	MSE	3.6578	0.2240	3.7886	0.2362	3.8156	0.3356	1.8320	0.3384	2.2145	0.5998	3.4799	1.7735
	Bias	11.9496	0.1478	12.0429	0.1714	11.6123	0.2466	7.5164	0.2618	8.0231	0.5805	8.2430	1.1118
	Variance	0.2495	0.2245	0.3414	0.2361	0.5174	0.3342	0.3091	0.3397	0.5257	0.5956	1.7313	1.7544
N =200	MSE	3.7978	0.1017	3.7635	0.1349	3.5556	0.1624	1.8287	0.1610	2.0457	0.2377	2.4639	0.8934
	Bias	12.4177	0.0794	12.2114	0.1747	11.7588	0.1296	7.8200	0.2081	8.3416	0.6901	7.9424	0.8348
	Variance	0.1270	0.1021	0.1565	0.1338	0.2198	0.1626	0.1565	0.1610	0.2264	0.2274	0.7795	0.8777

4.3.1.4. Application to PROVA data.

For illustration purposes we use data from a Danish clinical trial in cirrhosis of the liver (Andersen et al., 2000). The PROVA clinical trial was conceived to evaluate the effect of propanolol and/or sclerotherapy versus no treatment on risk of bleeding episodes and survival in patients with cirrhosis. In the period between 1985 and 1989, 286 patients entered the study. Some of these patients developed bleeding episodes, while others did not. The occurrence of these intermediate events may affect the patient outcome and can be included as a transient state in an illness-death multi-state model (see Figure 4.2).

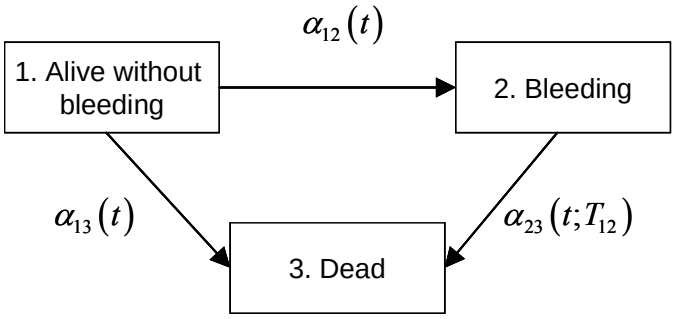


Figure 4.2. Illness-death model for PROVA trial.

From the total of 286 subjects, 50 developed bleeding episodes, and among these 29 died. On the other hand, 46 patients died without developing bleeding episodes. The rest of the patients remained alive and had no bleeding episodes up to the end of the follow-up.

PROVA trial was previously used by Andersen et al. (2000) to compare multi-state models and simple Cox regression models in regard to their simplicity and interpretation. Andersen et al. (2000) first used a Cox Markov model which turned out to fit poorly since the process was not Markov. To avoid the failure of the Markov assumption, Cox semi-Markov models were then used.

Here we present some figures to illustrate the differences between the Aalen-Johansen estimator and proposed non-Markov estimators of the transition probabilities. In Figure 4.3, we present estimated transition probabilities, $\widehat{p_{22}}(s,t)$, for a fixed value $s = 200$ and $s = 500$ (days).

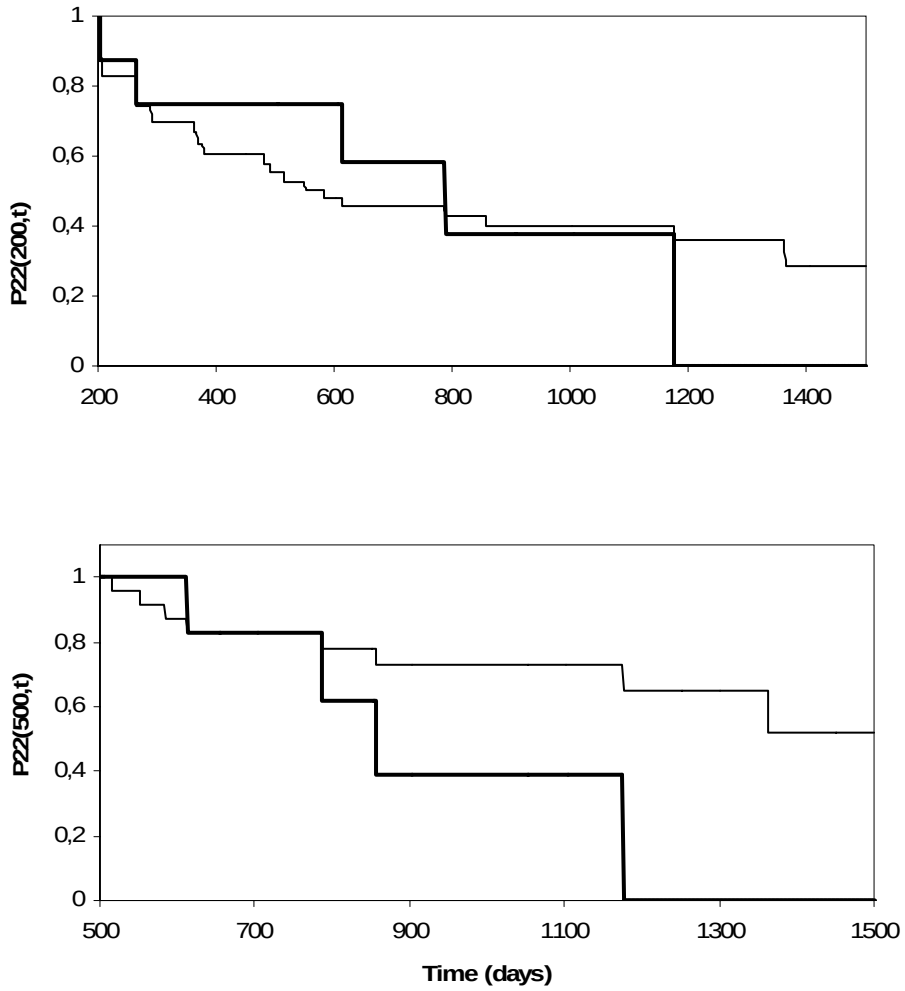


Figure 4.3. Estimated transition probabilities for $\widehat{p_{22}}(s,t)$, $s = 200$ (top) and $s = 500$ (bottom): Aalen-Johansen estimator (solid line) and non-Markov model (bold solid line).

4.3.1.5. Conclusions.

When using multi-state models for the analysis of survival data, some care is necessary in choosing an appropriate model. Traditionally, statistical methods for analyzing such models depend on the Markov assumption. We have shown through our simulation study that unless the process is, in fact, Markov, the non-Markovian estimator here proposed is a wise choice. Simulations suggest that the Aalen-Johansen estimator is highly susceptible to departures from the true transition probabilities when the process is not Markov. Moreover, for large survival data sets, our estimator always provides good results.

4.3.1.6. Cumulative incidence functions.

Another quantity of interest in multi-state modelling is the cause-specific cumulative incidence function, as defined by Kalbfleisch and Prentice (1980). In fact, if our interest is the estimation of failure probabilities, cumulative incidence functions are the appropriate quantity to use. We denote the cumulative incidence for disease as $CI_{12}(t) = \mathbb{P}(T_{12} \le t, T_{12} \le T_{13})$. Note that $CI_{12}(t)$ should not be confused with the transition probability, $p_{12}(0,t)$. The cumulative incidence function represents the probability of one individual's being or having been in the diseased state at time *t*, while $p_{12}(0,t)$ is the conditional probability of a healthy individual's being at entry time in the diseased state at later time *t*.

Note that,

$$CI_{12}(t) = \mathbb{P}(T_{12} \le t, T_{12} \le T_{13}) = \mathbb{E}\left[\varphi_t(T_{12}, T_{12}+T_{23})\mathbb{I}(T_{12} \le T_{13})\right],$$

where $\varphi_t(u,v) = \mathbb{I}(u \le t)$. In the same spirit as above, this quantity can be estimated by

$$\widehat{CI_{12}}(t) = \frac{1}{n}\sum_{i=1}^{n} \frac{\mathbb{I}(U_i \le t)\delta_i \rho_i}{1-\hat{G}\left((U_i+\delta_i V_i)^-\right)},$$

and it is consistent under the assumption of independence between (T_{12}, T_{13}, T_{23}) and *C*. The formal details are omitted.

For the uncensored case, it is simplified by

$$\widehat{CI_{12}}(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(U_i \le t)\delta_i .$$

The cumulative incidence function for death is the probability of one individual's being dead within time *t*. Because death is an absorbing state, the cumulative incidence for death is the transition probability from state 1 into state 3,

$$CI_{13}(t) = p_{13}(0,t).$$

4.3.1.7. Inference on the joint distribution function of (T_{12}, T_{13}, T_{23}).

Finally, we consider the problem of estimating the distribution functions of T_{13} and (T_{12}, T_{23}) in the illness-death model. For this purpose, we will assume (besides the independence between the censoring variable *C* and the process) that T_{13} and (T_{12}, T_{23}) are independent random variables. Let F_{13} be the distribution function of T_{13}. For the

estimation of F_{13}, we simply compute the Kaplan-Meier product limit $\widehat{F_{13}}$, based on $\left(U,(1-\delta)\eta\right)$. The consistency of this estimator follows from the independence of the variables T_{12} and T_{13}. On the other hand, the joint distribution function of $\left(T_{12},T_{23}\right)$ can be written as an expectation:

$$\mathbb{P}\left(T_{12}\le s,T_{23}\le t\right)=\mathbb{E}\left[\psi\left(T_{12},T_{23}\right)\right] \quad \text{for } \psi(u,v)=\mathbb{I}\left(u\le s,v\le t\right).$$

Of course, for $\widetilde{\psi}(u,v)=\psi\left(u,v-u\right)$ we have

$$\begin{aligned}\mathbb{P}\left(T_{12}\le s,T_{23}\le t\right)&=\mathbb{E}\left[\widetilde{\psi}\left(T_{12},T_{12}+T_{23}\right)\right]\\&=\mathbb{E}\left[\widetilde{\psi}\left(T_{12},T_{12}+T_{23}\right)\mathbb{I}\left(T_{12}\le T_{13}\right)\right]+\mathbb{E}\left[\widetilde{\psi}\left(T_{12},T_{12}+T_{23}\right)\mathbb{I}\left(T_{12}>T_{13}\right)\right]\end{aligned}$$

and hence it is seen that $\hat{S}\left(\widetilde{\psi}\right)$ as previously defined would underestimate the target.

Note that

$$\begin{aligned}\mathbb{P}\left(T_{12}\le s,T_{23}\le t\right)&=\mathbb{E}\left[\widetilde{\psi}\left(T_{12},T_{12}+T_{23}\right)\right]\\&=\mathbb{E}\left[\widetilde{\psi}\left(T_{12},T_{12}+T_{23}\right)\frac{\mathbb{I}\left(T_{12}\le T_{13}\right)\mathbb{I}\left(T_{12}+T_{23}\le C\right)}{\mathbb{E}\left[\mathbb{I}\left(T_{12}\le T_{13}\right)\mathbb{I}\left(T_{12}+T_{23}\le C\right)\middle|T_{12},T_{23}\right]}\right]\end{aligned}$$

Now, in order to introduce an estimator for the above expectation, besides the assumption that the censoring variable is independent of the process we will need the following assumptions:

H1: T_{12} independent of T_{13};

H2: $\mathbb{I}\left(T_{12}\le T_{13}\right)$ independent of T_{23} conditionally on T_{12}.

Of course, H1 and H2 are also valid if we assume independence between T_{13} and $\left(T_{12},T_{23}\right)$. Using these assumptions we have,

$$\begin{aligned}\mathbb{E}\left[\mathbb{I}\left(T_{12}\le T_{13}\right)\mathbb{I}\left(T_{12}+T_{23}\le C\right)\middle|T_{12},T_{23}\right]&=\mathbb{E}\left[\mathbb{I}\left(T_{12}\le T_{13}\right)\middle|T_{12},T_{23}\right]\times\mathbb{E}\left[\mathbb{I}\left(T_{12}+T_{23}\le C\right)\middle|T_{12},T_{23}\right]\\&=\left[1-F_{13}\left(T_{12}^{-}\right)\right]\times\left[1-G\left(\left(T_{12}+T_{23}\right)^{-}\right)\right]\end{aligned}$$

On the other hand, $\mathbb{I}\left(T_{12} \le T_{13}\right)\mathbb{I}\left(T_{12}+T_{23} \le C\right) = \delta\rho$. Since, under $\delta\rho = 1$, we have $U = T_{12}$ and $\delta V = T_{23}$. Then we obtain

$$\mathbb{P}\left(T_{12} \le s, T_{23} \le t\right) = \mathbb{E}\left[\widetilde{\psi}\left(T_{12}, T_{12}+T_{23}\right)\right]$$
$$= \mathbb{E}\left[\frac{\widetilde{\psi}\left(U, U+\delta V\right)\delta\rho}{\left[1-F_{13}\left(U^{-}\right)\right]\times\left[1-G\left(\left(U+\delta V\right)^{-}\right)\right]}\right]$$

which can be estimated through the sample average:

$$\widehat{F_{12,23}}\left(s,t\right) = \frac{1}{n}\sum_{i=1}^{n}\frac{\widetilde{\psi}\left(U_i, U_i+\delta_i V_i\right)\delta_i\rho_i}{\left[1-\widehat{F}_{13}\left(U_i^{-}\right)\right]\left[1-\hat{G}\left(\left(U_i+\delta_i V_i\right)^{-}\right)\right]}.$$

The formal derivation of the consistency (and further asymptotics) of $\widehat{F_{12,23}}\left(s,t\right)$ is of much interest, and it will be considered in future research.

4.3.2. Extension to more complex multi-state models.

Methods presented in the previous sections regarding the scope of the illness-death model can be extended to more complex multi-state models. Here we will consider, merely as an example, the bivariate model and the progressive four-state model. Note that the progressive three-state model is a particular case of the illness-death model when $T_{13} = \infty$. For the mortality model, for the survival analysis, and under our approach we obtain Kaplan-Meier estimates.

Note however that if the state structure has a large number of states, then these methods rapidly became complicated.

4.3.2.1 The bivariate Model.

The bivariate model, depicted in Figure 2.8 (page 34), is the multi-state model for bivariate parallel data, with states 'both alive', 'individual 1 dead', 'individual 2 dead' and 'both dead'. This model is described in detail in Hougaard (2000).

For this model, the stochastic behaviour of the process is represented by a random variable $(T_{12}, T_{13}, T_{24}, T_{34})$. Assuming that C is independent of $(T_{12}, T_{13}, T_{24}, T_{34})$, we may now denote the sample information as

$$(U_i, \delta_i, \eta_i, V_i, \rho_i),\ 1 \le i \le n$$

which are assumed to be independent and identically distributed copies of $(U, \delta, \eta, V, \rho)$, where:

$U = \min(T_{12}, T_{13}, C)$ is the observed sojourn time in state 1;

$\delta = \mathbb{I}(T_{12} \le \min(T_{13}, C))$ is an indicator of whether a transition $1 \to 2$ occurs;

$\eta = \mathbb{I}(T_{13} \le C)$, so that $(1-\delta)\eta$ is an indicator of whether a transition $1 \to 3$ occurs;

$V = \min(T_{24}, C - T_{12})\mathbb{I}(T_{12} \le \min(T_{13}, C)) + \min(T_{34}, C - T_{13})\mathbb{I}(T_{13} < \min(T_{12}, C))$ is the observed sojourn time in the intermediate state (state 2 if $\delta = 1$, or state 3 if $\delta = 0$ and $\eta = 1$);

$\rho = \mathbb{I}(T_{24} \le C - T_{12})\mathbb{I}(T_{12} \le \min(T_{13}, C)) + \mathbb{I}(T_{34} \le C - T_{13})\mathbb{I}(T_{13} < \min(T_{12}, C))$, so that $\delta\rho$ is an indicator of whether a transition $2 \to 4$ occurs; and $(1-\delta)\eta\rho$ is an indicator of whether a transition $3 \to 4$ occurs;

Note that, under $\delta = 0$, information on (V, ρ) is still available provided that $\eta = 1$; while if $\delta = 1$, then observation of η is not possible.

Again, we consider $Z=\min\left(T_{12},T_{13}\right)$. Furthermore, we assume that H is the distribution function of Z, and that $\widehat{H}$ is the Kaplan-Meier product-limit estimator of H. As for the illness-death model, $\widehat{H}$ is based on the $\left(U,\delta+\left(1-\delta\right)\eta\right)$.

For the bivariate model, it is enough to consider the estimation of the transition probabilities $p_{11}\left(s,t\right)$, $p_{12}\left(s,t\right)$, $p_{22}\left(s,t\right)$, $p_{13}\left(s,t\right)$ and $p_{33}\left(s,t\right)$. All the others can be obtained from these since $p_{14}\left(s,t\right)=1-p_{11}\left(s,t\right)-p_{12}\left(s,t\right)-p_{13}\left(s,t\right)$, $p_{24}\left(s,t\right)=1-p_{22}\left(s,t\right)$ and $p_{34}\left(s,t\right)=1-p_{33}\left(s,t\right)$.

Therefore, we have

$$p_{11}\left(s,t\right)=\mathbb{P}\left(T_{12}>t,T_{13}>t\middle|T_{12}>s,T_{13}>s\right),$$

$$p_{12}\left(s,t\right)=\mathbb{P}\left(T_{12}\le t,T_{12}\le T_{13},T_{12}+T_{24}>t\middle|T_{12}>s,T_{13}>s\right),$$

$$p_{22}\left(s,t\right)=\mathbb{P}\left(T_{12}\le t,T_{12}\le T_{13},T_{12}+T_{24}>t\middle|T_{12}\le s,T_{12}\le T_{13},T_{12}+T_{24}>s\right),$$

$$p_{13}\left(s,t\right)=\mathbb{P}\left(T_{13}\le t,T_{12}>T_{13},T_{13}+T_{34}>t\middle|T_{12}>s,T_{13}>s\right),$$

and

$$p_{33}\left(s,t\right)=\mathbb{P}\left(T_{13}\le t,T_{12}>T_{13},T_{13}+T_{34}>t\middle|T_{13}\le s,T_{12}>T_{13},T_{13}+T_{34}>s\right).$$

These quantities are determined by the joint distribution of $\left(T_{12},T_{13},T_{24},T_{34}\right)$. Specifically, knowledge of the distribution of $Z=\min\left(T_{12},T_{13}\right)$ is enough for the recovery of $p_{11}\left(s,t\right)$,

$$p_{11}\left(s,t\right)=\frac{\mathbb{P}\left(Z>t\right)}{\mathbb{P}\left(Z>s\right)},$$

while expectations of type $S\left(\phi\right)=\mathbb{E}\left[\phi\left(T_{12},T_{12}+T_{24}\right)\mathbb{I}\left(T_{12}\le T_{13}\right)\right]$ arise when handling $p_{12}\left(s,t\right)$ $\left(\phi\left(u,v\right)=\phi_{s,t}\left(u,v\right)=\mathbb{I}\left(s<u\le t,v>t\right)\right)$,

$$p_{12}(s,t)=\frac{S(\phi_{s,t})}{\mathbb{P}(Z>s)},$$

and $p_{22}(s,t)$ $\left(\phi(u,v)=\tilde{\phi}_{s,t}(u,v)=\mathbb{I}(u\le s, v>t)\right)$,

$$p_{22}(s,t)=\frac{S(\tilde{\phi}_{s,t})}{S(\tilde{\phi}_{s,s})}.$$

In the same spirit, expectations of type $D(\phi)=\mathbb{E}\left[\phi(T_{13},T_{13}+T_{34})\mathbb{I}(T_{12}>T_{13})\right]$ are needed for $p_{13}(s,t)$ $\left(\phi(u,v)=\phi_{s,t}(u,v)=\mathbb{I}(s<u\le t, v>t)\right)$

$$p_{13}(s,t)=\frac{D(\phi_{s,t})}{\mathbb{P}(Z>s)};$$

and for $p_{33}(s,t)$,

$$p_{33}(s,t)=\frac{D(\tilde{\phi}_{s,t})}{D(\tilde{\phi}_{s,s})}.$$

Using Lemma 2, expectations $S(\phi)$, can be estimated using sample averages [4.7], yielding the following estimates for the transition probabilities,

$$\widehat{p_{11}}(s,t)=\frac{1-\widehat{H}(t)}{1-\widehat{H}(s)},$$

$$\widehat{p_{12}}(s,t)=\frac{1}{n\left(1-\widehat{H}(s)\right)}\sum_{i=1}^{n}\frac{\mathbb{I}(s<U_i\le t)\mathbb{I}(U_i+V_i>t)\delta_i\rho_i}{1-\hat{G}\left((U_i+V_i)^-\right)},$$

and

$$\widehat{p_{22}}(s,t)=\frac{\sum_{i=1}^{n}\frac{\mathbb{I}(U_i\le s)\mathbb{I}(U_i+V_i>t)\delta_i\rho_i}{1-\hat{G}\left((U_i+V_i)^-\right)}}{\sum_{i=1}^{n}\frac{\mathbb{I}(U_i\le s)\mathbb{I}(U_i+V_i>s)\delta_i\rho_i}{1-\hat{G}\left((U_i+V_i)^-\right)}}.$$

Note that the Kaplan-Meier product-limit estimate of *C*, denoted by $\widehat{G}$, is now based on $(U+V, 1-\nu)$ with $\nu = \delta\rho + (1-\delta)\eta\rho$.

For the expectations $D(\phi)$, we need the following Lemma:

Lemma 4. For each function ϕ, we have

$$D(\phi) = \mathbb{E}\left[\frac{\phi(U, U+V)(1-\delta)\eta\rho}{1-G\left((U+V)^{-}\right)}\right].$$

Proof.

Analogous to Lemma 2. □

Then, estimators for the rest of the transition probabilities may be introduced:

$$\widehat{p_{13}}(s,t) = \frac{1}{n\left(1-\widehat{H}(s)\right)} \sum_{i=1}^{n} \frac{\mathbb{I}(s < U_i \le t)\mathbb{I}(U_i + V_i > t)(1-\delta_i)\eta_i\rho_i}{1-\hat{G}\left((U_i+V_i)^{-}\right)},$$

and

$$\widehat{p_{33}}(s,t) = \frac{\sum_{i=1}^{n} \frac{\mathbb{I}(U_i \le s)\mathbb{I}(U_i + V_i > t)(1-\delta_i)\eta_i\rho_i}{1-\hat{G}\left((U_i+V_i)^{-}\right)}}{\sum_{i=1}^{n} \frac{\mathbb{I}(U_i \le s)\mathbb{I}(U_i + V_i > s)(1-\delta_i)\eta_i\rho_i}{1-\hat{G}\left((U_i+V_i)^{-}\right)}}.$$

Estimation of other parameters under the bivariate model can be introduced following ideas similar to those discussed for the illness-death model. The formal derivation of the asymptotics pertaining to the introduced estimators is left for future research.

4.3.2.2 The progressive four-state model.

The four-state progressive model is fully characterized by three transition intensities. We therefore represent the stochastic behaviour of the process by a random variable (T_{12}, T_{23}, T_{34}), assuming that C is independent of the process. We denote the sample information as

$$(U_i, \delta_{12i}, \delta_{12i}.V_i, \delta_{12i}\delta_{23i}W, \delta_{12i}\delta_{23i}\rho_i),\ 1 \le i \le n$$

which are assumed to be independent and identically distributed copies of $(U, \delta_{12}, \delta_{12}.V, \delta_{12}\delta_{23}W, \delta_{12}\delta_{23}\rho)$, where

$U = \min(T_{12}, C)$ is the observed sojourn time in state 1;

$\delta_{12} = \mathbb{I}(T_{12} \le C)$ is an indicator of whether a transition $1 \to 2$ occurs;

$V = \min(T_{23}, C - T_{12})$ is the observed sojourn time in state 2;

$\delta_{23} = \mathbb{I}(T_{23} \le C - T_{12})$, so that $\delta_{12}\delta_{23}$ is an indicator of whether a transition $2 \to 3$ occurs;

$W = \min(T_{34}, C - T_{23})$ is the sojourn time in state 3;

$\delta_{34} = \mathbb{I}(T_{34} \le C - T_{12} - T_{23})$, so that $\delta_{12}\delta_{23}\delta_{34}$ is an indicator of whether a transition $3 \to 4$ occurs.

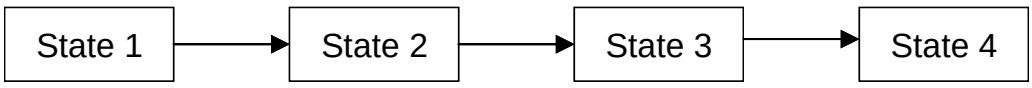


Figure 4.4. The four-state progressive model.

For the four-state progressive model, it is enough to consider the estimation of the following transition probabilities:

$$p_{11}(s,t)=\mathbb{P}\left(T_{12}>t\middle|T_{12}>s\right),$$

$$p_{12}(s,t)=\mathbb{P}\left(T_{12}\le t,T_{12}+T_{23}>t\middle|T_{12}>s\right),$$

$$p_{13}(s,t)=\mathbb{P}\left(T_{12}+T_{23}\le t,T_{12}+T_{23}+T_{34}>t\middle|T_{12}>s\right),$$

$$p_{22}(s,t)=\mathbb{P}\left(T_{12}\le t,T_{12}+T_{23}>t\middle|T_{12}\le s,T_{12}+T_{23}>s\right),$$

$$p_{23}(s,t)=\mathbb{P}\left(T_{12}+T_{23}\le t,T_{12}+T_{23}+T_{34}>t\middle|T_{12}\le s,T_{12}+T_{23}>s\right),$$

and

$$p_{34}(s,t)=\mathbb{P}\left(T_{12}+T_{23}+T_{34}\le t\middle|T_{12}+T_{23}\le s,T_{12}+T_{23}+T_{34}>s\right).$$

Let F_{12} be the distribution function of T_{12}, and let $\widehat{F_{12}}$ denote the Kaplan-Meier product-limit estimator of F_{12}. Then, we can introduce the estimator

$$\widehat{p_{11}}(s,t)=\frac{1-\widehat{F_{12}}(t)}{1-\widehat{F_{12}}(s)}.$$

Estimators for the rest of the transition probabilities can be obtained as for the illness-death model. Certainly,

$$\widehat{p_{12}}(s,t)=\frac{1}{n\left(1-\widehat{F_{12}}(s)\right)}\sum_{i=1}^{n}\frac{\mathbb{I}\left(s<U_i\le t\right)\mathbb{I}\left(U_i+\delta_{12i}V_i>t\right)\delta_{12i}\delta_{23i}}{1-\hat{G}\left(\left(U_i+\delta_{12i}V_i\right)^{-}\right)},$$

$$\widehat{p_{13}}(s,t)=\frac{1}{n\left(1-\widehat{F_{12}}(s)\right)}\sum_{i=1}^{n}\frac{\mathbb{I}\left(U_i>s\right)\mathbb{I}\left(U_i+\delta_{12i}V_i\le t\right)\mathbb{I}\left(U_i+\delta_{12i}V_i+\delta_{12i}\delta_{23i}W_i>t\right)\delta_{12i}\delta_{23i}\delta_{34i}}{1-\hat{G}\left(\left(U_i+\delta_{12i}V_i+\delta_{12i}\delta_{23i}W_i\right)^{-}\right)}$$

$$\widehat{p_{22}}(s,t)=\frac{\sum_{i=1}^{n}\frac{\mathbb{I}(U_i\le s)\mathbb{I}(U_i+\delta_{12i}V_i>t)\delta_{12i}\delta_{23i}}{1-\hat{G}\left((U_i+\delta_{12i}V_i)^-\right)}}{\sum_{i=1}^{n}\frac{\mathbb{I}(U_i\le s)\mathbb{I}(U_i+\delta_{12i}V_i>s)\delta_{12i}\delta_{23i}}{1-\hat{G}\left((U_i+\delta_{12i}V_i)^-\right)}},$$

$$\widehat{p_{23}}(s,t)=\frac{\sum_{i=1}^{n}\frac{\mathbb{I}(U_i\le s)\mathbb{I}(U_i+\delta_{12i}V_i\le t)\mathbb{I}(U_i+\delta_{12i}V_i+\delta_{12i}\delta_{23i}W_i>t)\delta_{12i}\delta_{23i}\delta_{34i}}{1-\hat{G}\left((U_i+\delta_{12i}V_i+\delta_{12i}\delta_{23i}W_i)^-\right)}}{\sum_{i=1}^{n}\frac{\mathbb{I}(U_i\le s)\mathbb{I}(U_i+\delta_{12i}V_i>s)\delta_{12i}\delta_{23i}}{1-\hat{G}\left((U_i+\delta_{12i}V_i)^-\right)}},$$

and

$$\widehat{p_{34}}(s,t)=\frac{\sum_{i=1}^{n}\frac{\mathbb{I}(U_i+\delta_{12i}V_i\le s)\mathbb{I}(s<U_i+\delta_{12i}V_i+\delta_{12i}\delta_{23i}W_i\le t)\delta_{12i}\delta_{23i}\delta_{34i}}{1-\hat{G}\left((U_i+\delta_{12i}V_i+\delta_{12i}\delta_{23i}W_i)^-\right)}}{\sum_{i=1}^{n}\frac{\mathbb{I}(U_i+\delta_{12i}V_i\le s)\mathbb{I}(U_i+\delta_{12i}V_i+\delta_{12i}\delta_{23i}W_i>s)\delta_{12i}\delta_{23i}\delta_{34i}}{1-\hat{G}\left((U_i+\delta_{12i}V_i+\delta_{12i}\delta_{23i}W_i)^-\right)}}.$$

where $\widehat{G}$ is the Kaplan-Meier product limit estimate based on $\left(U+\delta_{12}V+\delta_{12}\delta_{23}W,1-\nu\right)$ with $\nu=\delta_{12}\delta_{23}\delta_{34}$.

Investigation of the convergence properties of these estimators is not covered in this thesis. Finally, note that like in the previous sections different estimators for the transition probabilities could have been considered.

Chapter 5

Software

5.1. Introduction.

While the time-dependent Cox model (TDCM) can be fitted through all the major statistical packages, multi-state models need specialized software, being most of the programs written in FORTRAN, R or SAS. Marshall and Jones (1995) have developed a FORTRAN program called MARKOV for fitting multi-state Markov models with constant transition intensities and covariates. Later, Alioum and Commenges (2001) presented a new computer program, called MKVPCI, which extends MARKOV by allowing piecewise-constant intensities with different values at three time intervals at most. More recently, Jackson and Sharples (2002) developed the R package **msm**, implementing several functions for fitting continuous-time Markov multi-state models to categorical processes observed at arbitrary times. Presently, Paes and Lima (2004) developed a SAS macro, called PTRANSIT, for estimating transition probabilities in semi-parametric models for recurrent events. Hui-Min et al. (2004) have developed a SAS macro for estimating transition parameters in non-homogeneous (Weibull distributions, log-logistic, etc.) *k*-state progressive Markov models.

Although multi-state models may often be preferable to simple regression models, these models present, however, some limitations in practice. Most of the existent software for multi-state models assumes that the process is Markov and time-homogeneous, which can be very restrictive. In some clinical studies, a model with the Markov assumption may be appropriate, while in others the semi-Markov is preferable. In some instances, a homogeneous model will be satisfactory, while in others not. Furthermore, sufficient standard software for fitting non-homogeneous models or for semi-Markov (or non-Markov) models is not quite available yet. Possible comparisons between different multi-state models are rather difficult because each of the available

programs requires its own input data. In addition, most of the programs available only provide regression parameters estimates and do not supply graphical output for survival estimates and for transition probabilities estimates.

In this chapter, we present **tdc.surv**, a user-friendly R library for the analysis of survival data with time-dependent covariates. Specifically, the new software may be used to fit not only the time-dependent Cox model but also five different multi-state models including various transient states and one absorbing state. Advantages of this software include the same data input for fitting the different models while providing the corresponding numerical and graphical outputs obtained. In this way, users may easily analyze the results offered by the various models in order to compare them and make decisions accordingly.

Our software can be used to fit five different multi-state models in continuous time, namely: (a) Cox Markov model (CMM); (b) Cox semi-Markov model (CSMM); (c) Homogeneous Markov model (HMM); (d) Non-homogeneous piecewise model (NHM); and (e) Non-parametric Markov model (NPM).

After discussing the different proposed models in Chapter 3 and Chapter 4, we now focus on the description of **tdc.surv**, explaining how the data should be introduced and how to fit the different models. Finally, using Stanford Heart Transplant data and Stomach cancer data (see section 1.3) we provide some numerical results and some graphical outputs to illustrate the above methodologies.

5.2. Software description.

We have developed a library **tdc.surv** in R language to implement some of the proposed approaches presented in Chapter 3 and 4.

We consider the situation of having one categorical time-dependent covariate. Without loss of generality, we will assume that this covariate represents a treatment intervention (change) at a specified time. Continuous time-dependent covariates can be also accommodated in this framework, by transforming these variables into categorical ones.

By considering a total of T possible treatments over the whole study period, we can use the multi-state model of Figure 5.1, including a total of N states: an *initial* state (e.g., diagnosis of illness), a number of $T = N\text{-}2$ *transient* (intermediate) states, and an *absorbing* state representing a terminal event of interest (e.g., death). In the particular case of having only one transient state (that is, when N=3), we obtain the *illness-death* model.

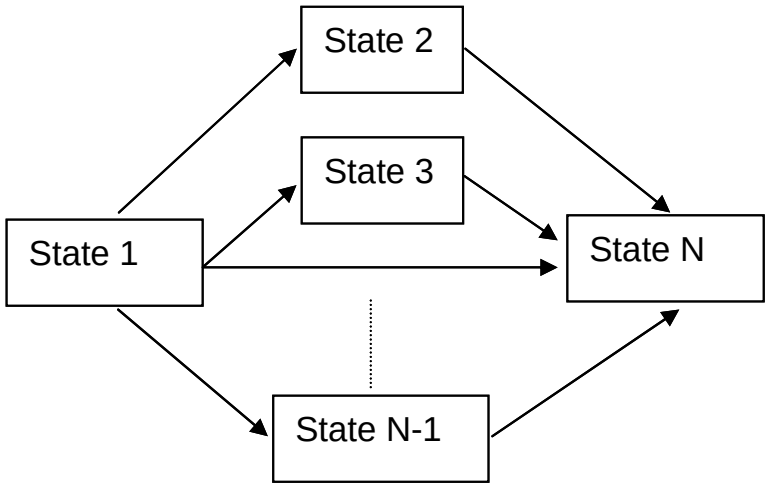


Figure 5.1: Extended illness-death model.

Library **tdc.surv** requires only one input data file. Records in the data file must contain the following variables: *id, start, stop, event, treat, covariate1, covariate2*,... and so on.

The structure of the data input is as follows: each patient is identified by *id*. Individuals without change in the time-dependent covariate are represented by only one line of data, whereas patients with a change in the time-dependent covariate must be represented by two lines. For these patients, the first line represents the time period until change of treatment, and then variable *treat* takes the value zero; the second line represents the time period that passes from the change to the end of the follow-up, and *treat* takes a new value (1 for the first alternative treatment, 2 for the second alternative treatment, and so on). The remaining covariates (*covariate1*, *covariate2*,…) are the same for the two lines. For each row, variables *start* and *stop* mark the time interval (*start*, *stop*] for the data, while *event* is an indicator variable taking value 1 if there was a death at time *stop*, and 0 otherwise.

The input data file contains the available information about these patients. For illustration purposes, in Table 5.1 we present a sample data for the Heart Transplantation data.

Table 5.1. Input data file for the Stanford Heart Transplantation data.

id	start	stop	event	treat	age	year	surgery
1	0	50	1	0	-17.15537	0.12320	0
2	0	6	1	0	3.83575	0.25462	0
3	0	1	0	0	6.29706	0.26557	0
3	1	16	1	1	6.29706	0.26557	0
4	0	36	0	0	-7.73717	0.49008	0
4	36	39	1	1	-7.73717	0.49008	0
:	:	:	:	:	:	:	:
15	0	1	1	0	5.81519	0.99110	1
:	:	:	:	:	:	:	:

In R language, programming is based on objects and computations are basically functions that are specialized to carry out specific calculations. Our library is composed of 11 functions and can fit all the proposed models by means of the following command line:

tdc.surv<-function(mydata, ncov, formula, fixedpars, model, cut, prevalences=FALSE, writedata=FALSE,covmat=FALSE, graphcov=0, surv.plot=FALSE, plot.trans=FALSE)
where:

mydata: data input as described above.

ncov: number of covariates included in the models.

formula: vector of length ncov, indicating the column position of the covariates in the data file input.

fixedpars: vector of indices of parameters whose values will be fixed at their initial values during the optimisation. By default, fixedpars=NULL.

model: vector of length 6, indicating the models to be fitted. Positions 1 to 6, correspond to TDCM, CMM, HMM, NHM, CSMM and NPM, respectively. By default, model=c(1,1,1,0,0,0). Elements of this vector indicate that only models TDCM, CMM and HMM, will be fitted.

cut: selected cutpoint for fitting the NHM. By default, cut=0.

prevalences: provides a rough indication of the goodness-of-fit of a multi-state model. By default, prevalences=FALSE.

writedata: provides generated data for the different models. By default, writedata=FALSE.

covmat: provides the variance-covariance matrices. By default, covmat=FALSE.

graphcov: displays the smoothed hazard ratio for continuous covariates. By default, graphcov=FALSE.

surv.plot: graphical output for survival estimates. By default, surv.plot=FALSE.

plot.trans: graphical output for transition probabilities. By default, plot.trans=FALSE.

Only mydata, ncov and formula are required arguments for **tdc.surv**.

Numerical results are printed on the screen. Among other results, the **tdc.surv** output includes parameter estimations with standard errors for the covariates (for models TDCM, CMM, CSMM, HMM, and NHM), transition rates (for HMM, NHM, and NPM), hazard rates with corresponding 95 percent confidence intervals (for TDCM, CMM, CSMM, HMM, and NHM), and prevalence tables (for HMM). Graphical output includes: survival estimates; transition probabilities estimates (for HMM, NHM and NPM); and the smoothed hazard ratio estimate for continuous covariates (for TDCM, CMM, and CSMM**).**

In case of HMM and NHM, Wald's test can be performed to check for differences in the mortality transitions or to assess the effect of a given covariate for two or more transitions.

To illustrate this, we have used **tdc.surv** to fit survival data from the Stanford Heart Transplant study and data from the Stomach Cancer study.

The fit for all the proposed models for the Stanford data can be obtained easily by using the following command,

tdc.surv(stanford, ncov=3, formula=c(6,7,8), model=c(1,1,1,1,1,1), cut=90)

This command line yields a large output with numerical results for all models. We have chosen to present some of the results for the first three models (i.e., TDCM in Table 5.2, HMM in Table 5.3 and CMM in Table 5.4). Results for all the output are shown in Appendix B. We can also obtain some graphical outputs, such as the estimates of survival or the smooth hazard ratio for age at acceptance (continuous covariate). This

output is shown in Figure 5.2. To obtain this graphical output, we use the following input command:

```
tdc.surv(stanford, ncov=3, formula=c(6,7,8), model=c(0,1,0,0,0,0), graphcov=1, surv.plot=T)
```

Table 5.2. Sample of the output for the time-dependent Cox regression model. Stanford Heart Transplantation data.

```
**********************************************************
     ********  Time-dependent Cox regression model  ********
**********************************************************
n=172
                coef      se(coef)        z       p-value
age          -0.0271      0.0134         2.02     0.043
year         -0.1463      0.0704        -2.08     0.038
surgery      -0.6376      0.3670        -1.74     0.082

                                95% Confidence Interval
             exp(coef)          lower     upper
age          1.027              1.001     1.055
year         0.864              0.753     0.992
surgery      0.529              0.257     1.085

Rsquare= 0.084  (max possible= 0.969)
Likelihood ratio test = 15.1 on 3 df, p-value =0.00172
Wald test = 14.5 on 3 df, p-value =0.00230
Score (logrank) test = 15.0 on 3 df, p-value =0.00179
-2*Log-likelihood: 581.1323
```

Table 5.3. Sample of the output for the homogeneous Markov model. Stanford Heart Transplantation data.

```
**************************************************
 ** MULTI-STATE HOMOGENEOUS MARKOV MODEL **
**************************************************
**** UNDERGOING TRANSITIONS ****
      To
from    1    2    3
  1     4   69   30
  2     0   24   45
*** convergence ***
-2*Log-likelihood: 1727.033
***** Estimated coefficients ****
Baseline:
              Stage 1        Stage 2        Stage 3
Stage 1  -0.01907839  -0.013679774   0.005398618
Stage 2  -0.00000000  -0.001761965   0.001761965
Stage 3  -0.00000000  -0.000000000   0.000000000

For age:
              Stage 1        Stage 2        Stage 3
Stage 1   0.00000000   0.06581947   0.05447583
Stage 2   0.00000000   0.00000000   0.07356488
Stage 3   0.00000000   0.00000000   0.00000000

                                  95% Confidence Interval
                        HR        Lower        Upper
Stage 1 - Stage 2  1.068034     1.039146     1.097725
Stage 1 - Stage 3  1.055987     1.019930     1.093318
Stage 2 - Stage 3  1.076338     1.030214     1.124528
```

Table 5.4. Sample of the output for the Cox Markov model. Stanford Heart Transplantation data.

```
*****************************************************************************
 ** COX MARKOV MODEL. SEPARATED FITTED COX MODELS FOR EACH TRANSITION **
*****************************************************************************
                          ***FROM STATE 1 TO STATE 2***
n=103
                coef      se(coef)        z          p-value
age           0.0311      0.014          2.226      0.026
year          0.0008      0.070          0.011      0.990
surgery       0.0473      0.315          0.150      0.880
                                  95% Confidence Interval
              exp(coef)        Lower      Upper
age           1.030            1.004      1.060
year          1.000            0.873      1.150
surgery       1.050            0.565      1.940

Rsquare= 0.054  (max possible= 0.993)
Likelihood ratio test = 5.77 on 3 df, p-value =0.123
Wald test = 5 on 3 df, p-value =0.172
Score (logrank) test = 5.06 on 3 df, p-value =0.167
-2*Log-likelihood: 509.5638

                          ***FROM STATE 1 TO STATE  3***
n=103
                coef      se(coef)        z          p-value
age           0.0198      0.018          1.094      0.270
year          -0.2833     0.111          -2.553     0.011
surgery       -0.2288     0.636          -0.360     0.720
                                  95% Confidence Interval
              exp(coef)        Lower      Upper
age           1.020            0.984      1.057
year          0.753            0.606      0.936
surgery       0.796            0.229      2.768

Rsquare= 0.08  (max possible= 0.886)
Likelihood ratio test = 8.62 on 3 df, p-value=0.0347
Wald test = 8.19 on 3 df, p-value=0.0422
Score (logrank) test = 8.67 on 3 df, p-value =0.034
-2*Log-likelihood: 214.9848

                          ***FROM STATE 2 TO STATE 3***
n=69
                coef      se(coef)        z          p-value
age           0.0496      0.021          2.318      0.020
year          -0.0230     0.097          -0.238     0.810
surgery       -0.8165     0.455          -1.795     0.073
                                  95% Confidence Interval
              exp(coef)        Lower      Upper
age           1.051            1.008      1.100
year          0.977            0.808      1.180
surgery       0.442            0.181      1.080

Rsquare= 0.151  (max possible= 0.987)
Likelihood ratio test = 11.3 on 3 df, p-value =0.0102
Wald test = 10.2 on 3 df, p-value =0.0167
Score (logrank) test = 10.7 on 3 df, p-value =0.0137
-2*Log-likelihood: 290.1922
```

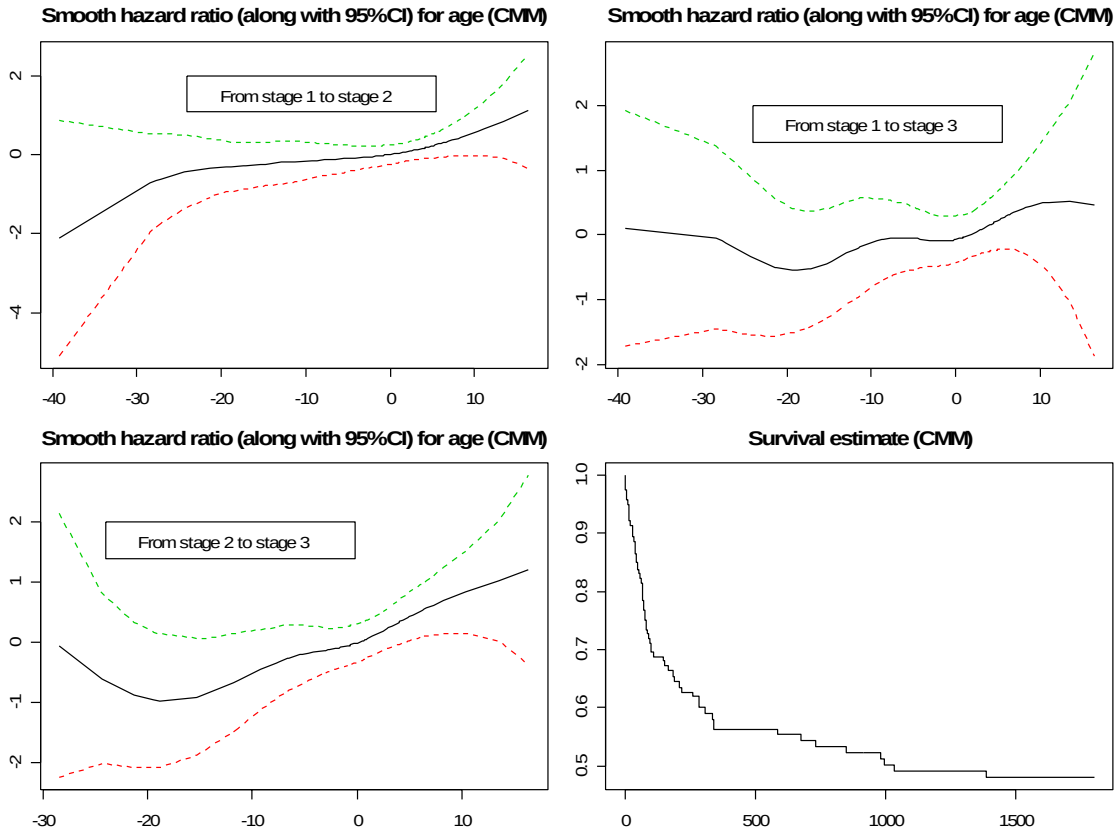


Figure 5.2. Graphical output for the Cox Markov model. Stanford Heart Transplantation data.

We should notice that, although the Stanford dataset requires a multi-state model with only two transient states, the software we offer can be used in more general situations, which can be formulated by the model represented in Figure 5.1. For this purpose, we shall use the stomach cancer data. In this study, patients may pass from the initial state (surgical intervention time) through one of two mutually exclusive states (‘metastases’ and ‘recurrence’) to an absorbing state (death). We therefore may use the multi-state model of Figure 5.3.

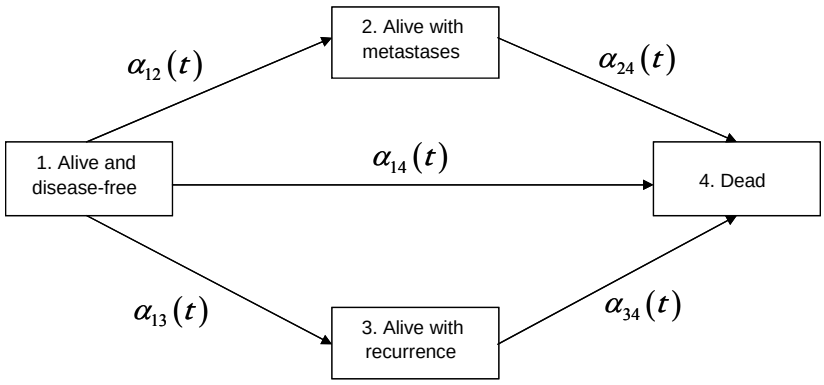


Figure 5.3. Multi-state model for Stomach cancer data.

For this database we present a sample of the output from the CMM (Table 5.5) and HMM (Table 5.6) approach. The complete output is shown in Appendix B. We also present some graphical outputs (Figure 5.4), such as the estimates of survival or the transition probability estimates.

Table 5.5. Sample of the output for the Cox Markov model. Stomach Cancer data.

```
*************************************************************************
***** COX MARKOV MODEL. SEPARATED FITTED COX MODELS FOR EACH TRANSITION *****
*************************************************************************


***** FROM STATE 1 TO STATE 2  *****
n= 345

          coef      exp(coef)  se(coef)   z        p
Sex       0.02482   1.03       0.2791     0.089    0.93
Age       0.00321   1.00       0.0107     0.300    0.76

          exp(coef) exp(-coef) lower.95   upper.95
Sex       1.03      0.975      0.593      1.77
Age       1.00      0.997      0.982      1.02

Rsquare= 0   (max possible= 0.807 )
Likelihood ratio test= 0.09  on 2 df,   p=0.955
Wald test            = 0.09  on 2 df,   p=0.955
Score (logrank) test = 0.09  on 2 df,   p=0.955

-2*Log-likelihood: 567.5803


***** FROM STATE 1 TO STATE 3  *****
n= 345

          coef      exp(coef)  se(coef)   z        p
Sex       0.6585    1.93       0.3391     1.94     0.052
Age       0.0169    1.02       0.0127     1.33     0.180


          exp(coef) exp(-coef) lower.95   upper.95
Sex       1.93      0.518      0.994      3.76
Age       1.02      0.983      0.992      1.04

Rsquare= 0.015   (max possible= 0.716 )
Likelihood ratio test= 5.21  on 2 df,   p=0.074
Wald test            = 4.9  on 2 df,   p=0.0862
Score (logrank) test = 5   on   2 df,   p=0.0822

-2*Log-likelihood: 428.8344
```

Table 5.6. Output for the homogeneous Markov model. Stomach Cancer data.

```
*********************************************************************************
            ***** MULTI-STATE HOMOGENEOUS MARKOV MODEL *****
*********************************************************************************

***** Estimated coefficients *****
$baseline
              Stage 1    Stage 2     Stage 3    Stage 4
Stage 1     -0.00066    0.00010   7.587e-05    0.00048
Stage 2      0.00000   -0.00189     0.00000    0.00189
Stage 3      0.00000    0.00000   -1.768e-03   0.00176
Stage 4      0.00000    0.00000     0.00000    0.00000

 $sex
                        HR        L95        U95
Stage 1 - Stage 2  1.06251    0.61473    1.83645
Stage 1 - Stage 3  2.01196    1.03213    3.92198
Stage 1 - Stage 4  0.97368    0.75767    1.25128
Stage 2 - Stage 4  0.41117    0.21701    0.77905
Stage 3 - Stage 4  1.22044    0.55678    2.67513

$age
                        HR        L95        U95
Stage 1 - Stage 2  1.01033    0.98880    1.03233
Stage 1 - Stage 3  1.02516    0.99925    1.05174
Stage 1 - Stage 4  1.01400    1.00398    1.02411
Stage 2 - Stage 4  1.00520    0.97407    1.03732
Stage 3 - Stage 4  1.00035    0.97341    1.02804
```

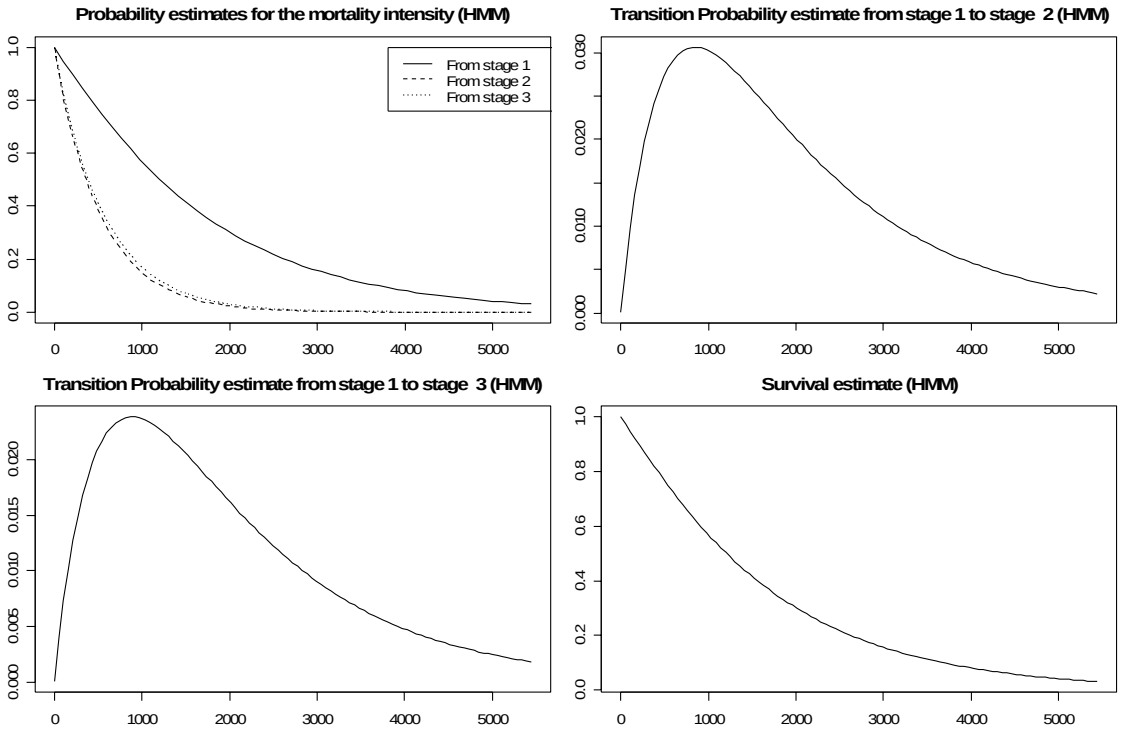


Figure 5.4. Graphical output for the homogeneous Markov model. Stomach Cancer data.

Chapter 6

Concluding remarks and future research

6.1 Concluding remarks

We have presented and discussed the application of event history analysis using multi-state models for several medical applications. The work described in this thesis contains several contributions to the development of multi-state models for survival data. First, we have presented a comprehensive survey of multi-state models. Secondly, this work contains a comparison of different multi-state models, including in its comparisons the commonly used Cox model. Advantages and disadvantages for both methods have been discussed. Possible advantages of using multi-state models are illustrated through simulation studies, addressing some issues that cannot be fully explored through the Cox model. Thirdly, new software has been developed to fit both Markov and semi-Markov models. These models have been applied to data from clinical studies, providing some guidance about the use of these methodologies for studying the evolution of the disease (e.g. hypothesis testing procedures, methods for model checking, goodness-of-fit, etc.). Finally, the main methodological contribution of this thesis has been the proposal of a new approach for the inference in multi-state modelling which overcomes the Markov assumption problem. We believe that this is the first non-parametric modelling approach completely free from the Markov assumption. This new approach has been evaluated through a simulation study, and applied to data from the PROVA clinical trial. Some asymptotic results were obtained for the proposed estimators.

6.2 Future research

As mentioned in Chapter 4, some asymptotic results for the introduced estimators are still missing. This constitutes a topic for our future research. Besides, we consider now some open questions that will be investigated in the future.

The inclusion of covariates in the estimation methods proposed for the transition probabilities. The estimation methods developed in Chapter 4 provide important measures for making long-term predictions. Since these quantities do not depend on patients' characteristics, we are assuming that they are the same for all subjects. In practical situations, however, it might be interesting to use the individual characteristics to obtain estimates of the transition probabilities conditionally on covariates. The issue of how to incorporate covariates in the transition probabilities has already been considered for the Aalen-Johansen estimators. For example, Helms et al., (2004) use pseudo-values to obtain a relationship between the Aalen-Johansen estimator and the covariates. These authors use the approach suggested by Andersen et al. (2003) for a direct estimation of the transition probabilities. This problem has also been studied previously using a Cox proportional hazards model (Andersen et al., 1991; Aalen et al., 2001). We believe that similar procedures can also be used for our estimators.

Non-Markovian estimation methods proposed for the transition probabilities. In Chapter 4, we introduced two different possibilities for the estimation of transition probabilities, denoted by $\widehat{p_{hj}}(s,t)$ and $\widehat{p^{*}_{hj}}(s,t)$. Remaining is the question of the relative efficiency of the different possible estimates.

Testing the Markov assumption. In multi-state modelling, the Markov assumption is commonly used. The simplest Markov assumption is that future evolution only depends on the current state at time *t*; in other words, the history of the process is summarized by the state occupied at time *t*.

Traditionally, the Markov assumption may be checked, among others, by including covariates in the modelling process. Such methods have been considered by Kay (1986). For simplicity, we consider the illness-death model, for which the Markov assumption is only present on transition from the disease state into the death state. We consider the null hypothesis that the data come from a Markov process:

H_0 : The Process is Markov ,

against the general alternative,

H_1 : The Process is not Markov .

Then we must show that the time spent in the healthy state (past) is not important on the transition from the disease state to death. For doing that, let $Z =$ "time spent in state 1", and t the current time. Fitting a model $\alpha_{23}(t) = \alpha_{230}(t)\exp\{\beta Z\}$, we now need to test $\beta = 0$, i.e., we have the null hypothesis,

H_0 : β=0 ,

against the general alternative,

H_1 : $\beta \neq 0$.

This would assess the assumption that the transition rate from the disease state into death is unaffected by the time spent in the previous state.

Our aim is to use the estimation methods developed in Chapter 4 for testing the Markov assumption. The purpose is to compare the estimated transition probabilities $p_{22}(s,t)$ under the Markov assumption (Aalen-Johansen estimator, see [3.7]),

expressed by $\widehat{p_{22}^{AJ}}(s,t)$, and through non-Markov approach (see [4.9]), denoted by $\widehat{p_{22}^{NM}}(s,t)$. We therefore consider the null hypothesis that the data come from a Markov process. Under the null hypothesis, it is expected that the difference between the two quantities, $\widehat{p_{22}^{AJ}}(s,t)$ and $\widehat{p_{22}^{NM}}(s,t)$ are, in some sense, smaller. Here, one main difficulty arises from the fact that those transition probabilities depend on both times *s* and *t*.

Software implementation. Multi-state non-Markov models have not been widely applied in medicine. Methodological contributions for analysing such models are still scarce in the literature. One main difficulty is the unavailability of software. Work is in progress to provide software for the estimation methods developed in Chapter 4.

Application to other datasets. It would also be interesting to apply some of the models described in this thesis to other datasets, particularly ones with non-linear covariate effect on the transition intensities. We plan to consider the application of these methods to a breast cancer dataset from Galicia and to a dataset on AIDS ("Registro Galego de Sida – Servicio Galego de Saúde. Junta da Galiza.").

A. Appendix of Chapter 4

A.1 Analytic expressions for the transition probabilities in the various setting schemes presented in section 4.3.1.3.

Here we present the analytic expressions for the transition probabilities, $p_{22}(s,t)$ in the various setting schemes presented in section 4.3.1.3. In all settings, T_{12} and T_{13} are exponentially distributed with rate parameter α_{12} and α_{13} respectively, assuming that $F_{12}(\cdot)$ and $F_{13}(\cdot)$ are the distribution functions of T_{12} and T_{13}, respectively.

Assuming that (T_{12},T_{13},T_{23}) are mutually independent times, the expression for the transition probabilities, $p_{22}(s,t)$ is

$$
\begin{aligned}
p_{22}(s,t) &= \frac{\mathbb{P}(T_{12}+T_{23}>t, T_{12}\le T_{13}, T_{12}\le s)}{\mathbb{P}(T_{12}+T_{23}>s, T_{12}\le T_{13}, T_{12}\le s)} \\
&= \frac{\int_0^s \mathbb{P}(T_{23}>t-u, T_{13}\ge u \,|\, T_{12}=u)\times dF_{12}(u)}{\int_0^s \mathbb{P}(T_{23}>s-u, T_{13}\ge u \,|\, T_{12}=u)\times dF_{12}(u)} \\
&= \frac{\int_0^s \mathbb{P}(T_{23}>t-u)\times \mathbb{P}(T_{13}\ge u)\times dF_{12}(u)}{\int_0^s \mathbb{P}(T_{23}>s-u)\times \mathbb{P}(T_{13}\ge u)\times dF_{12}(u)} \\
&= \frac{\int_0^s \mathbb{P}(T_{23}>t-u)\times \mathbb{P}(T_{13}\ge u)\times \mathbb{P}(T_{12}\ge u)\times dA_{12}(u)}{\int_0^s \mathbb{P}(T_{23}>s-u)\times \mathbb{P}(T_{13}\ge u)\times \mathbb{P}(T_{12}\ge u)\times dA_{12}(u)}
\end{aligned}
$$

where $A_{12}(\cdot)$ is the cumulative hazard of T_{12}.

Lemma. Assume that (T_{12},T_{13},T_{23}) are mutually independent times. Then, the process is Markovian if and only if T_{23} is exponentially distributed.

Proof.

Assume that the process is Markovian. Then, because equation [4.4] is valid for all $s_0 \in (0, s]$, and since (T_{12}, T_{13}, T_{23}) are mutually independent times, we obtain,

$$\begin{aligned}
\mathbb{P}(T_{23} > t \,|\, T_{23} > s) &= \mathbb{P}(T_{12} + T_{23} > t \,|\, T_{12} + T_{23} > s, T_{12} = 0, T_{12} \le T_{13}) \\
&= \mathbb{P}(T_{12} + T_{23} > t \,|\, T_{12} + T_{23} > s, T_{12} = s, T_{12} \le T_{13}) \\
&= \mathbb{P}(T_{23} > t - s \,|\, T_{23} > 0, T_{12} = s, T_{12} \le T_{13}) \\
&= \mathbb{P}(T_{23} > t - s)
\end{aligned}$$

This lack of memory is referred to as the Markovian property from the exponential distribution, the unique distribution function satisfying this property.

On the other hand, if we assume that T_{23} is exponentially distributed, then for any s_0 with $0 < s_0 \le s$,

$$\begin{aligned}
&\mathbb{P}(T_{12} + T_{23} > t \,|\, T_{12} + T_{23} > s, T_{12} \le s, T_{12} \le T_{13}) = \\
&\qquad = \frac{\mathbb{P}(T_{12} + T_{23} > t, T_{12} \le s \,|\, T_{12} + T_{23} > s, T_{12} \le T_{13})}{\mathbb{P}(T_{12} \le s \,|\, T_{12} + T_{23} > s, T_{12} \le T_{13})} \\
&\qquad = \frac{\int_0^s \mathbb{P}(T_{23} > t - u \,|\, T_{12} = u, T_{23} \ge s - u, u \le T_{13}) \, dF_{T_{12}|\bullet}(u)}{\int_0^s dF_{T_{12}|\bullet}(u)} \\
&\qquad = \frac{\int_0^s \mathbb{P}(T_{23} > t - u \,|\, T_{23} > s - u) \, dF_{T_{12}|\bullet}(u)}{\int_0^s dF_{T_{12}|\bullet}(u)} \\
&\qquad = \frac{\int_0^s \mathbb{P}(T_{23} > t - s) \, dF_{T_{12}|\bullet}(u)}{\int_0^s dF_{T_{12}|\bullet}(u)} \\
&\qquad = \mathbb{P}(T_{23} > t - s) \\
&\qquad = \mathbb{P}(T_{23} > (t - s_0) - (s - s_0)) \\
&\qquad = \mathbb{P}(T_{23} > t - s_0 \,|\, T_{23} > s - s_0) \\
&\qquad = \mathbb{P}(T_{23} > t - s_0 \,|\, T_{23} > s - s_0, T_{12} = s_0, T_{12} \le T_{13}) \\
&\qquad = \mathbb{P}(T_{12} + T_{23} > t \,|\, T_{12} + T_{23} > s, T_{12} = s_0, T_{12} \le T_{13})
\end{aligned}$$

where $F_{T_{12}|\cdot}$ is the conditional distribution function of T_{12} given that $T_{12}+T_{23}\geq s$ and $T_{13}\geq T_{12}$.

Thus, satisfying [4.4], the process is Markovian. □

Analytic expression for setting 1.

Assuming that (T_{12},T_{13},T_{23}) are mutually independent and T_{23} is exponentially distributed with rate parameter α_{23}, from the above expression we have

$$\begin{aligned}
p_{22}(s,t) &= \frac{\int_0^s \mathbb{P}(T_{23}>t\,|T_{23}>u)\times\mathbb{P}(T_{13}\geq u)\times\mathbb{P}(T_{12}\geq u)\times dA_{12}(u)}{\int_0^s \mathbb{P}(T_{23}>s\,|T_{23}>u)\times\mathbb{P}(T_{13}\geq u)\times\mathbb{P}(T_{12}\geq u)\times dA_{12}(u)} \\
&= \frac{\mathbb{P}(T_{23}>t)}{\mathbb{P}(T_{23}>s)}\times\frac{\int_0^s \frac{\mathbb{P}(T_{13}\geq u)\times\mathbb{P}(T_{12}\geq u)\times dA_{12}(u)}{\mathbb{P}(T_{23}>u)}}{\int_0^s \frac{\mathbb{P}(T_{13}\geq u)\times\mathbb{P}(T_{12}\geq u)\times dA_{12}(u)}{\mathbb{P}(T_{23}>u)}} \\
&= \mathbb{P}(T_{23}>t\,|T_{23}>s) \\
&= \mathbb{P}(T_{23}>t-s) \\
&= \exp(-\alpha_{23}(t-s)).
\end{aligned}$$

Note that according to the above Lemma, the process is now Markovian.

Analytic expression for setting 2.

According to the above Lemma this sampling scheme leads to a non-Markovian process, since we assume that T_{23} is Weibull with shape parameter 2 and scale parameter α_{23}^{-1}, with density function $f_{23}(x)=2\alpha_{23}^2 x e^{-(\alpha_{23}x)^2}$.

Assuming that (T_{12},T_{13},T_{23}) are mutually independent times, we maintain

$$p_{22}(s,t) = \frac{\int_0^s \mathbb{P}(T_{23} > t-u) \times \mathbb{P}(T_{13} \geq u) \times \mathbb{P}(T_{12} \geq u) \times dA_{12}(u)}{\int_0^s \mathbb{P}(T_{23} > s-u) \times \mathbb{P}(T_{13} \geq u) \times \mathbb{P}(T_{12} \geq u) \times dA_{12}(u)}$$
$$= \frac{\int_0^s e^{-\alpha_{23}{}^2(t-u)^2} \times e^{-(\alpha_{12}+\alpha_{13})u} \times \alpha_{12} du}{\int_0^s e^{-\alpha_{23}{}^2(s-u)^2} \times e^{-(\alpha_{12}+\alpha_{13})u} \times \alpha_{12} du}$$
$$= e^{-\alpha_{23}{}^2(t^2-s^2)} \frac{\int_0^s e^{-(\alpha_{23}u)^2 - (\alpha_{12}+\alpha_{13}-2t\alpha_{23}{}^2)u} du}{\int_0^s e^{-(\alpha_{23}u)^2 - (\alpha_{12}+\alpha_{13}-2s\alpha_{23}{}^2)u} du}$$

Analytic expression for setting 3.

Assuming that $T_{23} = c \times T_{12}$ for some constant c, $c > 0$, it is clear that the process is non-Markovian. Furthermore, times (T_{12}, T_{13}, T_{23}) are not mutually independent, since T_{12} and T_{23} are dependent. Then, the expression for the transition probabilities, $p_{22}(s,t)$ is now given by

$$p_{22}(s,t) = \frac{\mathbb{P}(T_{12} + T_{23} > t, T_{12} \leq T_{13}, T_{12} \leq s)}{\mathbb{P}(T_{12} + T_{23} > s, T_{12} \leq T_{13}, T_{12} \leq s)}$$
$$= \frac{\int_0^s \mathbb{P}(T_{23} > t-u, T_{13} \geq u \,|\, T_{12} = u) \times dF_{12}(u)}{\int_0^s \mathbb{P}(T_{23} > s-u, T_{13} \geq u \,|\, T_{12} = u) \times dF_{12}(u)}$$
$$= \frac{\int_{\min(s,t/(c+1))}^s \mathbb{P}(T_{13} > u \,|\, T_{12} = u) \times \mathbb{P}(T_{12} > u) \times dA_{12}(u)}{\int_{s/(c+1)}^s \mathbb{P}(T_{13} > u \,|\, T_{12} = u) \times \mathbb{P}(T_{12} > u) \times dA_{12}(u)}$$
$$= \frac{\int_{\min(s,t/(c+1))}^s e^{-(\alpha_{12}+\alpha_{13})u} \times \alpha_{12} du}{\int_{s/(c+1)}^s e^{-(\alpha_{12}+\alpha_{13})u} \times \alpha_{12} du}$$
$$= \frac{\int_{\min(s,t/(c+1))}^s e^{-(\alpha_{12}+\alpha_{13})u} du}{\int_{s/(c+1)}^s e^{-(\alpha_{12}+\alpha_{13})u} du}$$
$$= \frac{e^{-(\alpha_{12}+\alpha_{13}) \times \min(s,t/(c+1))} - e^{-(\alpha_{12}+\alpha_{13})s}}{e^{-(\alpha_{12}+\alpha_{13})\frac{s}{c+1}} - e^{-(\alpha_{12}+\alpha_{13})s}}.$$

A.2 Additional figures for section 4.3.1.3.

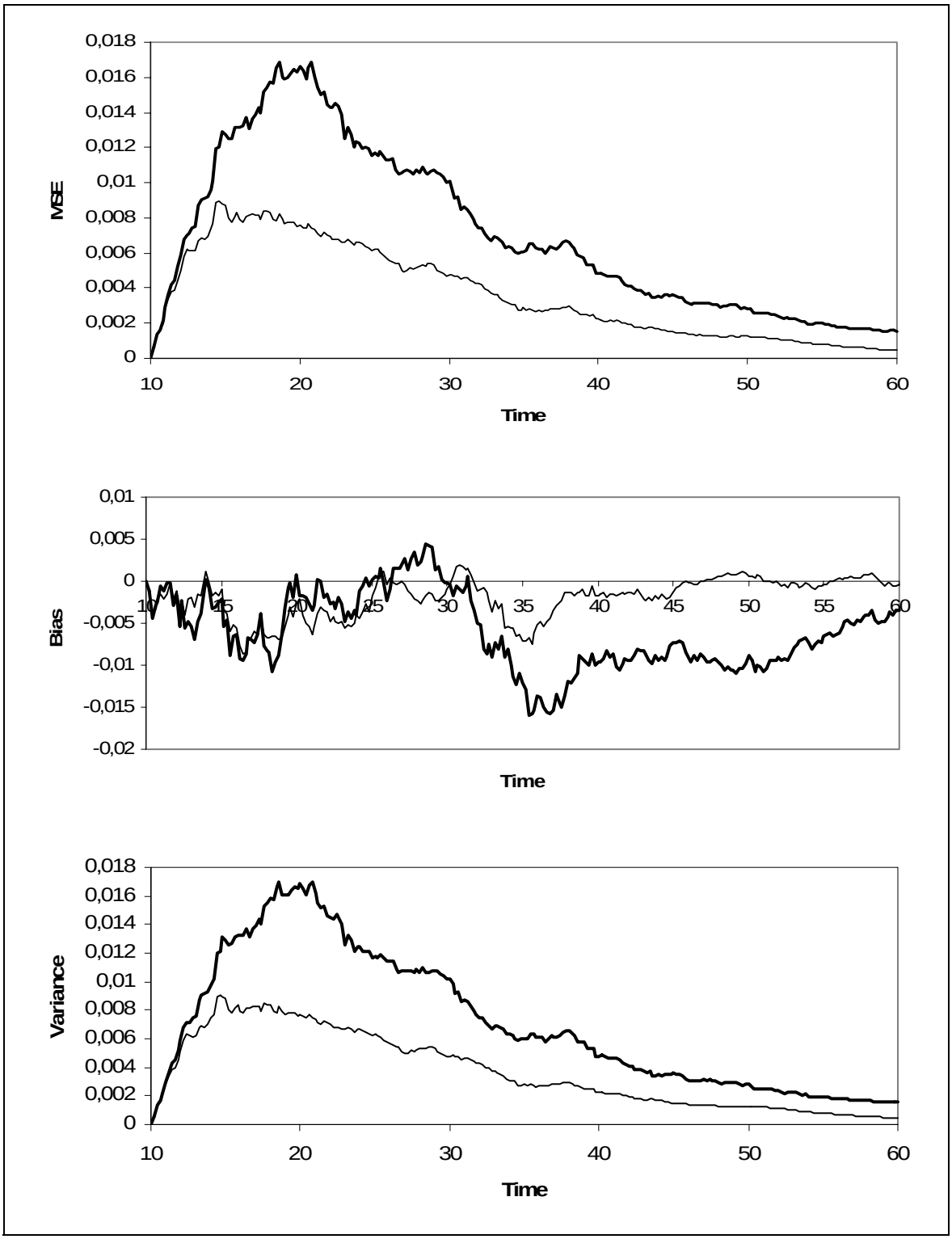


Figure A.1: Curves obtained for setting 1 with $s = 10$, $N = 100$ and 0% of censored observations under Aalen-Johansen estimator (solid line) and non-Markovian estimator (solid bold line).

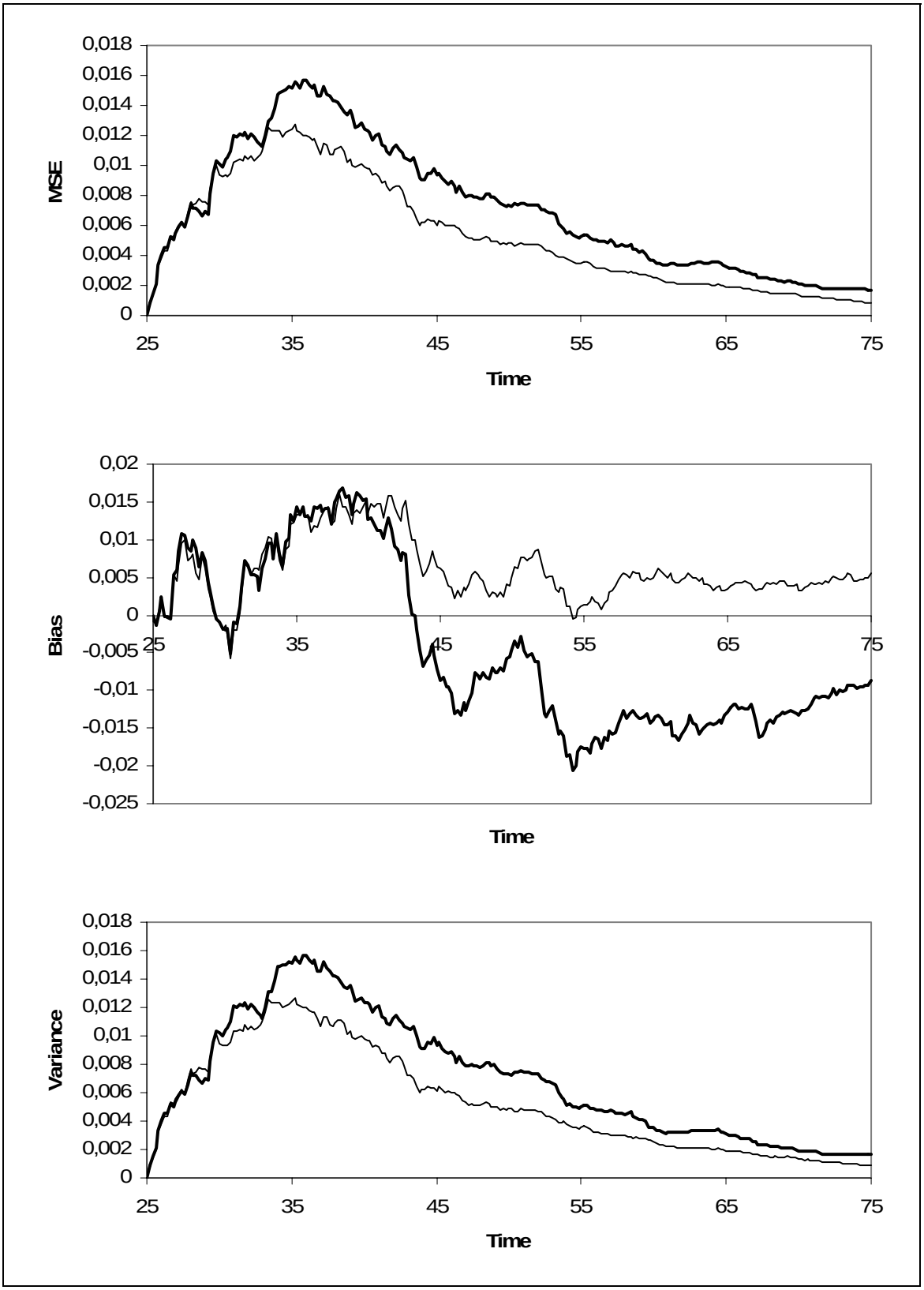


Figure A.2: Curves obtained for setting 1 with $s = 25$, $N = 100$ and 0% of censored observations under Aalen-Johansen estimator (solid line) and non-Markovian estimator (solid bold line).

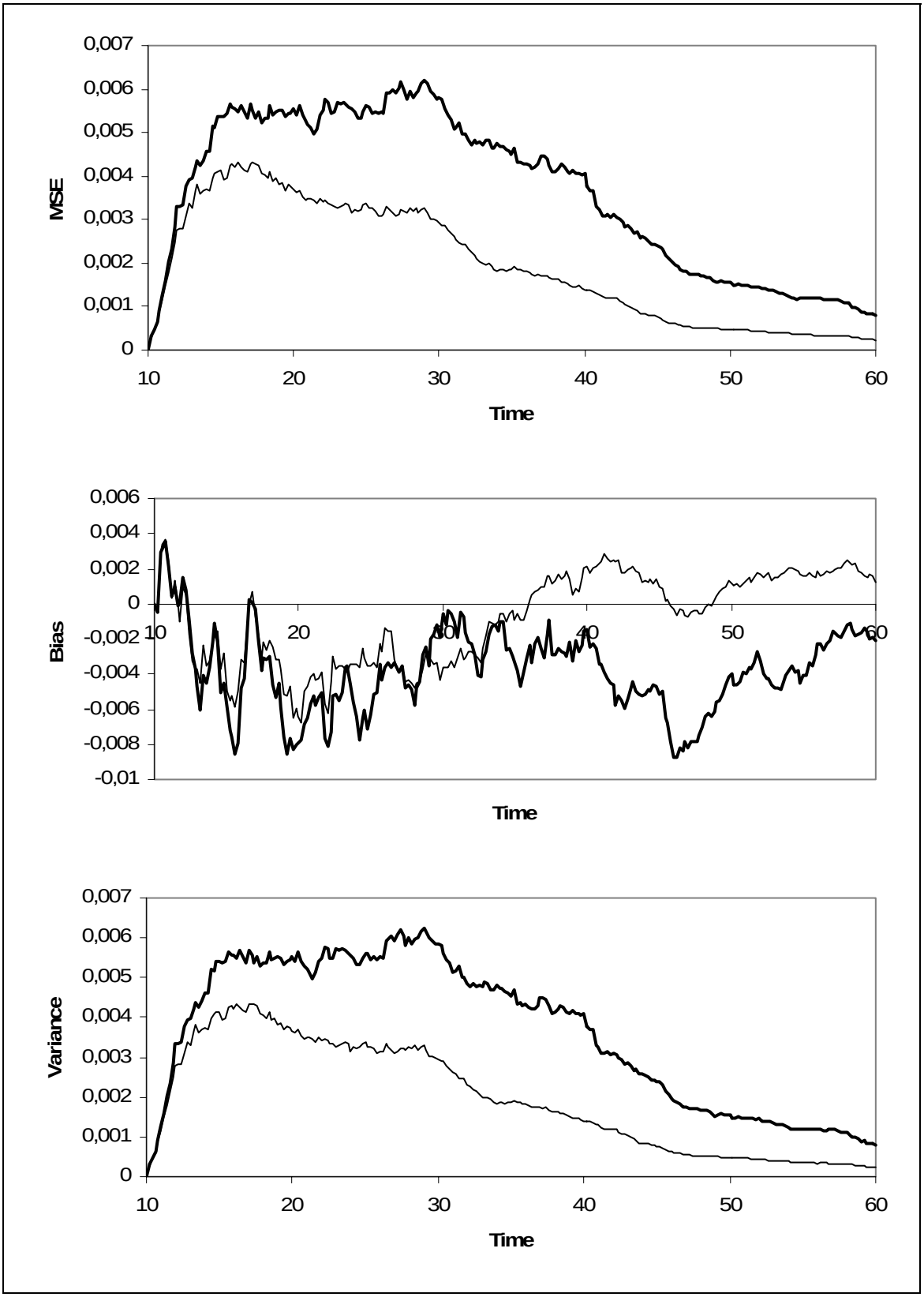


Figure A.3: Curves obtained for setting 1 with $s = 10$, $N = 200$ and 0% of censored observations under Aalen-Johansen estimator (solid line) and non-Markovian estimator (solid bold line).

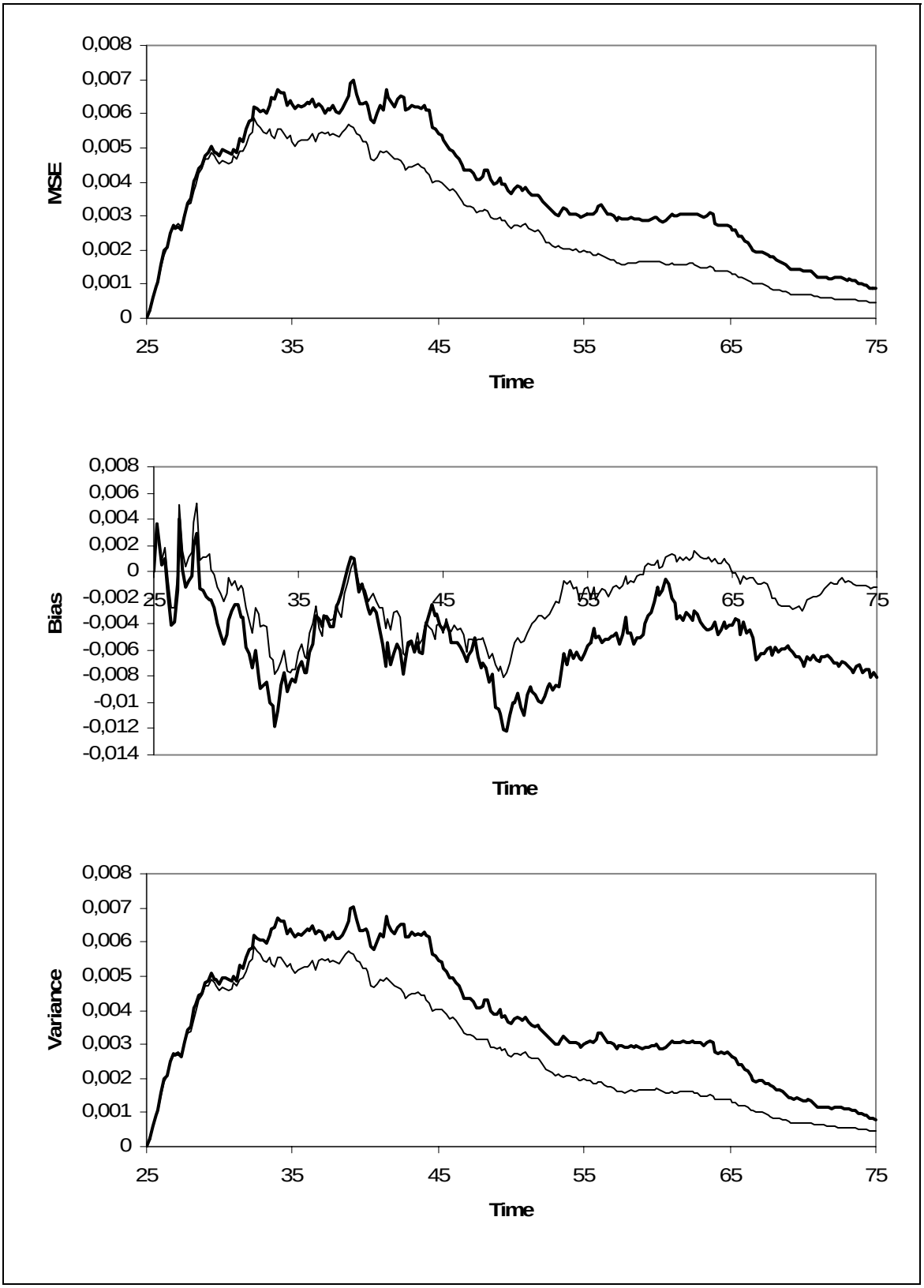


Figure A.4: Curves obtained for setting 1 with $s = 25$, $N = 200$ and 0% of censored observations under Aalen-Johansen estimator (solid line) and non-Markovian estimator (solid bold line).

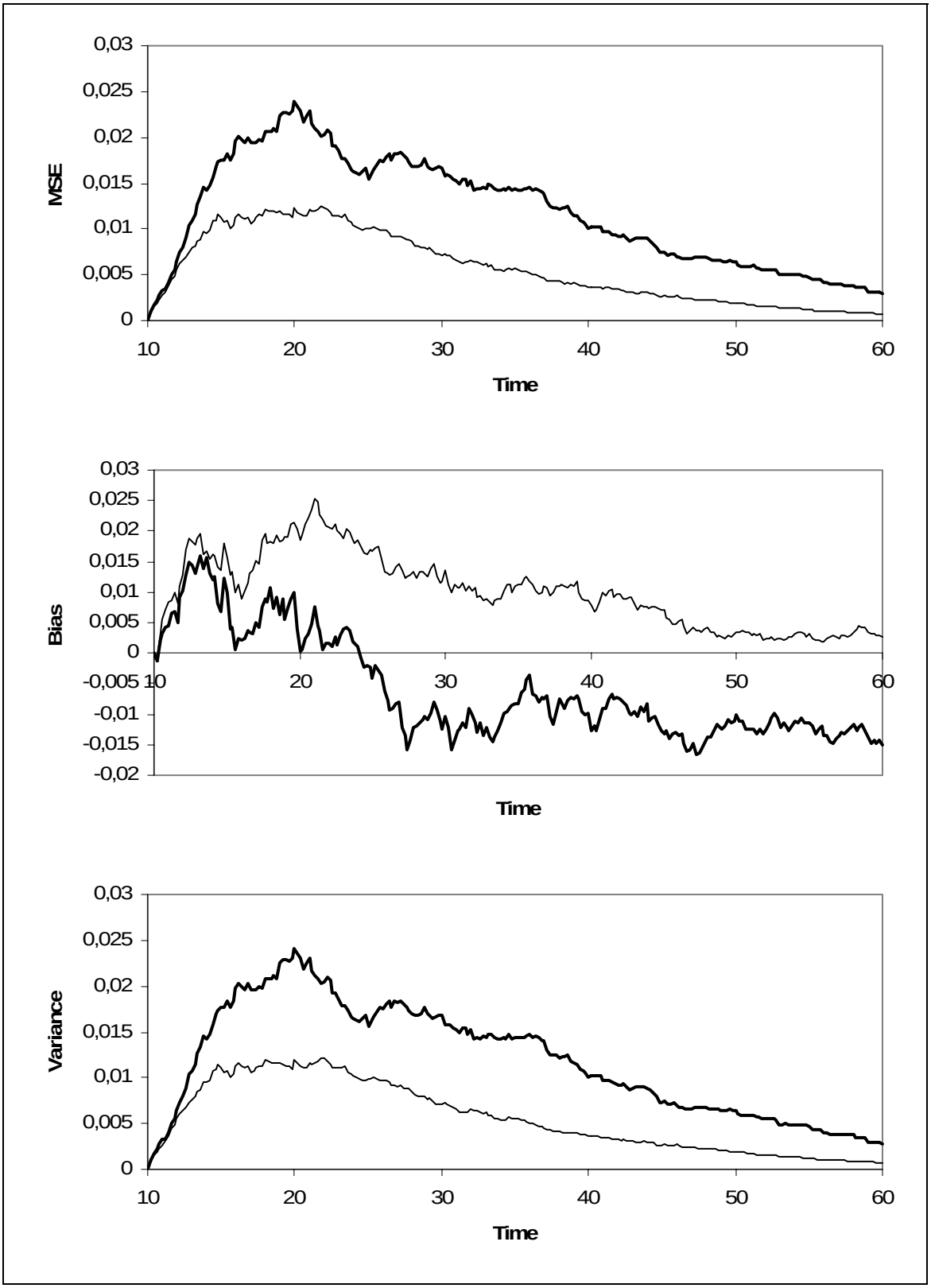


Figure A.5: Curves obtained for setting 1 with $s = 10$, $N = 100$ and 25% of censored observations under Aalen-Johansen estimator (solid line) and non-Markovian estimator (solid bold line).

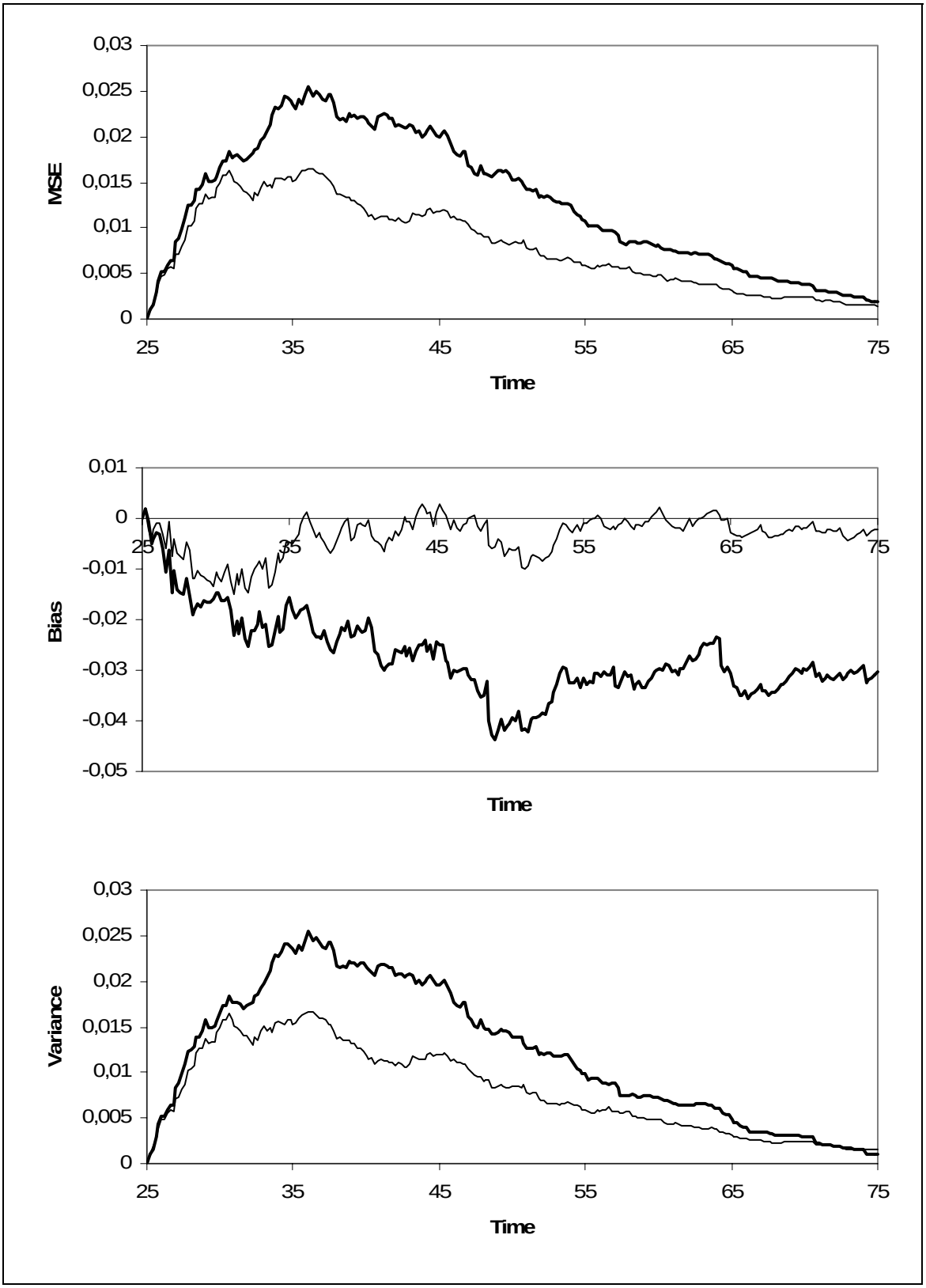


Figure A.6: Curves obtained for setting 1 with $s = 25$, $N = 100$ and 25% of censored observations under Aalen-Johansen estimator (solid line) and non-Markovian estimator (solid bold line).

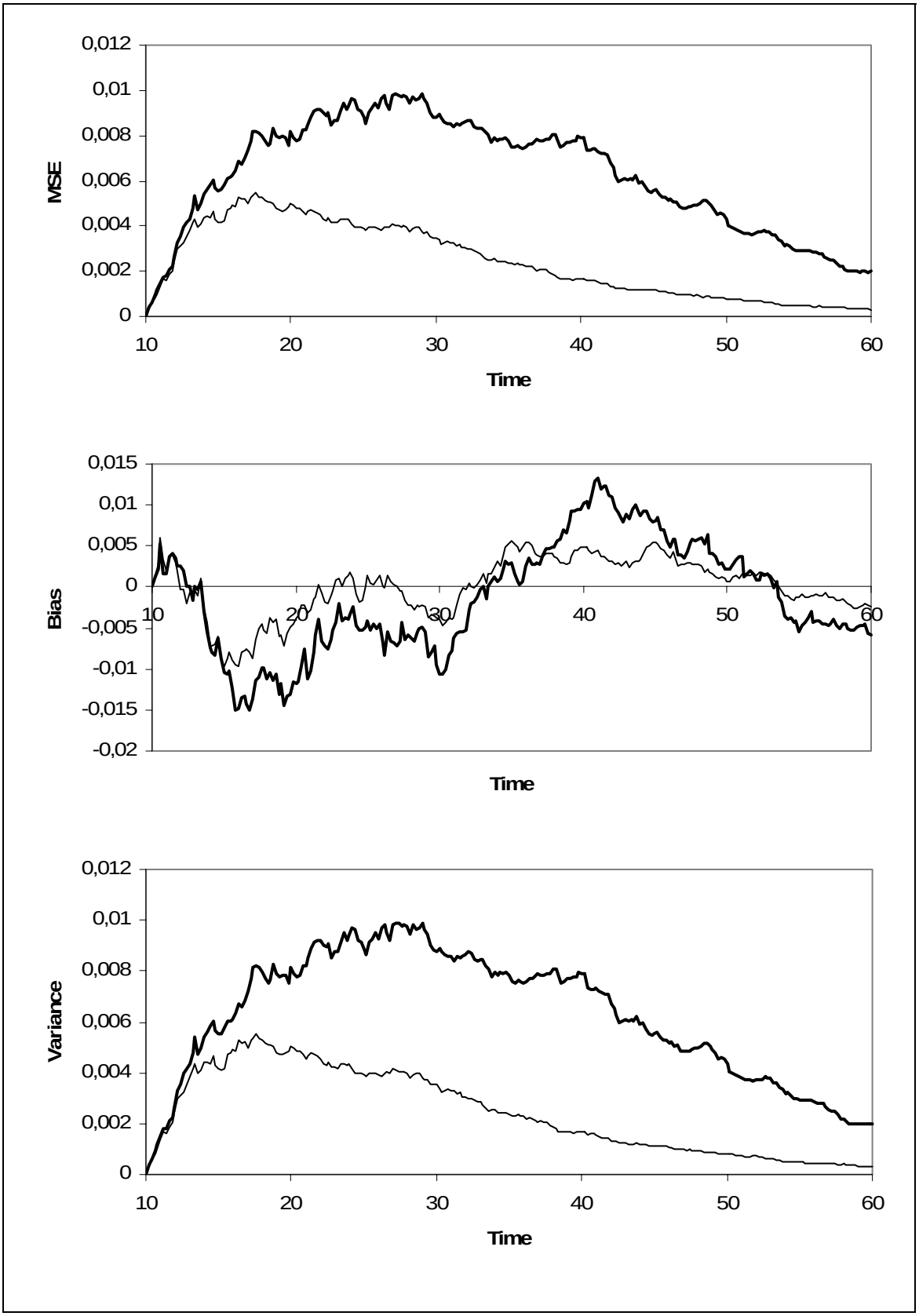


Figure A.7: Curves obtained for setting 1 with $s = 10$, $N = 200$ and 25% of censored observations under Aalen-Johansen estimator (solid line) and non-Markovian estimator (solid bold line).

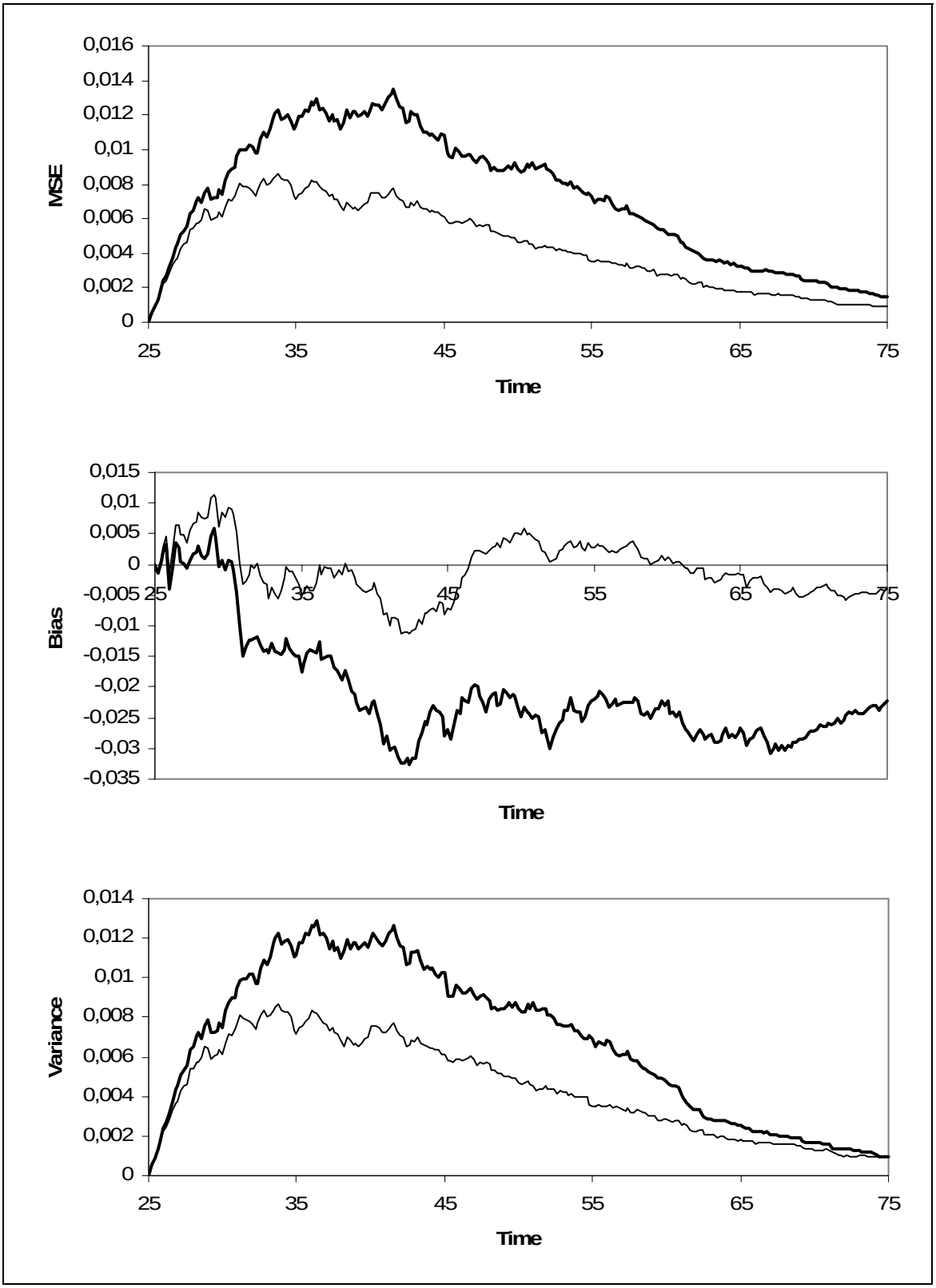


Figure A.8: Curves obtained for setting 1 with $s = 25$, $N = 200$ and 25% of censored observations under Aalen-Johansen estimator (solid line) and non-Markovian estimator (solid bold line).

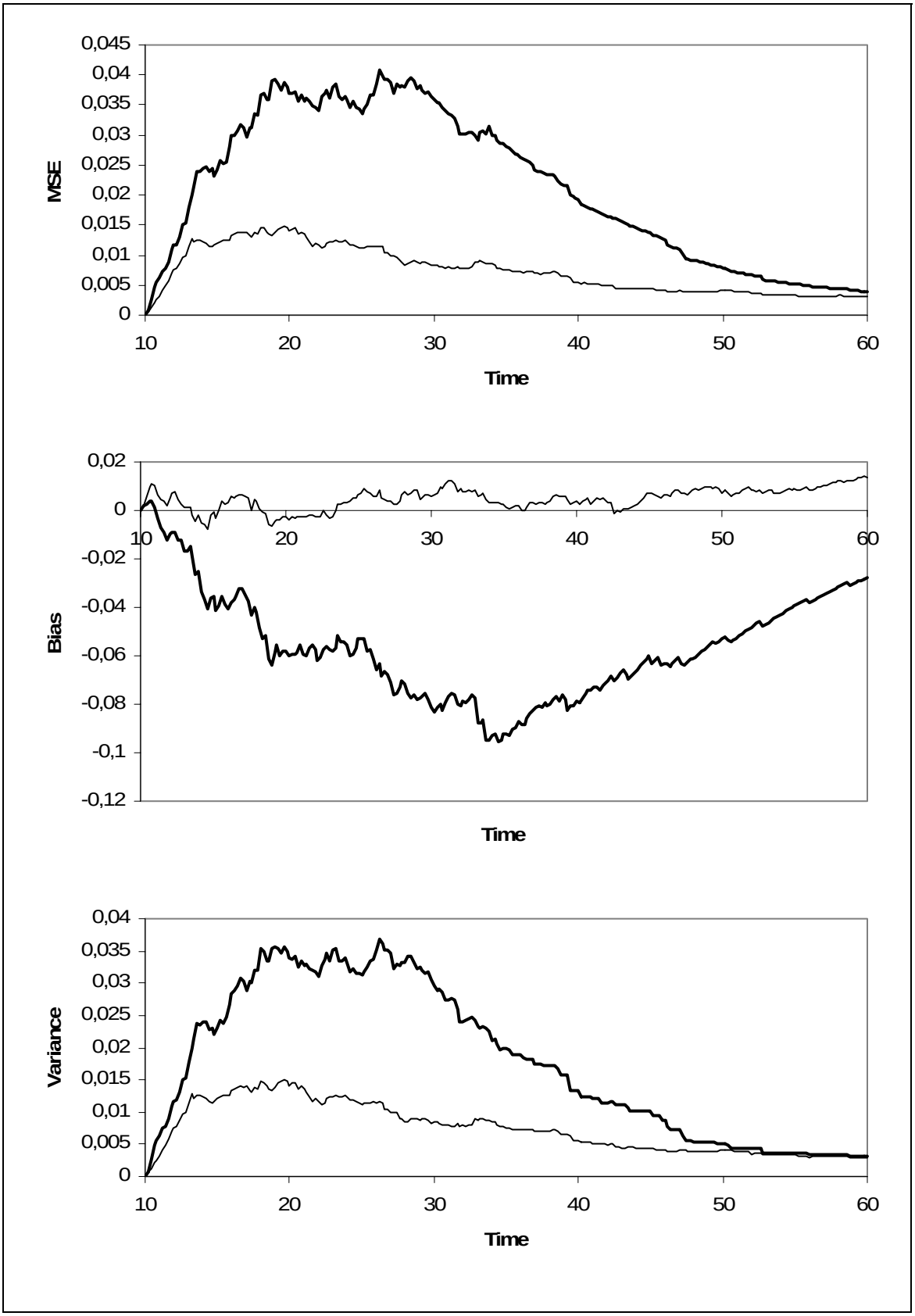


Figure A.9: Curves obtained for setting 1 with $s = 10$, $N = 100$ and 49% of censored observations under Aalen-Johansen estimator (solid line) and non-Markovian estimator (solid bold line).

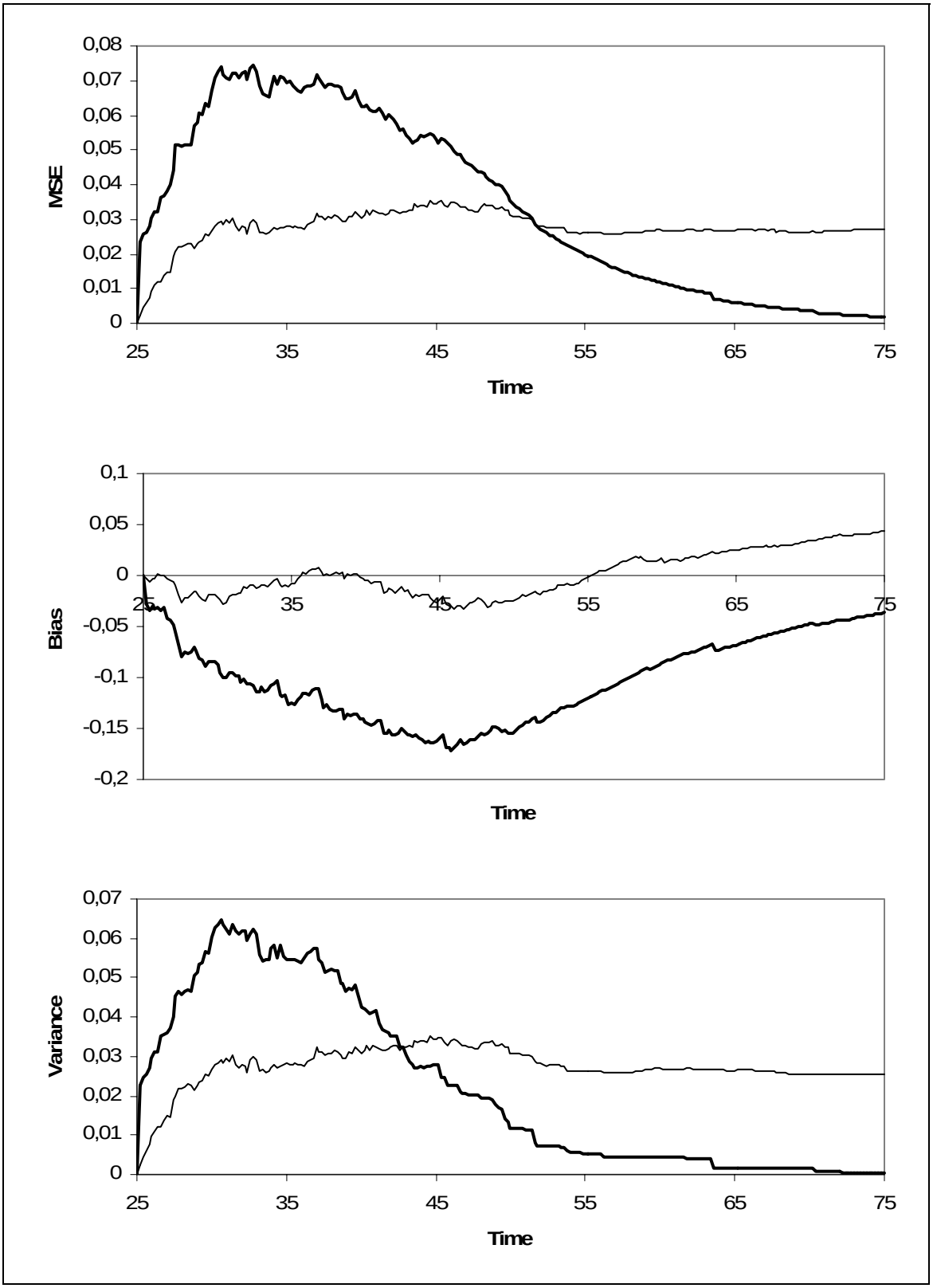


Figure A.10: Curves obtained for setting 1 with $s = 25$, $N = 100$ and 49% of censored observations under Aalen-Johansen estimator (solid line) and non-Markovian estimator (solid bold line).

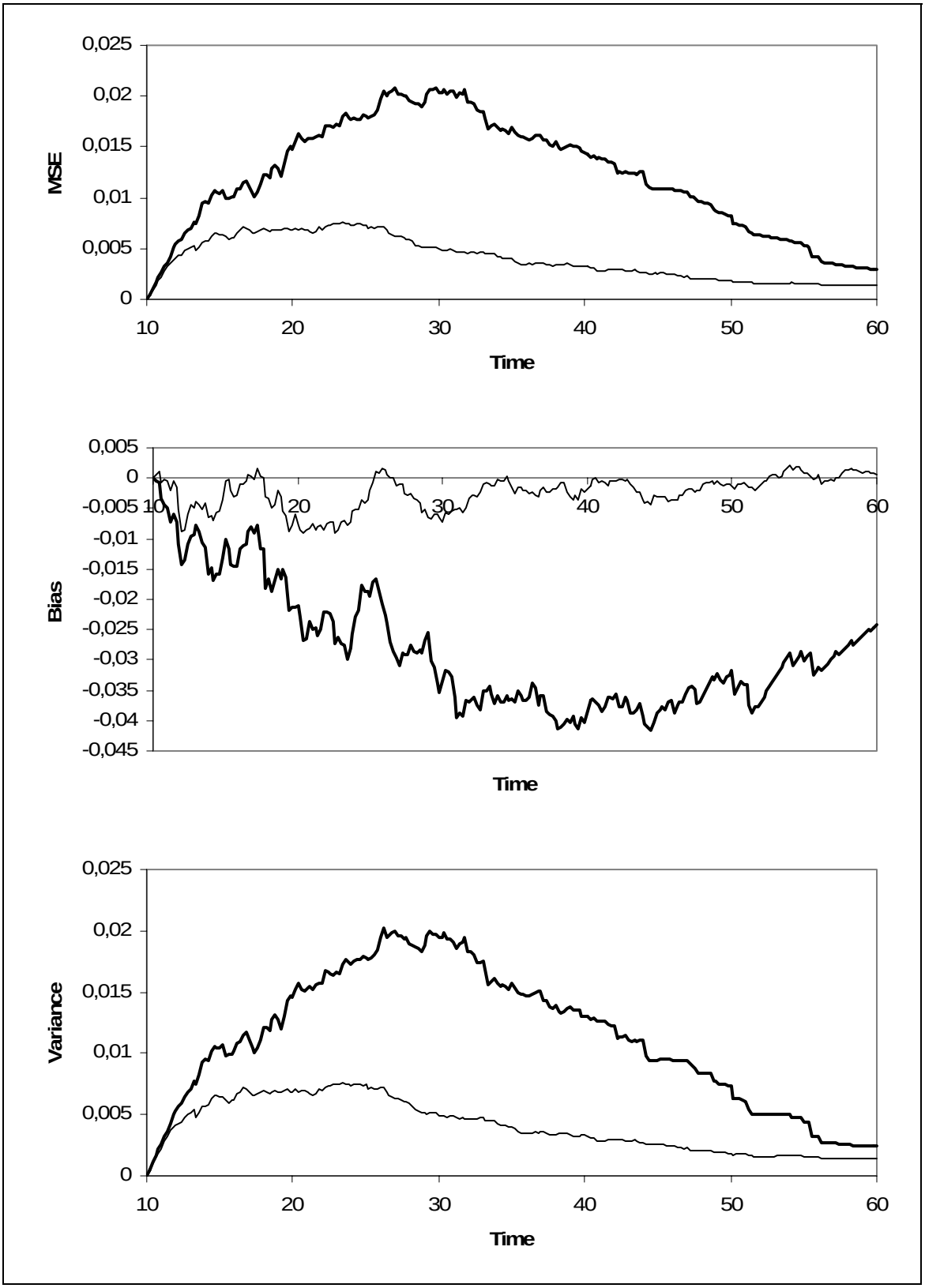


Figure A.11: Curves obtained for setting 1 with $s = 10$, $N = 200$ and 49% of censored observations under Aalen-Johansen estimator (solid line) and non-Markovian estimator (solid bold line).

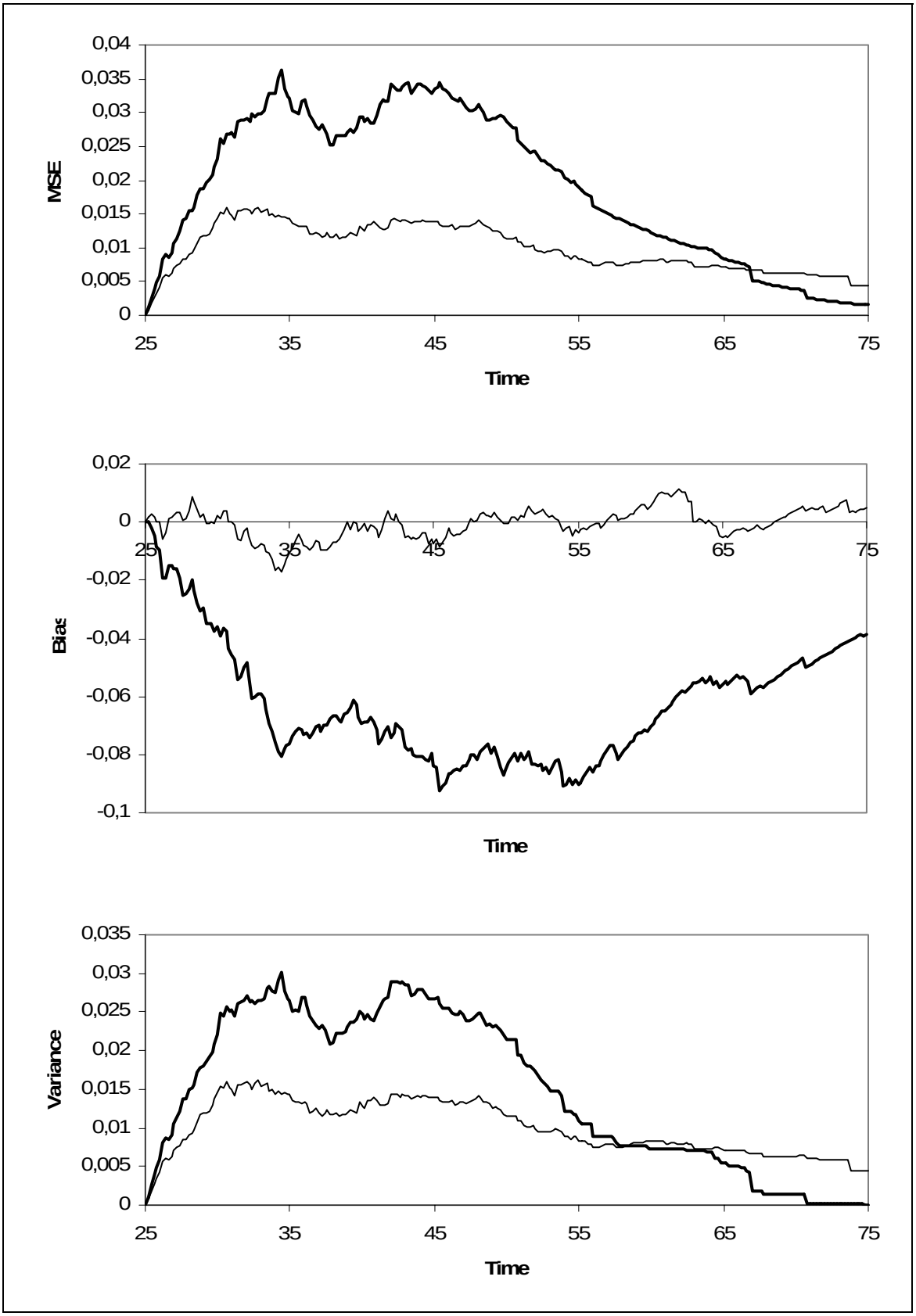


Figure A.12: Curves obtained for setting 1 with $s = 25$, $N = 200$ and 49% of censored observations under Aalen-Johansen estimator (solid line) and non-Markovian estimator (solid bold line).

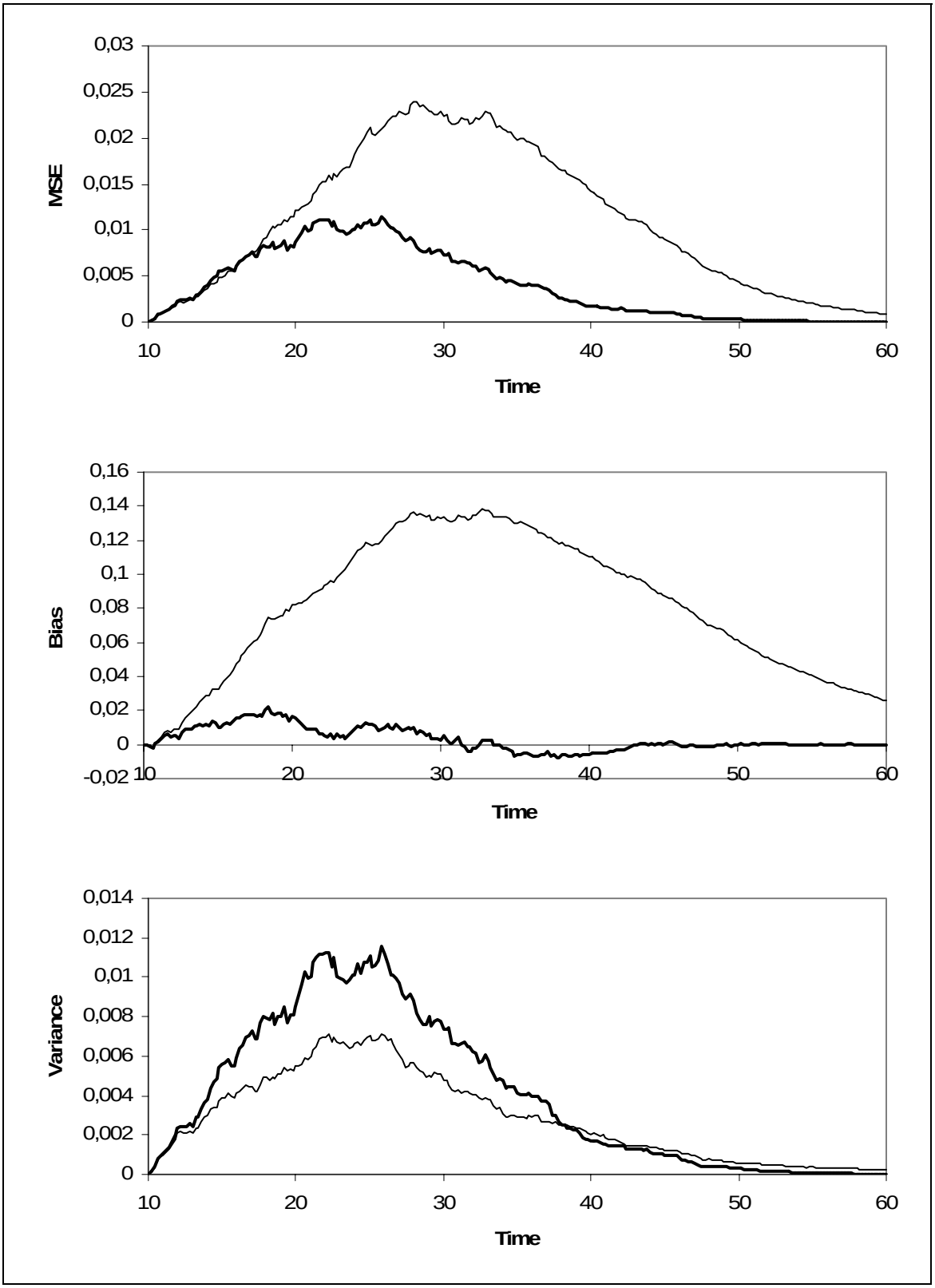


Figure A.13: Curves obtained for setting 2 with $s = 10$, $N = 100$ and 0% of censored observations under Aalen-Johansen estimator (solid line) and non-Markovian estimator (solid bold line).

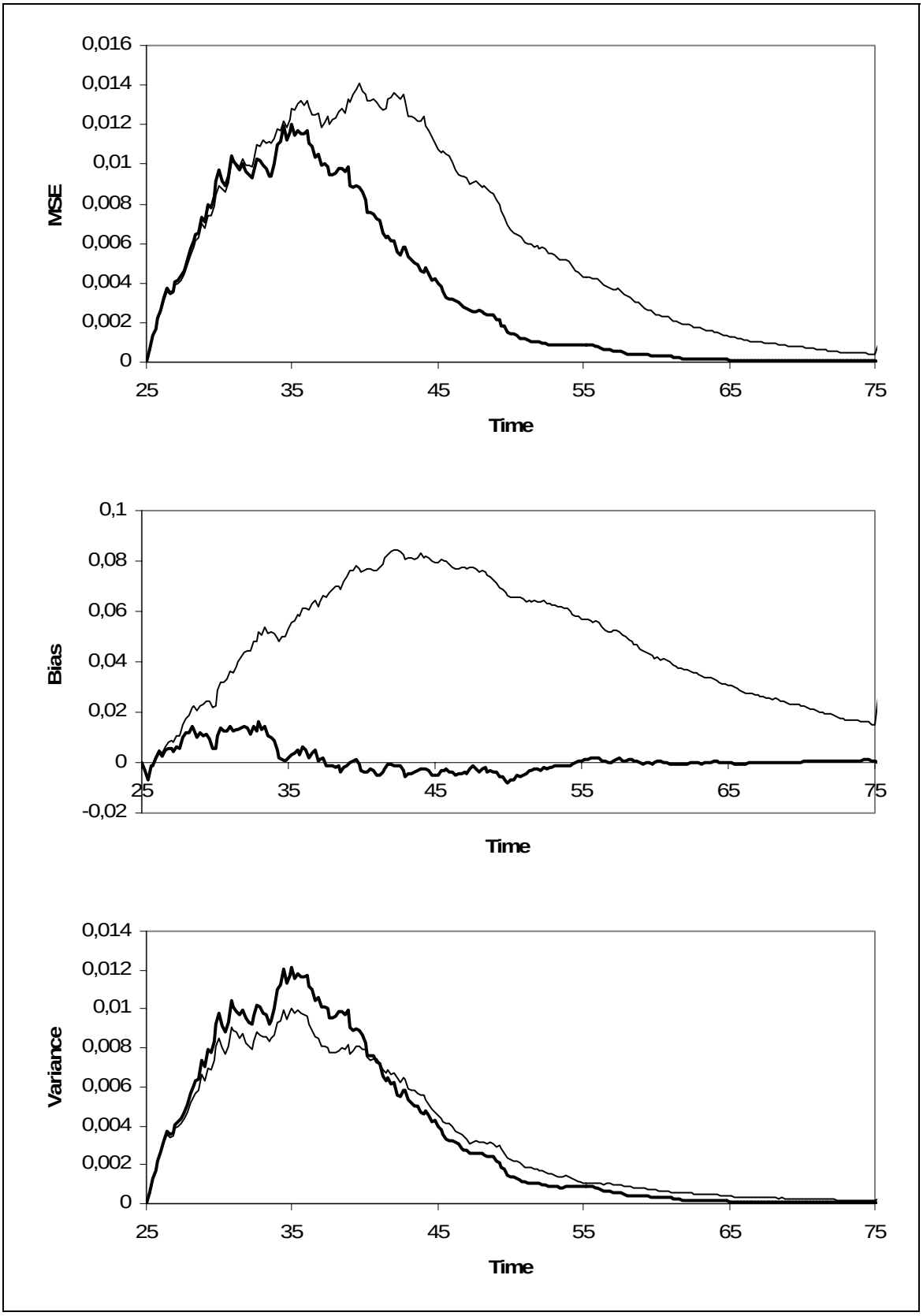


Figure A.14: Curves obtained for setting 2 with $s = 25$, $N = 100$ and 0% of censored observations under Aalen-Johansen estimator (solid line) and non-Markovian estimator (solid bold line).

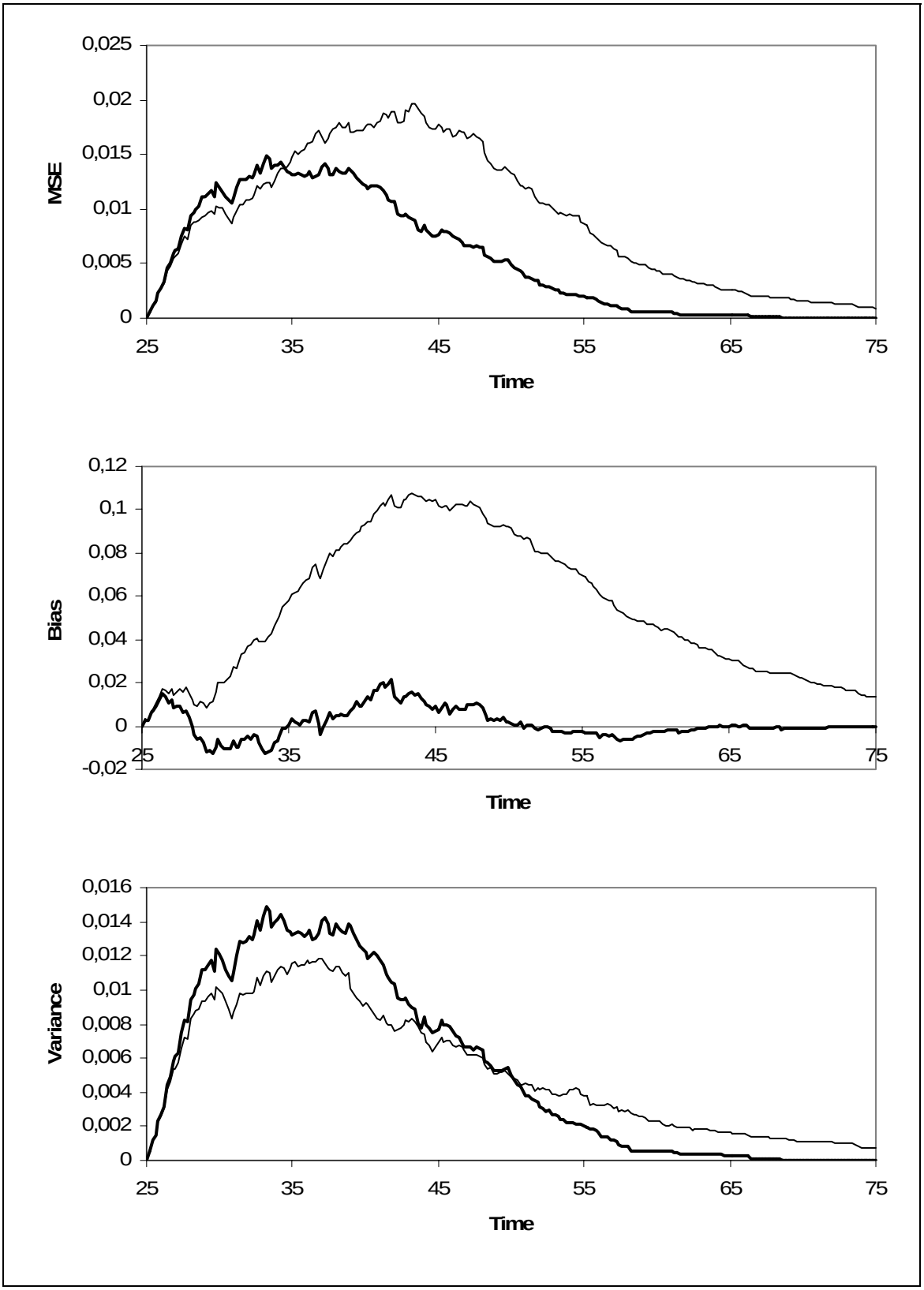


Figure A.15: Curves obtained for setting 2 with $s = 25$, $N = 100$ and 27% of censored observations under Aalen-Johansen estimator (solid line) and non-Markovian estimator (solid bold line).

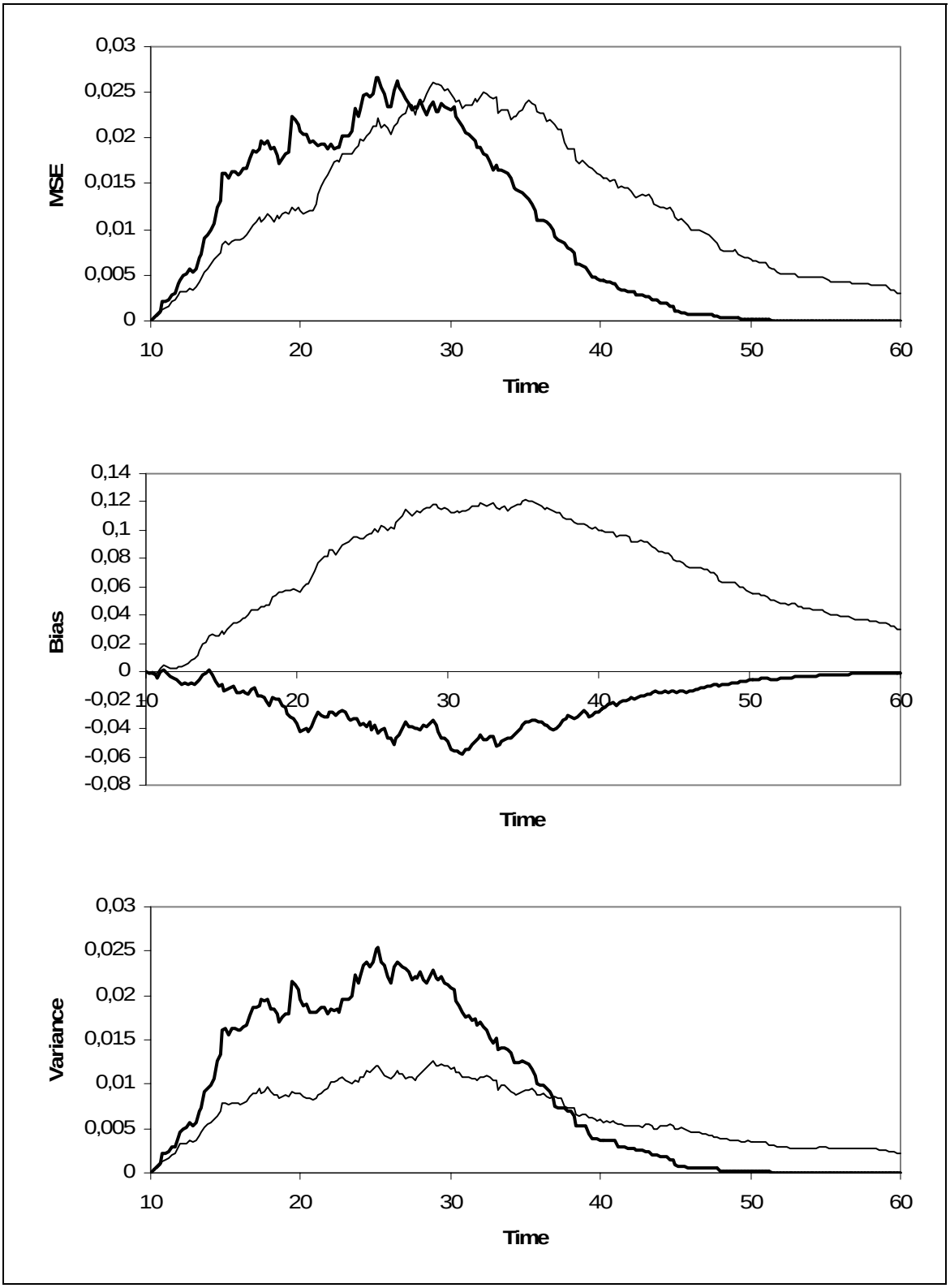


Figure A.16: Curves obtained for setting 2 with $s = 10$, $N = 100$ and 52% of censored observations under Aalen-Johansen estimator (solid line) and non-Markovian estimator (solid bold line).

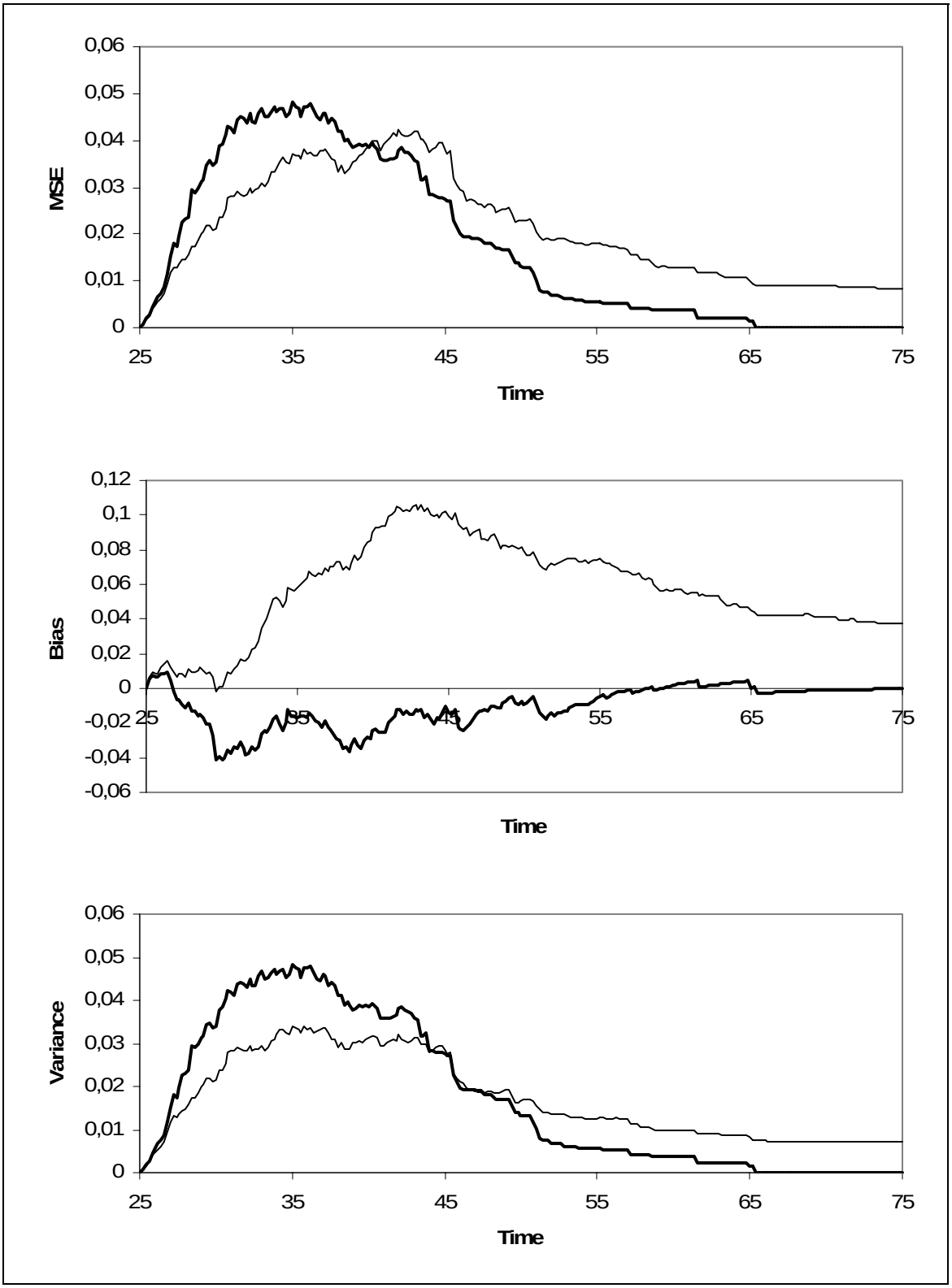


Figure A.17: Curves obtained for setting 2 with $s = 25$, $N = 100$ and 52% of censored observations under Aalen-Johansen estimator (solid line) and non-Markovian estimator (solid bold line).

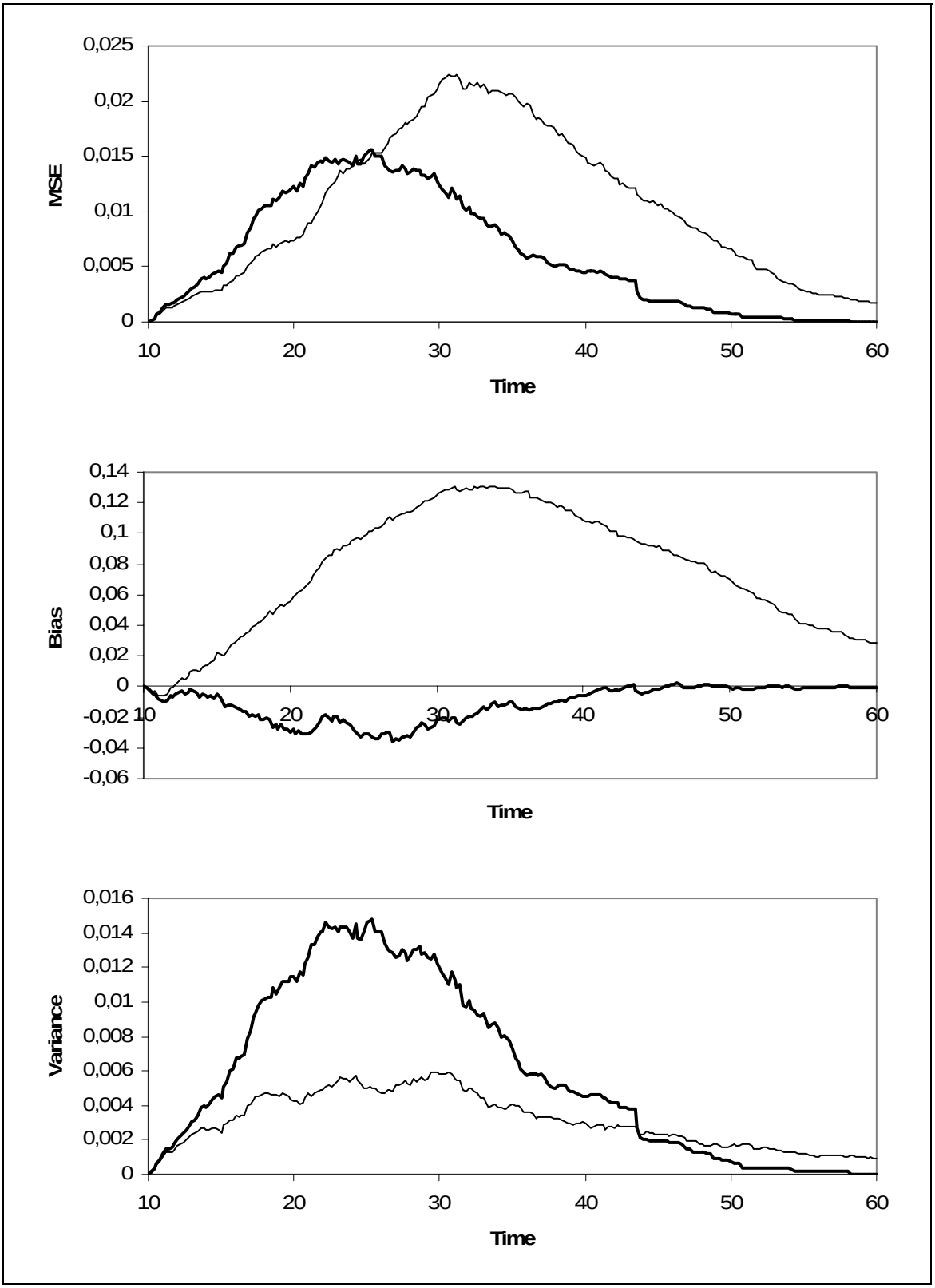


Figure A.18: Curves obtained for setting 2 with $s = 10$, $N = 200$ and 52% of censored observations under Aalen-Johansen estimator (solid line) and non-Markovian estimator (solid bold line).

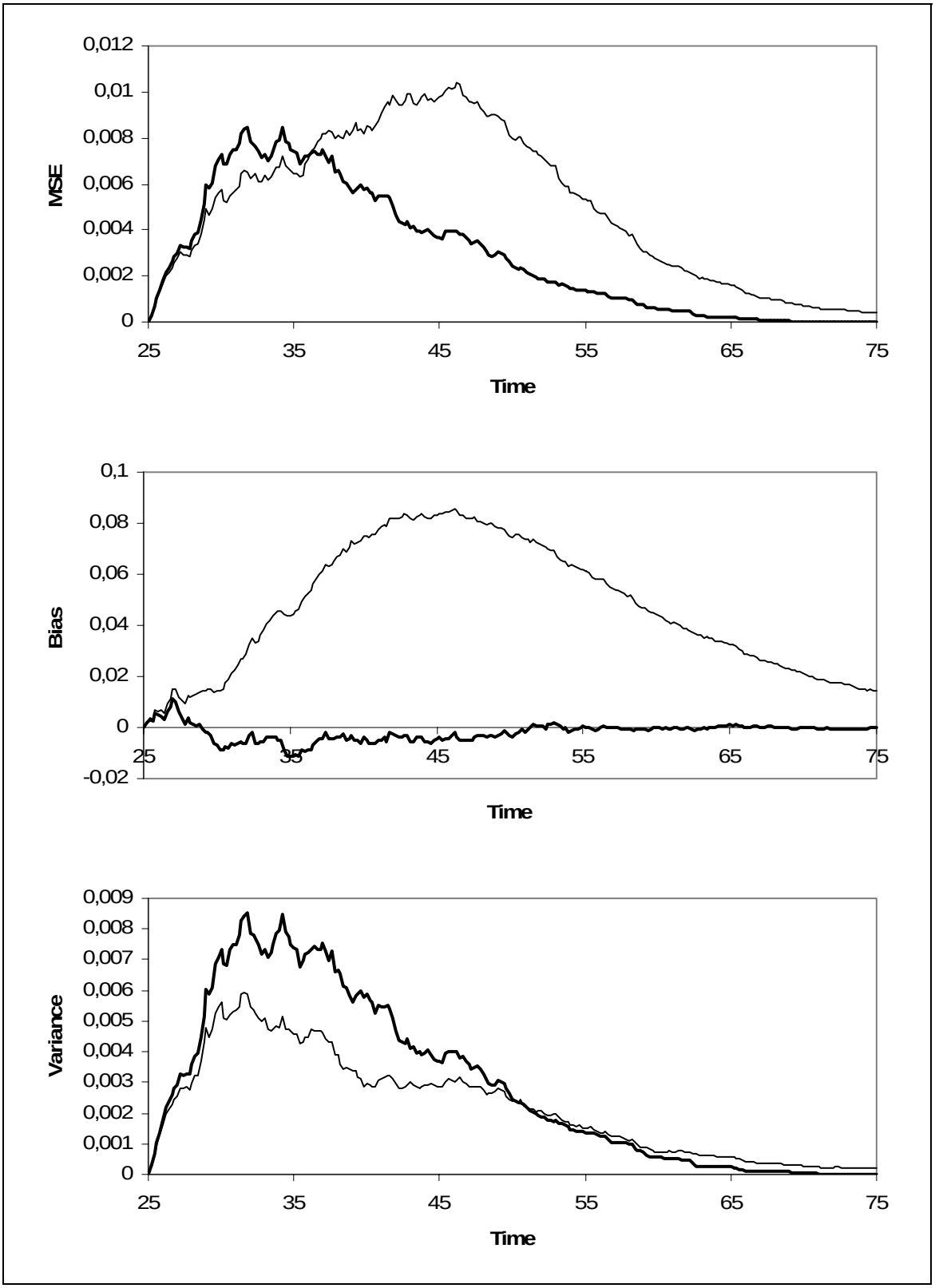


Figure A.19: Curves obtained for setting 2 with $s = 25$, $N = 200$ and 27% of censored observations under Aalen-Johansen estimator (solid line) and non-Markovian estimator (solid bold line).

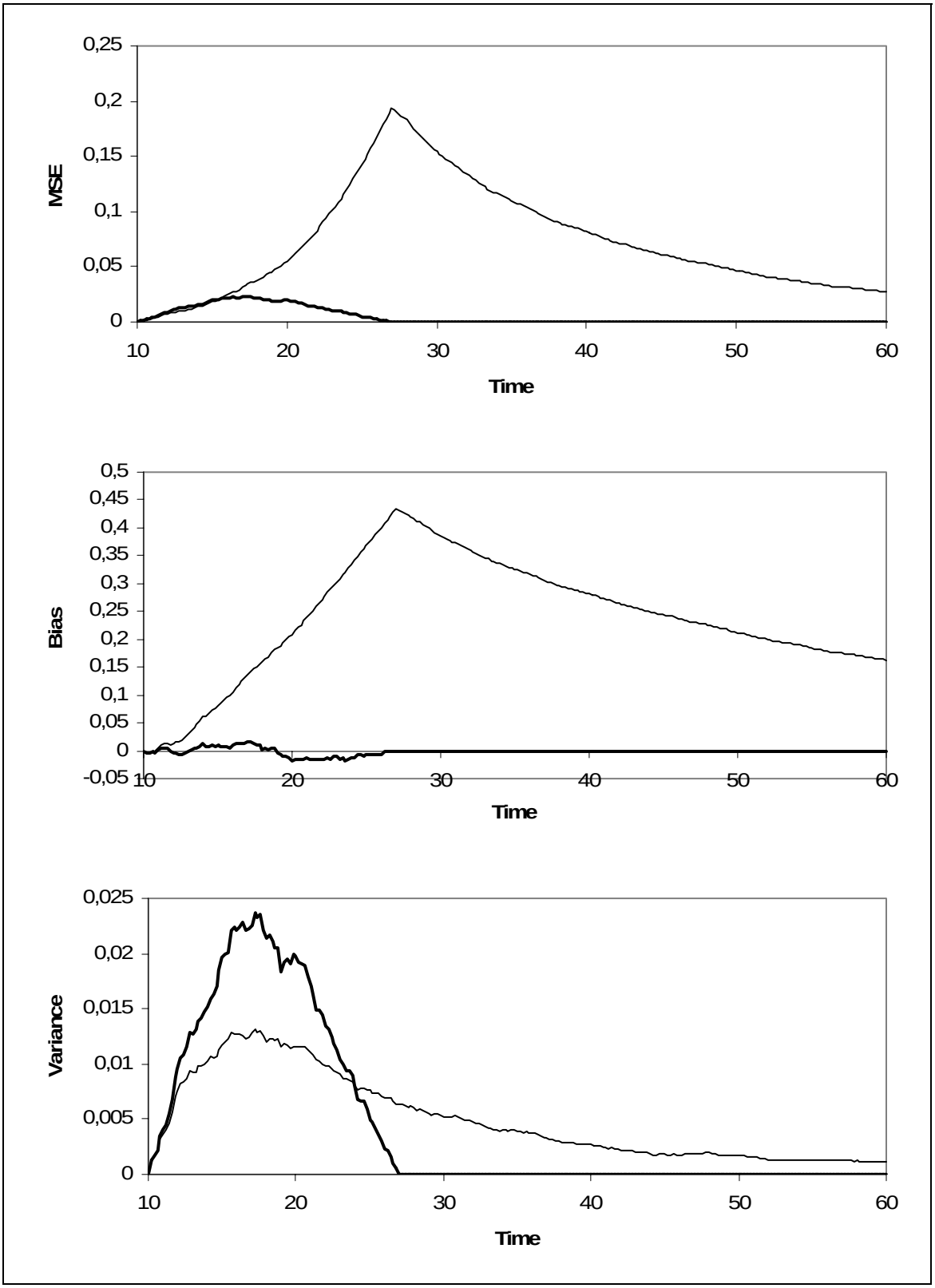


Figure A.20: Curves obtained for setting 3 with $s = 10$, $N = 100$ and 0% of censored observations under Aalen-Johansen estimator (solid line) and non-Markovian estimator (solid bold line).

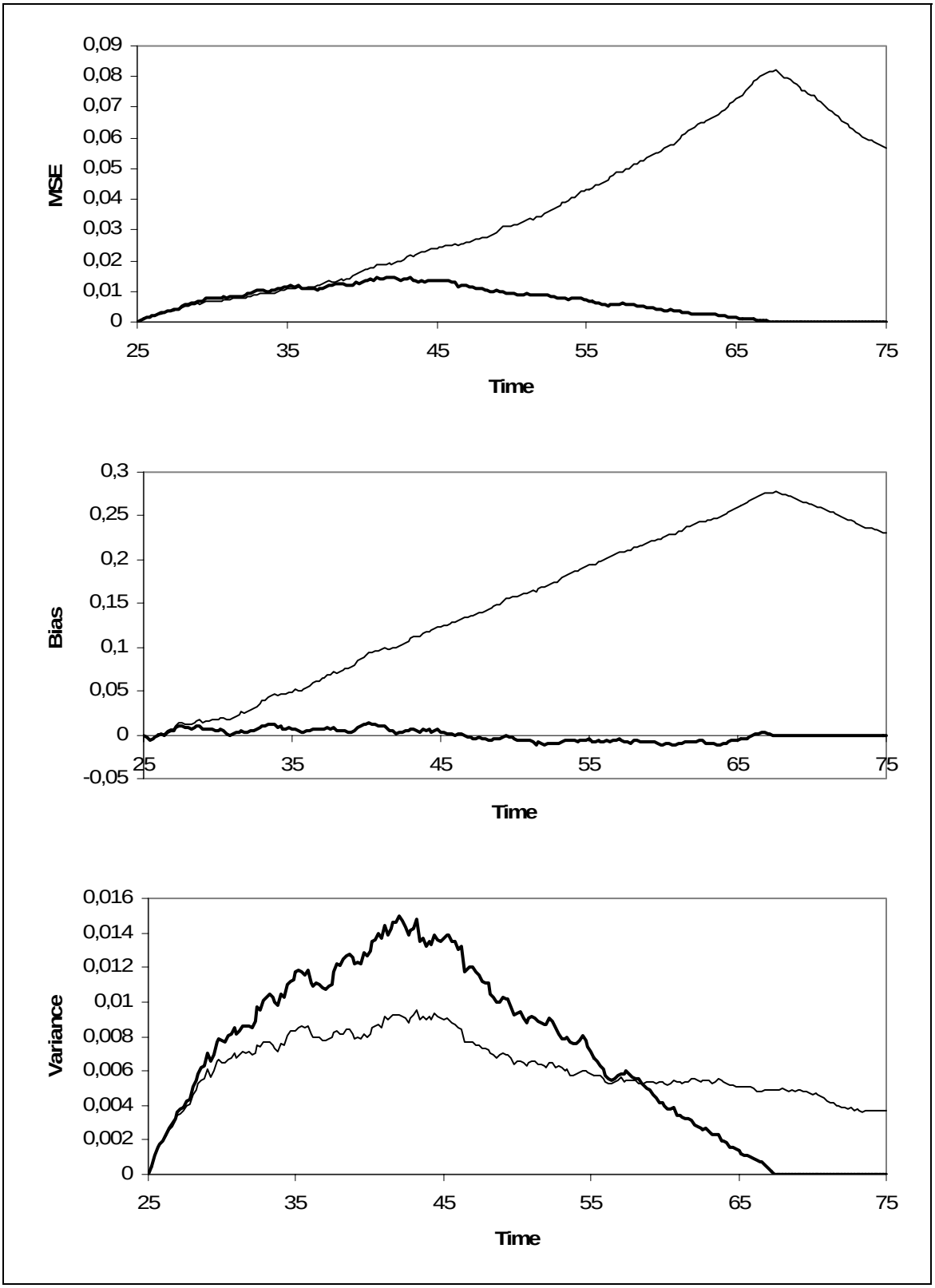


Figure A.21: Curves obtained for setting 3 with $s = 25$, $N = 100$ and 0% of censored observations under Aalen-Johansen estimator (solid line) and non-Markovian estimator (solid bold line).

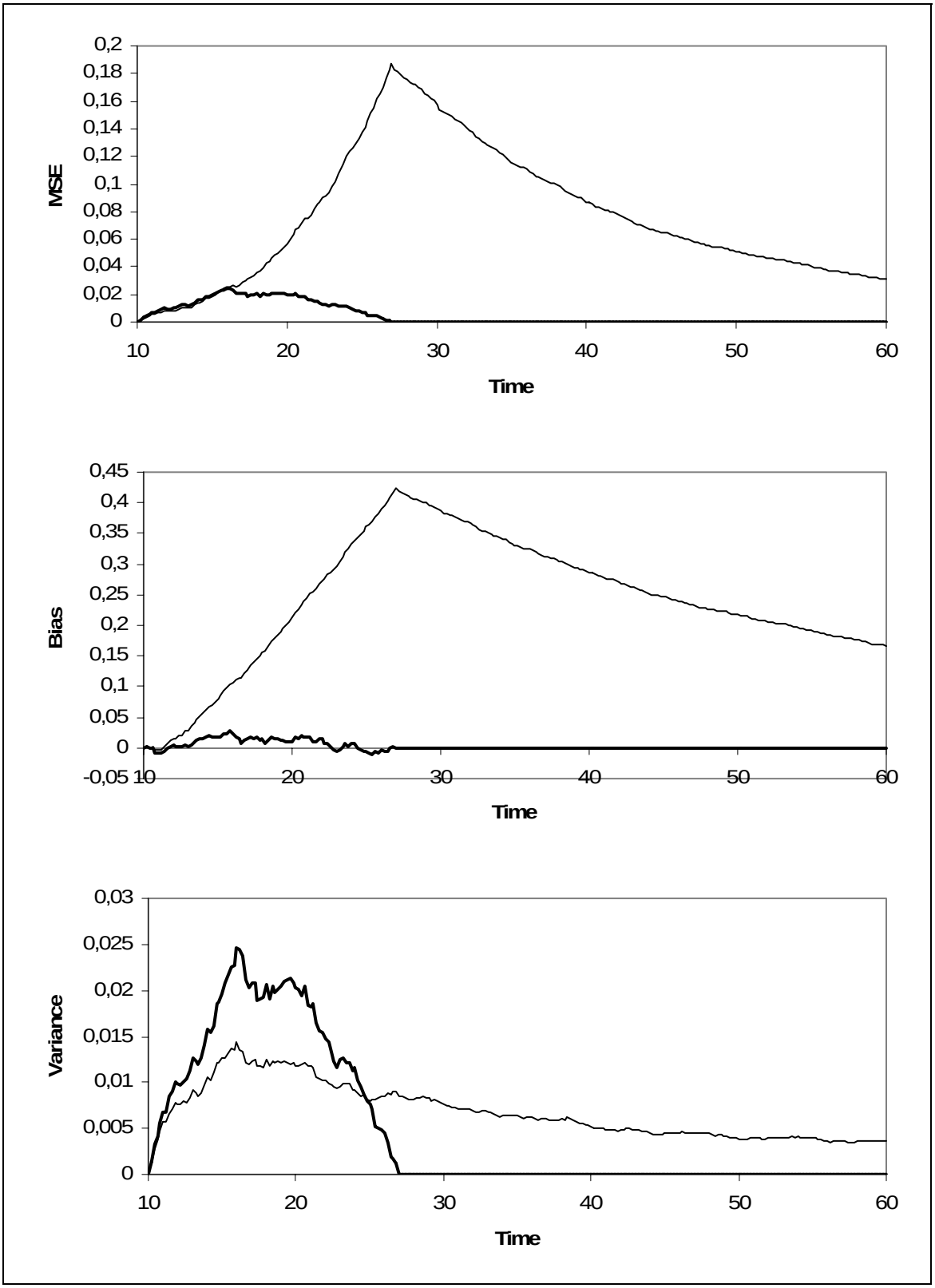


Figure A.22: Curves obtained for setting 3 with $s = 10$, $N = 100$ and 32% of censored observations under Aalen-Johansen estimator (solid line) and non-Markovian estimator (solid bold line).

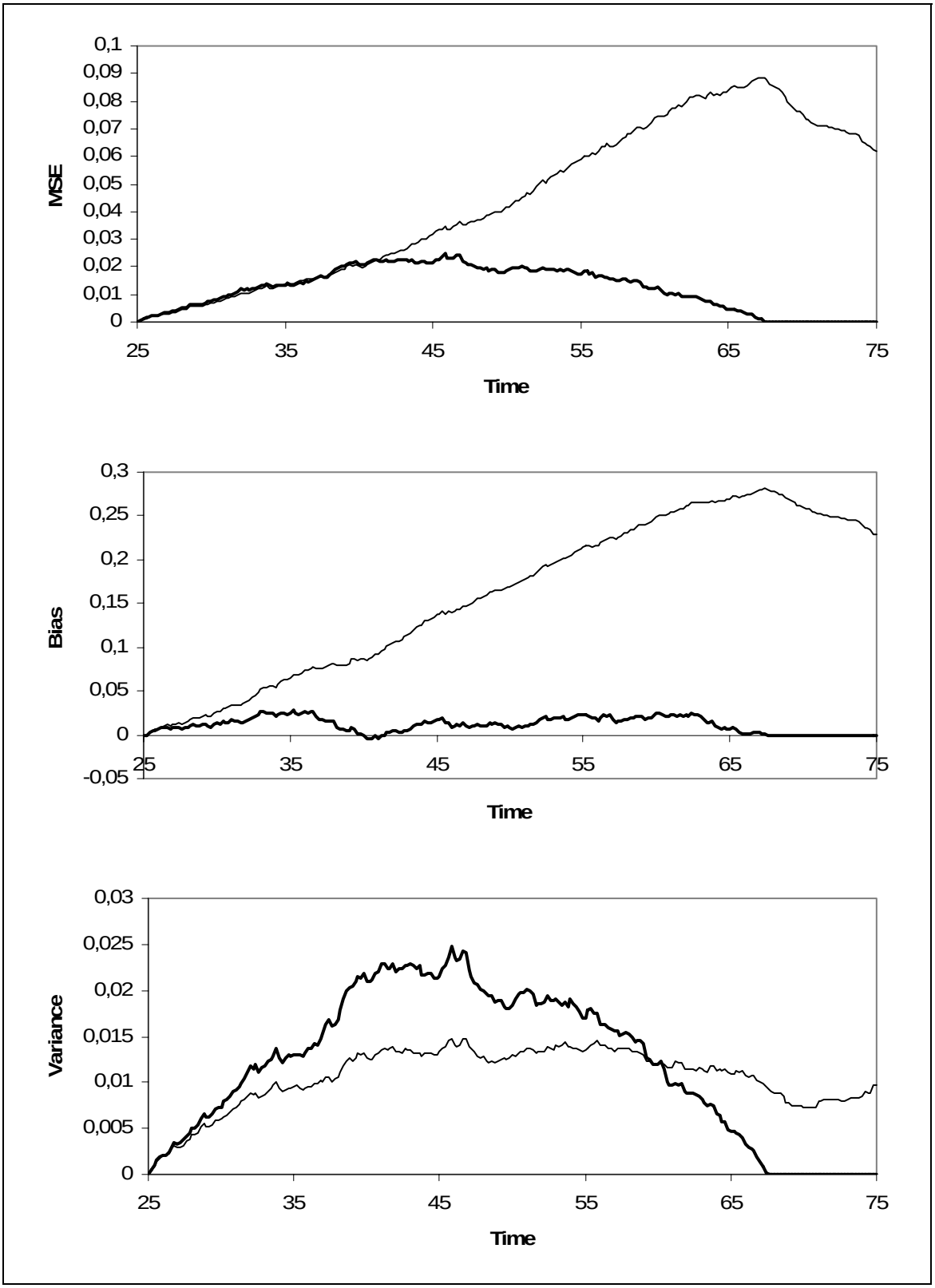


Figure A.23: Curves obtained for setting 3 with $s = 25$, $N = 100$ and 32% of censored observations under Aalen-Johansen estimator (solid line) and non-Markovian estimator (solid bold line).

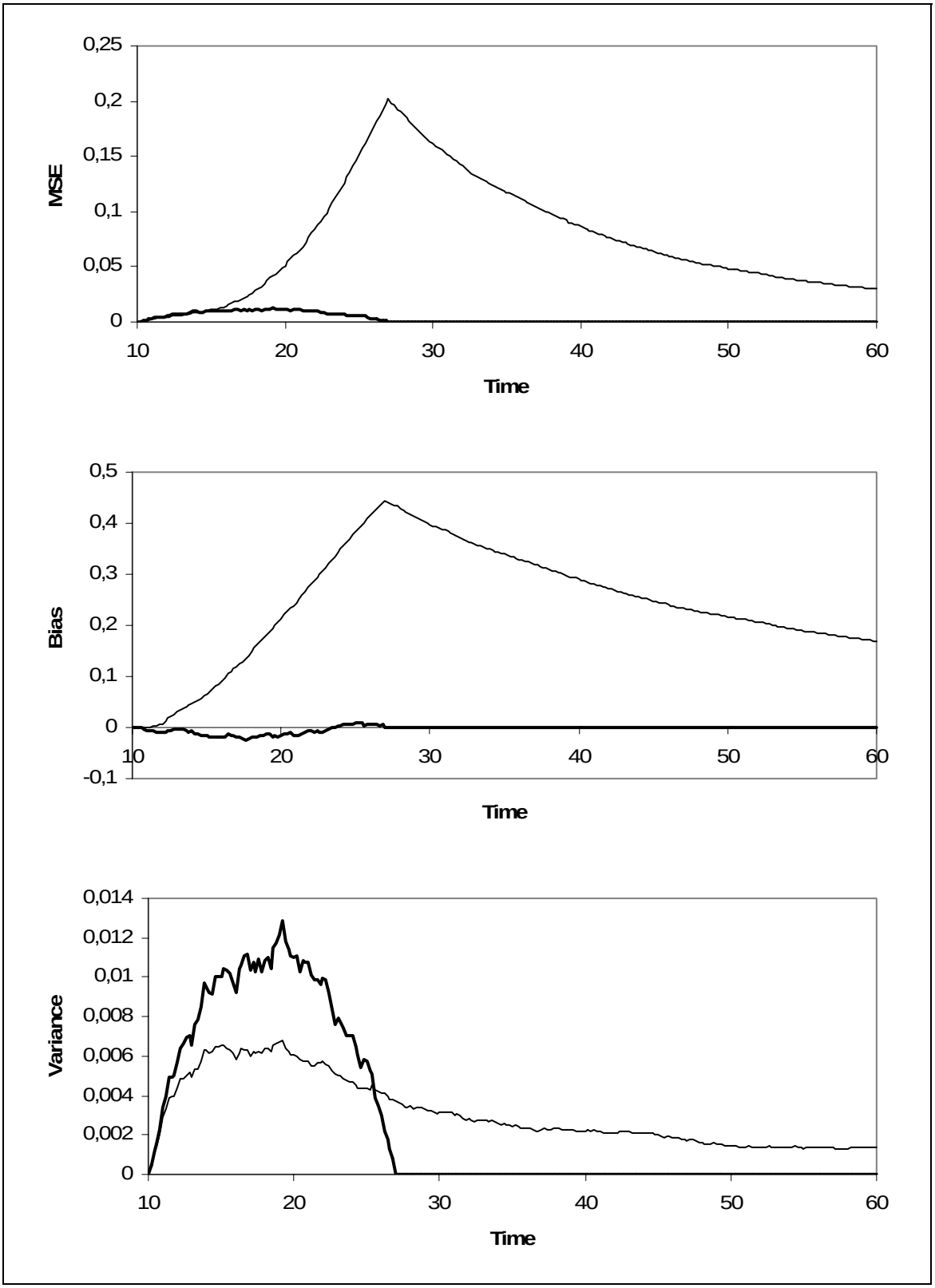


Figure A.24: Curves obtained for setting 3 with $s = 10$, $N = 200$ and 32% of censored observations under Aalen-Johansen estimator (solid line) and non-Markovian estimator (solid bold line).

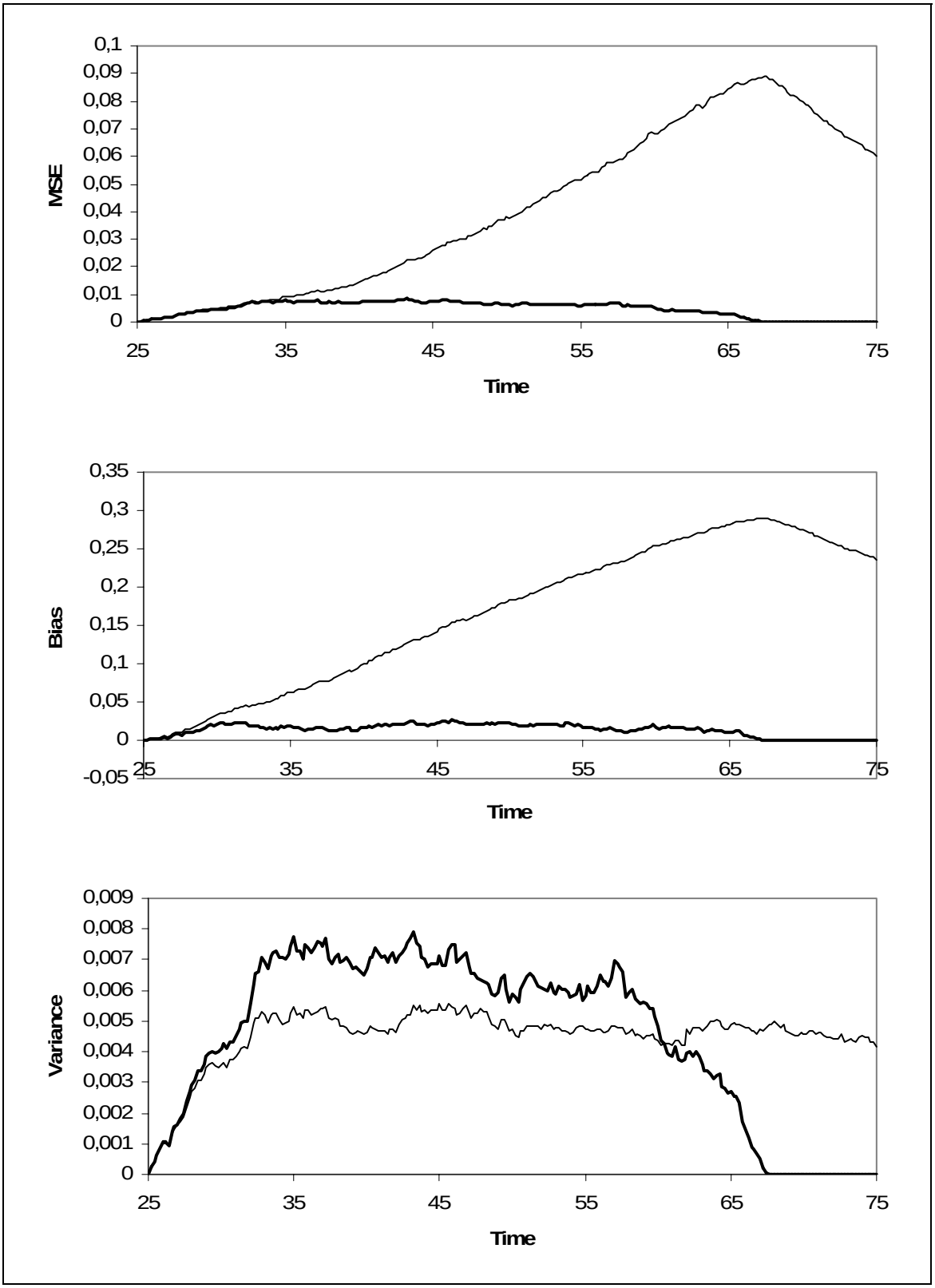


Figure A.25: Curves obtained for setting 3 with $s = 25$, $N = 200$ and 32% of censored observations under Aalen-Johansen estimator (solid line) and non-Markovian estimator (solid bold line).

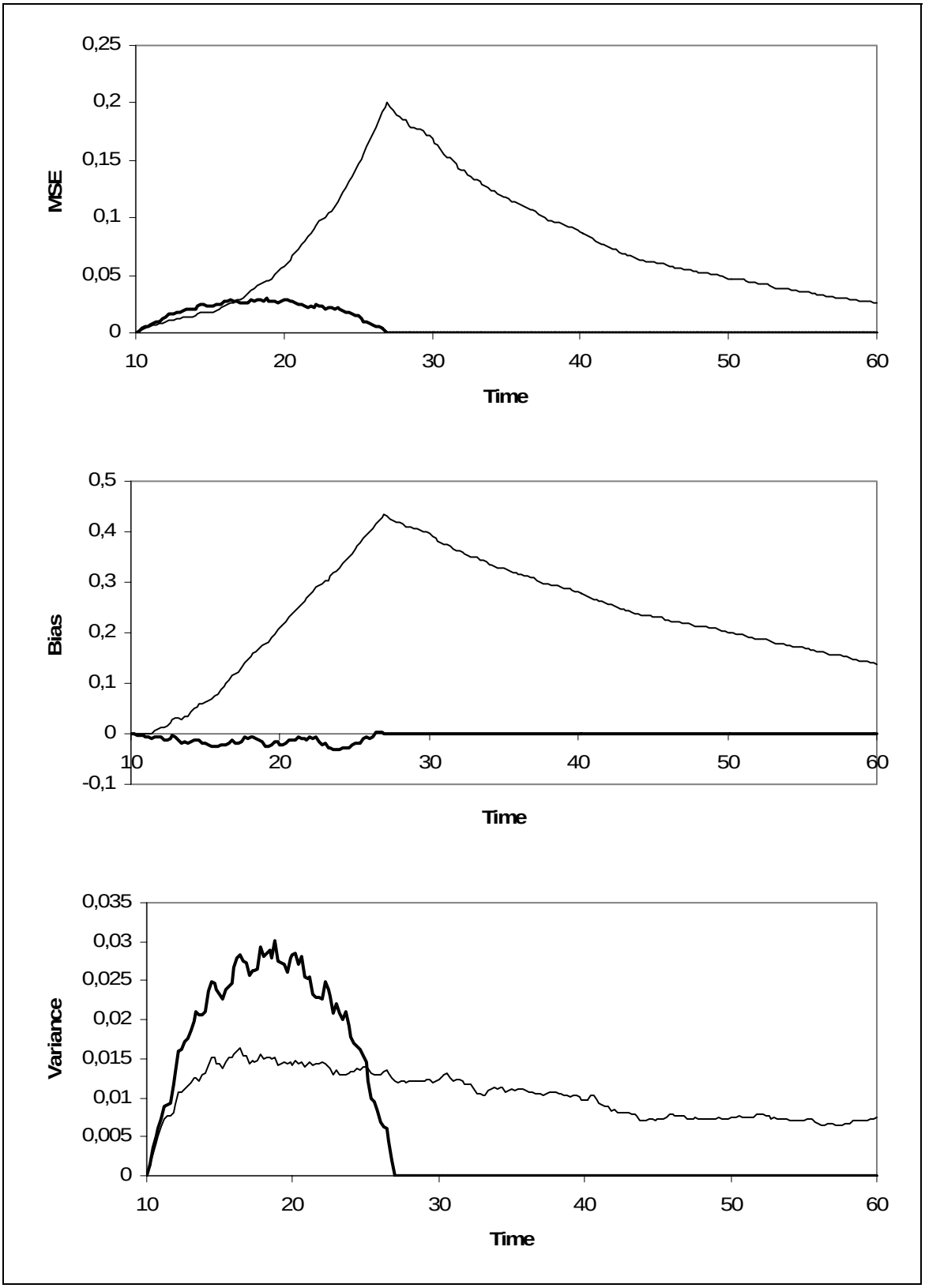


Figure A.26: Curves obtained for setting 3 with $s = 10$, $N = 100$ and 32% of censored observations under Aalen-Johansen estimator (solid line) and non-Markovian estimator (solid bold line).

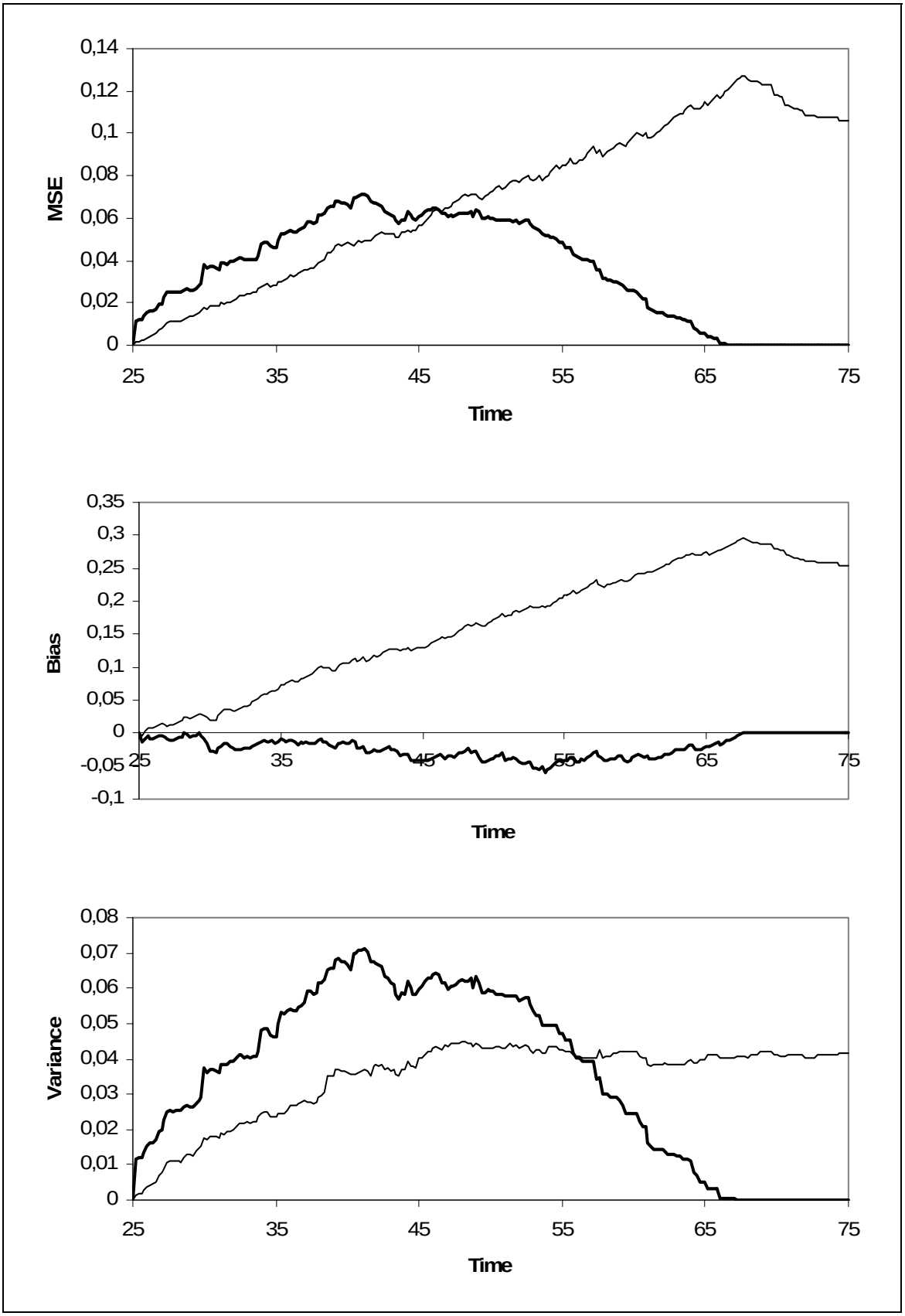


Figure A.27: Curves obtained for setting 3 with $s = 25$, $N = 100$ and 54% of censored observations under Aalen-Johansen estimator (solid line) and non-Markovian estimator (solid bold line).

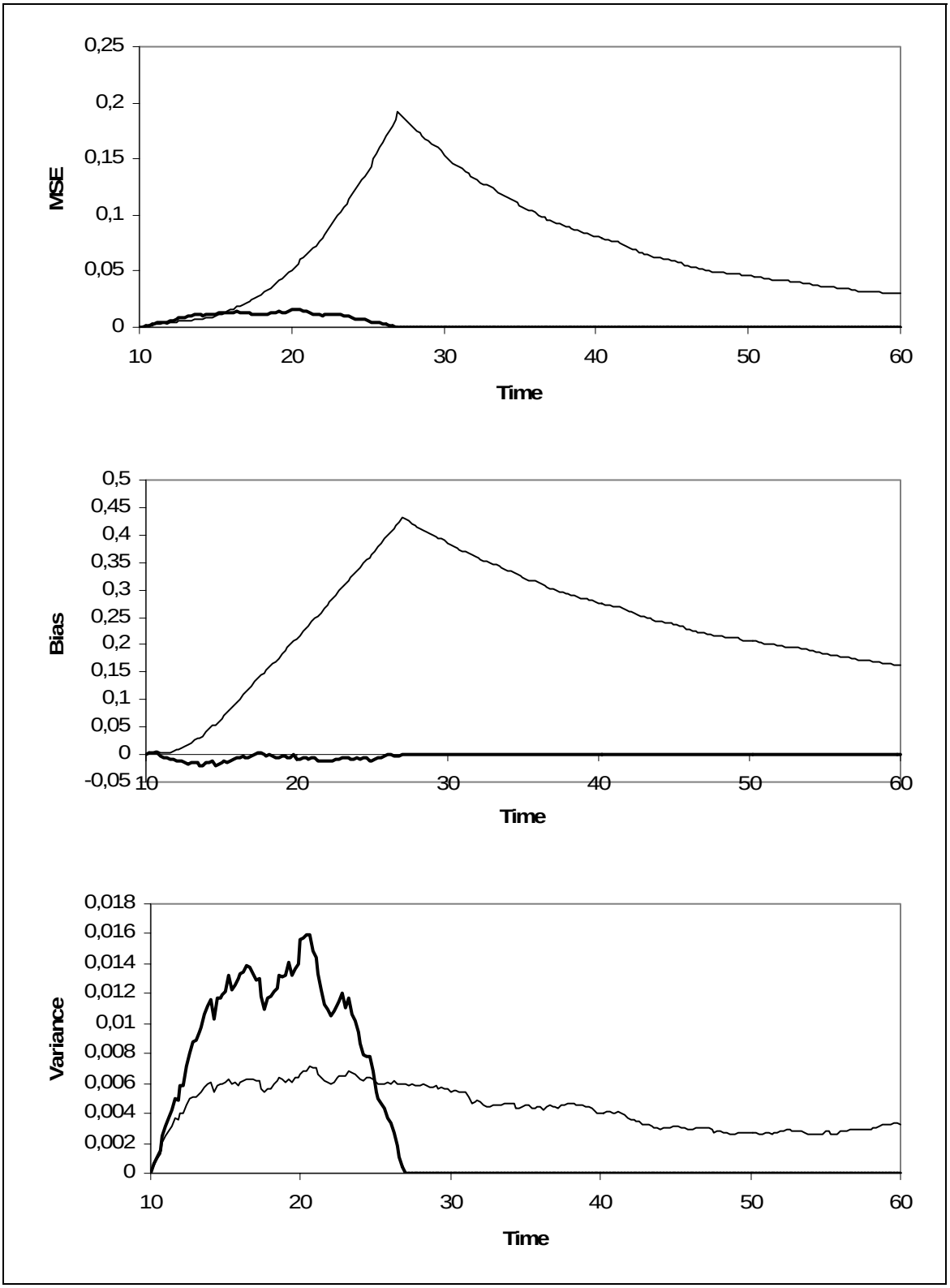


Figure A.28: Curves obtained for setting 3 with $s = 10$, $N = 200$ and 54% of censored observations under Aalen-Johansen estimator (solid line) and non-Markovian estimator (solid bold line).

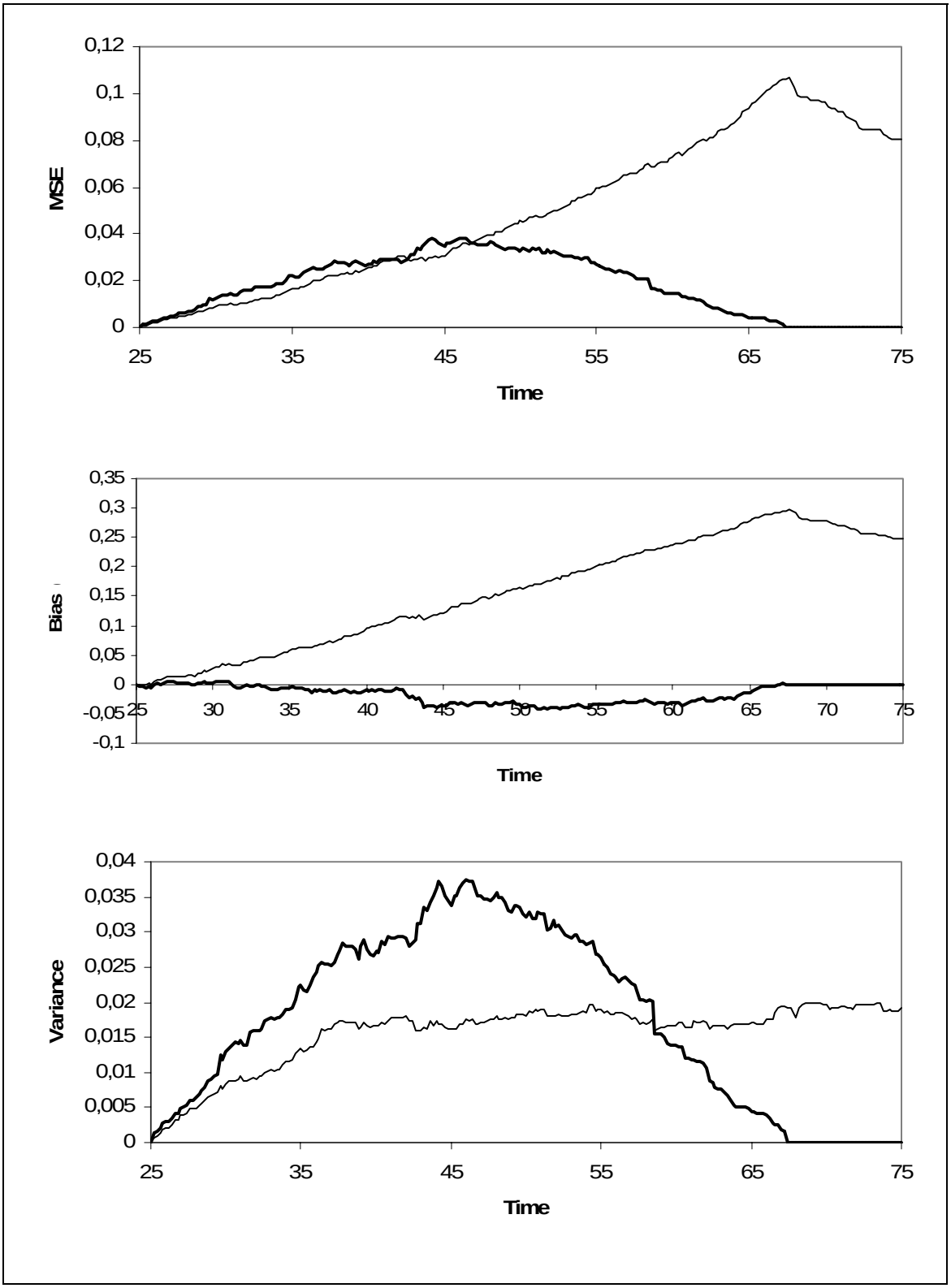


Figure A.29: Curves obtained for setting 3 with $s = 25$, $N = 200$ and 54% of censored observations under Aalen-Johansen estimator (solid line) and non-Markovian estimator (solid bold line).

B. Appendix of Chapter 5

B.1 Complete output for the Stanford Heart Transplantation study.

Command line:

tdc.surv(stanford,ncov=3,formula=c(6,7,8), model=c(1,0,0,0,0,0), surv.plot=TRUE, graphcov=1, plot.trans=TRUE)
Output:

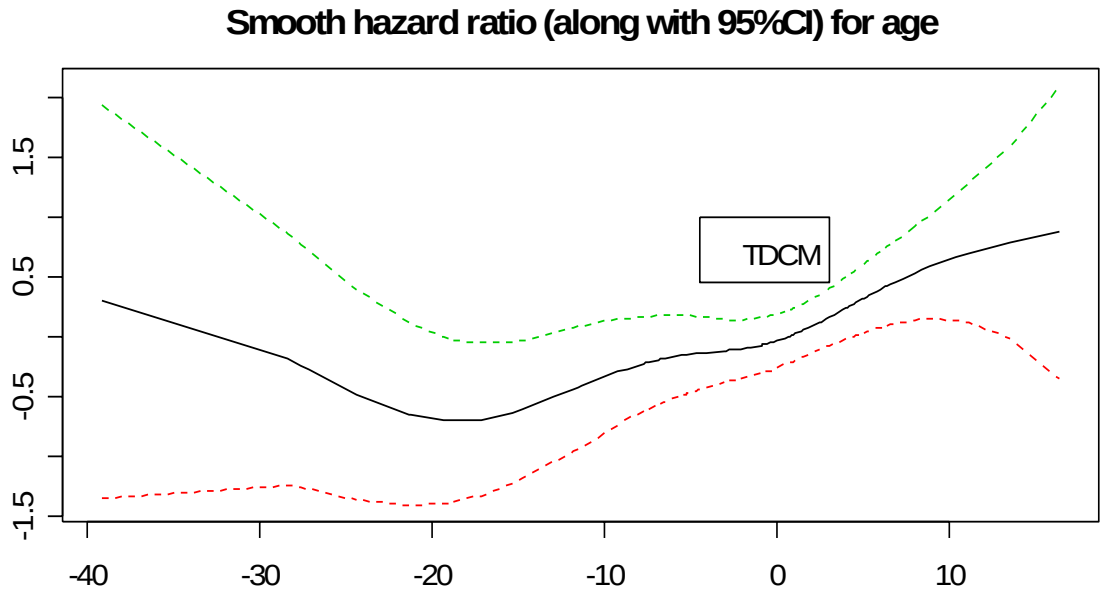
```
**********************************************************************
          ***** TIME-DEPENDENT COX REGRESSION MODEL *****
**********************************************************************
```

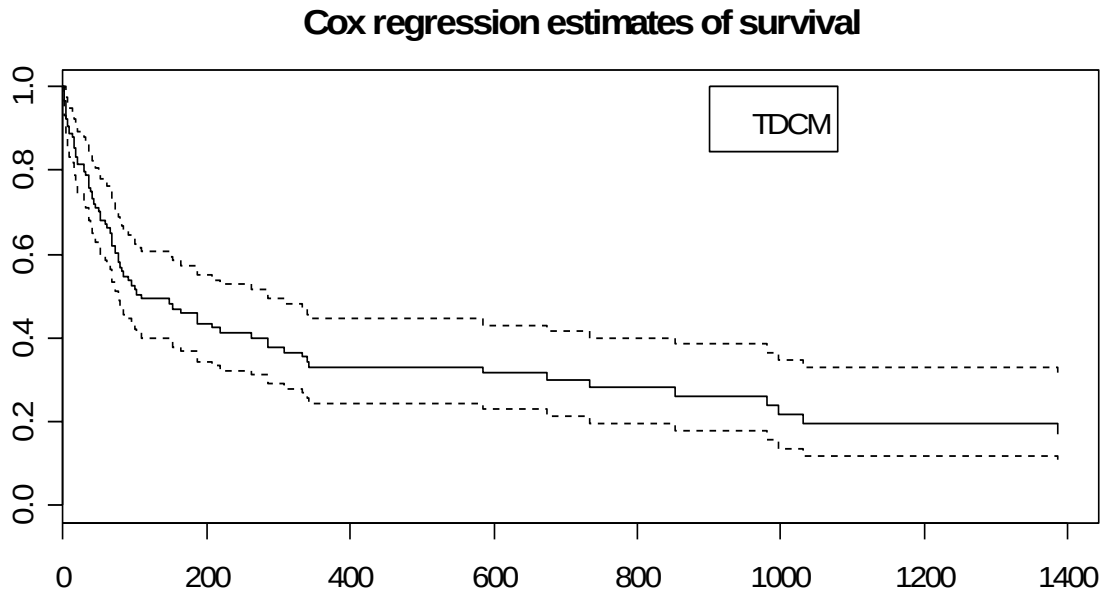
	coef	exp(coef)	se(coef)	z	p
Age	0.0271	1.027	0.0134	2.02	0.043
Year	-0.1463	0.864	0.0704	-2.08	0.038
Surgery	-0.6376	0.529	0.3670	-1.74	0.082

	exp(coef)	exp(-coef)	lower.95	upper.95
age	1.027	0.973	1.001	1.055
year	0.864	1.158	0.753	0.992
surgery	0.529	1.892	0.257	1.085

Rsquare= 0.084 (max possible= 0.969)
Likelihood ratio test= 15.1 on 3 df, p=0.00172
Wald test = 14.5 on 3 df, p=0.00230
Score (logrank) test = 15.0 on 3 df, p=0.00179

-2*Log-likelihood: 581.1323


Smooth hazard ratio (along with 95%CI) for age
TDCM
1.5
0.5
-0.5
-1.5
-40
-30
-20
-10
0
10


Cox regression estimates of survival
TDCM
1.0
0.8
0.6
0.4
0.2
0.0
0
200
400
600
800
1000
1200
1400

Command line:

tdc.surv(stanford, ncov=3,formula=c(6,7,8), model=c(0,1,0,0,0,0), surv.plot=TRUE, graphcov=1, plot.trans=TRUE)

Output:

```
*******************************************************************************
***** COX MARKOV MODEL. SEPARATED FITTED COX MODELS FOR EACH TRANSITION *****
*******************************************************************************


***** FROM STATE 1 TO STATE 2  *****
n= 103

          coef      exp(coef)  se(coef)  z       p
age       0.03111   1.03       0.0140    2.2255  0.026
year      0.00075   1.00       0.0695    0.0108  0.990
surgery   0.04734   1.05       0.3152    0.1502  0.880

          exp(coef) exp(-coef) lower.95  upper.95
age       1.03      0.969      1.004     1.06
year      1.00      0.999      0.873     1.15
surgery   1.05      0.954      0.565     1.94


Rsquare= 0.054   (max possible= 0.993 )
Likelihood ratio test= 5.77  on 3 df,   p=0.123
Wald test            = 5  on 3 df,   p=0.172
Score (logrank) test = 5.06  on 3 df,   p=0.167

-2*Log-likelihood: 509.5638

***** FROM STATE 1 TO STATE 3  *****
n= 103

          coef      exp(coef)  se(coef)  z       p
age       0.0198    1.020      0.0181    1.094   0.270
year      -0.2833   0.753      0.1110    -2.553  0.011
surgery   -0.2288   0.796      0.6361    -0.360  0.720

          exp(coef) exp(-coef) lower.95  upper.95
age       1.020     0.98       0.984     1.057
year      0.753     1.33       0.606     0.936
surgery   0.796     1.26       0.229     2.768

Rsquare= 0.08   (max possible= 0.886 )
Likelihood ratio test= 8.62  on 3 df,   p=0.0347
Wald test            = 8.19  on 3 df,   p=0.0422
Score (logrank) test = 8.67  on 3 df,   p=0.034

-2*Log-likelihood: 214.9848
```

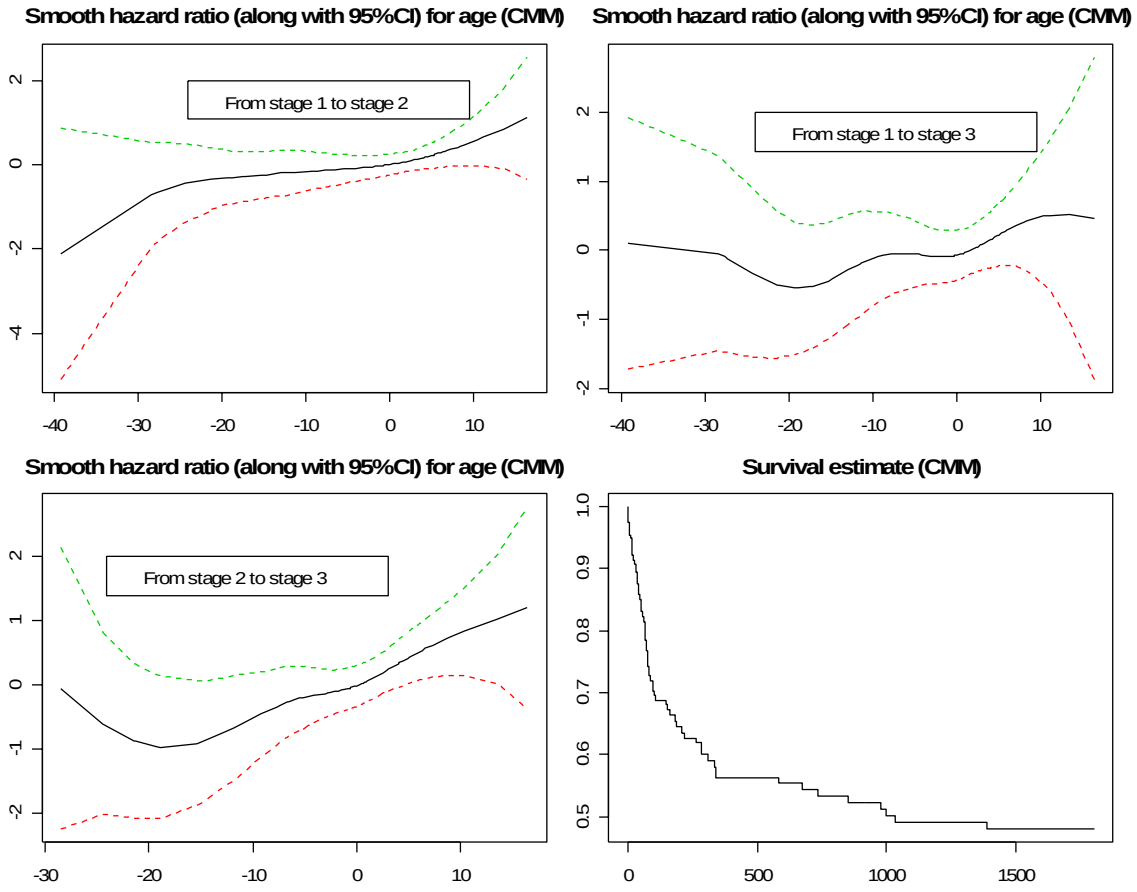
***** FROM STATE 2 TO STATE 3 *****
n= 69

	coef	exp(coef)	se(coef)	z	p
age	0.0496	1.051	0.0214	2.318	0.020
year	-0.0230	0.977	0.0969	-0.238	0.810
surgery	-0.8165	0.442	0.4549	-1.795	0.073

	exp(coef)	exp(-coef)	lower.95	upper.95
age	1.051	0.952	1.008	1.10
year	0.977	1.023	0.808	1.18
surgery	0.442	2.263	0.181	1.08

Rsquare= 0.151 (max possible= 0.987)
Likelihood ratio test= 11.3 on 3 df, p=0.0102
Wald test = 10.2 on 3 df, p=0.0167
Score (logrank) test = 10.7 on 3 df, p=0.0137

-2*Log-likelihood: 290.1922



Command line:

tdc.surv(stanford, ncov=3, formula=c(6,7,8), model=c(0,0,1,0,0,0), surv.plot=TRUE, graphcov=1, plot.trans=TRUE)

Output:

```
*****************************************************************************
          ***** MULTI-STATE HOMOGENEOUS MARKOV MODEL *****
*****************************************************************************

***** Undergoing Transitions *****
      to
from      1   2   3
     1    4  69  30
     2    0  24  45


-2*Log-likelihood: 1727.033

***** Estimated coefficients *****
$logbaseline
          Stage 1  Stage 2  Stage 3
Stage 1   0.00000  -4.29183 -5.22161
Stage 2   0.00000  0.00000  -6.34132
Stage 3   0.00000  0.00000  0.00000

 $age
          Stage 1  Stage 2  Stage 3
Stage 1   0.00000  0.06581  0.05447
Stage 2   0.00000  0.00000  0.07356
Stage 3   0.00000  0.00000  0.00000

$year
          Stage 1  Stage 2  Stage 3
Stage 1   0.00000  -0.02495 -0.30205
Stage 2   0.00000  0.00000  0.10324
Stage 3   0.00000  0.00000  0.00000

$surgery
          Stage 1  Stage 2  Stage 3
Stage 1   0.00000  0.31318  -0.04225
Stage 2   0.00000  0.00000  -1.18467
Stage 3   0.00000  0.00000  0.00000

 $baseline
          Stage 1  Stage 2  Stage 3
Stage 1   -0.01907 0.01367  0.00539
Stage 2   0.00000  -0.00176 0.00176
Stage 3   0.00000  0.00000  0.00000

$age
                      HR       L95      U95
Stage 1 - Stage 2  1.06803  1.03914  1.09772
Stage 1 - Stage 3  1.05598  1.01993  1.09331
Stage 2 - Stage 4  1.07633  1.03021  1.12452
```

$year

	HR	L95	U95
Stage 1 - Stage 2	0.97535	0.85212	1.11639
Stage 1 - Stage 3	0.73929	0.59477	0.91892
Stage 2 - Stage 4	1.10876	0.92804	1.32467

$surgery

	HR	L95	U95
Stage 1 - Stage 2	1.36776	0.73678	2.53911
Stage 1 - Stage 3	0.95862	0.27721	3.31492
Stage 2 - Stage 4	0.30584	0.12812	0.73005

***** Estimate of the ratio of the progression rate 1-3 *****
***** into death to the corresponding rate 2-3 *****
Estimate: 3.063975
SE: 0.7976103

**** WALDS TEST TO CHECK FOR DIFFERENCES BETWEEN MORTALITY INTENSITIES ****
LET H0 BE THE HYPOTHESIS OF NON-DIFFERENCES BETWEEN MORTALITY INTENSITIES
H0 produces a chisquare statistic of: 18.50131
with a p-value of: 1.697872e-05

**** WALDS TEST TO CHECK FOR DIFFERENCES BETWEEN COVARIATE 1 IN THE TRANSITION INTENSITIES ****
LET H0 BE THE HYPOTHESIS OF NON-DIFFERENCES BETWEEN COVARIATE 1 IN THE TRANSITION INTENSITIES
H0 produces a chisquare statistic of: 0.4884893
with a p-value of: 0.783296

**** WALDS TEST TO CHECK FOR DIFFERENCES BETWEEN COVARIATE 2 IN THE TRANSITION INTENSITIES ****
LET H0 BE THE HYPOTHESIS OF NON-DIFFERENCES BETWEEN COVARIATE 2 IN THE TRANSITION INTENSITIES
H0 produces a chisquare statistic of: 8.112579
with a p-value of: 0.01731314

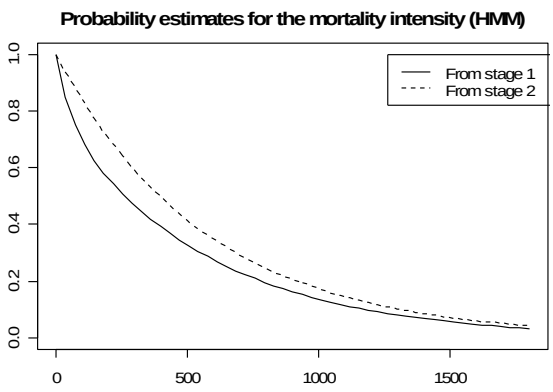
**** WALDS TEST TO CHECK FOR DIFFERENCES BETWEEN COVARIATE 3 IN THE TRANSITION INTENSITIES ****
LET H0 BE THE HYPOTHESIS OF NON-DIFFERENCES BETWEEN COVARIATE 3 IN THE TRANSITION INTENSITIES
H0 produces a chisquare statistic of: 7.608883
with a p-value of: 0.02227163

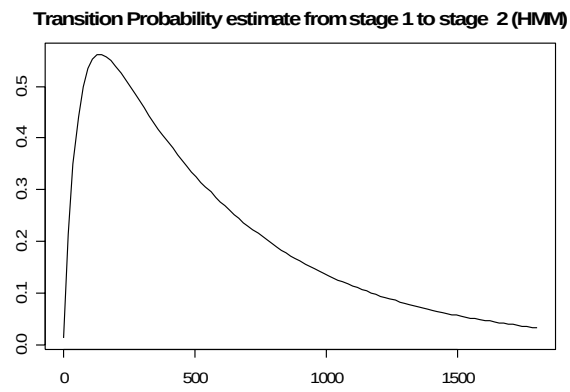
***** Sojourn times *****
$estimate

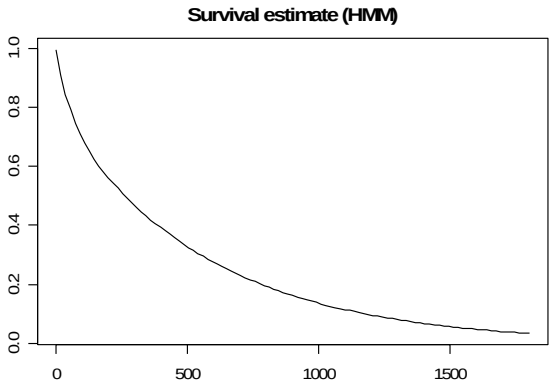
Stage 1	Stage 2
52.41532	567.54807

$SE

Stage 1	Stage 2
5.454848	93.233470







Command line:

tdc.surv(stanford,ncov=3, formula=c(6,7,8), model=c(0,0,0,1,0,0), surv.plot=TRUE, graphcov=1, cut=90, plot.trans=TRUE)

Output:

```
**************************************************************************
          ***** MULTI-STATE NON-HOMOGENEOUS MODEL *****
**************************************************************************

FIRST BRANCH

-2*Log-likelihood: 1251.956

***** Estimated coefficients *****
$logbaseline
          Stage 1   Stage 2   Stage 3
Stage 1   0.00000  -3.94184  -4.99322
Stage 2   0.00000   0.00000  -5.32174
Stage 3   0.00000   0.00000   0.00000

$age
          Stage 1   Stage 2   Stage 3
Stage 1   0.00000   0.02836   0.02265
Stage 2   0.00000   0.00000   0.05098
Stage 3   0.00000   0.00000   0.00000
```

$year

	Stage 1	Stage 2	Stage 3
Stage 1	0.00000	-0.02773	-0.32448
Stage 2	0.00000	0.00000	-0.02554
Stage 3	0.00000	0.00000	0.00000

$surgery

	Stage 1	Stage 2	Stage 3
Stage 1	0.00000	0.04078	-0.17449
Stage 2	0.00000	0.00000	-1.96772
Stage 3	0.00000	0.00000	0.00000

$baseline

	Stage 1	Stage 2	Stage 3
Stage 1	-0.02619	0.01941	0.00678
Stage 2	0.00000	-0.00488	0.00488
Stage 3	0.00000	0.00000	0.00000

$age

	HR	L95	U95
Stage 1 - Stage 2	1.02876	1.00070	1.05761
Stage 1 - Stage 3	1.02291	0.98461	1.06269
Stage 2 - Stage 4	1.05230	1.00008	1.10724

$year

	HR	L95	U95
Stage 1 - Stage 2	0.97264	0.84709	1.11679
Stage 1 - Stage 3	0.72290	0.57169	0.91410
Stage 2 - Stage 4	0.97477	0.78384	1.21221

$surgery

	HR	L95	U95
Stage 1 - Stage 2	1.04162	0.54989	1.97309
Stage 1 - Stage 3	0.83987	0.23966	2.94325
Stage 2 - Stage 4	0.13977	0.01881	1.03844

***** Estimate of the ratio of the progression rate 1-3 *****
***** into death to the corresponding rate 2-3 *****
Estimate: 1.388920
SE: 0.4672443

**** WALDS TEST TO CHECK FOR DIFFERENCES BETWEEN MORTALITY INTENSITIES ****
LET H0 BE THE HYPOTHESIS OF NON-DIFFERENCES BETWEEN MORTALITY INTENSITIES
H0 produces a chisquare statistic of: 0.9536886
with a p-value of: 0.3287822

**** WALDS TEST TO CHECK FOR DIFFERENCES BETWEEN COVARIATE 1 IN THE TRANSITION INTENSITIES ****
LET H0 BE THE HYPOTHESIS OF NON-DIFFERENCES BETWEEN COVARIATE 1 IN THE TRANSITION INTENSITIES
H0 produces a chisquare statistic of: 0.8074933
with a p-value of: 0.6678133

**** WALDS TEST TO CHECK FOR DIFFERENCES BETWEEN COVARIATE 2 IN THE TRANSITION INTENSITIES ****
LET H0 BE THE HYPOTHESIS OF NON-DIFFERENCES BETWEEN COVARIATE 2 IN THE TRANSITION INTENSITIES
H0 produces a chisquare statistic of: 4.945467
with a p-value of: 0.08435396

**** WALDS TEST TO CHECK FOR DIFFERENCES BETWEEN COVARIATE 3 IN THE TRANSITION INTENSITIES ****
LET H0 BE THE HYPOTHESIS OF NON-DIFFERENCES BETWEEN COVARIATE 3 IN THE TRANSITION INTENSITIES
H0 produces a chisquare statistic of: 3.500048
with a p-value of: 0.1737698

***** Sojourn times *****
$estimate

Stage 1	Stage 2
38.17357	204.74166

$SE

Stage 1	Stage 2
4.20200	50.88298

LAST BRANCH

*** convergence ***

-2*Log-likelihood: 392.1713

***** Estimated coefficients *****
$logbaseline

	Stage 1	Stage 2	Stage 3
Stage 1	0.00000	-9.42824	-7.92574
Stage 2	0.00000	0.00000	-7.32674
Stage 3	0.00000	0.00000	0.00000

$age

	Stage 1	Stage 2	Stage 3
Stage 1	0.00000	-0.05918	-0.03598
Stage 2	0.00000	0.00000	0.04552
Stage 3	0.00000	0.00000	0.00000

$year

	Stage 1	Stage 2	Stage 3
Stage 1	0.00000	2.76868	0.39456
Stage 2	0.00000	0.00000	-0.03432
Stage 3	0.00000	0.00000	0.00000

$surgery

	Stage 1	Stage 2	Stage 3
Stage 1	0.00000	1.29376	-5.42610
Stage 2	0.00000	0.00000	-0.23792
Stage 3	0.00000	0.00000	0.00000

$baseline

	Stage 1	Stage 2	Stage 3
Stage 1	-0.00044	0.00008	0.00036
Stage 2	0.00000	-0.00065	0.00065
Stage 3	0.00000	0.00000	0.00000

$age

	HR	L95	U95
Stage 1 - Stage 2	0.94253	0.80882	1.09835
Stage 1 - Stage 3	0.96465	0.84461	1.10175
Stage 2 - Stage 4	1.04657	0.97389	1.12468

$year

	HR	L95	U95
Stage 1 - Stage 2	15.93763	0.81377	312.1337
Stage 1 - Stage 3	1.48374	0.62133	3.54317
Stage 2 - Stage 4	0.96626	0.68381	1.36536

$surgery

	HR	L95	U95
Stage 1 - Stage 2	3.64650	2.937e-01	4.526e+01
Stage 1 - Stage 3	0.00440	1.024e-37	1.889e+32
Stage 2 - Stage 4	0.78825	2.637e-01	2.35627

***** Estimate of the ratio of the progression rate 1-3 *****
***** into death to the corresponding rate 2-3 *****
Estimate: 0.5493616
SE: 5.539668

**** WALDS TEST TO CHECK FOR DIFFERENCES BETWEEN MORTALITY INTENSITIES ****
LET H0 BE THE HYPOTHESIS OF NON-DIFFERENCES BETWEEN MORTALITY INTENSITIES
H0 produces a chisquare statistic of: 0.003528584
with a p-value of: 0.952632

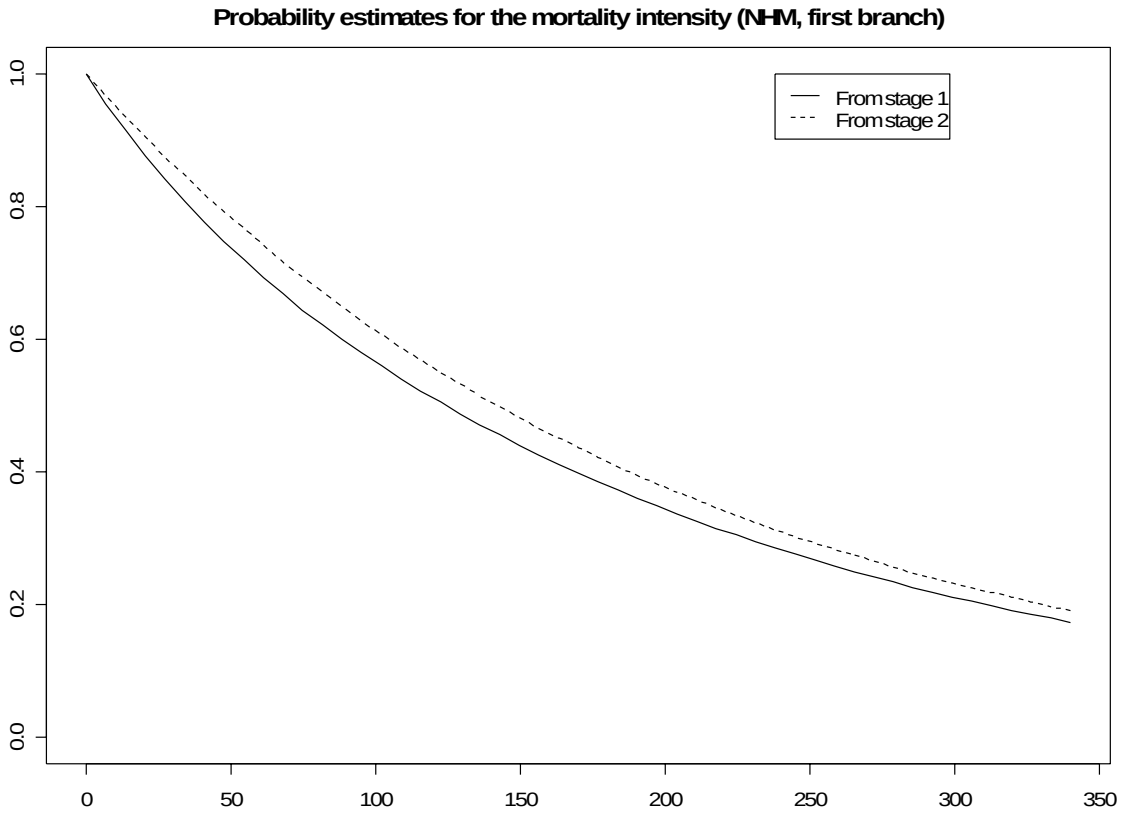
**** WALDS TEST TO CHECK FOR DIFFERENCES BETWEEN COVARIATE 1 IN THE TRANSITION INTENSITIES ****
LET H0 BE THE HYPOTHESIS OF NON-DIFFERENCES BETWEEN COVARIATE 1 IN THE TRANSITION INTENSITIES
H0 produces a chisquare statistic of: 2.159090
with a p-value of: 0.3397500

**** WALDS TEST TO CHECK FOR DIFFERENCES BETWEEN COVARIATE 2 IN THE TRANSITION INTENSITIES ****
LET H0 BE THE HYPOTHESIS OF NON-DIFFERENCES BETWEEN COVARIATE 2 IN THE TRANSITION INTENSITIES
H0 produces a chisquare statistic of: 4.037749
with a p-value of: 0.1328048

**** WALDS TEST TO CHECK FOR DIFFERENCES BETWEEN COVARIATE 3 IN THE TRANSITION INTENSITIES ****
LET H0 BE THE HYPOTHESIS OF NON-DIFFERENCES BETWEEN COVARIATE 3 IN THE TRANSITION INTENSITIES
H0 produces a chisquare statistic of: 1.212583
with a p-value of: 0.5453697

***** Sojourn times *****
$estimate
 Stage 1 Stage 2
2263.767 1520.424

$SE
  Stage 1   Stage 2
18698.422 486.838



Command line:

tdc.surv(stanford, ncov=3, formula=c(6,7,8), model=c(0,0,0,0,1,0), surv.plot=TRUE, graphcov=1, plot.trans=TRUE)

****
*** COX SEMI-MARKOV MODEL. SEPARATED FITTED COX MODELS FOR EACH TRANSITION ***
****

***** FROM STATE 1 TO STATE 2 *****
n= 103

	coef	exp(coef)	se(coef)	z	p
age	0.03111	1.03	0.0140	2.2255	0.026
year	0.00075	1.00	0.0695	0.0108	0.990
surgery	0.04734	1.05	0.3152	0.1502	0.880

	exp(coef)	exp(-coef)	lower.95	upper.95
age	1.03	0.969	1.004	1.06
year	1.00	0.999	0.873	1.15
surgery	1.05	0.954	0.565	1.94

Rsquare= 0.054 (max possible= 0.993)
Likelihood ratio test= 5.77 on 3 df, p=0.123
Wald test = 5 on 3 df, p=0.172
Score (logrank) test = 5.06 on 3 df, p=0.167

-2*Log-likelihood: 509.5638
***** FROM STATE 1 TO STATE 3 *****
n= 103

	coef	exp(coef)	se(coef)	z	p
age	0.0198	1.020	0.0181	1.094	0.270
year	-0.2833	0.753	0.1110	-2.553	0.011
surgery	-0.2288	0.796	0.6361	-0.360	0.720

	exp(coef)	exp(-coef)	lower.95	upper.95
age	1.020	0.98	0.984	1.057
year	0.753	1.33	0.606	0.936
surgery	0.796	1.26	0.229	2.768

Rsquare= 0.08 (max possible= 0.886)
Likelihood ratio test= 8.62 on 3 df, p=0.0347
Wald test = 8.19 on 3 df, p=0.0422
Score (logrank) test = 8.67 on 3 df, p=0.034

-2*Log-likelihood: 214.9848

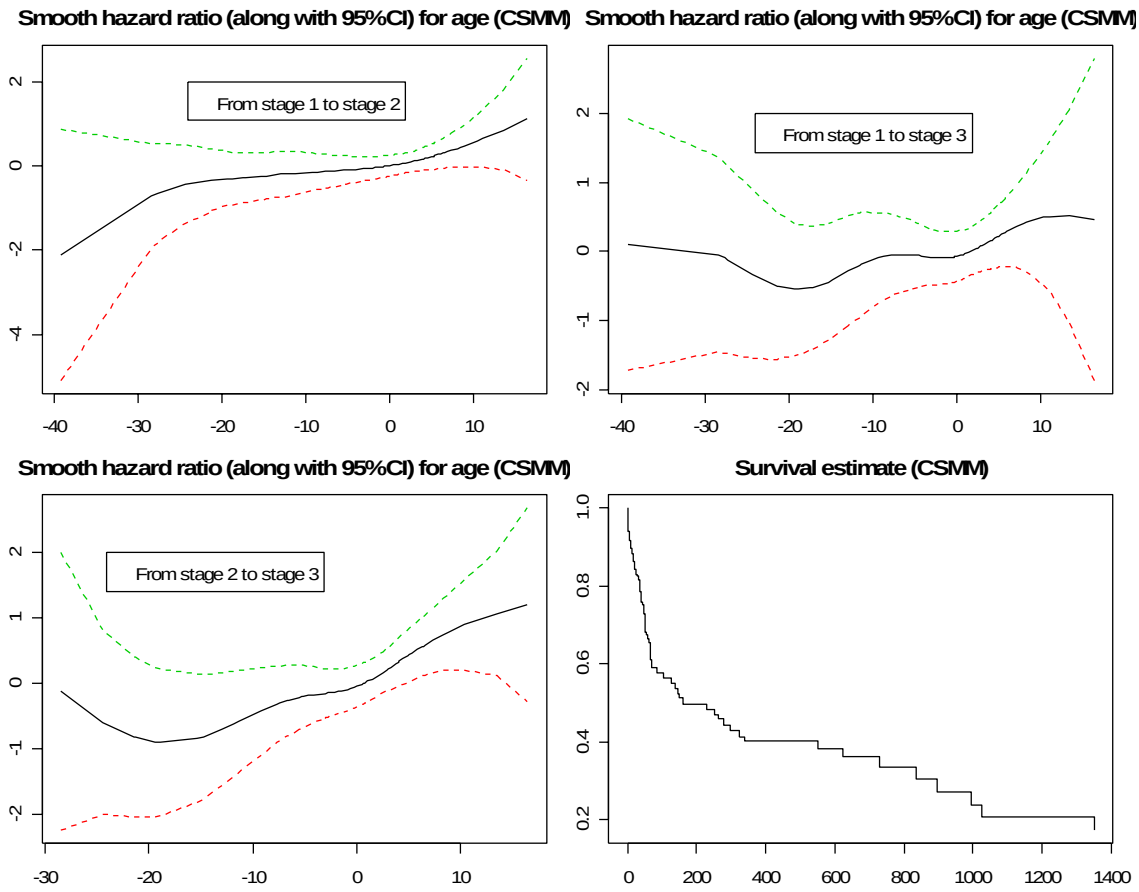
***** FROM STATE 2 TO STATE 3 *****
n= 69

	coef	exp(coef)	se(coef)	z	p
age	0.0477	1.049	0.0217	2.203	0.028
year	-0.0343	0.966	0.0966	-0.355	0.720
surgery	-0.8512	0.427	0.4522	-1.882	0.060

	exp(coef)	exp(-coef)	lower.95	upper.95
age	1.049	0.953	1.005	1.09
year	0.966	1.035	0.800	1.17
surgery	0.427	2.343	0.176	1.04

Rsquare= 0.154 (max possible= 0.991)
Likelihood ratio test= 11.6 on 3 df, p=0.00906
Wald test = 10.4 on 3 df, p=0.0155
Score (logrank) test = 10.9 on 3 df, p=0.0124

-2*Log-likelihood: 310.9225



Command line:

tdc.surv(stanford, ncov=3, formula=c(6,7,8), model=c(0,0,0,0,0,1), surv.plot=TRUE, graphcov=1, plot.trans=TRUE)

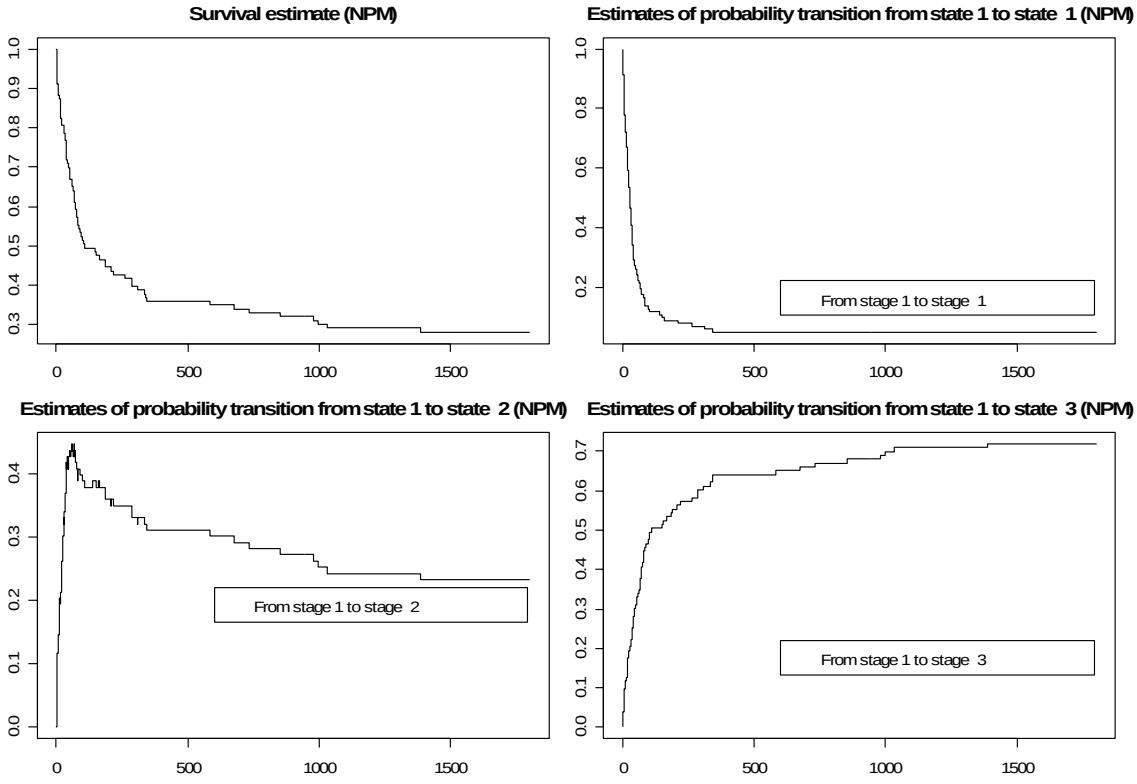
Output:

```
*****************************************************************************
                ***** NON-PARAMETRIC MULTI-STATE MODEL *****
*****************************************************************************
```

Times	Survival	P11	P12	P13
0	1	1	0	0
1	0,990291	0,970874	0,019417	0,009709
2	0,961165	0,912621	0,048544	0,038835
3	0,932039	0,854369	0,07767	0,067961
4	0,932039	0,834951	0,097087	0,067961
4,5	0,932039	0,825243	0,106796	0,067961
5	0,912621	0,796117	0,116505	0,087379
6	0,902913	0,776699	0,126214	0,097087
8	0,893204	0,747573	0,145631	0,106796
9	0,883495	0,737864	0,145631	0,116505
10	0,883495	0,718447	0,165049	0,116505

11	0,883495	0,718447	0,165049	0,116505
12	0,873786	0,679612	0,194175	0,126214
13	0,873786	0,669903	0,203884	0,126214
16	0,84466	0,650485	0,194175	0,15534
17	0,834952	0,631068	0,203884	0,165049
18	0,825243	0,61165	0,213592	0,174757
19	0,825243	0,601942	0,223301	0,174757
20	0,825243	0,592233	0,23301	0,174757
21	0,805825	0,543689	0,262136	0,194175
23	0,805825	0,533981	0,271845	0,194175
25	0,805825	0,524272	0,281553	0,194175
26	0,805825	0,504854	0,300971	0,194175
27	0,805825	0,485437	0,320388	0,194175
28	0,796117	0,466019	0,330097	0,203883
30	0,786408	0,466019	0,320388	0,213592
31	0,786408	0,446602	0,339806	0,213592
32	0,776699	0,417476	0,359223	0,223301
33	0,776699	0,407767	0,368932	0,223301
35	0,76699	0,398058	0,368932	0,23301
36	0,757282	0,368932	0,38835	0,242718
37	0,747573	0,339806	0,407767	0,252427
38	0,747573	0,330097	0,417476	0,252427
39	0,737864	0,330097	0,407767	0,262136
40	0,718447	0,31068	0,407767	0,281553
41	0,718447	0,291262	0,427184	0,281553
43	0,708738	0,291262	0,417476	0,291262
45	0,699029	0,291262	0,407767	0,300971
46	0,699029	0,271845	0,427184	0,300971
50	0,68932	0,262136	0,427184	0,31068
51	0,679612	0,242718	0,436893	0,320388
53	0,669903	0,242718	0,427184	0,330097
57	0,669903	0,23301	0,436893	0,330097
58	0,660194	0,223301	0,436893	0,339806
60	0,660194	0,213592	0,446602	0,339806
61	0,650485	0,213592	0,436893	0,349515
66	0,640777	0,213592	0,427184	0,359223
67	0,640777	0,194175	0,446602	0,359223
68	0,621359	0,194175	0,427184	0,378641
69	0,611651	0,184466	0,427184	0,38835
71	0,611651	0,174757	0,436893	0,38835
72	0,592233	0,174757	0,417476	0,407767
77	0,582524	0,174757	0,407767	0,417476
78	0,572816	0,165049	0,407767	0,427184
80	0,563107	0,165049	0,398058	0,436893
81	0,553398	0,165049	0,38835	0,446602
83	0,553398	0,145631	0,407767	0,446602
85	0,543689	0,135922	0,407767	0,456311
90	0,533981	0,135922	0,398058	0,466019
96	0,524272	0,126214	0,398058	0,475728
100	0,514563	0,126214	0,38835	0,485437
102	0,504854	0,116505	0,38835	0,495146

109	0,504854	0,116505	0,38835	0,495146
110	0,495146	0,116505	0,378641	0,504854
131	0,495146	0,116505	0,378641	0,504854
139	0,495146	0,106796	0,38835	0,504854
149	0,485437	0,097087	0,38835	0,514563
153	0,475728	0,097087	0,378641	0,524272
160	0,475728	0,087379	0,38835	0,524272
165	0,466019	0,087379	0,378641	0,533981
180	0,466019	0,087379	0,378641	0,533981
186	0,456311	0,087379	0,368932	0,543689
188	0,446602	0,087379	0,359223	0,553398
207	0,436893	0,087379	0,349515	0,563107
210	0,436893	0,07767	0,359223	0,563107
219	0,427185	0,07767	0,349515	0,572816
263	0,417476	0,067961	0,349515	0,582524
265	0,417476	0,067961	0,349515	0,582524
285	0,398058	0,067961	0,330097	0,601942
308	0,38835	0,067961	0,320388	0,61165
310	0,38835	0,058252	0,330097	0,61165
334	0,378641	0,058252	0,320388	0,621359
…	…	…	…	…
1800	0,281553	0,048544	0,23301	0,718447

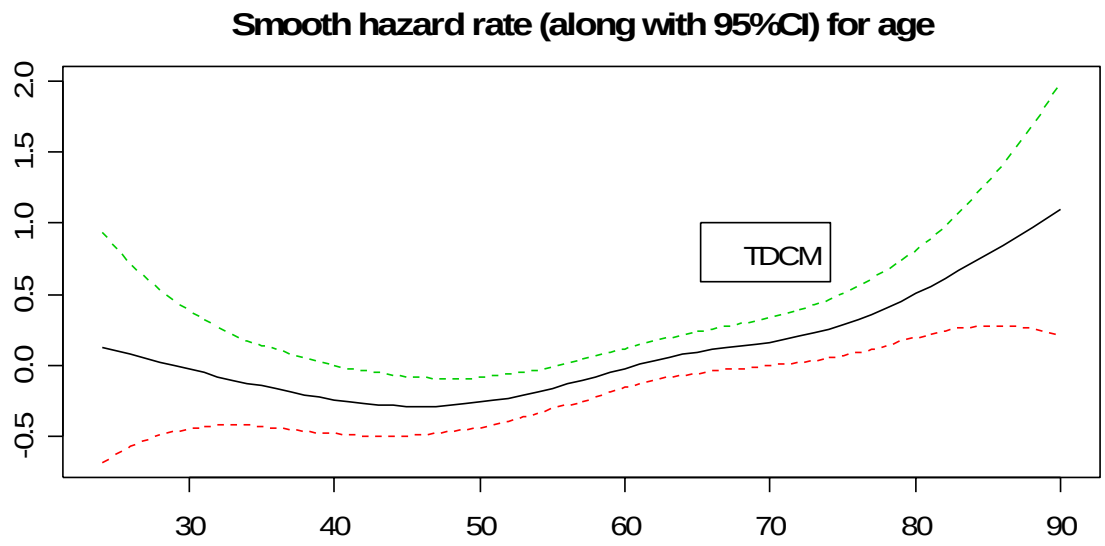


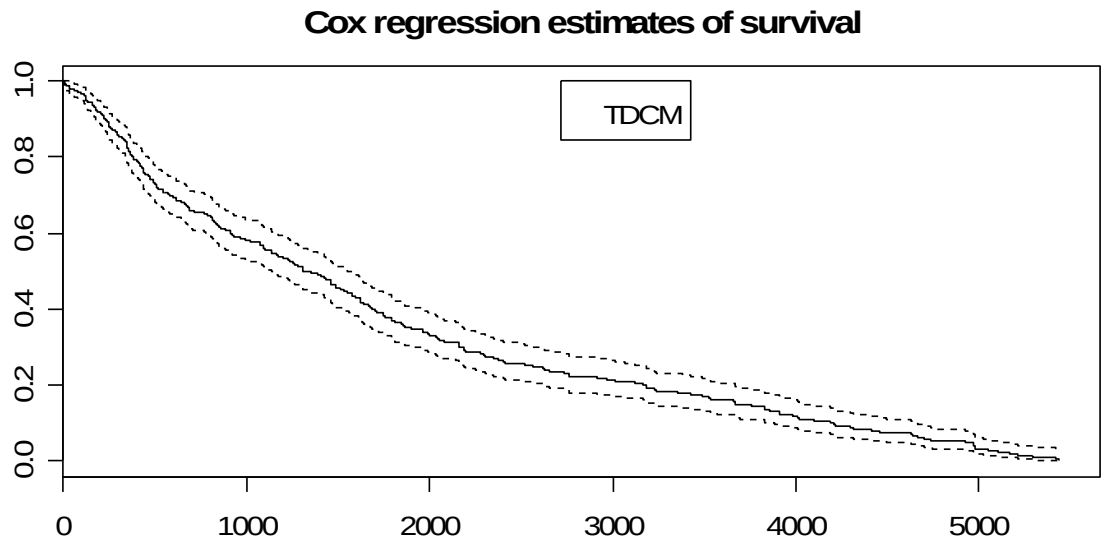
B.2 Complete output for the Stomach cancer study.

Command line:

tdc.surv(msmrecmet, ncov=2,formula=c(7,8), model=c(1,0,0,0,0,0), surv.plot=TRUE, graphcov=2, plot.trans=TRUE)

```
**********************************************************************
          ***** TIME-DEPENDENT COX REGRESSION MODEL *****
**********************************************************************
n= 439

        coef      exp(coef)  se(coef)   z        p
sex     -0.0327   0.968      0.11397    -0.287   0.7700
age     0.0147    1.015      0.00463    3.185    0.0014


        exp(coef) exp(-coef) lower.95   upper.95
sex     0.968     1.033      0.774      1.21
age     1.015     0.985      1.006      1.02

Rsquare= 0.024   (max possible= 0.999 )
Likelihood ratio test= 10.9  on 2 df,   p=0.00435
Wald test            = 10.8  on 2 df,   p=0.00464
Score (logrank) test = 10.8  on 2 df,   p=0.00456

-2*Log-likelihood: 3090.679
```





Command line:

tdc.surv(msmrecmet, ncov=2,formula=c(7,8), model=c(0,1,0,0,0,0), surv.plot=TRUE, graphcov=2, plot.trans=TRUE)

```
*******************************************************************************
***** COX MARKOV MODEL. SEPARATED FITTED COX MODELS FOR EACH TRANSITION *****
*******************************************************************************


***** FROM STATE 1 TO STATE 2  *****
n= 345
```

	coef	exp(coef)	se(coef)	z	p
sex	0.02482	1.03	0.2791	0.089	0.93
age	0.00321	1.00	0.0107	0.300	0.76

	exp(coef)	exp(-coef)	lower.95	upper.95
sex	1.03	0.975	0.593	1.77
age	1.00	0.997	0.982	1.02

Rsquare= 0 (max possible= 0.807)
Likelihood ratio test= 0.09 on 2 df, p=0.955
Wald test = 0.09 on 2 df, p=0.955
Score (logrank) test = 0.09 on 2 df, p=0.955

-2*Log-likelihood: 567.5803

***** FROM STATE 1 TO STATE 3 *****
n= 345

	coef	exp(coef)	se(coef)	z	p
sex	0.6585	1.93	0.3391	1.94	0.052
age	0.0169	1.02	0.0127	1.33	0.180

	exp(coef)	exp(-coef)	lower.95	upper.95
sex	1.93	0.518	0.994	3.76
age	1.02	0.983	0.992	1.04

Rsquare= 0.015 (max possible= 0.716)
Likelihood ratio test= 5.21 on 2 df, p=0.074
Wald test = 4.9 on 2 df, p=0.0862
Score (logrank) test = 5 on 2 df, p=0.0822

-2*Log-likelihood: 428.8344

***** FROM STATE 1 TO STATE 4 *****
n= 345

	coef	exp(coef)	se(coef)	z	p
sex	-0.0351	0.966	0.12846	-0.273	0.78000
age	0.0174	1.018	0.00525	3.307	0.00094

	exp(coef)	exp(-coef)	lower.95	upper.95
sex	0.966	1.036	0.75	1.24
age	1.018	0.983	1.01	1.03

Rsquare= 0.034 (max possible= 0.999)
Likelihood ratio test= 11.8 on 2 df, p=0.0028
Wald test = 11.6 on 2 df, p=0.00308
Score (logrank) test = 11.6 on 2 df, p=0.003

-2*Log-likelihood: 2336.422

***** FROM STATE 2 TO STATE 4 *****
n= 53

	coef	exp(coef)	se(coef)	z	p
sex	-0.636464	0.529	0.3491	-1.82302	0.068
age	-0.000147	1.000	0.0172	-0.00855	0.990

	exp(coef)	exp(-coef)	lower.95	upper.95
sex	0.529	1.89	0.267	1.05
age	1.000	1.00	0.967	1.03

Rsquare= 0.062 (max possible= 0.961)
Likelihood ratio test= 3.38 on 2 df, p=0.185
Wald test = 3.34 on 2 df, p=0.188
Score (logrank) test = 3.44 on 2 df, p=0.179

-2*Log-likelihood: 168.3983

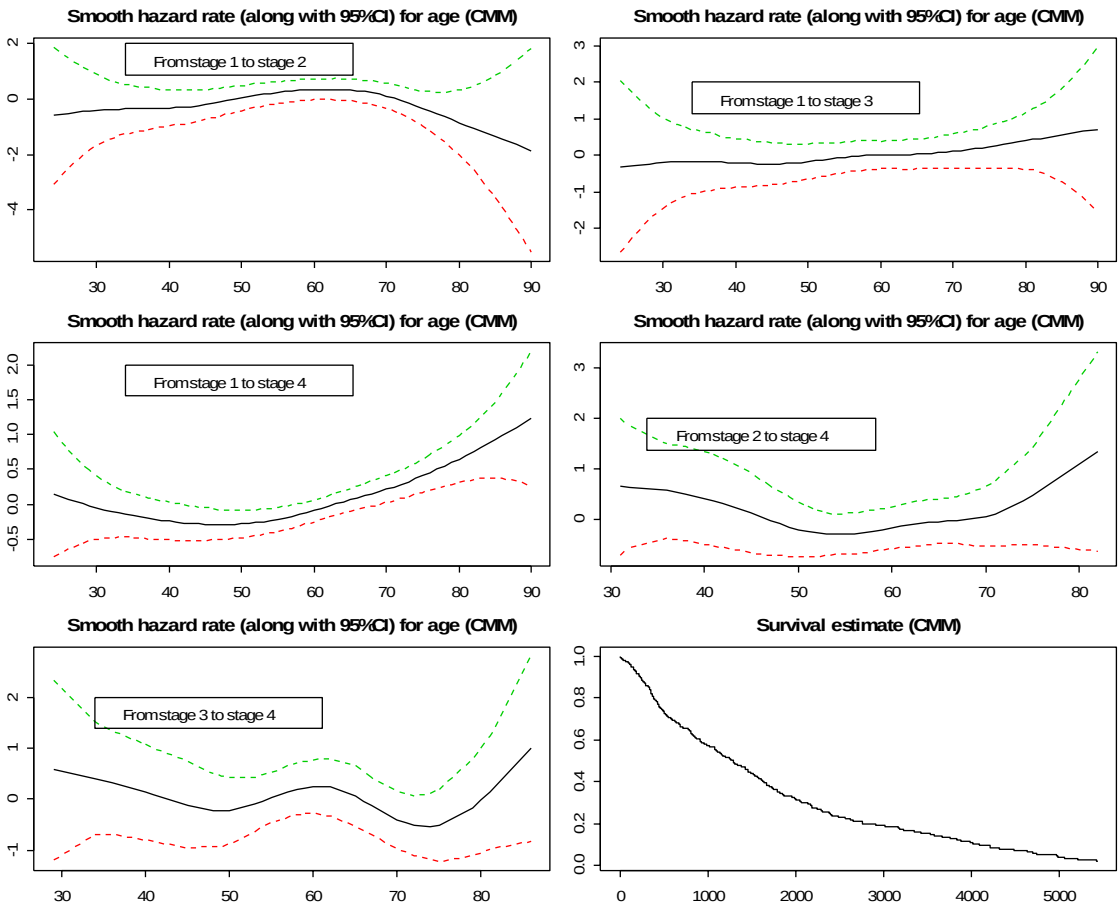
***** FROM STATE 3 TO STATE 4 *****
n= 41

	coef	exp(coef)	se(coef)	z	p
sex	0.59128	1.81	0.4233	1.397	0.16
age	-0.00926	0.99	0.0157	-0.591	0.55

	exp(coef)	exp(-coef)	lower.95	upper.95
sex	1.81	0.554	0.788	4.14
age	0.99	1.009	0.961	1.02

Rsquare= 0.061 (max possible= 0.949)
Likelihood ratio test= 2.57 on 2 df, p=0.277
Wald test = 2.45 on 2 df, p=0.294
Score (logrank) test = 2.51 on 2 df, p=0.285

-2*Log-likelihood: 119.7247



Command line:

tdc.surv(msmrecmet, ncov=2, formula=c(7,8), model=c(0,0,1,0,0,0), surv.plot=TRUE, graphcov=2, plot.trans=TRUE)

```
*********************************************************************************
                ***** MULTI-STATE HOMOGENEOUS MARKOV MODEL *****
*********************************************************************************

***** Undergoing Transitions *****
        to
from          1     2     3     4
           1  0    53    41   251
           2  0    15     0    38
           3  0     0    11    30


 *** convergence ***

-2*Log-likelihood: 7249.969
```

***** Estimated coefficients *****

$logbaseline

	Stage 1	Stage 2	Stage 3	Stage 4
Stage 1	0.00000	-9.18644	-9.48641	-7.63157
Stage 2	0.00000	0.00000	0.00000	-6.26611
Stage 3	0.00000	0.00000	0.00000	-6.33772
Stage 4	0.00000	0.00000	0.00000	0.00000

$sex

	Stage 1	Stage 2	Stage 3	Stage 4
Stage 1	0.00000	0.06063	0.69911	-0.02666
Stage 2	0.00000	0.00000	0.00000	-0.88873
Stage 3	0.00000	0.00000	0.00000	0.19921
Stage 4	0.00000	0.00000	0.00000	0.00000

$age

	Stage 1	Stage 2	Stage 3	Stage 4
Stage 1	0.00000	0.01028	0.02485	0.01390
Stage 2	0.00000	0.00000	0.00000	0.00519
Stage 3	0.00000	0.00000	0.00000	0.00035
Stage 4	0.00000	0.00000	0.00000	0.00000

$baseline

	Stage 1	Stage 2	Stage 3	Stage 4
Stage 1	-0.00066	0.00010	7.587e-05	0.00048
Stage 2	0.00000	-0.00189	0.00000	0.00189
Stage 3	0.00000	0.00000	-1.768e-03	0.00176
Stage 4	0.00000	0.00000	0.00000	0.00000

$sex

	HR	L95	U95
Stage 1 - Stage 2	1.06251	0.61473	1.83645
Stage 1 - Stage 3	2.01196	1.03213	3.92198
Stage 1 - Stage 4	0.97368	0.75767	1.25128
Stage 2 - Stage 4	0.41117	0.21701	0.77905
Stage 3 - Stage 4	1.22044	0.55678	2.67513

$age

	HR	L95	U95
Stage 1 - Stage 2	1.01033	0.98880	1.03233
Stage 1 - Stage 3	1.02516	0.99925	1.05174
Stage 1 - Stage 4	1.01400	1.00398	1.02411
Stage 2 - Stage 4	1.00520	0.97407	1.03732
Stage 3 - Stage 4	1.00035	0.97341	1.02804

**** WALDS TEST TO CHECK FOR DIFFERENCES BETWEEN MORTALITY INTENSITIES ****

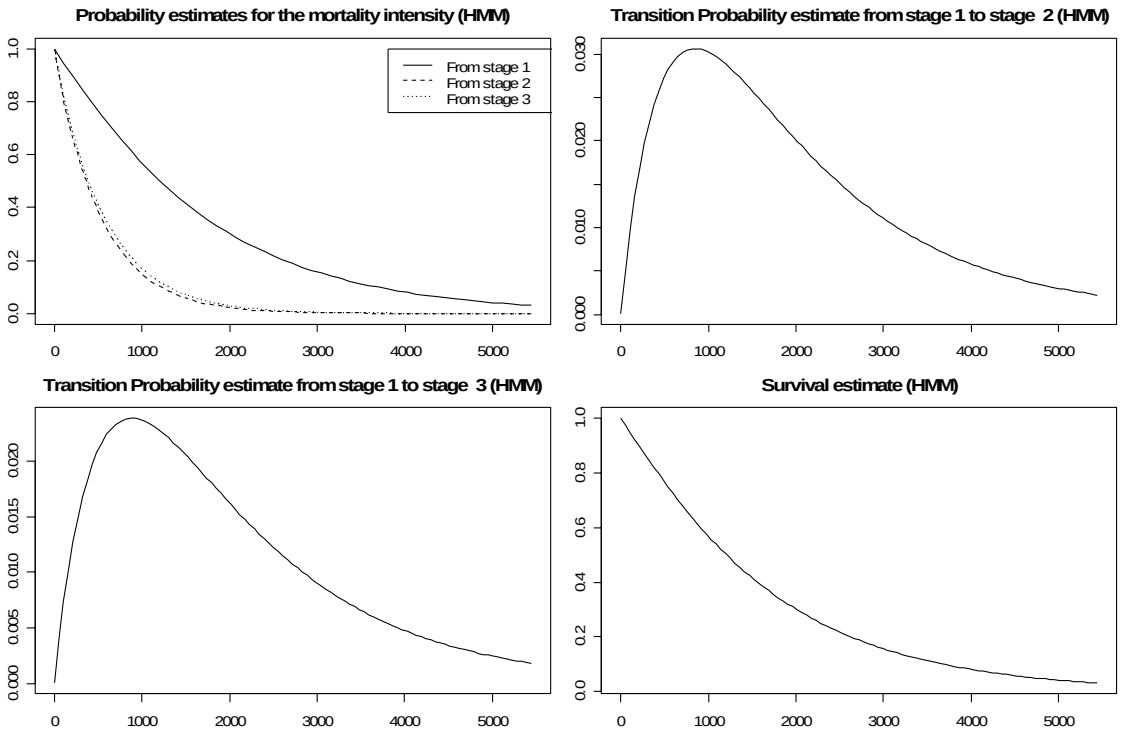
LET H0 BE THE HYPOTHESIS OF NON-DIFFERENCES BETWEEN MORTALITY INTENSITIES
H0 produces a chisquare statistic of: 88.81685
with a p-value of: 0

**** WALDS TEST TO CHECK FOR DIFFERENCES BETWEEN COVARIATE 1 IN THE TRANSITION INTENSITIES ****
LET H0 BE THE HYPOTHESIS OF NON-DIFFERENCES BETWEEN COVARIATE 1 IN THE TRANSITION INTENSITIES
H0 produces a chisquare statistic of: 33.07575
with a p-value of: 3.104323e-07

**** WALDS TEST TO CHECK FOR DIFFERENCES BETWEEN COVARIATE 2 IN THE TRANSITION INTENSITIES ****
LET H0 BE THE HYPOTHESIS OF NON-DIFFERENCES BETWEEN COVARIATE 2 IN THE TRANSITION INTENSITIES
H0 produces a chisquare statistic of: 2.485001
with a p-value of: 0.4780079

***** Sojourn times *****
$estimate
Stage 1	Stage 2	Stage 3
1507.8521	526.4247	565.5079

$SE
Stage 1	Stage 2	Stage 3
81.74408	88.43626	109.27736



Command line:

tdc.surv(msmrecmet, ncov=2, formula=c(7,8), model=c(0,0,0,0,1,0), surv.plot=TRUE, graphcov=2, plot.trans=TRUE)

Output:

```
*****************************************************************************
*** COX SEMI-MARKOV MODEL. SEPARATED FITTED COX MODELS FOR EACH TRANSITION ***
*****************************************************************************


***** FROM STATE 1 TO STATE 2  *****
n= 345

        coef      exp(coef)  se(coef)  z      p
sex     0.02482   1.03       0.2791    0.089  0.93
age     0.00321   1.00       0.0107    0.300  0.76

        exp(coef) exp(-coef) lower.95  upper.95
sex     1.03      0.975      0.593     1.77
age     1.00      0.997      0.982     1.02


Rsquare= 0   (max possible= 0.807 )
Likelihood ratio test= 0.09  on 2 df,   p=0.955
Wald test            = 0.09  on 2 df,   p=0.955
Score (logrank) test = 0.09  on 2 df,   p=0.955

-2*Log-likelihood: 567.5803


***** FROM STATE 1 TO STATE 3  *****
n= 345

        coef      exp(coef)  se(coef)  z      p
sex     0.6585    1.93       0.3391    1.94   0.052
age     0.0169    1.02       0.0127    1.33   0.180

        exp(coef) exp(-coef) lower.95  upper.95
sex     1.93      0.518      0.994     3.76
age     1.02      0.983      0.992     1.04

 Rsquare= 0.015   (max possible= 0.716 )
Likelihood ratio test= 5.21  on 2 df,   p=0.074
Wald test            = 4.9  on 2 df,   p=0.0862
Score (logrank) test = 5  on 2 df,   p=0.0822

-2*Log-likelihood: 428.8344

***** FROM STATE 1 TO STATE 4  *****
n= 345

        coef      exp(coef)  se(coef)  z       p
sex     -0.0351   0.966      0.12846   -0.273  0.78000
age     0.0174    1.018      0.00525   3.307   0.00094
```

	exp(coef)	exp(-coef)	lower.95	upper.95
sex	0.966	1.036	0.75	1.24
age	1.018	0.983	1.01	1.03

Rsquare= 0.034 (max possible= 0.999)
Likelihood ratio test= 11.8 on 2 df, p=0.0028
Wald test = 11.6 on 2 df, p=0.00308
Score (logrank) test = 11.6 on 2 df, p=0.003

-2*Log-likelihood: 2336.422

***** FROM STATE 2 TO STATE 4 *****
n= 53

	coef	exp(coef)	se(coef)	z	p
sex	-0.549876	0.577	0.3337	-16.476	0.099
age	-0.000862	0.999	0.0153	-0.0564	0.960

	exp(coef)	exp(-coef)	lower.95	upper.95
sex	0.577	1.73	0.30	1.11
age	0.999	1.00	0.97	1.03

Rsquare= 0.05 (max possible= 0.989)
Likelihood ratio test= 2.74 on 2 df, p=0.254
Wald test = 2.73 on 2 df, p=0.255
Score (logrank) test = 2.8 on 2 df, p=0.247

-2*Log-likelihood: 237.0028

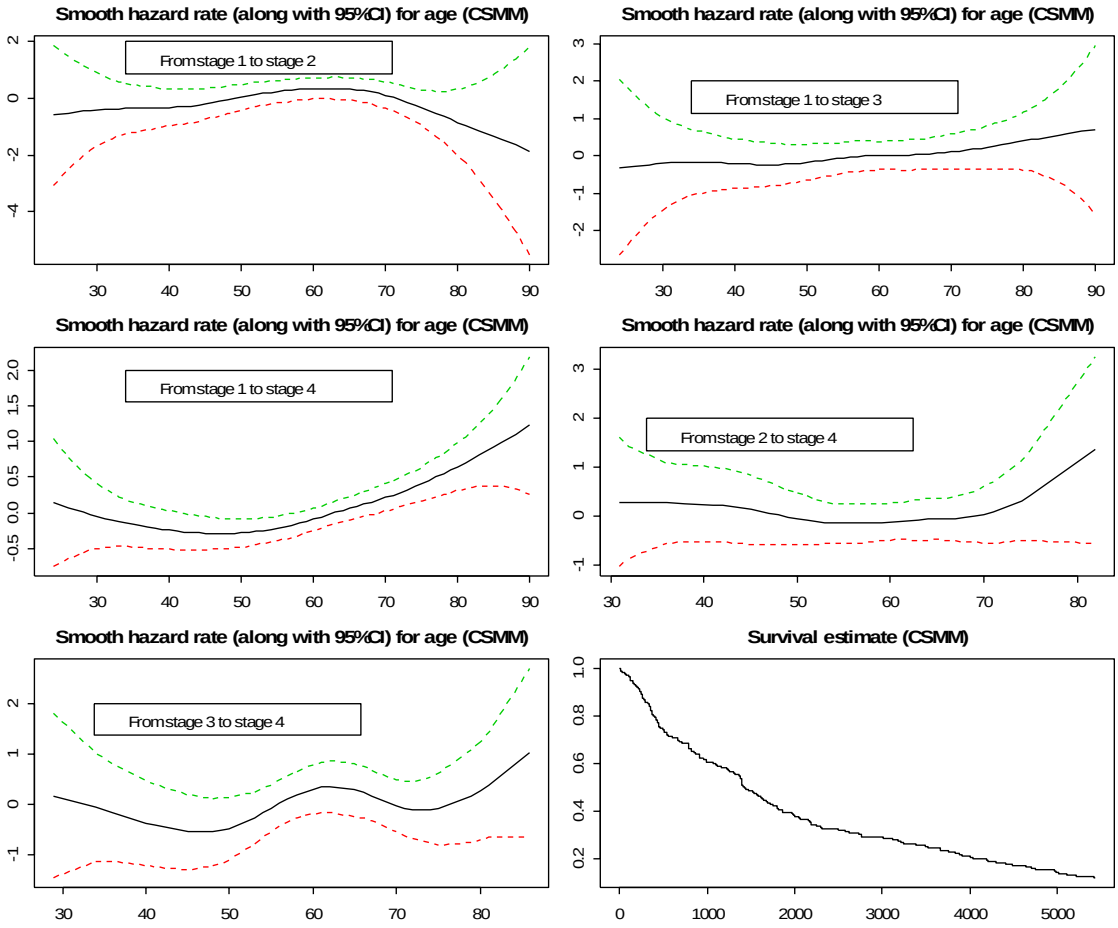
***** FROM STATE 3 TO STATE 4 *****
n= 41

	coef	exp(coef)	se(coef)	z	p
sex	0.5082	1.66	0.406	1.25	0.21
age	0.0165	1.02	0.015	1.10	0.27

	exp(coef)	exp(-coef)	lower.95	upper.95
sex	1.66	0.602	0.751	3.68
age	1.02	0.984	0.987	1.05

Rsquare= 0.059 (max possible= 0.986)
Likelihood ratio test= 2.47 on 2 df, p=0.290
Wald test = 2.42 on 2 df, p=0.298
Score (logrank) test = 2.43 on 2 df, p=0.297

-2*Log-likelihood: 171.7130



Command line:

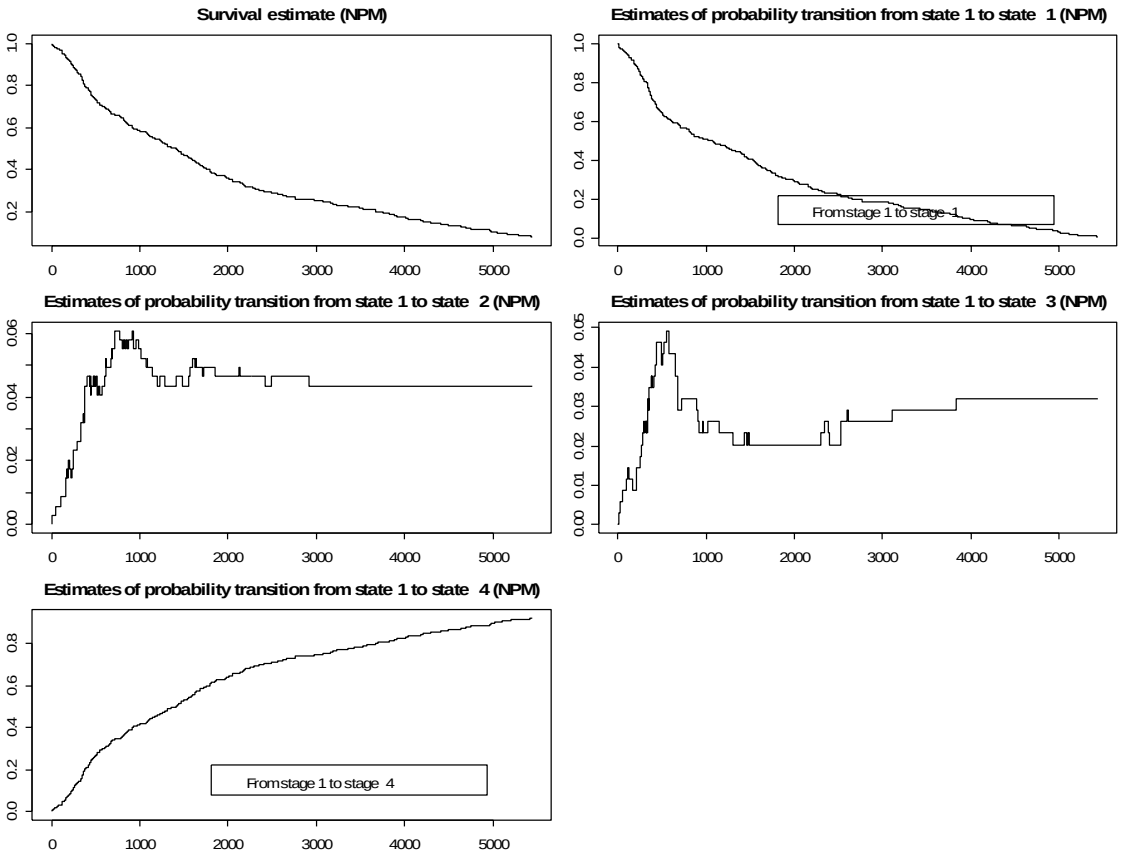
tdc.surv(msmrecmet, ncov=2,formula=c(7,8), model=c(0,0,0,0,0,1), surv.plot=TRUE, graphcov=2, plot.trans=TRUE)

Output:

```
*******************************************************************************
                ***** NON-PARAMETRIC MULTI-STATE MODEL *****
*******************************************************************************
```

Times	Survival	P11	P12	P13	P14
0	1	1	0	0	0
5	0.997101	0.997101	0	0	0.002899
8	0.994203	0.994203	0	0	0.005797
9	0.991304	0.991304	0	0	0.008696
10	0.991304	0.988406	0.002899	0	0.008696
11	0.988406	0.985507	0.002899	0	0.011594
14	0.988406	0.982609	0.002899	0.002899	0.011594
15	0.985507	0.97971	0.002899	0.002899	0.014493
21	0.985507	0.976812	0.002899	0.005797	0.014493

30	0.982609	0.973913	0.002899	0.005797	0.017391
36	0.97971	0.971014	0.002899	0.005797	0.02029
44	0.97971	0.968116	0.005797	0.005797	0.02029
48	0.97971	0.965217	0.005797	0.008696	0.02029
61	0.976812	0.962319	0.005797	0.008696	0.023188
63	0.973913	0.95942	0.005797	0.008696	0.026087
75	0.971014	0.956522	0.005797	0.008696	0.028986
88	0.968116	0.953623	0.005797	0.008696	0.031884
90	0.968116	0.950725	0.005797	0.011594	0.031884
99	0.965217	0.947826	0.005797	0.011594	0.034783
101	0.965217	0.944928	0.008696	0.011594	0.034783
104	0.965217	0.942029	0.008696	0.014493	0.034783
110	0.965217	0.942029	0.008696	0.014493	0.034783
112	0.962319	0.93913	0.008696	0.014493	0.037681
119	0.956522	0.936232	0.008696	0.011594	0.043478
120	0.953623	0.933333	0.008696	0.011594	0.046377
124	0.950725	0.930435	0.008696	0.011594	0.049275
129	0.947826	0.927536	0.008696	0.011594	0.052174
152	0.944928	0.924638	0.008696	0.011594	0.055072
156	0.942029	0.921739	0.008696	0.011594	0.057971
158	0.93913	0.918841	0.008696	0.011594	0.06087
159	0.936232	0.915942	0.008696	0.011594	0.063768
164	0.933333	0.913043	0.011594	0.008696	0.066667
165	0.933333	0.910145	0.014493	0.008696	0.066667
175	0.930435	0.907246	0.014493	0.008696	0.069565
180	0.927536	0.901449	0.017391	0.008696	0.072464
181	0.924638	0.901449	0.014493	0.008696	0.075362
184	0.924638	0.898551	0.017391	0.008696	0.075362
187	0.921739	0.892754	0.02029	0.008696	0.078261
190	0.918841	0.889855	0.02029	0.008696	0.081159
206	0.915942	0.889855	0.017391	0.008696	0.084058
209	0.913043	0.884058	0.017391	0.011594	0.086957
210	0.910145	0.884058	0.014493	0.011594	0.089855
213	0.910145	0.881159	0.014493	0.014493	0.089855
219	0.907246	0.878261	0.014493	0.014493	0.092754
220	0.904348	0.875362	0.014493	0.014493	0.095652
221	0.901449	0.872464	0.014493	0.014493	0.098551
227	0.898551	0.869565	0.014493	0.014493	0.101449
229	0.898551	0.866667	0.017391	0.014493	0.101449
239	0.895652	0.863768	0.017391	0.014493	0.104348
240	0.895652	0.857971	0.023188	0.014493	0.104348
244	0.892754	0.855072	0.023188	0.014493	0.107246
247	0.889855	0.852174	0.023188	0.014493	0.110145
…	…	…	…	…	…
5292	0.086957	0.011594	0.043478	0.031884	0.913043
5419	0.084058	0.008696	0.043478	0.031884	0.915942
5427	0.081159	0.005797	0.043478	0.031884	0.918841
5440	0.078261	0.002899	0.043478	0.031884	0.921739
5441	0.078261	0.002899	0.043478	0.031884	0.921739


Survival estimate (NPM)
Estimates of probability transition from state 1 to state 1 (NPM)
From stage 1 to stage 1
Estimates of probability transition from state 1 to state 2 (NPM)
Estimates of probability transition from state 1 to state 3 (NPM)
Estimates of probability transition from state 1 to state 4 (NPM)
From stage 1 to stage 4

C. Appendix with summary of the thesis in Spanish

Resumen de la tesis

La progresión de una cierta enfermedad puede describirse a través de los modelos multi-estado. Estos modelos pueden considerarse como una generalización del proceso de supervivencia donde varios eventos (intermedios) ocurren consecutivamente en el tiempo. En este contexto, algunos problemas de interés son: la estimación de tasas de progresión, la evaluación de los efectos de factores de riesgo individuales, o la estimación de la tasa de supervivencia. La influencia de estos eventos intermedios en la supervivencia es analizada habitualmente a través del modelo de regresión de Cox. Esta tesis contiene una revisión exhaustiva de los modelos multi-estado más habituales para estudiar la progresión de la enfermedad. Se discuten las diferencias entre estos modelos y el modelo de regresión de Cox, enfatizando las posibles ventajas y desventajas de cada método.

La supervivencia de un paciente puede pensarse como un proceso que consta de dos estados, con una posible transición del estado "vivo" al estado "muerto". En algunos estudios, sin embargo, el estado que representa a los pacientes "vivos" puede dividirse en dos o más estados intermedios, correspondiendo cada uno de ellos a una fase particular en el progreso natural de la enfermedad. A menudo, la influencia de estos eventos en la supervivencia es importante en el resultado del paciente y puede manejarse usando el modelo de regresión de Cox. En estos casos, el evento de interés se considera el evento principal, mientras que los eventos intermedios son incluidos frecuentemente como covariables dependientes del tiempo, en un modelo de riesgos proporcionales. El proceso termina cuando el paciente entra en el estado absorbente, "muerto", que corresponde al evento de interés. Ninguna otra observación es necesaria después de que este evento ocurra. Así, en análisis de supervivencia sólo considera dos estados, y el evento de interés es la transición de un estado a otro. Aunque el modelo de regresión de Cox presenta ciertas ventajas (fácil interpretación y disponibilidad en la mayoría de los paquetes estadísticos), este modelo proporciona estimaciones constantes de los efectos de las covariables a lo largo del estudio, lo que puede considerarse como una desventaja del modelo.

Los modelos multi-estado pueden ser considerados como una generalización de la herramienta básica que trata con datos de supervivencia. En estos modelos la supervivencia es el último resultado de interés, pero donde se identifican también estados intermedios. En contraste con los datos de supervivencia, en estos modelos se observa una sucesión de eventos y se registra más de una observación por individuo. En medicina, los estados intermedios podrían basarse en síntomas clínicos (episodios sangrantes), marcadores biológicos (recuentos de células CD4; el nivel de hemoglobina), alguna escala de la enfermedad (estadios de cáncer o de infección HIV) o una complicación no-fatal en el transcurso de la enfermedad (p.e., un transplante). Se llama transición, o evento, a un cambio de estado. Los estados pueden ser transitorios o absorbentes, si ninguna transición sale del estado (por ejemplo, muerte). La estructura de los estados, no necesariamente única, identifica los estados y las transiciones permitidas de un estado al otro.

Los modelos multi-estado presentan una serie de ventajas sobre el modelo de regresión de Cox. En primer lugar, aportan una perspectiva comprensiva del proceso de la enfermedad, y proporcionan una utilización más eficaz de la información incompleta, cuando solamente parte de la historia de la enfermedad de un individuo es conocida (manejando eficazmente datos con elevado porcentaje de censura). Además, con esta herramienta, las intensidades de transición (que proporcionan el riesgo instantáneo de pasar de un estado al otro) pueden utilizarse para determinar (entre otros parámetros) el tiempo de permanencia en un determinado estado de enfermedad, el número de individuos en diferentes estados en un determinado momento, y las tasas de supervivencia en cada estado. Finalmente, las covariables en las intensidades de transición también pueden explicar diferencias en el curso de la enfermedad entre la población. De este modo, los modelos multi-estado pueden revelar distintos efectos de un conjunto de covariables en las diferentes transiciones, algo que no sería posible con otros modelos como, por ejemplo, el modelo de regresión de Cox. De hecho, es muy improbable que el riesgo de muerte en pacientes que recibieron tratamientos distintos sea el mismo. Además, los factores de pronóstico asociados con el riesgo de muerte pueden ser diferentes en estos grupos de pacientes. En conclusión, los modelos multi-estado evalúan el progreso de la enfermedad del paciente dinámicamente dependiendo de la ocurrencia de eventos intermedios.

En los últimos años, existe un creciente interés en métodos estadísticos para estudiar la progresión de la enfermedad. Algunas enfermedades que se han estudiado usando los modelos multi-estado incluyen: Infección por VIH (Lagakos et al., 1988; Longini et al., 1989 y 1991; Gentleman et al., 1994; Satten y Longini, 1996; Aalen et al., 1997), cáncer de mama (Duffy y Chen, 1995; Chen et al., 1996; Pérez-Ocón et al., 2001), cirrosis (Kay, 1986; Andersen et al., 2000), leucemia (Klein et al., 1984 y 1994; Andersen et al., 1999; Keiding et al., 2001; Chevret et al., 2000), asma (Saint-Pierre et al., 2003), Alzheimer (Commenges et al., 2004), transplantes (Hansen et al., 1994; Klotz y Sharples, 1994; Jackson y Sharples, 2002), diabetes (Andersen, 1988), retinopatía diabética (Andersen, 1991; Marshall y Jones, 1995), malaria (Gottschau y Hogh, 1995) y esclerosis múltiple (Esbjerg et al., 1999). Estos modelos también se han utilizado en otros campos como el análisis de Fiabilidad (Su et al., 2000).

Una vez introducidos los modelos multi-estado, el Capítulo 2 de esta tesis proporciona una revisión más profunda de dichos modelos. En particular, en este capítulo introducimos estos modelos como procesos estocásticos; presentamos algunos de los modelos multi-estados más frecuentemente utilizados; consideramos diferentes patrones de censura y discutimos las hipótesis de simplificación más comunes.

En un modelo multi-estado general, los individuos cambian de un estado a otro a través de tiempo. El siguiente estado para cual el individuo cambia, y el tiempo de cambio, se especifican a través de intensidades de transición que proporcionan el riesgo instantáneo de movimiento de un estado al otro. Estos modelos se basan en procesos estocásticos en tiempo continuo permitiendo que los individuos se muevan entre un número finito de estados.

Formalmente, representaremos el estado de un individuo por un proceso estocástico $\{X_i(t),\ t \in \mathcal{T}\}$, donde $X_i(t)$ puede tomar un número finito de valores $\{1,...,N\}$, $\mathcal{T} = [0, \tau)$, $\tau \leq \infty$, y verificando ciertos presupuestos de simplificación. Así, $X_i(t)$ denota el estado ocupado por el i-ésimo individuo y $S = \{1,...,N\}$ es un espacio de estados finito. Por consiguiente, $\{X_i(t),\ t \in \mathcal{T}\}$ contiene la información de las diferentes transiciones del individuo a lo largo del tiempo, así como el tiempo en que estas transiciones ocurrieron. Con la evolución del proceso en tiempo *t*, una historia, o

filtración, $\mathcal{F}_{t-}$, se generará conteniendo toda la información sobre el proceso en el intervalo $[0,t)$.

Este proceso multi-estado se caracteriza completamente por intensidades de transición o por probabilidades de transición entre estados *h* y *j*, que nosotros expresaremos respectivamente por $\alpha_{hj}\left(t\middle|\mathcal{F}_{t-}\right)$ y $p_{hj}(s,t)=\mathbb{P}\left(X(t)=j\middle|X(s)=h,\mathcal{F}_{s-}\right)$, siendo $\mathcal{F}_{l-}$ la historia observada del proceso en tiempo *l*. Así, mientras que las probabilidades de transición proporcionan medidas importantes para hacer predicciones a largo plazo, cada intensidad de transición, $\alpha_{hj}\left(t\middle|\mathcal{F}_{t-}\right)$ representa el riesgo instantáneo de progresión de un estado al otro.

La complejidad de un modelo multi-estado depende del número de estados y de las transiciones permitidas entre dichos estados. En análisis de supervivencia, la forma más simple de modelo multi-estado es el modelo de mortalidad. Este modelo se representa en la Figura 1.

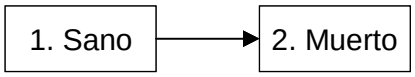


Figura 1: modelo de mortalidad para el análisis de supervivencia.

Otro modelo posible para describir la progresión de la enfermedad es el modelo enfermedad-muerte (illness-death). Este modelo se caracteriza totalmente por sus tres intensidades de transición, cada una de ellas describiendo el riesgo instantáneo de progresión de un estado al otro: la intensidad de la enfermedad, la intensidad de mortalidad sin la enfermedad, y la intensidad de mortalidad después de la ocurrencia de la enfermedad. Estos modelos se usan ampliamente en la literatura médica y pueden utilizarse en el estudio de la incidencia de enfermedad y/o muerte. Estos modelos serán tratados extensamente a lo largo de esta tesis.

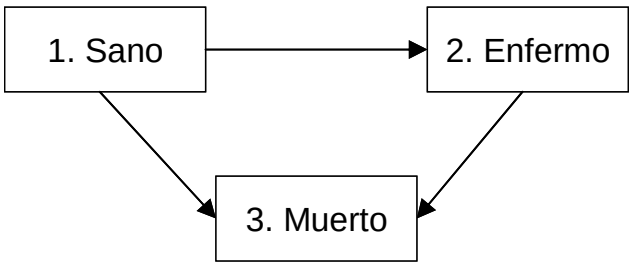


Figura 2: modelo enfermedad-muerte (Illness-death).

En el Capítulo 3 revisamos varios métodos que siguen la metodología de los modelos multi-estado con el presupuesto de Markov. Específicamente, revisamos los modelos siguientes: (i) modelo de Cox Markov (Andersen et al., 2000), donde las intensidades de transición usan separadamente diferentes modelos de Cox; (ii) modelos de Markov homogéneos en tiempo (Kay, 1986); y (iii) modelos de Markov con intensidades constantes a trozos (Pérez-Ocón et al., 2001). Los métodos revisados se ilustran con los datos de transplante de corazón de Stanford, proporcionando un guía sobre el uso de estas metodologías para el estudio de la evolución de la enfermedad.

El estudio de transplante de corazón de Stanford comenzó en octubre de 1967. Los datos disponibles en el artículo de Crowley y Hu (1977) cubren el período hasta el 1 de abril de 1974. Algunos pacientes murieron antes de que se dispusiera de un corazón apropiado para el transplante. De los 103 pacientes, 69 recibieron un transplante de corazón. El número total de muertes fue de 75. Para cada individuo, se encuentra registrado un indicador de su estado vital final (censura o no), los tiempos de supervivencia desde la entrada del paciente en el estudio (en días), y un vector de covariables incluyendo la edad a la aceptación, año de aceptación, cirugía anterior, y transplante. Transplante es la única covariable dependiente en tiempo. En el contexto de un planteamiento multi-estado podemos considerar la covariable "transplante" como un estado de riesgo asociado y entonces utilizar un modelo multi-estado para investigar el efecto del transplante en la supervivencia. Para los datos de Stanford, es interesante comparar el riesgo de muerte antes y después del transplante. También podemos explorar los efectos de las restantes covariables en cada una de las transiciones. En este capítulo, también utilizamos estos datos para discutir posibles contrastes de hipótesis y métodos para chequear el modelo, como el presupuesto de homogeneidad en el tiempo o la propiedad de Markov. Además, al examinar los datos, es interesante comprobar diferentes hipótesis sobre el modelo (hipótesis sobre las intensidades de transición,

coeficientes de regresión, bondad de ajuste, etc.). Se discuten diferencias entre estos modelos y el modelo de regresión de Cox con covariables dependientes en tiempo y enfocamos las posibles ventajas y desventajas para cada método. A través de esta ilustración mostramos como los modelos multi-estado pueden proporcionar nuevas perspectivas, al tiempo que confirman los resultados obtenidos mediante el modelo de regresión de Cox.

Una atención especial debe prestarse cuando se está interesado en evaluar el efecto de una covariable. Como se ha mencionado anteriormente, el modelo de Cox proporciona estimaciones constantes del efecto de la covariable a lo largo del periodo de estudio. Para evitar este problema utilizamos métodos de suavización spline (P-splines) para obtener un modelo de Cox dinámico. Además, introducimos estos métodos spline en modelos multi-estado para averiguar los posibles efectos no-lineales de las covariables en las intensidades de transición. El uso de estos métodos en modelos multi-estado es novedoso.

Mientras que los modelos multi-estado proporcionan estimaciones no-sesgadas de la importancia y capacidad de las covariables para explicar el curso de la enfermedad, estos problemas no pueden ser totalmente explorados utilizando modelos de Cox. Para ilustrar los posibles beneficios del uso de los modelos multi-estado, se realizaron varios estudios de simulación. A través de estos estudios, se explica por qué el modelo de Cox puede no ser apropiado (con interpretación difícil) cuando se utiliza en presencia de covariables dependientes del tiempo. De este modo, mostramos cómo los modelos multi-estado permiten un análisis más concluyente que el modelo de Cox sobre los efectos de las covariables.

Al analizar los datos de transplante de corazón de Stanford utilizando la metodología multi-estado, el presupuesto de Markov se reveló satisfactorio. Se utilizó un test de bondad de ajuste para comparar los distintos modelos utilizados, indicando que el modelo de Cox Markov y el modelo homogéneo a trozos son más apropiados que el modelo homogéneo. Al utilizar el modelo homogéneo se verificó que dicho modelo infravalora a corto plazo la mortalidad. La razón por la cual este modelo ajustó pobremente a los datos de Stanford se debe al hecho de que el proceso de supervivencia no es homogéneo a lo largo del tiempo. Mientras que el modelo de regresión de Cox con covariables dependientes en tiempo presentó un efecto despreciable del trasplante, el modelo de Markov homogéneo en el tiempo indicó que el trasplante se asocia

significativamente a una disminución en el riesgo de mortalidad. La aplicación del modelo homogéneo a trozos sólo confirmó tal asociación cuando se analizaba la supervivencia a largo plazo.

A través de un estudio de simulación, hemos mostrado que el efecto de las covariables en la supervivencia puede verse profundamente afectado por el efecto de dichas covariables a través de una única transición. De hecho, al analizar los datos de Stanford, verificamos que el efecto de la covariable "tiempo de aceptación" (año), considerado el predictor más importante en el modelo de Cox, sólo obtuvo un efecto estadísticamente significativo en la transición de mortalidad para los pacientes sin trasplante. Algunos de los modelos multi-estado estudiados mostraron una influencia negativa de la covariable "cirugía" en el riesgo para la transición $2 \to 3$. Por otro lado, la edad a la aceptación, ha presentado un efecto significativo en la supervivencia en cualquiera de los modelos estudiados, y su efecto positivo en el riesgo indica que los pacientes más jóvenes tienen una supervivencia mejor.

Tradicionalmente, los métodos estadísticos para analizar los modelos multi-estado dependen del presupuesto de Markov. Bajo la propiedad de Markov, las intensidades de transición dependen del tiempo actual y del estado actualmente ocupado pero no dependen de la historia del paciente (el tiempo de permanencia en el estado actual; los tiempos de transición de un estado al otro, etc.). Ignorando la historia de la enfermedad, estos modelos pueden presentar serias limitaciones, llevando entonces a una mala especificación. Un planteamiento alternativo consiste en utilizar el presupuesto de semi-Markov, con el cual el futuro del proceso no depende del tiempo actual, si no tan sólo de la duración en el estado actual. En el Capítulo 4 se revisa el modelo de Cox semi-Markov y se propone un nuevo planteamiento no-Markoviano, el cual permite que las intensidades de transición puedan depender no sólo del tiempo actual, sino también del tiempo de transición al estado actual. Nuestra investigación sobre modelos no-Markovianos tiene dos objetivos principales. El primer objetivo es desarrollar un nuevo planteamiento basado en presupuestos menos restrictivos que aquéllos basados en la propiedad de Markov. El segundo objetivo es comparar los estimadores desarrollados aquí para las probabilidades de la transición con las estimaciones de Aalen-Johansen (derivadas bajo el presupuesto de Markov). Para este propósito, consideramos el modelo enfermedad-muerte (illness-death). Posteriormente, consideramos la extensión de esta metodología a modelos multi-estado más complejos.

Asumamos un modelo enfermedad-muerte (ver Figura 2). Sea $\{X(t), t \geq 0, X(0) = 1\}$ un proceso estocástico no-homogéneo y asumamos que se observa un número finito de historias independientes del proceso. Representamos el comportamiento estocástico del proceso por un vector aleatorio (T_{12}, T_{13}, T_{23}) donde T_{hj} es el tiempo potencial de permanencia en el estado *h* antes de la transición al estado *j*. Obsérvese que los tiempos de permanencia en estado 1 y 2 son dados por $Z = \min(T_{12}, T_{13})$ y T_{23} respectivamente. Un individuo en el estado sano se expone a dos eventos mutuamente excluyentes, "enfermedad" y "muerte", que no pueden ocurrir simultáneamente, observando tan sólo el primero de estos eventos. Solamente los individuos para los cuales $T_{12} \leq T_{13}$ entran en estado 2 en algún tiempo T_{12}. Por supuesto, existen varios problemas que afectan a la observación de las variables T_{hj}. Puede ocurrir una censura por la derecha debido a la limitación de tiempo en el seguimiento, pérdida de casos, etc. Por otro lado, siempre que $T_{13} < T_{12}$, se obtiene un valor censurado por la derecha para T_{12}, no disponiéndose de ninguna información sobre T_{23}. De lo mismo modo, siempre que $T_{12} \leq T_{13}$ se obtiene un valor censurado por la derecha de T_{13}. Al largo de este trabajo denotamos por C el tiempo potencial de censura. De este modo, los tiempos de los individuos pueden estar censurados antes de la enfermedad $(C < \min(T_{12}, T_{13}))$ o después de la enfermedad $(C < T_{12} + T_{23}$ and $T_{12} \leq \min(T_{13}, C))$. Asumiremos que la variable *C* es independiente de (T_{12}, T_{13}, T_{23}).

Nuestro objetivo es la estimación de la probabilidad de transición $p_{hj}(s,t) = \mathbb{P}(X(t) = j \,|\, X(s) = h)$, $s < t$, que representa la probabilidad condicional de que el individuo esté en el estado *j* en tiempo *t* sabiendo que estaba en el estado *h* en tiempo *s*. Para el modelo de enfermedad-muerte, es suficiente considerar la estimación de las siguientes probabilidades de transición: $p_{11}(s,t)$, $p_{12}(s,t)$ y $p_{22}(s,t)$. Las probabilidades restantes pueden obtenerse a partir de estas: $p_{13}(s,t) = 1 - p_{11}(s,t) - p_{12}(s,t)$ y $p_{23}(s,t) = 1 - p_{22}(s,t)$.

Nótese que

$$p_{11}(s,t) = \mathbb{P}\left(Z > t \middle| Z > s\right),$$

$$p_{12}(s,t) = \mathbb{P}\left(T_{12} \le t, T_{12} \le T_{13}, T_{12} + T_{23} > t \middle| Z > s\right),$$

$$p_{22}(s,t) = \mathbb{P}\left(T_{12} + T_{23} > t \middle| T_{12} \le s, T_{12} \le T_{13}, T_{12} + T_{23} > s\right).$$

Estas cantidades son determinadas por la distribución conjunta de (T_{12}, T_{13}, T_{23}). En particular, el conocimiento de la distribución de *Z* es suficiente para $p_{11}(s,t)$:

$$p_{11}(s,t) = \frac{\mathbb{P}(Z > t)}{\mathbb{P}(Z > s)},$$

mientras que esperanzas del tipo $S(\phi) = \mathbb{E}\left[\phi(T_{12}, T_{12} + T_{23})\mathbb{I}(T_{12} \le T_{13})\right]$ surgen para $p_{12}(s,t)$ $\left(\phi(u,v) = \phi_{s,t}(u,v) = \mathbb{I}(s < u \le t, v > t)\right)$:

$$p_{12}(s,t) = \frac{S(\phi_{s,t})}{\mathbb{P}(Z > s)},$$

y para $p_{22}(s,t)$ $\left(\phi(u,v) = \tilde{\phi}_{s,t}(u,v) = \mathbb{I}(u \le s, v > t)\right)$,

$$p_{22}(s,t) = \frac{S(\tilde{\phi}_{s,t})}{S(\tilde{\phi}_{s,s})}.$$

Para la estimación de estas probabilidades, consideremos $U = \min(T_{12}, T_{13}, C)$, $\delta = \mathbb{I}(T_{12} \le \min(T_{13}, C))$, $V = \min(T_{23}, C - T_{12})$, $\rho = \mathbb{I}(T_{23} \le C - T_{12})$ y $\eta = \mathbb{I}(T_{13} \le C)$. Las variables *U* y *V* representan los tiempos de permanencia en estados 1 y 2, respectivamente; δ nos indica si ocurrió una transición del estado 1 al estado 2; el indicador ρ nos ofrece la misma información para la transición $2 \to 3$; finalmente, cuando $\delta = 0$, el evento $\eta = 1$ nos indica que ocurrió una transición $1 \to 3$. La muestra observada se representa por $(U_i, \delta_i, \delta_i V_i, \delta_i \rho_i, (1-\delta_i)\eta_i), 1 \le i \le n.$

Consideremos $\gamma_i = \delta_i + (1-\delta_i)\eta_i$. Como *Z* y *C* son independientes, el estimador Kaplan-Meier basado en los (U_i, γ_i)'s, digamos $\widehat{H}$, estima consistentemente la distribución *H* de *Z*. Entonces un posible estimador para $p_{11}(s,t)$ viene dado por,

$$\widehat{p_{11}}(s,t) = \frac{1 - \widehat{H}(t)}{1 - \widehat{H}(s)}.$$

Para introducir un estimador para $S(\phi)$ notemos que

$$S(\phi) = \mathbb{E}\left[\frac{\phi(U, U+\delta V)\delta\rho}{1-G\left((U+\delta V)^{-}\right)}\right],$$

donde G denota la distribución de la variable C. Además, si $T = \mathbb{I}(T_{12} \le T_{13})(T_{12}+T_{23}) + \mathbb{I}(T_{12} > T_{13})(T_{13})$ denota el tiempo de supervivencia del proceso, entonces tenemos (i) $\mathbb{I}(T \le C) = (1-\delta)\eta + \delta\rho$; y (ii) $U + \delta V = \min(T, C)$. Como T y C son independientes, el estimador Kaplan-Meier basado en los $(U_i + \delta_i V_i, 1-\nu_i)$'s con $\nu_i = (1-\delta_i)\eta_i + \delta_i\rho_i$ estima consistentemente G. Denotamos este estimador por $\widehat{G}$. Entonces, proponemos estimar $S(\phi)$ por

$$\hat{S}(\phi) = \frac{1}{n}\sum_{i=1}^{n} \frac{\phi(U_i, U_i+\delta_i V_i)\delta_i\rho_i}{1-\hat{G}\left((U_i+\delta_i V_i)^{-}\right)}.$$

Podemos ahora introducir estimadores para las restantes probabilidades de transición:

$$\widehat{p_{12}}(s,t) = \frac{1}{n\left(1-\widehat{H}(s)\right)}\sum_{i=1}^{n} \frac{\mathbb{I}(s < U_i \le t)\mathbb{I}(U_i+\delta_i V_i > t)\delta_i\rho_i}{1-\hat{G}\left((U_i+\delta_i V_i)^{-}\right)},$$

$$\widehat{p_{22}}(s,t) = \frac{\displaystyle\sum_{i=1}^{n} \frac{\mathbb{I}(U_i \le s)\mathbb{I}(U_i+\delta_i V_i > t)\delta_i\rho_i}{1-\hat{G}\left((U_i+\delta_i V_i)^{-}\right)}}{\displaystyle\sum_{i=1}^{n} \frac{\mathbb{I}(U_i \le s)\mathbb{I}(U_i+\delta_i V_i > s)\delta_i\rho_i}{1-\hat{G}\left((U_i+\delta_i V_i)^{-}\right)}}.$$

Los estimadores aquí propuestos difieren de los propuestos por Aalen y Johansen (1978), excepto para el caso de $p_{11}(s,t)$. De hecho, los estimadores de Aalen-Johansen presentan un sesgo sistemático en situaciones no-Markovianas, debido a que están basados en la suposición de Markov. Nuestra propuesta de estimación vence esta dificultad, si bien a costa de resultar en una estimación con mayor varianza.

Otro punto de interés son los resultados asintóticos para las probabilidades de transición estimadas. Se estudiaran: (i) consistencia; (ii) convergencia en distribución a la distribución normal; y (iii) estimación (límite) de la varianza.

Se realizaron estudios de simulación para comparar los estimadores no-Markovianos propuestos con los estimadores de Aalen-Johansen (Markovianos) para el modelo de enfermedad-muerte. A través de nuestro estudio de simulación hemos

verificado que, a menos que el proceso satisfaga el presupuesto de Markov, los estimadores no-Markovianos aquí propuestos constituyen una buena opción. Estos métodos fueron ilustrados utilizando datos de un ensayo clínico danés en cirrosis del hígado.

Una limitación importante para la aplicación de modelos multi-estado es la poca disponibilidad de software "amigable" para estos modelos. La mayoría del software disponible presenta algunas dificultades y limitaciones en práctica. Además, en algunos estudios clínicos el presupuesto de Markov puede ser apropiado, mientras que para otros un modelo con la propiedad de semi-Markov (o no-Markov) es preferible. En algunos casos, un modelo homogéneo en tiempo es satisfactorio, mientras que en otros no. Además, las posibles comparaciones entre los distintos modelos multi-estado son difíciles de llevar a cabo, ya que cada uno de ellos tiene su propia estructura de datos de entrada (input). Además, la mayoría de los programas disponibles sólo proporcionan estimaciones de los parámetros de regresión y no proporcionan gráficas para la estimación de supervivencia y para las probabilidades de transición estimadas. Por todo ello, hemos desarrollado un programa en R, llamado **tdc.surv,** que puede ser utilizado para ajustar de una manera sencilla y compacta la mayoría de los modelos estudiados. Las ventajas de este software incluyen la misma entrada de los datos para ajustar a los distintos modelos, proporcionando los resultados numéricos y gráficos correspondientes. De este modo, los usuarios pueden analizar los resultados ofrecidos a través de diferentes modelos, compararlos entre sí y tomar decisiones. En el Capítulo 5 se hace una descripción detallada de **tdc.surv**, con aplicaciones a distintas bases de datos reales.

Bibliography

Aalen, O. (1976). Nonparametric inference in connection with multiple decrement models. *Scandinavian Journal of Statistics* **3**, 15-27.

Aalen, O., Borgan, O., and Fekjaer, H. (2001). Covariate adjustment of event histories estimated from Markov chains: the additive approach. *Biometrics* **57**, 993-1001.

Aalen, O. and Johansen, S. (1978). An empirical transition matrix for nonhomogeneous Markov chains based on censored observations. *Scandinavian Journal of Statistics* **5**, 141-150.

Aalen, O., Vernon, T.F., De Angelis, D., Day, N.E. and Gill, O.N. (1997). A Markov model for HIV disease progression including the effect of HIV diagnosis and treatment: Application to AIDS prediction in England and Wales. *Statistics in Medicine* **16**, 2191-2210.

Alioum, A. and Commenges, D. (2001). MKVPCI: A computer program for Markov models with piecewise constant intensities and covariates. *Computer Methods and Programs in Biomedicine* **64**, 109-119.

Andersen, P.K. (1988). Multistate models in survival analysis: a study of nephropathy and mortality in diabetes. *Statistics in Medicine* **7**, 661-670.

Andersen, P.K., Abildstrom, S.Z. and Rosthøj, S. (2002). Competing Risks as a Multi-State Model. *Statistical Methods in Medical Research* **11,** 203-215.

Andersen, P.K., Borgan, O., Gill, R.D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer, New York.

Andersen, P.K., Esbjerj, S. and Sorensen T.I.A. (2000). Multistate models for bleeding episodes and mortality in liver cirrhosis. *Statistics in Medicine* **19**, 587-599.

Andersen, P.K., Hansen, L.S. and Keiding, N. (1991). Assessing the influence of reversible disease indicators on survival. *Statistics in Medicine* **10**, 1061-1067.

Andersen, P.K., Hansen, L.S. and Keiding, N. (1991b). Non- and semi-parametric estimation of transition probabilities from censored observations of a non-homogeneous Markov process. *Scandinavian Journal of Statistics* **18**, 153-167.

Andersen, P.K., Horowitz, M.M., Klein, J.P., Socie, G., Stone, J.V. and Zhang, M.J. (1999). Modelling covariate adjusted mortality relative to a standard population. *Statistics in Medicine* **18**, 1529-1540.

Borgan O. (1998). Aalen-Johansen estimator. *Encyclopedia of Biostatistics* (eds. P. Armitage and T. Colton), vol **1**, pp 5-10, Wiley: Chichester.

Borgan O. (1998b). Kaplan-Meier estimator. *Encyclopedia of Biostatistics* (eds. P. Armitage and T. Colton), vol **1**, pp 2154-2160, Wiley: Chichester.

Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics* **30**, 89–99.

Chen, H.H., Duffy, S.W. and Tabar, L. (1996). A Markov chain method to estimate the tumour progression rate from preclinical to clinical phase, sensitivity and positive predictive value for mammography in breast cancer screening. *Statistician* **45**, 307–317.

Chevret, S., Leporrier, M. and Chastang, C. (2000). Measures of treatment effectiveness on tumour response and survival: a multi-state model approach. *Statistics in Medicine* **19**, 837-848.

Commenges, D. (2002). Inference for multi-state models from interval-censored data. *Statistical Methods in Medical Research* **11**, 167-182.

Commenges, D., Joly, P., Letyenneur, L. and Dartigues J.F. (2004). Incidence and mortality of Alzheimer's disease or dementia using an illness-death model. *Statistics in Medicine* **23**, 199-210.

Couper D, Pepe MS. (1997). Modelling prevalence of a condition: chronic graft-versus-host disease after bone marrow transplantation. *Statistics in Medicine* **16**, 1551-1571.

Cox D.R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society* Series B **34**, 187-220.

Cox, D.R. and Miller, H.D. (1965). *The theory of stochastic processes*. Chapman and Hall, London.

Cox, D.R. and Oakes, D. (1984). *Analysis of survival data*. Chapman and Hall, London.

Crowley, J. and Hu, M. (1977). Covariance Analysis of Heart Transplant Survival Data. *Journal of the American Statistical Association* **72**, 27-36.

Datta, S. and Satten, G.A. (2001). Validity of the Aalen-Johansen estimators of stage occupation probabilities and Nelson Aalen integrated transition hazards for non-Markov models. *Statistics and Probability Letters* **55**, 403-411.

De Boor, C. (1978). *A practical guide to splines,* Berlin: Springer.

Duffy, S.W. and Chen, H.H. (1995). Estimation of mean sojourn time in breast cancer screening using a Markov chain model of entry to and exit from preclinical detectable phase. *Statistics in Medicine* **14**, 1531–1543.

Durrleman, S. and Simon, R. (1989). Flexible regression models with cubic splines. *Statistics in Medicine* **8**, 551–561.

Efron B. (1977). The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistical Association* **72**, 557-565.

Eilers, P.H.C. and Marx, B.D. (1996). Flexible smoothing with B-splines and penalities. *Statistical Science* **11**, 89-121.

Esbjerg, S., Keiding, N. and Koch-Henriksen, N. (1999). Reporting delay and corrected incidence of multiple sclerosis. *Statistics in Medicine* **18**, 1691-1706.

Escolano, S., Golmard, J.L., Korinek, A.M. and Mallet, A. (2000). A multi-state model for evolution of intensive care unit patients: prediction of nosocomial infections and deaths. *Statistics in Medicine* **19**, 3465-3482.

Fleming, T. R. and Harrington, D. P. (1984). Nonparametric estimation of the survival distribution in censored data. *Communications in Statistics: Theory and Methods*, **13**, 2469–2486.

Fleming, T. R. and Harrington, D. P. (1991). *Counting processes and survival analysis*. Wiley, New York.

Frydman, H. (1995). Nonparametric estimation of a Markov illness-death process from interval-censored observations, with application to diabetes survival data. *Biometrika* **82**, 773-789.

Gentleman, R. and Crowley, J. (1991). Local full likelihood estimation for the proportional hazards model, *Biometrics* **47**, 1283-1296.

Gentleman, R.C., Lawless, J.F., Lindsey, J.C. and Yan, P. (1994). Multi-state Markov models for analysing incomplete disease history data with illustrations for HIV disease. *Statistics in Medicine* **13**, 805-821.

Glidden D. (2002). Robust inference for event probabilities with non-Markov event data. *Biometrics* **58**, 361-368.

Gottschau, A. and Hogh, B. (1995). Interval censored survival data and multistate compartmental models in the analysis of first appearance of Plasmodium falciparum parasites in infants. *Statistics in Medicine* **14**, 2727-2736.

Gray R.J. (1994). Spline-based tests in survival analysis. *Biometrics* **50**:640– 652.

Greenwood, M. (1926). The natural duration of cancer. *Reports on public Health and Medical Subjects*. London: Her Majesty's Stationery Office **33**: 1-26.

Hansen, B.E., Thorogood, J., Hermans, J., Ploeg, R.J., Van Bockel, J.H. and Van Houwelingen, J.C. (1994). Multistate modelling of liver transplantation data. *Statistics in Medicine* **13**, 2517-2529.

Harrel, F.E. and Lee, K.L. (1986). Verifying assumptions of the Cox proportional hazards model. *Proceedings of the eleventh international conference of the SAS user's group.* Atlanta Georgia, February 9-12, 823-828.

Hastie, T.J. and Tibshirani, R.J. (1986). Generalized additive models. *Statistical Science* **1**, 297–318.

Hastie, T.J. and Tibshirani, R.J. (1987). Generalized additive models: some applications. *Journal of the American Statistical Association* **82**, 371-386.

Hastie, T.J. and Tibshirani, R.J. (1990). Exploring the nature of covariate effects in the proportional hazards model. *Biometrics* **46**, 1005-1016.

Helms F., Czado C. and Gschlobl S. (2004). Calculation of LTC Premiums based on direct estimates of transition probabilities. 10 June 2005 <http://www-m4.ma.tum.de/Papers/Czado/paper393.ps>.

Hougaard P. *Analysis of Multivariate Survival Data*. Springer, New York, 2000.

Hui-Min, W., Ming-Fang, Y. and Hsiu-Hsi, C. (2004). SAS macro program for non-homogeneous Markov process in modelling multi-state disease progression. *Computer Methods and Programs in Biomedicine* **75**, 95-105.

Jackson, C.H., Sharples, L.D. (2002). Hidden Markov models for the onset and progression of bronchiolitis obliterans syndrome in lung transplant recipients. *Statistics in Medicine* **21**, 113-128.

Joly, P., Commenges, D., Helmer, C. and Letenneur, L. (2002). A penalized likelihood approach for an illness-death model with interval censored data: application to age-specific incidence of dementia. *Biostatistics* **3**, 433-443.

Juang, B.H. and Rabiner, L.R. (1991). Hidden Markov models for speech recognition. *Tecnometrics* **33**, 251-272.

Kalbfleisch, J.D. and Lawless, J. (1985). The analysis of panel data under a Markov assumption. *Journal of the American Statistical Association* **80**, 863-871.

Kalbfleisch, J.D. and Prentice, R.L. (1980). *The Statistical Analysis of Failure Time Data*. Wiley: New York.

Kaplan, E.L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **53**, 457-481.

Kay, R. (1986). A Markov model for analysing cancer markers and disease states in survival studies. *Biometrics* **42**, 855-865.

Keiding N. (1991). Age-specific incidence and prevalence: A statistical perspective. *Journal of the Royal Statistical Society Series A* **154**, 371–396.

Keiding, N., Klein, J.P. and Horowitz, M. (2001). Multi-state models and outcome prediction in bone marrow transplantation. *Statistics in Medicine* **20**, 1871-1885.

Klein, J.P., Klotz, J.H. and Grever M.R. (1984). A biological marker model for predicting disease transitions. *Biometrics* **40**, 927-936.

Klein, J.P., Keiding, N. and Copelan, E.A. (1994). Plotting summary predictions in multistate survival models: probabilities of relapse and death in remission for bone marrow transplantation patients. *Statistics in Medicine* **13**, 2315-2332.

Klotz, J.H. and Sharples, L.D. (1994). Estimation for a Markov heart transplant model. *Statistician* **43**, 431– 436.

Lagakos, S.W., Barraj, L.M. and De Gruttola, V. (1988). Nonparametric analysis of truncated survival data, with applications to AIDS. *Biometrika* **75**, 515-523.

Longini, I.R., Clark, W.S., Byers, R.H., Ward, J.W., Darrow, W.W., Lemp, G.F. and Hethcote, H.W. (1989). Statistical analysis of the stages of HIV infection using Markov model. *Statistics in Medicine* **8**, 831– 843.

Longini, I.M., Clark, W.S., Gardner, L.I. and Brundage, J.F. (1991). The dynamics of CD4+ T-lymphocyte decline in HIV infected individuals: a Markov modelling approach. *Journal of Acquired Immune Deficiency Syndromes* **4**, 1141–1147.

Mantel, N. and Byar, D.P. (1974). Evaluation of Response-Time Data Involving Transient States: An Illustration Using Heart-Transplant Data. *Journal of the American Statistical Association* **69**, 81-86.

Marshall, G. and Jones, R.H. (1995). MARKOV: a computer program for multi-state Markov models with covariables. *Computer Methods and Programs in Biomedicine* **47**, 147-156.

Marshall, G. and Jones, R.H. (1995) Multi-state models and diabetic retinopathy. *Statistics in Medicine* **14**, 1975-1983.

Nelson,W. (1969). Hazard plotting for incomplete failure data. *Journal of Quality Technology* **1**, 27–52.

O´Sullivan, F. (1988). Nonparametric estimation of relative risk using splines and cross-validation, *SIAM journal of Scientific and Statistical Computation* **9**, 531-542.

Paes, A. and Lima, A. (2004). A SAS macro for estimating transition probabilities in semiparametric models for recurrent events. *Computer Methods and Programs in Biomedicine* **75**, 59-65.

Pérez-Ocón, R., Ruiz-Castro, J.E. and Gámiz-Pérez, M.L. (2001). A piecewise Markov process for analysing survival from breast cancer in different risk groups. *Statistics in Medicine* **20**, 109-122.

Peto, R. (1972). Contribution to the discussion of the paper by D. R. Cox. *Journal of the Royal Statistical Society Series B*, **34**, 205–207.

Saint-Pierre, P., Combescure, C., Daurès, J.P. and Godard, P. (2003). The analysis of asthma control under a Markov assumption with use of covariates. *Statistics in Medicine* **22**, 3775-3770.

Satten, G.A. and Longini, I.M. (1996). Markov chains with measurement error: Estimating the 'true' course of a marker of the progression of human immunodeficiency virus disease. *Journal of the Royal Statistical Society Series C* **45**, 275–295.

Serfling, R.J. (1980). *Approximation theorems of mathematical statistics*. New York, Wiley.

Strauss, D. and Shavell, R. (1998). An extended Kaplan-Meier estimator and its applications. *Statistics in Medicine* **17**, 971-982.

Stute, W. (1993). Consistent estimation under random censorship when covariables are present. *Journal of Multivariate Analysis* **45**, 89-103.

Stute, W. (1995). The central limit theorem under random censorship. *Annals of Statistics* **23**, 422-439.

Stute, W. (1996). Distributional convergence under random censorship when covariables are present. *Scandinavian Journal Statistics* **23**, 461-471.

Stute, W. and Wang, J.L. (1993). The strong law under random censorship. *Annals of Statistics* **21**, 1591-1607.

Su, C.T., Wu, S.C. and Chang, C.C. (2000). Multiaction maintenance subject to action-dependent risk and stochastic failure. *European Journal of Operational Research* **125**, 133-148.

Turnbull, B.W., Brown, B.W. and Hu, M. (1974). Survivorship Analysis of Heart Transplant Data. *Journal of the American Statistical Association* **69**, 74-80.