University

This is to certify that the

dissertation entitled

Regression Models for
Analysis of Medical Costs

presented by

Elena Polverejan

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Statistics

Joseph Gardiner
Major professor
Joseph Gardiner

Date July 31, 2001

REGRESSION MODELS FOR ANALYSIS OF MEDICAL COSTS

By

Elena Polverejan

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Statistics and Probability

2001

REGRES

Risin

interventions

Hospital cos

Knowing the

decisions on

Incre

variables joi

assess the im

parametric o

regression co

correlation b

In an

function of p

LOS and cos

expected cum

We ex

studies when

are incurred i

ABSTRACT

REGRESSION MODELS FOR ANALYSIS OF MEDICAL COSTS

By

Elena Polverejan

Rising cost of health care and the need for evaluating costs of new medical interventions have led to interest in developing methods for medical cost analysis. Hospital costs constitute a significant proportion of overall expenditures in health care. Knowing the correlates of in-hospital length of stay (LOS) and cost is important for decisions on allocating resources.

Increasing availability of patient specific LOS and cost permits analysis of these variables jointly, accounting for their likely correlation. A bivariate model is used to assess the impact of covariates on these outcomes. Under marginal specification through parametric or Cox regression models for LOS and cost, standard errors of estimates of regression coefficients are obtained using a robust covariance matrix to account for correlation between LOS and cost that is otherwise left unspecified.

In another model, we use a conditional approach to estimate mean costs as a function of patient hospital stay and adjusts for the influence of patient characteristics on LOS and cost. The mean cost over a specified duration is a weighted average of the expected cumulative cost, with weighting determined by the distribution of LOS.

We extend this model to address costs and resource utilization in longitudinal studies when patient histories evolve through several health states. In these studies costs are incurred in random amounts at random times as patients transit through different

health states

Markov pro

patient char

of expendit

sojourns in

of all costs

initial state

expected ne

Our

controlling

competing

statistics an

health states. We describe the evolution of a patient's health history by a continuous time Markov process with finite state space. Dependence of the transition intensities on patient characteristics is modeled through semiparametric regression models. Two types of expenditures are incurred, one at transitions between health states and the other for sojourns in a health state. Over a fixed follow up period, we consider net present values of all costs incurred in this period for the two types of expenditures. Conditional on the initial state and a specified covariate vector, we obtain consistent estimates of the expected net present values and derive their asymptotic distributions.

Our methods provide flexible approaches to estimating medical costs while controlling for the effects of covariates. In addition, for economic evaluation studies of competing medical interventions, our methods can be applied to estimate summary statistics and cost-effectiveness ratios.

The beauty of Science is that one can
say something (useful) without having
to say everything.

To my husband and parents, for all of their love and support!

I wo

guidance a

research to

advice on

members

Agency fo

T

Adrian a

encourag

N

underst

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Joseph C. Gardiner for all of his patient guidance and support during my graduate studies. The freedom he has given me in my research to investigate different areas is greatly appreciated and I am also thankful for his advice on the non-statistics related issues. I also want to express my thanks to the other members of my committee for their friendly support. My research was supported by the Agency for Healthcare Research & Quality, under Grant 1R01HS09514.

To my family, I extend my love and thanks for being so supportive. My parents, Adrian and Valentina Feleaga, my sister Anamaria have continually offered me encouragement.

My deepest gratitude goes to my husband Mihai, for all the love and understanding he has given me over the years.

TABLE OF CONTENTS

Table 1

Table 1

Table 1.

Table 2.

LIST OF TABLES

LIST OF FIGURES

AB

CER

CABG

CATH

LOS

DRG

AMI

ICD

i.i.d.

□

■

CLT

SLLN

ABBREVIATIONS

CER	-Cost-Effectiveness Ratio
CABG	-Coronary Artery Bypass Grafting
CATH	-Catheterization
LOS	-Length of Stay
DRG	-Diagnosis Related Group
AMI	-Acute Myocardial Infarction
ICD	-Implantable Cardioverter Defibrillator
i.i.d.	-Independent Identically Distributed
□	-End of Statement
■	-End of Proof
CLT	-Central Limit Theorem
SLLN	-Strong Law of Large Numbers

INTRODUCTION

Over the past decade the need to control health care expenditures in an environment of limited budgets has led health care providers and government planners to turn to cost analyses and cost-effectiveness analyses as an aid to decision making in allocation of health care dollars. While the primary goals of clinical studies are centered on patient outcomes, more attention has being paid to collecting economic data alongside traditional clinical investigations of efficacy of interventions such as randomized clinical trials and prospective cohort studies. Discrete patient-level cost and resource use data will become increasingly available. Therefore there is interest in developing rigorous statistical techniques to analyze both cost and health outcomes.

In many situations two competing interventions need to be compared on their health benefits and costs. When an intervention is more effective and more costly than its comparator, the cost-effectiveness ratio (CER) is defined as the ratio of the incremental cost relative to the incremental benefit. With benefits measured in their natural units such as years of life saved or number of lives saved and costs measured in dollars, the CER is stated in dollars per unit of effectiveness. When health benefit is measured by gain in life expectancy, the cost-effectiveness ratio is the additional cost of the new intervention to deliver one unit of benefit and is expressed in dollars per life year saved. Assessment and estimation of the CER are an important part of conducting economic evaluations of health care programs. A goal of our research is to address the specification, estimation and evaluation of statistical methods for costs and to demonstrate

developing

appropriate

Usu

resource and

role of time i

manifest in ea

accumulating

random even

in the treatm

death and th

pharmacolo

mg/dL, the

so $C(T)$ is

stay (denot

Est

some of the

demograph

severity of

next goal m

$E(C(T)|Z)$

hospital cost

patient, age 6

demonstrate their application to cost-effectiveness analysis. In this thesis we focus on developing rigorous models for analysis of cost data. Integrating these cost models with appropriate models for assessing health outcomes is the subject of future research.

Usually costs accrue due to resource use over time. Both the amount of the resource and the time at which it is used vary across individuals. To fully integrate the role of time into analyses of costs we would need the cumulative cost histories as they manifest in each patient. For a treatment or intervention under study, let $C(t)$ denote the accumulating cost over time t in an individual patient. Expenditures terminate at a random event time T that signals the occurrence of some health outcome. For example, in the treatment of cancer patients following diagnosis, the study endpoint T is the time of death and the cost at the endpoint $C(T)$ is the lifetime cost from diagnosis. In a pharmacological intervention in patients with serum cholesterol elevated above 240 mg/dL, the endpoint T might be the first time the cholesterol level falls below 200 mg/dL, so $C(T)$ is the total treatment cost. For hospital cost studies the endpoint is the length of stay (denoted LOS), and $C(T)$ is the cumulative cost from admission through discharge.

Estimating the distribution of the total cost $C(T)$ or assessing its correlates are some of the objectives of cost studies. Correlates (called also covariates) might be demographic factors such as age, gender, race, education or clinical factors such as severity of the disease. Once the significant correlates for total cost are determined, the next goal might be the estimation of a summary statistic such as the expected total cost $E(C(T)\,|\,Z)$ or the median cost $m(Z)$ for specified covariate profiles. For example, in a hospital cost study one might want to estimate the expected total hospital cost for a male patient, age 65 at admission, hospitalized for acute myocardial infarction (AMI), undergoing

patient cost

$E(C(t)|Z)$

cost example

cost after 9 d

for AMI, und

Statist

include right-

censored cost

observations

several perio

has to cope v

The

simple meth

$C=C(T)$ will

used to miti

applied to th

effects of pa

Using an ob

model $g(C,$

undergoing coronary artery bypass surgery (CABG). If there is interest in modeling patient cost histories, one objective might be the estimation of the expected cost $E(C(t) \mid Z)$ over a given fixed duration *t*, for specified covariate profiles. In the hospital cost example, the objective of interest could be the estimation of the expected hospital cost after 9 days of hospitalization for a male patient, age 65 at admission, hospitalized for AMI, undergoing CABG.

Statistical analyses of cost data must address several technical problems.[1] There include right-skewed cost data, a significant proportion of zero observations, right-censored cost data, correlation between time and cost outcomes and dependent observations when costs are ascertained at multiple time points (for instance in each of several periods during the course of an intervention). Every statistical model for costs has to cope with at least one of these issues.

Skewness and Transformation

The distribution of costs might exhibit a considerable degree of skewness. Then simple methods of analysis based on an assumed normal distribution of the cost variable *C=C(T)* will not be tenable. A transformation *g* such as logarithm or square-root can be used to mitigate the effects of this skewness. Standard regression analyses may then be applied to the transformed dependent variable *g(C)*, which permit assessment of the effects of patient and intervention characteristics *Z* that influence the cost distribution. Using an observed sample $\{(C_i, Z_i) : 1 \le i \le n\}$, least-squares estimation of β in the model $g(C_i) = Z_i'\beta + \varepsilon_i$ needs only a simple moment structure for the errors ε_i.[2, 3] A retransform

of measurem

$\log(C_i) = Z'$

cost at a spe

$\varepsilon_i \sim N(0,\sigma^2$

mean cost E

normality is

residuals log

estimation of

and median c

suspect, the e

inference. W

of C and a par

simplicity of t

approach that

When s

during a given

have incurred a

only for users a

retransformation then reproduces the results of these analyses back in their original units of measurement, permitting easy interpretation. For example, we could use the model $\log(C_i) = Z_i'\beta + \varepsilon_i$, where the errors ε_i have zero mean and variance σ^2. Then the mean cost at a specified covariate profile Z_0 is $E(C \mid Z_0) = \exp(\beta' Z_0) E(\exp \varepsilon)$. If $\varepsilon_i \sim N(0, \sigma^2)$, so the costs are log-normally distributed, the simple closed form for the mean cost $E(C \mid Z_0) = \exp(\beta' Z_0 + 0.5\sigma^2)$ make the analyses quite straightforward. If normality is untenable, one can use a smearing estimate for $E(\exp \varepsilon)$, based on the residuals $\log C_i - \hat{\beta}' Z_i = \hat{\varepsilon}_i$.[4] In the absence of covariates, classical maximum likelihood estimation of parameters in log-normal distributions have been used to compare mean and median costs in two independent samples.[5] When the parametric assumptions are suspect, the estimates of the standard error of β could be imprecise, leading to invalid inference. When transformations can substantially eliminate the skew in the distribution of C and a parametric distribution can be assumed for ε, the advantage lies in the relative simplicity of the analysis and greater efficiency of estimates compared to a nonparametric approach that leaves the distribution of ε unspecified.[4]

Two-Part Models

When sampling an eligible population for assessing the costs of medical services during a given period, a large proportion would not have used any services so would not have incurred any medical costs. In these circumstances, the cost measure C is positive only for users and the zero costs cannot be ignored econometrically. The *two-part model* assumes that $P(C>0|Z)$ is governed by a parametric binary probability model like logit or probit (part one) and that $E(g(C)|C>0,Z)$ is a linear function of Z (part two), where g is a transformation applied to moderate the effects of skewness. The objective is to obtain an estimate of the overall mean $E(C|Z)$.[6] The first part governs the probability of some expenditure and the second part models the level of the expenditure, given that there is a positive expense.

Two-part models are used not only for medical costs, but also for many other outcomes, such as measures of health care utilization (e.g. number of physician visits over a specified period), health care outcomes (claims data) or measures of use of substance use / abuse (tobacco, alcohol, illicit drugs).

Right-Censored Cost Observations

Due to incomplete patient follow-up, in many studies the endpoint T and the total cost $C(T)$ for some patients are right-censored. For example, in clinical trials with staggered entry of patients, the signaling event of interest might not have occurred by the close of the trial. If U denotes the follow-up time for a subject, with right-censorship the observable data are restricted to $X=\min(T,U)$, the smaller of T and U, the indicator of non-censoring $\delta=[T\leq U]$ that denotes whether T (if $\delta=1$) or U (if $\delta=0$) was observed, and the covariate vector Z. If T is not censored we observe the true cost $C(T)$. If T is censored we observe the cost up to the follow-up time U, but we know that $C(T)>C(U)$.

Because of the analogy with censored survival times, censored medical costs have been analyzed by Cox regression and other survival analysis techniques. An earlier work by Dudley *et al.*[7] explored the idea by comparing different analytic models for the cost of CABG surgery. In addition to the Cox model, these investigators studied the OLS method, with and without a log transformation of the cost variable C, a parametric Weibull regression model, and a binary logistic model using a dichotomization of C. In their analysis, the Cox model provided the most accurate estimates of the mean and median cost of CABG surgery, and the proportion of patients with high cost (>$20,000). This method was successfully reapplied to assess the determinants of costs in CABG surgery.[8]

Recent articles[9-12] have questioned the appropriateness of survival analytic methods for medical costs particularly in the treatment of censoring. The total costs $C(T)$ would not, in general, be independent of $C(U)$ even if T and U are independent. To apply standard survival analysis methods, we would need the independence of $C(T)$ and $C(U)$ given the covariate profile Z.

When cumulative cost histories are available over time, greater flexibility in modeling is possible that skirt the issue of censored costs. In a discussion of different models for predicting the cost of illness, Lipscomb *et al.*[1] applied a proportional hazards model for the cost intensity $\alpha(c \mid Z)$ on a large data set of Medicare patients hospitalized for stroke. Costs were analyzed for a 36-month period following hospital discharge. The unit of analysis was a patient-month making the potential cost incurred in month (j) right censored if the patient died during that month. If this occurred, only costs through the first ($j-1$) months were considered, thereby skirting the issue of censored costs. Any dependence of costs in month (j) on the stroke patient's cost history is captured in the regression model by using as covariates the initial cost of hospitalization and costs in follow up months ($j-1$) and ($j-2$). Also included were patient characteristics such as age, race, gender and economic status. The investigators found that the Cox proportional hazards model and the two-part model were superior in their ability to predict accurately the distribution of costs, based on a logarithmic scoring rule to compare models. For predictions of mean and median costs these models and log-transformed linear models performed equally well.

Because $C(T)$ and $C(U)$ cannot in general be independent, Lin *et al.*[10] proposed two alternative ways to analyze cost in trials with incomplete patient follow-up. An assumption is made that patients are not censored because they accrue unusually high or low costs. Under one approach, if cost histories are available for each patient, an estimate called the Kaplan-Meier Sampling Average (KMSA) estimate of the average cost $EC(T)$ is computed. It is essentially an average of costs incurred in each of several time periods in $[0,T]$, weighted by the Kaplan-Meier estimate of survival at the start of each period. If cost histories are not available, the second approach bases cost estimates on the subset of patients who experience the "event" at issue (in their example death). The properties of these estimators are dependent on the assumption of discrete censoring times, which is not true in general. Bang and Tsiatis[13] introduced a class of weighted estimators for mean medical costs, which account appropriately for censoring. Besides the consistency and the asymptotic normality, the efficiency of their estimators was also studied. None of these methods incorporate covariates in the cost modeling.

The KMSA technique to defining average cost is similar to the approach taken by Gardiner *et al.*[14, 15] in evaluating the cost-effectiveness of the Implantable Cardioverter Defibrillator (ICD). Expected total cost over a fixed time interval $[0, t_0]$ was defined by $\int_0^{t_0} e^{-rt} S(t) dC(t)$, where r is the discount rate, S the survival function and $C(t)$ the value of cumulative resource use up to time t. Apart from the discounting, the integrand weights by $S(t)$ the incremental expenditure $dC(t)$ in the small interval $[t, t+dt]$. The cost $C(.)$ was assumed to be nonstochastic. It was derived from Medicare payments, drug charges and physician fees for services that were associated with the interventions. In their application to the cost-effectiveness of the ICD, cost comprised of expenditures over 72, 30-day periods.

Very recently methods have been proposed to explicitly account for cost censoring and also adjust for patient characteristics. Lin[16] modified in several ways the familiar normal equations for least-squares estimation. The total cost in subjects with complete follow-up is used in the proposed methodology. More efficient estimators are provided when cost data are recorded in multiple time intervals. Lin[11] also developed a methodology that specifies multiplicative rather than additive covariate effects on the mean medical cost. He proposed a semiparametric proportional means regression model for the cumulative medical cost. This model specifies that the mean cost function over time, conditional on a set of covariates, is equal to an arbitrary baseline mean function multiplied by an exponential regression function. The corresponding inference procedures are based on possibly censored observations of the lifetime cost.

Objectives and Structure of Thesis

The existing literature on cost analysis is still in its infancy. Several issues of practical importance have yet to be address. We believe that our research will help fill some methodological gaps on modeling and estimation of medical costs, particularly in longitudinal studies.

With the growing availability of large databases on patient-level health care utilization and outcomes, there is need to develop statistical techniques to analyze *jointly* both costs and patient outcomes. Current methods generally focus on a single measure of cost or health outcome and do not fully exploit the longitudinal dynamic mechanisms that engender cost and health outcome data. How individual characteristics might impact summary statistics such as mean, median cost and survival are key to predicting resource utilization and informing policy on allocating health care dollars.

A simple, although important situation in which our proposed methods would apply is in assessing the correlates of hospital cost and LOS jointly. Several studies have used LOS as a proxy for resource utilization[17, 18], while others have focused on the correlates of hospital cost or charge[19-22]. Because of the likely correlation between cost and LOS, it is unclear if a model that explicitly recognizes this correlation would lead to different quantitative results. Other issues such as the skewness in the distribution of cost, whether or not in-hospital deaths should be regarded as censoring events also add to the complexity of the joint analyses of LOS and hospital cost.

In **Chapter 1** we develop a bivariate model for the two outcomes, LOS and cost, from a common constellation of covariates that might influence their joint distribution.

Cost per patient, measured in monetary units, is the total resource use from admission to discharge, and LOS, measured in time units, is the duration of the hospital stay. Our regression analyses account for three important aspects. First, because the distributions of cost and LOS are skewed, for each outcome we use either a semiparametric Cox model or a linear model applied to a transformation of the outcome. Second, the model accounts for incomplete observations in both outcomes. Finally, although the correlation between cost and LOS is not a primary concern, we use methods developed for multiple failure times[23] to adjust for its impact on the standard errors of regression coefficients. The proposed methods are applied to assess the influence of comorbidity and demographic factors on LOS and hospital costs in a cohort of patients who underwent CABG surgery.

The longitudinal framework that underlies survival analytic techniques provides a natural setting for a complete specification of alternative models for estimating costs. In Chapters 2 and 3 we develop several models of increasing complexity for health care costs and outcomes. In these models costs are considered to dynamically evolve over time.

In **Chapter 2** we focus on cost alone including LOS among other patient characteristics as potential correlates of the accumulating hospital cost. The model permits estimation of mean cost over a given duration of hospital stay. The mean cost over a specified duration is a weighted average of the expected cumulative cost, with weighting determined by the distribution of LOS. We demonstrate the application of this technique using the same study as in Chapter 1. The described model can be incorporated in a more general setup, in longitudinal studies with multiple health states and transitions between them.

When a health care intervention is deployed, costs are engendered through the use of resources. These occur in random amount at random times that might differ among patients. Incorporating these components into statistical models that accurately reflect the patient health histories permits consideration of health outcomes and costs jointly. In the actuarial literature it is common to model the stochastic mechanism governing the events that trigger the payment of life insurance benefits as a Markov chain with finite state space, where each sample path is interpreted as the life history of the insured.[24-28] We extend and adapt these models to our context.

In **Chapter 3** we propose longitudinal stochastic models that reflect the experience of patients in sustaining and changing states of health. We use a Markov model[29] to describe the evolution of patient histories over time. Dependence of the transition intensities on patient characteristics is modeled through the Cox regression model. We consider two types of costs that might be incurred in the course of follow-up: costs at transitions between health states (eg., cost of diagnosis of a condition) and costs of sojourns in a health state (eg., cost of the treatment of that particular condition). Present values are obtained by discounting all expenditures at a fixed rate. Conditional on the initial state and a specified covariate profile, we provide estimators of the expected present value of these two types of costs incurred over a fixed follow-up period. Under additional assumptions, these estimators can be shown to be consistent and asymptotically normal, which sets the stage necessary for statistical inference.

Our proposed methods have the capability of incorporating concomitant covariate information for the time and cost outcomes. Both fully parametric and semiparametric models were studied, including regression models for the transformed response variables, Cox regression and Markov models that specify covariate effects in transition intensities. These models are based on a natural time-cost setting and have easy interpretation. As a result, the methodologies developed in this thesis are very promising to provide a flexible, unified framework for statistical inference on summary statistics (such as CER) used in cost-effectiveness analysis.

CHAPTER 1

A BIVARIATE MODEL FOR HOSPITAL LENGTH OF STAY AND COST

Hospitalizations constitute a significant proportion of overall expenditure in health care. Length of stay (LOS) is often used as a surrogate for hospital cost or charge. However, the increasing availability of databases with patient specific LOS and cost permits analyses of these variables jointly, accounting for their likely correlation. In this chapter we explore the differences and advantages of using a bivariate model, compared to separate univariate models for assessing the impact of covariates on LOS and cost. Under marginal specification through semiparametric or fully parametric regression models for LOS and cost, standard errors of estimates of regression coefficients are obtained using a robust covariance matrix[23, 30] to account for correlation between LOS and cost. These models account for incomplete observations in both outcomes.

In Section 1.1 we use a Cox regression model for each of the LOS and cost outcomes and in Section 1.2 a parametric model. In Section 1.3 we apply the proposed methods to LOS and hospital cost in a cohort of patients who underwent coronary artery bypass surgery (CABG).

1.1 Semiparametric Marginal Models

Consider n individuals in a study. Censoring occurs when assessment of LOS and cost are made at some fixed calendar time. This is called administrative censoring. At that time some patients may not have completed their LOS or incurred all their costs.

For the i-th patient we observe the true LOS T_i or the censoring time T_i', whichever occurs first, and $Z_{1i}(.)$, a vector of p explanatory variables that depends on the current time t since admission. We restrict the time t to a finite interval $[0,\tau_1]$, $\tau_1 < \infty$. The nonnegative variables T_i, T_i' and the process $Z_{1i}(.)$ are defined on the probability space $(\Omega_1, \mathcal{F}_1, P_1)$. Let $X_{1i} = \min(T_i, T_i')$, $\delta_{1i} = [T_i \le T_i']$, where $[.]$ is the indicator function of the displayed event.

For the i-th patient we also observe $X_{2i} = \min(C_i, C_i')$, $\delta_{2i} = [C_i \le C_i']$ and $Z_{2i}(.)$, where C_i is the total hospital cost, C_i' is the censoring cost and $Z_{2i}(.)$ a vector of p explanatory variables that depends on the current cost c since admission. The costs are restricted to a finite interval $[0,\tau_2]$, $\tau_2 < \infty$. The nonnegative variables C_i, C_i' and the process $Z_{2i}(.)$ are defined on the probability space $(\Omega_2, \mathcal{F}_2, P_2)$.

The following hazard functions α_{1i}, α_{2i} relate the covariates to the distributions of T_i and C_i.

For LOS: $\quad \alpha_{1i}(t, \beta_{10}) = \alpha_{10}(t)\exp(\beta_{10}' Z_{1i}(t))$. (1.1)

For cost: $\quad \alpha_{2i}(c, \beta_{20}) = \alpha_{20}(c)\exp(\beta_{20}' Z_{2i}(c))$. (1.2)

We assume the vectors of true values of the regression parameters β_{10}, β_{20} to have dimension p. The underlying intensities $\alpha_{10}(.)$, $\alpha_{20}(.)$ are the baseline intensities corresponding to zero covariates and they are left completely unspecified. The subscripts will allow us to distinguish between covariate effects on LOS and cost. Given a fixed covariate profile Z_0 and $T_i \geq t$, the hazard $\alpha_{1i}(t, \beta_{10} \mid Z_0)$ is the instantaneous probability that LOS would end just after time t. Similarly, given Z_0 and $C_i \geq c$, the hazard $\alpha_{2i}(c, \beta_{20} \mid Z_0)$ is interpreted as the instantaneous probability that total cost would be realized just above the level c.[1]

With independent identically distributed (i.i.d.) data $(X_{1i}, \delta_{1i}, Z_{1i}(.))$ on n patients we obtain on the basis of (1.1) the estimate $\hat{\beta}_1$ of β_{10} by maximum partial likelihood estimation and the Nelson-Aalen estimate $\hat{A}_{10}(t, \hat{\beta}_1)$ of the integrated baseline hazard $A_{10}(t) = \int_0^t \alpha_{10}(u)du$. Analogously, with i.i.d. data $(X_{2i}, \delta_{2i}, Z_{2i}(.))$ we use (1.2) to obtain the corresponding estimates $\hat{\beta}_2$ and $\hat{A}_{20}(c, \hat{\beta}_2)$. Following Wei *et al.* (1989)[23], $(\hat{\beta}_1', \hat{\beta}_2')'$ has an asymptotic 2p-variate normal distribution whose covariance matrix can be consistently estimated.

The survival distributions of LOS and cost at a fixed (i.e. time-independent) covariate profile Z_0 are respectively, $S_1(t \mid Z_0) = P(T_1 > t \mid Z_0) = \exp(-A_{10}(t)e^{\beta_{10}' Z_0})$ and $S_2(c \mid Z_0) = P(C_1 > c \mid Z_0) = \exp(-A_{20}(c)e^{\beta_{20}' Z_0})$. Their estimates, denoted by $\hat{S}_1(t \mid Z_0)$ and $\hat{S}_2(c \mid Z_0)$, are obtained by replacing the unknown quantities by the aforementioned estimates. We will show that for fixed time t and cost c, given a fixed covariate profile

Z_0, $\left\{n^{1/2}\left(\hat{S}_1(t\,|\,Z_0)-S_1(t\,|\,Z_0)\right), n^{1/2}\left(\hat{S}_2(c\,|\,Z_0)-S_2(c\,|\,Z_0)\right)\right\}'$ is asymptotically bivariate normal, with zero-mean vector and a covariance matrix *CS* that can be consistently estimated from the data. We call *CS* the adjusted covariance matrix. Approximate pointwise 95% confidence intervals for $S_1(t\,|\,Z_0)$ and $S_2(c\,|\,Z_0)$ are calculated and point estimates and approximate 95% confidence intervals for the median LOS and median cost are obtained from the estimated survival curves by the procedure described in p511-512, Andersen *et al.* (1993).[29]

1.1.1 Estimation of the Regression Parameters and Integrated Baseline Hazards

Define of each patient the processes $N_{1i}(t)=[X_{1i}\le t,\delta_{1i}=1]$ and $Y_{1i}(t)=[X_{1i}\ge t]$. Aggregated over all *n* patients, the processes $N_1(t)=\sum_{i=1}^{n}N_{1i}(t)$ and $Y_1(t)=\sum_{i=1}^{n}Y_{1i}(t)$ denote respectively the number of patients with completed LOS by time *t* and the number who have not completed their hospital stay at time just prior to time *t* from admission. Similarly define $N_{2i}(c)$, $Y_{2i}(c)$. Here the aggregated process $N_2(c)$ denotes the number of patients whose total hospital costs, completely observed, do not exceed *c*, and $Y_2(c)$ the number of patients whose current hospital cost is least *c*. In the sequel for notational convenience we will use a single generic argument *u* for all processes remembering that the subscript 1 is associated with time and the subscript 2 with cost.

We need some standard notation. Let

$$S_k^{(0)}(u,\beta_k)=\sum_{i=1}^{n} Y_{ki}(u)\exp(\beta_k' Z_{ki}(u)),$$

$$S_k^{(1)}(u,\beta_k)=\sum_{i=1}^{n} Y_{ki}(u)Z_{ki}(u)\exp(\beta_k' Z_{ki}(u)),$$

$$S_k^{(2)}(u,\beta_k)=\sum_{i=1}^{n} Y_{ki}(u)Z_{ki}(u)Z_{ki}'(u)\exp(\beta_k' Z_{ki}(u)),\ k=1,2.$$

Note that $S_k^{(0)}$ is a scalar, $S_k^{(1)}$ a p-dimensional vector and $S_k^{(2)}$ a $p\times p$ matrix.

Define

$$E_k(u,\beta_k)=S_k^{(1)}(u,\beta_k)/S_k^{(0)}(u,\beta_k),$$

$$V_k(u,\beta_k)=\left\{S_k^{(2)}(u,\beta_k)/S_k^{(0)}(u,\beta_k)\right\}-E_k(u,\beta_k)E_k(u,\beta_k)',$$

$$I_k(\beta_k)=\int_0^{\tau_k} V_k(u,\beta_k)dN_k(u).$$

Let

$$s_k^{(0)}(u,\beta_k)=E\left(Y_{k1}(u)\exp(\beta_k' Z_{k1}(u))\right),$$

$$s_k^{(1)}(u,\beta_k)=E\left(Y_{k1}(u)Z_{k1}(u)\exp(\beta_k' Z_{k1}(u))\right),$$

$s_k^{(2)}(u,\beta_k)=E\left(Y_{k1}(u)Z_{k1}(u)Z_{k1}(u)'\exp(\beta_k' Z_{k1}(u))\right)$, where $E(.)$ denotes the expectation of the displayed quantity with respect to the corresponding probability.

Under some regularity conditions [31], $n^{-1}S_k^{(m)}(u,\beta_k)$ converges in probability to $s_k^{(m)}(u,\beta_k)$, uniformly in some neighborhood of β_{k0} and in $u\in[0,\tau_k]$. We will provide later some details. Define

$$e_k(u,\beta_k)=s_k^{(1)}(u,\beta_k)/s_k^{(0)}(u,\beta_k),$$

$$v_k(u,\beta_k)=\left\{s_k^{(2)}(u,\beta_k)/s_k^{(0)}(u,\beta_k)\right\}-e_k(u,\beta_k)e_k(u,\beta_k)',$$

$$\Sigma_k(\beta_k) = \int_0^{\tau_k} v_k(u,\beta_k) s_k^{(0)}(u,\beta_k)\alpha_{k0}(u)du .$$

We formulate our models and prove many of the results in the framework of multivariate counting processes. A survey of this theory can be found in Andersen *et al.* (1993).[29] Our notations follow the ones of this reference.

The following list of conditions will be assumed to hold throughout this section:

Model Assumptions:

A.1 Conditional on $Z_{1i}(.)$, T_i and T_i' are independent and conditional on $Z_{2i}(.)$, C_i and C_i' are independent;

A.2 $\left\{ \begin{pmatrix} X_{1i} \\ X_{2i} \end{pmatrix}, \begin{pmatrix} \delta_{1i} \\ \delta_{2i} \end{pmatrix}, \begin{pmatrix} Z_{1i}(.) \\ Z_{2i}(.) \end{pmatrix} \right\}, 1 \le i \le n$ are i.i.d.;

A.3 $A_{10}(\tau_1) = \int_0^{\tau_1} \alpha_{10}(t)dt < \infty$, $A_{20}(\tau_2) = \int_0^{\tau_2} \alpha_{20}(c)dc < \infty$;

A.4 $Z_{1i}(.)$, $Z_{2i}(.)$ are bounded;

A.5 $Z_{1i}(.)$, $Z_{2i}(.)$ are adapted, left-continuous with right-hand limits processes;

A.6 $P_k(Y_{k1}(u) = 1, \forall u \in [0,\tau_k]) = P(Y_{k1}(\tau_k) = 1) > 0$;

A.7 $\Sigma_k \underset{notation}{\overset{by}{=}} \Sigma_k(\beta_{k0})$ is positive definite.

Note:

A.1 is also called the independent censoring assumption. A.2 implies that $\left(N_{ki}(u), Y_{ki}(u), Z_{ki}(u), u \in [0,\tau_k]\right)$, $1 \le i \le n$ are i.i.d. A.4 and A.5 assure that $Z_{1i}(.)$,

$Z_{2i}(.)$ are bounded, predictable processes. For $k=1$, the interpretation of A.6 is that there is a positive probability that at any time t from admission a subject might not have completed his/her hospital stay. For $k=2$, the interpretation of A.6 is that there is a positive probability that the total cost of any subject might be larger than any cost c from admission. Assumption A.7 is crucial for the (asymptotic) existence of the regression parameter estimator. □

Consider on the probability space $(\Omega_1, \mathcal{F}_1, P_1)$ the right-continuous nondecreasing family $\left(\mathcal{F}_{1i}(t), t \in [0, \tau_1]\right)$, where $\mathcal{F}_{1i}(t)$ represents everything that happens up to the time t for the i-th patient. Formally $\mathcal{F}_{1i}(t) = \mathcal{F}_{1i}^Z(t) \vee \mathcal{F}_{1i}^N(t)$, where $\mathcal{F}_{1i}^N(t) = \sigma\{N_{1i}(s), s \le t\}$, $\mathcal{F}_{1i}^Z(t) = \sigma\{Z_{1i}(s), s \le t\}$. Similarly we define the filtration $\left(\mathcal{F}_{2i}(c), c \in [0, \tau_2]\right)$ on the probability space $(\Omega_2, \mathcal{F}_2, P_2)$.

Under the independent censoring assumption, $N_{ki}(.)$ is a counting process on $(\Omega_k, \mathcal{F}_k, P_k)$ with the $\mathcal{F}_{ki}(.)$ – intensity process $\lambda_{ki}(v, \beta_k) = \alpha_{ki}(v, \beta_k) Y_{ki}(v)$, where the hazard function $\alpha_{ki}(., \beta_k)$ was defined in (1.1), (1.2). The processes M_{ki} defined by

$$M_{ki}(u) = N_{ki}(u) - \int_0^u \lambda_{ki}(v, \beta_k) dv, \; u \in [0, \tau_k] \tag{1.3}$$

are $\mathcal{F}_{ki}(.)$ – local square integrable martingales on the interval $[0, \tau_k]$, with $\langle M_{ki} \rangle (u) = \int_0^u \lambda_{ki}(v, \beta_k) dv$ and $\langle M_{ki}, M_{kj} \rangle = 0$ for $i \ne j$, i.e. M_{ki} and M_{kj} are orthogonal for $i \ne j$. The process $\Lambda_{ki}(u, \beta_k) = \int_0^u \lambda_{ki}(v, \beta_k) dv$ is called the compensator of the counting process $N_{ki}(.)$. For details of this result see Andersen *et al.* (1993)[29], Sections II.4.1 and III.2.2.

Let $\Omega_k^{(n)} = \Omega_k \otimes ... \otimes \Omega_k$, $\mathcal{F}_k^{(n)} = \mathcal{F}_k \otimes ... \otimes \mathcal{F}_k$, $P_k^{(n)} = P_k \otimes ... \otimes P_k$ and $\mathcal{F}_k^{(n)}(u) = \mathcal{F}_{k1} \otimes ... \otimes \mathcal{F}_{kn}$, where the product $\otimes ... \otimes$ is over n factors. The family $\left(\mathcal{F}_k^{(n)}(u), u \in [0, \tau_k]\right)$ is a filtration on the n-th sample space $(\Omega_k^{(n)}, \mathcal{F}_k^{(n)}, P_k^{(n)})$. Then (see Andersen *et al.* (1993)[29], Section I.4.3) N_{ki} has the same compensator Λ_{ki} with respect to the product sample space $(\Omega_k^{(n)}, \mathcal{F}_k^{(n)}, P_k^{(n)})$ and the filtration $\left(\mathcal{F}_k^{(n)}(u), u \in [0, \tau_k]\right)$. When LOS and cost quantities are considered jointly, the stochastic properties are relative to the filtration $\left(\mathcal{F}_1^{(n)}(t) \otimes \mathcal{F}_2^{(n)}(c), (t, c) \in [0, \tau_1] \times [0, \tau_2]\right)$ on the product space $(\Omega_1^{(n)} \otimes \Omega_2^{(n)}, \mathcal{F}_1^{(n)} \otimes \mathcal{F}_2^{(n)}, P_1^{(n)} \otimes P_2^{(n)})$.

An estimator $\hat{\beta}_1$ of β_{10} is obtained by maximizing the Cox partial likelihood (see Andersen *et al.* (1993)[29], p483-484). The log-partial likelihood evaluated at time/cost u has the form

$$C_k(u, \beta_k) = \sum_{i=1}^n \int_0^u \beta_k' Z_{ki}(v) dN_{ki}(v) - \int_0^u \log S_k^{(0)}(v, \beta_k) dN_k(v).$$

The vector $U_k(u, \beta_k)$ of derivatives of $C_k(u, \beta_k)$ with respect to β_k is

$$U_k(u, \beta_k) = \sum_{i=1}^n \int_0^u Z_{ki}(v) dN_{ki}(v) - \int_0^u \frac{S_k^{(1)}(v, \beta_k)}{S_k^{(0)}(v, \beta_k)} dN_k(v). \tag{1.4}$$

The **maximum partial likelihood estimator** $\hat{\beta}_k$ of β_{k0} is defined as the solution of the likelihood equation $U_k(\tau_k, \beta_k) = 0$. Then the **Nelson-Aalen estimator** of the integrated baseline hazard $A_{k0}(u) = \int_0^u \alpha_{k0}(v)dv$ is given by

$$\hat{A}_{k0}(u, \hat{\beta}_k) = \int_0^u \frac{J_k(u)}{S_k^{(0)}(v, \hat{\beta}_k)} dN_k(v),$$

where $J_k(v) = [Y_k(v) > 0]$ (see Andersen *et al.* (1993)[29], Sections IV.1 and VII.2.1).

1.1.2 Large Sample Properties of the Estimators of Regression Parameters and Integrated Baseline Hazards

The following conditions are necessary for the asymptotic properties of our estimators. They were first introduced by Andersen and Gill (1982).[31] We use their formulation from Andersen *et al.* (1993).[29] Throughout this chapter the norm of a vector $a = (a_i)$ or a matrix $A = (a_{ij})$ is $\|a\| = \sup_i |a_i|$ and $\|A\| = \sup_{i,j} |a_{ij}|$, respectively.

Conditions C.a-C.f:

There exist a compact neighborhood $\mathcal{B}_k$ of β_{k0}, with $\beta_{k0} \in \overset{\circ}{\mathcal{B}}_k$ (the interior of $\mathcal{B}_k$), and scalar, vector and matrix functions $s_k^{(0)}$, $s_k^{(1)}$, $s_k^{(2)}$ defined on $[0, \tau_k] \times \mathcal{B}_k$ such that for $m \in \{0, 1, 2\}$:

C.a $$\sup_{(u, \beta_k) \in [0, \tau_k] \times \mathcal{B}_k} \left\| \frac{1}{n} S_k^{(m)}(u, \beta_k) - s_k^{(m)}(u, \beta_k) \right\| \xrightarrow{P} 0;$$

C.b $s_k^{(m)}(.,.)$ is a uniformly continuous bounded function of $(u, \beta_k) \in [0, \tau_k] \times \mathcal{B}_k$;

C.c $s_k^{(0)}(.,.)$ is bounded away from zero;

C.d $s_k^{(1)}(u, \beta_k) = \dfrac{\partial}{\partial \beta_k} s_k^{(0)}(u, \beta_k)$, $s_k^{(2)}(u, \beta_k) = \dfrac{\partial}{\partial \beta_k} s_k^{(1)}(u, \beta_k)$;

C.e Σ_k is positive definite;

C.f $\int_0^{\tau_k} \alpha_{k0}(u) du < \infty$.

Under our model assumptions A.1-A.7 the conditions C.a-C.f are verified for the functions $s_k^{(0)}$, $s_k^{(1)}$, $s_k^{(2)}$ defined in the previous sub-section. For a proof and a discussion about these regularity conditions see Section 4 of Andersen and Gill (1982).[31] Under these general conditions, with a probability tending to one, there exists a unique consistent solution $\hat{\beta}_k$ of the likelihood equation:

Theorem 1 (Theorem VII.2.1, p497, Andersen *et al.* (1993)[29])

Under the assumptions A.1, A.2 and conditions C.a-C.f, the probability that the equation $U_k(\tau_k, \beta_k) = 0$ has a unique solution $\hat{\beta}_k$ tends to 1 and $\hat{\beta}_k \xrightarrow{P} \beta_{k0}$ as $n \to \infty$. □

The assumption of bounded covariates is very important to prove the asymptotic normality of $(\hat{\beta}_1', \hat{\beta}_2')'$. This assumption implies the Lindeberg-type condition used in Andersen *et al.* (1993).[29] Wei *et al.* (1989)[23] proved the following theorem. Because some intermediate steps of the proof of this theorem will be used later, we sketch them below.

Theorem 2

Under the model assumptions A.1-A.7, $n^{1/2}\left(\hat{\beta}_1' - \beta_{10}', \hat{\beta}_2' - \beta_{20}'\right)'$ converges in distribution to a zero mean normal $2p$-dimensional random vector with covariance matrix $Q = \left(D_{kl}, k, l \in \{1,2\}\right)$, where the $p \times p$ matrix D_{kl} is

$$D_{kl} = \Sigma_k^{-1} E\left(w_{k1}(\beta_{k0}) w_{l1}(\beta_{l0})'\right) \Sigma_l^{-1},$$

with $w_{ki}(\beta_{k0})$ a p-dimensional vector,

$$w_{ki}(\beta_{k0}) = \int_0^{\tau_k} \left(Z_{ki}(u) - e_k(u, \beta_{k0})\right) dM_{ki}(u). \square$$

Sketch of the proof:

By (1.4) the score function $U_k(u, \beta_k)$ has the form

$$U_k(u, \beta_k) = \sum_{i=1}^n \int_0^u \left(Z_{ki}(v) - E_k(v, \beta_k)\right) dN_{ki}(v).$$

Replacing $N_{ki}(v)$ by (1.3), it follows immediately that

$$U_k(u, \beta_k) = \sum_{i=1}^n \int_0^u \left(Z_{ki}(v) - E_k(v, \beta_k)\right) dM_{ki}(v).$$

By Taylor expansion of $U_k(\tau_k, \beta_k)$ around β_{k0}, we have

$$n^{-1/2} U_k(\tau_k, \beta_{k0}) = \left(n^{-1} I_k(\beta_k^*)\right) n^{1/2}\left(\hat{\beta}_k - \beta_{k0}\right),$$

where $-I_k(\beta_k)$ is the matrix of derivatives of $U_k(\tau_k, \beta_k)$ with respect to β_k and β_k^* is on the line segment between $\hat{\beta}_k$ and β_{k0}.

Step 1:

$n^{-1/2}\left(U_1(\tau_1,\beta_{10})', U_2(\tau_2,\beta_{20})'\right)'$ converges in distribution to a zero mean normal $2p$-dimensional random vector. The asymptotic covariance is $B = \left(B_{kl}, k,l \in \{1,2\}\right)$, where $B_{kl} = E\left(w_{k1}(\beta_{k0})w_{l1}(\beta_{l0})'\right)$.

Step 2:

$n^{-1}I_k(\beta_k^*) \xrightarrow{P} \Sigma_k$ for any random β_k^* such that $\beta_k^* \xrightarrow{P} \beta_{k0}$.

Step 3:

$n^{1/2}\left(\hat{\beta}_1' - \beta_{10}', \hat{\beta}_2' - \beta_{20}'\right)'$ converges in distribution to a zero mean normal $2p$-dimensional random vector with asymptotic covariance matrix $Q = A^{-1}BA^{-1}$, where $A = diag(\Sigma_1, \Sigma_2)$ and B was defined in Step 1.■

The next theorem gives a consistent estimator $\hat{Q}$ of the asymptotic covariance matrix Q. We follow closely the notations of Wei *et al.* (1989).[23]

Theorem 3

Under the model assumptions A.1-A.7, the asymptotic covariance matrix $Q = \left(D_{kl}, k,l \in \{1,2\}\right)$ of $n^{1/2}\left(\hat{\beta}_1' - \beta_{10}', \hat{\beta}_2' - \beta_{20}'\right)'$ is consistently estimated by $\hat{Q} = \left(\hat{D}_{kl}, k,l \in \{1,2\}\right)$, with

$$\hat{D}_{kl} = n^2 I_k^{-1}(\hat{\beta}_k)\hat{B}_{kl}I_l^{-1}(\hat{\beta}_l),$$

where $\hat{B}_{kl} = n^{-1}\sum_{i=1}^{n} \hat{w}_{ki}(\hat{\beta}_k)\hat{w}_{li}(\hat{\beta}_l)'$ and $\hat{w}_{ki}(\hat{\beta}_k)$ is a p-dimensional vector,

$$\hat{w}_{ki}(\beta_k) = \int_0^{\tau_k} \left(Z_{ki}(u) - E_k(u,\beta_k)\right) d\tilde{M}_{ki}(u,\beta_k),$$

$$\tilde{M}_{ki}(u,\beta_k) = N_{ki}(u) - \int_0^u Y_{ki}(v)\exp(\beta_k' Z_{ki}(v)) \frac{J_k(v)}{S_k^{(0)}(v,\beta_k)} dN_k(v). \square$$

Wei *et al.* (1989)[23] paper does not provide a proof of this result. Although they mention that the proof essentially uses the same techniques as in the proofs of Theorem 1 of Wei and Lachin (1984)[32] and of Theorem 3.2 and Corollary 3.3 of Andersen and Gill (1982)[31], these references are not in the context of our model. In the parametric section of this chapter we need a similar theorem and we provide all the details of its proof.

In the following we present several results related to the large sample properties of the integrated baseline hazards estimators. As mentioned in Section 1.1.1, the estimator of the integrated baseline hazard $A_{k0}(u) = \int_0^u \alpha_{k0}(v)dv$ is given by

$$\hat{A}_{k0}(u,\hat{\beta}_k) = \int_0^u \frac{J_k(u)}{S_k^{(0)}(v,\hat{\beta}_k)} dN_k(v),$$

where $J_k(v) = [Y_k(v) > 0]$.

As shown in the proof of Theorem VII.2.3, p504, Andersen *et al.* (1993)[29], $n^{1/2}\left(\hat{A}_{k0}(u,\hat{\beta}_k) - A_{k0}(u)\right)$ can be expanded as

$$\hat{W}_k(u) - n^{1/2}(\hat{\beta}_k - \beta_{k0})' \int_0^u e_k(v,\beta_{k0})\alpha_{k0}(v)dv + o_p(1),$$

where $\hat{W}_k(u) = n^{1/2} \int_0^u \frac{J_k(v)}{S_k^{(0)}(v,\beta_{k0})} dM_k(v)$ and

$$M_k(u) = \sum_{i=1}^n M_{ki}(u) = N_k(u) - \int_0^u S_k^{(0)}(v,\beta_{k0})\alpha_{k0}(v)dv.$$

Therefore

$$n^{1/2}\left(\hat{A}_{k0}(u,\hat{\beta}_k)-A_{k0}(u)\right)+n^{1/2}(\hat{\beta}_k-\beta_{k0})'\int_0^u e_k(v,\beta_{k0})\alpha_{k0}(v)dv \tag{1.5}$$

has the same asymptotic distribution as $\hat{W}_k(u)$.

The following theorem describes the asymptotic distribution of $\left(\hat{W}_1(t),\hat{W}_2(c)\right)'$ for fixed t, c.

Theorem 4

Under the model assumptions A.1-A.7, $\left(\hat{W}_1(t),\hat{W}_2(c)\right)'$ converges in distribution to a zero mean bivariate normal random vector with covariance matrix $Q^* = \left(D^*_{kl}(u_k,u_l), k,l\in\{1,2\}\right)$, where $D^*_{kl}(u_k,u_l)=E\left(w^*_{k1}(u_k,\beta_{k0})w^*_{l1}(u_l,\beta_{l0})\right)$, u_k is equal to t or c according as $k=1$ or 2 and

$$w^*_{ki}(u,\beta_{k0})=\int_0^u \frac{dM_{ki}(v)}{s_k^{(0)}(v,\beta_{k0})}.\square$$

Proof of Theorem 4:

Let u denoting t or c, depending on the index $k\in\{1,2\}$.

We can write $\hat{W}_k(u)=n^{-1/2}\int_0^u J_k(v)\dfrac{n}{S_k^{(0)}(v,\beta_{k0})}dM_k(v)=$

$$=n^{-1/2}\int_0^u J_k(v)\left[\frac{n}{S_k^{(0)}(v,\beta_{k0})}-\frac{1}{s_k^{(0)}(v,\beta_{k0})}\right]dM_k(v)-$$

$$-n^{1/2}\int_0^u(1-J_k(v))\frac{dM_k(v)}{s_k^{(0)}(v,\beta_{k0})}+n^{-1/2}\int_0^u\frac{dM_k(v)}{s_k^{(0)}(v,\beta_{k0})}.$$

We will show

$$n^{-1/2}\int_0^u J_k(v)\left[\frac{n}{S_k^{(0)}(v,\beta_{k0})}-\frac{1}{s_k^{(0)}(v,\beta_{k0})}\right]dM_k(v)\xrightarrow{P}0, \tag{1.6}$$

$$n^{1/2}\int_0^u (1-J_k(v))\frac{dM_k(v)}{s_k^{(0)}(v,\beta_{k0})}\xrightarrow{P}0. \tag{1.7}$$

For this we use the Lenglart's Inequality for local square integrable martingales and the following proposition:

Lenglart's Inequality (see p86, Andersen *et al.* (1993)[29])

For every $\eta>0,\delta>0$

$$P(\sup_{u\in[0,\tau]}|M(u)|>\eta)\le\frac{\delta}{\eta^2}+P(\langle M\rangle(\tau)>\delta),$$

where $\langle M\rangle$ is the compensator of the martingale M. □

Proposition (see Proposition II.4.1, p78, Andersen *et al.* (1993)[29])

Let the counting process N have intensity process λ, let $M=N-\int\lambda$ and let H be locally bounded and predictable. Then M and $\int HdM$ are local square integrable martingales with

$$\langle M\rangle=\int\lambda,$$

$$\left\langle\int HdM\right\rangle=\int H^2\lambda. \ \square$$

In our case $N_k(u)=M_k(u)-\int_0^u S_k^{(0)}(v,\beta_{k0})\alpha_{k0}(v)dv$ and the quantities in the integrals of (1.6) and (1.7) are bounded and predictable by our model assumptions. Consequently both expressions in (1.6) and (1.7) are local square integrable martingales, of the type $\int HdM$, so we can calculate their compensators and apply the Lenglart's Inequality.

Let $\eta > 0, \delta > 0$.

For (1.6):

$$P(\sup_{u\in[0,\tau_k]}\left|\int_0^u n^{-1/2}J_k(v)\left(\frac{n}{S_k^{(0)}(v,\beta_{k0})}-\frac{1}{s_k^{(0)}(u,\beta_{k0})}\right)dM_k(v)\right| > \eta) \le$$

$$\le \frac{\delta}{\eta^2} + P(\int_0^{\tau_k} J_k(v)\left(\frac{n}{S_k^{(0)}(v,\beta_{k0})}-\frac{1}{s_k^{(0)}(u,\beta_{k0})}\right)^2 n^{-1}S_k^{(0)}(u,\beta_{k0})\alpha_{k0}(v)dv > \delta).$$

By conditions C.a , C.b, C.c, C.f and Dominated Convergence Theorem,

$$\int_0^{\tau_k} J_k(v)\left(\frac{n}{S_k^{(0)}(v,\beta_{k0})}-\frac{1}{s_k^{(0)}(u,\beta_{k0})}\right)^2 n^{-1}S_k^{(0)}(u,\beta_{k0})\alpha_{k0}(v)dv \xrightarrow{P} 0,$$

so (1.6) is proved.

For (1.7):

$$P(\sup_{u\in[0,\tau_k]}\left|\int_0^u n^{-1/2}(1-J_k(v))\frac{dM_k(v)}{s_k^{(0)}(v,\beta_{k0})}\right| > \eta) \le$$

$$\le \frac{\delta}{\eta^2} + P(\int_0^{\tau_k}(1-J_k(v))\frac{1}{s_k^{(0)}(v,\beta_{k0})^2}n^{-1}S_k^{(0)}(v,\beta_{k0})\alpha_{k0}(v)dv > \delta).$$

But $1-J_k(v)=[Y_k(u)=0]=[Y_{ki}(u)=0\forall i]=\left[S_k^{(0)}(u,\beta_{k0})=0\right]$, so the integral in the previous relation is zero. Consequently (1.7) follows.

Relations (1.6) and (1.7) imply that $\hat{W}_k(u)$ is asymptotically equivalent to

$$n^{-1/2}\int_0^u \frac{dM_k(v)}{s_k^{(0)}(v,\beta_{k0})} = n^{-1/2}\sum_{i=1}^n w_{ki}^*(u,\beta_{k0}), \tag{1.8}$$

a sum of n i.i.d. random variables $w_{ki}^{*}(u,\beta_{k0}) = \int_0^u \frac{dM_{ki}(v)}{s_k^{(0)}(v,\beta_{k0})}$.

The quantity $s_k^{(0)}(.,\beta_{k0})$ is predictable, bounded away from 0, so, by the stated Proposition, $w_{ki}^{*}(u,\beta_{k0})$ are zero-mean martingales. Because $w_{ki}^{*}(u,\beta_{k0})$ are also i.i.d. random variables, by the Multivariate Central Limit Theorem, $\left(\hat{W}_1(t),\hat{W}_2(c)\right)'$ converges in distribution to a zero mean bivariate normal random vector with covariance matrix $Q^{*} = \left(D_{kl}^{*}(u_k,u_l), k,l \in \{1,2\}\right)$, where $D_{kl}^{*}(u_k,u_l) = E\left(w_{k1}^{*}(u_k,\beta_{k0})w_{l1}^{*}(u_l,\beta_{l0})\right)$ and u_k is equal to t or c according as $k=1$ or 2. Therefore Theorem 4 is proved. ■

The following theorem gives a consistent estimator $\hat{Q}^{*}$ of the asymptotic covariance matrix Q^{*}. As previously mentioned, the techniques of the proof for this type of theorem will be provided in a similar theorem in the parametric section of this chapter.

Theorem 5

Under the model assumptions A.1-A.7, the asymptotic covariance matrix $Q^{*} = \left(D_{kl}^{*}(u_k,u_l), k,l \in \{1,2\}\right)$ of $\left(\hat{W}_1(t),\hat{W}_2(c)\right)'$ is consistently estimated by $\hat{Q}^{*} = \left(\hat{D}_{kl}^{*}(u_k,u_l), k,l \in \{1,2\}\right)$, with

$$\hat{D}_{kl}^{*}(u_k,u_l) = n^{-1}\sum_{i=1}^{n} \hat{w}_{k1}^{*}(u_k,\hat{\beta}_k)\hat{w}_{l1}^{*}(u_l,\hat{\beta}_l) \text{ and}$$

$$\hat{w}_{k1}^{*}(u,\beta_k) = \int_0^u J_k(v)\frac{n}{S_k^{(0)}(v,\beta_k)}d\tilde{M}_{ki}(v,\beta_k),$$

$$\tilde{M}_{ki}(u,\beta_k) = N_{ki}(u) - \int_0^u Y_{ki}(v)\exp(\beta_k' Z_{ki}(v))\frac{J_k(v)}{S_k^{(0)}(v,\beta_k)}dN_k(v).\square$$

1.1.3 Large Sample Properties of the Estimators of Survival Functions

The survival distribution of LOS or cost at a fixed covariate profile Z_0 $S_k(u\,|\,Z_0) = \exp(-A_{k0}(u)e^{\beta_k' Z_0})$ is estimated by $\hat{S}_k(u\,|\,Z_0) = \exp(-\hat{A}_{k0}(u,\hat{\beta}_k)e^{\hat{\beta}_k' Z_0})$. We want to determine the asymptotic joint distribution of $\hat{S}_1(t\,|\,Z_0)$ and $\hat{S}_2(c\,|\,Z_0)$ for fixed t and c. Because by (1.5) $n^{1/2}\left(\hat{A}_{k0}(u,\hat{\beta}_k) - A_{k0}(u)\right) + n^{1/2}(\hat{\beta}_k - \beta_{k0})' \int_0^u e_k(v,\beta_{k0})\alpha_{k0}(v)dv$ is asymptotically equivalent to $\hat{W}_k(u)$, a first step would be to consider the joint distribution of the vectors $\left(\hat{W}_1(t), \hat{W}_2(c)\right)'$ and $n^{1/2}\left(\hat{\beta}_1' - \beta_{10}', \hat{\beta}_2' - \beta_{20}'\right)'$.

Theorem 6

Under the model assumptions A.1-A.7,

$$(\hat{W}_1(t),\, \hat{W}_2(c),\, n^{1/2}(\hat{\beta}_1 - \beta_{10})',\, n^{1/2}(\hat{\beta}_2 - \beta_{20})')'$$

converges in distribution to a zero-mean normal random vector with covariance matrix $\tilde{Q} = \begin{pmatrix} Q^* & P' \\ P & Q \end{pmatrix}$. The matrices Q and Q^* and their consistent estimators are described in Theorems 2-5. The matrix P has the form

$$P = \begin{pmatrix} \underline{0} & P_{21}(c) \\ P_{12}(t) & \underline{0} \end{pmatrix},$$

where $P_{kl}(u_k) = \Sigma_l^{-1} E\left(w_{k1}^*(u_k, \beta_{k0}) w_{l1}(\beta_{l0}) \right)$ for $k \neq l$ is consistently estimated by

$\hat{P}_{kl}(u_k) = I_l^{-1}(\hat{\beta}_l) \cdot \sum_{i=1}^{n} \hat{w}_{ki}^*(u_k, \hat{\beta}_k) \hat{w}_{li}(\hat{\beta}_l)$. □

(See Theorems 2-5 for notations.)

Proof Theorem 6:

As shown in the steps of the proof of Theorem 2,

$$n^{1/2}(\hat{\beta}_k - \beta_{k0}) = nI_k^{-1}(\beta_k^*)(n^{-1/2} \sum_{i=1}^{n} w_{ki}(\beta_{k0})) + o_p(1),$$

where $nI_k^{-1}(\beta_k^*) \xrightarrow{P} \Sigma_k^{-1}$ and $w_{ki}(\beta_{k0}) = \int_0^{\tau_k} \left(Z_{ki}(u) - e_k(u, \beta_{k0}) \right) dM_{ki}(u), 1 \leq i \leq n$ are i.i.d. zero mean random vectors. By (1.8)

$$\hat{W}_k(u) = n^{-1/2} \sum_{i=1}^{n} w_{ki}^*(u, \beta_{k0}),$$

where $w_{ki}^*(u, \beta_{k0}) = \int_0^u \frac{dM_{ki}(v)}{s_k^{(0)}(v, \beta_{k0})}$ are i.i.d. zero mean random variables. Consequently

$(\hat{W}_1(t), \hat{W}_2(c), n^{1/2}(\hat{\beta}_1 - \beta_{10})', n^{1/2}(\hat{\beta}_2 - \beta_{20})')'$ is asymptotically equivalent to the

($2p+2$)-dimensional vector

$$diag\left(1, 1, nI_1^{-1}(\beta_1^*), nI_2^{-1}(\beta_2^*)\right) \times n^{-1/2} \sum_{i=1}^{n} p_i(t, c, \beta_{10}, \beta_{20}),$$

where the i.i.d. vectors p_i have the components $w_{1i}^*(t, \beta_1), w_{2i}^*(c, \beta_2), w_{1i}(\beta_{10}), w_{2i}(\beta_{20})$,

the first two being scalars and the last two p-dimensional vectors.

It follows from the Multivariate Central Limit Theorem and Slutsky's Lemma that

$(\hat{W}_1(t), \hat{W}_2(c), n^{1/2}(\hat{\beta}_1 - \beta_{10})', n^{1/2}(\hat{\beta}_2 - \beta_{20})')'$ converges in distribution to a zero-mean normal random vector with covariance matrix $\tilde{Q} = \begin{pmatrix} Q^* & P' \\ P & Q \end{pmatrix}$, where the matrices

Q and Q^* and their consistent estimators are described in Theorems 2-5. Write

$$P = \begin{pmatrix} P_{11}(t) & P_{21}(c) \\ P_{12}(t) & P_{22}(c) \end{pmatrix},$$

where $P_{kl}(u_k)$ is the asymptotic covariance between $\hat{W}_k(u_k)$ and $n^{1/2}(\hat{\beta}_l - \beta_{l0})$,

$k, l \in \{1,2\}$.

For $k \neq l$, $P_{kl}(u_k)$ has the form $P_{kl}(u_k) = \Sigma_l^{-1} E\left(w_{k1}^*(u_k, \beta_{k0}) w_{l1}(\beta_{l0})\right)$ and is

consistently estimated by $\hat{P}_{kl}(u_k) = I_l^{-1}(\hat{\beta}_l) \sum_{i=1}^n \hat{w}_{ki}^*(u_k, \hat{\beta}_k) \hat{w}_{li}(\hat{\beta}_l)$,

where $\hat{w}_{li}(\hat{\beta}_l)$ and $\hat{w}_{ki}^*(u_k, \hat{\beta}_k)$ are described in Theorems 3 and 5, respectively.

We will show that $P_{kk}(u_k)$ has only zero components, so that

$$P = \begin{pmatrix} \underline{0} & P_{21}(c) \\ P_{12}(t) & \underline{0} \end{pmatrix}.$$

For $k = 1$,

$$\begin{aligned} P_{11}(t) &= \Sigma_1^{-1} E\left(w_{11}^*(t, \beta_{10}) w_{11}(\beta_{10})\right) \\ &= \Sigma_1^{-1} E\left(\int_0^t \frac{dM_{k1}(u)}{s_1^{(0)}(u, \beta_{10})} \int_0^{\tau_1} (Z_{11}(u) - e_1(u, \beta_{10})) dM_{11}(u) \right). \end{aligned}$$

Both factors of the product in the expectation are local square integrable martingales. By the definition of the predictable covariation process $\langle M, M' \rangle$ of two local square integrable martingales M and M' (see p68, Andersen *et al.* (1993)[29]), $EMM' = E\langle M, M' \rangle$. If H, K are bounded predictable processes and M is a counting process martingale with $\langle M \rangle = \int \lambda$ then $\left\langle \int HdM, \int KdM \right\rangle = \int HK\lambda$ (see p78, Andersen *et al.* (1993)[29]). Therefore, using these results, for $j \in \{1,\dots,p\}$ we obtain that

$$E\left(\int_0^t \frac{dM_{k1}(u)}{s_1^{(0)}(u,\beta_{10})} \int_0^{\tau_1} (Z_{11}(u) - e_1(u,\beta_{10}))dM_{11}(u) \right)_j =$$

$$= E\left(\int_0^t \frac{dM_{k1}(u)}{s_1^{(0)}(u,\beta_{10})} \int_0^{\tau_1} \left(Z_{11j}(u) - e_{1j}(u,\beta_{10})\right) dM_{11}(u) \right)$$

$$= E\left\langle \int_0^{\tau_1} I_{[0,t]}(u) \frac{dM_{k1}(u)}{s_1^{(0)}(u,\beta_{10})}, \int_0^{\tau_1} \left(Z_{11j}(u) - e_{1j}(u,\beta_{10})\right) dM_{11}(u) \right\rangle$$

$$= E\left(\int_0^{\tau_1} I_{[0,t]}(u) \frac{1}{s_1^{(0)}(u,\beta_{10})} \left(Z_{11j}(u) - e_{1j}(u,\beta_{10})\right) \alpha_{10}(u) \exp(\beta_{10}' Z_{11}(u)) Y_{11}(u) du \right)$$

By Fubini's Theorem the last quantity is equal to

$$\int_0^t \frac{\alpha_{10}(u)}{s_1^{(0)}(u,\beta_{10})} \left(s_1^{(1)}(u,\beta_{10})_j - \frac{s_1^{(1)}(u,\beta_{10})_j}{s_1^{(0)}(u,\beta_{10})} s_1^{(0)}(u,\beta_{10}) \right) du = 0.$$

Hence we proved $P_{11}(t) = \underline{0}$. Similarly we can show $P_{22}(c) = \underline{0}$. ■

Now we have the tools to prove the following theorem about the asymptotic joint distribution of $\hat{S}_1(t \mid Z_0)$ and $\hat{S}_2(c \mid Z_0)$, for fixed time t and cost c.

Theorem 7

Under the model assumptions A.1-A.7,

$$n^{1/2}\left(\hat{S}_1(t \mid Z_0) - S_1(t \mid Z_0), \hat{S}_2(c \mid Z_0) - S_2(c \mid Z_0)\right)'$$

converges in distribution to a zero mean bivariate normal random vector with covariance matrix $CS = \left(CS_{kl}(u_k, u_l), k, l \in \{1,2\}\right)$, where

$$CS_{kk}(u_k) = S_k(u_k \mid Z_0)^2 \exp(2\beta_{k0}' Z_0) \times$$

$$\times\{D_{kk}^*(u_k) + \left(\int_0^{u}(Z_0 - e_k(v, \beta_{k0}))\alpha_{k0}(v)dv\right)' D_{kk}\left(\int_0^{u}(Z_0 - e_k(v, \beta_{k0}))\alpha_{k0}(v)dv\right)\},$$

$$CS_{12}(t,c) = S_1(t \mid Z_0) S_2(c \mid Z_0) \exp(\beta_{10}' Z_0 + \beta_{20}' Z_0) \times$$

$$\times\{D_{12}^*(t,c) + \left(\int_0^{t}(Z_0 - e_1(v, \beta_{10}))\alpha_{10}(v)dv\right)' P_{21}(c) +$$

$$+\left(\int_0^{c}(Z_0 - e_2(v, \beta_{20}))\alpha_{20}(v)dv\right)' P_{12}(t) +$$

$$+\left(\int_0^{t}(Z_0 - e_1(v, \beta_{10}))\alpha_{10}(v)dv\right)' D_{12}\left(\int_0^{c}(Z_0 - e_2(v, \beta_{20}))\alpha_{20}(v)dv\right)\}.$$

□ (See Theorem 2-6 for notations.)

Proof Theorem 7:

We first state the Delta Method, a popular and elementary tool of asymptotic statistics, which we will apply repeatedly in the proof of this theorem. We will use the following version, stated at p109, Andersen *et al.* (1993)[29]:

Delta Method

Suppose for some random p-vectors T_n and a sequence of numbers $a_n \to \infty$,

$$a_n(T_n - \theta) \xrightarrow{D} Z \text{ as } n \to \infty,$$

where $\theta \in \mathbb{R}^p$ is fixed. Suppose $\phi : \mathbb{R}^p \to \mathbb{R}^q$ is differentiable at θ with $q \times p$ matrix $\phi'(\theta)$ of partial derivatives. Then

$$a_n\left(\phi(T_n)-\phi(\theta)\right)\xrightarrow{D}\phi'(\theta)\times Z \text{ in } \mathbb{R}^q$$

and, indeed, $a_n\left(\phi(T_n)-\phi(\theta)\right)$ is asymptotically equivalent to $\phi'(\theta)\times a_n\left(T_n-\theta\right)$. □

Recall that for $k\in\{1,2\}$ the survival function at a fixed covariate profile Z_0 is $S_k(u\mid Z_0)=\exp(-A_k(u,\beta_{k0}\mid Z_0))$, where $A_k(u,\beta_{k0}\mid Z_0)=A_{k0}(u)\exp(\beta_{k0}'Z_0)$ is the integrated hazard. The survival function is estimated by

$\hat S_k(u\mid Z_0)=\exp(-\hat A_k(u,\hat\beta_k\mid Z_0))$, where $\hat A_k(u,\hat\beta_k\mid Z_0)=\hat A_{k0}(u,\hat\beta_k)\exp(\hat\beta_k'Z_0)$.

By the Delta Method $n^{1/2}\left(\hat S_1(t\mid Z_0)-S_1(t\mid Z_0),\hat S_2(c\mid Z_0)-S_2(c\mid Z_0)\right)'$ has the same limiting distribution as

$$-\begin{pmatrix} S_1(t\mid Z_0) & 0\\ 0 & S_2(c\mid Z_0)\end{pmatrix} n^{1/2}\begin{pmatrix}\hat A_1(t,\hat\beta_1\mid Z_0)-A_1(t,\beta_{10}\mid Z_0)\\ \hat A_2(c,\hat\beta_2\mid Z_0)-A_2(c,\beta_{20}\mid Z_0)\end{pmatrix}. \tag{1.9}$$

Therefore we will determine first the asymptotic distribution of

$$n^{1/2}\begin{pmatrix}\hat A_1(t,\hat\beta_1\mid Z_0)-A_1(t,\beta_{10}\mid Z_0)\\ \hat A_2(c,\hat\beta_2\mid Z_0)-A_2(c,\beta_{20}\mid Z_0)\end{pmatrix}=n^{1/2}\begin{pmatrix}\exp(\hat\beta_1'Z_0)\hat A_{10}(t,\hat\beta_1)-\exp(\beta_{10}'Z_0)A_{10}(t)\\ \exp(\hat\beta_2'Z_0)\hat A_{20}(c,\hat\beta_2)-\exp(\beta_{20}'Z_0)A_{20}(c)\end{pmatrix}.$$

We apply again the Delta Method. Let $T_n=\left(\hat A_{10}(t,\hat\beta_1),\hat A_{20}(c,\hat\beta_2),\hat\beta_1',\hat\beta_2'\right)'$, $\theta=\left(A_{10}(t),A_{20}(c),\beta_{10}',\beta_{20}'\right)'$ and $a_n=n^{1/2}$. A linear combination of the elements of the vector $a_n\left(T_n-\theta\right)$ has the same limiting distribution as a linear combination of the elements of $(\hat W_1(t),\ \hat W_2(c),\ n^{1/2}(\hat\beta_1-\beta_{10})',\ n^{1/2}(\hat\beta_2-\beta_{20})')'$. Consequently, by Theorem 6, $a_n\left(T_n-\theta\right)$ converges weakly to a multivariate normal vector.

We define the mapping $\phi:\mathbb{R}\times\mathbb{R}\times\mathbb{R}^p\times\mathbb{R}^p\to\mathbb{R}^2$,

$$\phi(a,b,c,d) = \begin{pmatrix} a\exp(c'Z_0) \\ b\exp(d'Z_0) \end{pmatrix}.$$

The function ϕ is everywhere differentiable, with the $2\times(2p+2)$ matrix $\phi'(a,b,c,d)$ of partial derivatives

$$\phi'(a,b,c,d) = \begin{pmatrix} \exp(c'Z_0) & 0 & a\exp(c'Z_0)Z_0' & \underline{0} \\ 0 & \exp(d'Z_0) & \underline{0} & b\exp(d'Z_0)Z_0' \end{pmatrix}.$$

By the stated Delta Method,

$$n^{1/2}\begin{pmatrix} \hat{A}_1(t,\hat{\beta}_1 \mid Z_0) - A_1(t,\beta_{10} \mid Z_0) \\ \hat{A}_2(c,\hat{\beta}_2 \mid Z_0) - A_2(c,\beta_{20} \mid Z_0) \end{pmatrix} = a_n\left(\phi(T_n) - \phi(\theta)\right)$$

is asymptotically equivalent to $\phi'(\theta)\times a_n\left(T_n - \theta\right) =$

$$= n^{1/2}\begin{pmatrix} \exp(\beta_{10}' Z_0)\left(\hat{A}_{10}(t,\hat{\beta}_1) - A_{10}(t)\right) + \exp(\beta_{10}' Z_0)A_{10}(t)Z_0'(\hat{\beta}_1 - \beta_{10}) \\ \exp(\beta_{20}' Z_0)\left(\hat{A}_{20}(c,\hat{\beta}_2) - A_{20}(c)\right) + \exp(\beta_{20}' Z_0)A_{20}(c)Z_0'(\hat{\beta}_2 - \beta_{20}) \end{pmatrix}.$$

By (1.5) this quantity has the same limiting distribution as

$$\begin{pmatrix} \exp(\beta_{10}' Z_0) & 0 & \exp(\beta_{10}' Z_0)F_1(t,\beta_{10})' & \underline{0} \\ 0 & \exp(\beta_{20}' Z_0) & \underline{0} & \exp(\beta_{20}' Z_0)F_2(c,\beta_{20})' \end{pmatrix}\times$$

$\times(\hat{W}_1(t), \hat{W}_2(c), n^{1/2}(\hat{\beta}_1 - \beta_{10})', n^{1/2}(\hat{\beta}_2 - \beta_{20})')'$, where

$F_k(u,\beta_{k0}) = \int_0^u \left(Z_0 - e_k(v,\beta_{k0})\right)\alpha_{k0}(v)dv$, $k \in \{1,2\}$. Therefore, by Theorem 6,

$n^{1/2}\begin{pmatrix} \hat{A}_1(t,\hat{\beta}_1 \mid Z_0) - A_1(t,\beta_{10} \mid Z_0) \\ \hat{A}_2(c,\hat{\beta}_2 \mid Z_0) - A_2(c,\beta_{20} \mid Z_0) \end{pmatrix}$ converges in distribution to a zero mean normal random vector, with covariance matrix denoted $C = \left(C_{kl}(u_k,u_l), k,l \in \{1,2\}\right)$. By the asymptotic independence of $\hat{W}_k(u_k)$ and $n^{1/2}(\hat{\beta}_k - \beta_{k0})$,

$$C_{kk}(u_k) = \exp(2\beta'_{k0}Z_0)\times$$

$$\times\{D^*_{kk}(u_k) + \left(\int_0^{u}(Z_0 - e_k(v,\beta_{k0}))\alpha_{k0}(v)dv\right)' D_{kk}\left(\int_0^{u}(Z_0 - e_k(v,\beta_{k0}))\alpha_{k0}(v)dv\right)\},$$

$$C_{12}(t,c) = \exp(\beta'_{10}Z_0 + \beta'_{20}Z_0)\times$$

$$\times\{D^*_{12}(t,c) + \left(\int_0^{u}(Z_0 - e_1(v,\beta_{10}))\alpha_{10}(v)dv\right)' P_{21}(c) +$$

$$+\left(\int_0^{c}(Z_0 - e_2(v,\beta_{20}))\alpha_{20}(v)dv\right)' P_{12}(t) +$$

$$+\left(\int_0^{t}(Z_0 - e_1(v,\beta_{10}))\alpha_{10}(v)dv\right)' D_{12}\left(\int_0^{c}(Z_0 - e_2(v,\beta_{20}))\alpha_{20}(v)dv\right)\}.$$

By (1.9), $n^{1/2}\left(\hat{S}_1(t\,|\,Z_0) - S_1(t\,|\,Z_0), \hat{S}_2(c\,|\,Z_0) - S_2(c\,|\,Z_0)\right)'$ converges in distribution to a zero mean bivariate normal random vector with covariance matrix $CS = \left(CS_{kl}(u_k,u_l), k,l \in \{1,2\}\right)$, where

$$CS_{kk}(u_k) = S_k(u_k\,|\,Z_0)^2 C_{kk}(u_k),$$

$$CS_{12}(t,c) = S_1(t\,|\,Z_0)S_2(c\,|\,Z_0)C_{12}(t,c). \blacksquare$$

A consistent estimator $\hat{CS} = \left(\hat{CS}_{kl}(u_k,u_l), k,l \in \{1,2\}\right)$ of the asymptotic covariance matrix $CS = \left(CS_{kl}(u_k,u_l), k,l \in \{1,2\}\right)$ of $n^{1/2}\left(\hat{S}_1(t\,|\,Z_0) - S_1(t\,|\,Z_0), \hat{S}_2(c\,|\,Z_0) - S_2(c\,|\,Z_0)\right)'$ can be obtained replacing β_{k0} by $\hat{\beta}_k$, $S_k(u_k\,|\,Z_0)$ by the estimator $\hat{S}_k(u_k\,|\,Z_0)$, $D_{kl}, D^*_{kl}(u_k,u_l), P_{21}(c), P_{12}(t)$ by their consistent estimates and $\int_0^{u}(Z_0 - e_k(v,\beta_{k0}))\alpha_{k0}(v)dv$ by the quantity

$$\int_0^{u}\left(Z_0 - E_k(v,\hat{\beta}_k)\right)\frac{J_k(v)}{S_k^{(0)}(v,\hat{\beta}_k)}dN_k(v).$$

1.1.4 Point Estimates and Confidence Intervals for Median LOS and Median Cost

The median LOS or cost for a fixed covariate profile Z_0 is defined as

$$m_k(Z_0) = \inf\{u : S_k(u \mid Z_0) \leq .5\}$$

and estimated by

$$\hat{m}_k(Z_0) = \inf\{u : \hat{S}_k(u \mid Z_0) \leq .5\}.$$

For constructing a confidence interval for the median we will follow the approach described in Andersen *et al.* (1993)[29], p511-512. The advantage of this procedure is that the estimation of the density of S_k is not needed. The confidence interval for the median $m_k(Z_0)$ can be read directly from the lower and upper pointwise confidence limits for the survival distribution in exactly the same manner as $\hat{m}_k(Z_0)$ can be read from the curve $\hat{S}_k(. \mid Z_0)$ itself.

Let us first consider pointwise confidence intervals for $S_k(u \mid Z_0)$. By Theorem 6, $n^{1/2}\left(\hat{S}_k(u \mid Z_0) - S_k(u \mid Z_0)\right)$ converges weakly to a zero mean normal distribution with variance $CS_{kk}(u)$. The "standard" asymptotic $100(1-\alpha)\%$ interval

$$\left[\hat{S}_k(u \mid Z_0) - z_{\alpha/2}\left(\hat{CS}_{kk}(u)/n\right)^{1/2}, \hat{S}_k(u \mid Z_0) + z_{\alpha/2}\left(\hat{CS}_{kk}(u)/n\right)^{1/2}\right],$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ quantile of the standard normal distribution, might not be completely satisfactory for small sample size.[33] Kalbfleisch and Prentice (1980)[34] and Thomas and Grunkemeier (1975)[33] suggested that using transformations such as

$g(x) = \log(-\log x), x \in (0,1)$ might improve the small sample properties of the confidence intervals. Other transformations for constructing confidence intervals are described in Klein and Moeschberger (1997).[35]

If $CS_{kk}(u) \neq 0$ and g is a real function differentiable in the neighborhood of $S_k(u \mid Z_0)$, with continuous derivative g' different from zero at $S_k(u \mid Z_0)$, then by the Delta Method and Slutsky's Lemma,

$$\frac{g\left(\hat{S}_k(u \mid Z_0)\right) - g\left(S_k(u \mid Z_0)\right)}{\left|g'\left(\hat{S}_k(u \mid Z_0)\right)\right| \times \left(\hat{CS}_{kk}(u)/n\right)^{1/2}} \xrightarrow{D} N(0,1).$$

Asymptotic $100(1-\alpha)\%$ confidence limits for $g\left(S_k(u \mid Z_0)\right)$ are

$$g\left(\hat{S}_k(u \mid Z_0)\right) \pm z_{\alpha/2}\left|g'\left(\hat{S}_k(u \mid Z_0)\right)\right| \times \left(\hat{CS}_{kk}(u)/n\right)^{1/2}. \tag{1.10}$$

For the transformation $g(x) = \log(-\log x), x \in (0,1)$ the derivative g' exists and is continuous everywhere on $(0,1)$, $g'(x) = \left(x \log x\right)^{-1}$. Also the inverse of this function has the form $g^{-1}(y) = \exp(-\exp y), y \in \mathbb{R}$. If $S_k(u \mid Z_0) \in (0,1)$ then, by retransforming (1.10), we obtain the following $100(1-\alpha)\%$ confidence interval for $S_k(u \mid Z_0)$:

$\left[\hat{S}_{k1}(u \mid Z_0), \hat{S}_{k2}(u \mid Z_0)\right]$, where

$$\hat{S}_{k1}(u \mid Z_0) = \hat{S}_k(u \mid Z_0)^{\exp(-PS_k(u))},$$

$$\hat{S}_{k2}(u \mid Z_0) = \hat{S}_k(u \mid Z_0)^{\exp(PS_k(u))}, \tag{1.11}$$

$$PS_k(u) = z_{\alpha/2}\frac{\left(\hat{CS}_{kk}(u)/n\right)^{1/2}}{\hat{S}_k(u \mid Z_0)\log \hat{S}_k(u \mid Z_0)}.$$

We now turn to the construction of a confidence interval for the median $m_k(Z_0)$. By the procedure described in Andersen *et al.* (1993)[29], p277, we can take as an approximate $100(1-\alpha)\%$ confidence interval for $m_k(Z_0)$ all values u which satisfy

$$\frac{\left|g\left(\hat{S}_k(u \mid Z_0)\right)-g(.5)\right|}{\left|g'\left(\hat{S}_k(u \mid Z_0)\right)\right| \times \left(\hat{CS}_{kk}(u)/n\right)^{1/2}} \leq z_{\alpha/2},$$

where $g(x)=\log(-\log x), x \in (0,1)$, i.e. all hypothesized values u of $m_k(Z_0)$ which are not rejected when testing $H_0: m_k(Z_0)=u$ against $H_1: m_k(Z_0) \neq u$ at level α, based on the asymptotic normality of $g\left(\hat{S}_k(u \mid Z_0)\right)$. Therefore the approximate confidence interval for $m_k(Z_0)$ is

$$\left\{u: g(.5) \text{ is between } g\left(\hat{S}_k(u \mid Z_0)\right) \pm z_{\alpha/2}\left|g'\left(\hat{S}_k(u \mid Z_0)\right)\right| \times \left(\hat{CS}_{kk}(u)/n\right)^{1/2}\right\} =$$

$=\left\{u: 0.5 \text{ is between } \hat{S}_{k1}(u \mid Z_0) \text{ and } \hat{S}_{k2}(u \mid Z_0)\right\}$, where $\hat{S}_{k1}(u \mid Z_0)$ and $\hat{S}_{k2}(u \mid Z_0)$ are given in (1.11).

Define $\hat{m}_{k1}(Z_0)=\inf\left\{u: \hat{S}_{k1}(u \mid Z_0) \leq .5\right\}$, $\hat{m}_{k2}(Z_0)=\inf\left\{u: \hat{S}_{k2}(u \mid Z_0) \leq .5\right\}$. Then $[\hat{m}_{k1}(Z_0), \hat{m}_{k2}(Z_0)]$ is an approximate $100(1-\alpha)\%$ confidence interval for $m_k(Z_0)$.

This completes our construction of point estimates and confidence intervals for the median LOS and median cost in the case of semiparametric marginal models. An application of these results will be presented in Section 1.3.

1.2 Parametric Marginal Models

Using the same notations as in Section 1.1, suppose we observe for the i-th patient of a study with n individuals $X_{1i} = \min(T_i, T_i')$, $\delta_{1i} = [T_i \le T_i']$, Z_{1i}, with $T_i = g(T_i^*)$, $T_i' = g(T_i'^*)$, where T_i^* is the true LOS, $T_i'^*$ is the censoring time and Z_{1i} is vector of p time-independent explanatory variables. The monotonic transformation g (with inverse g^{-1}) is chosen to mitigate the effects of skewness that might be present in the data. For example, the log or square-root transformations are used when the data are right-skewed and have the advantage of permitting easy interpretation of the model. All variables $T_i^*, T_i'^*, Z_{1i}$ are defined on the probability space $(\Omega_1, \mathcal{F}_1, P_1)$. The transformation g is chosen such that both T_i and T_i' are strictly positive variables.

In a similar manner we consider $X_{2i} = \min(C_i, C_i')$, $\delta_{2i} = [C_i \le C_i']$ and the p-vector Z_{2i} of explanatory variables, where $C_i = g(C_i^*)$, $C_i' = g(C_i'^*)$, C_i^* is the true hospital cost, $C_i'^*$ is the censoring cost and both $C_i, C_i' > 0$. All these variables are defined on the probability space $(\Omega_2, \mathcal{F}_2, P_2)$. The transformation applied to cost need not be the same as that applied to LOS.

The relationship of explanatory variables to the time-cost observations is modeled through the following linear regression model:

$$
\begin{aligned}
T_i &= \beta_{10}' Z_{1i} + \sigma_{10}\varepsilon_{1i} \\
C_i &= \beta_{20}' Z_{2i} + \sigma_{20}\varepsilon_{2i},
\end{aligned}
\tag{1.12}
$$

where $(\varepsilon_{1i},\varepsilon_{2i})'$, $1 \le i \le n$ are i.i.d., with distribution function $F(.,.|\rho_0)$. Here ρ_0 is the true value of a nuisance parameter ρ. Let $f(.,.|\rho_0)$ and $\alpha(.,.|\rho_0)$ be the density and the hazard function associated to the distribution function $F(.,.|\rho_0)$.

As in Section 1.1, the subscript $k = 1$ is associated with time and the subscript $k = 2$ is associated with cost. We assume the $\varepsilon_{ki}, 1 \le i \le n$ are i.i.d., with zero mean and distribution function F_0 that does not depend on ρ_0. This parameter is associated with the joint distribution function of $(\varepsilon_{1i},\varepsilon_{2i})'$. Let f_0 be the density and α_0 the hazard of the distribution of ε_{ki}.

Example 1.2.1:

Suppose $\begin{pmatrix}\varepsilon_{1i}\\ \varepsilon_{2i}\end{pmatrix}$, $1 \le i \le n$ are i.i.d. $N\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}1 & \rho_0\\ \rho_0 & 1\end{pmatrix}\right)$.

Marginally ε_{ki}, $1 \le i \le n$ are i.i.d. $N(0,1)$, so $\alpha_0(u) = \dfrac{\varphi}{1-\Phi}(u)$, where φ and Φ are the hazard and distribution function of the standard normal distribution.□

Example 1.2.2:

Suppose $\begin{pmatrix}\varepsilon_{1i}\\ \varepsilon_{2i}\end{pmatrix}$, $1 \le i \le n$ are i.i.d. $F(y_1, y_2) = \dfrac{1}{1+e^{-y_1}+e^{-y_2}}$, $y_1, y_2 \in \mathbb{R}$,

so $(\varepsilon_{1i},\varepsilon_{2i})'$ has the bivariate logistic distribution, with no parameter affecting the shape of the joint distribution of ε_{1i} and ε_{2i}. Marginally ε_{ki}, $1 \le i \le n$ are i.i.d.

$F_0(u) = \alpha_0(u) = \dfrac{1}{1+e^{-u}}$, $u \in \mathbb{R}$, with mean $E(\varepsilon_{ki}) = 0$ and variance $\mathrm{var}(\varepsilon_{ki}) = \pi^2/6$.□

For the model (1.12) we denoted by $\beta_{10}, \beta_{20}, \sigma_{10}, \sigma_{20}$ the true values of the regression and scale parameters, reserving $\beta_1, \beta_2, \sigma_1, \sigma_2$ for the free parameters in the likelihood function. The vectors β_{10}, β_{20} have each dimension p and σ_{10}, σ_{20} are strictly positive scalars. Denote by θ_{k0} the (p+1)- dimensional vector $\theta'_{k0} = (\beta'_{k0}, \sigma_{k0})'$.

By the specification of the linear model (1.12), given the covariates Z_{1i}, Z_{2i}, the time-cost vector $(T_i, C_i)'$ has the distribution function $F^*(.,.,\theta_{10}, \theta_{20}, \rho_0)$ and density $f^*(.,.,\theta_{10}, \theta_{20}, \rho_0)$, where

$$\begin{aligned} F^*(t,c,\theta_{10},\theta_{20},\rho_0) &= F\left(\frac{t-\beta'_{10}Z_{1i}}{\sigma_{10}}, \frac{c-\beta'_{20}Z_{2i}}{\sigma_{20}} \mid \rho_0\right), \\ f^*(t,c,\theta_{10},\theta_{20},\rho_0) &= (\sigma_{10}\sigma_{20})^{-1} f\left(\frac{t-\beta'_{10}Z_{1i}}{\sigma_{10}}, \frac{c-\beta'_{20}Z_{2i}}{\sigma_{20}} \mid \rho_0\right). \end{aligned} \tag{1.13}$$

Marginally, given Z_{1i}, T_i has distribution function $F_{1i}(.,\theta_{10})$, density $f_{1i}(.,\theta_{10})$ and hazard function $\alpha_{1i}(.,\theta_{10})$, where

$$\begin{aligned} F_{1i}(t,\theta_{10}) &= F_0\left(\frac{t-\beta'_{10}Z_{1i}}{\sigma_{10}}\right), \\ f_{1i}(t,\theta_{10}) &= \frac{1}{\sigma_{10}} f_0\left(\frac{t-\beta'_{10}Z_{1i}}{\sigma_{10}}\right), \\ \alpha_{1i}(t,\theta_{10}) &= \frac{1}{\sigma_{10}} \alpha_0\left(\frac{t-\beta'_{10}Z_{1i}}{\sigma_{10}}\right). \end{aligned} \tag{1.14}$$

Similarly, given Z_{2i}, we define for the costs C_i the functions $F_{2i}(.,\theta_{20}), f_{2i}(.,\theta_{20})$, $\alpha_{2i}(.,\theta_{20})$.

With i.i.d. data $(X_{ki}, \delta_{ki}, Z_{ki})$ on n patients we estimate θ_k by the maximum partial likelihood estimator $\hat{\theta}_k$. We will show that $\left(\hat{\theta}_1', \hat{\theta}_2'\right)'$ has an asymptotic $2(p+1)$-variate normal distribution and we will provide a consistent estimator of the asymptotic covariance matrix. Then point estimates and confidence intervals for the median LOS and median cost will be obtained for a specified covariate profile. Detailed formulas will be provided for the bivariate normal case (see Example 1.2.1).

1.2.1 Estimation of the Model Parameters

Consider the study observation period restricted to the finite time interval $[0, \tau_1]$, $\tau_1 < \infty$. Similarly costs are assumed to be upper-bounded by the finite cost τ_2.

The processes $N_{ki}(u), Y_{ki}(u), N_k(u)$ have the same definition and interpretation as in Section 1.1.1. The argument u is again used to denote either t or c, depending on the subscript $k \in \{1,2\}$. As mentioned, u takes values in the finite interval $[0, \tau_k]$.

Define the $(p+1)\times(p+1)$ matrix

$$\Sigma_k = \begin{pmatrix} \Sigma_k^{11} & \Sigma_k^{12} \\ \left(\Sigma_k^{12}\right)' & \sigma_k^{22} \end{pmatrix},$$

where the $p \times p$ matrix Σ_k^{11} is

$$\Sigma_k^{11} = \sigma_{k0}^{-3} E\left\{ Z_{k1} Z_{k1}' \int_0^{\tau_k} \frac{(\alpha_0')^2}{\alpha_0} \left(\frac{u - \beta_{k0}' Z_{k1}}{\sigma_{k0}} \right) Y_{k1}(u) du \right\},$$

the p-dimension vector Σ_k^{12} has the form

$$\Sigma_k^{12} = \sigma_{k0}^{-3} E\left\{ Z_{k1} \int_0^{\tau_k} \alpha_0'\left(\frac{u-\beta_{k0}'Z_{k1}}{\sigma_{k0}}\right)\left[1+\left(\frac{u-\beta_{k0}'Z_{k1}}{\sigma_{k0}}\right)\frac{\alpha_0'}{\alpha_0}\left(\frac{u-\beta_{k0}'Z_{k1}}{\sigma_{k0}}\right)\right] Y_{k1}(u)du \right\}$$

and the scalar σ_k^{22} is

$$\sigma_k^{22} = \sigma_{k0}^{-3} E\left\{ \int_0^{\tau_k} \left[1+\left(\frac{u-\beta_{k0}'Z_{k1}}{\sigma_{k0}}\right)\frac{\alpha_0'}{\alpha_0}\left(\frac{u-\beta_{k0}'Z_{k1}}{\sigma_{k0}}\right)\right]^2 \alpha_0\left(\frac{u-\beta_{k0}'Z_{k1}}{\sigma_{k0}}\right) Y_{k1}(u)du \right\}.$$

The following assumptions, similar to those used in Section 1.1, will be assumed to hold throughout this section:

Model Assumptions:

B.1 Conditional on $Z_{1i}(.)$, T_i and T_i' are independent and conditional on $Z_{2i}(.)$, C_i and C_i' are independent; (Independent Censoring Assumption)

B.2 $\left\{\begin{pmatrix} X_{1i} \\ X_{2i}\end{pmatrix}, \begin{pmatrix} \delta_{1i} \\ \delta_{2i}\end{pmatrix}, \begin{pmatrix} Z_{1i}(.) \\ Z_{2i}(.)\end{pmatrix}\right\}, 1 \le i \le n$ are i.i.d.;

B.3 Z_{1i}, Z_{2i} are bounded;

B.4 The hazard function α_0 is strictly positive and its derivatives of first, second and third order exist and are continuous;

B.5 The matrices Σ_1, Σ_2 are nonsingular.

Under the independent censoring assumption, the counting process $N_{ki}(.)$ has the compensator $\Lambda_{ki}(u,\theta_{k0}) = \int_0^u \lambda_{ki}(v,\theta_{k0})dv$, where the intensity process $\lambda_{ki}(v,\theta_{k0})$ has the form $\lambda_{ki}(v,\theta_{k0}) = \alpha_{ki}(v,\theta_{k0})Y_{ki}(v)$, with the hazard function $\alpha_{ki}(.,\theta_{k0})$ defined in (1.14).

Let M_{ki} denote the local square integrable martingales $M_{ki}(u) = N_{ki}(u) - \Lambda_{ki}(u, \theta_{k0})$ and $M_k = \sum_{i=1}^{n} M_{ki}$.

Properties of stochastic processes, such as being a local martingale, are relative to a filtration $\left(\mathcal{F}_k^{(n)}(u), u \in [0, \tau_k]\right)$ of sub σ-algebras on the n-th sample space $(\Omega_k^{(n)}, \mathcal{F}_k^{(n)}, P_k^{(n)})$; $\mathcal{F}_k^{(n)}(u)$ represents everything that happens up to the point u in the n-th model. For the joint time-cost stochastic processes, we consider the filtration $\left(\mathcal{F}_1^{(n)}(t) \otimes \mathcal{F}_2^{(n)}(c), (t, c) \in [0, \tau_1] \times [0, \tau_2]\right)$ on the product space $(\Omega_1^{(n)} \otimes \Omega_2^{(n)}, \mathcal{F}_1^{(n)} \otimes \mathcal{F}_2^{(n)}, P_1^{(n)} \otimes P_2^{(n)})$. Details of the filtrations definitions are provided in Section 1.1.1.

An estimator $\hat{\theta}_k$ of θ_{k0} is obtained by maximizing the partial likelihood

$$L_{\tau_k}(\theta_k) = \prod_{i=1}^{n} \prod_{u \in [0, \tau_k]} \lambda_{ki}(u, \theta_k)^{\Delta N_{ki}} \exp\left(-\int_0^{\tau_k} \lambda_{ki}(v, \theta_k) dv\right).$$

A summary of the main results on partial likelihood for counting processes can be found in Section II.7, Andersen *et al.* (1993).[29] Likelihood representations for general counting process models were first given by Jacod (1973, 1975).[36, 37] Use of the product integral concept to make the otherwise rather involved formulas more interpretable goes back to Johansen (1983).[38] Arjas and Haara (1984)[39] were the first to describe the notations of independent censoring rigorously by formulating these in terms of likelihoods of intensity processes of general marked point processes.

The log-partial likelihood is

$$C_{\tau_k}(\theta_k) = \sum_{i=1}^{n} \left\{ \int_0^{\tau_k} \log \lambda_{ki}(u,\theta_k) dN_{ki}(u) - \int_0^{\tau_k} \lambda_{ki}(u,\theta_k) du \right\}.$$

Because $\int_0^{\tau_k} \log Y_{ki}(u)\, dN_{ki}(u) = [0 \le X_{ki} \le \tau_k]\delta_i \log[X_{ki} \ge X_{ki}] = 0$, we can also write

$$C_{\tau_k}(\theta_k) = \sum_{i=1}^{n} \left\{ \int_0^{\tau_k} \log \alpha_{ki}(u,\theta_k) dN_{ki}(u) - \int_0^{\tau_k} \lambda_{ki}(u,\theta_k) du \right\}. \qquad (1.15)$$

Assuming we may interchange the order of differentiation and integration, the vector $U_{\tau_k}(\theta_k)$ of derivatives of $C_{\tau_k}(\theta_k)$ with respect to θ_k has the components

$$U_{\tau_k}^{j}(\theta_k) = \sum_{i=1}^{n} \left\{ \int_0^{\tau_k} \frac{\partial}{\partial \theta_{kj}} \log \alpha_{ki}(u,\theta_k) dN_{ki}(u) - \int_0^{\tau_k} \frac{\partial}{\partial \theta_{kj}} \alpha_{ki}(u,\theta_k) Y_{ki}(u) du \right\},$$

$j \in \{1,\ldots,q\}$, $q = p+1$.

The log-partial likelihood may have a number of local maxima, so the equation $U_{\tau_k}(\theta_k) = 0$ may have multiple solutions. We will consider the maximum partial likelihood estimator $\hat{\theta}_k$ given as a solution to the equation $U_{\tau_k}(\theta_k) = 0$. If more than one solution is found in a concrete situation, one could then check which of these gives the largest value of the log-partial likelihood function.

1.2.2 Large Sample Properties of the Parameter Estimators

The asymptotic properties of the parameter estimators $\left(\hat{\theta}_1', \hat{\theta}_2'\right)'$ hold under some general "regularity" conditions D.a – D.d. These conditions are stated in p420-421, Andersen *et al.* (1993).[29] They were used by Borgan (1984)[40], who studied the maximum likelihood estimation for the multiplicative intensity model. We adopt the notation

$\frac{\partial}{\partial\theta_{kj}} g(\theta_{k0})$ for $\frac{\partial}{\partial\theta_{kj}} g(\theta_k)\big|_{\theta_k=\theta_{k0}}$. The dimension of the vector θ_k of parameters is $q = p+1$.

Conditions D.a-D.d:

D.a There exists a neighborhood Θ_{k0} of θ_{k0} such that for every n, $\theta_k \in \Theta_{k0}$ and for almost all $u \in [0,\tau_k]$, the partial derivatives of $\alpha_{ki}(u,\theta_k)$ and $\log\alpha_{ki}(u,\theta_k)$ of first, second and third order with respect to θ_k exist and are continuous in θ_k for $\theta_k \in \Theta_{k0}$.

Moreover, the log-partial likelihood may be differentiated three times with respect to $\theta_k \in \Theta_{k0}$ by interchanging the order of integration and differentiation.

D.b There exist a sequence $\{a_n, n \ge 1\}$ of non-negative constants increasing to infinity as $n \to \infty$ and finite functions $\sigma_k^{jl}(\theta_k)$ defined on Θ_{k0} such that for all $j,l \in \{1,\dots,q\}$:

$$a_n^{-2} \int_0^{\tau_k} \sum_{i=1}^{n} \frac{\partial}{\partial\theta_{kj}} \log\alpha_{ki}(u,\theta_{k0}) \frac{\partial}{\partial\theta_{kl}} \log\alpha_{ki}(u,\theta_{k0}) \lambda_{ki}(u,\theta_{k0}) du \xrightarrow{P} \sigma_k^{jl}(\theta_{k0}) \text{ as}$$

$n \to \infty$.

D.c The matrix $\Sigma_k = \left(\sigma_k^{jl}(\theta_{k0}), j,l \in \{1,\dots,q\}\right)$, with $\sigma_k^{jl}(\theta_{k0})$ defined in condition D.b, is positive definite.

D.d For every n there exist predictable processes G_{ki} and H_{ki} not depending on θ_k such that for all $u \in [0,\tau_k]$:

D.d.1 $\sup_{\theta_k \in \Theta_{k0}} \left| \frac{\partial^3}{\partial\theta_{kj}\partial\theta_{kl}\partial\theta_{km}} \alpha_{ki}(u,\theta_k) \right| \leq G_{ki}(u)$

D.d.2 $\sup_{\theta_k \in \Theta_{k0}} \left| \frac{\partial^3}{\partial\theta_{kj}\partial\theta_{kl}\partial\theta_{km}} \log\alpha_{ki}(u,\theta_k) \right| \leq H_{ki}(u)$

for all $j,l,m \in \{1,\ldots,q\}$.

Moreover:

D.d.3 $a_n^{-2} \int_0^{\tau_k} \sum_{i=1}^n G_{ki}(u)du$

D.d.4 $a_n^{-2} \int_0^{\tau_k} \sum_{i=1}^n H_{ki}(u)\lambda_{ki}(u,\theta_{k0})du$

D.d.5 $a_n^{-2} \int_0^{\tau_k} \sum_{i=1}^n \left\{ \frac{\partial^2}{\partial\theta_{kj}\partial\theta_{kl}} \log\alpha_{ki}(u,\theta_{k0}) \right\}^2 \lambda_{ki}(u,\theta_{k0})du$

converge in probability to finite quantities as $n \to \infty$, and for all $\varepsilon > 0$:

D.d.6 $a_n^{-2} \int_0^{\tau_k} \sum_{i=1}^n H_{ki}(u)\left[a_n^{-1}H_{ki}(u)^{1/2} > \varepsilon\right]\lambda_{ki}(u,\theta_{k0})du \xrightarrow{P} 0.$

Note:

Deriving the statistical properties of the maximum likelihood estimators $\left(\hat{\theta}_1', \hat{\theta}_2'\right)'$ involves martingale results and Taylor expansions. Condition D.b ensures the convergence in probability of the predictable covariation processes of certain martingales. Conditions D.b and D.c are crucial in the proof of the existence and consistency of the parameter estimators. By condition D.a the Taylor expansions are valid, whereas D.d ensures the remainder terms in these expressions will behave properly. □

Proposition

Under our model assumptions B.1-B.5, the conditions D.a-D.d are verified. □

(The matrix Σ_k from conditions D.b, D.c is the one defined at the beginning of Section 1.2.1.)

Proof Proposition:

We consider $\Theta_{k0} = \Theta_{pk0} \times I_{k0}$ a neighborhood of θ_{k0} such that $\Theta_{pk0} \subset \mathbb{R}^p$, $I_{k0} \subset (0,\infty)$ are compact sets and $\beta_{k0} \in \overset{\circ}{\Theta}_{pk0}, \sigma_{k0} \in \overset{\circ}{I}_{k0}$. We denote by $\overset{\circ}{A}$ the largest open set included in A, also called the interior of A.

Condition D.a:

Recall that $\alpha_{ki}(u,\theta_k) = \sigma_k^{-1}\alpha_0\left(\dfrac{u-\beta_k' Z_{ki}}{\sigma_k}\right)$, where $\theta_k' = (\beta_k', \sigma_k)' \in \mathbb{R}^p \times (0,\infty)$.

Then $\log \alpha_{ki}(u,\theta_k) = -\log \sigma_k + \log \alpha_0\left(\dfrac{u-\beta_k' Z_{ki}}{\sigma_k}\right)$. The log function is infinitely differentiable on $(0,\infty)$, with continuous derivatives. By assumption B.4, the first part of condition D.a is verified.

The log-partial likelihood was given in (1.15):

$$C_{\tau_k}(\theta_k) = \sum_{i=1}^n \left\{ \int_0^{\tau_k} \log \alpha_{ki}(u,\theta_k) dN_{ki}(u) - \int_0^{\tau_k} \alpha_{ki}(u,\theta_k) Y_{ki}(u) du \right\},$$

where $\int_0^{\tau_k} \log \alpha_{ki}(u,\theta_k) dN_{ki}(u) = [0 \le X_{ki} \le \tau_k] \delta_i \log \alpha_{ki}(X_{ki},\theta_k)$. By the differentiability properties of $\log \alpha_{ki}$, what it is left to be proved is that $\int_0^{\tau_k} \alpha_{ki}(u,\theta_k) Y_{ki}(u) du$ can be dif

int

th

sig

(1

su

i)

ii

iii

T

differentiated three times with respect to $\theta_{k0} \in \Theta_{k0}$ by interchanging the order of integration and differentiation. We will show only that

$$\frac{\partial}{\partial \beta_k} \int_0^{\tau_k} \alpha_{ki}(u,\theta_k) Y_{ki}(u) du = \int_0^{\tau_k} \frac{\partial}{\partial \beta_k} \alpha_{ki}(u,\theta_k) Y_{ki}(u) du, \tag{1.16}$$

the proofs of the other relations having similar arguments.

The next theorem gives sufficient conditions for differentiation under the integral sign. For a proof for the 1-dimensional case see Theorem 4, p30, Fabian and Hannan (1985)[41]:

Theorem (Differentiation under the integral sign)

Let a compact set $\Theta \subset \mathbb{R}^q$, θ_0 a point in Θ and $f(u,\theta)$ a function on $[0,\tau] \times \Theta$ such that

i) $f(.,\theta)$ is Lebesgue measurable for every $\theta \in \Theta$;

ii) $f(.,\theta_0)$ is integrable;

iii) for every $u \in [0,\tau]$ the partial derivative $\frac{\partial}{\partial \theta_j} f(u,\theta)$ exists on Θ for $j \in \{1,...,q\}$ and there exist integrable functions g_j, $j \in \{1,...,q\}$ such that

$$\left| \frac{\partial}{\partial \theta_j} f(u,\theta) \right| \le g_j(u) \quad \forall u \in [0,\tau], \theta \in \Theta.$$

Then $f(.,\theta)$ is integrable for every $\theta \in \Theta$ and for every $j \in \{1,...,q\}$

$$\frac{\partial}{\partial \theta_j} \int_0^{\tau} f(u,\theta) du = \int_0^{\tau} \frac{\partial}{\partial \theta_j} f(u,\theta) du. \square$$

For every fixed θ_k, $\alpha_{ki}(u,\theta_k) Y_{ki}(u)$ is integrable because

$$\int_0^{\tau_k} \alpha_{ki}(u,\theta_k) Y_{ki}(u)du \le \int_0^{\tau_k} \alpha_{ki}(u,\theta_k)du$$

and $\alpha_{ki}(.,\theta_k)$ is continuous on $[0,\tau_k]$, so it is bounded. Also, for $\theta_k = (\beta_k,\sigma_k) \in \Theta_{k0}$ and $j \in \{1,\dots,q\}$:

$$\left| \frac{\partial}{\partial \beta_{kj}} \alpha_{ki}(u,\theta_k) Y_{ki}(u) \right| \le \sup_{\theta_k \in \Theta_{k0}} \left| \frac{\partial}{\partial \beta_{kj}} \alpha_{ki}(u,\theta_k) \right|.$$

Using $\frac{\partial}{\partial \beta_{kj}} \alpha_{ki}(u,\theta_k) = -\sigma_k^{-2} \alpha_0' \left(\frac{u - \beta_k' Z_{ki}}{\sigma_k} \right) Z_{kij}$ and the boundedness of the components of Z_{ki}, we obtain

$$\left| \frac{\partial}{\partial \beta_{kj}} \alpha_{ki}(u,\theta_k) Y_{ki}(u) \right| \le \sup_{\theta_k \in \Theta_{k0}} \left| \sigma_k^{-2} \alpha_0' \left(\frac{u - \beta_k' Z_{ki}}{\sigma_k} \right) \right| \underset{notation}{\overset{by}{=}} s(u).$$

By the following Lemma (see Lemma 1, p635, Jennrich (1969)[42]), $s(.)$ is a continuous function on $[0,\tau_k]$. Consequently it is bounded and hence integrable over any finite interval.

Lemma 1:

If g is a real valued function continuous on the Cartesian product $\mathcal{X} \times \mathcal{Y}$ of two Euclidian spaces and if Y is a bounded subset of $\mathcal{Y}$ then $\sup_{y \in \mathcal{Y}} g(x,y)$ is a continuous function of x. □

The conditions i-iii of the theorem for differentiation under the integral sign are therefore verified and (1.16) is proved.

Condition D.b:

Let $a_n = n^{1/2}$. We will show

D.b.1 $\int_0^{\tau_k} n^{-1}\sum_{i=1}^{n}\left(\frac{\partial}{\partial\beta_k}\log\alpha_{ki}(u,\theta_{k0})\right)\left(\frac{\partial}{\partial\beta_k}\log\alpha_{ki}(u,\theta_{k0})\right)'\lambda_{ki}(u,\theta_{k0})du \xrightarrow{P} \Sigma_k^{11}$, where

all the elements of the $p\times p$ matrix Σ_k^{11} are finite;

D.b.2 $\int_0^{\tau_k} n^{-1}\sum_{i=1}^{n}\left(\frac{\partial}{\partial\sigma_k}\log\alpha_{ki}(u,\theta_{k0})\right)^2\lambda_{ki}(u,\theta_{k0})du \xrightarrow{P} \sigma_k^{22} < \infty$;

D.b.3 $\int_0^{\tau_k} n^{-1}\sum_{i=1}^{n}\frac{\partial}{\partial\beta_k}\log\alpha_{ki}(u,\theta_{k0})\frac{\partial}{\partial\sigma_k}\log\alpha_{ki}(u,\theta_{k0})\lambda_{ki}(u,\theta_{k0})du \xrightarrow{P} \Sigma_k^{12}$, where all

the elements of the $p\times 1$ matrix Σ_k^{12} are finite.

The derivative functions involved in the relations D.b.1-D.b.3 are

$$\frac{\partial}{\partial\beta_k}\log\alpha_{ki}(u,\theta_k) = \frac{\alpha_0'}{\alpha_0}\left(\frac{u-\beta_k' Z_{ki}}{\sigma_k}\right)\left(-\sigma_k^{-1}Z_{ki}\right),$$

$$\frac{\partial}{\partial\sigma_k}\log\alpha_{ki}(u,\theta_k) = -\sigma_k^{-1}\left\{1+\left(\frac{u-\beta_k' Z_{ki}}{\sigma_k}\right)\frac{\alpha_0'}{\alpha_0}\left(\frac{u-\beta_k' Z_{ki}}{\sigma_k}\right)\right\}.$$

Relations D.b.1-D.b.3 can be proved through a version of Strong Law of Large Numbers (SLLN). We will show the details of the proof for D.b.1.; the other two can be proved using similar arguments.

Consider the $p\times p$ matrix

$$V_i(u,\theta_{k0}) = \left(\frac{\partial}{\partial\beta_k}\log\alpha_{ki}(u,\theta_{k0})\right)\left(\frac{\partial}{\partial\beta_k}\log\alpha_{ki}(u,\theta_{k0})\right)'\lambda_{ki}(u,\theta_{k0}) =$$

$$= \sigma_{k0}^{-3}\frac{(\alpha_0')^2}{\alpha_0}\left(\frac{u-\beta_{k0}' Z_{ki}}{\sigma_{k0}}\right)Y_{ki}(u)Z_{ki}Z_{ki}', u\in[0,\tau_k].$$

The function $\frac{(\alpha_0')^2}{\alpha_0}$ is continuous and $Y_{ki}(.)$ is right continuous with left hand limits. Let

$V_i^*(u,\theta_{k0}) = V_i(\tau_k - u, \theta_{k0}), u \in [0,\tau_k]$. Then the p^2 components of the matrix $V_i^*(u,\theta_{k0})$ are random components of $D[0,\tau_k]$, the set of right-continuous real valued functions with left-hand limits on $[0,\tau_k]$. The space $D[0,\tau_k]$ is endowed with the Skorohod topology. We will apply an extension of SLLN for $D[0,\tau_k]$. For the proof for $D[0,1]$, see Rao R.R. (1963).[43]

SLLN for $D[0,t_k]$

Let $X_i, i \geq 1$ be i.i.d. random elements of $D[0,\tau_k]$. Suppose that

$E \sup_{u\in[0,\tau_k]} |X_1(u)| < \infty$. Then $\sup_{u\in[0,\tau_k]} \left| \frac{1}{n} \sum_{i=1}^{n} X_i(u) - EX_1(u) \right| \to 0$ a.e. as $n \to \infty$. □

Let $V_{ijl}^*(.,\theta_{k0})$ the (j,l) component of the matrix $V_i^*(.,\theta_{k0})$. Because Z_{ki} are bounded derivatives,

$$E \sup_{u\in[0,\tau_k]} \left| V_{ijl}^*(u,\theta_{k0}) \right| = E \sup_{u\in[0,\tau_k]} \left| V_{1jl}(u,\theta_{k0}) \right| \leq$$

$$\leq E \sup_{u\in[0,\tau_k]} \left| \sigma_{k0}^{-3} \frac{(\alpha_0')^2}{\alpha_0} \left(\frac{u - \beta_{k0}' Z_{k1}}{\sigma_{k0}} \right) \right| \overset{by}{\underset{notation}{=}} Eg(Z_{k1}).$$

By Lemma 1, the function g is continuous so the boundedness of Z_{k1} implies $Eg(Z_{k1}) < \infty$. By the extended SLLN we obtain

$$\sup_{u\in[0,\tau_k]} \left| \frac{1}{n} \sum_{i=1}^{n} V_{ijl}^*(u,\theta_{k0}) - EV_{1jl}^*(u,\theta_{k0}) \right| \to 0 \text{ a.e. as } n \to \infty, \text{ that implies}$$

$$\sup_{u\in[0,\tau_k]} \left| \frac{1}{n} \sum_{i=1}^{n} V_{ijl}(u,\theta_{k0}) - EV_{1jl}(u,\theta_{k0}) \right| \to 0 \text{ a.e. as } n \to \infty.$$

Because $\tau_k < \infty$, the previous relation implies

$$\int_0^{\tau_k} \frac{1}{n} \sum_{i=1}^{n} V_{ijl}(u, \theta_{k0}) du \xrightarrow{P} \int_0^{\tau_k} EV_{1jl}(u, \theta_{k0}) du = (\Sigma_k^{11})_{jl}.$$

Every (j, l) component of the matrix Σ_k^{11} is finite:

$$(\Sigma_k^{11})_{jl} \le \tau_k E \sup_{u \in [0, \tau_k]} V_{1jl}(u, \theta_{k0}) < \infty.$$

Therefore condition D.b.1 holds.

Condition D.c:

By assumption B.5, the matrix Σ_k is nonsingular. Consequently we need to show that Σ_k is positive semidefinite. Let $x \in \mathbb{R}^p, y \in \mathbb{R}$. It is easy to show that

$$(x', y)\Sigma_k \begin{pmatrix} x \\ y \end{pmatrix} = E \int_0^{\tau_k} \left[\left(x' \frac{\partial}{\partial \beta_k} \log \alpha_{k1}(u, \theta_{k0}) + y \frac{\partial}{\partial \sigma_k} \log \alpha_{k1}(u, \theta_{k0}) \right)^2 \lambda_{k1}(u, \theta_{k0}) \right] du \ge 0.$$

Condition D.d:

Let $u \in [0, \tau_k]$ and $j, l, m \in \{1, \ldots, p+1\}$.

$$\sup_{\theta_k \in \Theta_{k0}} \left| \frac{\partial^3}{\partial \theta_{kj} \partial \theta_{kl} \partial \theta_{km}} \lambda_{ki}(u, \theta_{k0}) \right| \le$$

$$\le \sup_{u \in [0, \tau_k]} \sup_{\theta_k \in \Theta_{k0}} \left| \frac{\partial^3}{\partial \theta_{kj} \partial \theta_{kl} \partial \theta_{km}} \lambda_{ki}(u, \theta_{k0}) \right| \underset{notation}{\overset{by}{=}} g_k(Z_{ki}) = G_{ki}.$$

Variables G_{ki} are bounded because g_k is continuous by Lemma 1 and Z_{ki} are bounded. Therefore condition D.d.1 is verified. Because variables G_{ki} do not depend on the argument *u*, condition D.d.3 follows by regular SLLN. Similar arguments can be used for the rest of the D.d conditions.

This completes the proof of the Proposition. ■

The next results state that, with a probability tending to one, there exists a solution of the likelihood equation and this solution is consistent. However, this does not rule out the possibility of the likelihood equation having other, possibly inconsistent, solutions.

Under condition D.a the vector $U_{\tau_k}(\theta_k)$ of score statistics has the components

$$U_{\tau_k}^j(\theta_k) = \sum_{i=1}^n \int_0^{\tau_k} \frac{\partial}{\partial \theta_{kj}} \log \alpha_{ki}(u, \theta_k) dN_{ki}(u) - \sum_{i=1}^n \int_0^{\tau_k} \frac{\partial}{\partial \theta_{kj}} \alpha_{ki}(u, \theta_{k0}) Y_{ki}(u) du .$$

Let $-\mathcal{P}_{\tau_k}^{jl}(\theta_k)$, $\mathcal{R}_{\tau_k}^{jlm}(\theta_k)$ denote the second and third order partial derivatives of the log-partial likelihood $C_{\tau_k}(\theta_k)$.

Theorem 1 (Theorem VI.1.1, p422, Andersen *et al.* (1993)[29])

Under the assumptions B.1, B.2 and conditions D.a-D.d, with a probability tending to one, the equation $U_{\tau_k}(\theta_k) = 0$ has a solution $\hat{\theta}_k$ and $\hat{\theta}_k \xrightarrow{P} \theta_{k0}$ as $n \to \infty$.□

Sketch of the proof:

By condition D.a, a Taylor expansion gives for every $\theta_k \in \Theta_{k0}$, $j \in \{1, \ldots, q\}$ that

$$U_{\tau_k}^j(\theta_k) = U_{\tau_k}^j(\theta_{k0}) - \sum_{l=1}^q (\theta_{kl} - \theta_{k0l}) \mathcal{P}_{\tau_k}^{jl}(\theta_{k0}) +$$
$$+ \frac{1}{2} \sum_{l=1}^q \sum_{m=1}^q (\theta_{kl} - \theta_{k0l})(\theta_{km} - \theta_{k0m}) \mathcal{R}_{\tau_k}^{jlm}(\theta_k^*),$$

where θ_k^* is on the line segment joining θ_k and θ_{k0}.

Step 1:

$$a_n^{-2} U_{\tau_k}^j(\theta_{k0}) \xrightarrow{P} 0 .$$

Essential in this step is that

$$U_{\tau_k}^{j}(\theta_{k0}) = \sum_{i=1}^{n} \int_0^{\tau_k} \frac{\partial}{\partial \theta_{kj}} \log \alpha_{ki}(u, \theta_{k0}) dM_{ki}(u).$$

Step 2:

$$a_n^{-2} \mathcal{P}_{\tau_k}^{jl}(\theta_{k0}) \xrightarrow{P} \sigma_k^{jl}(\theta_{k0}).$$

Step 3:

There exists $M_k < \infty$, not depending on θ_k, such that

$$\lim_{n\to\infty} P(A_n) \underset{notation}{\overset{by}{=}} \lim_{n\to\infty} P\left(\left|a_n^{-2} \mathcal{R}_{\tau_k}^{jlm}(\theta_k)\right| < M_k \ \forall j,l,m, \ \forall \theta_k \in \Theta_{k0}\right) = 1.$$

Step 4:

Combine the previous steps and finish the proof of the Theorem. ■

Next we prove the following result about the joint distribution of the maximum partial-likelihood estimators:

Theorem 2

Assume that our model assumptions B.1-B.5 hold. Let $\hat{\theta}_k$ be a consistent solution of the equation $U_{\tau_k}(\theta_k) = 0$. Then $n^{1/2}\left(\hat{\theta}_1' - \theta_{10}', \hat{\theta}_2' - \theta_{20}'\right)'$ converges in distribution to a $2(p+1)$-dimensional normal random vector with zero mean and covariance matrix $C = \left(C_{kl},\, k,l \in \{1,2\}\right)$, where the $(p+1)\times(p+1)$ matrix C_{kl} is

$$C_{kl} = \Sigma_k^{-1} B_{kl} \Sigma_l^{-1},$$

with $B_{kl} = E\left(v_{k1}(\theta_{k0}) v_{l1}(\theta_{l0})'\right)$ and the $(p+1)$-dimensional vectors $v_{ki}(\theta_{k0})$ given by

$$v_{ki}(\theta_{k0}) = \int_0^{\tau_k} \frac{\partial}{\partial\theta_k} \log\alpha_{ki}(u,\theta_{k0}) dM_{ki}(u). \square$$

Proof Theorem 2:

For n sufficiently large $\hat{\theta}_k \in \Theta_{k0}$. Expanding $U_{\tau_k}(\hat{\theta}_k)$ around θ_{k0} gives

$$n^{-1/2}U_{\tau_k}(\theta_{k0}) = n^{-1}\mathcal{P}_{\tau_k}(\theta_k^*)n^{1/2}\left(\hat{\theta}_k - \theta_{k0}\right).$$

Step 1:

$n^{-1/2}\left(U_{\tau_1}(\theta_{10})', U_{\tau_2}(\theta_{20})'\right)'$ converges in distribution to a $2q$-dimensional normal random vector ($q = p+1$), with mean zero and covariance matrix $B = \left(B_{kl}, k,l \in \{1,2\}\right)$.

Because $n^{-1/2}U_{\tau_k}(\theta_{k0}) = n^{-1/2}\sum_{i=1}^n v_{ki}(\theta_{k0})$ is a sum of i.i.d. zero mean random vectors, the result of this step follows immediately from the Multivariate Central Limit Theorem.

Step 2:

$n^{-1}\mathcal{P}_{\tau_k}(\theta_k^*) \xrightarrow{P} \Sigma_k$ for any random θ_k^* such that $\theta_k^* \xrightarrow{P} \theta_{k0}$.

With probability approaching 1, θ_k^* lies in Θ_{k0}. When $\theta_k^* \in \Theta_{k0}$, by the Taylor expansion, for every $j,l \in \{1,\dots,q\}$,

$$n^{-1}\mathcal{P}_{\tau_k}^{jl}(\theta_k^*) = n^{-1}\mathcal{P}_{\tau_k}^{jl}(\theta_{k0}) - n^{-1}\sum_{m=1}^q \left(\theta_{km}^* - \theta_{k0m}\right)\mathcal{R}_{\tau_k}^{jlm}(\tilde{\theta}_k),$$

where $\tilde{\theta}_k$ is on the line segment joining θ_k^* and θ_{k0}.

By Step 2 of the proof of Theorem 1, $n^{-1}\mathcal{P}_{\tau_k}^{jl}(\theta_{k0}) \xrightarrow{P} \sigma_k^{jl}(\theta_{k0})$. We will show that the second term of the right-hand side of the previous expression converges in probability to zero.

Let the constant M_k and the sequence of events A_n be defined as in the Step 3 of Theorem 1. We have that $P(A_n) \to 1$. Consider an arbitrary $\varepsilon > 0$ and the sequence of events

$$B_n = \left\{ \left| n^{-1} \sum_{m=1}^{q} \left(\theta_{km}^{*} - \theta_{k0m} \right) \mathcal{R}_{\tau_k}^{jlm}(\tilde{\theta}_k) \right| > \varepsilon \right\}.$$

We need to show that $P(B_n) \to 0$. This follows from the inequality

$$0 \le P(B_n) = P(B_n \cap A_n) + P(B_n \cap A_n^c),$$

where A_n^c is the complementary set, and the results

$$P(B_n \cap A_n^c) \le P(A_n^c) \to 0 \text{ and}$$

$$P(B_n \cap A_n) \le P\left(M_k \left| \sum_{m=1}^{q} \left(\theta_{km}^{*} - \theta_{k0m} \right) \right| > \varepsilon \right) \le$$

$$\le P\left(\left\| \theta_k^{*} - \theta_{k0} \right\| > \frac{\varepsilon}{q M_k} \right) \to 0 \text{ because } \theta_k^{*} \xrightarrow{P} \theta_{k0}.$$

We denote by $\| \cdot \|$ the supremum norm.

Step 3:

$n^{1/2}\left(\hat{\theta}_1' - \theta_{10}', \hat{\theta}_2' - \theta_{20}'\right)'$ converges in distribution to a $2(p+1)$-dimensional normal random vector with mean zero and covariance matrix $C = A^{-1} B A^{-1}$, where $A = diag(\Sigma_1, \Sigma_2)$ and $B = \left(B_{kl}, k, l \in \{1,2\}\right)$.

By a Taylor expansion we can write

$$n^{-1/2}\begin{pmatrix} U_{\tau_1}(\theta_{10}) \\ U_{\tau_2}(\theta_{20}) \end{pmatrix} = diag(n^{-1}\mathcal{P}_{\tau_1}(\theta_1^*), n^{-1}\mathcal{P}_{\tau_2}(\theta_2^*))n^{1/2}\begin{pmatrix} \hat{\theta}_1 - \theta_{10} \\ \hat{\theta}_2 - \theta_{20} \end{pmatrix}.$$

We apply the following lemma:

Lemma 2 (Theorem 10.1, p62, Billingsley (1961)[44])

Let E an Euclidian space. Suppose u_n is a random vector in E^r satisfying $u_n \xrightarrow{D} \mu$, where μ is a probability measure in E^r. Suppose further that v_n is a second random vector in E^r satisfying either

$$|u_n - v_n| \le \varepsilon_n |u_n| \text{ or } |u_n - v_n| \le \varepsilon_n' |v_n|,$$

where $\varepsilon_n \xrightarrow{P} 0$, $\varepsilon_n' \xrightarrow{P} 0$. Then v_n has the same limiting distribution μ as u_n. □

We have

$$\left\| n^{-1/2}\begin{pmatrix} U_{\tau_1}(\theta_{10}) \\ U_{\tau_2}(\theta_{20}) \end{pmatrix} - An^{1/2}\begin{pmatrix} \hat{\theta}_1 - \theta_{10} \\ \hat{\theta}_2 - \theta_{20} \end{pmatrix} \right\| \le$$

$$\le \left\| diag(n^{-1}\mathcal{P}_{\tau_1}(\theta_1^*), n^{-1}\mathcal{P}_{\tau_2}(\theta_2^*)) - A \right\| \times \left\| n^{1/2}\begin{pmatrix} \hat{\theta}_1 - \theta_{10} \\ \hat{\theta}_2 - \theta_{20} \end{pmatrix} \right\|.$$

By Step 2 the first factor of the right-hand side converges to zero in probability.

Applying Lemma 2 we obtain that $An^{1/2}\begin{pmatrix} \hat{\theta}_1 - \theta_{10} \\ \hat{\theta}_2 - \theta_{20} \end{pmatrix}$ has the same asymptotic distribution as $n^{-1/2}\begin{pmatrix} U_{\tau_1}(\theta_{10}) \\ U_{\tau_2}(\theta_{20}) \end{pmatrix}$. By Step 1, the result of Step 3 is shown and therefore Theorem 2 is proved. ■

The next theorem provides a consistent estimator for the asymptotic covariance matrix in Theorem 2.

Theorem 3

Under the model assumptions B.1-B.5, the asymptotic covariance matrix $\mathcal{C}=\left(C_{kl},k,l\in\{1,2\}\right)$ of $n^{1/2}\left(\hat{\theta}_1'-\theta_{10}',\hat{\theta}_2'-\theta_{20}'\right)'$ is consistently estimated by $\hat{\mathcal{C}}=\left(\hat{C}_{kl},k,l\in\{1,2\}\right)$, with

$$\hat{C}_{kl}=n^2\mathcal{P}_{\tau_k}^{-1}(\hat{\theta}_k)\hat{B}_{kl}\mathcal{P}_{\tau_l}^{-1}(\hat{\theta}_l),$$

where $\hat{B}_{kl}=n^{-1}\sum_{i=1}^{n}\hat{V}_{ki}(\hat{\theta}_k)\hat{V}_{li}(\hat{\theta}_l)'$ and $\hat{V}_{ki}(\hat{\theta}_k)$ is a $(p+1)$-dimensional vector,

$$\hat{V}_{ki}(\hat{\theta}_k)=\int_0^{\tau_k}\frac{\partial}{\partial\theta_k}\log\alpha_{ki}(u,\hat{\theta}_k)d\tilde{M}_{ki}(u),$$

$$\tilde{M}_{ki}(u)=N_{ki}(u)-\int_0^u Y_{ki}(s)\alpha_{ki}(s,\hat{\theta}_k)ds\,.\square$$

Proof Theorem 3:

By Step 2 of the previous theorem and Slutsky's Lemma, what is left to be proved is the convergence $\hat{B}_{kl}\xrightarrow{P}B_{kl}$, $k,l\in\{1,2\}$. Recall that $B_{kl}=E\left(v_{k1}(\theta_{k0})v_{l1}(\theta_{l0})'\right)$, where the $(p+1)$-dimensional vectors $v_{ki}(\theta_{k0})$ have the form

$$v_{ki}(\theta_{k0})=\int_0^{\tau_k}\frac{\partial}{\partial\theta_k}\log\alpha_{ki}(u,\theta_{k0})dM_{ki}(u).$$

Therefore we will show that for every $j,m\in\{1,\dots,p+1\}$

$$n^{-1}\sum_{i=1}^{n}\hat{V}_{ki}^{j}(\hat{\theta}_k)\hat{V}_{li}^{m}(\hat{\theta}_l)'\xrightarrow{P}E\left(v_{k1}^{j}(\theta_{k0})v_{l1}^{m}(\theta_{l0})\right). \tag{1.17}$$

The j-th components of the vectors $\hat{V}_{ki}(\hat{\theta}_k)$ and $v_{k1}(\theta_{k0})$ are

$$\hat{V}_{ki}^j(\hat{\theta}_k) = \delta_{ki}[0 \le X_{ki} \le \tau_k]\frac{\partial}{\partial \theta_{kj}}\log \alpha_{ki}(X_{ki},\hat{\theta}_k) - \int_0^{\tau_k}\frac{\partial}{\partial \theta_{kj}}\alpha_{ki}(u,\hat{\theta}_k)Y_{ki}(u)du\,,$$

$$v_{k1}^j(\theta_{k0}) = \delta_{ki}[0 \le X_{k1} \le \tau_k]\frac{\partial}{\partial \theta_{kj}}\log \alpha_{k1}(X_{k1},\theta_{k0}) - \int_0^{\tau_k}\frac{\partial}{\partial \theta_{kj}}\alpha_{k1}(u,\theta_{k0})Y_{k1}(u)du\,.$$

Consequently $n^{-1}\sum_{i=1}^n \hat{V}_{ki}^j(\hat{\theta}_k)\hat{V}_{li}^m(\hat{\theta}_l)'$ can be expanded as $R1 - R2 - R3 + R4$, where

$$R1 = n^{-1}\sum_{i=1}^n \delta_{ki}\delta_{li}[0 \le X_{ki} \le \tau_k][0 \le X_{li} \le \tau_l]\frac{\partial}{\partial \theta_{kj}}\log \alpha_{ki}(X_{ki},\hat{\theta}_k)\frac{\partial}{\partial \theta_{lm}}\log \alpha_{li}(X_{li},\hat{\theta}_l)\,,$$

$$R2 = n^{-1}\sum_{i=1}^n \delta_{ki}[0 \le X_{ki} \le \tau_k]\frac{\partial}{\partial \theta_{kj}}\log \alpha_{ki}(X_{ki},\hat{\theta}_k)\int_0^{\tau_l}\frac{\partial}{\partial \theta_{lm}}\alpha_{li}(s,\hat{\theta}_l)Y_{li}(s)ds\,,$$

$$R3 = n^{-1}\sum_{i=1}^n \delta_{li}[0 \le X_{li} \le \tau_l]\frac{\partial}{\partial \theta_{lm}}\log \alpha_{li}(X_{li},\hat{\theta}_l)\int_0^{\tau_k}\frac{\partial}{\partial \theta_{kj}}\alpha_{ki}(u,\hat{\theta}_k)Y_{ki}(u)du\,,$$

$$R4 = n^{-1}\sum_{i=1}^n \int_0^{\tau_k}\frac{\partial}{\partial \theta_{kj}}\alpha_{ki}(u,\hat{\theta}_k)Y_{ki}(u)du\int_0^{\tau_l}\frac{\partial}{\partial \theta_{lm}}\alpha_{li}(s,\hat{\theta}_l)Y_{li}(s)ds\,.$$

Similarly $E\left(v_{k1}^j(\theta_{k0})v_{l1}^m(\theta_{l0})\right)$ can be expanded as $L1 - L2 - L3 + L4$, where

$$L1 = E\left(\delta_{k1}\delta_{l1}[0 \le X_{k1} \le \tau_1][0 \le X_{l1} \le \tau_l]\frac{\partial}{\partial \theta_{kj}}\log \alpha_{k1}(X_{k1},\theta_{k0})\frac{\partial}{\partial \theta_{lm}}\log \alpha_{l1}(X_{l1},\theta_{l0})\right),$$

$$L2 = E\left(\delta_{k1}[0 \le X_{k1} \le \tau_k]\frac{\partial}{\partial \theta_{kj}}\log \alpha_{k1}(X_{k1},\theta_{k0})\int_0^{\tau_l}\frac{\partial}{\partial \theta_{lm}}\alpha_{l1}(s,\theta_{l0})Y_{l1}(s)ds\right),$$

$$L3 = E\left(\delta_{l1}[0 \le X_{l1} \le \tau_l]\frac{\partial}{\partial \theta_{lm}}\log \alpha_{l1}(X_{l1},\theta_{l0})\int_0^{\tau_k}\frac{\partial}{\partial \theta_{kj}}\alpha_{k1}(u,\theta_{k0})Y_{k1}(u)du\right),$$

$$L4 = E\left(\int_0^{\tau_k}\frac{\partial}{\partial \theta_{kj}}\alpha_{k1}(u,\theta_{k0})Y_{k1}(u)du\int_0^{\tau_l}\frac{\partial}{\partial \theta_{lm}}\alpha_{l1}(s,\theta_{l0})Y_{l1}(s)ds\right).$$

By Slutsky's Lemma, for showing (1.17) it is sufficient to prove that $Ra \xrightarrow{P} La$ for $a \in \{1,2,3,4\}$ and $k,l \in \{1,2\}$. We will provide the details for $k = 1, l = 2$, the other cases being similar.

We start by proving $R1 \xrightarrow{P} L1$.

Recall that $\alpha_{1i}(t,\theta_1) = \sigma_1^{-1}\alpha_0\left(\frac{t-\beta_1' Z_{1i}}{\sigma_1}\right)$. Let

$\xi(t,c,\theta_1,\theta_2,z_1,z_2) = \frac{\partial}{\partial\theta_{1j}}\log\left(\sigma_1^{-1}\alpha_0\left(\frac{t-\beta_1' z_1}{\sigma_1}\right)\right)\frac{\partial}{\partial\theta_{2m}}\log\left(\sigma_2^{-1}\alpha_0\left(\frac{c-\beta_2' z_2}{\sigma_2}\right)\right)$ and

$R_{1i}(\theta_1,\theta_2) = \delta_{1i}\delta_{2i}[0 \le T_i \le \tau_1][0 \le C_i \le \tau_2]\xi(T_i,C_i,\theta_1,\theta_2,Z_{1i},Z_{2i})$. Then

$R1 = n^{-1}\sum_{i=1}^n R_{1i}(\hat\theta_1,\hat\theta_2)$ and $L1 = E\left(R_{11}(\theta_{10},\theta_{20})\right)$.

First we show that

$$\sup_{(\theta_1,\theta_2)\in\Theta_{10}\times\Theta_{20}}\left|n^{-1}\sum_{i=1}^n R_{1i}(\theta_1,\theta_2) - E\left(R_{11}(\theta_1,\theta_2)\right)\right| \to 0 \text{ a.e. as } n\to\infty. \qquad (1.18)$$

We will apply a SLLN for separable Banach spaces, first proved by Mourier (1953)[45]:

SLLN 1:

If $(X,\|\,.\,\|)$ is a separable Banach space and $\{V_n\}$ a sequence of i.i.d. random elements in X such that $E\|V_1\| < \infty$ then $\left\|n^{-1}\sum_{i=1}^n V_i - EV_1\right\| \to 0$ a.e. as $n\to\infty$. □

Consider the separable Banach space of continuous functions on the compact set $\Theta_{10}\times\Theta_{20}$, endowed with the supremum norm. Then $R_{1i}(.,.)$ are i.i.d. random elements of this space. The convergence (1.18) follows from the direct application of the stated SLLN 1 if we show that

$$E\left(\sup_{(\theta_1,\theta_2)\in\Theta_{10}\times\Theta_{20}}\left|R_{11}(\theta_1,\theta_2)\right|\right)<\infty. \tag{1.19}$$

Let $\xi_1(t,c,z_1,z_2)=\sup_{(\theta_1,\theta_2)\in\Theta_{10}\times\Theta_{20}}\left|\xi(t,c,\theta_1,\theta_2,z_1,z_2)\right|$. Then

$$E\left(\sup_{(\theta_1,\theta_2)\in\Theta_{10}\times\Theta_{20}}\left|R_{11}(\theta_1,\theta_2)\right|\right)\le E\left([0\le T_1\le\tau_1][0\le C_1\le\tau_2]\xi_1(T_1,C_1,Z_{11},Z_{21})\right)=$$

$$=E\left[E\left([0\le T_1\le\tau_1][0\le C_1\le\tau_2]\xi_1(T_1,C_1,Z_{11},Z_{21})\mid Z_{11},Z_{21}\right)\right].$$

Given the covariates Z_{11},Z_{21}, the distribution of $(T_1,C_1)'$ has distribution function $\boldsymbol{F}^*(.,.,\theta_{10},\theta_{20},\rho_0)$ and density $f^*(.,.,\theta_{10},\theta_{20},\rho_0)$. Therefore the right hand side of the previous inequality has the form

$$E\left(\int_0^{\tau_1}\int_0^{\tau_2}\xi_1(t,c,Z_{11},Z_{21})f^*(t,c,\theta_{10},\theta_{20},\rho_0)dtdc\right)\le$$

$$\le E\left(\sup_{(t,c)\in[0,\tau_1]\times[0,\tau_2]}\xi_1(t,c,Z_{11},Z_{21})F^*(\tau_1,\tau_2,\theta_{10},\theta_{20},\rho_0)\right)\le$$

$$\le E\left(\sup_{(t,c)\in[0,\tau_1]\times[0,\tau_2]}\ \sup_{(\theta_1,\theta_2)\in\Theta_{10}\times\Theta_{20}}\xi(t,c,\theta_1,\theta_2,Z_{11},Z_{21})\right)\underset{notation}{\overset{by}{=}}Eg(Z_{11},Z_{21}).$$

Because the function ξ is continuous in all its six arguments, by Lemma 1 stated in the proof of Proposition, $g(.,.)$ is also continuous. By the assumed boundedness of covariates it follows that $Eg(Z_{11},Z_{21})<\infty$ so (1.19) and therefore (1.18) are proved.

We assumed θ_{k0} in the interior of the set Θ_{k0}. Because $\hat{\theta}_k\xrightarrow{P}\theta_{k0}$ as $n\to\infty$ then, with a probability approaching one, $\hat{\theta}_k\in\Theta_{k0}$. Consider an arbitrary $\varepsilon>0$. By (1.19),

$$\left|n^{-1}\sum_{i=1}^{n}R_{1i}(\hat{\theta}_1,\hat{\theta}_2)-E\left(R_{11}(\hat{\theta}_1,\hat{\theta}_2)\right)\right|<\varepsilon/2 \tag{1.20}$$

with probability tending to one as $n\to\infty$.

By Dominated Convergence Theorem, $ER_{11}(.,.)$ is a continuous function on the compact set $\Theta_{10} \times \Theta_{20}$. Then

$$\left| ER_{11}(\hat{\theta}_1, \hat{\theta}_2) - ER_{11}(\theta_{10}, \theta_{20}) \right| < \varepsilon / 2 \tag{1.21}$$

with probability tending to one as $n \to \infty$.

The inequalities (1.20) and (1.21) imply that

$$\left| n^{-1} \sum_{i=1}^{n} R_{1i}(\hat{\theta}_1, \hat{\theta}_2) - ER_{11}(\theta_{10}, \theta_{20}) \right| < \varepsilon$$

with probability tending to one as $n \to \infty$. Hence we proved $R1 \xrightarrow{P} L1$.

Next we show that $R2 \xrightarrow{P} L2$.

Let $\eta(t, c, \theta_1, \theta_2, z_1, z_2)$ the function obtained from $\frac{\partial}{\partial \theta_{1j}} \log \alpha_{11}(t, \theta_1) \frac{\partial}{\partial \theta_{2m}} \alpha_{21}(c, \theta_2)$ by replacing the covariates Z_{11}, Z_{21} with the arguments z_1, z_2. Define also $R_{2i}(c, \theta_1, \theta_2) = \delta_{1i}[0 \le T_i \le \tau_1] \eta(T_i, c, \theta_1, \theta_2, Z_{1i}, Z_{2i}) Y_{2i}(c)$. By Fubini's Theorem $R2 = \int_0^{\tau_2} (n^{-1} \sum_{i=1}^{n} R_{2i}(c, \hat{\theta}_1, \hat{\theta}_2)) dc$ and $L2 = \int_0^{\tau_2} ER_{21}(c, \theta_{10}, \theta_{20}) dc$.

First we show that

$$\sup_{c \in [0, \tau_2]} \sup_{(\theta_1, \theta_2) \in \Theta_{10} \times \Theta_{20}} \left| n^{-1} \sum_{i=1}^{n} R_{2i}(c, \theta_1, \theta_2) - ER_{21}(c, \theta_1, \theta_2) \right| \to 0 \text{ a.e. as } n \to \infty. \tag{1.22}$$

We will apply an extension of SLLN to $D_E[0, \tau_2]$, the set of right-continuous functions with left-hand limits on $[0, \tau_2]$, taking values in a separable Banach space E. The space $D_E[0, \tau_2]$ is endowed with the Skorohod topology. For a proof of this result see Andersen and Gill (1982).[31]

SLLN 2:

Let $\{V_n\}$ a sequence of i.i.d. random elements of $D_E[0,\tau_2]$. Suppose

$E\|V_1\| = E \sup_{c\in[0,\tau_2]} \|V_1(c)\| < \infty$. Then $\left\|n^{-1}\sum_{i=1}^{n} V_i - EV_1\right\| \to 0$ a.e. as $n \to \infty$. □

In our case E is the space of continuous real functions on $\Theta_{10} \times \Theta_{20}$, endowed with the supremum norm, and $V_i(c) = R_{2i}(\tau_2 - c,.,.)$. For applying SLLN 2 we need to show

$$E\left(\sup_{c\in[0,\tau_2]} \sup_{(\theta_1,\theta_2)\in\Theta_{10}\times\Theta_{20}} \left|R_{21}(c,\theta_1,\theta_2)\right|\right) < \infty. \tag{1.23}$$

Let $\eta_1(t,z_1,z_2) = \sup_{c\in[0,\tau_2]} \sup_{(\theta_1,\theta_2)\in\Theta_{10}\times\Theta_{20}} \left|\eta(t,c,\theta_1,\theta_2,z_1,z_2)\right|$. Then

$$E\left(\sup_{c\in[0,\tau_2]} \sup_{(\theta_1,\theta_2)\in\Theta_{10}\times\Theta_{20}} \left|R_{21}(c,\theta_1,\theta_2)\right|\right) \le E\left([0 \le T_1 \le \tau_1]\eta_1(T_1,Z_{11},Z_{21})\right) =$$

$$= E\left[E\left([0 \le T_1 \le \tau_1]\eta_1(T_1,Z_{11},Z_{21}) \mid Z_{11},Z_{21}\right)\right] =$$

$$= E\left(\int_0^{\tau_1}\int_{-\infty}^{\infty} \eta_1(t,Z_{11},Z_{21}) f^*(t,c,\theta_{10},\theta_{20},\rho_0)dtdc\right) \le$$

$$\le E\left(\sup_{t\in[0,\tau_1]} \eta_1(t,Z_{11},Z_{21})\right) \underset{notation}{\overset{by}{=}} Eg_1(Z_{11},Z_{21}).$$

By the same argument used to prove (1.19), $Eg_1(Z_{11},Z_{21}) < \infty$. Therefore (1.23) is shown and, by the stated SLLN 2, (1.22) follows.

Consider an arbitrary $\varepsilon > 0$. With probability tending to one, $\hat{\theta}_k \in \Theta_{k0}$. Then, as a consequence of (1.22),

$$\sup_{c\in[0,\tau_2]} \left|n^{-1}\sum_{i=1}^{n} R_{2i}(c,\hat{\theta}_1,\hat{\theta}_2) - ER_{21}(c,\hat{\theta}_1,\hat{\theta}_2)\right| < \varepsilon / 2\tau_2$$

with probability tending to one as $n \to \infty$. Then

$$\left|\int_0^{\tau_2} n^{-1}\sum_{i=1}^n R_{2i}(c,\hat{\theta}_1,\hat{\theta}_2)dc - \int_0^{\tau_2} ER_{21}(c,\hat{\theta}_1,\hat{\theta}_2)dc\right| \le$$
$$\le \tau_2 \sup_{c\in[0,\tau_2]}\left|n^{-1}\sum_{i=1}^n R_{2i}(c,\hat{\theta}_1,\hat{\theta}_2) - ER_{21}(c,\hat{\theta}_1,\hat{\theta}_2)\right| < \tau_2 \frac{\varepsilon}{2\tau_2} = \varepsilon/2 \tag{1.24}$$

with probability approaching one as $n \to \infty$.

By (1.23) and Dominated Convergence Theorem, the function

$$(\theta_1,\theta_2) \to \int_0^{\tau_2} ER_{21}(c,\theta_1,\theta_2)dc = E\int_0^{\tau_2} R_{21}(c,\theta_1,\theta_2)dc$$

is continuous on $\Theta_{10} \times \Theta_{20}$, so

$$\left|\int_0^{\tau_2} ER_{21}(c,\hat{\theta}_1,\hat{\theta}_2)dc - \int_0^{\tau_2} ER_{21}(c,\theta_{10},\theta_{20})dc\right| < \varepsilon/2 \tag{1.25}$$

with probability tending to one as $n \to \infty$.

The inequalities (1.24) and (1.25) imply that

$$\left|\int_0^{\tau_2} n^{-1}\sum_{i=1}^n R_{2i}(c,\hat{\theta}_1,\hat{\theta}_2)dc - \int_0^{\tau_2} ER_{21}(c,\theta_{10},\theta_{20})dc\right| < \varepsilon$$

with probability approaching one as $n \to \infty$. Therefore the convergence $R2 \xrightarrow{P} L2$ is proved. The proof for $R3 \xrightarrow{P} L3$ is identical.

Finally we will show that $R4 \xrightarrow{P} L4$.

Let $\zeta(t,c,\theta_1,\theta_2,z_1,z_2)$ the function obtained from $\frac{\partial}{\partial\theta_{1j}}\alpha_{11}(t,\theta_1)\frac{\partial}{\partial\theta_{2m}}\alpha_{21}(c,\theta_2)$ by replacing the covariates Z_{11}, Z_{21} with the arguments z_1, z_2. Define also $R_{4i}(t,c,\theta_1,\theta_2) = \zeta(t,c,\theta_1,\theta_2,Z_{1i},Z_{2i})Y_{1i}(t)Y_{2i}(c)$. By Fubini's Theorem

$R4 = \int_0^{\tau_1}\int_0^{\tau_2}(n^{-1}\sum_{i=1}^n R_{4i}(t,c,\hat{\theta}_1,\hat{\theta}_2))dtdc$ and $L4 = \int_0^{\tau_1}\int_0^{\tau_2} ER_{41}(t,c,\theta_{10},\theta_{20})dtdc$.

The proof follows the same steps as in the proof of $R2 \xrightarrow{P} L2$. We will show only that

$$\sup_{(t,c)\in[0,\tau_1]\times[0,\tau_2]}\ \sup_{(\theta_1,\theta_2)\in\Theta_{10}\times\Theta_{20}} \left| n^{-1}\sum_{i=1}^{n} R_{4i}(t,c,\theta_1,\theta_2) - ER_{41}(t,c,\theta_1,\theta_2) \right| \to 0 \text{ a.e.} \qquad (1.26)$$

as $n \to \infty$. The SLLN that has to be applied in this case is an extension to the space $D_E([0,\tau_1]\times[0,\tau_2])$ and it is proved in Appendix A.

SLLN 3:

Let $\{V_n\}$ a sequence of i.i.d. random elements of $D_E([0,\tau_1]\times[0,\tau_2])$. Suppose $E\|V_1\| = E \sup_{(t,c)\in[0,\tau_1]\times[0,\tau_2]} \|V_1(t,c)\| < \infty$. Then $\left\| n^{-1}\sum_{i=1}^{n} V_i - EV_1 \right\| \to 0$ a.e. as $n \to \infty$. □

In our case E is the space of continuous real functions on $\Theta_{10}\times\Theta_{20}$, endowed with the supremum norm, and $V_i(t,c) = R_{4i}(\tau_1 - t, \tau_2 - c, .,.)$. The SLLN 3 assumption is satisfied:

$$E\left(\sup_{(t,c)\in[0,\tau_1]\times[0,\tau_2]}\ \sup_{(\theta_1,\theta_2)\in\Theta_{10}\times\Theta_{20}} \left| R_{41}(t,c,\theta_1,\theta_2) \right| \right) \le$$

$$\le E\left(\sup_{(t,c)\in[0,\tau_1]\times[0,\tau_2]}\ \sup_{(\theta_1,\theta_2)\in\Theta_{10}\times\Theta_{20}} \left| \zeta(t,c,\theta_1,\theta_2,Z_{11},Z_{21}) \right| \right) \underset{notation}{\overset{by}{=}} Eg_2(Z_{11},Z_{21}) < \infty,$$

because g_2 is a continuous function and Z_{11}, Z_{21} are bounded. Thus (1.26) is proved.

This concludes the proof of Theorem 3. ■

1.2.3 Point Estimates and Confidence Intervals

for the Median LOS and Median Cost;

Application for the Bivariate Normal Case

In Section 1.2 joint analysis of hospital LOS and cost is based on the linear model (1.12). Let $\varepsilon_{.5}$ the median of the distribution of the marginal errors ε_{ki} and Z_0 a specified covariate profile. Denote by $Z_{0\varepsilon}$ the vector $\left(Z_0', \varepsilon_{.5}\right)'$.

By the model (1.12), we consider the point estimates for the g(LOS) median and g(cost) median (denoted $T_{.5}$ and $C_{.5}$, respectively) to be

$$\begin{aligned} \hat{T}_{.5} &= \hat{\beta}_1' Z_0 + \hat{\sigma}_1 \varepsilon_{.5} = \hat{\theta}_1' Z_{0\varepsilon}, \\ \hat{C}_{.5} &= \hat{\beta}_2' Z_0 + \hat{\sigma}_2 \varepsilon_{.5} = \hat{\theta}_2' Z_{0\varepsilon}, \end{aligned} \tag{1.27}$$

where $\hat{\theta}_k$ is a consistent estimator of $\theta_{k0} = \left(\beta_{k0}', \sigma_{k0}\right)'$.

The standard errors of the estimators $\hat{T}_{.5}$ and $\hat{C}_{.5}$ are

$$SE(\hat{T}_{.5}) = \left[Z_{0\varepsilon}'(n^{-1}C_{11})Z_{0\varepsilon}\right]^{1/2},$$

$$SE(\hat{C}_{.5}) = \left[Z_{0\varepsilon}'(n^{-1}C_{22})Z_{0\varepsilon}\right]^{1/2},$$

where C_{kk} is the asymptotic covariance matrix of $n^{1/2}\left(\hat{\theta}_k - \theta_{k0}\right)$, as defined in the Theorem 2 of the sub-section 1.2.2. Consistent estimators of the standard errors $SE(\hat{T}_{.5})$ and $SE(\hat{C}_{.5})$ are

$$\hat{SE}(\hat{T}_{.5}) = \left[Z'_{0\varepsilon}(n^{-1}\hat{C}_{11})Z_{0\varepsilon}\right]^{1/2},$$
$$\hat{SE}(\hat{C}_{.5}) = \left[Z'_{0\varepsilon}(n^{-1}\hat{C}_{22})Z_{0\varepsilon}\right]^{1/2}, \tag{1.28}$$

where $\hat{C}_{kk}$ is the consistent estimator of the asymptotic covariance matrix C_{kk} (see Theorem 3).

Then, retransforming the results in the original unit of measurement, given a covariate profile Z_0, we consider for the median LOS and median cost the following point estimates:

$$\hat{T}^*_{.5} = g^{-1}(\hat{T}_{.5}),$$
$$\hat{C}^*_{.5} = g^{-1}(\hat{C}_{.5}).$$

Considering g^{-1} nondecreasing, the confidence intervals are defined as

$$\left[g^{-1}\left(\hat{T}_{.5} - z_\alpha \hat{SE}(\hat{T}_{.5})\right), g^{-1}\left(\hat{T}_{.5} + z_\alpha \hat{SE}(\hat{T}_{.5})\right)\right],$$

$$\left[g^{-1}\left(\hat{C}_{.5} - z_\alpha \hat{SE}(\hat{C}_{.5})\right), g^{-1}\left(\hat{C}_{.5} + z_\alpha \hat{SE}(\hat{C}_{.5})\right)\right],$$

where z_α is the upper α quantile of the standard normal distribution and $\hat{T}_{.5}, \hat{C}_{.5}$ and $\hat{SE}(\hat{T}_{.5}), \hat{SE}(\hat{C}_{.5})$ are given in (1.27) and (1.28), respectively.

Application: Bivariate Normal Case

In the setting of Example 1.2.1, ε_{ki} are i.i.d. $N(0,1)$, which is a symmetric distribution with median $\varepsilon_{.5} = 0$.

Using the notations of Theorem 2, consider

$$C_{kk}=\begin{pmatrix} C_{kk}^{11} & C_{kk}^{12} \\ \left(C_{kk}^{12}\right)' & C_{kk}^{22} \end{pmatrix},$$

where C_{kk}^{11} is the $p\times p$ asymptotic covariance matrix of $n^{1/2}\left(\hat{\beta}_k-\beta_{k0}\right)$ and C_{kk}^{22} is the asymptotic variance of $n^{1/2}(\hat{\sigma}_k-\sigma_{k0})$. Similarly define the estimators $\hat{C}_{kk}^{11}$, $\hat{C}_{kk}^{12}$, $\hat{C}_{kk}^{22}$.

In this case the estimates and confidence intervals for the median LOS and median cost, at a specified covariate profile Z_0 are

$$\hat{T}_{.5}^{*}=g^{-1}(\hat{\beta}_1'Z_0),$$
$$\hat{C}_{.5}^{*}=g^{-1}(\hat{\beta}_2'Z_0),$$

and $$\left[g^{-1}\left(\hat{\beta}_1'Z_0-z_\alpha\left(Z_0'(n^{-1}\hat{C}_{11}^{11})Z_0\right)^{1/2}\right),g^{-1}\left(\hat{\beta}_1'Z_0+z_\alpha\left(Z_0'(n^{-1}\hat{C}_{11}^{11})Z_0\right)^{1/2}\right)\right],$$

$$\left[g^{-1}\left(\hat{\beta}_2'Z_0-z_\alpha\left(Z_0'(n^{-1}\hat{C}_{22}^{11})Z_0\right)^{1/2}\right),g^{-1}\left(\hat{\beta}_2'Z_0+z_\alpha\left(Z_0'(n^{-1}\hat{C}_{22}^{11})Z_0\right)^{1/2}\right)\right],$$

respectively. Again, we considered g^{-1} nondecreasing.

In the following we will provide the details of the calculation of the exact form of the estimator $\hat{C}=\left(\hat{C}_{kl},k,l\in\{1,2\}\right)$ of the asymptotic covariance matrix of

$n^{1/2}\left(\hat{\theta}_1'-\theta_{10}',\hat{\theta}_2'-\theta_{20}'\right)'$. By Theorem 3, $\hat{C}_{kl}=n^2\mathcal{P}_{\tau_k}^{-1}(\hat{\theta}_k)\hat{B}_{kl}\mathcal{P}_{\tau_l}^{-1}(\hat{\theta}_l)$, where $-\mathcal{P}_{\tau_k}^{-1}(\theta_k)$ is the matrix of the second order partial derivatives of the log-partial likelihood and

$$\hat{B}_{kl}=n^{-1}\sum_{i=1}^{n}\hat{V}_{ki}(\hat{\theta}_k)\hat{V}_{li}(\hat{\theta}_l)',$$

with $\hat{V}_{ki}(\hat{\theta}_k)$ is a (p+1)-dimensional vector,

$$\hat{V}_{ki}(\hat{\theta}_k) = \int_0^{\tau_k} \frac{\partial}{\partial \theta_k} \log \alpha_{ki}(u, \hat{\theta}_k) d\tilde{M}_{ki}(u),$$

$$\tilde{M}_{ki}(u) = N_{ki}(u) - \int_0^u Y_{ki}(s) \alpha_{ki}(s, \hat{\theta}_k) ds.$$

Because PROC LIFEREG in SAS provides the values of $\mathcal{P}_{\tau_k}^{-1}(\hat{\theta}_k)$, we compute only $\hat{B}_{kl}$. We again adopt the notation $\frac{\partial}{\partial \theta_{kj}} g(a)$ for $\frac{\partial}{\partial \theta_{kj}} g(\theta_k)\big|_{\theta_k = a}$. The vector $\hat{V}_{ki}(\hat{\theta}_k)$ has the components:

for $j \in \{1, \ldots, p\}$:

$$\hat{V}_{ki}^{j}(\hat{\theta}_k) = \delta_{ki}[0 \le X_{ki} \le \tau_k] \frac{\partial}{\partial \beta_{kj}} \log\left(\hat{\sigma}_k^{-1} \alpha_0 \left(\frac{X_{ki} - \hat{\beta}_k' Z_{ki}}{\hat{\sigma}_k} \right)\right) -$$

$$- \int_0^{\tau_k} \frac{\partial}{\partial \beta_{kj}} \left(\hat{\sigma}_k^{-1} \alpha_0 \left(\frac{u - \hat{\beta}_k' Z_{ki}}{\hat{\sigma}_k} \right)\right) Y_{ki}(u) du;$$

for $j = p + 1$:

$$\hat{V}_{ki}^{p+1}(\hat{\theta}_k) = \delta_{ki}[0 \le X_{ki} \le \tau_k] \frac{\partial}{\partial \sigma_k} \log\left(\hat{\sigma}_k^{-1} \alpha_0 \left(\frac{X_{ki} - \hat{\beta}_k' Z_{ki}}{\hat{\sigma}_k} \right)\right) -$$

$$- \int_0^{\tau_k} \frac{\partial}{\partial \sigma_k} \left(\hat{\sigma}_k^{-1} \alpha_0 \left(\frac{u - \hat{\beta}_k' Z_{ki}}{\hat{\sigma}_k} \right)\right) Y_{ki}(u) du.$$

In the next calculations we will assume that in the study $0 \le X_{ki} \le \tau_k$ and we will use the notation $\hat{S}_{ki} = \frac{X_{ki} - \hat{\beta}_k' Z_{ki}}{\hat{\sigma}_k}$.

Then, for $j \in \{1, \ldots, p\}$:

$$\hat{V}_{ki}^{j}(\hat{\theta}_k) = -\delta_{ki}\frac{\alpha_0'}{\alpha_0}(\hat{S}_{ki})\hat{\sigma}_k^{-1}Z_{kij} + \int_0^{X_{ki}}\hat{\sigma}_k^{-2}\alpha_0'\left(\frac{u-\hat{\beta}_k'Z_{ki}}{\hat{\sigma}_k}\right)Z_{kij}du =$$

$$= \hat{\sigma}_k^{-1}\left(\int_{-\frac{\hat{\beta}_k'Z_{ki}}{\hat{\sigma}_k}}^{\hat{S}_{ki}}\alpha_0'(v)dv - \delta_{ki}\frac{\alpha_0'}{\alpha_0}(\hat{S}_{ki})\right)Z_{kij}.$$

For the standard normal distribution $\frac{\alpha_0'}{\alpha_0}(x) = \alpha_0(x) - x$. As a result

$$\hat{V}_{ki}^{j}(\hat{\theta}_k) = \hat{\sigma}_k^{-1}\left[\alpha_0\left(\hat{S}_{ki}\right) - \alpha_0\left(-\frac{\hat{\beta}_k'Z_{ki}}{\hat{\sigma}_k}\right) - \delta_{ki}\alpha_0\left(\hat{S}_{ki}\right) + \delta_{ki}\hat{S}_{ki}\right]Z_{kij} =$$

$$= \hat{\sigma}_k^{-1}\left[(1-\delta_{ki})\alpha_0\left(\hat{S}_{ki}\right) - \alpha_0\left(-\frac{\hat{\beta}_k'Z_{ki}}{\hat{\sigma}_k}\right) + \delta_{ki}\hat{S}_{ki}\right]Z_{kij}. \tag{1.29}$$

For $j = p+1$:

$$\hat{V}_{ki}^{p+1}(\hat{\theta}_k) = -\delta_{ki}\hat{\sigma}_k^{-1} - \delta_{ki}\frac{\alpha_0'}{\alpha_0}(\hat{S}_{ki})\hat{\sigma}_k^{-1}\hat{S}_{ki} +$$

$$+ \int_0^{X_{ki}}\left[\hat{\sigma}_k^{-2}\alpha_0\left(\frac{u-\hat{\beta}_k'Z_{ki}}{\hat{\sigma}_k}\right) + \hat{\sigma}_k^{-2}\alpha_0'\left(\frac{u-\hat{\beta}_k'Z_{ki}}{\hat{\sigma}_k}\right)\left(\frac{u-\hat{\beta}_k'Z_{ki}}{\hat{\sigma}_k}\right)\right]du =$$

$$= \hat{\sigma}_k^{-1}\left[\int_{-\frac{\hat{\beta}_k'Z_{ki}}{\hat{\sigma}_k}}^{\hat{S}_{ki}}\alpha_0(v)dv + \int_{-\frac{\hat{\beta}_k'Z_{ki}}{\hat{\sigma}_k}}^{\hat{S}_{ki}}\alpha_0'(v)vdv - \delta_{ki} - \delta_{ki}\left(\alpha_0\left(\hat{S}_{ki}\right) - \hat{S}_{ki}\right)\hat{S}_{ki}\right].$$

By assumption B.4 and integration by parts:

$$\int_a^b \alpha_0(v)dv + \int_a^b \alpha_0'(v)vdv = \alpha_0(b)b - \alpha_0(a)a.$$

Therefore

$$\hat{V}_{ki}^{p+1}(\hat{\theta}_k) = \hat{\sigma}_k^{-1}\left[\alpha_0\left(\hat{S}_{ki}\right)\hat{S}_{ki} + \alpha_0\left(-\frac{\hat{\beta}_k'Z_{ki}}{\hat{\sigma}_k}\right)\frac{\hat{\beta}_k'Z_{ki}}{\hat{\sigma}_k} - \delta_{ki} - \delta_{ki}\alpha_0\left(\hat{S}_{ki}\right)\hat{S}_{ki} + \delta_{ki}\hat{S}_{ki}^2\right] =$$

$$= \hat{\sigma}_k^{-1}\left[(1-\delta_{ki})\alpha_0\left(\hat{S}_{ki}\right)\hat{S}_{ki} + \alpha_0\left(-\frac{\hat{\beta}_k'Z_{ki}}{\hat{\sigma}_k}\right)\frac{\hat{\beta}_k'Z_{ki}}{\hat{\sigma}_k} - \delta_{ki} + \delta_{ki}\hat{S}_{ki}^2\right]. \tag{1.30}$$

Let

$$\hat{a}_{ki} = (1-\delta_{ki})\alpha_0\left(\hat{S}_{ki}\right) - \alpha_0\left(-\frac{\hat{\beta}_k' Z_{ki}}{\hat{\sigma}_k}\right) + \delta_{ki}\hat{S}_{ki},$$

$$\hat{b}_{ki} = (1-\delta_{ki})\alpha_0\left(\hat{S}_{ki}\right)\hat{S}_{ki} + \alpha_0\left(-\frac{\hat{\beta}_k' Z_{ki}}{\hat{\sigma}_k}\right)\frac{\hat{\beta}_k' Z_{ki}}{\hat{\sigma}_k} - \delta_{ki} + \delta_{ki}\hat{S}_{ki}^2.$$

By (1.29) and (1.30), the (p+1)-dimensional vector $\hat{V}_{ki}(\hat{\theta}_k)$ has the form

$$\hat{V}_{ki}(\hat{\theta}_k) = \hat{\sigma}_k^{-1}\begin{pmatrix} \hat{a}_{ki} Z_{ki} \\ \hat{b}_{ki} \end{pmatrix}.$$

Consequently

$$\hat{B}_{kl} = n^{-1}\sum_{i=1}^{n} \hat{V}_{ki}(\hat{\theta}_k)\hat{V}_{li}(\hat{\theta}_l)'$$

$$= \hat{\sigma}_k^{-1}\hat{\sigma}_l^{-1}\begin{pmatrix} n^{-1}\sum_{i=1}^{n}\hat{a}_{ki}\hat{a}_{li}Z_{ki}Z_{li}' & n^{-1}\sum_{i=1}^{n}\hat{a}_{ki}\hat{b}_{li}Z_{ki} \\ n^{-1}\sum_{i=1}^{n}\hat{a}_{li}\hat{b}_{ki}Z_{li}' & n^{-1}\sum_{i=1}^{n}\hat{b}_{ki}\hat{b}_{li} \end{pmatrix}.$$

1.3 Application

We demonstrate the application of the methods proposed in Sections 1.1 and 1.2 to hospital LOS and costs in a cohort of patients admitted for coronary artery bypass surgery (CABG). Data are from a tertiary care academically affiliated medical center. In this study there were 1268 consecutive admissions for CABG from August 23, 1993 to December 29, 1994. Computerized files were maintained on demographic characteristics, medical history, clinical outcomes, resource use and costs.

Length of stay was defined as the number of days from admission to discharge, inclusive of the admission day. Costs were derived from services for operating room, nursing, laboratory, and pharmacy as well as room and board and convenience items. All professional fees were excluded.

The complete dataset was available, including cost histories for each patient. As an example for our techniques that incorporate censoring and assume that only total costs per patient were observed, we reconstructed the dataset six months after the study started and we computed the total costs. At that time some patients were still in hospital and some costs were not as yet incurred. This resulted in a dataset of 465 subjects, with 7.53% of the LOS and 9.68% of the hospital costs being censored, respectively. Our objective is to examine in our sample the health care utilization, as measured by LOS and costs. Using both approaches (semiparametric and parametric), the medians of these two outcomes will be estimated and confidence intervals will be provided.

CORRELATES OF LOS AND COST

Potential correlates of LOS and costs included demographic and clinical variables that could be identified at admission, the use of cardiac catheterization during hospital stay and discharge status, an indicator of whether the patient was alive at discharge. The variables available at admission were age at admission, gender, race (White, African-American or Other), marital status (Married, Alone or Unknown), insurance status, comorbidity, ejection fraction and history of prior CABG. Insurance status was categorized as Medicare, private, Medicaid or other. Ejection fraction, a useful measure of cardiac function, is the volume of blood expelled at each systole as a fraction of the volume of blood contained in the ventricle at the end of the diastole. A value less than

50% is generally considered abnormal. Values of ejection fraction were grouped as below 35%, 35% to 49% and 50% and above.

Charlson Comorbidity Index (CCI)[46] was used to assess comorbidity. It is a weighted sum of the presence of 19 specified medical conditions at admission. These conditions include diabetes, liver disease, congestive heart failure, peripheral vascular disease, prior myocardial infarction, cerebrovascular disease, connective tissue disease, dementia, chronic obstructive pulmonary disease, hemiplegia, tumor, and acquired immunodeficiency syndrome (AIDS) or AIDS related complex. We formed three comorbidity groups based on CCI scores 0-1, 2-3 and 4 or more.

Diagnosis Related Group variable (DRG) is in this case a binary variable that indicates if a patient has cardiac catheterization during the hospital stay. Table 1.1 shows the characteristics of the 465 patients in our sample. The mean age was 63.4 years and the median age 65 years.

RESULTS

There is a high correlation between LOS and cost (Spearman $r = .79$, $n = 465$) even if all censored observations are omitted (Spearman $r =.77$, $n = 420$). The range of the uncensored LOS observations was 3 to 35 days, and the uncensored costs ranged from $11,669 to $90,088. Figure 1.1 shows a plot of LOS and cost (circles represent censored cost observations).

The bivariate models we use for LOS and cost allow inferences about regression parameters simultaneously for these two outcomes. For example, suppose we are interested in the effects of a covariate with three categories, such as CCI, on both LOS and cost. Two dummy variables are created for these three categories. Denote by

$$\eta_0 = \left((\beta_{10})_1, (\beta_{10})_2, (\beta_{20})_1, (\beta_{20})_2\right)'$$

the subvector of the vector of all true regression coefficients $\left(\beta_{10}', \beta_{20}'\right)'$ corresponding to the dummy variables for both LOS and cost, and by $\hat{\eta}$ the vector of estimators

$$\hat{\eta} = \left(\hat{\beta}_{11}, \hat{\beta}_{12}, \hat{\beta}_{21}, \hat{\beta}_{22}\right)'.$$

Let $\hat{\psi}$ be the consistent estimator of the asymptotic covariance matrix of $\hat{\eta}$, as defined in Theorems 3, Sections 1.1 and 1.2. By the asymptotic normality of the regression parameter estimators, the quadratic form $W = \hat{\eta}'\hat{\psi}^{-1}\hat{\eta}$ is asymptotically chi-square distributed with 4 degrees of freedom and can be used to test jointly the null hypotheses:

$$H_{kj} : \beta_{k0j} = 0, k \in \{1,2\}, j \in \{1,2\}.$$

In the covariate selection procedure, correlates of LOS and cost were tested jointly so that the resulting model would have the same constellation of significant variables. Each potential covariate was assessed individually and then in combination with all others that were found to be significant by univariate analysis (*p-value*< 0.20). Only age at admission was regarded as a continuous independent variable. In the final regression model we retained only variables that were significant at *p-value*< 0.10. All analyses were performed with SAS software version 8 (SAS Institute Inc., Cary NC).

Semiparametric Model

To implement the method developed in Section 1.1, we create two records for each patient, one for LOS and one for cost. For the categorical covariates sixteen dummy variables d_i, $2 \le i \le 16$ were created and type-specific covariates were defined as follows:

For LOS:

$Z_{11} = \text{age} * [\text{type} = 1]$ and $Z_{i1} = d_i * [\text{type} = 1], 2 \le i \le 16$;

For cost:

$Z_{12} = \text{age} * [\text{type} = 0]$ and $Z_{i2} = d_i * [\text{type} = 0], 2 \le i \le 16$.

After the covariate selection procedure, the significant correlates in the final model were age at admission (p=.0958), DRG (p<.0001), indicator of being discharged alive (p<.0001), history of prior CABG (p=.0006), ejection fraction (three categories, p=.0246) and Charlson Comorbidity Index (three categories p<.0001). Age was regarded as a continuous variable and seven dummy variables d_i, $2 \le i \le 8$ correspond to the significant categorical covariates.

Consider the vectors $Z_k = (Z_{1k}, Z_{2k}, \ldots, Z_{8k})'$, $\beta_k = (\beta_{1k}, \beta_{2k}, \ldots, \beta_{8k})'$, $1 \le k \le 2$ and $Z = (Z_1', Z_2')'$, $\beta = (\beta_1', \beta_2')'$. Note that for the LOS data: $\beta' Z = \beta_1' Z_1$ and for the cost data: $\beta' Z = \beta_2' Z_2$. We are essentially fitting two separate Cox models to LOS and cost, but this formulation is amenable to the SAS PHREG procedure and permits simultaneous estimation of β_1 and β_2 as well as direct estimation of the correlation between the two estimators.

We present a part of the final model estimates. For the three CCI categories, the two (out of seven) dummy variables d_2, d_3 were defined as following:

for CCI 0-1: $d_2 = 1, d_3 = 0$;

for CCI 2-3: $d_2 = 0, d_3 = 1$;

for CCI 4 or more $d_2 = 0, d_3 = 0$.

The LOS observations have $Z_{21} = d_2, Z_{31} = d_3, Z_{22} = 0, Z_{32} = 0$ and the cost observations have $Z_{21} = 0, Z_{31} = 0, Z_{22} = d_2, Z_{32} = d_3$. Estimates of the corresponding regression parameters $\beta_{21}, \beta_{31}, \beta_{22}, \beta_{32}$ are

$$\hat{\beta}_{21} = .683, \hat{\beta}_{31} = .230, \hat{\beta}_{22} = .766, \hat{\beta}_{32} = .140.$$

The estimated adjusted covariance matrix of these four beta estimators is

$$\begin{pmatrix} 0.0155 & 0.0105 & 0.0140 & 0.0101 \\ & 0.0165 & 0.0102 & 0.0163 \\ & & 0.0219 & 0.0150 \\ & & & 0.0230 \end{pmatrix}.$$

When fitting independent proportional hazard models for LOS and cost, the estimate for the so-called naïve covariance matrix is

$$\begin{pmatrix} 0.0224 & 0.0148 & 0 & 0 \\ & 0.0200 & 0 & 0 \\ & & 0.0229 & 0.0151 \\ & & & 0.0206 \end{pmatrix}.$$

Estimates and approximate 95% pointwise confidence intervals for the LOS and cost survival distribution functions were calculated for the 6 covariate profiles defined by CCI and discharge status (alive or dead), for a patient age 65 years at admission, with ejection fraction above 50, who had catheterization and with no history of prior CABG. As presented in Section 1.1.4, confidence intervals were obtained using the log(-log) transformation and both the adjusted and naïve variances were used. Figures 1.2 and 1.3 depict the LOS and cost survival function estimates and approximate confidence intervals for one of these profiles (for CCI = 4+, discharged alive). The naïve and adjusted confidence intervals are different, but close.

Median LOS and cost were estimated from the corresponding survival distributions (see Section 1.1.4). Table 1.2 shows these estimates for the 6 covariate profiles, previously presented. The adjusted and naïve confidence intervals for the median LOS are the same for patients discharged alive, but they differ substantially for patients who died during their hospital stay. For the median cost the two types of confidence intervals were different for all profiles, possibly due to more variation in the cost data. For the survivor profiles the naïve confidence intervals wholly contained the corresponding adjusted confidence intervals but this pattern changed completely for the non-survivor profiles. In our sample of 465 subjects, 12 patients did not survive their hospital stay. The small number of deaths and the large variability of the outcomes in these 12 patients might be a reason for the instability of our estimates for the non-survivor profiles.

In Table 1.2 we notice how the LOS and cost median estimates increase with larger comorbidity. Patient who survived their hospital stay had larger LOS and smaller costs than those who died.

Parametric Model

Both LOS and cost exhibit right skewness but this is not severe. No simple transformation could eliminate it. In our analyses outcomes are in their original scale and we assume bivariate normality.

The significant correlates in the final joint model were DRG (p<.0001), indicator of being discharged alive (p<.0001), history of prior CABG (p=.0282), ejection fraction (p=.0032) and Charlson Comorbidity Index (p<.0001). This parametric model has the same set of covariates as the semiparametric one, except age, which is not significant in this case. Again, these assessments were made jointly for both LOS and cost.

Following the calculations of the Section 1.2.3, we computed the adjusted estimator of the asymptotic covariance matrix of the regression parameters and the point estimates and confidence intervals for the medians of both outcomes. Since no transformation was used, g was the identity function. Estimates and confidence intervals of the median LOS and median cost by comorbidity and discharge status are shown in Table 1.3 for patients with ejection fraction above 50, no history of prior CABG, who had cardiac catheterization. The estimates are larger than in the semiparametric case, but similar patterns are noticed. The confidence intervals are wider for non-survivors than for survivors and non-survivors have lower LOS and larger costs than survivors. As expected, LOS and cost increase with comorbidity.

DISCUSSION

We applied two models to estimate the median LOS and the median cost for patients hospitalized for CABG. Each model permits adjustments for correlates and recognizes the correlation between the dependent variables. The semiparametric approach specifies marginal models for LOS and cost and yields consistent estimates of the regression parameters as long as the marginal models are correctly specified. The adjusted covariance matrix then accounts for the correlation between the outcomes, without explicitly specifying a joint distribution for them. The parametric approach specifies a joint distribution for LOS and cost but the parameters related only to the joint distribution and not to the marginal models are considered nuisance parameters and are left unspecified. In our study, using the adjusted instead of the naïve covariance estimates that do not address the correlation between LOS and cost, gave qualitatively different results with respect to the confidence intervals for the median cost.

The final sets of covariates in both models were essentially the same: comorbidity, discharge status (alive or dead), DRG, history of prior CABG, ejection fraction and age at admission, that was significant only in the semiparametric model. Previous studies of LOS and hospital cost have shown the importance of these covariates and their statistical significance regardless of the model used.[8, 22, 47]

A problem with LOS and in-hospital cost data is the appropriate treatment of in-hospital deaths. Previously, several authors[7, 8] have regarded deaths as an early curtailment of LOS and costs. In these studies observed cost and LOS of non-survivors is considered right censored because if they had survived their hospital stay, their costs would have been higher and LOS longer. In this approach the independent censoring assumption is not verified. Besides this, for many applications estimates of costs for those who died are just as important as for survivors. By censoring at death no model can be used to derive predictions for a decision model or cost-effectiveness analysis in which death is an explicit outcome. If only the observations of those who died are used in the analysis, the investigator generally sacrifices considerable efficiency, since the majority of the total sample typically survives. We regarded an in-hospital death along with other demographic and clinical characteristics as potential correlates of LOS and total hospital costs. Our finding was that the non-survivors had lower LOS and larger costs than survivors. Treating non-survivor costs as censored would have increased the bias of the cost estimators.

We have several limitations in the analyses of our application. One is the strong distributional assumption needed in the parametric model. We made a normality assumption and even though the resulting model did not provide a very good fit for our data, we used it as a comparison for the semiparametric model. When a distributional assumption is plausible for a study, the parametric models have the advantage of the simplicity of calculations and the efficiency of the estimators. Another limitation is the problem of censoring for costs. The use of survival analysis techniques to analyze medical care costs is relatively new and has sparked a lively debate[1, 9, 11, 48] on its applicability, given the assumptions that underlie traditional survival models for duration times. Patients who accumulate costs over time at relatively higher rates tend to generate larger cumulative costs at both the event time and the censoring time, leading to dependent censoring. This contradicts the usual assumption of independent censoring made in standard survival analyses. While no single approach can be expected to perform in all situations, we believe the traditional survival methods will still have a useful role in the cost analyses, especially of cost histories are not available.

TABLE 1.1: Characteristics of Patients

Variable	Subgroup	N	Percent
Discharged Alive	Yes	453	97.41
Gender	Male	314	67.53
Race	White	398	85.59
	African Amer.	36	7.74
	Other	22	4.73
	Unknown	9	1.94
Marital Status	Married	333	71.61
	Alone	112	24.09
	Unknown	20	4.30
DRG	with CATH	235	50.54
Ejection Fraction	< 35	58	12.47
	35 - 49	126	27.10
	50 +	281	60.43
History of prior CABG	Yes	36	7.74
Charlson Comorbidity Index	0-1	198	42.58
	2-3	189	40.65
	4 +	78	16.77
Insurance	Medicare	275	59.14
	Private	133	28.60
	Medicaid	30	6.45
	Other	27	5.81

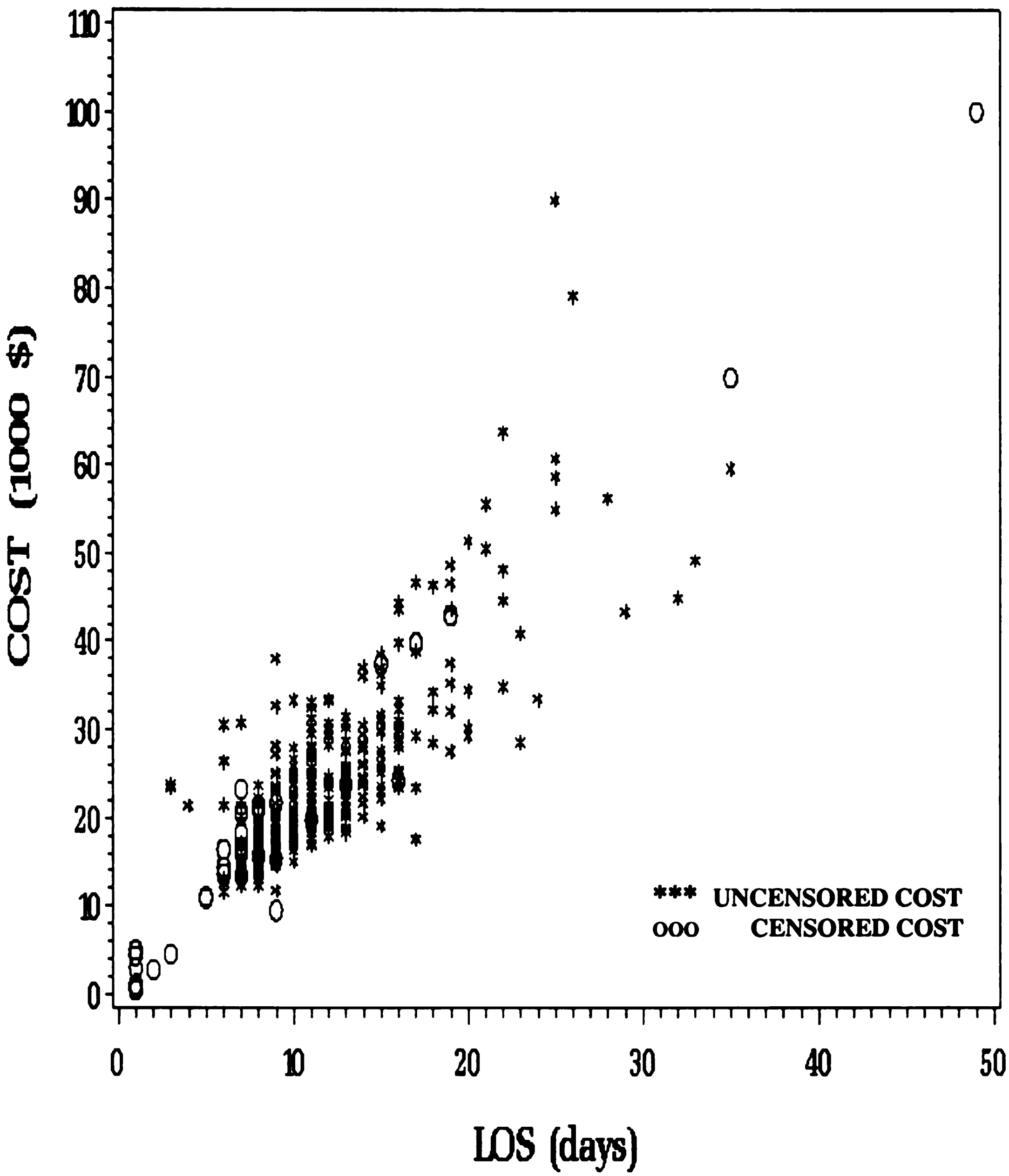


FIGURE 1.1: Distribution of Costs and LOS

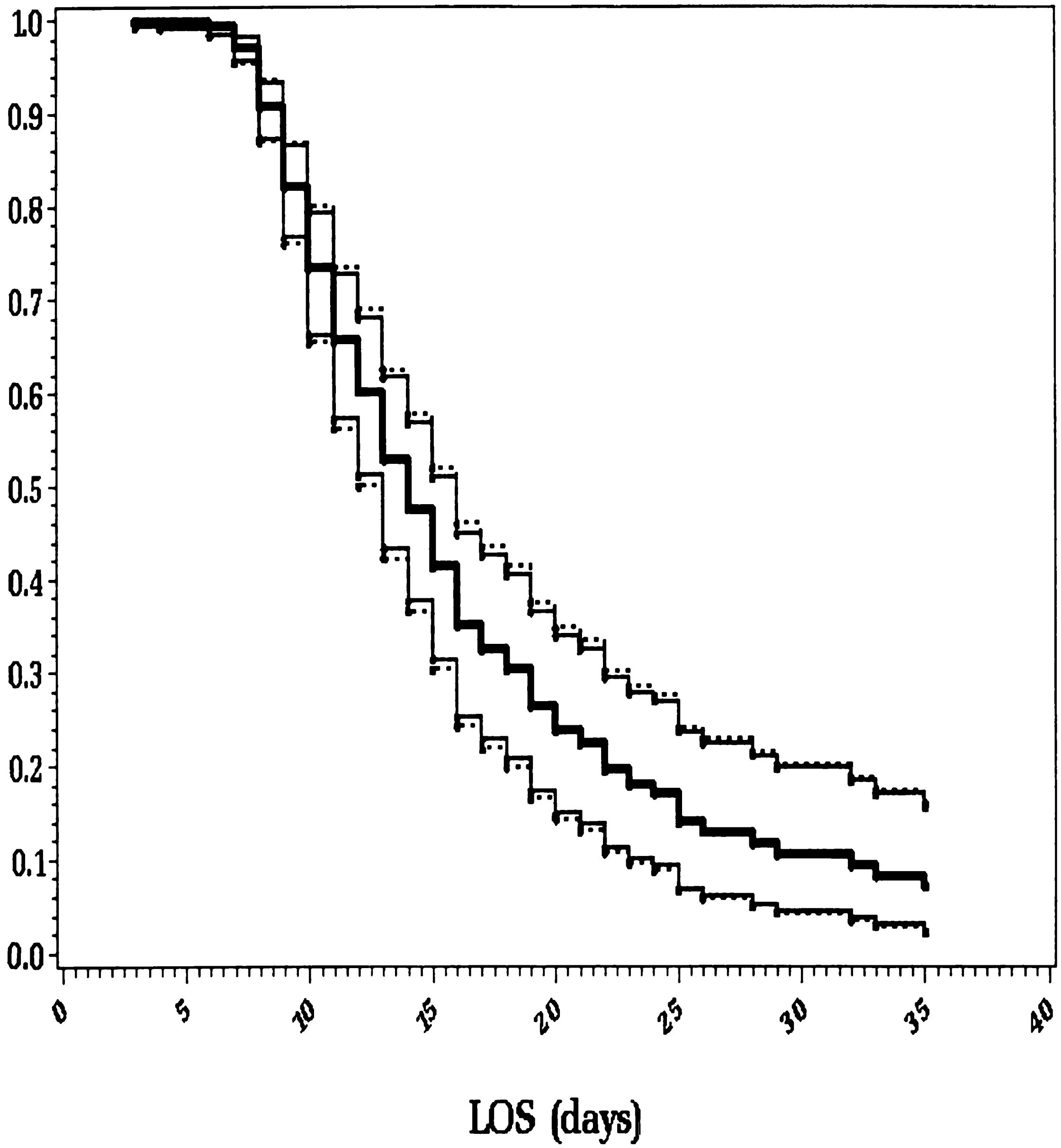


FIGURE 1.2: Estimated LOS survival function and approximate 95% adjusted (——) and naïve (- - - -) pointwise confidence intervals. Estimates were made for a patient discharged alive, who underwent CATH, with a CCI of 4+, ejection fraction 50+, age 65 at admission and no history of prior CABG.

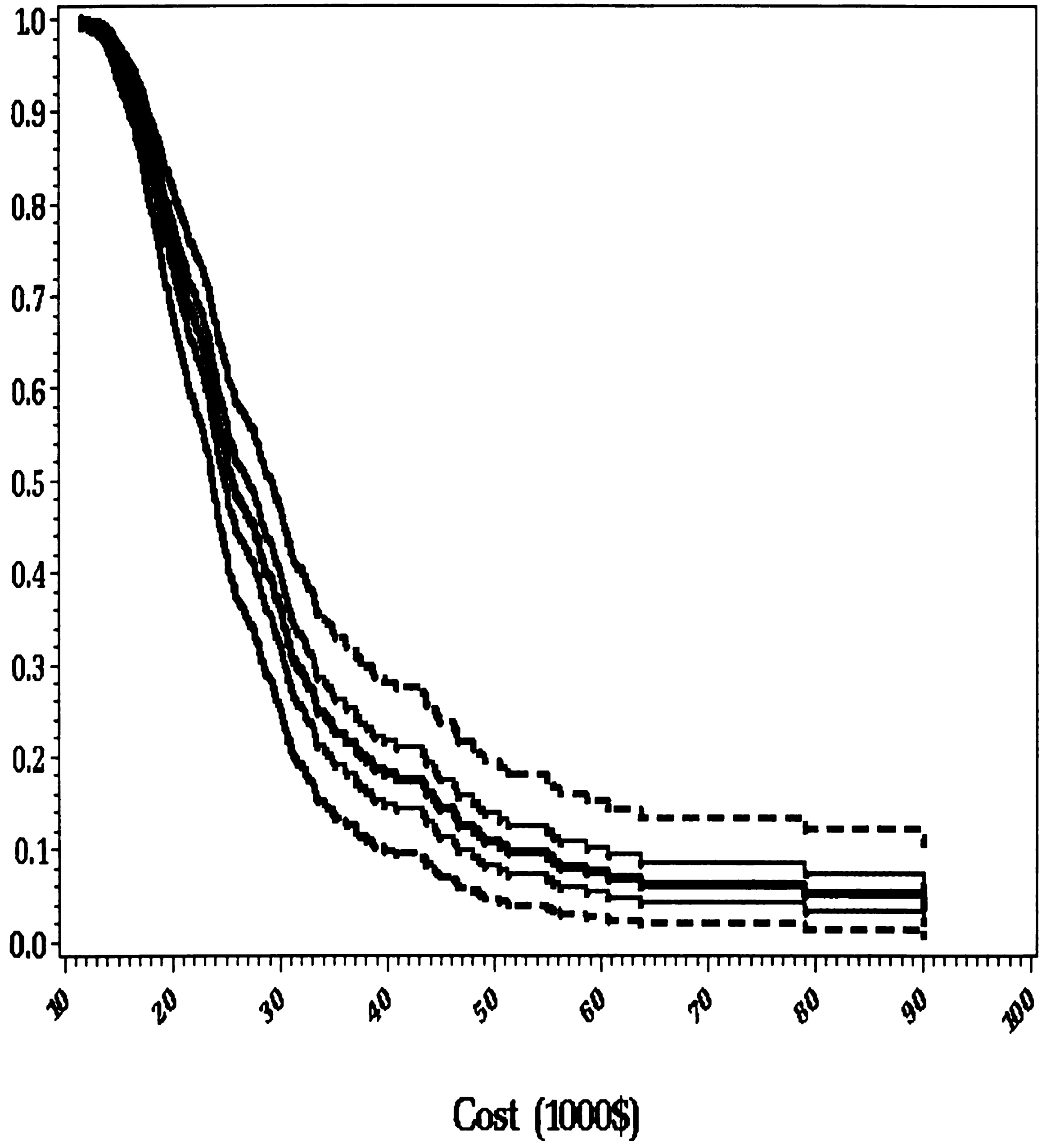


FIGURE 1.3: Estimated cost survival function and approximate 95% adjusted (——) and naïve (- - - -) pointwise confidence intervals. Estimates were made for a patient discharged alive, who underwent CATH, with a CCI of 4+, ejection fraction 50+, age 65 at admission and no history of prior CABG.

TABLE 1.2: Length of Stay and Costs by Comorbidity and Discharge Status (Semiparametric Model)

		Length of Stay, days			Cost, $		
CCI	Status	Median	Adjusted (naïve) 95% LCL	Adjusted (naïve) 95% UCL	Median	Adjusted (naïve) 95% LCL	Adjusted (naïve) 95% UCL
0-1	Alive	11	10 (10)	11 (11)	20,819	20,378 (20,356)	21,315 (21,382)
	Dead	10	8 (9)	15 (12)	24,167	18,922 (20,682)	34,884 (29,422)
2-3	Alive	13	12 (12)	14 (14)	24,666	24,140 (24,116)	25,149 (25,192)
	Dead	12	9 (10)	19 (15)	30,155	21,415 (24,201)	79,103 (44,637)
4 +	Alive	14	13 (13)	16 (16)	25,660	24,913 (24,728)	27,083 (27,177)
	Dead	13	9 (10)	22 (16)	31,276	23,048 (25,399)	79,103 (46,662)

Estimates for a patient who underwent CATH, with ejection fraction 50+, age 65 at admission and no history of prior CABG.

TABLE 1.3: Length of Stay and Costs by Comorbidity and Discharge Status (Parametric Model)

		Length of Stay, days			Cost, $		
CCI	Status	Median	Adjusted (naïve) 95% LCL	Adjusted (naïve) 95% UCL	Median	Adjusted (naïve) 95% LCL	Adjusted (naïve) 95% UCL
0-1	Alive	11.54	10.98 (10.73)	12.10 (13.35)	23,237	22,094 (21,493)	24,380 (24,981)
	Dead	8.59	4.27 (5.92)	12.91 (11.27)	29,705	17,778 (23,945)	41,632 (35,466)
2-3	Alive	13.47	12.47 (12.62)	14.47 (14.32)	27,841	25,790 (26,001)	29,892 (29,681)
	Dead	10.52	6.13 (7.89)	14.92 (13.16)	34,309	22,380 (28,636)	46,238 (39,982)
4 +	Alive	14.87	13.51 (13.66)	16.23 (16.07)	29,893	27,058 (27,287)	32,728 (32,499)
	Dead	11.92	7.35 (9.20)	16.49 (14.64)	36,361	23,278 (30,497)	49,445 (42,226)

Estimates for a patient who underwent CATH, with ejection fraction 50+ and no history of prior CABG.

CHAPTER 2

ESTIMATING HOSPITAL COST OVER A SPECIFIED DURATION

Considering accumulating cost as a process evolving over time, we construct in this chapter a regression model that permits the estimation of the mean cost over a specified duration of hospital stay and also adjusts for the influence of patient characteristics on LOS and cost. The proposed methods that model the relationship between hospital cost and LOS are applied to assess the mean cost in a cohort of hospitalized patients who underwent CABG surgery.

2.1 Model Description

Let T denote the LOS and Z_1 a vector of p explanatory variables that might have an impact on the distribution of T. The cumulative cost $C(t)$ through t days of hospital stay is only observed at $t = T$ and therefore we will focus on the influence of covariates Z_2 on the distribution of the total cost $C = C(T)$.

To fully integrate the role of time into analyses of costs we would need the cumulative cost histories as they manifest in each patient. Suppose that $C(t) = \int_0^t B(u)du$, so that cost is incurred at the rate $B(u)$ at time u. The rate of hospital cost accumulation is observed at time u only if u is less or equal than the hospital stay T. Therefore the observed cost has the form

$$\int_0^t [T \geq u] B(u) du \ .$$

We want to estimate the expected value of the observed cost over a duration t for a specified covariate profile Z_0. Given this covariate profile, we assume that T is independent of the rate process $\{B(u), u > 0\}$. Then, by Fubini's Theorem, the mean observed cost is

$$MC(t \mid Z_0) = \int_0^t b(u \mid Z_0) S(u \mid Z_0) du \ , \tag{2.1}$$

where $b(u \mid Z_0) = E(B(u) \mid Z_0)$ is the average potential rate at time u and $S(u \mid Z_0) = P(T > u \mid Z_0)$ the survival function of T, both for the specified profile Z_0.

Consider n individuals in the study. Using the same notations of Chapter 1, for the i-th patient let T_i denote the true LOS, T_i' the censoring time, $X_{1i} = \min(T_i, T_i')$, $\delta_{1i} = [T_i \leq T_i']$ the indicator of non-censoring for time and Z_{1i} a vector of p explanatory variables that influence T_i. Consider also the true hospital cost C_i, the censored cost C_i', $X_{2i} = \min(C_i, C_i')$, $\delta_{2i} = [C_i \leq C_i']$ the indicator of non-censoring for cost and Z_{2i} a vector of p explanatory variables that influence C_i.

Let $\hat{b}(. \mid Z_0)$, $\hat{S}(. \mid Z_0)$ be estimators of $b(. \mid Z_0), S(. \mid Z_0)$, respectively, obtained from the data $\{(X_{1i}, \delta_{1i}, Z_{1i}), (X_{2i}, \delta_{2i}, Z_{2i}), 1 \leq i \leq n\}$. Then the mean observed cost is estimated by

$$\hat{MC}(t \mid Z_0) = \int_0^t \hat{b}(u \mid Z_0)\hat{S}(u \mid Z_0)du. \quad (2.2)$$

The construction of the estimators $\hat{b}(. \mid Z_0)$, $\hat{S}(. \mid Z_0)$ relies heavily on the study data. In the application described below, a linear relationship between time and cost is considered appropriate and a linear regression model is used for the estimation of the expected rate. The survival function is estimated through a proportional hazard model.

In Chapter 3 we extend the model proposed here to a more general setup in which patients pass through different health states. In the present chapter a patient makes a single transition from the state of being hospitalized to the state of being discharged. A rate model is used for describing the accumulating hospital costs. With transitions between several states, costs might be incurred at transition between health states and also during a sojourn in a health state. The latter costs can be described also through a rate model, whereas the "jump" costs at transition times between states can be described using marked point processes. For these cases we provide in Chapter 3 our methods of estimation of mean cost.

2.2 Application

We apply the presented method to the hospital costs of the patients from the same sample used in Section 1.3. In that section we described the study, the available potential correlates and the covariate selection procedure.

The proportional hazard model provided a good fit to the data. The significant correlates of LOS were age at admission (p=.0366), DRG (p<.0001), history of prior CABG (p=.0685), ejection fraction (p=.0650) and Charlson Comorbidity Index

(p<.0001). The estimator $\hat{S}(.\mid Z_0)$ of the LOS survival function was derived from the Cox regression model, as described in Section 1.1.

The plot of costs versus LOS (see Figure 1.1) suggests a linear relationship. The variance seems to change a little over time but for simplicity we do not consider this fact in the analyses of this example.

As in the parametric model used in Section 1.3, we assume costs approximately normally distributed. For modeling total costs as a function of time, we allow LOS and LOS^2 to compete for inclusion in the final model. We also regard the indicator of non-censoring for cost along with the other demographic and clinical characteristics as potential correlates of total hospital cost. The significant correlates are LOS (p<.0001), DRG (p=.0636), history of prior CABG (p=.0213) and indicator of being discharged alive (p<.0001). Even though the indicator of cost non-censoring is not significant (p=.6138), we include it in the final list of covariates because we want to be able to distinguish between the censored and the non-censored cost observations. We then consider the model

$$C_i(t) = \beta_1\delta_{2i} + \beta_2' Z_{2i} + \beta_3 t + \sigma\varepsilon_i,$$

in which the parameters $\beta_1, \beta_2, \beta_3, \sigma$ are estimated by maximum likelihood, assuming the errors ε_i independently normally distributed with zero mean and unit variance. Consequently, for a given profile Z_0, we estimate the expected (uncensored) cost

$$C_0(t\mid Z_0) = E(C(t)\mid Z_0) = E\left(\int_0^t B(u)du \mid Z_0\right) = \int_0^t b(u\mid Z_0)du$$

by $\quad \hat{C}_0(t\mid Z_0) = \hat{\beta}_1 + \hat{\beta}_2' Z_0 + \hat{\beta}_3 t$, for $t > 0$ and $\hat{C}_0(0\mid Z_0) = 0$. (2.3)

The above model for $C_i(t)$ incorporates the dynamics of time into the accumulating hospital cost. In other applications the dependence on time might be more complex than the simple linear relation used here. However, in practice a polynomial in t should adequately capture the dependence on the time t. In a related study of hospital charges in patients undergoing cardiac procedures, a quadratic in t provided reasonable fit to the data.[49]

Suppose we want to estimate the mean observed cost over the fixed duration t. The estimated survival function $\hat{S}(.\,|\,Z_0)$ is a step function that jumps at the observed LOS times. Let $T_{(l)}, l \geq 1$ denote the ordered observed LOS times in our sample. Then the fixed duration t is located between two of these times: $T_{(l-1)} < t \leq T_{(l)}$ and, by (2.2),

$$\hat{MC}(t\,|\,Z_0) = \sum_{j=1}^{l-1} \hat{S}(T_{(j-1)}\,|\,Z_0)\Big(\hat{C}_0(T_{(j)}\,|\,Z_0) - \hat{C}_0(T_{(j-1)}\,|\,Z_0)\Big) + \\ + \hat{S}(T_{(l-1)}\,|\,Z_0)\Big(\hat{C}_0(t\,|\,Z_0) - \hat{C}_0(T_{(l-1)}\,|\,Z_0)\Big),$$

where $\hat{C}_0(.\,|\,Z_0)$ is given by (2.3). Thus the mean observed cost is an average of costs increments, weighted by the likelihood of surviving through each incremental period.

Table 2.1 shows the estimated mean costs at the LOS median and at the largest observed LOS (35 days) by comorbidity and discharge status (alive or dead). The estimates are for survivors, age 65 at admission, ejection fraction 50+, with no prior history of CABG, who had a cardiac catheterization during their hospital stay. Medians of the LOS distribution for different covariate profiles were estimated from the Cox model. The discharge status was not significant in the model for LOS, so the LOS medians do not differ between the survivors and non-survivors with the same characteristics. As expected, the mean costs increase with comorbidity. We also notice that non-surv

the estimated

and $34,030 f

by category o

for Table 2.1.

noticed in the

the linear tren

In conc

stays of specif

survival analys

potential corre

way we do not

valid in many s

cost accumulat

absence of the

cost and LOS i

increments. Th

adopted in eval

In that study su

estimated by th

incomplete fol

which gave the

$\int_0 e^{-ru} S(u|Z_0)$

that non-survivors have larger costs than survivors. Thus for patients with CCI of 2 or 3, the estimated mean cost through their median LOS of 12 days was \$21,651 for survivors and \$34,030 for non-survivors. Figure 2.1 is a plot of the estimated mean observed cost by category of comorbidity for patients discharged alive, with the same profile described for Table 2.1. The fact that a model linear in time was used for the accumulating cost is noticed in the figure. After approximately 9 days, the survival function weights change the linear trend and make the estimated mean cost different by comorbidity.

In conclusion, the presented model estimates the mean observed cost for hospital stays of specified duration. The censored LOS observations were analyzed by standard survival analysis methods. The indicator of non-censoring for cost was regarded as a potential correlate of total cost, even though it was not statistically significant. In this way we do not use the assumption of independent cost censoring, which might not be valid in many situations. When assessing mean costs over a given duration, the potential cost accumulation $C(t)$ through time t was modeled by a linear function of time. In the absence of the cost histories of the patients, the model was fitted using the observed total cost and LOS in our sample. The overall mean cost was a weighted average of cost increments. This approach is similar to that Gardiner *et al.* (1995, 1999)[14, 15] previously adopted in evaluating the cost-effectiveness of the implantable cardioverter defibrillator. In that study survival time was the underlying stochastic variable whose distribution was estimated by the Kaplan-Meier method or Cox regression. Censoring occurred due to incomplete follow-up of some patients. The cumulative cost $C(t)$ was assumed known, which gave the expected total cost over a fixed time interval $[0,t]$ as

$\int_0^t e^{-ru} S(u \mid Z_0) dC(u)$, where r is the discount rate and S the survival function. Our

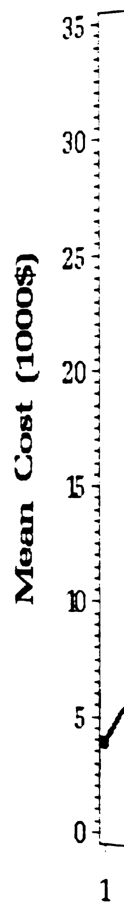
proposed me
the stochastic
short hospita
costs over [0
now interpret
underlying re
estimator is i
properties de

proposed method extends this analysis in a very important practical way by incorporating the stochastic element in costs. Ignoring discounting (which was irrelevant for relative short hospital stays considered in this study) we constructed an estimator of the mean costs over $[0,t]$ as $\int_0^t S(u \mid Z_0) dC_0(u \mid Z_0)$, where $C_0(t \mid Z_0) = E(C(t) \mid Z_0)$ and $C(t)$ is now interpreted as the potential cumulative cost up to time t. This estimator exploits the underlying relationship between total hospital cost and LOS. While consistency of the estimator is immediate, other distributional properties will follow as special cases of the properties developed in the more general set-up of Chapter 3.

TABLE 2.1: Estimates of mean cost at duration times by comorbidity and discharge status

CCI	Status	LOS Median (days)	Mean Cost ($) at LOS Median	Mean Cost ($) at Overall Follow-up
0-1	Alive	10	19,190	22,759
	Dead	10	31,569	35,138
2-3	Alive	12	21,651	26,697
	Dead	12	34,030	39,076
4 +	Alive	13	23,020	29,272
	Dead	13	35,399	41,651

Estimates for patients with age 65 at admission, ejection fraction 50+, no history of prior CABG, who underwent catheterization during their hospital stay.



FIGUR

(for surv

of prior (

stay)

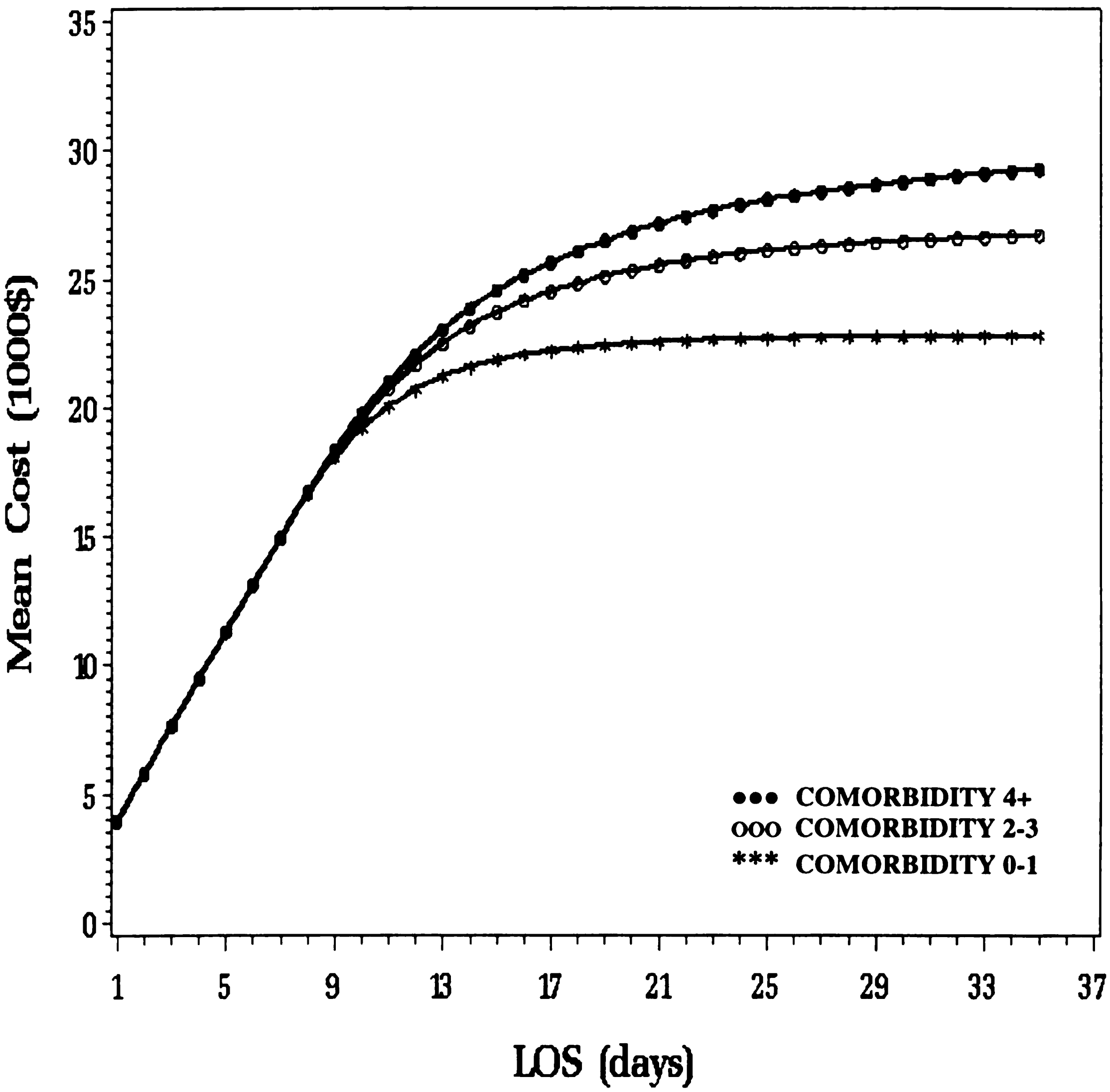


FIGURE 2.1: Estimated mean cost at duration times by comorbidity (for survivors with age 65 at admission, ejection fraction 50+, no history of prior CABG, who underwent catheterization during their hospital stay)

In st

periods of fo

example, in

count. A le

and 500, an

costs are in

will depend

variables o

Mu

have becon

disease pro

decision a

of patients

degree of

between t

random.

entails th

CHAPTER 3

ESTIMATING MEDICAL COSTS IN LONGITUDINAL STUDIES

In studies of the natural history of diseases and in medical interventions with long periods of follow-up, clinical conditions define several health conditions or states. For example, in the progression of HIV disease[50-52] stages can be defined by the CD4 cell count. A level below $200(\times 10^6 / L)$ triggers aggressive treatment and levels between 200 and 500, and above 500 have recognizable clinical interpretation. In these situations costs are incurred in each state and at transitions between states through resource use that will depend on treatment and patient attributes. Assessing the influence of these variables on costs is one of the objectives of this chapter.

Multistate Markov models which have their theoretical origin in survival models have become the standard for modeling health related outcomes, specially in studying disease progression in patients.[53-58] They are also framework for cost-effectiveness and decision analyses (see p152-153, Gold *et al.* (1996)[59]). By describing the event histories of patients as sojourns through different health states, Markov models provide a sufficient degree of flexibility to model the probabilistic mechanisms that underlie the transitions between these states. The sojourn times in the states and transitions between them are random. The Markov assumption restricts their dependency on the past information and entails the conditional independence of sojourn times given the states.

In Section 3.1 of this chapter we use a Markov model to describe the experience of patients in sustaining and changing states of health. Two types of costs are considered: costs incurred at transition between health states and costs of sojourns in a health state. Then present values are computed by discounting all costs at a fixed rate. In Section 3.2 we provide estimators of the mean present value of these two types of costs incurred over a fixed duration. The estimators are obtained conditional on an initial state and a given covariate profile. Large sample properties of these estimators are presented in Section 3.3.

3.1 Model Description

3.1.1 A Markov Model for Describing Patient Health Histories

Let $(\Omega, \mathcal{F}, P)$ a probability space and let $\{X(t),\ t \in \mathcal{T}\}$ with $\mathrm{T} = [0, \tau], \tau < \infty$, a non-homogeneous continuous time Markov process with finite state space $E = \{1, 2, \ldots, k\}$, having transition probabilities $P_{hj}(s,t)$ and transition intensities $\alpha_{hj}(t)$. This Markov process describes the evolution of one patient's health history, with $X(t)$ the patient health state occupied at time t. Typically E consists of several transient states, such as "well", "ill", "recovery", "relapse", and one or more absorbing states such as "disabled" or "dead".

Let $\alpha = (\alpha_{hj}), h, j \in \{1, 2, \ldots, k\}$ be the matrix of these transition intensities, where $\alpha_{hh} = -\sum_{j \neq h} \alpha_{hj}$. Thus, starting from the time of entry into state h, the sojourn times in the given state h are

process jumps ou

Let $A_{hj}(t$

the integrated in

called the negati

$A = (A_{hj}, h, j \in \{$

Let $P(s,$

$P(s,t) =$

is the $k \times k$ tran

The inte

registering each

Jacobsen (1982

Theore

Let A co

$\mathcal{N}_t = \sigma\{X(s):$

$Y_h(t) =$

$N_{hj}(t) =$

given state h are continuously distributed, with hazard rate function $-\alpha_{hh}$. Given that the process jumps out of state h at time t, it jumps into state $j \neq h$ with probability $\alpha_{hj} / -\alpha_{hh}$.

Let $A_{hj}(t) = \int_0^t \alpha_{hj}(s)ds$ and $A_{hh} = -\sum_{j \neq h} A_{hj}$. For $h \neq j$ the function A_{hj} is called the integrated intensity function for transitions from state h to state j, whereas A_{hh} is called the negative integrated intensity function for transitions out of state h. The matrix $A = (A_{hj}, h, j \in \{1, 2, ..., k\})$ is also called the intensity measure of the Markov process X.

Let $P(s,t) = \prod_{(s,t]} (I + dA)$ for $s < t, s, t \in \mathrm{T}$. The matrix

$$P(s,t) = (P_{hj}(s,t), h, j \in \{1, 2, ..., k\})$$

is the $k \times k$ transition matrix of the Markov process.

The intensity measure reappears in the compensator of the counting process registering each type of jump of the process. The next result was first proved by Jacobsen (1982)[60]:

Theorem (Theorem II.6.8, p94, Andersen *et al.* (1993)[29])

Let A correspond to the intensity measure of a Markov process X and $\mathcal{N}_t = \sigma\{X(s) : s \leq t\}$. Define

$$Y_h(t) = [X(t-) = h],$$

$$N_{hj}(t) = \#\{s \leq t : X(s-) = h, X(s) = j\}, h \neq j.$$

Then $N = (N_{hj}, h \neq j)$ is a multivariate counting process and its compensator with respect to $(\mathcal{F}_t) = \left(\sigma(X(0)) \vee \mathcal{N}_t\right)$ has components $\Lambda_{hj}(t) = \int_0^t Y_h(s) A_{hj}(ds)$.

Equivalently, the processes M_{hj} defined by $M_{hj} = N_{hj} - \int Y_h dA_{hj}$ are martingales.□

Here $Y_h(t)$ is the indicator that the process was in state h just prior to time t. In our absolutely continuous case with transition intensities α_{hj} and $A_{hj}(t) = \int_0^t \alpha_{hj}(s) ds$, we say that the multivariate counting process N has intensity $\lambda = \left(\lambda_{hj}, h \neq j\right)$, with $\lambda_{hj}(t) = Y_h(t)\alpha_{hj}(t)$.

Let Z_0 a given fixed vector of basic covariates. The p-dimensional vectors Z_{hj0} of type-specific covariates are computed from the vector Z_0, reflecting that some of these basic covariates may affect the different transition intensities differently. We assume that for an individual with given basic covariates Z_0 and corresponding type-specific covariates Z_{hj0} for $h \neq j$, the transition intensity $\alpha_{hj}(t \mid Z_0)$ in the Markov process has the form

$$\alpha_{hj}(t \mid Z_0) = \alpha_{hj0}(t) \exp(\beta_0' Z_{hj0}),$$

where β_0 is the true parameter value and α_{hj0} the intensity corresponding to $Z_{hj0} = 0$.

Let $A_{hj0}(t) = \int_0^t \alpha_{hjo}(s) ds$ be the integrated baseline intensity for transitions from state h to state j and $A_{hh0} = -\sum_{j \neq h} A_{hj0}$.

Given a random sample of patient histories, we describe in Section 3.2.1 how suitable estimates $\hat{\beta}$ and $\hat{A}_{hj0}(t, \hat{\beta})$ can be obtained for β_0 and $A_{hj0}(t)$.

3.1.2 Incorporating Costs in the Markov Model

As previously mentioned, we consider two types of costs that might be incurred in the course of follow-up: costs at transition between health states and costs of sojourns in a particular health state.

Suppose an amount $C_{hj}(t)$ is incurred just after time t if a transition h to j takes place at time t. The present value of expenditures in $(0,t]$ associated with these transitions is $C_{hj}^{(1)}(t) = \int_0^t e^{-rs} C_{hj}(s) dN_{hj}(s)$, where r is the discount rate. In economic studies expenditures to be incurred in the future are discounted to present value. A dollar spent now is worth more than a dollar that would be spent later. The discount rates used for the US have usually been between 3% and 5% per year, reflecting the rates on savings accounts or certificates of deposit.

Conditional on the initial state, given the vector Z_0 of basic covariates with the corresponding type-specific covariates Z_{hj0}, the mean of this present value is:

$$MPV_{hj}^{(1)}(t \mid i, Z_0) = E(C_{hj}^{(1)}(t) \mid X_0 = i, Z_0) = E(\int_0^t e^{-rs} C_{hj}(s) dN_{hj}(s) \mid X_0 = i, Z_0).$$

We assume that:

A.0.1 $C_{hj}(.)$ are bounded, non-negative real stochastic processes over $\mathcal{T}$, adapted to $(\mathcal{F}_t)$, left continuous with right hand limits (so $C_{hj}(.)$ are bounded, predictable processes).

A.0.2 $E(C_{hj}(t) \mid X_0 = i, X(t-) = h, Z_0) = E(C_{hj}(t) \mid X(t-) = h, Z_0)$ for $t > 0$, so the expected transition cost at any $t > 0$ does not depend on the initial health state.

It is known that if N is a counting process with intensity process λ, $M = N - \int \lambda$ and H is locally bounded and predictable, then M and $\int HdM$ are local square integrable martingales, with $E(M) = E\left(\int HdM\right) = 0$ (see Proposition II.4.1, p70, Andersen *et al.* (1993)[29]). Then, by assumption (A.0.1),

$$MPV_{hj}^{(1)}(t \mid i, Z_0) = E(\int_0^t e^{-rs} C_{hj}(s) \lambda_{hj}(s) ds \mid X_0 = i, Z_0) =$$

$$= E(\int_0^t e^{-rs} C_{hj}(s) Y_h(s) \alpha_{hj0}(s) \exp(\beta_0' Z_{hj0}) ds \mid X_0 = i, Z_0).$$

By Fubini's Theorem:

$$MPV_{hj}^{(1)}(t \mid i, Z_0) = \int_0^t e^{-rs} E(C_{hj}(s) Y_h(s) \mid X(0) = i, Z_0) \alpha_{hj0}(s) \exp(\beta_0' Z_{hj0}) ds.$$

We can write

$$E(C_{hj}(s) Y_h(s) \mid X_0 = i, Z_0) = E(C_{hj}(s) \mid X_0 = i, X(s-) = h, Z_0) P(X(s-) = h \mid X_0 = i, Z_0).$$

By the assumption (A.0.2), $MPV_{hj}^{(1)}(t \mid i, Z_0)$ has the form

$$MPV_{hj}^{(1)}(t \mid i, Z_0) = \int_0^t e^{-rs} c_{hj}(s \mid Z_0) P_{ih}(0, s \mid Z_0) \alpha_{hj0}(s) \exp(\beta_0' Z_{hj0}) ds, \tag{3.1}$$

where $c_{hj}(s \mid Z_0) = E(C_{hj}(s) \mid X(s-) = h, Z_0)$.

Estimation of the transition probabilities is described in Section 3.2.2 and the method of estimation of $c_{hj}(s \mid Z_0)$ is presented in Section 3.2.3.

We now turn to the cost of sojourns in a health state. Suppose that the cost in state h is incurred at the rate $B_h(u)$ at time u. The observed rate is zero at time u whenever, just before u, the patient is not in state h anymore, so $[X(u-)=h]=0$. Then the observed present value of all expenditures in state h, started at time s and ended after the duration time d is given by

$$C_h^{(2)}(s,d)=\int_s^{s+d} e^{-ru} B_h(u) Y_h(u)\,du,$$

where r is the discount rate and $Y_h(u)=[X(u-)=h]$.

Conditional on the initial state, given the vector Z_0 of basic covariates, the mean of this present value is

$$\begin{aligned} MPV_h^{(2)}(s,d \mid i,Z_0) &= E(C_h^{(2)}(s,d) \mid X_0=i,Z_0)= \\ &= \int_s^{s+d} e^{-ru} E(B_h(u)Y_h(u) \mid X_0=i,Z_0)\,du. \end{aligned}$$

Conditions similar to (A.0.1) and (A.0.2) are assumed for $B_h(.)$:

A.0.3 $B_h(.)$ are bounded, non-negative real stochastic processes over $[0,\tau]$, adapted to $(\mathcal{F}_t)$.

A.0.4 $E(B_h(u) \mid X_0=i, X(u-)=h, Z_0)=E(B_h(u) \mid X(u-)=h, Z_0)$ for all $u \in [0,\tau]$.

Denote $b_h(u \mid Z_0)=E(B_h(u) \mid X(u-)=h,Z_0)$. We can write

$$E(B_h(u)Y_h(u) \mid X_0=i,Z_0)=E(B_h(u) \mid X(u-)=h, X_0=i,Z_0)P(X(u-)=h \mid X_0=i,Z_0).$$

By assumption (A.0.4):

$$MPV_h^{(2)}(s,d\,|\,i,Z_0)=\int_s^{s+d}e^{-ru}b_h(u\,|\,Z_0)P_{ih}(0,u\,|\,Z_0)du\,. \tag{3.2}$$

The method of estimation of $b_h(.\,|\,Z_0)$ is presented in Section 3.2.4.

3.2 Estimation of the Mean Transition Cost and Mean Sojourn Cost

3.2.1 Estimation of the Regression Parameters and Integrated Baseline Intensities

Consider n individuals in the study. We assume that given the random vector $X_0=(X_{10},\ldots,X_{n0})$ and the random processes $Z(.)=(Z_1(.),\ldots,Z_n(.))$, independent Markov processes $X_1(.),\ldots,X_n(.)$ are constructed, with $X_i(0)=X_{i0}, 1\le i\le n$. Each process $X_i(.)$ has the same description as that of $X(.)$ previously presented in Section 3.1. For the i-th individual, $Z_i(t)$ is the p-dimensional vector of covariates measured at time t and X_{i0} the initial state.

A multivariate counting process $N_i=(N_{hji},h\neq j)$ is defined from $X_i(.)$:

$$N_{hji}(t)=\#\{0\le s\le t: X_i(s-)=h, X_i(s)=j\},\ h\neq j\,.$$

Let $\mathcal{F}^0=\sigma\{X_0\}$, $\mathcal{F}^Z(t)=\sigma\{Z(s),s\le t\}$, $\mathcal{N}_i(t)=\sigma\{N_{hji}(s),s\le t,h\neq j\}$ and $\mathcal{F}_i(t)=\mathcal{F}^0\vee\mathcal{F}^Z(t)\vee\mathcal{N}_i(t)$. By Theorem II.6.8, p94, Andersen *et al.* (1993)[29], $N_i=(N_{hji},h\neq j)$ is a multivariate counting process with its $\mathcal{F}_i(t)$- transition intensities

$\lambda_{hji}(t) = \alpha_{hji}(t)Y_{hi}(t)$, where $Y_{hi}(t) = [X_i(t-) = h]$ and α_{hji} is the transition intensity from state h to state j for the Markov process $X_i(.)$.

We assume the transition intensities α_{hji} have the form

$$\alpha_{hji}(t, \beta_0) = \alpha_{hj0}(t)\exp(\beta_0' Z_{hji}(t)),\ t \in \mathcal{T},$$

where the type-specific covariate vector $Z_{hji}(t)$ is computed from the vector $Z_i(t)$ of basic covariates for the i-th individual. This is the standard proportional hazard model.

In the following we present the construction and the large sample properties of the estimators $\hat{\beta}$ and $\hat{A}_{hj0}(., \hat{\beta})$ of the true value of the p-dimensional regression parameter β_0 and of the integrated baseline intensity $A_{hj0}(t) = \int_0^t \alpha_{hj0}(u)du$, respectively. Most of this is the development of Cox regression model from Andersen *et al.* (1993)[29]. The stated main results will be needed for the proofs presented in Section 3.3 on the mean present values.

Let $(\Omega^{(n)}, \mathcal{F}^{(n)}, P^{(n)})$ denote the product probability space and we define the filtration $\mathcal{F}(t) = \mathcal{F}^0 \vee \mathcal{F}^Z(t) \vee \{\mathcal{N}_1(t), \ldots, \mathcal{N}_n(t)\}$. This filtration is the same as the one generated by the covariate vectors and all n Markov processes.

By the conditional independence of $X_i(.)$ and by the product construction (see Section II.4.3, Andersen *et al.* (1993)[29]), the multivariate counting process

$$N = \left(N_{hji}; i \in \{1, \ldots, n\}, h, j \in \{1, \ldots, k\}, h \neq j\right)$$

has the intensity process $\left(\lambda_{hji}; i \in \{1, \ldots, n\}, h, j \in \{1, \ldots, k\}, h \neq j\right)$ with respect to the combined filtration $\left(\mathcal{F}(t), t \in \mathcal{T}\right)$.

Next, suppose the observation of $N_i = \left(N_{hji}, h \neq j\right)$ is ceased after some random time $U_i > 0$. We say the process N_i is right censored at U_i. Define the censoring indicator process $C_i(t) = [U_i \geq t]$, the filtration $\mathcal{G}_i(t) = \mathcal{F}_i(t) \vee \sigma\left\{C_i(s), s \leq t\right\}$ and the right censored counting process $N_i^c = \left(N_{hji}^c, h \neq j\right)$, where $N_{hji}^c(t) = \int_0^t C_i(s) dN_{hji}(s) =$ $= \int_0^{t \wedge U_i} dN_{hji}(s)$. The censoring process $C_i(.)$ is $\mathcal{G}_i$-predictable.

Assume that given $Z_i(.)$, U_i is independent of $X_i(.)$. Then $N_i(.)$ has the same compensator both with respect to $\left(\mathcal{F}_i(t), t \in \mathcal{T}\right)$ and with respect to $\left(\mathcal{G}_i(t), t \in \mathcal{T}\right)$. This is the assumption of independent right censoring, referred in this sequel as "independent censoring".

Each N_{hji} with $h \neq j$ has the decomposition

$$N_{hji}(t) = \Lambda_{hji}(t) + M_{hji}(t),$$

where M_{hji} is a local square integrable martingale with respect to $\left(\mathcal{G}_i(t), t \in \mathcal{T}\right)$. Then

$$N_{hji}^c(t) = \int_0^t C_i(s) dN_{hji}(s) = \int_0^t C_i(s) d\Lambda_{hji}(s) + \int_0^t C_i(s) dM_{hji}(s) = \Lambda_{hji}^c(t) + M_{hji}^c(t).$$

By the predictability and boundedness of $C_i(.)$, M_{hji}^c is a local square integrable martingale with respect to $\left(\mathcal{G}_i(t), t \in \mathcal{T}\right)$. Thus, under independent censoring, N_{hji}^c has the $\left(\mathcal{G}_i(t), t \in \mathcal{T}\right)$ - compensator $\Lambda_{hji}^c(t) = \int_0^t C_i(s) d\Lambda_{hji}(s)$, so N_i^c has the intensity process $\lambda_i^c = \left\{\lambda_{hji}^c, h \neq j\right\}$, where $\lambda_{hji}^c(t) = \alpha_{hji}(t) Y_{hi}^c(t)$ and

$Y^c_{hi}(t) = C_i(t)Y_{hi}(t) = [X_i(t-) = h, U_i \geq t]$. Therefore N^c_{hji} has the same "individual intensity" α_{hji} as the uncensored process and the proportional hazards assumption is preserved for N^c_{hji}. Also Y^c_{hi} is interpreted as the predictable indicator process for the i-th individual being observed in state h just before time t.

For n independent processes $N^c_i, 1 \leq i \leq n$, $N^c = \left(N^c_i, i \in \{1,\ldots,n\}\right)$ has intensity process $\lambda^c = \left(N^c_i, i \in \{1,\ldots,n\}\right)$ (see Section II.4.3, Andersen *et al.* (1993)[29]).

From now on the superscript "c" will be dropped and although not explicit in the notation, N_{hji} and Y_{hi} are derived from censored observations.

The following standard notations, similar to those of Chapter 1, will be used. For $h \neq j$:

$$Z_{hji}(t)^{\otimes m} = Z_{hji}(t)Z_{hji}(t)', \text{ if } m = 2$$

$$= Z_{hji}(t), \text{ if } m = 1$$

$$= 1, \text{ if } m = 0;$$

$$S^{(m)}_{hj}(t,\beta) = \sum_{i=1}^{n} Y_{hi}(t)Z_{hji}(t)^{\otimes m} \exp(\beta' Z_{hji}(t)),\ m \in \{0,1,2\};$$

$$E_{hj}(t,\beta) = S^{(1)}_{hj}(t,\beta) / S^{(0)}_{hj}(t,\beta);$$

$$V_{hj}(t,\beta) = S^{(2)}_{hj}(t,\beta) / S^{(0)}_{hj}(t,\beta) - E_{hj}(t,\beta)^{\otimes 2};$$

$$I(t,\beta) = \sum_{h \neq j} \int_0^t V_{hj}(u,\beta) dN_{hj}(u), \text{ with } N_{hj} = \sum_{i=1}^{n} N_{hji};$$

$$s^{(m)}_{hj}(t,\beta) = E\left[Y_{h1}(t)Z_{hj1}(t)^{\otimes m} \exp(\beta' Z_{hj1}(t))\right],\ m \in \{0,1,2\};$$

$$e_{hj}(t,\beta) = s_{hj}^{(1)}(t,\beta) / s_{hj}^{(0)}(t,\beta);$$

$$\hat{A}_{hj}(t \mid Z_0) = \hat{A}_{hj0}(t,\hat{\beta})\exp(\hat{\beta}' Z_{hj0}) \;,$$

$$\Sigma(t,\beta) = \sum_{h \neq j} \int_0^t v_{hj}(u,\beta)\, s_{hj}^{(0)}(u,\beta)\alpha_{hj0}(u)du.$$

Consider the vector $Y_i = \left(Y_{hi}, h \in \{1,\ldots,k\}\right)$, where $1,\ldots,k$ label all the health states. As in Chapter 1, similar assumptions will be adopted throughout this chapter:

Model Assumptions:

A.1 Conditional on $Z_i(.)$, U_i is independent of $X_i(.)$;

A.2 $\left(N_i(.), Y_i(.), Z_i(.)\right), 1 \leq i \leq n$ are i.i.d.;

For $h \neq j$:

A.3 $A_{hj0}(\tau) = \int_0^\tau \alpha_{hj0}(t)dt < \infty$;

A.4 $Z_{hji}(.)$ are bounded;

A.5 $Z_{hji}(.)$ are adapted, left continuous with right hand limits processes (so $Z_{hji}(.)$ are predictable processes);

A.6 $P\left(Y_{h1}(t) = 1, \forall t \in [0,\tau]\right) > 0$;

A.7 $\Sigma_\tau \underset{notation}{\overset{by}{=}} \Sigma(\tau,\beta_0)$ is positive definite.

The form of the partial likelihood is functionally the same as in the case of the ordinary survival Cox proportional hazards model. Thus the log-partial likelihood evaluated at time t (see p483, Andersen *et al.* (1993)[29]) is:

$$C(t,\beta)=\sum_{i=1}^{n}\sum_{\substack{h,j=1\\h\neq j}}^{k}\int_0^t\left[\beta' Z_{hji}(u)-\log S_{hj}^{(0)}(t,\beta)\right]dN_{hji}(u).$$

Since $S_{hj}^{(1)}(t,\beta)$ is the vector of first partial derivatives of $S_{hj}^{(0)}(t,\beta)$ with respect to β, the vector $U(t,\beta)$ of partial derivatives of $C(t,\beta)$ with respect to β is

$$U(t,\beta)=\sum_{i=1}^{n}\sum_{\substack{h,j=1\\h\neq j}}^{k}\int_0^t\left[Z_{hji}(u)-E_{hj}(u,\beta)\right]dN_{hji}(u).$$

The **maximum partial likelihood estimator** $\hat{\beta}$ of β_0 is defined as the solution of the likelihood equation $U(\tau,\beta)=0$. For $h\neq j$ we estimate $A_{hj0}(t)$ by the **Nelson-Aalen estimator**

$$\hat{A}_{hj0}(t,\hat{\beta})=\int_0^t\frac{J_h(u)}{S_{hj}^{(0)}(u,\hat{\beta})}dN_{hj}(u),$$

where $N_{hj}=\sum_{i=1}^{n}N_{hji}$, $J_h(u)=\left[Y_h(u)>0\right]$, $Y_h=\sum_{i=1}^{n}Y_{hi}$. We use the convention $\frac{0}{0}=0$. Let $\hat{A}_{hh0}(t,\hat{\beta})=-\sum_{j\neq h}\hat{A}_{hj0}(t,\hat{\beta})$. Thus the matrix of integrated baseline intensities

$A_0(t)=\left(A_{hj0}(t),h,j\in\{1,\ldots,k\}\right)$ is estimated by $\hat{A}_0(t,\hat{\beta})=\left(\hat{A}_{hj0}(t,\hat{\beta}),h,j\in\{1,\ldots,k\}\right)$.

As we have also seen in Chapter 1, under our model assumptions A.1-A.7, the following conditions necessary for the asymptotic properties of our estimators are verified. We denote by $\|\,.\,\|$ the supremum norm of a vector or a matrix.

Conditions C.a-C.f:

There exist a compact neighborhood $\mathcal{B}$ of β_0, with $\beta_0 \in \overset{\circ}{\mathcal{B}}$ (the interior of $\mathcal{B}$), and scalar, p-vector and $p \times p$ matrix functions $s_{hj}^{(0)}$, $s_{hj}^{(1)}$ and $s_{hj}^{(2)}$, $h \neq j$, defined on $\hat{A}(.\,|\,Z_0) = \left(\hat{A}_{hj}(.\,|\,Z_0), h, j \in \{1,...,k\}\right)$ such that for $m \in \{0,1,2\}$ and $h, j \in \{1,...,k\}$, $h \neq j$:

C.a $\sup_{(t,\beta)\in[0,\tau]\times\mathcal{B}} \left\| \frac{1}{n} S_{hj}^{(m)}(t,\beta) - s_{hj}^{(m)}(t,\beta) \right\| \xrightarrow{P} 0$;

C.b $s_{hj}^{(m)}(.,.)$ are uniformly continuous bounded functions of $(t,\beta) \in [0,\tau]\times\mathcal{B}$;

C.c $s_{hj}^{(0)}(.,.)$ is bounded away from zero;

C.d $s_{hj}^{(1)}(t,\beta) = \frac{\partial}{\partial\beta} s_{hj}^{(0)}(t,\beta), s_{hj}^{(2)}(t,\beta) = \frac{\partial}{\partial\beta} s_{hj}^{(1)}(t,\beta)$;

C.e Σ_τ is positive definite;

C.f $\int_0^\tau \alpha_{hj0}(t)dt < \infty$.

Theorem 1 (see Theorem VII.2.1, p497, Andersen *et al.* (1993)[29])

Under the assumptions A.1, A.2 and conditions C.a-C.f, the probability that the equation $U(\tau,\beta) = 0$ has a unique solution $\hat{\beta}$ tends to one and $\hat{\beta} \xrightarrow{P} \beta_0$ as $n \to \infty$. □

The next theorem gives the asymptotic normality of $\hat{\beta}$ and an estimator of the asymptotic covariance:

Theorem 2 (see Theorem VII.2.2, p498, Andersen *et al.* (1993)[29])

Assume A.1, A.2, A.4 and C.a-C.f. Then $n^{1/2}(\hat{\beta}-\beta_0)$ converges in distribution to a zero mean normal p-dimensional random vector with covariance matrix Σ_τ^{-1} and $\sup_{t\in[0,\tau]}\left\| n^{-1}I(t,\hat{\beta})-\Sigma(t,\beta_0)\right\| \xrightarrow{P} 0$. In particular $\hat{\Sigma}_\tau \underset{notation}{\overset{by}{=}} n^{-1}I(\tau,\hat{\beta}) \xrightarrow{P} \Sigma_\tau$. □

The next theorem provides a description of the asymptotic joint distribution of the estimators $\hat{A}_{hj0}(t,\hat{\beta}), h, j \in \{1,...,k\}$ and $\hat{\beta}$.

First we need to state some definitions. We denote by $\langle M \rangle$ and $[M]$ the predictable and the optional variation process of a martingale M, respectively.

Definition (see p83, Andersen *et al.* (1993)[29]):

A continuous k-dimensional vector martingale $M = (M(t), t\in \mathcal{T}), \mathcal{T} = [0,\tau)$, $\tau \in \bar{\mathbb{R}}$ is called Gaussian if:

i) $\langle M \rangle = V$, a continuous deterministic $k \times k$ positive semidefinite matrix valued function on $\mathcal{T}$, with positive definite increments, zero at time zero;

ii) $M(t)-M(s)$ has a multivariate normal distribution with zero mean and covariance matrix $V(t)-V(s)$ and is independent of $(M(u), u\leq s)$, for all $0\leq s\leq t$ in $\mathcal{T}$. □

Definition

Two sequences of processes $\left(X_{n1}(.), n \geq 1\right)$ and $\left(X_{n2}(.), n \geq 1\right)$ are called asymptotic independent if $\left(X_{n1}, X_{n2}\right)(.)$ converges weakly to a process $(X_1, X_2)(.)$ with $X_1(.)$ independent of $X_2(.)$. □

Recall that $E = \{1, \ldots, k\}$ denotes the state space of all the Markov processes. Let $E^* = \left\{(h, j), h, j \in \{1, \ldots, k\}, h \neq j\right\}$.

Theorem 3 (see Theorem VII.2.3, p503, Andersen *et al.* (1993)[29])

Assume A.1, A.2, A.4 and C.a-C.f. Then $n^{1/2}(\hat{\beta} - \beta_0)$ and the processes

$W_{hj}(.) = n^{1/2}(\hat{A}_{hj0}(., \hat{\beta}) - A_{hj0}(.)) + n^{1/2}(\hat{\beta} - \beta_0)' \int_0 e_{hj}(u, \beta_0)\alpha_{hj0}(u)du$, $(h, j) \in E^*$ are independent. Let $W_{hh}(.) = -\sum_{j \neq h} W_{hj}(.)$.

The limiting distribution of the $k \times k$ matrix-valued process $W(.) = \left(W_{hj}(.), h, j \in \{1, \ldots, k\}\right)$ is that of a $k \times k$ matrix-valued process $U_0^*(.) = \left(U_{0hj}^*(.), h, j \in \{1, \ldots, k\}\right)$, where $U_{0hh}^* = -\sum_{j \neq h} U_{0hj}^*$ and $\left\{U_{0hj}^*(.), (h, j) \in E^*\right\}$ is a continuous Gaussian vector martingale, with

i) $U_{0hj}^*(0) = 0$,

ii) $\left\langle U_{0hj}^*, U_{0mr}^* \right\rangle = 0$ for $(h, j) \neq (m, r), (h, j), (m, r) \in E^*$,

iii) $\left\langle U_{0hj}^* \right\rangle(t) = \omega_{hj}^2(t) \underset{notation}{\overset{by}{=}} \int_0^t \frac{\alpha_{hj0}(u)}{s_{hj}^{(0)}(u, \beta_0)} du$. □

Notes:

1) The sequence of vector processes $\left\{W_{hj}(.),(h,j)\in E^{*}\right\}$ weakly converges to $\left\{U^{*}_{0hj}(.),(h,j)\in E^{*}\right\}$ in $D[0,\tau]^{k(k-1)}$, the space of $\mathbb{R}^{k(k-1)}$-valued right-continuous with left-hand limits functions on $[0,\tau]$, endowed with the Skorohod topology.

2) Relation ii) implies that the processes $U^{*}_{0hj}(.),(h,j)\in E^{*}$ are independent.

3) For $s,t\in[0,\tau]$ and $(h,j)\in E^{*}$ we have that

$$Cov(U^{*}_{0hj}(s),U^{*}_{0hj}(t))=\omega^{2}_{hj}(s\wedge t),$$

where $\omega^{2}_{hj}(.)$ is defined in iii).□

3.2.2 Estimation of the Transition Probabilities

The matrix of transition probabilities

$$P(s,t\mid Z_0)=\left(P_{hj}(s,t\mid Z_0),h,j\in\{1,\ldots,k\}\right)$$

for individuals with given fixed basic covariates Z_0 and corresponding type-specific covariates Z_{hj0} is defined as the product integral $P(s,t\mid Z_0)=\prod_{(s,t]}(I+dA(u\mid Z_0))$, for $s\le t$, $s,t\in[0,\tau]$, where the matrix of integrated intensity functions

$$A(.\mid Z_0)=\left(A_{hj}(.\mid Z_0),h,j\in\{1,\ldots,k\}\right)$$

has elements $A_{hj}(t\mid Z_0)=A_{hj0}(t)\exp(\beta_0' Z_{hj0})$ for $h\neq j$ and $A_{hh}(.\mid Z_0)=-\sum_{j\neq h}A_{hj}(.\mid Z_0)$.

For a review of the definition and properties of the product integration, see Section II.6 of Andersen *et al.* (1993).[29]

We consider the estimators $\hat{A}_{hj}(t \mid Z_0) = \hat{A}_{hj0}(t, \hat{\beta})\exp(\hat{\beta}' Z_{hj0})$ for $h \neq j$, $\hat{A}_{hh}(. \mid Z_0) = -\sum_{j \neq h} \hat{A}_{hj}(. \mid Z_0)$ and the matrix $\hat{A}(. \mid Z_0) = \left(\hat{A}_{hj}(. \mid Z_0), h, j \in \{1, \ldots, k\}\right)$. Then the matrix of transition probabilities $P(s,t \mid Z_0)$ is estimated by the product integral $\hat{P}(s,t \mid Z_0) = \prod_{(s,t]} (I + d\hat{A}(u \mid Z_0))$, this estimate being meaningful as long as $\Delta \hat{A}_{hh}(u \mid Z_0) \geq -1$ on $(s,t]$. A jump process ΔX is defined by $\Delta X(t) = X(t) - X(t-)$.

Next we state and prove the asymptotic properties of $\hat{P}(s,. \mid Z_0)$ for a given $s \in [0, \tau)$. We provide the details of the proofs because some results and intermediate steps, such as representations or expansions of certain entities, will be used in Section 3.3. Even though we follow the development offered by Andersen *et al.* (1993)[29] on p521-516, clear statements and details of needed results were not provided in any reference.

Let $(h, j) \in E^*$. First we want to show that

$n^{1/2}(\hat{A}_{hj}(. \mid Z_0) - A_{hj}(. \mid Z_0))$ is asymptotically equivalent to $X^n_{1hj}(.) + X^n_{2hj}(.)$, (3.3)

where $X^n_{1hj}(t) = \exp(\beta_0' Z_{hj0}) n^{1/2} (\hat{\beta} - \beta_0)' \int_0^t (Z_{hj0} - e_{hj}(u, \beta_0)) \alpha_{hj0}(u) du$ and $X^n_{2hj}(t) = \exp(\beta_0' Z_{hj0}) \hat{W}_{hj}(t)$.

The proof of (3.3) is very similar to the one of Theorem VII.2.3, Andersen *et al.* (1993).[29] The process $\hat{W}_{hj}(.)$ defined as

$$\hat{W}_{hj}(t) = n^{1/2} \int_0^t \frac{J_h(u)}{S^{(0)}_{hj}(u, \beta_0)} dM_{hj}(u)$$

is asymptotically equivalent to the process $W_{hj}(.)$ defined in Theorem 3. "Asymptotically equivalence" means convergence in probability to zero of the supremum norm of the difference.

We use the expansion:

$$n^{1/2}(\hat{A}_{hj}(t\,|\,Z_0)-A_{hj}(t\,|\,Z_0))=n^{1/2}(\hat{A}_{hj0}(t,\hat{\beta})\exp(\hat{\beta}'Z_{hj0})-A_{hj0}(t)\exp(\beta_0'Z_{hj0}))=$$

$$=n^{1/2}\int_0^t J_h(s)\left\{\exp(\hat{\beta}'Z_{hj0})S_{hj}^{(0)}(s,\hat{\beta})^{-1}-\exp(\beta_0'Z_{hj0})S_{hj}^{(0)}(s,\beta_0)^{-1}\right\}dN_{hj}(s)+$$

$$+\exp(\beta_0'Z_{hj0})n^{1/2}\int_0^t J_h(s)\left\{S_{hj}^{(0)}(s,\beta_0)^{-1}dN_{hj}(s)-\alpha_{hj0}(s)ds\right\}+$$

$$+\exp(\beta_0'Z_{hj0})n^{1/2}\int_0^t (J_h(s)-1)\alpha_{hj0}(s)ds.$$

The third term above converges in probability to zero and the second term is $X_{2hj}^n(t)$. By Taylor expansion around β_0, the first term equals

$$\exp(\beta^{*\prime}Z_{hj0})n^{1/2}(\hat{\beta}-\beta_0)'\int_0^t J_h(s)(Z_{hj0}-E_{hj}(u,\beta^*))S_{hj}^{(0)}(s,\beta^*)^{-1}dN_{hj}(s),$$

with β^* on the line segment between $\hat{\beta}$ and β_0. It can be shown that

$$\sup_{t\in[0,\tau]}\left\|\int_0^t J_h(s)(Z_{hj0}-E_{hj}(u,\beta^*))S_{hj}^{(0)}(s,\beta^*)^{-1}dN_{hj}(s)-\int_0^t (Z_{hj0}-e_{hj}(s,\beta_0))\alpha_{hj0}(s)du\right\|\xrightarrow{P}0$$

for any $\beta^*\xrightarrow{P}\beta_0$. Then the first term of the expanded $n^{1/2}(\hat{A}_{hj}(t\,|\,Z_0)-A_{hj}(t\,|\,Z_0))$ is asymptotically equivalent to $X_{1hj}^n(.)$, so (3.3) follows.

Then $n^{1/2}(\hat{A}_{hh}(.\,|\,Z_0)-A_{hh}(.\,|\,Z_0))=-n^{1/2}(\sum_{j\neq h}\hat{A}_{hj}(.\,|\,Z_0)-\sum_{j\neq h}A_{hj}(.\,|\,Z_0))$ is

asymptotically equivalent to $\left(-\sum_{j\neq h}X_{1hj}^n(.)\right)+\left(-\sum_{j\neq h}X_{2hj}^n(.)\right)\underset{notation}{\overset{by}{=}}X_{1hh}^n(.)+X_{2hh}^n(.)$.

Let $X_m^n(.)=\left(X_{mhj}^n(.),h,j\in\{1,\ldots,k\}\right),m\in\{1,2\}$. Thus

$n^{1/2}(\hat{A}(.|Z_0) - A(.|Z_0))$ is asymptotically equivalent to $X_1^n(.) + X_2^n(.)$. (3.4)

Now we will show the uniform consistency of the estimator $\hat{P}(s,.|Z_0)$ to $P(s,.|Z_0)$. From (3.3) we have that for $(h,j) \in E^*$, $\hat{A}_{hj}(.|Z_0) - A_{hj}(.|Z_0)$ is asymptotically equivalent to $B_{hj}^n(.) = B_{1hj}^n(.) + B_{2hj}^n(.) = n^{-1/2}\left(X_{1hj}^n(.) + X_{2hj}^n(.)\right)$. We have that

$$B_{1hj}^n(t) = \exp(\beta_0' Z_{hj0})(\hat{\beta} - \beta_0)' \int_0^t (Z_{hj0} - e_{hj}(u,\beta_0))\alpha_{hj0}(u)du,$$

$$B_{2hj}^n(t) = \exp(\beta_0' Z_{hj0}) \int_0^t \frac{J_h(u)}{S_{hj}^{(0)}(u,\beta_0)} dM_{hj}(u).$$

By the boundedness of Z_{hj0}, $e_{hj}(.,\beta_0)$, assumption A.3: $\int_0^\tau \alpha_{hj0}(t)dt < \infty$ and consistency of $\hat{\beta}$ to β_0, we have that $\sup_{t\in[0,\tau]} |B_{1hj}^n(t)| \xrightarrow{P} 0$. By the Lenglart's Inequality for local square integrable martingales (see p86, Andersen *et al.* (1993)[29]), for every $\eta, \delta > 0$:

$$P\left(\sup_{t\in[0,\tau]} \left| \int_0^t \frac{J_h(u)}{S_{hj}^{(0)}(u,\beta_0)} dM_{hj}(u) \right| > \eta \right) \le \frac{\delta}{\eta^2} + P\left(\int_0^\tau \frac{J_h(u)}{S_{hj}^{(0)}(u,\beta_0)^2} S_{hj}^{(0)}(u,\beta_0)\alpha_{hj0}(u)du > \eta \right)$$

and $n^{-1} \int_0^\tau J_h(u) \frac{n}{S_{hj}^{(0)}(u,\beta_0)} \alpha_{hj0}(u)du \xrightarrow{P} 0$. Thus $\sup_{t\in[0,\tau]} |B_{2hj}^n(t)| \xrightarrow{P} 0$.

Therefore we have proved $\sup_{t\in[0,\tau]} |B_{hj}^n(t)| \xrightarrow{P} 0$, from which

$\sup_{t\in[0,\tau]} \left\| \hat{A}(t|Z_0) - A(t|Z_0) \right\| \xrightarrow{P} 0$, where $\| . \|$ is the supremum norm.

Consider a fixed $s \in [0,\tau)$. By the continuity of the product integral (see Appendix B), the previous uniform convergence implies the uniform consistency of the estimator $\hat{P}(s,.\mid Z_0)$ to $P(s,.\mid Z_0)$:

$$\sup_{t\in[0,\tau]} \left\| \hat{P}(s,t\mid Z_0) - P(s,t\mid Z_0) \right\| \xrightarrow{P} 0.$$

In the following we will describe the asymptotic distribution of $n^{1/2}(\hat{P}(s,.\mid Z_0) - P(s,.\mid Z_0))$. In (3.4) we have seen that $n^{1/2}(\hat{A}(.\mid Z_0) - A(.\mid Z_0))$ was asymptotically equivalent to $X_1^n(.) + X_2^n(.)$.

By Theorem 2, $n^{1/2}(\hat{\beta} - \beta_0)$ converges in distribution to a zero mean normal distributed p-dimensional random vector ξ, with covariance matrix Σ_τ^{-1}. Then

$X_1^n(.)$ converges weakly to a $k \times k$ matrix-valued process

$$U_1^*(.\mid Z_0) = \left(U_{1hj}^*(.\mid Z_0), h, j \in \{1,\ldots,k\}\right), \tag{3.5}$$

where $U_{1hj}^*(t\mid Z_0) = \xi' W_{hj}^*(.\mid Z_0)$, with

$$W_{hj}^*(t\mid Z_0) = \exp(\beta_0' Z_{hj0}) \int_0^t (Z_{hj0} - e_{hj}(u,\beta_0))\alpha_{hj0}(u)du \text{ for } h \neq j \text{ and}$$

$$W_{hh}^*(.\mid Z_0) = -\sum_{j\neq h} W_{hj}^*(.\mid Z_0).$$

The process $\hat{W}_{hj}(.)$ is asymptotically equivalent to the process $W_{hj}(.)$ described in Theorem 3. Then

$X_2^n(.)$ converges weakly to a $k \times k$ matrix-valued process

$$U_2^*(.\mid Z_0) = \left(U_{2hj}^*(.\mid Z_0), h, j \in \{1,\ldots,k\}\right), \tag{3.6}$$

where $U_2^*(t \mid Z_0) = \exp(\beta_0' Z_{hj0}) U_0^*(t)$ and the process $U_0^*(.)$ is defined in the statement of Theorem 3.

Theorem 3 also implies that $X_1^n(.)$ and $X_2^n(.)$ are asymptotically independent. Then, by (3.5) and (3.6), $n^{1/2}(\hat{A}(.\mid Z_0) - A(.\mid Z_0))$ converges weakly (in the Skorohod sense) to $U^*(.\mid Z_0) = U_1^*(.\mid Z_0) + U_2^*(.\mid Z_0)$, where the processes $U_1^*(.\mid Z_0)$ and $U_2^*(.\mid Z_0)$ are independent, with continuous sample paths. Thus $n^{1/2}(\hat{A}(.\mid Z_0) - A(.\mid Z_0))$ converges weakly to $U^*(.\mid Z_0)$ in the supremum norm sense (see Appendix B). By the compact differentiability of the product integral and the Functional Delta Method (see Appendix B),

$n^{1/2}(\hat{P}(s,.\mid Z_0) - P(s,.\mid Z_0))$ converges weakly to

$$U(s,.\mid Z_0) = \int_s P(s,u \mid Z_0) dU^*(u \mid Z_0) P(u,.\mid Z_0) = U_1(s,.\mid Z_0) + U_2(s,.\mid Z_0), \tag{3.7}$$

where $U_1(s,.\mid Z_0)$ and $U_2(s,.\mid Z_0)$ are independent,

$$U_m(s,.\mid Z_0) = \int_s P(s,u \mid Z_0) dU_m^*(u \mid Z_0) P(u,.\mid Z_0), \; m \in \{1,2\}.$$

We can write

$$U_m(s,t \mid Z_0)_{hj} = \sum_{g=1}^{k} \sum_{l \neq g} \int_s^t P_{hg}(s,u \mid Z_0) dU_{mgl}^*(u \mid Z_0) P_{lj}(u,t \mid Z_0) +$$

$$+ \sum_{g=1}^{k} \int_s^t P_{hg}(s,u \mid Z_0)(-\sum_{l \neq g} dU_{mgl}^*(u \mid Z_0)) P_{gj}(u,t \mid Z_0) =$$

$$= \sum_{g=1}^{k} \sum_{l \neq g} \int_s^t P_{hg}(s,u \mid Z_0)\{P_{lj}(u,t \mid Z_0) - P_{gj}(u,t \mid Z_0)\} dU_{mgl}^*(u \mid Z_0),$$

where the matrix-valued processes $U_1^*(.\,|\,Z_0)$ and $U_2^*(.\,|\,Z_0)$ are described in (3.5) and (3.6).

3.2.3 Estimation of Mean Transition Cost

Consider all *h* to *j* transition expenses that occur in the follow-up period $[0,\tau]$. We assume that each patient has at least one of these expenses. In the comments at the end of this subsection we suggest two possible approaches for including in the analysis patients without such expenses. For a single individual *i* of our sample we consider the vector $C_i = (C_{i1},\ldots,C_{in_i})'$ of all *h* to *j* transition costs, where C_{il} (with $l \in \{1,\ldots,n_i\}$) denotes the observed expense incurred by the *i*-th patient at the *h* to *j* transition time t_{il}.

We will use a two-stage analysis that will finally result in a general linear mixed model (see Section 3.3, Verbeke and Molenberghs (1997)[61]). The first-stage regression model is given by

$$C_i = Z_i\beta_i + \varepsilon_i\,, \tag{3.8}$$

where β_i is a $q\times 1$ vector of regression parameters, Z_i is a $n_i \times q$ matrix of covariates and ε_i a vector of error terms with mean zero and covariance matrix Σ_i, all for the *i*-th subject. Typically the first column of Z_i is a vector of ones for the intercept while the rest of the columns are variables that vary within the subject.

We assume that every individual *h* to *j* transition cost profile can be well approximated by a polynomial function of time. For example, consider the cost quadratic in time. In this case Z_i is a $n_i \times 3$ matrix with ones in the first column, the time points

t_{il}, $l \in \{1,\ldots,n_i\}$ in the second column and the squared time points t_{il}^2, $l \in \{1,\ldots,n_i\}$ in the third column. Elementwise,

$$C_{il} = \beta_{1i} + \beta_{2i}t_{il} + \beta_{3i}t_{il}^2 + \varepsilon_{il},\ l \in \{1,\ldots,n_i\}, \tag{3.9}$$

where $\beta_i = (\beta_{1i}, \beta_{2i}, \beta_{3i})'$.

In the second-stage regression, β_i are regarded as n independent q-dimensional random vectors, called random regression coefficients. One goal is to investigate what subject-level characteristics affect these regression coefficients. Thus we assume:

$$\beta_i = B_i\beta + b_i, \tag{3.10}$$

where B_i is a $q \times p$ matrix of subject-level covariates, β is a p-dimensional vector of fixed-effects regression coefficients and b_i are i.i.d., with mean zero and covariance matrix D.

Replacing in (3.8) the random regression coefficients β_i by (3.10) yields the model

$$C_i = X_i\beta + Z_ib_i + \varepsilon_i, \tag{3.11}$$

where $X_i = Z_iB_i$ is a $n_i \times p$ matrix of covariates and all the other components are as defined before. The model (3.11) is a linear mixed model with fixed effects β and random effects b_i. We follow the notations and the theory development from Verbeke and Molenberghs (1997).[61] Other references with reviews of the linear mixed model theory are Verbeke and Molenberghs (2000)[62] and Diggle, Liang and Zeger (1994).[63]

Note:

The notation β applied conforms to usage in the mixed model theory and does not refer to the previous usage as regression parameter for transition intensities. □

For our example we assume the following second-stage (3.10)-type regression model:

$$\begin{aligned}\beta_{1i} &= \beta_1 F_{1i} + \beta_2 F_{2i} + \ldots + \beta_r F_{ri} + b_{1i} \\ \beta_{2i} &= \beta_{r+1} F_{1i} + \beta_{r+2} F_{2i} + \ldots + \beta_{2r} F_{ri} + b_{2i} \\ \beta_{3i} &= \beta_{2r+1} F_{1i} + \beta_{2r+2} F_{2i} + \ldots + \beta_{3r} F_{ri} + b_{3i},\end{aligned} \tag{3.12}$$

where the subject-level covariates, considered independent of time, might be age at diagnosis, indicator variables for disease, severity group etc. Usually $F_{1i} = 1$ for the intercept. In this section we want to estimate the expected h to j transition cost, conditional on a given vector Z_0 of covariates. Without loss of generality we can assume that the subject-level covariates in the (3.12) model are among the ones available in Z_0.

Let $\beta_{1,r} = (\beta_1, \ldots, \beta_r)'$, $\beta_{r+1,2r} = (\beta_{r+1}, \ldots, \beta_{2r})'$, $\beta_{2r+1,3r} = (\beta_{2r+1}, \ldots, \beta_{3r})'$ and $F_i = (F_{1i}, \ldots, F_{ri})'$. Then we write (3.12) as

$$\begin{aligned}\beta_{1i} &= \beta'_{1,r}\, F_i + b_{1i} \\ \beta_{2i} &= \beta'_{r+1,2r}\, F_i + b_{2i} \\ \beta_{3i} &= \beta'_{2r+1,3r}\, F_i + b_{3i}.\end{aligned}$$

Replacing (3.12) in (3.9), the resulting model for C_{il}, $l \in \{1, \ldots, n_i\}$ is

$$C_{il} = \left(\beta'_{1,r}\, F_i\right) + \left(\beta'_{r+1,2r}\, F_i\right)\, t_{il} + \left(\beta'_{2r+1,3r}\, F_i\right)\, t_{il}^2 + b_{1i} + b_{2i} t_{il} + b_{3i} t_{il}^2 + \varepsilon_{il}. \tag{3.13}$$

Notes:

1) Cnaan *et al.* (1997)[64] provides useful guidelines for modeling X_i and Z_i. One of these is that, in general, the columns of Z_i should be a subset of the columns of X_i. In our example we assume that a quadratic curve adequately models the time dependence on the mean response for each subject. We would not want to include a quadratic effect in Z_i and omit it in the mean modeled by X_i, since this model would imply that each subject had a quadratic curve but that the population curve was linear. The two-stage approach has the advantage that Z_i are necessarily a subset of X_i and any within-subject variables modeled in X_i are also contained in Z_i.

2) Linear mixed model literature is vast. Verbeke and Molenberghs (1997, 2000)[62] and Diggle, Liang and Zeger (1994)[63] provide reviews of the general theory of mixed models and also guidance on its application in practice.

3) Depending on the data, a simple transformation such as log or square root might be applied to C_{il} to attenuate the effects of cost skewness. □

The underlying assumptions of the linear mixed model (3.11) are that $b_i \sim MVN(0,D)$, $\varepsilon_i \sim MVN(0,\Sigma_i)$ and $b_1,...,b_n,\varepsilon_1,...,\varepsilon_n$ are independent. We denote by $MVN(\mu,\Sigma)$ a multivariate normal distribution with mean vector μ and covariance matrix Σ. The matrix D is a $q \times q$ covariance matrix and Σ_i is a $n_i \times n_i$ covariance matrix that depends on the index i only through its dimension n_i. Thus the unknown parameters in Σ_i do not depend upon i.

The mean and the covariance matrix of the cost vector C_i are:

$$E(C_i) = X_i\beta$$

$$\text{cov}(C_i) = Z_i D Z_i' + \Sigma_i \underset{notation}{\overset{by}{=}} V_i.$$

Conditional on the random effect b_i, C_i is normally distributed with mean vector $X_i\beta + Z_i b_i$ and with covariance matrix Σ_i. Further $b_i \sim MVN(0, D)$. Let $f(c_i \mid b_i)$ and $f(b_i)$ be the corresponding density functions. The marginal density function of C_i is $f(c_i) = \int f(c_i \mid b_i) f(b_i) db_i$, which is the density function for $MVN(X_i\beta, V_i)$. Statistical inference is based on these marginal distributions of the response variables C_i. Let α denote the vector of all variance and covariance parameters (called variance components) found in V_i. Assuming the cost vectors C_i independent, the marginal likelihood function is

$$L_{ML}(\beta,\alpha) = \prod_{i=1}^{n} \left\{ (2\pi)^{-n_i/2} \left| V_i(\alpha) \right|^{-1/2} \exp\left(-.5(C_i - X_i\beta)' V_i^{-1}(\alpha)(C_i - X_i\beta) \right) \right\},$$

where | . | denotes the determinant of the matrix.

First assume the parameter α known. Conditional on α, the maximum likelihood estimator (MLE) of β is given by

$$\hat{\beta}_{MLE}(\alpha) = (\sum_{i=1}^{n} X_i' V_i^{-1}(\alpha) X_i)^{-} \sum_{i=1}^{n} X_i' V_i^{-1}(\alpha) C_i$$

and follows a multivariate normal distribution with mean β and covariance matrix $\text{cov}\left(\hat{\beta}_{MLE}(\alpha)\right) = (\sum_{i=1}^{n} X_i' V_i^{-1}(\alpha) X_i)^{-}$.[65] Then the parameter α is estimated by its maximum likelihood (ML) or restricted maximum likelihood (REML) estimator.

Linear mixed models often contain many fixed effects and in such cases it might by important for the variance component estimation to explicitly take into account the loss of degrees of freedom involved in estimating the fixed effects. This can be done via restricted maximum likelihood estimation. The REML estimator for the variance components α is obtained from maximizing the likelihood function of error contrasts $U = A'Y$, where $Y = (Y_1',...,Y_n')'$ and A is a $(n \times (n-p))$ full-rank matrix with columns orthogonal to the columns of X, the matrix obtained from stacking the matrices X_i underneath each other. This likelihood can be written as:

$$L_{REML}(\alpha) = C\left|\sum_{i=1}^{n} X_i' V_i^{-1}(\alpha) X_i\right|^{-1/2} \times L_{ML}(\hat{\beta}_{MLE}(\alpha), \alpha),$$

where C is a constant which does not depend on α, so the resulting REML estimator does not depend on the error contrasts (i.e. on the choice of A). See p43-47, Verbeke and Molenberghs (2000)[62] or Diggle, Liang and Zeger (1994)[63] for reviews of the REML results and comparisons between ML and REML estimators.

Let $\hat{\alpha}$ denote the ML or REML estimator of α and $\hat{V}_i$ the estimator of V_i obtained by replacing the variance components α from D and Σ_i by $\hat{\alpha}$. We will then estimate $\hat{\beta}_{MLE}(\alpha)$ by

$$\hat{\beta} = (\sum_{i=1}^{n} X_i' \hat{V}_i^{-1} X_i)^{-} \sum_{i=1}^{n} X_i' \hat{V}_i^{-1} C_i \tag{3.14}$$

and $\text{cov}\left(\hat{\beta}_{MLE}(\alpha)\right)$ by $\widehat{\text{cov}}(\hat{\beta}_{MLE}) = (\sum_{i=1}^{n} X_i' \hat{V}_i^{-1} X_i)^{-}$.

It follows from classical likelihood theory (see for example Chapter 9, Cox and Hinkley (1990)[66]) that under some regularity conditions the REML estimator $\hat{\alpha}$ is consistent and its distribution can be well approximated by a normal distribution with mean vector α and covariance matrix given by the inverse of the Fisher information matrix.

Given α, suppose there exists the asymptotic matrix

$\Sigma(\alpha) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i' V_i^{-1}(\alpha) X_i$. Then $n^{1/2}(\hat{\beta}_{MLE}(\alpha) - \beta)$ converges weakly to a multivariate normal distribution $MVN(0, \Sigma(\alpha)^{-1})$.

In many situations (i.e. different covariance structures) the consistency of $\hat{\alpha}$ implies the consistency of our estimator $\hat{\beta} = \hat{\beta}_{MLE}(\hat{\alpha})$ (called the feasible generalized least squares estimator in the econometrics literature) and also implies that $n^{1/2}(\hat{\beta} - \beta)$ and $n^{1/2}(\hat{\beta}_{MLE}(\alpha) - \beta)$ have the same asymptotic distribution $MVN(0, \Sigma(\alpha)^{-1})$. Amemiya (1985)[67] (see p186-222) provides in detailed proofs of these results for several models with economic interpretation: 1) serial correlation; 2) seemingly unrelated regression models; 3) heteroscedasticity; 4) error components model and 5) random coefficients model. We assume that the above stated results hold in our situation. Thus

$\hat{\beta} \xrightarrow{P} \beta$ and

$$n^{1/2}(\hat{\beta} - \beta) \xrightarrow{D} \zeta_1, \text{ where } \zeta_1 \sim MVN(0, \Sigma(\alpha)^{-1}). \tag{3.15}$$

Now we need to interpret $c_{hj}(s \mid Z_0) = E(C_{hj}(s) \mid X(s-) = h, Z_0)$ and get a suitable estimator. For ease of explanation we will use the quadratic time model described previously as an example throughout this section.

Recall that for transitions of the h to j type we recorde $(C_{il}, l \in \{1,\ldots,n_i\})$, the costs incurred at the i-th individual's transition times $(t_{il}, l \in \{1,\ldots,n_i\})$. Based on our example model (3.13), the expected cost for the i-th individual with subject level covariates $F_{1i},\dots,F_{ri}$, at his/her h to j transition time t_{il} is

$$\begin{aligned}E(C_{il}) &= (\beta_1 F_{1i} + \beta_2 F_{2i} + \dots + \beta_r F_{ri}) + (\beta_{r+1} F_{1i} + \beta_{r+2} F_{2i} + \dots + \beta_{2r} F_{ri}) t_{il} + \\ &\quad + (\beta_{2r+1} F_{1i} + \beta_{2r+2} F_{2i} + \dots + \beta_{3r} F_{ri}) t_{il}^2 = \\ &= \beta'_{1,r} F_i + \beta'_{r+1,2r} F_i \, t_{il} + \beta'_{2r+1,3r} F_i \, t_{il}^2.\end{aligned} \tag{3.16}$$

In practice, after the covariate selection procedure, the model (3.13) might be reduced so some of the $3r$ regression coefficients in (3.16) might be zero.

For an individual with given subject covariates Z_0, $c_{hj}(s \mid Z_0)$ is the expected cost of the h to j transition time s. Thus we assume

$$c_{hj}(s \mid Z_0) = \beta'_{1,r} Z_0^* + \beta'_{r+1,2r} Z_0^* \, s + \beta'_{2r+1,3r} Z_0^* \, s^2,$$

where Z_0^* is the r-dimensional covariate vector obtained by substituting $F_{1i},\dots,F_{ri}$ by their correspondent covariates in Z_0.

Replacing the fixed effect regression parameter vector $\beta = (\beta'_{1,r}, \beta'_{r,2r}, \beta'_{2r,3r})'$ by its estimator given in (3.14), we estimate $c_{hj}(s \mid Z_0)$ by

$$\hat{c}_{hj}(s \mid Z_0) = \hat{\beta}'_{1,r} Z_0^* + \hat{\beta}'_{r+1,2r} Z_0^* \, s + \hat{\beta}'_{2r+1,3r} Z_0^* \, s^2. \tag{3.17}$$

Denote $Z_{01}^*(s) = \left(\left(Z_0^*\right)', s\left(Z_0^*\right)', s^2\left(Z_0^*\right)' \right)'$, so we can write $\hat{c}_{hj}(s \mid Z_0) = \hat{\beta}' Z_{01}^*(s)$.

The asymptotic properties (3.15) of the estimator $\hat{\beta}$ imply that

$$\sup_{s \in [0,\tau]} \left| \hat{c}_{hj}(s \mid Z_0) - c_{hj}(s \mid Z_0) \right| \xrightarrow{P} 0 \text{ and}$$

$$n^{1/2}(\hat{c}_{hj}(. \mid Z_0) - c_{hj}(. \mid Z_0)) \xrightarrow{D} c_{hj0}(. \mid Z_0), \text{ where } c_{hj0}(s \mid Z_0) \overset{D}{=} \zeta_1' Z_{01}^*(s). \tag{3.18}$$

Therefore the process $c_{hj0}(.\,|\,Z_0)$ is Gaussian, with mean zero and

$$\mathrm{cov}(c_{hj0}(s\,|\,Z_0), c_{hj0}(t\,|\,Z_0)) = Z_{01}^{*}(s)'\Sigma(\alpha)^{-1}Z_{01}^{*}(t).$$

The matrix $\Sigma(\alpha)^{-1}$ is consistently estimated by $\Sigma(\hat{\alpha})^{-1}$.

Comments

Suppose there are some patients that do not incur any h to j transition costs.

One possible approach for this situation is the use of a two-part model.[6, 68, 69] For each individual i in the sample we observe the binary variable δ_{ci} that indicates if the subject incurred any h to j transition expenses. Then we assume that $P(\delta_{ci} = 1) = \pi_i(\alpha)$ is governed by a parametric binary probability model (part one) and we consider the mixed-effects model (part two)

$$C_i = X_i\beta + Z_i b_i + \varepsilon_i,$$

where all the assumed error and random effects distributions are conditional on the realization of the event $(\delta_{ci} = 1)$. Then

$$E(C_i) = P(\delta_{ci} = 1)E(C_i\,|\,\delta_{ci} = 1) = \pi_i(\alpha)X_i\beta.$$

The component $\pi_i(\alpha)$ can be specified through either a logit model:

$\pi_i(\alpha) = \dfrac{\exp(\alpha' Z_i)}{1+\exp(\alpha' Z_i)}$ or a probit model $\pi_i(\alpha) = \Phi(\alpha' Z_i)$, where Z_i is a set of covariates and Φ the standard normal distribution function. The parameter α is estimated using all the data. Then one could use the available costs C_i to estimate the parameters β in the linear mixed models.

A second possible approach is to consider for the individual i all transition expenses $C_i = (C_{i1},\ldots,C_{in_i})'$ that occur in the follow-up period $[0,\tau]$ and not only the h to

j transition costs. We denote by C_{il} the observed expense incurred by the i-th patient at time t_{il}. Then, in the first-stage model (3.8), the matrix Z_i of covariates includes also classification variables for the origination and destination states of the transitions. For example,

$$C_{il} = \beta_{1i} + \sum_{a=1}^{k-1} \beta_{2i}^{a}[X(t_{il}-) = a] + \sum_{\delta=1}^{k-1} \beta_{3i}^{\delta}[X(t_{il}) = \delta] + \beta_{4i}t_{il} + \beta_{5i}t_{il}^{2} + \varepsilon_{il}\,,\ l \in \{1,\ldots,n_i\},$$

where s_{il} denotes the originating health state and d_{il} the destination state. After the second-stage regression, the resulting mean transition costs have the form

$$E(C_{il}) = \beta_1' F_i + \sum_{a=1}^{k-1}\left(\beta_2^{a}\right)' F_i[X(t_{il}-) = a] + \sum_{\delta=1}^{k-1}\left(\beta_{3i}^{\delta}\right)' F_i[X(t_{il}) = \delta] + \beta_4' F_i\, t_{il} + \beta_5' F_i\, t_{il}^{2},$$

$l \in \{1,\ldots,n_i\}$. Hence we assume

$$c_{hj}(s \mid Z_0) = \beta_1' Z_0^{*} + \left(\beta_2^{h}\right)' Z_0^{*} + \left(\beta_3^{j}\right)' Z_0^{*} + \beta_4' Z_0^{*} s + \beta_5' Z_0^{*} s^2$$

and we estimate it by replacing the fixed-effect regression parameter vector β with its estimator given in (3.14).

This strategy of utilizing the entire sample to estimate simultaneously all $\left(c_{hj}(.\mid Z_0), h \neq j\right)$ has the advantage of drawing strength from other parts of the data set, when some patients do not have observed transitions of a specific type. Its limitation is that, when considering all types of transition costs, it might be difficult to distinguish any pattern in time that approximates the individual transition cost profiles. □

3.2.4 Estimation of Mean Sojourn Rate

For a given fixed covariate vector Z_0, we defined in Section 3.1.2

$$b_h(u \mid Z_0) = E(B_h(u) \mid X(u-) = h, Z_0)$$

as the expected expense rate at time u for sojourns in state h. This quantity is never truly observed unless we have a very fine time scale on which the accumulating cost history is observed. For example, daily or weekly costs incurred while sojourning in a state might provide in same applications an adequate representation of the rate of expenditures. We assume we do not have a detailed cost history and observation is restricted to the total expenditures for each sojourn, together with the time of entry and duration of the sojourns. The cumulative expected expense of a sojourn in state h with entry time s, after duration d is $C_h(s,d) = \int_s^{s+d} b_h(u \mid Z_0)du$. We assume the rate of accumulating costs in a sojourn does not depend on the entry time in that sojourn.

For the i-th individual, all observed total costs incurred in sojourns in state h are collected into a single vector $C_i = (C_{i1}, \dots, C_{in_i})'$. For $l \in \{1, \dots, n_i\}$, C_{il} is the total cost (up to transition in another state or up to censoring) of the l-th sojourn in state h that had entry time s_{il} and duration d_{il}. We assume the vector C_i contains at least one cost value, for all i. The comments from the end of the previous section apply for the situation when this assumption does not hold.

Our approach is similar to our use of linear mixed models to derive estimates of the expected transition costs. We use the same two-stage analysis and we also make the assumption that the individual sojourn cumulative profile can be well approximated by a polynomial function of the duration of the sojourn.

As an example we consider again a quadratic curve model. For the first stage model, let

$$C_{il} = \beta_{1i} + \beta_{2i} I_{il}^{c} + \beta_{3i} s_{il} + \beta_{4i} d_{il} + \beta_{5i} d_{il}^{2} + \varepsilon_{il}, \tag{3.19}$$

where I_{il}^{c} is the indicator that the l-th sojourn in state h of the i-th individual is completely observed. The second stage model is similar to (3.12):

$$\beta_{ki} = \beta'_{(k-1)r+1,kr} F_i + b_{ki},\ k \in \{1,2,3,4,5\}. \tag{3.20}$$

We use the notations from the previous section. Let $\beta = (\beta'_{1,r}, \beta'_{r+1,2r}, \ldots, \beta'_{4r+1,5r})'$, $b_i = (b_{1i}, \ldots, b_{5i})'$.

Replacing (3.20) in (3.19) we obtain the following final model:

$$\begin{aligned} C_{il} = {} & \beta'_{1,r} F_i + \beta'_{r+1,2r} F_i\, I_{il}^{c} + \beta'_{2r+1,3r} F_i\, s_{il} + \beta'_{3r+1,4r} F_i\, d_{il} + \beta'_{4r+1,5r} F_i\, d_{il}^{2} + \\ & + b_{1i} + b_{2i} I_{il}^{c} + b_{3i} s_{il} + b_{4i} d_{il} + b_{5i} d_{il}^{2} + \varepsilon_{il}. \end{aligned} \tag{3.21}$$

Note:

In practice, after the model selection procedure, some of the regression coefficients might be zero. □

Based on the model (3.21), an estimator $\hat{\beta}$ analogous to (3.14) can be calculated.

For an individual i with an observed l-th sojourn in state h with entry time s_{il} and duration d_{il}, the expected cost is

$$\begin{aligned} E(C_{il}) &= \beta'_{1,r} F_i + \beta'_{r+1,2r} F_i + \beta'_{2r+1,3r} F_i s_{il} + \beta'_{3r+1,4r} F_i d_{il} + \beta'_{4r+1,5r} F_i d_{il}^{2} \\ &= \beta' \left(F_i', F_i', s_{il} F_i', d_{il} F_i', d_{il}^{2} F_i' \right)'. \end{aligned}$$

Accordingly, for an individual with given subject covariates Z_0, we assume that

$$C_h(s,d) = \beta'\left(\left(Z_0^*\right)',\left(Z_0^*\right)',s\left(Z_0^*\right)',d\left(Z_0^*\right)',d^2\left(Z_0^*\right)'\right)'$$

and we estimate $C_h(s,d)$ by

$$\hat{C}_h(s,d) = \hat{\beta}'\left(\left(Z_0^*\right)',\left(Z_0^*\right)',s\left(Z_0^*\right)',d\left(Z_0^*\right)',d^2\left(Z_0^*\right)'\right)',$$

where, as defined in the previous section, Z_0^* is the r-dimensional covariate vector obtained on substituting the elements of F_i by their corresponding covariates in Z_0.

The rate $b_h(.\,|\,Z_0)$ was assumed not to depend on the entry time s. Therefore we estimate it by

$$\hat{b}_h(u\,|\,Z_0) = \frac{\partial}{\partial d}\hat{C}_h(0,d)\big|_{d=u} = \hat{\beta}' Z_{02}^*(u), \tag{3.22}$$

where $Z_{02}^*(u) = \left(0' \quad 0' \quad 0' \quad \left(Z_0^*\right)' \quad u\left(Z_0^*\right)'\right)'$.

As described in the previous section, under some regularity conditions

$$\sup_{u\in[0,\tau]}\left|\hat{b}_h(u\,|\,Z_0) - b_h(u\,|\,Z_0)\right| \xrightarrow{P} 0 \text{ and}$$

$$n^{1/2}\left(\hat{b}_h(.\,|\,Z_0) - b_h(.\,|\,Z_0)\right) \xrightarrow{D} b_{h0}(.\,|\,Z_0),\text{ with } b_{h0}(u\,|\,Z_0) \overset{D}{=} \zeta_2' Z_{02}^*(u), \tag{3.23}$$

where $n^{1/2}(\hat{\beta}-\beta) \xrightarrow{D} \zeta_2$, $\zeta_2 \sim MVN(0,\Sigma(\alpha)^{-1})$ and $\Sigma(\alpha) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} X_i' V_i^{-1}(\alpha) X_i$. We use the same notations from the previous section. This will not create confusion because we derive asymptotic properties separately for the mean transition cost and the mean sojourn rate. The process $b_{h0}(.\,|\,Z_0)$ is Gaussian, with zero mean and

$$\text{cov}(b_{h0}(u \mid Z_0), b_{h0}(w \mid Z_0)) = Z_{02}^*(u)'\Sigma(\alpha)^{-1} Z_{02}^*(w).$$

3.3 Large Sample Properties of the Mean Cost Estimators

3.3.1 Uniform Consistency of the Mean Cost Estimators

By (3.1), conditional on the initial state i, given the vector Z_0 of basic covariates, the mean present value of all expenditures associated with the h to j transitions in $(0,t]$ is

$$MPV_{hj}^{(1)}(t \mid i, Z_0) = \int_0^t \gamma_{hji}(s)\, dA_{hj}(s \mid Z_0),$$

where $\gamma_{hji}(s) = e^{-rs} c_{hj}(s \mid Z_0) P_{ih}(0, s \mid Z_0)$ and $A_{hj}(s \mid Z_0) = \int_0^s \alpha_{hj0}(u) \exp(\beta_0' Z_{hj0}) du$ is the integrated intensity function of a h to j transition. We estimate this quantity by

$$M\hat{P}V_{hj}^{(1)}(t \mid i, Z_0) = \int_0^t \hat{\gamma}_{hji}(s) d\hat{A}_{hj}(s \mid Z_0),$$

where $\hat{\gamma}_{hji}(s) = e^{-rs} \hat{c}_{hj}(s \mid Z_0) \hat{P}_{ih}(0, s \mid Z_0)$ and $\hat{A}_{hj}(s \mid Z_0) = \hat{A}_{hj0}(s, \hat{\beta}) \exp(\hat{\beta}' Z_{hj0})$. See Sections 3.2.1, 3.2.2 and 3.2.2 for the definitions of the estimators $\hat{A}_{hj}(. \mid Z_0)$, $\hat{P}_{ih}(0,. \mid Z_0)$ and $\hat{c}_{hj}(. \mid Z_0)$.

We will prove the uniform consistency of the mean transition cost estimator, that is

$$\sup_{t \in [0,\tau]} \left| M\hat{P}V_{hj}^{(1)}(t \mid i, Z_0) - MPV_{hj}^{(1)}(t \mid i, Z_0) \right| \xrightarrow{P} 0. \tag{3.24}$$

We first prove

$$\sup_{t \in [0,\tau]} \left| \hat{\gamma}_{hji}(t) - \gamma_{hji}(t) \right| \xrightarrow{P} 0. \tag{3.25}$$

By the definition of $\hat{\gamma}_{hji}(.)$ and $\gamma_{hji}(.)$,

$$\sup_{t\in[0,\tau]}\left|\hat{\gamma}_{hji}(t)-\gamma_{hji}(t)\right| \le \sup_{t\in[0,\tau]}\left|\hat{c}_{hj}(t\,|\,Z_0)\hat{P}_{ih}(0,t\,|\,Z_0)-c_{hj}(t\,|\,Z_0)P_{ih}(0,t\,|\,Z_0)\right| \le$$

$$\le \sup_{t\in[0,\tau]}\left|\hat{c}_{hj}(t\,|\,Z_0)\right| \sup_{t\in[0,\tau]}\left|\hat{P}_{ih}(0,t\,|\,Z_0)-P_{ih}(0,t\,|\,Z_0)\right| + \sup_{t\in[0,\tau]}\left|\hat{c}_{hj}(t\,|\,Z_0)-c_{hj}(t\,|\,Z_0)\right|.$$

By (3.18), $\sup_{t\in[0,\tau]}\left|\hat{c}_{hj}(t\,|\,Z_0)-c_{hj}(t\,|\,Z_0)\right|\xrightarrow{P}0$ and by the assumption A.0.1, $c_{hj}(.\,|\,Z_0)$ is bounded. We have shown in Section 3.2.2 that $\sup_{t\in[0,\tau]}\left|\hat{P}_{ih}(0,t\,|\,Z_0)-P_{ih}(0,t\,|\,Z_0)\right|\xrightarrow{P}0$.

Consequently (3.25) follows.

For $t\in[0,\tau]$:

$$\left|M\hat{P}V_{hj}^{(1)}(t\,|\,i,Z_0)-MPV_{hj}^{(1)}(t\,|\,i,Z_0)\right| \le \int_0^t\left|\hat{\gamma}_{hji}(s)-\gamma_{hji}(s)\right|d\hat{A}_{hj}(s\,|\,Z_0)+$$

$$+\left|\int_0^t\gamma_{hji}(s)\left(d\hat{A}_{hj}(s\,|\,Z_0)-dA_{hj}(s\,|\,Z_0)\right)\right|, \text{ so}$$

$$\sup_{t\in[0,\tau]}\left|M\hat{P}V_{hj}^{(1)}(t\,|\,i,Z_0)-MPV_{hj}^{(1)}(t\,|\,i,Z_0)\right| \le \sup_{s\in[0,\tau]}\left|\hat{\gamma}_{hji}(s)-\gamma_{hji}(s)\right|\hat{A}_{hj}(\tau\,|\,Z_0)+$$

$$+\sup_{t\in[0,\tau]}\left|\int_0^t\gamma_{hji}(s)\left(d\hat{A}_{hj}(s\,|\,Z_0)-dA_{hj}(s\,|\,Z_0)\right)\right|.$$

By (3.25) and the fact $\hat{A}_{hj}(\tau\,|\,Z_0)\xrightarrow{P}A_{hj}(\tau\,|\,Z_0)<\infty$, the first term of the right hand side of the above inequality converges to zero in probability.

By our model assumptions $\gamma_{hji}(.)$ is bounded on $[0,\tau]$. Then, as in the proof of (3.3), we can show that $\int_0^t\gamma_{hji}(s)\left(d\hat{A}_{hj}(s\,|\,Z_0)-dA_{hj}(s\,|\,Z_0)\right)$ is asymptotically equivalent to

$$B_{hji}(t) = \exp(\beta_0' Z_{hj0}) \int_0^t \gamma_{hji}(s) \frac{J_h(s)}{S_{hj}^{(0)}(s,\beta_0)} dM_{hj}(s) +$$

$$+ (\hat{\beta} - \beta_0)' \int_0^t \gamma_{hji}(s)(Z_{hj0} - e_{hj}(s,\beta_0))\alpha_{hj0}(s)ds.$$

Using Lenglart's Inequality for square integrable martingales, consistency of $\hat{\beta}$ to β_0, $\int_0^\tau \alpha_{hj0}(s)\,ds < \infty$ and our model boundedness assumptions, one can prove that $\sup_{t\in[0,\tau]} \left|B_{hji}(t)\right| \xrightarrow{P} 0$. Thus $\sup_{t\in[0,\tau]} \left| \int_0^t \gamma_{hji}(s)\left(d\hat{A}_{hj}(s \mid Z_0) - dA_{hj}(s \mid Z_0)\right) \right| \xrightarrow{P} 0$ and (3.24) follows.

Next we will show the uniform consistency of the mean sojourn cost estimator. By (3.2), conditional on the initial state *i*, given the vector Z_0 of basic covariates, the mean present value of expenditures after duration time *d* for the sojourns in state *h* with entry time *s* is

$$MPV_h^{(2)}(s,d \mid i, Z_0) = \int_s^{s+d} e^{-ru} b_h(u \mid Z_0) P_{ih}(0,u \mid Z_0) du .$$

We estimate this quantity by

$$M\hat{P}V_h^{(2)}(s,d \mid i, Z_0) = \int_s^{s+d} e^{-ru} \hat{b}_h(u \mid Z_0) \hat{P}_{ih}(0,u \mid Z_0) du ,$$

where the estimators $\hat{P}_{ih}(0,. \mid Z_0)$ and $\hat{b}_h(. \mid Z_0)$ are defined in Sections 3.2.2 and 3.2.4, respectively.

Using an argument similar to the one used to show (3.25), it can be shown that for every duration *d* such that $0 < s + d \le \tau$:

$$\sup_{a\in[0,d]} \left| M\hat{P}V_h^{(2)}(s,a \mid i, Z_0) - MPV_h^{(2)}(s,a \mid i, Z_0) \right| \le$$

$$\le \text{constant} \times \sup_{u\in[s,s+d]} \left| \hat{b}_h(u \mid Z_0)\hat{P}_{ih}(0,u \mid Z_0) - b_h(u \mid Z_0) P_{ih}(0,u \mid Z_0) \right| = o_P(1).$$

Therefore the uniform consistency of the mean sojourn cost estimator holds:

$$\sup_{a\in[0,d]}\left|M\hat{P}V_h^{(2)}(s,a\,|\,i,Z_0)-MPV_h^{(2)}(s,a\,|\,i,Z_0)\right|\xrightarrow{P}0$$

for all d such that $0<s+d\le\tau$.

3.3.2 Asymptotic Distribution of the Mean Transition Cost

In this section we will assess the asymptotic normality of $n^{1/2}\left(M\hat{P}V_{hj}^{(1)}(t\,|\,i,Z_0)-MPV_{hj}^{(1)}(t\,|\,i,Z_0)\right)$ using the Functional Delta Method. Appendix B provides a statement of this method, the concept of Hadamard differentiability and other related results that we will use in this section.

Fix the time t. Consider the functional

$$\varphi_t : E\rightarrow \mathbb{R},\ \varphi_t(x,y,z)=\int_0^t x(s)y(s)dz(s),$$

where E is a subset of $D[0,\tau]^3$ such that φ_t is well defined. Notice we can write $MPV_{hj}^{(1)}(t\,|\,i,Z_0)$ as

$$\varphi_t(x_0,y_0,z_0)=\varphi_t(e^{-r\cdot}c_{hj}(.\,|\,Z_0),P_{ih}(0,.\,|\,Z_0),A_{hj}(.\,|\,Z_0)).$$

If φ_t has an extension to $D[0,\tau]^3$ that is Hadamard differentiable in (x_0,y_0,z_0) then, under some extra-conditions, we can apply the Functional Delta Method to obtain our desired result.

First we recall our convergence results from the previous sections.

From (3.3), (3.5) and (3.6), we obtained in Section 3.2.2 that for $(h,j)\in E^{*}$, $n^{1/2}(\hat{A}_{hj}(.\,|\,Z_0)-A_{hj}(.\,|\,Z_0))$ converged weakly in $D[0,\tau]$ (in the Skorohod sense) to the process $U^{*}_{hj}(.\,|\,Z_0)=U^{*}_{1hj}(.\,|\,Z_0)+U^{*}_{2hj}(.\,|\,Z_0)$, where $U^{*}_{1hj}(.\,|\,Z_0)$ and $U^{*}_{2hj}(.\,|\,Z_0)$ were independent. We showed that

$$U^{*}_{1hj}(t\,|\,Z_0)\overset{D}{=}\xi' W^{*}_{hj}(t\,|\,Z_0), \tag{3.26}$$

where $\xi \sim MVN(0,\Sigma_\tau^{-1})$, the matrix Σ_τ being defined in assumption A.7, Section 3.2.1, and $W^{*}_{hj}(t\,|\,Z_0)=\exp(\beta_0' Z_{hj0})\int_0^t (Z_{hj0}-e_{hj}(u,\beta_0))\alpha_{hj0}(u)du$. The other process is defined as

$$U^{*}_{2hj}(t\,|\,Z_0)\overset{D}{=}\exp(\beta_0' Z_{hj0})U^{*}_{0hj}(t), \tag{3.27}$$

where the matrix-valued process $U^{*}_0(.)$ is described in Theorem 3, Section 3.2.1.

By (3.7), $n^{1/2}(\hat{P}_{ih}(0,.\,|\,Z_0)-P_{ih}(0,.\,|\,Z_0))$ converges weakly to the process $U_{ih}(0,.\,|\,Z_0)=U_{1ih}(0,.\,|\,Z_0)+U_{2ih}(0,.\,|\,Z_0)$, where $U_{1ih}(0,.\,|\,Z_0)$ and $U_{2ih}(0,.\,|\,Z_0)$ are independent and for $m\in\{1,2\}$

$$U_{mih}(0,s\,|\,Z_0)=\sum_{g=1}^{k}\sum_{l\neq g}\int_0^s P_{ig}(0,u\,|\,Z_0)\{P_{lh}(u,s\,|\,Z_0)-P_{gh}(u,s\,|\,Z_0)\}dU^{*}_{mgl}(u\,|\,Z_0).$$

Using (3.26) and (3.27), we write

$$U_{1ih}(0,s\,|\,Z_0)\overset{D}{=}\xi' F_{ih}(0,s\,|\,Z_0), \tag{3.28}$$

$$\begin{aligned}U_{2ih}(0,s\,|\,Z_0)&=\exp(\beta_0' Z_{hj0})\times\\ &\times\sum_{g=1}^{k}\sum_{l\neq g}\int_0^s P_{ig}(0,u\,|\,Z_0)\{P_{lh}(u,s\,|\,Z_0)-P_{gh}(u,s\,|\,Z_0)\}dU^{*}_{0gl}(u)'\end{aligned} \tag{3.29}$$

where $F_{ih}(0,s\,|\,Z_0)=\sum_{g=1}^{k}\sum_{l\neq g}\int_0^s P_{ig}(0,u\,|\,Z_0)\{P_{lh}(u,s\,|\,Z_0)-P_{gh}(u,s\,|\,Z_0)\}dW_{gl}^{*}(u\,|\,Z_0)$.

By (3.18), $n^{1/2}(\hat{c}_{hj}(.\,|\,Z_0)-c_{hj}(.\,|\,Z_0))$ converges weakly to the Gaussian process $c_{hj0}(.\,|\,Z_0)$ described in Section 3.2.3:

$$c_{hj0}(s\,|\,Z_0)=\zeta_1{}'Z_{01}^{*}(s)\,, \tag{3.30}$$

where $\zeta_1 \sim MVN(0,\Sigma(\alpha)^{-1})$, $\Sigma(\alpha)=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}X_i'V_i^{-1}(\alpha)X_i$.

Next we prove the following lemma. We denote by $\int_a^b|dz|$ the total variation of the function z.

Lemma

Let $E=\left\{(x,y,z)\in D[0,\tau]^3:\int_0^\tau|\,dz\,|\leq C\right\}$, where $0<C<\infty$. For a fixed time $t\in(0,\tau]$ we define $\varphi_t:E\to\mathbb{R}$ by $\varphi_t(x,y,z)=\int_0^t xydz$.

Let (x_0,y_0,z_0) be a fixed point of E such that $\int_0^\tau|\,d(x_0y_0)\,|<\infty$. Then φ_t can be extended to the space $D[0,\tau]^3$ so as to be Hadamard differentiable at (x_0,y_0,z_0), with derivative

$$d\varphi_t(x_0,y_0,z_0).(h,k,l)=\int_0^t hy_0dz_0+\int_0^t x_0kdz_0+\int_0^t x_0y_0dl\,, \tag{3.31}$$

where the integral with respect to l is defined by the integration by parts formula if l is not of finite variation. □

Notes:

1) We interpret integration from 0 to τ as being over the interval $(0,\tau]$.

2) The integration by parts formula gives

$$\int_0^t x_0 y_0 dl = x_0(t) y_0(t) l(t) - x_0(0) y_0(0) l(0) - \int_{(0,t]} l_- d(x_0 y_0).$$

3) The extension assessed in this Lemma is not necessarily unique and the differentiability is shown only for the fixed point (x_0, y_0, z_0). □

Proof Lemma:

Obviously the hypothesized derivative $d\varphi_t(x_0, y_0, z_0)$ is a linear map. We will show that $d\varphi_t(x_0, y_0, z_0)$ is also continuous. Let $(h,k,l) \in D[0,\tau]^3$ be a fixed arbitrary point. Consider sequences $t_n \in \mathbb{R}^+$, $h_n, k_n, l_n \in D[0,\tau]$ that satisfy $t_n \to 0$, $h_n \xrightarrow{\|\cdot\|} h$, $k_n \xrightarrow{\|\cdot\|} k$, $l_n \xrightarrow{\|\cdot\|} l$, where $\|.\|$ is the supremum norm, and define

$$x_n = x_0 + t_n h_n$$
$$y_n = y_0 + t_n k_n$$
$$z_n = z_0 + t_n l_n$$

and suppose $(x_n, y_n, z_n) \in E$ for each *n*. We need to prove that

$$d\varphi_t(x_0, y_0, z_0).(h_n, k_n, l_n) \to d\varphi_t(x_0, y_0, z_0).(h,k,l) \text{ as } n \to \infty. \quad (3.32)$$

The sequence $d\varphi_t(x_0, y_0, z_0).(h_n, k_n, l_n)$ has the form

$$d\varphi_t(x_0, y_0, z_0).(h_n, k_n, l_n) = \int_0^t h_n y_0 dz_0 + \int_0^t x_0 k_n dz_0 + \int_0^t x_0 y_0 dl_n,$$

where the integral with respect to l_n is defined by integration by parts formula.

$$\left|\int_0^t h_n y_0 dz_0 - \int_0^t h y_0 dz_0\right| = \left|\int_0^t (h_n - h) y_0 dz_0\right| \le \|h_n - h\| \times \|y_0\| \times \int_0^\tau |dz_0|.$$

By the hypothesis of the Lemma we prove $\int_0^\tau |dz_0| \le C$ and we also have $\|y_0\| < \infty$

because $y_0 \in D[0,\tau]$. Then the convergence $h_n \xrightarrow{\|\cdot\|} h$ implies that

$$\left|\int_0^t h_n y_0 dz_0 - \int_0^t h y_0 dz_0\right| \to 0 \text{ as } n \to \infty.$$

Similarly

$$\left|\int_0^t x_0 k_n dz_0 - \int_0^t x_0 k dz_0\right| \to 0 \text{ as } n \to \infty.$$

By the integration by parts formula

$$\left|\int_0^t x_0 y_0 dl_n - \int_0^t x_0 y_0 dl\right| \le |x_0(t)| \times |y_0(t)| \times \|l_n - l\| + |x_0(0)| \times |y_0(0)| \times \|l_n - l\| +$$

$$+ \|l_{n-} - l_-\| \times \int_0^\tau |d(x_0 y_0)|.$$

But $\|l_n - l\| = \|l_{n-} - l_-\|$, so

$$\left|\int_0^t x_0 y_0 dl_n - \int_0^t x_0 y_0 dl\right| \le \|l_n - l\| \left(2\|x_0\| \times \|y_0\| + \int_0^\tau |d(x_0 y_0)|\right).$$

Because $\int_0^\tau |d(x_0 y_0)| < \infty$, $l_n \xrightarrow{\|\cdot\|} l$ and $\|x_0\|, \|y_0\| < \infty$, the right hand side of the previous

inequality converges to zero as $n \to \infty$, so

$$\left|\int_0^t x_0 y_0 dl_n - \int_0^t x_0 y_0 dl\right| \to 0 \text{ as } n \to \infty.$$

We proved (3.32) that implies the continuity of the mapping $d\varphi_t(x_0, y_0, z_0)$.

Next we apply the Lemma stated in Appendix B. In its context B_1 and B_2 are

normed vector spaces, endowed with σ – algebras $\mathcal{B}_1, \mathcal{B}_2$, respectively, where

$\mathcal{B}_i' \subseteq \mathcal{B}_i \subseteq \mathcal{B}_i''$, $i = 1,2$. The σ – algebras $\mathcal{B}_i'$, $\mathcal{B}_i''$ are generated by the open balls and the open sets of B_i, respectively. In our case $B_2 = \mathbb{R}$, $\mathcal{B}_2' = \mathcal{B}_2''$ and $B_1 = D[0,\tau]^3$ is endowed with the open ball topology.

According to the Appendix B Lemma, if

$$t_n^{-1}\{\varphi_t(x_n, y_n, z_n) - \varphi_t(x_0, y_0, z_0)\} - d\varphi_t(x_0, y_0, z_0).(h,k,l) \to 0 \text{ as } n \to \infty \quad (3.33)$$

then φ_t can be extended to $D[0,\tau]^3$ in such a way that it is differentiable at (x_0, y_0, z_0), with derivative $d\varphi_t(x_0, y_0, z_0).(h,k,l)$ as defined in (3.31), so our Lemma holds. Therefore we will prove (3.33).

By the continuity of $d\varphi_t(x_0, y_0, z_0)$, for (3.33) it is sufficient to prove that

$$S_n \underset{notation}{\overset{by}{=}} t_n^{-1}\{\varphi_t(x_n, y_n, z_n) - \varphi_t(x_0, y_0, z_0)\} - d\varphi_t(x_0, y_0, z_0).(h_n,k_n,l_n) \to 0$$

as $n \to \infty$. This sequence is equal to

$$t_n^{-1}\int_0^t x_n y_n dz_n - t_n^{-1}\int_0^t x_0 y_0 dz_0 - \int_0^t h_n y_0 dz_0 - \int_0^t x_0 k_n dz_0 - \int_0^t x_0 y_0 dl_n .$$

We expand

$$t_n^{-1}\int_0^t x_n y_n dz_n = \int_0^t (t_n^{-1}x_0 + h_n)(y_0 + t_n k_n) d(z_0 + t_n l_n) =$$

$$= t_n^{-1}\int_0^t x_0 y_0 dz_0 + \int_0^t x_0 y_0 dl_n + \int_0^t x_0 k_n dz_0 + t_n\int_0^t x_0 k_n dl_n + \int_0^t h_n y_0 dz_0 + t_n\int_0^t h_n y_0 dl_n +$$

$$+ t_n\int_0^t h_n k_n dz_0 + t_n^2\int_0^t h_n k_n dl_n$$

Thus

$$S_n = t_n\int_0^t x_0 k_n dl_n + t_n\int_0^t h_n y_0 dl_n + t_n\int_0^t h_n k_n dz_0 + t_n^2\int_0^t h_n k_n dl_n =$$

$$= \int_0^\tau x_0 k_n d(z_n - z_0) + \int_0^\tau h_n y_0 d(z_n - z_0) + \int_0^\tau (x_n - x_0) k_n dz_0 + \int_0^\tau h_n (y_n - y_0) d(z_n - z_0) =$$
$$= T_{n1} + T_{n2} + T_{n3} + T_{n4}.$$

We will prove that $T_{ni} \to 0$ as $n \to \infty$ for all $i \in \{1,2,3,4\}$.

We start with $T_{n1} = \int_0^\tau x_0 k_n d(z_n - z_0)$.

$$\left| \int_0^\tau x_0 (k_n - k) d(z_n - z_0) \right| \le \|x_0\| \times \|k_n - k\| \times \left(\int_0^\tau | dz_n | + \int_0^\tau | dz_0 | \right) \le 2C \|x_0\| \times \|k_n - k\| \to 0,$$

because $k_n \xrightarrow{\|\cdot\|} k$ and $\|x_0\| < \infty$. The fact $\int_0^\tau | dz_n | < C$ follows from our assumption that $(x_n, y_n, z_n) \in E$. Thus it is sufficient to show

$$\int_0^\tau x_0 k d(z_n - z_0) \to 0 \text{ as } n \to \infty. \tag{3.34}$$

Let $f_0 = x_0 k$. Then f_0 is an element of $D[0,\tau]$, so for every $\varepsilon > 0$ there exists $f_0' \in D[0,\tau]$ such that f_0' is a step function with a finite number (say N) of jumps and $\|f_0 - f_0'\| \le \varepsilon$. We have

$$\left| \int_0^\tau (f_0 - f_0') d(z_n - z_0) \right| \le \|f_0 - f_0'\| \left(\int_0^\tau | dz_n | + \int_0^\tau | dz_0 | \right) \le 2C\varepsilon. \tag{3.35}$$

By partial integration

$$\left| \int_0^\tau f_0' d(z_n - z_0) \right| \le 2 \|f_0'\| \times \|z_n - z_0\| + \|z_n - z_0\| \times \int_0^\tau | df_0' |.$$

The mapping f_0' is a step function with N jump points, say $s_1, \dots, s_N$, so

$$\int_0^\tau | df_0' | \le \sum_{i=1}^N \left| f_0'(s_n) - f_0'(s_n -) \right| \le 2N \|f_0'\|.$$

Thus

$$\left|\int_0^{\tau} f_0' d(z_n - z_0)\right| \leq 2(N+1)\|f_0'\| \times \|z_n - z_0\| \tag{3.36}$$

and this quantity converges to zero because $\|f_0'\| < \infty$ and

$$\|z_n - z_0\| = t_n\|l_n\| \leq t_n(\|l_n - l\| + \|l\|) \to 0 \text{ as } n \to \infty.$$

Notice that similarly $\|x_n - x_0\| \to 0$ and $\|y_n - y_0\| \to 0$ as $n \to \infty$.

By (3.35) and (3.36) we obtain $\limsup_{n\to\infty}\left|\int_0^{\tau} f_0 d(z_n - z_0)\right| \leq 2C\varepsilon$. Since ε was arbitrary chosen, (3.34) is proved, so the convergence $T_{n1} \to 0$ follows.

The proof of $T_{n2} \to 0$ is identical.

For $T_{n3} = \int_0^{\tau}(x_n - x_0)k_n dz_0$ we have that

$$|T_{n3}| \leq \|x_n - x_0\| \times \|k_n\| \times \int_0^{\tau} | dz_0 |.$$

Because $\|x_n - x_0\| \to 0$, $\|k_n\| \leq \|k_n - k\| + \|k\| < \infty$ and $\int_0^{\tau} | dz_0 | < C$, the convergence $T_{n3} \to 0$ is an immediate consequence.

For $T_{n4} = \int_0^{\tau} h_n(y_n - y_0)d(z_n - z_0)$,

$$|T_{n4}| \leq \|h_n\| \times \|y_n - y_0\| \times (\int_0^{\tau} | dz_n | + \int_0^{\tau} | dz_0 |) \leq 2C\|h_n\| \times \|y_n - y_0\| \to 0 \text{ as } n \to \infty.$$

This completes the proof of Lemma. ■

Recall that in Section 3.2.3 we considered $C_i' = \left(C_{1i},\ldots,C_{in_i}\right)'$ to denote the vector of all h to j transition costs related to the i-th subject. The cost vectors $C_1,\ldots,C_n$ are assumed to be independent.

For technical reasons the following extra-assumptions are considered:

EA.1 $c_{hj}(.\mid Z_0)$ is of finite variation over $[0,\tau]$. We write $\int_0^\tau \left|dc_{hj}(s\mid Z_0)\right| < \infty$.

EA.2 The cost vectors $C_1,\ldots,C_n$ are independent of $(N_i,Y_i,Z_i), 1\le i\le n$.

Comments:

1) In the proof of the next theorem we will need

$$\int_0^\tau \left|d(e^{-rs}c_{hj}(s\mid Z_0)P_{ih}(0,s\mid Z_0))\right| < \infty. \tag{3.37}$$

The function $s \to e^{-rs}$ is monotone, so of bounded variation on $[0,\tau]$. The matrix-valued process $P(0,.\mid Z_0)$ is (componentwise) right continuous with left hand limits and of bounded variation (see Theorem II.6.1, p90, Andersen *et al.* (1993)[29]). By assumption EA.1, $c_{hj}(.\mid Z_0)$ is of finite variation over $[0,\tau]$. A product of finite variation functions is also a function of finite variation, so (3.37) follows.

2) The estimator $\hat{c}_{hj}(.\mid Z_0)$ of $c_{hj}(.\mid Z_0)$ was obtained from the cost vectors $C_1,\ldots,C_n$. The estimators $\hat{A}_{hj}(.\mid Z_0), \hat{P}_{ih}(0,.\mid Z_0)$ were calculated from $(N_i,Y_i,Z_i), 1\le i\le n$. By assumption EA.2, we can consider that $\hat{c}_{hj}(.\mid Z_0)$ is independent of $\left(\hat{A}_{hj}(.\mid Z_0), \hat{P}_{ih}(0,.\mid Z_0)\right)'$. This implies that

$n^{1/2}\left(\hat{c}_{hj}(.\,|\,Z_0)-c_{hj}(.\,|\,Z_0)\right)$ is asymptotically independent of

$$n^{1/2}\begin{pmatrix}\hat{A}_{hj}(.\,|\,Z_0)-A_{hj}(.\,|\,Z_0)\\ \hat{P}_{ih}(0,.\,|\,Z_0)-P_{ih}(0,.\,|\,Z_0)\end{pmatrix}.\square \tag{3.38}$$

Theorem 4

Under the assumptions A.0-A.7 and the extra-assumptions EA.1 and EA.2, for a fixed time t:

$$n^{1/2}\left(M\hat{P}V_{hj}^{(1)}(t\,|\,i,Z_0)-MPV_{hj}^{(1)}(t\,|\,i,Z_0)\right)=$$

$$=n^{1/2}[\varphi_t\left(e^{-r\cdot}\hat{c}_{hj}(.\,|\,Z_0),\hat{P}_{ih}(0,.\,|\,Z_0),\hat{A}_{hj}(.\,|\,Z_0)\right)-$$

$$-\varphi_t\left(e^{-r\cdot}c_{hj}(.\,|\,Z_0),P_{ih}(0,.\,|\,Z_0),A_{hj}(.\,|\,Z_0)\right)]\xrightarrow{D}$$

$$\xrightarrow{D}d\varphi_t\left(e^{-r\cdot}c_{hj}(.\,|\,Z_0),P_{ih}(0,.\,|\,Z_0),A_{hj}(.\,|\,Z_0)\right).\left(e^{-r\cdot}c_{hj0}(.\,|\,Z_0),U_{ih}(0,.\,|\,Z_0),U_{hj}^{*}(.\,|\,Z_0)\right)=$$

$$=\int_0^t e^{-rs}c_{hj0}(s\,|\,Z_0)P_{ih}(0,s\,|\,Z_0)dA_{hj}(s\,|\,Z_0)+\int_0^t e^{-rs}c_{hj}(s\,|\,Z_0)U_{ih}(0,s\,|\,Z_0)dA_{hj}(s\,|\,Z_0)+$$

$$+\int_0^t e^{-rs}c_{hj}(s\,|\,Z_0)P_{ih}(0,s\,|\,Z_0)dU_{hj}^{*}(s\,|\,Z_0)\underset{notation}{\overset{by}{=}}P(t).\square$$

Proof Theorem 4:

The mapping $\varphi_t:E\to\mathbb{R}$ is defined by $\varphi_t(x,y,z)=\int_0^t x(s)y(s)dz(s)$, where

$E=\{(x,y,z)\in D[0,\tau]^3:\int_0^\tau |dz|\le C\}$ and $C=A_{hj}(\tau\,|\,Z_0)+1<\infty$ by assumption A.3.

Let $x_0(s) = e^{-rs} c_{hj}(s \mid Z_0), y_0(s) = P_{ih}(0, s \mid Z_0), z_0(s) = A_{hj}(s \mid Z_0)$. All $x_0, y_0, z_0 \in D[0,\tau]$ and

$$\int_0^\tau \left| dA_{hj}(s \mid Z_0) \right| = \exp(\beta_0' Z_{hj0}) \int_0^\tau \left| dA_{hj0}(s) \right| = \exp(\beta_0' Z_{hj0}) A_{hj0}(\tau) = A_{hj}(\tau \mid Z_0) < C.$$

Therefore $(x_0, y_0, z_0) \in E$. By (3.37), we have that $\int_0^\tau \left| d(x_0 y_0) \right| < \infty$.

The previous Lemma implies that φ_t can be extended to $D[0,\tau]^3$ so as to be Hadamard differentiable at (x_0, y_0, z_0), with derivative

$$d\varphi_t(x_0, y_0, z_0).(h,k,l) = \int_0^t h y_0 dz_0 + \int_0^t x_0 k dz_0 + \int_0^t x_0 y_0 dl,$$

where the integral with respect to l is defined by the integration by parts formula if l is not of finite variation. Denote by φ_t^E the extension of φ_t to $D[0,\tau]^3$.

Define $\hat{x}_n(s) = e^{-rs} \hat{c}_{hj}(s \mid Z_0), \hat{y}_n(s) = \hat{P}_{ih}(0, s \mid Z_0), \hat{z}_n(s) = \hat{A}_{hj}(s \mid Z_0), s \in [0,\tau]$.

We have $(\hat{x}_n, \hat{y}_n, \hat{z}_n) \in D[0,\tau]^3$ for every n and

$$P(\int_0^\tau d \left| z_n \right| < C) = P(\int_0^\tau d \mid \hat{A}_{hj0}(s, \hat{\beta}) \mid < C) = P(\hat{A}_{hj}(\tau \mid Z_0) < C) \to 1 \text{ as } n \to \infty$$

because $\hat{A}_{hj}(\tau \mid Z_0) \xrightarrow{P} A_{hj}(\tau \mid Z_0) < C$. Thus $(\hat{x}_n, \hat{y}_n, \hat{z}_n) \in E$ with probability tending to one.

We have

$$n^{1/2}(\hat{x}_n(.) - x_0(.)) \xrightarrow{D} X_0(.) = e^{-r.} c_{hj0}(. \mid Z_0),$$

$$n^{1/2}(\hat{y}_n(.) - y_0(.)) \xrightarrow{D} Y_0(.) = U_{ih}(0,. \mid Z_0),$$

$$n^{1/2}(\hat{z}_n(.) - z_0(.)) \xrightarrow{D} Z_0(.) = U_{hj}^*(. \mid Z_0).$$

The processes X_0, Y_0, Z_0 are Gaussian and hence have versions that are almost surely continuous. Let $C[0,\tau]$ the set of continuous functions on $[0,\tau]$. The subset $C[0,\tau]^3 \subset D[0,\tau]^3$ is separable, so (X_0, Y_0, Z_0) has separable support.

Because $\hat{P}(0,.\,|\,Z_0) = \prod_{(0,.]} (I + d\hat{A}(.\,|\,Z_0))$ and $P(0,.\,|\,Z_0) = \prod_{(0,.]} (I + dA(.\,|\,Z_0))$, the matrices $\hat{P}(0,.\,|\,Z_0), P(0,.\,|\,Z_0)$ are functionals of $\hat{A}(.\,|\,Z_0), A(.\,|\,Z_0)$, respectively. It can be easily shown that jointly:

$$n^{1/2}\begin{pmatrix} \hat{P}_{ih}(0,.\,|\,Z_0) - P_{ih}(0,.\,|\,Z_0) \\ \hat{A}_{hj}(.\,|\,Z_0) - A_{hj}(.\,|\,Z_0) \end{pmatrix} \xrightarrow{D} \begin{pmatrix} U_{ih}(0,.\,|\,Z_0) \\ U^*_{hj}(.\,|\,Z_0) \end{pmatrix}.$$

Consequently, by (3.38),

$$n^{1/2}\left[(\hat{x}_n(.), \hat{y}_n(.), \hat{z}_n(.))' - (x_0(.), y_0(.), z_0(.))' \right] \xrightarrow{D} (X_0(.), Y_0(.), Z_0(.))'.$$

By the Functional Delta Method stated in Appendix B,

$n^{1/2}\left(\varphi_t(\hat{x}_n, \hat{y}_n, \hat{z}_n) - \varphi_t(x_0, y_0, z_0)\right) \xrightarrow{D} d\varphi_t^E(x_0, y_0, z_0).(X_0, Y_0, Z_0) = P(t)$ defined in the theorem statement. This completes the proof of Theorem 4. ■

By (3.26)-(3.30), we can write the limiting process $P(t)$ as

$$P(t) = \zeta_1' \int_0^t e^{-rs} Z^*_{01}(s) P_{ih}(0, s\,|\,Z_0) \alpha_{hj0}(s) \exp(\beta_0' Z_{hj0}) ds$$

$$+\xi' \int_0^t e^{-rs} c_{hj}(s\,|\,Z_0) F_{ih}(0, s\,|\,Z_0) \alpha_{hj0}(s) \exp(\beta_0' Z_{hj0}) ds$$

$$+\sum_{g=1}^{k} \sum_{l \neq g} \int_0^t e^{-rs} c_{hj}(s\,|\,Z_0) \left[\int_0^s P_{ig}(0, u\,|\,Z_0) \left(P_{lh}(u, s\,|\,Z_0) - P_{gh}(u, s\,|\,Z_0) \right) dU^*_{0gl}(u) \right]$$

$$\alpha_{hj0}(s) \exp(2\beta_0' Z_{hj0}) ds$$

$$+\xi' \int_0^t e^{-rs} c_{hj}(s \mid Z_0) P_{ih}(0,s \mid Z_0) \exp(\beta_0' Z_{hj0}) \big(Z_{hj0} - e_{hj}(s,\beta_0) \big) \alpha_{hj0}(s) ds$$

$$+ \int_0^t e^{-rs} c_{hj}(s \mid Z_0) P_{ih}(0,s \mid Z_0) \exp(\beta_0' Z_{hj0}) dU_{0hj}^*(s).$$

By Theorem 3, $\{U_{0hj}^*, (h,j) \in E^*\}$ are independent, continuous Gaussian martingales. The integrals with respect to U_{0hj}^* are Ito integrals and the theory from Appendix C can be applied. Therefore, by the Fubini-type Theorem from Appendix C applied to the bounded functions

$$H_{gl}(s,u \mid Z_0) = I_{(0,s]}(u) e^{-rs} c_{hj}(s \mid Z_0) P_{ig}(0,u \mid Z_0) \big(P_{lh}(u,s \mid Z_0) - P_{gh}(u,s \mid Z_0) \big)$$

and the finite measure defined by $\mu(0,s] = \int_0^s \alpha_{hj0}(u) du$, the third term of the previous sum can be written as

$$\exp(2\beta_0' Z_{hj0}) \sum_{g=1}^{k} \sum_{l \neq g} \int_0^t P_{ig}(0,u \mid Z_0) \times$$
$$\times \left[\int_u^t \big(P_{lh}(u,s \mid Z_0) - P_{gh}(u,s \mid Z_0) \big) \; e^{-rs} c_{hj}(s \mid Z_0) \alpha_{hj0}(s) ds \right] dU_{0gl}^*(u).$$

Therefore the process $P(t)$ has the form:

$$P(t) = P_1(t) + P_2(t) + P_3(t) + P_4(t),$$

where $P_1(t) = \zeta_1' T_1(t)$,

$$P_2(t) = \xi' T_2(t),$$

$$P_3(t) = \sum_{g=1}^{k} \sum_{\substack{l \neq g \\ (l,g) \neq (h,j)}} \int_0^t f_{3gl}(u) dU_{0gl}^*(u),$$

$$P_4(t) = \int_0^t f_{4hj}(u) dU_{0hj}^*(u)$$

and we denote by $T_1(.), T_2(.), f_{3gl}(.), f_{4hj}(.)$ the following expressions:

$$T_1(t) = \exp(\beta_0' Z_{hj0}) \int_0^t e^{-rs} Z_{01}^*(s) P_{ih}(0,s \mid Z_0) \alpha_{hj0}(s) ds$$

$$T_2(t) = \exp(\beta_0' Z_{hj0}) \times$$
$$\times \int_0^t e^{-rs} c_{hj}(s \mid Z_0) \left[F_{ih}(0,s \mid Z_0) + P_{ih}(0,s \mid Z_0) \left(Z_{hj0} - e_{hj}(s, \beta_0) \right) \right] \alpha_{hj0}(s) ds$$

$$f_{3gl}(u) = \exp(2\beta_0' Z_{hj0}) P_{ig}(0,u \mid Z_0) \left[\int_u^t \left(P_{lh}(u,s \mid Z_0) - P_{gh}(u,s \mid Z_0) \right) e^{-rs} c_{hj}(s \mid Z_0) \alpha_{hj0}(s) ds \right]$$

$$f_{4hj}(u) = \exp(2\beta_0' Z_{hj0}) P_{ih}(0,u \mid Z_0) \int_u^t \left(P_{jh}(u,s \mid Z_0) - P_{hh}(u,s \mid Z_0) \right) e^{-rs} c_{hj}(s \mid Z_0) \alpha_{hj0}(s) ds +$$
$$+ \exp(\beta_0' Z_{hj0}) e^{-ru} c_{hj}(u \mid Z_0) P_{ih}(0,u \mid Z_0).$$

The functions $T_1(.)$, $T_2(.)$, $f_{3gl}(.)$, $f_{4hj}(.)$ are deterministic. Property iv) stated in Appendix C and the fact that $\left\{ U_{0hj}^*, (h,j) \in E^* \right\}$ are independent, Gaussian martingales imply that

$P_3(t) + P_4(t)$ is normally distributed, with mean zero and variance

$$\sum_{g=1}^{k} \sum_{\substack{l \neq g \\ (l,g) \neq (h,j)}} \int_0^t f_{3gl}^2(u) d\left\langle U_{0gl}^* \right\rangle(u) + \int_0^t f_{4hj}^2(u) d\left\langle U_{0hj}^* \right\rangle(u). \tag{3.39}$$

By (3.30), ζ_1 is multivariate normally distributed, with mean vector 0 and covariance $\Sigma(\alpha)^{-1}$. Thus

$P_1(t)$ is normally distributed, with mean zero and variance $T_1(t)' \Sigma(\alpha)^{-1} T_1(t)$. (3.40)

Similarly, by (3.26),

$P_2(t)$ is normally distributed, with mean zero and variance $T_2(t)' \Sigma_\tau^{-1} T_2(t)$. (3.41)

By Theorem 3, ξ and $\left\{ U_{0hj}^*(.), (h,j) \in E^* \right\}$ are independent, so

$P_2(t)$ is independent of $P_3(t)+P_4(t)$. (3.42)

In Section 3.2.3 the regression parameter estimator $\hat{\beta}$ was computed from the cost vectors $C_1,...,C_n$. Then, by assumption EA.2, we can consider ζ_1, the asymptotic limit of $n^{1/2}(\hat{\beta}-\beta)$, independent of ξ and $\left\{U^*_{0hj}(.),(h,j)\in E^*\right\}$. Consequently,

$P_1(t)$ is independent of $P_2(t)$ and $P_3(t)+P_4(t)$. (3.43)

By (3.39)-(3.43), $P(t)$ is normally distributed, with mean zero and variance

$$Var(P(t)) = T_1(t)'\Sigma(\alpha)^{-1}T_1(t) + T_2(t)'\Sigma_\tau^{-1}T_2(t) +$$

$$+\sum_{g=1}^{k}\sum_{\substack{l\neq g\\(l,g)\neq(h,j)}}\int_0^t f_{3gl}^2(u)d\left\langle U^*_{0gl}\right\rangle(u) + \int_0^t f_{4hj}^2(u)d\left\langle U^*_{0hj}\right\rangle(u).$$

By Theorem 3, for $h\neq j$: $\left\langle U^*_{0hj}\right\rangle(t) = \int_0^t \frac{\alpha_{hj0}(u)}{s_{hj}^{(0)}(u,\beta_0)}du$. A consistent estimator of the variance of $P(t)$ is obtained replacing all unknown quantities by their corresponding consistent estimators.

3.3.3 Asymptotic Distribution of the Mean Sojourn Cost

The asymptotic normality of $n^{1/2}\left(M\hat{P}V_h^{(2)}(s,d\,|\,i,Z_0) - MPV_h^{(2)}(s,d\,|\,i,Z_0)\right)$ will be assessed also by the Functional Delta Method. The entry time s and duration time d are considered fixed throughout this sub-section.

Consider the functional $\psi_{s,d}: D[0,\tau]^2 \to \mathbb{R}$, $\psi_{s,d}(x,y) = \int_s^{s+d} x(u)y(u)du$. The mean present value $MPV_h^{(2)}(s,d\,|\,i,Z_0)$ can be written as

$$MPV_h^{(2)}(s,d \mid i,Z_0) = \psi_{s,d}(x_0, y_0) = \psi_{s,d}(e^{-r\cdot}b_h(.\mid Z_0), P_{ih}(0,.\mid Z_0)).$$

We will show that $\psi_{s,d}$ is Hadamard differentiable in (x_0, y_0).

Two convergence results from the previous sections will be needed. One is that $n^{1/2}\left(\hat{P}_{ih}(0,.\mid Z_0) - P_{ih}(0,.\mid Z_0)\right)$ converges weakly to the process

$$U_{ih}(0,.\mid Z_0) = U_{1ih}(0,.\mid Z_0) + U_{2ih}(0,.\mid Z_0),$$

where $U_{1ih}(0,.\mid Z_0)$, $U_{2ih}(0,.\mid Z_0)$ are described in (3.28) and (3.29). The second is that, by (3.23), $n^{1/2}\left(\hat{b}_h(.\mid Z_0) - b_h(.\mid Z_0)\right)$ converges weakly to the Gaussian process $b_{h0}(.\mid Z_0)$ described in Section 3.2.4:

$$b_{h0}(.\mid Z_0) = \zeta_2' Z_{02}^*(.), \tag{3.44}$$

where $Z_{02}^*(.)$ is a deterministic vector function, $\zeta_2 \sim MVN(0, \Sigma(\alpha)^{-1})$ and

$$\Sigma(\alpha) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i' V_i^{-1}(\alpha) X_i.$$

First we prove the following lemma:

Lemma

For a fixed $s \in [0,\tau)$ and $d > 0$ such that $s + d \le \tau$ we define $\psi_{s,d} : D[0,\tau]^2 \to \mathbb{R}$ by $\psi_{s,d}(x,y) = \int_s^{s+d} x(u)y(u)du$. Then $\psi_{s,d}$ is Hadamard differentiable at every point $(x_0, y_0) \in D[0,\tau]^2$, with derivative

$$d\psi_{s,d}(x_0, y_0).(h,k) = \int_s^{s+d} h(u)y_0(u)du + \int_s^{s+d} x_0(u)k(u)du. \square$$

Proof Lemma:

Let (x_0, y_0) be an arbitrary element of the space $D[0,\tau]^2$. It is straightforward that $d\psi_{s,d_t}(x_0, y_0)$ is a continuous, linear mapping.

Consider an arbitrary $(h,k) \in D[0,\tau]^2$ and the sequences $t_n \in \mathbb{R}^+, h_n, k_n \in D[0,\tau]$ that satisfy $t_n \to 0,\ h_n \xrightarrow{\|\cdot\|} h,\ k_n \xrightarrow{\|\cdot\|} k$, where $\|\,.\,\|$ is the supremum norm. Define

$$x_n = x_0 + t_n h_n$$

$$y_n = y_0 + t_n k_n.$$

For each n, $(x_n, y_n) \in D[0,\tau]^2$.

We want to show that $\psi_{s,d}$ is Hadamard differentiable at (x_0, y_0), so that

$$t_n^{-1}\{\psi_{s,d}(x_n, y_n) - \psi_{s,d}(x_0, y_0)\} - d\psi_{s,d}(x_0, y_0).(h,k) \to 0 \text{ as } n \to \infty. \qquad (3.45)$$

By the continuity of $d\psi_{s,d}(x_0, y_0)$, it is sufficient to prove that

$$S_n \underset{notation}{\overset{by}{=}} t_n^{-1}\{\psi_{s,d}(x_n, y_n) - \psi_{s,d}(x_0, y_0)\} - d\psi_{s,d}(x_0, y_0).(h_n, k_n) \to 0$$

as $n \to \infty$. This sequence is equal to

$$t_n^{-1}\int_s^{s+d} x_n(u)y_n(u)du - t_n^{-1}\int_s^{s+d} x_0(u)y_0(u)du - \int_s^{s+d} h_n(u)y_0(u)du - \int_s^{s+d} x_0(u)k_n(u)du.$$

We expand

$$t_n^{-1}\int_s^{s+d} x_n(u)y_n(u)du = \int_s^{s+d}(t_n^{-1}x_0 + h_n)(u)(y_0 + t_n k_n)(u)du =$$

$$= t_n^{-1}\int_s^{s+d} x_0(u)y_0(u)du + \int_s^{s+d} h_n(u)y_0(u)du + \int_s^{s+d} x_0(u)k_n(u)du + t_n\int_s^{s+d} h_n(u)k_n(u)du.$$

As a result

$$S_n = t_n\int_s^{s+d} h_n(u)k_n(u)du = \int_s^{s+d}(x_n - x_0)(u)k_n(u)du.$$

We have that

$$|S_n| \le d\,\|x_n - x_0\| \times \|k_n\| \to 0 \text{ as } n \to \infty$$

because $\|x_n - x_0\| \to 0$, $\|k_n - k\| \to 0$ and $\|k_n\| \le \|k_n - k\| + \|k\| < \infty$. Therefore (3.45) follows. This completes the proof of the stated Lemma.■

In Section 3.2.4 we considered $C_i' = \left(C_{1i}, \ldots, C_{in_i}\right)'$ to denote the vector of all total observed costs incurred in sojourns in state h by the i-th individual. The cost vectors $C_1, \ldots, C_n$ are assumed independent. The same notations are used in both Sections 3.2.3 and 3.2.4 but in the context we are easily able to recognize between them.

A similar extra-assumption as EA.2 is considered:

EA.3 The cost vectors $C_1, \ldots, C_n$ are independent of $(N_i, Y_i, Z_i), 1 \le i \le n$.

The estimator $\hat{b}_h(.\,|\,Z_0)$ of $b_h(.\,|\,Z_0)$ was obtained from the cost vectors $C_1, \ldots, C_n$ and $\hat{P}_{ih}(0,.\,|\,Z_0)$ is calculated from $(N_i, Y_i, Z_i), 1 \le i \le n$. By assumption EA.3, we can consider that $\hat{b}_h(.\,|\,Z_0)$ is independent of $\hat{P}_{ih}(0,.\,|\,Z_0)$. This implies that

$$n^{1/2}\left(\hat{b}_h(.\,|\,Z_0) - b_h(.\,|\,Z_0)\right) \text{ is asymptotically independent of}$$

$$n^{1/2}\left(\hat{P}_{ih}(0,.\,|\,Z_0) - P_{ih}(0,.\,|\,Z_0)\right). \tag{3.46}$$

Theorem 5

Under the assumptions A.0-A.7 and the extra-assumption EA.3, for a fixed entry time s and fixed duration d:

$$n^{1/2}\left(M\hat{P}V_h^{(2)}(s,d\,|\,i,Z_0)-MPV_h^{(2)}(s,d\,|\,i,Z_0)\right)=$$

$$=n^{1/2}\left[\psi_{s,d}\left(e^{-r\cdot}\hat{b}_h(.\,|\,Z_0),\hat{P}_{ih}(0,.\,|\,Z_0)\right)-\psi_{s,d}\left(e^{-r\cdot}b_h(.\,|\,Z_0),P_{ih}(0,.\,|\,Z_0)\right)\right]\xrightarrow{D}$$

$$\xrightarrow{D}d\psi_{s,d}\left(e^{-r\cdot}b_h(.\,|\,Z_0),P_{ih}(0,.\,|\,Z_0)\right).\left(e^{-r\cdot}b_{h0}(.\,|\,Z_0),U_{ih}(0,.\,|\,Z_0)\right)=$$

$$=\int_s^{s+d}e^{-ru}b_{h0}(u\,|\,Z_0)P_{ih}(0,u\,|\,Z_0)du+\int_s^{s+d}e^{-ru}b_h(s\,|\,Z_0)U_{ih}(0,u\,|\,Z_0)du\underset{notation}{\overset{by}{=}}R(s,d)\,.\square$$

Proof Theorem 5:

We defined $\psi_{s,d}:D[0,\tau]^2\to\mathbb{R}$ by $\psi_{s,d}(x,y)=\int_s^{s+d}x(u)y(u)du$. Let $x_0(u)=e^{-ru}b_h(u\,|\,Z_0)$ and $y_0(u)=P_{ih}(0,u\,|\,Z_0)$, where both x_0,y_0 are elements of $D[0,\tau]$.

By the previous Lemma, $\psi_{s,d}$ is Hadamard differentiable at (x_0,y_0), with derivative $d\psi_{s,d}(x_0,y_0).(h,k)=\int_s^{s+d}h(u)y_0(u)du+\int_s^{s+d}x_0(u)k(u)du$.

For every n define $(\hat{x}_n,\hat{y}_n)\in D[0,\tau]^2$, $\hat{x}_n(u)=e^{-ru}\hat{b}_h(u\,|\,Z_0)$, $\hat{y}_n(u)=\hat{P}_{ih}(0,u\,|\,Z_0)$, $u\in[0,\tau]$. By (3.46),

$$n^{1/2}\left[\begin{pmatrix}\hat{x}_n(.)\\\hat{y}_n(.)\end{pmatrix}-\begin{pmatrix}x_0(.)\\y_0(.)\end{pmatrix}\right]\xrightarrow{D}\begin{pmatrix}X_0(.)\\Y_0(.)\end{pmatrix},$$

where $X_0(.)=e^{-r\cdot}b_{h0}(.\,|\,Z_0)$, $Y_0(.)=U_{ih}(0,.\,|\,Z_0)$.

The processes X_0,Y_0 are Gaussian, so they have versions that are almost surely continuous. Let $C[0,\tau]$ the set of continuous functions on $[0,\tau]$. The subset $C[0,\tau]^2\subset D[0,\tau]^2$ is separable, so (X_0,Y_0) has separable support.

By the Functional Delta Method (see Appendix B),

$n^{1/2}\left(\psi_{s,d}(\hat{x}_n,\hat{y}_n)-\psi_{s,d}(x_0,y_0)\right)\xrightarrow{D} d\psi_{s,d}(x_0,y_0).(X_0,Y_0)=R(s,d)$, so Theorem 5 is proved. ■

By (3.28), (3.29) and (3.44), the limiting process $R(s,d)$ can be written as

$$R(s,d)=R_1(s,d)+R_2(s,d)+R_3(s,d),$$

where $R_1(s,d)=\zeta_2{}'S_1(s,d)$,

$$R_2(s,d)=\xi' S_2(s,d),$$

$$R_3(s,d)=$$
$$=\sum_{g=1}^{k}\sum_{l\neq g}\int_s^{s+d}e^{-ru}b_h(u\,|\,Z_0)\left[\int_0^u P_{ig}(0,v\,|\,Z_0)\left(P_{lh}(v,u\,|\,Z_0)-P_{gh}(v,u\,|\,Z_0)\right)dU^*_{0gl}(v)\right]du$$

and we denote by $S_1(s,d)$, $S_2(s,d)$ the expressions:

$$S_1(s,d)=\int_s^{s+d}e^{-ru}Z^*_{02}(u)P_{ih}(0,u\,|\,Z_0)du,$$

$S_2(s,d)=\int_s^{s+d}e^{-ru}b_h(u\,|\,Z_0)F_{ih}(0,u\,|\,Z_0)du$. (See p139-140 for the definition of $F_{ih}(0,u\,|\,Z_0)$.)

By the Fubini-type Theorem from Appendix C, applied to the bounded functions

$H_{gl}(u,v)=I_{[s,s+d]}(u)e^{-ru}b_h(u\,|\,Z_0)I_{[0,u]}(v)P_{ig}(0,v\,|\,Z_0)\left(P_{lh}(v,u\,|\,Z_0)-P_{gh}(v,u\,|\,Z_0)\right)$,

$u,v\in[0,\tau]$ and the Lebesgue measure on $[0,\tau]$, $R_3(s,d)$ can be written as

$$R_3(s,d)=\exp(\beta_0'Z_{hj0})\sum_{g=1}^{k}\sum_{l\neq g}\int_0^\tau P_{ig}(0,v\,|\,Z_0)\times$$
$$\times\left[\int_0^\tau I_{[s,s+d]}(u)I_{[v,\tau]}(u)\left(P_{lh}(v,u\,|\,Z_0)-P_{gh}(v,u\,|\,Z_0)\right)e^{-ru}b_h(u\,|\,Z_0)du\right]dU^*_{0gl}(v),$$

so $R_3(s,d)=\sum_{g=1}^{k}\sum_{l\neq g}\int_0^{s+d} S_{3gl}(v)dU_{0gl}^*(v)$,

where

$$S_{3gl}(v)=\exp(\beta_0' Z_{hj0})P_{ig}(0,v\,|\,Z_0)\times$$
$$\times\int_s^{s+d} I_{[v,s+d]}(u)\Big(P_{lh}(v,u\,|\,Z_0)-P_{gh}(v,u\,|\,Z_0)\Big)\, e^{-ru}b_h(u\,|\,Z_0)du.$$

Using the same approach from Section 3.3.3, we obtain that

$R(s,d)$ is normally distributed, with mean zero and variance

$$Var(R(s,d))=S_1(s,d)'\Sigma(\alpha)^{-1}S_1(s,d)+S_2(s,d)'\Sigma_\tau^{-1}S_2(s,d)+$$
$$+\sum_{g=1}^{k}\sum_{l\neq g}\int_0^{s+d} S_{3gl}^2(v)d\left\langle U_{0gl}^*\right\rangle(v),$$

where for $g\neq l$: $\left\langle U_{0gl}^*\right\rangle(t)=\int_0^t \frac{\alpha_{gl0}(u)}{s_{gl}^{(0)}(u,\beta_0)}du$. A consistent estimator of the variance of $R(s,d)$ is obtained replacing all unknown quantities by their corresponding consistent estimators.

Comments

1) The technique proposed in this chapter separates the temporal dynamics of movement between states from the actual expenses. Transition probabilities and intensities that capture the former are estimated by Markov models, while the level of expense is modeled through mixed models.

2) Consider there are only two states: the initial state '0' and the state we label as '1'. Denote by T the random time of transition from the initial state to the state '1'. For a

given profile Z_0, we have that the mean present value of all expenditures in $(0,t]$ associated with expenditures in state '0' is

$$MPV_0^{(2)}(t \mid Z_0) = \int_0^t e^{-ru} b_0(u \mid Z_0) S(u \mid Z_0) du,$$

where $S(u \mid Z_0) = P_{00}(0,u \mid Z_0) = P(T \geq u \mid Z_0)$ is the survival function and $b_0(u \mid Z_0) = E(B_0(u) \mid T > u, Z_0)$. Under the assumption that T is independent of the rate process $\{B_0(u), u > 0\}$, $b_0(u \mid Z_0) = E(B_0(u) \mid Z_0)$ is the expected rate of the accumulating cost. Therefore, if '0' and '1' are labels for the states of a patient being 'in-hospital' and 'discharged', the model described in Chapter 3 for sojourn costs with no discounting reduces to the model proposed in Chapter 2.

APPENDIX A

EXTENSION OF SLLN ON $D_E([0,1]^2)$

Let $I_2 = [0,1]^2$ and $(E, \| . \|)$ a separable Banach space. Following Neuhaus (1971)[70] we will introduce the space $D_E(I_2)$.

Let $| . |$ be the maximum norm in $\mathbb{R}^2$. For $A \subset \mathbb{R}^2$, $\bar{A}$ denotes the closure and $\overset{o}{A}$ the interior of A in the $\mathbb{R}^2$-topology. Let $\mathcal{P} = \{\rho = (\rho_1, \rho_2); \rho_1, \rho_2 \in \{0,1\}\}$ the set consisting of the four vertices of I_2.

Consider $t = (t_1, t_2) \in I_2, \rho = (\rho_1, \rho_2) \in \mathcal{P}$. We define the quadrants $Q(\rho, t)$ and $\tilde{Q}(\rho, t)$ in I_2 with vertex t by:

$$Q(\rho,t) = I(\rho_1, t_1) \times I(\rho_2, t_2),$$

where $I(0,t_k) = [0,t_k), I(1,t_k) = (t_k, 1]$, $k \in \{1,2\}$ and

$$\tilde{Q}(\rho,t) = \tilde{I}(\rho_1, t_1) \times \tilde{I}(\rho_2, t_2),$$

where $\tilde{I}(0,t_k) = \begin{cases} [0,t_k) & \text{if } t_k < 1 \\ [0,1] & \text{if } t_k = 1 \end{cases}$, $\tilde{I}(0,t_k) = \begin{cases} [t_k,1] & \text{if } t_k < 1 \\ \Phi & \text{if } t_k = 1 \end{cases}$, $k \in \{1,2\}$, where Φ is the null set. Figures A.1-A.4 provide a visualization of the defined quadrants.

The following properties are immediate consequences of the above definitions:

$Q(\rho,t) \subset \tilde{Q}(\rho,t) \subset \bar{Q}(\rho,t)$;

$Q(\rho,t) = 0$ if and only if $\tilde{Q}(\rho,t) = 0$;

$\tilde{Q}(\rho,t) \cap \tilde{Q}(\rho',t) = \Phi$ if $\rho \neq \rho'$;

$\sum_{\rho\in\mathcal{P}} \tilde{Q}(\rho,t) = I_2$ for every $t \in I_2$.

Also, for every $t \in I_2$ there exists one and only one $\rho = \rho(t) \in \mathcal{P}$ (denoted σ) with $t \in \tilde{Q}(\sigma,t)$. The quadrants $\tilde{Q}(\sigma,t)$ and $Q(\sigma,t)$ are called continuity quadrants in t. For these quadrants $\overset{o}{Q}(\sigma,t) \neq \Phi$ and $\bar{Q}(\sigma,t) = \tilde{Q}(\sigma,t)$.

Definition of the "quadrant limit"

Consider the function $f: I_2 \to E$. If for the point $t \in I_2$, the vertex $\rho \in \mathcal{P}$ with $Q(\rho,t) \neq \Phi$ and for every sequence $\{t_n\} \subset Q(\sigma,t)$ with $t_n \to t$, the sequence $\{f(t_n)\}$ converges then the limit (not necessarily unique) is denoted $f(t+0_\rho)$ and it is called a **ρ-limit of f in t** or a **"quadrant limit"**. □

Definition of the space $D_E(I_2)$

The space $D_E(I_2)$ is the set of all functions $f: I_2 \to E$ for which the ρ-limit of f in t exists for every $\rho \in \mathcal{P}, t \in I_2$ for which $Q(\rho,t) \neq \Phi$ and which are "continuous from above", in the sense that $f(t) = f(t+0_\sigma)$ for every t. □

Definition of a partition generated by points of I_2

Let $t_1,\ldots,t_r \in I_2$. The collection of all rectangles R of the form:

$R = [u_1, u_1'\rangle \times [u_2, u_2'\rangle$, where $u_j, u'_j \in K_j = \{t_{1j}, ..., t_{rj}\} \cup \{0,1\}$, $u_j < u'_j$, $(u_j, u'_j) \cap K_j = \Phi$, $j \in \{1,2\}$ is called the **partition generated by** $t_1, ..., t_r$ and it is denoted $\mathcal{R} = \mathcal{R}(t_1, ..., t_r)$. The symbol "$\rangle$" means ")" or "]" if the right endpoint of the interval is less than 1 or equal to one, respectively. □

Neuhaus (1971)[70] generalized the Skorohod metrics d, d_0 on the space $D_{\mathbb{R}}[0,1]$ to the metrics d, d_0 on $D_{\mathbb{R}}(I_2)$ (actually on $D_{\mathbb{R}}(I_k), I_k = [0,1] \times ... \times [0,1]$). The space $D_{\mathbb{R}}(I_2)$ is separable and complete with respect to the metric d_0. Then, just like $D_{\mathbb{R}}(I_2)$, $D_E(I_2)$ is also separable and complete with respect to d_0, replacing the absolute value on $\mathbb{R}$ with the norm on the space E. The metrics d and d_0 are equivalent.

The characterizations of compact sets of $D_{\mathbb{R}}[0,1]$ given by Theorems 14.3 and 14.4 in Billingsley (1968)[71] and generalized by Neuhaus (1971)[70] to $D_{\mathbb{R}}(I_2)$ do not carry over directly to $D_E(I_2)$, since in E a closed, bounded set is not necessarily compact. However, the given conditions are still necessary for compactness, even if not sufficient anymore.[31]

Necessary condition for compactness

If $K \subset D_E(I_2)$ is a compact then

$$\lim_{\delta \to 0} \sup_{x \in K} w''_x(\delta) = 0,$$

where $w''_x(\delta) = \sup_{\substack{t\in[t_1,t_2] \\ |t_2-t_1|<\delta}} \min\left(\|x(t)-x(t_1)\|, \|x(t)-x(t_2)\|\right)$, with $[t_1,t_2]$ for $t_1,t_2 \in I_2$ denoting the Cartesian product $[t_{11},t_{21}]\times[t_{12},t_{22}]$. □

On $\mathbb{R}^2$ we say that $t \le u$ if and only if $t_1 \le u_1$ and $t_2 \le u_2$ (same if we replace "$\le$" by the strictly inequality sign "$<$"). This is not an well defined order relationship.

Definition of a $D_E(I_2)$-valued random variable

By a $D_E(I_2)$-valued random variable we understand a function $X = X(t,\omega)$ such that

1) for each fixed $t \in I_2$, $X(t,\omega)$ is a random variable;

2) $X(t,\omega) \in D_E(I_2)$ for almost all ω. □

For $x \in D_E(I_2)$ we define $\|x\|_d = \sup_{t\in I_2}\|x(t)\|$. Then $\left(D_E(I_2), \|\cdot\|_d\right)$ is a normed space.

The following lemma is a generalized version of Lemma 2, Rao R.R. (1963).[43]

Lemma

Let X be a $D_E(I_2)$-valued random variable such that $E\|X\|_d < \infty$. Then, for each $\varepsilon > 0$, there exists a partition of I_2 generated by some points $\tau_1,\ldots,\tau_r$ such that

$$\sup_{t,t'\in R} E\|X(t)-X(t')\| \le \varepsilon$$

for every rectangle R of the partition $\mathcal{R}(\tau_1,...,\tau_r)$. □

Proof Lemma:

For $a<b$ we define $\rho(a,b)=\sup_{t,t'\in[0,b)\setminus[0,a)} E\|X(t)-X(t')\|$. Recall that for $a\in I_2$, $[0,a)=[0,a_1)\times[0,a_2)$.

Consider the set $Diag=\{t\in I_2 : t_1=t_2\}$ endowed with the order relationships "$\leq$", "$<$", where

$t\leq u$ $(t<u)$ if and only if $t_1\leq u_1$ $(t_1<u_1)$.

Therefore we can consider the infimum or supremum of subsets of *Diag*.

Consider an arbitrary $\varepsilon>0$.

If $\rho\left((0,0),(1,1)\right)\leq\varepsilon$ then define the point $\tau_1=(1,1)$. As t "travels" on *Diag* from (0,0) to (1,1), the function $t\to\rho\left((0,0),t\right), t\in Diag$ is increasing (with respect to the order relation "$\leq$" on *Diag*) and also continuous, by the Lemma hypothesis. As a result we can define $\tau_1=\inf\left\{t\in Diag:\rho\left((0,0),t\right)>\varepsilon\right\}$.

Generally, define $\tau_j=(1,1)$ if $\rho(\tau_{j-1},1)\leq\varepsilon$ and otherwise let

$\tau_j=\inf\left\{t\in Diag:\tau_{j-1}<t,\rho\left(\tau_{j-1},t\right)>\varepsilon\right\}$.

Next we show that $\tau_j=(1,1)$ for some *j*. If this is not true, there would exist a sequence $\{t_n\}\subset Diag, \tau_n\leq t_n<\tau_{n+1}$ such that for each n

$$E\|X(t_n)-X(\tau_{n+1})\|>\varepsilon/2. \tag{A.1}$$

The sequence $\{\tau_n\}$ is increasing (with respect to the order relation "$\leq$" on *Diag*) and bounded, so there exists $\tau \in Diag$ such that $\tau_n \to \tau$. Then $X(t_n) - X(\tau_n) \to 0$ in E as $n \to \infty$. We also have that $E\|X(t_n) - X(\tau_n)\| \leq 2E\|X\|_d < \infty$. By Dominated Convergence Theorem, (A.1) is then not possible.

Consequently there exists r such that $\tau_{r+1} = (1,1)$ and we consider the partition generated by $\tau_1, ..., \tau_r$. The stated Lemma is proved. ■

Now we state and prove a Strong Law of Large Numbers (SLLN) on $D_E(I_2)$. We followed the ideas of the proof of the SLLN on $D_E[0,1]$ done by Andersen and Gill (1982).[31]

SLLN on $D_E(I_2)$

Let $X, X_1, X_2, ...$ a sequence of i.i.d. $D_E(I_2)$-valued random variables such that $E\|X\|_d < \infty$. Then

$$\left\| n^{-1} \sum_{i=1}^{n} X_i - EX \right\|_d \to 0 \text{ a.e. as } n \to \infty. \square$$

Proof SLLN:

The space $D_E(I_2)$ is separable and complete with respect to the Skorohod d_0 metric. Then any random element of $D_E(I_2)$ is tight. This result is true by the following statement:

Proposition (Theorem 1.4, p10, Billingsley (1968)[71])

If $(S,\mathcal{S})$ is a metric space with $\mathcal{S}$ the class of Borel sets in S and S is separable and complete then each probability measure on $(S,\mathcal{S})$ is tight. □

Therefore, because $E\|X\|_d < \infty$,

i) for every $\varepsilon > 0$ there exists a compact set $K \subset D_E(I_2)$ such that

$E\|X\|[X \notin K] < \varepsilon$.

Next we show that

ii) for every $\varepsilon > 0$ and every compact set $K \subset D_E(I_2)$ there exists $\delta > 0$ such that if $x \in K$ and $\alpha \le t < \beta \le \alpha + (\delta,\delta)$ then

$\|x(t) - x(\alpha)\| \le \|x(\beta + 0_{\rho_0}) - x(\alpha)\| + \varepsilon$,

where the vertex $\rho_0 = (0,0)$.

Let $\varepsilon > 0$ and $K \subset D_E(I_2)$ a compact set. Using the previously stated necessary condition for compactness, there exists $\delta > 0$ such that $\sup_{x \in K} w_x''(\delta) \le \varepsilon$. By the definition of $w_x''(\delta)$, if $x \in K$ and $\alpha \le t < \beta \le \alpha + (\delta,\delta)$ then

$$\min\left(\|x(t) - x(\alpha)\|, \|x(\beta - \varepsilon') - x(t)\|\right) \le \varepsilon$$

for every $\varepsilon' = (\varepsilon_1', \varepsilon_2'), \varepsilon_1', \varepsilon_2' > 0$ such that $\alpha \le t \le \beta - \varepsilon' < \beta$. Because $x \in D_E(I_2)$, $\lim_{\varepsilon' \to (0,0)} x(\beta - \varepsilon') = x(\beta + 0_{\rho_0})$. Consequently

$$\min\left(\|x(t) - x(\alpha)\|, \|x(\beta + 0_{\rho_0}) - x(t)\|\right) \le \varepsilon.$$

If $\|x(t) - x(\alpha)\| \le \varepsilon$ then ii) is obviously satisfied. If $\|x(\beta + 0_{\rho_0}) - x(t)\| \le \varepsilon$ then

$$\|x(t)-x(\alpha)\| \le \|x(\beta+0_{\rho_0})-x(\alpha)\| + \|x(\beta+0_{\rho_0})-x(t)\| \le \|x(\beta+0_{\rho_0})-x(\alpha)\| + \varepsilon,$$

so ii) is again verified.

The last property we prove is:

iii) for every $\varepsilon > 0$ and every $\delta > 0$ there exists a partition $\mathcal{R}$ of I_2 generated by some points $t_1,\ldots,t_{N-1}$ such that for each rectangle $R \in \mathcal{R}$, $R = [\alpha,\beta)$ we have

$$|\beta-\alpha| < \delta \text{ and } E\|X(\beta+0_{\rho_0}) - X(\alpha)\| \le \varepsilon.$$

By the stated Lemma, there exists a partition of I_2 generated by some points $\tau_1,\ldots,\tau_r \in Diag$ such that $\sup_{t,t'\in R} E\|X(t)-X(t')\| \le \varepsilon$ for every rectangle R of the partition $\mathcal{R}(\tau_1,\ldots,\tau_r)$. Taking on *Diag* intermediate points between τ_j, τ_{j+1}, we define a finer partition $\mathcal{R} = \mathcal{R}(t_1,\ldots,t_{N-1})$ of I_2 such that for every $R \in \mathcal{R}$, $R = [\alpha,\beta)$ we have $|\beta-\alpha| < \delta$ and $E\|X(t)-X(t')\| \le \varepsilon$ for all $t,t' \in R$. Then, taking ρ_0-limits of X in β and α in the previous relation (possible because $E\|X\|_d < \infty$), we obtain $E\|X(\beta+0_{\rho_0}) - X(\alpha)\| \le \varepsilon$, so iii) follows.

In the following we will use the properties i)-iii) to prove the SLLN.

Consider an arbitrary $\varepsilon > 0$. We choose a compact set K by i), a $\delta > 0$ by ii) and finally the partition $\mathcal{R} = \mathcal{R}(t_1,\ldots,t_{N-1})$ by iii).

First we show that

$$\sup_{t\in[0,1)\times[0,1)} \left\| n^{-1}\sum_{i=1}^{n} X_i(t) - EX(t)\right\| \to 0 \text{ a.e. as } n \to \infty. \qquad \text{(A.2)}$$

Let t a point of $[0,1)\times[0,1)$. Then there exists a rectangle R in the partition $\mathcal{R}$ such that $t \in R$. Denote $R = [\alpha,\beta)$.

We have that

$$\left\|n^{-1}\sum_{i=1}^{n} X_i(t) - EX(t)\right\| \le \varepsilon_n^K(t) + n^{-1}\sum_{i=1}^{n}\|X_i\|_d\,[X_i \notin K] + E\|X\|_d\,[X \notin K], \qquad \text{(A.3)}$$

where $\varepsilon_n^K(t) = \left\|n^{-1}\sum_{i=1}^{n} X_i(t)[X_i \in K] - EX(t)[X \in K]\right\|$.

The quantity

$$\varepsilon_n^K(t) \le \varepsilon_n^K(\alpha) + n^{-1}\sum_{i=1}^{n}\|X_i(t) - X_i(\alpha)\|[X_i \in K] + E\|X(t) - X(\alpha)\|[X \in K].$$

Then, by ii) and iii),

$$\begin{aligned}\varepsilon_n^K(t) &\le \varepsilon_n^K(\alpha) + n^{-1}\sum_{i=1}^{n}\left\|X_i(\beta + 0_{\rho_0}) - X_i(\alpha)\right\| + \varepsilon + E\left\|X(\beta + 0_{\rho_0}) - X(\alpha)\right\| + \varepsilon \\ &\le \varepsilon_n^K(\alpha) + n^{-1}\sum_{i=1}^{n}\left\|X_i(\beta + 0_{\rho_0}) - X_i(\alpha)\right\| + 3\varepsilon.\end{aligned}$$

We will apply a SLLN on separable Banach spaces, first proved by Mourier (1953)[45]:

SLLN on Banach spaces:

If $(X, \|\,.\,\|)$ is a separable Banach space and $\{V_n\}$ a sequence of i.i.d. random elements in X such that $E\|V_1\| < \infty$ then $\left\|n^{-1}\sum_{i=1}^{n} V_i - EV_1\right\| \to 0$ a.e. as $n \to \infty$.□

By this SLLN, $\varepsilon_n^K(\alpha) \to 0$ a.e. and by the regular SLLN (for the real valued random variables)

$$n^{-1}\sum_{i=1}^{n}\left\|X_i(\beta + 0_{\rho_0}) - X_i(\alpha)\right\| \to E\left\|X(\beta + 0_{\rho_0}) - X(\alpha)\right\| \text{ a.e. as } n \to \infty.$$

Therefore

$$\limsup_{n\to\infty}\ \sup_{t\in[0,1)\times[0,1)} \varepsilon_n^K(t) \le 0 + E\left\|X(\beta + 0_{\rho_0}) - X(\alpha)\right\| + 3\varepsilon \overset{\text{by iii)}}{\le} 4\varepsilon. \qquad \text{(A.4)}$$

We apply again the SLLN on Banach spaces and then, by (A.3),

$$\limsup_{n\to\infty} \sup_{t\in[0,1)\times[0,1)} \left\| n^{-1}\sum_{i=1}^{n} X_i(t) - EX(t) \right\| \le$$
$$\le \limsup_{n\to\infty} \sup_{t\in[0,1)\times[0,1)} \varepsilon_n^K(t) + 2E\|X\|_d [X \in K].$$

By (A.4) and property i),

$$\limsup_{n\to\infty} \sup_{t\in[0,1)\times[0,1)} \left\| n^{-1}\sum_{i=1}^{n} X_i(t) - EX(t) \right\| \le 6\varepsilon .$$

Taking ε to converge to zero, we obtain (A.2).

The following is the extension of SLLN to the space $D_E[0,1]$, proved by Andersen and Gill (1982)[31]:

SLLN on $D_E[0,1]$

Let $\{V_n\}$ a sequence of i.i.d. random elements of $D_E[0,\tau_2]$. Suppose

$E\|V_1\| = E\left(\sup_{c\in[0,\tau_2]} \|V_1(c)\| \right) < \infty$. Then $\left\| n^{-1}\sum_{i=1}^{n} V_i - EV_1 \right\| \to 0$ a.e. as $n \to \infty$.□

By this extended SLLN:

$$\sup_{t\in\{(t_1,1),t_1\in[0,1]\}} \left\| n^{-1}\sum_{i=1}^{n} X_i(t) - EX(t) \right\| =$$
$$= \sup_{t_1\in[0,1]} \left\| n^{-1}\sum_{i=1}^{n} X_i(t_1,1) - EX(t_1,1) \right\| \to 0 \text{ a.e.} \qquad \text{(A.5)}$$

because $X_i(.,1)$ are $D_E[0,1]$-valued random variables and $E\|X_1(t_1,1)\| \le E\|X_1\|_d < \infty$.

Similarly

$$\sup_{t\in\{(1,t_2),t_2\in[0,1]\}} \left\| n^{-1}\sum_{i=1}^{n} X_i(t) - EX(t) \right\| \to 0 \text{ a.e.} \qquad \text{(A.6)}$$

By (A.2), (A.5) and (A.6), the SLLN on $D_E(I_2)$ follows.■

$\rho=(0,0)$ $\qquad \tilde{Q}(\rho,t)=[0,t_1)\times[0,t_2)$

$\rho=(0,1)$ $\qquad \tilde{Q}(\rho,t)=[0,t_1)\times[t_2,1]$

$\rho=(1,0)$ $\qquad \tilde{Q}(\rho,t)=[t_1,1]\times[0,t_2)$

$\rho=(1,1)$ $\qquad \tilde{Q}(\rho,t)=[t_1,1]\times[t_2,1]$

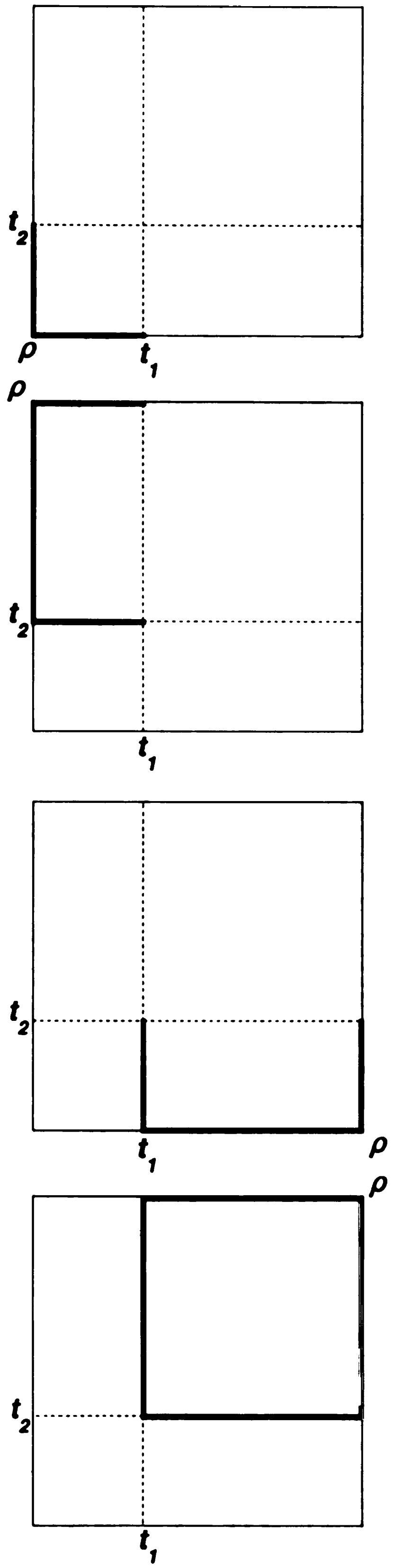
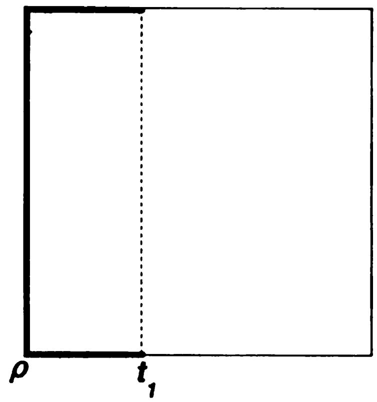


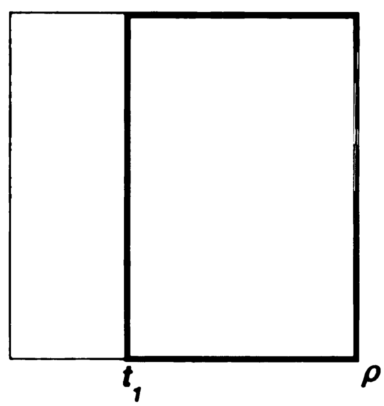
Figure A.1: Definition of the quadrants $\tilde{Q}(\rho,t)$ when $t\in[0,1)\times[0,1)$

$\rho = (0,0)$



$\tilde{Q}(\rho,t) = [0,t_1) \times [0,1]$

$\rho = (1,0)$



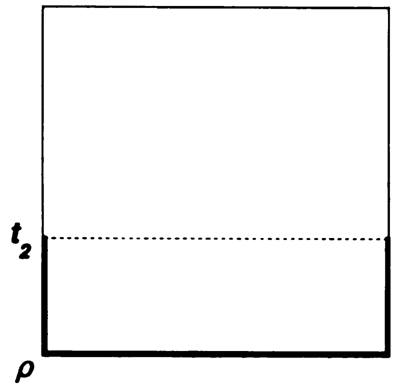
$\tilde{Q}(\rho,t) = [t_1,1] \times [0,1]$

$\rho \in \{(0,1),(1,1)\}$

$\tilde{Q}(\rho,t) = \Phi$

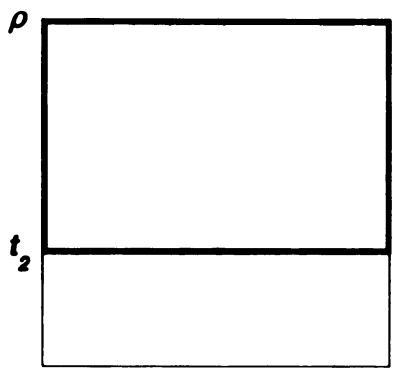
Figure A.2: Definition of the quadrants $\tilde{Q}(\rho,t)$ when $t \in \{(t_1,1),\ t_1 \in [0,1)\}$

$\rho = (0,0)$



$\tilde{Q}(\rho,t) = [0,1] \times [0,t_2)$

$\rho = (0,1)$



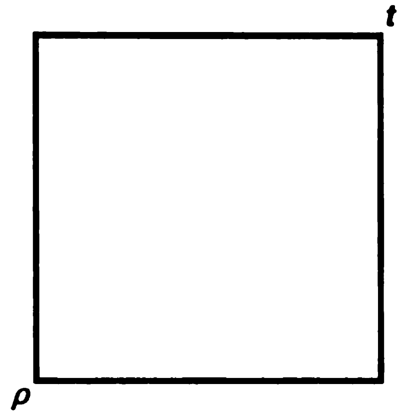
$\tilde{Q}(\rho,t) = [0,1] \times [t_2,1]$

$\rho \in \{(1,0),(1,1)\}$

$\tilde{Q}(\rho,t) = \Phi$

Figure A.3: Definition of the quadrants $\tilde{Q}(\rho,t)$ when $t \in \{(1,t_2),\ t_2 \in [0,1)\}$

$\rho = (0,0)$



$\tilde{Q}(\rho,t) = [0,1]\times[0,1]$

$\rho \in \{(0,1),(1,0),(1,1)\}$ $\qquad$ $\tilde{Q}(\rho,t) = \Phi$

Figure A.4: Definition of the quadrants $\tilde{Q}(\rho,t)$ when $t = (1,1)$

APPENDIX B

The Functional Delta Method

We will briefly review some results from Gill (1989).[72]

A concept of differentiability that allows a generalization of the usual Delta Method is the one of Hadamard or compact differentiability.

Let B_1, B_2 denote two normed vector spaces.

Definition

The functional $\varphi : B_1 \to B_2$ is **compactly or Hadamard differentiable** at a point $\theta \in B_1$ if and only if a continuous linear map $d\varphi : B_1 \to B_2$ exists, such that for all real sequences $a_n \to \infty$ and all convergent sequences $h_n \to h \in B_1$,

$$a_n\left(\varphi(\theta + a_n{}^{-1}h_n) - \varphi(\theta)\right) \to d\varphi(\theta).h \text{ as } n \to \infty.$$

Here $d\varphi(\theta)$ is called the derivative of φ at the point θ.

(See Definitions 1-3, p100 and the characterizations of differentiability, p102, Gill (1989)[72])□

An important property of Hadamard differentiation is that it satisfies the chain rule: if $\varphi : B_1 \to B_2$ and $\psi : B_2 \to B_3$ are Hadamard differentiable at $x \in B_1$ and

$\varphi(x) \in B_2$ respectively, then $\psi \circ \varphi : B_1 \to B_3$ is Hadamard differentiable at x, with derivative $d\psi(\varphi(x)).d\varphi(x)$.

Next we define the concept of weak convergence in normed vector spaces. Let $(B, \|\cdot\|)$ be a normed vector space endowed with a σ – algebra $\mathcal{B}$, such that $\mathcal{B}' \subseteq \mathcal{B} \subseteq \mathcal{B}''$, where $\mathcal{B}'$ and $\mathcal{B}''$ are the σ – algebras generated by the open balls and the open sets of B, respectively. Thus $\mathcal{B}''$ is the Borel σ – algebra ; when B is separable, $\mathcal{B}' = \mathcal{B}''$.

Definition (See Definition 4, Gill (1989)[72])

Let X_n be a sequence of random elements of $(B, \mathcal{B})$ and let X be another random element of that space. We say X_n converges weakly (or in distribution) to X and we write $X_n \xrightarrow{D} X$ if and only if $Ef(X_n) \xrightarrow{D} Ef(X)$ for all bounded, norm-continuous, $\mathcal{B}$-measurable $f : B \to \mathbb{R}$. □

The full functional version of Delta Method is given by Gill (1989)[72], Theorem 3:

Theorem (Functional Delta Method)

Suppose $\varphi : B_1 \to B_2$ is compactly differentiable at a point $\mu \in B_1$ and both it and its derivative are measurable with respect to the σ – algebras $\mathcal{B}_1$ and $\mathcal{B}_2$ (each nested between the open ball and Borel σ – algebras). Suppose X_n is a sequence of random elements of B_1 such that $Z_n = n^{1/2}(X_n - \mu) \xrightarrow{D} Z$ in B_1, where the distribution of Z is concentrated on a separable subset of B_1. Suppose addition: $B_2 \times B_2 \to B_2$ is measurable (see Remark 2 below). Then

(1) $\quad \left[n^{1/2}(X_n - \mu), n^{1/2}(\varphi(X_n) - \varphi(\mu)) - d\varphi(\mu).n^{1/2}(X_n - \mu)\right] \xrightarrow{D} (Z,0)$ in $B_1 \times B_2$

and consequently (in particular)

(2) $\quad n^{1/2}(\varphi(X_n) - \varphi(\mu)) - d\varphi(\mu).n^{1/2}(X_n - \mu) \xrightarrow{P} 0,$

(3) $\quad n^{1/2}(\varphi(X_n) - \varphi(\mu)) \xrightarrow{D} d\varphi(\mu).Z$. □

Remark 1:

Measurability of $d\varphi(\mu): B_1 \to B_2$ can often be shown to follow from measurability of φ (see Lemmas 4.4.3 and 4.4.4, van der Vaart (1988)[73]). □

Remark 2:

For $x = (x_1, x_2) \in B_1 \times B_2$ we define $\|x\| = \max(\|x_1\|, \|x_2\|)$ and we give product spaces $B_1 \times B_2$ and $B_2 \times B_2$ their product σ-algebras. If B_1 and B_2 are $D[0,\tau]^p \times \mathbb{R}^q$ for some finite p, q and $\mathcal{B}_1$, $\mathcal{B}_2$ are the open-ball σ-algebras then all product σ-algebras are also the open-ball σ-algebras with respect to the max norm. If one is only interested in getting (3), it suffices (qua measurability) to assume that left and right hand sides here are random elements of B_2. □

The following is a useful lemma. In many applications the mapping φ is only a priori defined on certain members of B_1 and one could set about choosing a particular extension to all of B_1 such that the hypotheses of the Functional Delta Method are satisfied in each particular application.

Lemma (see Lemma 1, Gill (1989)[72])

Consider $x \in E \subset B_1$ and $\varphi : E \to B_2$. Suppose there exists a continuous linear map $d\varphi(x) : B_1 \to B_2$ such that for all $t_n \to 0$ ($t_n \in \mathbb{R}$) and $h_n \to h \in B_1$ such that $x_n = x + t_n h_n \in E$ for all n, we have:

$$t_n^{-1}\left(\varphi(x + t_n h_n) - \varphi(x)\right) \to d\varphi(x).h \text{ as } n \to \infty.$$

Then φ can be extended to B_1 in such a way that it is differentiable at x, with derivative $d\varphi(x)$. The derivative is unique if the closed linear span of possible limit points h equals B_1. □

Comments:

Let $D[0,\tau]$ the space of real functions, right-continuous with left-hand limits, defined on $[0,\tau]$. We endow this space with $\|\,.\,\|_\infty$, the supremum norm. In Chapter 3 we consider spaces like $B = \left(D[0,\tau]\right)^p$. Under the Skorohod topology these spaces are Banach and separable. Under the max-supremum norm the separability is not valid any more. In these spaces, if the limiting process has continuous sample paths then weak convergence in the sense of the Skorohod metric and in the sense of the supremum norm are exactly equivalent. Otherwise, supremum norm convergence is stronger. □

The next proposition characterizes the compact differentiability of the product integral. As a reference see Theorem 8, Gill and Johansen (1990).[74] We state the result in the Andersen *et al.* (1993)[29] form (see Proposition II.8.7, p114):

Proposition

Let $E_M^{k^2} \subset (D[0,\tau])^{k^2}$ be the set of $k \times k$ matrix cadlag functions with components of total variation bounded by the constant M. Let $\varphi : E_M^{k^2} \to D[0,\tau]^{k^2}$ be defined by $\varphi(X) = \prod_{[0,.)}(I + dX)$.

Let X be a fixed point of $E_M^{k^2}$. Then φ can be extended to $D[0,\tau]^{k^2}$ so as to be compactly differentiable at X, with derivative

$$(d\varphi(X).H)(t) = \int_{s\in[0,t]} \prod_{[0,s)}(I + dX)H(ds)\prod_{(s,t]}(I + dX),$$

where, when H is not of bounded variation, the last integral is defined by the application (twice) of the integration by parts formula and the forward and backward integral equations (see Theorem 5, Gill and Johansen (1990)[74]).□

The following continuity result is proved in Theorem 7, Gill and Johansen (1990)[74]:

If X_n, X in $E_M^{k^2}$ are such that $X_n \to X$ in supremum norm then $\prod(I + dX_n) \to \prod(I + dX)$ in supremum norm.

APPENDIX C

Results on Ito Integration

Let $(\Omega, \mathcal{F}, P)$ a filtered, complete probability space, with filtration $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$, $\mathcal{F}_0$ containing all null sets of $\mathcal{F}$. Let M a continuous Gaussian martingale on this space, $M = \{M(t), t \in \mathcal{T}\}$, with $\mathcal{T} = [0, \tau], \tau < \infty$ (see definition of Gaussian martingales in Section 3.2.1). Let $\langle M \rangle = V$ a continuous, deterministic, positive, increasing function on $\mathcal{T}$, zero at time zero.

Note:

The results in stochastic calculus are usually for the standard Brownian motion (e.g. Harrison (1985)[75], Oksendal (1995)[76]). They can be easily translated for continuous Gaussian martingales which are called "time-transformed" Brownian motions. □

Let H^2 be the set of all adapted processes X on $((\Omega, \mathcal{F}, P), \mathbb{F})$ satisfying $E \int_0^t X^2(s) V(ds) < \infty$ for all $t \in \mathcal{T}$.

For any fixed t, one can defined the Ito integral $I_t(X) = \int_0^t X dM$ for $X \in H^2$.

This integral is unique up to a null set, with $E(I_t(X)) = 0$, $E(I_t(X)^2) = E \int_0^t X^2(s) V(ds)$.

As a process in t, $I_{\cdot}(X)$ is adapted, continuous, with $I_0(X)=0$. We will give a sketch of this definition, as presented in p56-61, Harrison (1985).[75] The stochastic integral $I_t(X)$ can be defined for a more general class of processes X: set of all adapted processes X such that

$P\left(\int_0^t X^2(s)V(ds) < \infty\right) = 1$ for all $t \in \mathcal{T}$. In our applications $X \in H^2$, so we will not explore the definition further than H^2.

Let S^2 be the set of all simple $X \in H^2$. A process X is called simple if there exist times $\{t_k\}$ such that $0 = t_0 < t_1 < ... < t_k \to \infty$ and $X(t,\omega) = X(t_k,\omega)$ for all $t \in [t_k, t_{k+1})$, $k \geq 0$. Note that the times $\{t_k\}$ do not depend on the argument ω.

Let L^2 denote as usual the set of all random variables ξ on $(\Omega, \mathcal{F}, P)$ such that $\|\xi\| = \left[E\xi^2\right]^{1/2} < \infty$. For $X \in H^2$ and fixed $t \in \mathcal{T}$, let $\|X\| = E\left[\int_0^t X^2(s)V(ds)\right]^{1/2}$. The same symbol $\|\,.\,\|$ will be used to denote both a norm on L^2 and a norm on H^2.

We fix $t \in \mathcal{T}$ and to simplify notation, set $I(X) = I_t(X)$ until t is freed.

If X is simple then one can define $I(X)$ in the Riemann-Stieltjes sense for almost every ω. Thus if $X = \sum_{i=1}^{n-1} X(t_k) I[t_k, t_{k+1})$, with $0 = t_0 < t_1 < ... < t_k = t$ then

$I(X) = \sum_{i=1}^{n-1} X(t_k)[M(t_{k+1}) - M(t_k)]$. By Proposition 10, p57, Harrison (1985)[75], if $X \in S^2$ then $E[I(X)] = 0$ and $\|I(X)\| = \|X\|$.

The space S^2 is dense in H^2. That is, for each $X \in H^2$, there exist simple processes $\{X_n\}$ such that $X_n \to X$ in H^2. As a reference see p92-95, Liptser-Shiryayev (1977).[77]

For $X \in H^2$ there exists a random variable $I(X) \in L^2$, unique up to a null set, such that $I(X_n) \to I(X)$ in L^2 for each simple sequence $\{X_n\}$ satisfying $X_n \to X$ in H^2. Furthermore, $E[I(X)] = 0$ and $\|I(X)\| = \|X\|$ (see Proposition 11, p58-59, Harrison (1985)[75]).

This concludes our short sketch of the definition of $I_t(X)$ for $X \in H^2$ and a fixed time $t \geq 0$.

Properties of the Ito integral:

Let $X, Y \in H^2$ and let $0 \leq s < u < t,\ s, u, t \in \mathcal{T}$. Then

i) $\int_s^t XdM = \int_s^u XdM + \int_u^t XdM$ for a.a. ω;

ii) $\int_s^t (cX + Y)dM = c\int_s^t XdM + \int_s^t YdM$ for a.a. ω;

iii) $\int_s^t XdM$ is $\mathcal{F}_t$ − measurable;

iv) If $X(t,\omega) = X(t)$ only depends on t (so *X(t)* is deterministic) then $I_t(X)$ is normally distributed, with mean 0 and variance $\int_0^t X^2(s)V(ds)$. □

Proof:

For i)-iii) see Theorem 3.9, p27, Oksendal (1995).[76]

For iv) consider first X simple, so $I_t(X) = \sum_{i=0}^{n-1} X(t_k)[M(t_{k+1}) - M(t_k)]$. Because $X(t)$ is deterministic, by the properties of the Gaussian martingales, $I_t(X)$ is normally distributed. We have seen that $EI_t(X) = 0$ and the standard deviation is $std(I_t(X)) = \|X\|$.

Let $X \in H^2$ and a simple sequence $\{X_n\}$ such that $X_n \to X$ in H^2 (i.e. $\|X_n\| \to \|X\|$). Then $I_t(X_n) \to I_t(X)$ in L^2, so $I_t(X_n)$ converges in distribution to $I_t(X)$. The distribution function of $I_t(X_n)$ is $\Phi\left(\frac{x}{\|X_n\|}\right)$ (where Φ is the standard normal distribution function) which converges to $\Phi\left(\frac{x}{\|X\|}\right)$, the distribution function of $I_t(X)$. Thus iv) follows. ■

The next theorem is a type of Fubini's Theorem for stochastic integration.

Fubini-type Theorem

Let $(s,u) \to H(s,u)$, $(s,u) \in \mathcal{T} \times \mathcal{T}$ be a bounded, $\mathcal{B} \times \mathcal{B}$ measurable function, where $\mathcal{B}$ is the set of Borelians on $\mathcal{T}$. Let μ a finite measure on the space $(\mathcal{T}, \mathcal{B})$.

Then, for every $t \in \mathcal{T}$, $A \in \mathcal{B}$:

$$\int_A \left[\int_0^t H(s,u)dM(u) \right] \mu(ds) = \int_0^t \left[\int_A H(s,u)\mu(ds) \right] dM(u) \text{ for a.a.} \omega. \ \square$$

For more general versions of this theorem and their proofs, see p159-161, Protter (1990).[78]

REFERENCES

1. Lipscomb J, Ancukiewicz M, Parmigiani G, Hasselblad V, Samsa G, Matchar DB. Predicting the Cost of Illness: A comparison of alternative models applied to stroke. *Medical Decision Making.* 1998;18 suppl:S39-S56.
2. Intrillagator MD, Bodkin RG, Hsaio C. *Econometric Models, Techniques, and Applications.* Second ed. Upper Saddle River: Prentice Hall; 1996.
3. Wooldridge JM. *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA: MIT Press; 1999.
4. Duan N. Smearing estimate: A nonparametric retransformation method. *Journal of the American Statistical Association.* 1983;78(383):605-610.
5. Zhou XH, Melfi CA, Hui SL. Methods for comparison of cost data. *Annals of Internal Medicine.* 1997;127(8):752-756.
6. Mullahy J. Much ado about two: Reconsidering retransformation and the two-part model in health econometrics. *Journal of Health Economics.* 1998;17:247-281.
7. Dudley RA, Frank E. Harrell J, Smith LR, et al. Comparison of analytic models for estimating the effect of clinical factors on the cost of coronary artery bypass graft surgery. *Journal of Clinical Epidemiology.* 1993;46(3):261-271.

8. Smith LR, Milano CA, Molter BS, Elbeery JR, Sabiston DC, Smith PK. Preoperative determinants of postoperative costs associated with coronary artery bypass graft surgery. *Circulation.* 1994;90(5, Part 2):124-128.

9. Etzioni RD, Feuer EJ, Sullivan SD, Lin D, Hu C, Ramsey SD. On the use of survival analysis techniques to estimate medical care costs. *Journal of Health Economics.* 1999;18:365-380.

10. Lin DY, Feuer EJ, Etzioni R, Wax Y. Estimating medical costs from incomplete follow-up data. *Biometrics.* 1997;53:419-434.

11. Lin DY. Proportional means regression for censored medical costs. *Biometrics.* 2000;56:775-778.

12. Hallstrom AP, Sullivan SD. On estimating costs for economic evaluation in failure time studies. *Medical Care.* 1998;36(3):433-436.

13. Bang H, Tsiatis AA. Estimating medical costs with censored data. *Biometrika.* 2000;87(2):329-343.

14. Gardiner J, Hogan A, Holmes-Rovner M, Rovner D, Griffith L, Kupersmith J. Confidence intervals for cost-effectiveness ratios. *Medical Decision Making.* 1995;15:254-263.

15. Gardiner J, Holmes-Rovner M, Goddeeris J, Rovner D, Kupersmith J. Covariate-adjusted cost-effectiveness ratios. *Journal of Statistical Planning and Inference.* 1999;75:291-304.

16. Lin DY. Linear regression of censored medical costs. *Biostatistics.* 2000;1:35-47.

17. Rapoport J, Teres D, Lemeshow S. Explaining variability of cost using a severity-of-illness measure for ICU patients. *Medical Care.* 1990;28:338-348.

18. Jones KR. Predicting hospital charge and stay variation: the role of patient teaching status, controlling for diagnosis related groups, demographic characteristics, and severity of illness. *Medical Care.* 1995;23:220-235.

19. Silberbach M, Shumaker D, Menashe V, Cobanoglu A, Morris C. Predicting hospital charge and length of stay for congenital heart disease surgery. *Am J Cardiol.* 1993;72:958-963.

20. Calvin JE, Klein LW, VandenBerg BJ, Meyer P, Ramirez-Morgen LM, Parrillo JE. Clinical predictors easily obtained at presentation predict resource utilization in unstable angina. *American Heart Journal.* 1998;136:373-381.

21. Benzaquen BS, Eisenberg MJ, Challapalli R, Nguyen T, Brown KJ, Topol EJ. Correlates of in-hospital cost among patients undergoing abdominal aortic aneurysm repair. *American Heart Journal.* 1998;136:696-702.

22. Krumholz HM, Chen J, Murillo JE, Cohen DJ, Radford MJ. Clinical correlates of in-hospital costs for acute myocardial infarction in patients 65 years of age and older. *American Heart Journal.* 1998;135:523-531.

23. Wei LJ, Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distritutions. *Journal of the American Statistical Association.* 1989;84(408):1065-1073.

24. Hoem JM, Aalen OO. Actuarial values and payment streams. *Scandinavian Actuarial Journal.* 1978:38-47.

25. Praestgaard J. Nonparametric estimation of actuarial values. *Scandinavian Actuarial Journal.* 1991;2:129-143.

26. Norberg R. Payment measures, interest, and discounting - an axiomatic approach with applications to insurance. *Scandinavian Actuarial Journal.* 1990:14-33.

27. Norberg R. Reserves in life and pension insurance. *Scandinavian Actuarial Journal.* 1991:3-24.

28. Norberg R. Hattendorff's Theorem and Thiele's Differential Equation Generalized. *Scandinavian Actuarial Journal.* 1992:2-14.

29. Andersen PK, Borgan O, Gill RD, Keiding N. *Statistical Models Based on Counting Processes*. New York: Springer-Verlag; 1993.

30. Lin DY, Wei LJ. The robust inference for the cox proportional hazards model. *Journal of the American Statistical Association.* 1989;84(408):1074-1078.

31. Andersen PK, Gill RD. Cox's regression model for counting processes: A large sample study. *Annals of Statistics.* 1982;10(4):1100-1120.

32. Wei LJ, Lachin JM. Two-sample Asymptotically Distribution-Free Tests for Incomplete Multivariate Observations. *Journal of the American Statistical Association.* 1984;79:653-661.

33. Thomas DR, Grunkemeier GL. Confidence interval estimation of survival probabilities for censored data. *Journal of American Statistical Association.* 1975;70:865-871.

34. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. New York: John Wiley & Sons; 1980.

35. Klein JP, Moeschberger ML. *Survival Analysis: Techniques for Censored and Truncated Data*. New York: Springer-Verlag; 1997.

36. Jacod J. Multivariate point processes: Predictable projection, Radon-Nikodym derivatives, representation of martingales. *Z. Wahrsch. verw. Geb.* 1975;31:235-253.

37. Jacod J. On the stochastic intensity of a random point process over the half-line. *Technical Report 15, Department of Statistics, Princeton University.* 1973.

38. Johansen S. An extension of Cox's regression model. *International Statistical Review.* 1983;51:258-262.

39. Arjas E, Haara P. A marked point process approach to censored failure time data with complicated covariates. *Scandinavian Journal of Statistics.* 1984;11:193-209.

40. Borgan O. Maximum likelihood estimation in parametric counting process models, with applications to censored failure time data. *Scandinavian Journal of Statistics.* 1984;11:1-16.

41. Fabian V, Hannan J. Introduction to probability and mathematical statistics. 1985.

42. Jennrich RI. Asymptotic properties of non-linear least squares estimators. *Annals of Mathematical Statistics.* 1969;40(2):633-643.

43. Rao RR. The law of large numbers for D[0,1]-valued random variables. *Theory of Probabability with Applications.* 1963;8:70-74.

44. Billingsley P. *Statistical Inferences for Markov Processes.* Chicago: University of Chicago Press; 1961.

45. Mourier E. Elements aleatoires dans un espace de Banach. *Ann. Inst. H. Poincare.* 1953;13:161-244.

46. Charlson ME, Pompei P, Ales KL, Mackenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of Chronic Diseases.* 1987;5:373-383.

47. Matsui K, Goldman L, Johnson PA, Kuntz KM, Cook F, Lee TH. Comorbidity as a correlate of length of stay for hospitalized patients with acute chest pain. *Journal of General Internal Medicine.* 1996;11:262-268.

48. Fenn P, McGuire A, Backhouse M, Jones D. Modelling programme costs in economic evaluation. *Journal of Health Economics.* 1996;15:115-125.

49. Polverejan E, Gardiner JC, Bradley CJ, Holmes-Rovner M, Rovner D. Estimating mean hospital cost as a function of length of stay and patient characteristics. *In review.* 2001.

50. Longini IR, Byers RH, Hessol NA, Tan WY. Estimating the stage-specific numbers of HIV infection using a Markov model and backcalculation. *Statistics in Medicine.* 1992;11:831-843.

51. Satten GA, Longini IR. Markov chains with measurement error: estimating the 'true' course of a marker of the progression of HIV disease. *Applied Statistics.* 1996;45:275-309.

52. Aalen OO, Farewell VT, de-Angelis D, Day NE, Gill ON. 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.* 1997;16:2191-2210.

53. Longini IMJ, Clark WS, Byers RH, et al. Statistical analysis of the stages of HIV infection using a markov model. *Statistics in Medicine.* 1989;8:831-843.

54. Gentleman RC, Lawless JF, Lindsey JC, Yan P. Multi-state markov models for analysing incomplete disease history data with illustrations for HIV disease. *Statistics in Medicine.* 1994;13:805-821.

55. Hansen BE, Thorogood J, Hermans J, Ploeg RJ, Bockel JHV, Houwelingen JCV. Multistate modelling of liver transplantation data. *Statistics in Medicine.* 1994;13:2517-2529.

56. Dabrowska DM, Guo-wen S, Horowitz MM. Cox regression in a Markov renewal model: an application to the analysis of bone marrow transplant data. *Journal of the American Statistical Association.* 1994;89:876-877.

57. Wanek LA, Elashoff RM, Goradia TM, Morton DL, Cochran AJ. Application of multistage markov modeling to malignant melanoma progression. *Cancer.* 1994;73:336-343.

58. Perez-Ocon R, Ruiz-Castro JE, Gamiz-Rerez ML. A multivariate model to measure the effect of treatments in survival to breast cancer. *Biometrical Journal.* 1998;40:703-715.

59. Gold MR, Siegel JE, Russell LB, Weinstein MC, eds. *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press; 1996.

60. Jacobsen M. *Statistical Analysis of Counting Processes*. New York: Springer-Verlag; 1982.

61. Verbeke G, Molenberghs G, eds. *Linear Mixed Models in Practice: A SAS-Oriented Approach*. New York: Springer-Verlag; 1997.

62. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data*. New York: Springer-Verlag; 2000.

63. Diggle PJ, Liang KY, Zeger SL. *The Analysis of Longitudinal Data*. Oxford, UK: Oxford University Press; 1994.

64. Cnaan A, Laird NM, Slasor P. Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Statistics in Medicine.* 1997;16:2349-2380.

65. Laird NM, Ware JH. Random-Effects Models for Longitudinal Data. *Biometrics.* December 1982 1982;38:963-974.

66. Cox DR, Hinkley DV. *Theoretical Statistics*. London: Chapman & Hall; 1990.

67. Amemiya T. *Advanced Econometrics*. Cambridge, MA: Harvard University Press; 1985.

68. Manning WG, Duan N, Rogers WH. Monte Carlo evidence on the choice between sample selection and two-part models. *Journal of Econometrics.* 1987;35:59-82.

69. Duan N, Willard G. Manning J, Morris CN, Newhouse JP. Choosing between the sample-selection model and the multi-part model. *Journal of Business & Economic Statistics.* 1984;2(3):283-289.

70. Neuhaus G. On weak convergence of stochastic processes with multidimensional time parameter. *Annals of Mathematical Statistics.* 1971;42(4):1285-1295.

71. Billingsley P. *Convergence of probability measures*. New York: Wiley; 1968.

72. Gill RD. Non- and Semi-parametric Maximum Likelihood Estimators and the von Mises Method (Part 1). *Scandinavian Journal of Statistics.* 1989;16:97-128.

73. van der Vaart AW. *Statistical estimation in large parameter spaces.* Vol 44. Amsterdam: Centrum voor Wiskunde en Informatica; 1988.

74. Gill RD, Johansen S. A survey of product-integration with a view towards application in survival analysis. *Annals of Statistics.* 1990;18:1501-1555.

75. Harrison JM. *Brownian Motion and Stochastic Flow Systems*. New York: John Wiley; 1985.

76. Oksendal B. *Stochastic Differential Equations*. Fourth ed. New York: Springer-Verlag; 1995.

77. Liptser RS, Shiryayev AN. *Statistics of Random Processes I.* New york: Springer-Verlag; 1977.

78. Protter P. *Stochastic Integration and Differential Equations*. New York: Springer-Verlag; 1990.