This is to certify that the
dissertation entitled

ASSESSING MEDICAL COSTS FROM A LONGITUDINAL MODEL

presented by

CORINA MIHAELA SIRBU

has been accepted towards fulfillment
of the requirements for the

Doctoral degree in Statistics and Probability

Joseph C Gardiner
Major Professor's Signature

July 2, 2004
Date

ASSESSING MEDICAL COSTS FROM A LONGITUDINAL MODEL

By

Corina Mihaela Sirbu

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

2004

ABSTRACT

ASSESSING MEDICAL COSTS FROM A LONGITUDINAL MODEL

By

Corina Mihaela Sirbu

The United States spends a larger share of its gross domestic product (GDP) on health care than any other major industrialized country. Expenditures for health care represent nearly one-seventh of the Nation's GDP, and they continue to be one of the fastest growing components of the Federal budget. In 1960, for example, health care expenditures accounted for about 5 percent of the GDP; by 2000, that figure had grown to more than 13 percent. Although the rate of growth in health care costs slowed somewhat in the mid-1990s, it has once again started to rise at a rate that exceeds other sectors of the economy. Thus, identifying methods to accurate estimate health care costs continues to be a priority for policymakers and public and private payers.

In medical follow up studies incomplete observation due to censoring would preclude ascertainment of outcomes in some subjects. Standard assumptions used in survival analysis do not apply to medical costs because the cumulative cost at the endpoint of interest will generally be correlated with the cumulative cost at the time of censoring.

We use a dynamic regression model in which costs are incurred in random amounts at transition times between and during sojourn in health states. A Markov model describes the unfolding over time of individual patient event histories, with transition

intensities depending on patient specific demographic and clinical characteristics through a multiplicative intensity model. A random effects model is used for transition and sojourn costs. We then estimate the net present value of expenditures incurred over a finite time horizon. While incorporating explanatory variables, the joint model can accommodate heteroscedasticity, skewness and censoring in cost and health outcome data and provides a flexible approach to analyses of health care costs and outcomes.

Our transition model can be viewed as an extension of the simpler two state model, case in which we obtain and revise already developed techniques for regression analysis of medical costs with the focus being on estimation in the presence of time censoring that might result in incomplete costs data on some patients. Using the 2000 Nationwide Inpatient Sample data set of Health Care Utilization Project we focus on estimating costs for patients admitted in the hospital with acute myocardial infarction (AMI), a common high-mortality condition whose outcomes are affected by the process of care.

Our methods provide flexible approaches to estimating medical costs. Estimates from cost studies are not only needed to determine the economic burden of disease, to predict the economic consequences of new medical interventions, but also for comparative purposes such as cost-effectiveness analysis. Other possible extensions of our methods are in this area.

To my husband, parents and brother for all his love and support!

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Joseph C. Gardiner, for having the patience to allow me to pursue my own interests, with the occasional push when required to keep me on track. I felt very privileged to have his guidance and support during my graduate studies. I also want to express my thanks to the other committee members for their friendly support. My research was supported by the Agency for healthcare Research & Quality, under Grant 1R01HS09514.

I thank my family for always believing in me. My parents, Rodica and Ion Radut, my brother, Radu have continually offered me support and encouragement. I also want to thank my friends at Michigan State University for always helping me to turn worries into laughs.

Last and most of all, I thank my dear husband, George. This last year was trying for both of us, but his unwavering support and encouragement certainly made my struggles more bearable and I am forever grateful to him.

TABLE OF CONTENTS

LIST OF TABLES ix

LIST OF FIGURES x

ABBREVIATIONS xi

INTRODUCTION 1

CHAPTER 1
ESTIMATING MEDICAL COSTS FROM A TRANSITION MODEL 13
1.1 A Markov model for describing patient health histories 14
1.1.1 Marked point processes 18
1.1.2 Incorporating censoring 20
1.2 Incorporating costs in the Markov model 22
1.3 Estimation of transition probabilities 31
1.4 Estimation of average transition costs 51
1.4.1 Modification for censoring 53
1.4.2 Consistency and asymptotic normality of $\hat{\beta}_w$ 56
1.4.3 Computation of A_w and B_w 57
1.4.4 Estimation of $G(\cdot \mid \mathbf{z})$ 58
1.5 Asymptotic distribution of the mean transition cost 61

CHAPTER 2
ESTIMATING MEAN COST FOR THE TWO-STATE CASE 78
2.1 Introduction and Background 79
2.2 Applying general transition model methods to the 2-state model 86
2.2.1 Mean cost in the minimal cost data case 87
2.2.2 Mean cost in the interval cost data case 97
2.3 Asymptotic normality of the mean cost 101
2.3.1 Regression model for log-cost 110
2.4 Parametric estimation of the survival distribution for censoring time 114

2.4.1 Minimal cost data 114
2.4.2 Interval cost data 121

CHAPTER 3
ESTIMATING HOSPITAL COST FOR AMI PATIENTS IN THE NIS 2000 130
3.1 Description of the data set and background 130
3.2 Creating a working subset data set 144
3.3 Estimation methods 148
3.3.1 Unconditional means model 149
3.3.2 Including effects of hospital level (level 2) correlates 150
3.3.3 Including effects of discharge level (level 1) correlates 151
3.3.4 Including both effects of discharge level (level 1) and hospital level (level 2) correlates 153
3.3.5 Inverse probability weighting 155
3.4 Results of estimation 159

CHAPTER 4
CONCLUSIONS AND FUTURE WORK 176

APPENDIX
Mathematical background 184

BIBLIOGRAPHY 192

LIST OF TABLES

Table 2.1 Estimating the survival distribution for censoring time 92

Table 3.1 Primary procedures 164

Table 3.2 Descriptive statistics (N=58469) 166

Table 3.3 Parameter estimates in multilevel regression 170

Table 3.4 Estimates of total charges at median LOS by procedure types, comorbidity levels (0, 1+) and regions. 173

LIST OF FIGURES

Figure 1.1 Transition diagram for a multi-state Markov model 16

Figure 3.1 Creating a working subset data set 165

Figure 3.2 Histogram of Log (TOTCHG).....167

Figure 3.3 Graph of Log (TOTCHG) versus LOS.....168

Figure 3.4 Estimation of G: Graphs of the Kaplan-Meier estimates of the cumulative hazard versus the regression model estimates of the cumulative hazard from both log-normal and gamma models.....169

Figure 3.5 Estimation of S: Graphs of the Kaplan-Meier estimates of the cumulative hazard versus the regression model estimates of the cumulative hazard from both log-logistic and gamma models 172

Figure 3.6 Estimates of total charges at median LOS – no comorbidities.....174

Figure 3.7 Estimates of total charges at median LOS – at least one comorbidity 175

ABBREVIATIONS

CEA	-Cost-Effectiveness Analysis
CER	-Cost-Effectiveness Ratio
NHC	-Net Health Cost
NHB	-Net Health Benefit
NPV	-Net Present Value
QALY	-Quality Adjusted Life Years
LE	-Life Expectancy
CABG	-Coronary Artery Bypass Grafting
PTCA	-Percutanerous Transluminal Coronary Angioplasty
CATH	-Cardiac Catheterization
LOS	-Length of Stay
TOTCHG	-Total Charges
AMI	-Acute Myocardial Infarction
DRG	-Diagnosis Related Group
HMO	-Health Maintenance Organization
NIS	-Nationwide Inpatient Sample data
HCUP	-Health Care Utilization Project
AHRQ	-Agency for Healthcare, Research and Quality
CCS	-Clinical Classifications Software
CCI	-Charlson Comorbidity Index.
IDC-9-CM	-International Classification of Diseases 9th Revision, Clinical Modification
OECD	-Organization for Economic Co-operation and Development
WHO	-World Health Organization
iid	-Independent Identically Distributed
CLT	-Central Limit Theorem
□	-End of Statement

INTRODUCTION

Economic evaluations of health care interventions are increasingly important in an era of constrained health care budgets. As policymakers seek to prioritize health care expenditures, an accurate assessment of costs and health benefits of competing interventions and treatments is critical in informing resource allocation decisions in health care. A recent report[1] from the Office of the Actuary at the Center for Medicare and Medicaid Services projects that the national health expenditures would reach $3.4 trillion in 2013, growing at an average annual rate of 7.3 percent during the forecast period 2002-2013. As a share of gross domestic product (GDP), health care spending is projected to reach 18.4 percent by 2013, up from its 2002 level of 14.9 percent. This demands extraordinary restructuring of the organization and financing of U.S health services.

Over the past decade there has been an explosion of research on methodology for economic evaluations in health care. With increasing availability of large databases on patient outcomes and costs, statistical methods for comparing outcomes with costs need to be developed. The field is young, however, and there are many important and challenging problems that remain unresolved.

In this dissertation we will address several statistical issues with analysis of medical cost data. We adopt a longitudinal framework in which costs incurred as the individual level are random quantities associated with events that occur as an individual's

health history unfolds over time. A Markov process is used to describe the dynamics of movement of an individual through different health states. Costs are incurred at transition times, and in sojourn in health states. Total expenditures over a finite period of time is then defined as an 'expected value' called a net present value (NPV). Because individual characteristics such as demographics (age, gender, race) and clinical factors (treatments, comorbidities) can influence NPV, we incorporate covariate effects into our model for estimation of NPV.

Our longitudinal framework provides a natural setting for estimating medical costs. We will demonstrate how some recent approaches to analysis of costs[2-11] can be subsumed into this framework.

Importance of Cost-Effectiveness Analysis

In addition to evidence of clinical effectiveness of treatments, evidence of their cost-effectiveness has become an important consideration as policy-makers world-wide face decisions in allocating resources for health care services. In Australia, the Pharmaceutical Benefits Advisory Committee makes recommendations, based on effectiveness and cost-effectiveness evidence, on drug products that should be subsidized and placed in the Pharmaceutical Benefits Scheme[12]. The National Institute of Clinical Excellence[13] in the UK makes similar requirements for use of new healthcare technologies in the National Health Service, and in Ontario, Canada, the Drug Benefits Plan uses economic data when supporting new additions to its formulary[14]. The phenomenal penetration of HMOs into the US health care market has heightened awareness of cost-effectiveness among providers and consumers of healthcare services.

The US Preventive Services Task Force and the Panel of Cost-Effectiveness in Health and Medicine have urged consideration of cost-effectiveness in addition to clinical effectiveness to help inform investment of health care dollars[15].

Measures used in cost-effectiveness analysis

Cost-effectiveness analysis (CEA) has been promoted as a useful tool in the effort to prioritize expenditure on health care programs[15]. By quantifying the trade-offs between resources that need to be deployed and health benefits that accrue from use of alternative interventions, CEA offers guidance in decision-making by structuring comparisons between these interventions. A *cost-identification analysis* is often conducted for treatments and procedures that are believed to be equivalent in their clinical efficacy. For example, if two competing programs do not differ on average in their health benefits, then the one with the lower average cost will be preferred. On the other hand, if the costs of two programs are judged equivalent, the intervention with the greater health benefit will be preferred. An intervention that delivers higher benefit at lower cost than its competitor is said to be *dominant.* A decision has to be made when one program has both higher cost and greater benefit than does its competitor. Is there a critical value below which society would consider the more costly intervention still "cost-effective"? In this situation, the *cost-effectiveness ratio* (CER) becomes a useful summary statistic for ranking competing interventions. It is the ratio of the incremental cost relative to the incremental benefit. With costs measured in dollars and health benefits measured in their natural units such as life expectancy, number of lives saved, or preferably *quality-adjusted life years* (QALYs), the CER is stated in dollars per unit of effectiveness. In

CEAs conducted with a societal perspective that accounts for all costs of the interventions, whether borne by the recipient of care, the provider or the insurer, the critical value of a CER is the upper limit of what society is willing to pay for an additional unit of health benefit.

Other summary statistics used in CEA are the *net health benefit* (NHB) or *net health cost* (NHC)[16]. Suppose the incremental health benefit is monetized using a value for each additional unit of health benefit. This could be the upper limit of the CER as judged by what society is willing to pay for adopting the competing intervention. The NHC, expressed in dollars, is the difference between the incremental cost and the monetized incremental benefit. The net health benefit can be defined in an entirely analogous manner, and would be expressed in units of effectiveness. Many researchers have pointed out that the CER has undesirable properties that make its use in decision making problematic.

Analysis of medical costs

Incomplete data are likely to arise in longitudinal studies, because patient follow up will not be complete in all subjects. In survival analysis, censoring occurs when the time to event variable T is not observed in some individuals because a censoring event occurs first at time U, that is $U < T$. Most survival analysis models assume T, U independent, or, when covariates $\mathbf{z}$ are present, that conditionally on $\mathbf{z}$, T and U are independent. This is the usual random censorship model. With accumulating costs this assumption is untenable. If $y(t)$ is the accumulated cost up to time t, then $y(T)$ and

$y(U)$ are generally correlated. Therefore analyzing costs by traditional survival analysis techniques is not possible.

In this dissertation, we first consider the situation in which a single cost variable y is observed together with covariate information $\mathbf{z}$ in a sample of subjects. In our general framework costs can potentially be accrued over a fixed time period $[0,\tau]$ with expenditure terminating at some event time T so that complete cost observation occurs if a patient is followed through time $T^* = \min(T,\tau)$. Suppose $y(t)$ is a right-continuous process that represents the cumulative cost up to time t (including time t) for a typical patient in the population under study. If lifetime cost is of interest then T denotes survival time. Since costs do not accumulate after T, $y(t) = y(T)$ for all $t \geq T$. The cumulative cost $y(\tau)$ at time τ is the principal random variable of interest, so inference focuses on the mean cost, $\mu = E(y(\tau)) = E(y(T^*))$. With lifetime medical cost, the cumulative cost is $y(T)$ and estimating the average $E(y(T))$ is of interest.

Because of possible censoring at time U, $y(T)$ is not observed if $T > U$. If this is the case we observe $y(U)$. A simple sample average of the observed costs in the patient sample would underestimate the true expected medical cost for the treatment under study. Also using the average in the sub-sample of patients with complete costs would be inefficient.

Even if complete costs were available, standard regression techniques for assessing the influence of covariates on costs can not be directly applied. Cost data are often very skewed, usually to the right. They also exhibit considerable heterogeneity across patients. Standard assumptions used in ordinary least squares (OLS) for example can not be applied. To mitigate the effects of skewness, the log-transformation of costs

might be considered. However, this too has adherents and non-adherents as explained in Manning (2001)[17]. Even if a transformation were feasible, a retransformation would be needed to obtain estimates of mean costs (and other statistics) across specified covariate profiles: retransformation itself presents some methodological challenges[18-21].

Extreme form of skewness occurs in cost data when proportion of subjects in the sample have zero costs. For example if we examine costs of office visits to a doctor or other health professional in the year 2000, 20% of adults 18 years of age and over did not make any office visit[22] and therefore incurred zero expense. This creates a 2-part distribution for costs, one part for the sub-population with an expense, and the second part for those without. These groups differ considerably in their demographic characteristics and medical history. In a two-part model, one has a model for the likelihood of expense, for example, a probit or logit model for $P(y>0|\mathbf{z})$, where $\mathbf{z}$ is a vector of covariates, and then a second model for $E(g(y)|\, y>0,\mathbf{z})$, where g is a transformation, such as the logarithmic. Debate continues on the proper analysis of the two-part models and comparisons to other models such as the Heckman model and sample selection models[23, 24].

Cox regression has been the mainstay for analysis of censored time to event data. However using this method directly with costs is not possible. As noted earlier, cost at censoring time and cost at event time are correlated. In this dissertation we maintain the traditional use of Cox regression for time to event analysis. It is used to model covariate effects on the transition intensities as patients move from one health state to another. We then combine this with a linear mixed effects model for costs (incurred at transition times

or sojourn in states) conditional on event times. Finally we derive estimators of NPV given a covariate profile and develop the asymptotic theory of these estimators.

A transition model for analysis of medical costs: Outline of Dissertation

When an intervention is deployed costs are incurred in random amounts at random points in time. Typically these costs are associated with health states that a patient might visit in the course of the intervention, and the different lengths of time spent in each state. The probabilistic mechanism that governs transition between these states and the distribution sojourn times in health states vary at the individual level depending on patient specific demographic and clinical characteristics. This thesis concerns the development of new statistical methodologies for estimating medical costs with censored data in both this multiple states setting and the two-state case.

In **Chapter 1** we describe the evolution of a patient's health as the unfolding in time of a finite state stochastic process. A non-homogeneous Markov process $X = \{X(t) : t \in \mathcal{T}\}$ with finite state space $E = \{1, 2, ..., k\}$, provides a natural setting to describe the probabilistic mechanisms that govern transitions between states, where $X(t)$ is the patient status or health state occupied at time $t \in \mathcal{T} = [0, \tau]$, and $\tau \leq \infty$. Transition probabilities are denoted by $P_{hj}(s,t)$ and transition intensities by $\alpha_{hj}(t)$. The state space of X typically consists of several transient states, such as "well", "recovery", "relapse" and one or more absorbing states such as "dead" or "disabled". Over the follow up period the typical patient would transit to other health states, $X_1 = X(T_1)$, $X_2 = X(T_2), ...$ at random times $T_1, T_2, ...$ and these transition times and health states

describe the event history of each patient. If observation of X is ceased after some random time U, independent of X, then we will need to account for censoring accordingly.

The survival model is an example of a two-state process with a single transient state "alive" and a single absorbing state "dead" with survival time $T = T_1$ and $T_n = \infty$ for $n \geq 2$. The multi-state analog of survival time is the time to absorption in state k given by $\tau_k = \inf\{t > 0 : X(t) = k\}$.

Having described the evolution of a patient history by the finite state space non-homogeneous Markov process X, we now 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 health state. Incorporating costs in the model enlarges the usual σ-field used in the multiple states survival theory, namely $\mathcal{F}_t^c = \sigma\{(T_n, X_n) : 0 \leq T_n \leq t \wedge U\}$, by adding cost information. Under specified assumptions the martingale theory still obtains and the compensators remain unchanged.

The estimation of $P_{hj}(s,t \mid Z_0)$ from a Cox regression model (multiplicative intensity model) for the $\alpha_{hj}(t \mid Z_0)$, has been very well developed by Andersen *et al* (1993)[25]. Numerous applications of this method are published regularly in the medical and epidemiologic literature.

To analyze costs incurred at transition times, we adopt a mixed model approach. If $T_{i1}, T_{i2}, ..., T_{in_i}$ denotes the observed sequence of n_i transition times in the ith individual and $\mathbf{Y}_i = (y_{i1}, y_{i2}, ... y_{in_i})'$ the associated vector of costs (or transformed costs), then the random-effects model $\mathbf{Y}_i = \mathbf{X}_i\beta + \mathbf{Z}_i v_i + \mathbf{u}_i$ is the basis for estimation of β. Here the covariate matrix $\mathbf{X}_i$ will include terms for the times $T_{i1}, T_{i2}, ..., T_{in_i}$, individual patient

characteristics, and the matrix $\mathbf{Z}_i$ will include a subset of these factors, most likely variables for modeling the effect of transition times such as T_{ij} and T_{ij}^2. The unobserved heterogeneity is the vector v_i, inducing dependence among the y_{ij}'s and $\mathbf{u}_i$ is the residual error. We will derive the NPV for all transition costs in the interval [0, τ], in the form $\text{NPV}(Z_0) = \sum_{h \neq j} \int_0^{\tau} e^{-rt} c_{hj}(t \mid Z_0) \alpha_{hj}(t \mid Z_0) P_{ih}(0,t \mid Z_0) dt$ conditional on the initial state $X_0 = i$ and a specified covariate profile Z_0, where r denotes the discount rate. Here $c_{hj}(t \mid Z_0)$ is the expected cost incurred at time t if the transition $h \rightarrow j$ occurred and it can be obtained from the components of $E(\mathbf{Y}_i \mid \mathbf{X}_i)$. Similarly we can define NPV for all sojourn costs. The mathematical form for NPV depends on the underlying transition probabilities and intensities.

Following Andersen *et al* (1993)[25] we account for heterogeneity across patients by semiparametric modeling of the transition intensities of the process through patient-specific covariates. We estimate transition probabilities using a Cox regression model. We combine the two parts to form an estimate of the mean present value of all expenditures and use the inverse-probability of censoring-weighted (IPCW) technique to account for censored observations. The estimators are obtained conditional on an initial state and given a covariate profile. By applying the delta-method to functionals that arise in the estimation of the NPV, we obtain large sample properties of these estimators.

This approach is new and builds upon a similar idea used by Praestgaard (1991)[26] to estimate actuarial values in life insurance. In that context the benefit (cost) is fixed and the stochastic elements are the sojourns in policy states or transitions between policy states. The 'cost' at transitions (called assurances in the life-insurance literature) are fixed

by the terms of the policy. Also the 'unit cost of sojourn' (called annuity payments) are also fixed[27-29]. In our context these quantities are no longer fixed. With longitudinal data these costs are observed and vary across patients.

In **Chapter 2,** we show that our transition model described in Chapter 1 also captures costs under the simpler two state survival model with a single transition time and sojourn. In this case several investigators have developed techniques for regression analysis of medical costs with the focus being on estimation in the presence of time censoring that might result in incomplete cost data on some patients. Our transition model can be therefore viewed as an extension of this methodology to multiple transition times and sojourns. If we specialize our multiple transition model methods from Chapter 1 to a two state model with patients starting in state '0' (alive) and followed until they reach a terminal state '1' (death) at time T, the total cost for a patient can be interpreted as a sojourn cost that ends at time T or τ whichever occurs first or as a transition cost at time T if the patient dies in the interval of time $[0, \tau]$.

Current methods for estimation of the population mean cost are both nonparametric and semi-parametric. The key references are Lin *et al* (1997, 2000, 2003)[4,5,30], Bang and Tsiatis (2000)[31], Strawderman (2000)[32], Willan *et al* (2002, 2003)[2,3], Wooldridge (2002, 2003) [33,34]. Semi-parametric models would assume a special parametric form for the distribution of cost. For a single cost, Zhou *el al* (2000)[35] primarily uses a log-normal regression model to assess covariate effects on mean costs. He also discusses approaches to deal with heteroscedasticity, skewness, censoring and zero costs, all in the context of a parametric model[35-41].

Our approach uses the same inverse-probability of censoring-weighted (IPCW) technique to derive consistent and asymptotically normal estimators of regression parameters and for the net present values. We discuss parametric methods for estimating the survival distribution for censoring time, and therefore the weights in the IPCW technique, as well as methods of estimation for the survival distribution for event time.

Chapter 3 focuses on applications to real data: We use the inpatient utilization data from the Nationwide Inpatient Sample (NIS) of Health Care and Utilization Project (HCUP), a database of all hospital inpatient stays drawn from a stratified sample of approximately 1,000 community hospitals in the US. For 2000, the NIS contains over 7.4 million discharges from 28 states. Total charge and length of stay (LOS) are the main healthcare utilization variables for each hospital stay. There is a growing literature on use of the NIS in health services research that we use for guidance [42-47]. Following our experience with analyzing charges and LOS of acute myocardial infarction (AMI) patients in the MICH study [48-50], we will focus on patients admitted in the hospital with AMI that have undergone either no procedures or Coronary Artery Bypass Grafting (CABG), Cardiac Catheterization (CATH) or Percutaneous Transluminal Coronary Angioplasty (PTCA) as a primary procedure.

Hospital characteristics, such as quality of service or managerial performance may impose distinct effects on the costs of treating patients. Rice *et al* (1997)[51], Carey (2000, 2002) [52, 53] and Goldstein (2002)[54] insist on the usefulness of multilevel methods in studies where data on cost are collected over multiple sites (hospitals in our data). In such circumstances it can be expected that hospitals may have an impact on the cost regardless of treatment the patient receives. The inclusion of hospital as a level in a

multilevel analysis will ensure that the clustering effects within hospitals will be adequately controlled for. Traditional estimation procedures such as OLS, which is used for example in multiple regression, are inapplicable because of the existence of a non-zero intra-hospital correlation, resulting from the presence of more than one residual term in the model.

We use a multilevel modeling technology to estimate costs for patients diagnosed with acute myocardial infarction as they relate to both patient and hospital level characteristics. Patients transferred to a short-term hospital, as well as other transfers, including skilled nursing facilities (SNF), intermediate care, home health care have incomplete total charges and length of stay so they will be treated as censored. In our working data set 32% of the discharges are censored. Selection probabilities from the censored sample of event times are estimated using parametric estimators of the censoring distribution. We then estimate total charges at the median LOS for specific covariate profiles. Our method accounts for the existence of a non-zero intra-hospital correlation, censoring and skewness of total charges.

In the concluding **Chapter 4** we outline some extensions of our work particularly to estimation of summary measures in CEA such as the CER, net health benefit (NHB) or cost (NHC), net present value (NPV), life-expectancy (LE) and quality-adjusted life years (QALY).

CHAPTER 1

ESTIMATING MEDICAL COSTS FROM A TRANSITION MODEL

Economic evaluations of health care interventions are increasingly important in an era of constrained health care budgets. As policymakers seek to prioritize health care expenditures, an accurate assessment of costs and health benefits of competing interventions and treatments is critical in informing resource allocation decisions in health care.

A multi-state model is defined as a model for a stochastic-process, which at any time occupies one of a set of discrete states. The states can describe conditions like healthy, diseased, diseased with complications and dead. When an intervention is deployed costs are incurred in random amounts at random points in time. Typically these costs are associated with health states that a patient might visit in the course of the intervention, and the different lengths of time spent in each state. The probabilistic mechanism that governs transition between these states and the distribution sojourn times in health states vary at the individual level depending on patient specific demographic and clinical characteristics.

The main themes of this chapter are arranged as follows. In section 1.1 we provide a description of patient history as the unfolding in time of a finite state stochastic process. A non-homogeneous Markov process describes the probabilistic mechanisms that govern transitions between states. Following Andersen *et al* (1993)[25] we account for heterogeneity across patients by semiparametric modeling of the transition intensities of the process through patient-specific covariates. We then describe how costs are incorporated into this framework. We consider two types of costs, costs while sojourning in a health state and costs incurred at transitions between health states. Net present values of all expenditures incurred over a finite time horizon are then defined and have mathematical forms that depend on the underlying transition probabilities and intensities. We will write each term of these mathematical forms as sum of independent and identically distributed variables. Conditional on the initial state i, given the vector Z_0 of basic covariates, we will assess the asymptotic normality of the net present value of all expenditures associated with the h to j transitions in $(0,t]$, i.e. the asymptotic distribution of

$$n^{1/2}\left(N\hat{P}V_{hj}^{(1)}(t\,|\,i,Z_0)-NPV_{hj}^{(1)}(t\,|\,i,Z_0)\right)$$

using the Functional Delta Method.

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 $\mathcal{T}=[0,\tau],$, a non-homogeneous continuous time Markov process with finite state space $E=\{1,2,...,k\}$,

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 $\alpha = (\alpha_{hj}), h, j \in \{1,2,...,k\}$ be the matrix of these transition intensities,

$$\alpha_{hj}(t) = \lim_{\Delta t \downarrow 0} P[X(t+\Delta t) = j \mid X(t) = h]/\Delta t, \ j \neq h$$

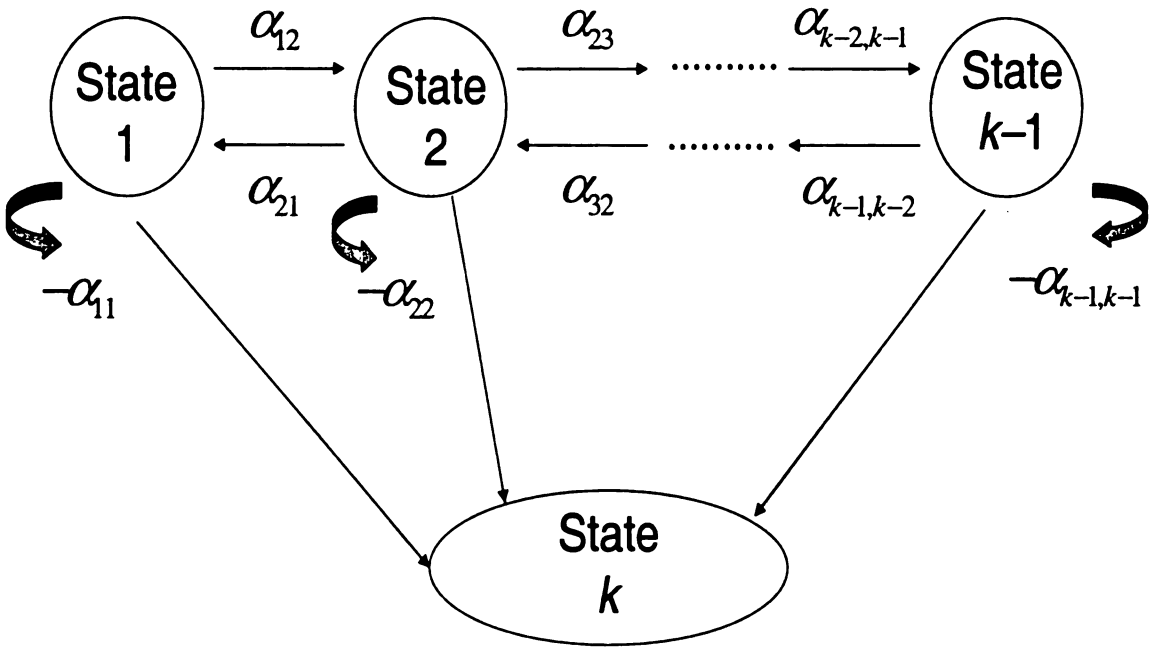
and $\alpha_{hh} = -\sum_{j \neq h} \alpha_{hj}$. Thus, starting from the time of entry into state h, the sojourn times in the given state h are continuously distributed, with hazard rate function $-\alpha_{hh}$. Given that the process jumps out of state h at time t, it jumps into state $j \neq h$ with 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. Hereafter integrated intensity functions $A_{hj}, h, j \in \{1,2,...,k\}$ are supposed continuous, unless otherwise mentioned. Let

$$\mathbf{P}(s,t) = \prod_{(s,t]} (\mathbf{I} + d\mathbf{A}) \text{ for } s < t, s, t \in \mathcal{T}.$$

The matrix $\mathbf{P}(s,t) = (P_{hj}(s,t), h, j \in \{1,2,...,k\})$ is the $k \times k$ transition matrix of the Markov process.

FIGURE 1.1: Transition diagram for a multi-state Markov model



To incorporate heterogeneity between patients we let the transition intensities depend on a covariate vector $\mathbf{z}(t)$ through a Cox regression model

$$\alpha_{hj}(t \mid \mathbf{z}(t)) = \alpha_{hj0}(t)\exp(\beta' \mathbf{z}_{hj}(t))$$

with one vector of regression coefficients $\beta = (\beta_1, \beta_2, \ldots \beta_p)$ and type-specific covariate $\mathbf{z}_{hj}(t) = (z_{hj1}(t), z_{hj2}(t), \ldots, z_{hjp}(t))$ computed from the vector $\mathbf{z}(t)$ of basic covariates, consistent with the results of Chapter VII in Andersen *et al*[25]. Technically, $\mathbf{z}_{hj}(t)$ are everywhere assumed to be predictable and locally bounded, but precise assumptions will

be presented later in Section 1.3 of this chapter. Let $A_{hj0}(t) = \int_0^t \alpha_{hj0}(s)ds$ be the integrated baseline intensity for transitions from state h to state j and $A_{hh0} = -\sum_{j \neq h} A_{hj0}$.

Associated with X is a counting process $N_{hj}(t)$ which denotes the number of direct transitions from state h to j in the time interval [0, t],

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

The cumulative information revealed up to time t is the sigma-algebra $\mathcal{F}_t$ generated by $X(0)$ and $\{N_{hj}(s), s \leq t, h \neq j, h, j \in \mathrm{E}\}$. Introduce the indicator function $Y_h(t) = [X(t-) = h]$ to denote whether the process is in state h just prior to time t. Then with respect to the filtration $\{\mathcal{F}_t : t \geq 0\} : t \geq 0\}$, the multivariate counting process $N = \{N_{hj}, h \neq j\}$ has random intensity process $\{\lambda_{hj}, h \neq j, h, j \in \mathrm{E}\}$, where $\lambda_{hj}(t) = \alpha_{hj}(t)Y_h(t)$. Moreover,

$$M_{hj}(t) = N_{hj}(t) - \int_0^t Y_h(u)\alpha_{hj}(u)du, \; h \neq j, h, j \in \mathrm{E}$$

are zero mean local square integrable martingales. These results were first proved by Jacobsen (1982)[55] and are summarized in Theorem II.6.8, p94, Andersen *et al.* (1993)[25]. Using the continuity of $A_{hj}, h, j \in \{1, 2, ..., k\}$, it can be shown that $\langle M_{hj}, M_{qr} \rangle = 0$ for all pairs (h, j) and (q, r) with (h, j) $\neq$ (q, r). Here we denote by $\langle M, M' \rangle$ the predictable covariation process of M and M'. 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 $\lambda = (\lambda_{hj}, h \neq j)$, with $\lambda_{hj}(t) = Y_h(t)\alpha_{hj}(t)$.

1.1.1 Marked point processes

Over the follow up period the subject transits to other health states, $X_1, X_2, \ldots$ at random times $T_1, T_2, \ldots$ and these *transition* or *epoch* times $\{T_n : n \geq 0\}$ and health states $\{X_n : n \geq 0\}$ describe the event history of each patient. Formally, these are defined in terms of the *forward recurrence time* $W(t) = \inf\{s > 0 : X(t+s) \neq X(t)\}$, which is the waiting time from t until the next transition out of the state $X(t)$. Having set $X_0 = X(0), T_0 = 0, W(\infty) = \infty$, we define for all $n \geq 0$:

$$T_{n+1} = T_n + W(T_n) \text{ with } X_{n+1} = X(T_{n+1}) \text{ if } T_{n+1} < \infty \text{ and } X_{n+1} = X_n \text{ if } T_{n+1} = \infty.$$

The sojourn at the nth transition is $W(T_n)$ in state X_n. The survival model is an example of a two-state process with a single transient state "alive" and a single absorbing state "dead" with survival time $T = T_1$ and $T_n = \infty$ for $n \geq 2$. The multi-state analog of survival time is the time to absorption in state k given by $\tau_k = \inf\{t > 0 : X(t) = k\}$.

Suppose that R_n is a general mark corresponding to T_n and we define the marked point process

$$N(t, A) = \sum_{n \geq 1} [T_n \leq t, R_n \in A]$$

where A is a subset in the range of the R_n. For our case we could think of $R_n = (X_{n-1}, X_n)$ with values in $\mathrm{R} = \{(h, j) : h, j \in E, h \neq j\}$. Our previously described counting process $N_{hj}(t)$ can be identified with $N(t, A)$ by taking $A = \{h, j\}$. The natural filtration $\mathcal{F}_t$ is generated by $\{N(s, A) : 0 \leq s \leq t, A \subset \mathrm{R}\}$.

We also have[56]

$$\mathcal{F}_t = \sigma\{N(s,A): 0 \le s \le t, A \subset \mathrm{R}\} = \sigma\{(T_n, X_n): 0 \le T_n \le t\}.$$

We can endow R with an appropriate σ-field and regard $A \to N(t,A)$ for each t as a random measure, and $t \to N(t,A)$ as a counting process for each A. $\{\mathcal{F}_t, t \ge 0\}$ is a right-continuous filtration and $\mathcal{F}_\infty = \sigma(\bigcup_{t \ge 0} \mathcal{F}_t)$, where $\mathcal{F}_0 = \sigma(X(0), \mathbf{z})$.

For a stopping time T with respect to $\mathcal{F}_t, t \ge 0$ we define

$$\mathcal{F}_T = \{B \in \mathcal{F}_\infty : B \cap [T \le t] \in \mathcal{F}_t \ \textit{for all}\ t\}$$

$$\mathcal{F}_{T-} = \{B \cap [T > t] : t \ge 0,\ B \in \mathcal{F}_t\}$$

We have $\mathcal{F}_{t-} = \mathcal{F}_{T_{n-1}}$ for all $T_{n-1} < t \le T_n$ and

$$\mathcal{F}_{T_n} = \sigma\{(T_k, X_k): 1 \le k \le n\},$$

$$\mathcal{F}_{T_n-} = \sigma\{(T_k, X_k): 1 \le k \le n-1; T_n\}.$$

We have also $Y_h(t) = [X_{n-1} = h]$ for $t \in (T_{n-1}, T_n]$. Then with respect to the filtration $\{\mathcal{F}_t : t \ge 0\}$ the compensator $\Lambda_{hj}(t)$ of the process $N_{hj}(t)$ is given by

$$\Lambda_{hj}(dt) = \frac{P[V_n \in du, X_n = j \mid \mathcal{F}_{T_{n-1}}]}{P[V_n \ge u \mid \mathcal{F}_{T_{n-1}}]}[X_{n-1} = h] \text{ on}$$

where $V_n = T_n - T_{n-1}$ and $u = t - T_{n-1}$ is the duration of the current sojourn at time t.

Our Markov assumption gives

$$\Lambda_{hj}(dt) = [X_{n-1} = h] P[T_n \in [t, t+dt), X_n = j \mid T_{n-1}, X_{n-1}] = [X_{n-1} = h]\alpha_{hj}(t)dt,$$

for $t \in (T_{n-1}, T_n]$.

1.1.2 Incorporating censoring

If observation of X is ceased after some random time U, independent of X, we will need to replace $N_{hj}(t)$ by the censored process

$N_{hj}^{c}(t) = \#\{s \le t \wedge U : X(s-) = h, X(s) = j\}, h \neq j$ and

$Y_h(t)$ by $Y_h^{c}(t) = [X(t-) = h, U \ge t]$.

Then, with respect to an expanded filtration the aforementioned martingale property still obtains. For the ith patient we observe a basic covariate $\mathbf{z}_i(t)$, an initial state $X_i(0)$, the state indicator $Y_{hi}^{c}(t) = [X_i(t-) = h, U_i \ge t]$ and $N_{hji}^{c}(t)$, the number of event transition times before t from state h to j. Let $Y_h^{c}(t) = \sum_{i=1}^{n} Y_{hi}^{c}(t)$ the number of subjects who were not censored at time t and just before t were in state h and $N_{hj}^{c}(t) = \sum_{i=1}^{n} N_{hji}^{c}(t)$.

In terms of marked-point processes we define the marked point process $N^{c}(t, A) = \sum_{n \ge 1} [T_n \le t \wedge U, R_n \in A]$ where A is a subset in the range of the R_n. Our previously described counting process $N_{hj}^{c}(t)$ can be identified with $N^{c}(t, A)$ by taking $A = \{h, j\}$. The natural filtration $\mathcal{F}_t^{c}$ is $\mathcal{F}_t^{c} = \sigma\{N^{c}(s, A) : 0 \le s \le t, A \subset \mathrm{R}\}$ $= \sigma\{(T_n, X_n) : 0 \le T_n \le t \wedge U\}$.

With respect to the filtration $\{\mathcal{F}_t^{c} : t \ge 0\}$ the compensator $\Lambda_{hj}^{c}(t)$ of the multivariate counting process $N_{hj}^{c}(t)$ is given by

$$\Lambda_{hj}^{c}(dt) = \frac{P[V_n \in du, X_n = j, T_{n-1} \le U \mid \mathcal{F}_{T_{n-1}}^{c}]}{P[V_n \ge u, T_{n-1} \le U \mid \mathcal{F}_{T_{n-1}}^{c}]}[X_{n-1} = h]$$

on $t \in (T_{n-1}, T_n]$. Equivalently, the processes M_{hj}^{c} defined by $M_{hj}^{c} = N_{hj}^{c} - \Lambda_{hj}^{c}$ are martingales.

We assume that the censoring variable, U, is independent of everything else in the model. Then

$$\Lambda_{hj}^{c}(dt) = \frac{P[V_n \in du, X_n = j, T_{n-1} \le U \mid \mathcal{F}_{T_{n-1}}^{c}]}{P[V_n \ge u, T_{n-1} \le U \mid \mathcal{F}_{T_{n-1}}^{c}]}[X_{n-1} = h] =$$

$$= \frac{G(T_{n-1})P[V_n \in du, X_n = j \mid T_{n-1}, X_{n-1}]}{G(T_{n-1})P[V_n \ge u \mid T_{n-1}, X_{n-1}]}[X_{n-1} = h] = \Lambda_{hj}(dt)$$

and $N_{hj}^{c}(t)$ has the same compensator as $N_{hj}(t)$. Therefore conform to the Definition III.2.1 in Andersen *et al.* (1993)[25], the right-censoring of the process N generated by U is independent.

In the sequel we will assume that censoring has been accommodated in this way. Next we will incorporate costs in the model, the assumption of independence of costs of everything else in the model is often violated, for example the longer the length of stay in the hospital, the higher the costs. We will incorporate censoring without assuming independence.

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.

Let Z_0 be 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. 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:

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

We assume that:

A1 $C_{hj}(.)$ are bounded, non-negative processes over $\mathcal{T}$, adapted to $\{\mathcal{F}_t : t \in \mathcal{T}\}$, left continuous with right hand limits (so $C_{hj}(.)$ are bounded, predictable processes).

A2 For $t \in (0, \tau]$,

$$E(C_{hj}(t) \mid X_0 = i, X(t-) = h, Z_0) = E(C_{hj}(t) \mid X(t-) = h, Z_0),$$

so that at any $t > 0$ the expected transition costs do 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)[25]). Then, by assumption **A1**,

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

$$= E(\int_0^t e^{-rs} C_{hj}(s) Y_h(s) \alpha_{hj}(s) ds \mid X_0 = i, Z_0).$$

By Fubini's Theorem:

$$NPV_{hj}^{(1)}(t \mid i, Z_0) = \int_0^t e^{-rs} E(C_{hj}(s) Y_h(s) \mid X(0) = i, Z_0) \alpha_{hj}(s) ds.$$

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 assumption **A2**, $NPV_{hj}^{(1)}(t \mid i, Z_0)$ has the form

$$NPV_{hj}^{(1)}(t \mid i, Z_0) = \int_0^t e^{-rs} c_{hj}(s \mid Z_0) P_{ih}(0, s \mid Z_0) \alpha_{hj}(s) ds, \qquad (1.1)$$

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

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

$$NPV_h^{(2)}(s,d\mid i,Z_0)=E(C_h^{(2)}(s,d)\mid X_0=i,Z_0)=$$

$$=\int_s^{s+d} e^{-ru}E(B_h(u)Y_h(u)\mid X_0=i,Z_0)du.$$

Conditions similar to **A1** and **A2** are assumed for $B_h(.)$:

A3 $B_h(.)$ are bounded, non-negative real stochastic processes over $[0,\tau]$, adapted to $(\mathcal{F}_t)$.

A4 $E(B_h(u)\mid X_0=i,X(u-)=h,Z_0)=E(B_h(u)\mid X(u-)=h,Z_0)$ 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 **A4**:

$$NPV_h^{(2)}(s,d \mid i,Z_0) = \int_s^{s+d} e^{-ru} b_h(u \mid Z_0) P_{ih}(0,u \mid Z_0) du \qquad (1.2)$$

The right hand side of (1.1) may be interpreted intuitively as follows. Starting at s=0 in state i, a patient is in state h at time s with probability $P_{ih}(0,s \mid Z_0)$. Conditional on being in state h just prior to s, suppose a transition to state j occurs at s with intensity $\alpha_{hj}(s \mid Z_0)$ and this transition incurs a cost. Then (1.1) is the NPV for all $h \rightarrow j$ transition costs in [0, t]. Similarly, for the right hand of (1.2) consider the cost of sojourn in state h in the interval (s, s+ds]. This is $b_h(s \mid Z_0)ds$, conditional on reaching state h at s. To incur this cost a patient must move from the initial state i to h by time s, with probability $P_{ih}(0,s \mid Z_0)$. So (1.2) is the NPV of the total sojourn cost in state h in [0, t].

Suppose costs potentially accrue up to a fixed time horizon τ for sojourns in, and transitions among the transient states. If k is the only absorbing state costs would cease at the absorption time $\tau_k = \inf\{t > 0 : X(t) = k\}$ or τ whichever is observed first. The net present value of all expenditures is

$$NPV(i,Z_0) = \sum_{h \neq j} \int_0^{\tau} e^{-rs} c_{hj}(s \mid Z_0) \alpha_{hj}(s \mid Z_0) P_{ih}(0,s \mid Z_0) ds +$$

$$+ \sum_h \int_0^{\tau} e^{-rs} b_h(s \mid Z_0) P_{ih}(0,s \mid Z_0) ds \qquad (1.3)$$

where the dependence on the initial state $X_0 = i$ and the covariate profile is shown. The unconditional version is obtained by averaging over the initial distribution $\pi_j(0 \mid Z_0) = P(X_0 = j \mid Z_0)$ which yields

$$NPV(Z_0) = \sum_{i \in E} \pi_i(0 \mid Z_0) NPV(i,Z_0).$$

Comments

In the absence of covariates, the quantities (1.1) and (1.2) are ubiquitous in the insurance and actuarial literature[29] where $C_{hj}(t)$ is an assurance amount paid to the insured upon transition at time t from the insurance policy state h to state j. Then $C_{hj}^{(1)}(t)$ is the discounted value (at time 0) of all assurance benefits received in [0, t] for transitions $h \rightarrow j$ and (1.1) is its corresponding actuarial value given $X_0 = i$. Likewise, $B_h(t)$ is the annuity payment rate at time t in policy state h, $C_h^{(2)}(t)$ is the discounted value (at time 0) of all annuity payments received in [0, t] while the insured is in policy state h and (1.2) is its associated actuarial value. Usually $C_{hj}(t)$ and $B_h(t)$ are known non-random functions and one is interested in the total payment function $\sum_{h \neq j} C_{hj}^{(1)}(t) + \sum_h C_h^{(2)}(t)$. In this context Praestgaard (1991)[26] considers the estimation of (1.1) and (1.2) using a framework very similar to that we have described.

However, because costs are incurred in random amounts at random points in time during the course of a health care intervention the average expense functions $b_h(t \mid \mathbf{z})$ and $c_{hj}(t \mid \mathbf{z})$ are no longer known and need to be estimated from appropriate data along with the transition probabilities $P_{hj}(0, t \mid Z_0)$ and integrated transition functions $A_{hj}(t \mid Z_0) = \int_0^t \alpha_{hj}(s \mid Z_0) ds$. For easiness of notations we will assume the discount rate is null, i.e. $r = 0$, unless otherwise specified.

Insights into assumption A1

The assumption A1: $C_{hj}(.)$ are bounded, non-negative processes over $\mathcal{T}$, adapted to $\{\mathcal{F}_t : t \in \mathcal{T}\}$, left continuous with right hand limits (so $C_{hj}(.)$ are bounded, predictable processes) is not as naïve as it seems. Without it we would have to extend the observed history at time t to $\tilde{\mathcal{F}}_t$, the minimal σ-field generated by $\mathcal{F}_t$ and $\sigma\{C_{hj}(s) : s \le t,\ h \neq j,\ h, j \in \mathrm{E}\}$. With respect to this new σ-field one should wonder whether the compensator $\Lambda_{hj}(t)$ of the process $N_{hj}(t)$ can be estimated in the same way etc.

Some insights into assumption **A1** can be gained by viewing the cost C_n as a mark associated with the transition time T_n and describing the underlying process as a marked point process. Suppose that R_n is a general mark corresponding to T_n and we define the marked point process

$$N(t, A) = \sum_{n \ge 1} [T_n \le t, R_n \in A]$$

where A is a subset in the range of the R_n. For our case we could think of $R_n = (X_{n-1}, X_n, C_n)$ with values in $\mathrm{R} = \{(h, j, c) : h, j \in \mathrm{E},\ h \neq j, c > 0\}$. Our previously described counting process $N_{hj}(t)$ can be identified with $N(t, A)$ by taking $A = \{h, j, (0, \infty)\}$. The natural filtration $\tilde{\mathcal{F}}_t$ is generated by $\{N(s, A) : 0 \le s \le t, A \subset \mathrm{R}\}$. We can endow R with an appropriate σ-field and regard $A \to N(t, A)$ for each t as a random measure, and $t \to N(t, A)$ as a counting process for

each A. Moreover with respect to $\{\tilde{\mathcal{F}}_t, t \geq 0\}$ the compensator $\tilde{\Lambda}(t,A)$ of $N(t,A)$ is given by

$$\tilde{\Lambda}(t,A) = \tilde{\Lambda}(T_{n-1},A) + \int_0^{t-T_{n-1}} \frac{dF_n(u,A)}{1-F_n(u-,\mathrm{R})},\ u \in (0, T_n - T_{n-1}]$$

where $F_n(u,A)$ is the conditional distribution $F_n(u,A) = P[T_n - T_{n-1} \leq u, R_n \in A \mid \tilde{\mathcal{F}}_{T_{n-1}}]$ and $\tilde{\mathcal{F}}_{T_{n-1}} = \sigma\{(T_j,R_j): 1 \leq j \leq n-1\}$. For $t \in (T_{n-1}, T_n]$ we have

$$d\tilde{\Lambda}(t,A) = \frac{dF_n(u,A)}{dF_n(u,\mathrm{R})} \cdot \frac{dF_n(u,\mathrm{R})}{1-F_n(u-,\mathrm{R})}$$

where $u = t - T_{n-1}$ is the duration of the current sojourn at time t. The first term $\frac{dF_n(u,A)}{dF_n(u,\mathrm{R})}$ can be interpreted as the conditional probability of $R_n \in A$ given $\tilde{\mathcal{F}}_{T_{n-1}}$ and $T_n - T_{n-1} = u$. The second term $\frac{dF_n(u,\mathrm{R})}{1-F_n(u-,\mathrm{R})}$ is the conditional hazard rate for the sojourn $T_n - T_{n-1}$ given $\tilde{\mathcal{F}}_{T_{n-1}}$.

In the case of interest $R_n = (X_{n-1}, X_n, C_n)$. Then writing $V_n = T_n - T_{n-1}$ and taking $A = \{h, j, (y, y+dy]\}$ we can express

$$\frac{dF_n(u,A)}{dF_n(u,\mathrm{R})} = [X_{n-1} = h] P[C_n \in dy \mid V_n = u, X_n = j, \tilde{\mathcal{F}}_{T_{n-1}}] P[X_n = j \mid V_n = u, \tilde{\mathcal{F}}_{T_{n-1}}]$$

and

$$\frac{dF_n(u,\mathrm{R})}{1-F_n(u-,\mathrm{R})} = [X_{n-1} = h] \frac{P[V_n \in du \mid \tilde{\mathcal{F}}_{T_{n-1}}]}{P[V_n \geq u \mid \tilde{\mathcal{F}}_{T_{n-1}}]}$$

so that on $t \in (T_{n-1}, T_n]$ and recalling that $u = t - T_{n-1}$,

$$\tilde{\Lambda}(dt,A) = P[C_n \in dy \mid V_n = u, X_n = j, \tilde{\mathcal{F}}_{T_{n-1}}] \frac{P[V_n \in du, X_n = j \mid \tilde{\mathcal{F}}_{T_{n-1}}]}{P[V_n \geq u \mid \tilde{\mathcal{F}}_{T_{n-1}}]} [X_{n-1} = h]. \quad (1.4)$$

With respect to the filtration $\{\mathcal{F}_t : t \geq 0\}$ the compensator $\Lambda_{hj}(t)$ of the process $N_{hj}(t)$ is given by

$$\Lambda_{hj}(dt) = \frac{P[V_n \in du, X_n = j \mid \mathcal{F}_{T_{n-1}}]}{P[V_n \geq u \mid \mathcal{F}_{T_{n-1}}]} [X_{n-1} = h] \quad \text{on} \quad t \in (T_{n-1}, T_n]. \quad (1.5)$$

Now $\tilde{\mathcal{F}}_{T_{n-1}} = \sigma\{(T_j, X_j, C_j) : j \leq n-1\}$ whereas $\mathcal{F}_{T_{n-1}} = \sigma\{(T_j, X_j) : j \leq n-1\}$. It is reasonable to assume that

(T_n, X_n) is independent of $\{C_j : j \leq n-1\}$ given $\mathcal{F}_{T_{n-1}}$,

or, at least that

$\{C_j : j \leq n-1\}$ is $\mathcal{F}_{T_{n-1}}$ -measurable.

If (T_n, X_n) is independent of $\{C_j : j \leq n-1\}$ then

$$\frac{P[V_n \in du, X_n = j \mid \tilde{\mathcal{F}}_{T_{n-1}}]}{P[V_n \geq u \mid \tilde{\mathcal{F}}_{T_{n-1}}]} [X_{n-1} = h] =$$

$$\frac{P[V_n \in du, X_n = j \mid \mathcal{F}_{T_{n-1}}, \{C_j, j \leq n-1\}]}{P[V_n \geq u \mid \mathcal{F}_{T_{n-1}}, \{C_j, j \leq n-1\}]} [X_{n-1} = h] =$$

$$= \frac{P[V_n \in du, X_n = j \mid \mathcal{F}_{T_{n-1}}]}{P[V_n \geq u \mid \mathcal{F}_{T_{n-1}}]} [X_{n-1} = h] = \Lambda_{hj}(dt).$$

Since a similar conclusion can be found if we suppose $\{C_j : j \leq n-1\}$ is $\mathcal{F}_{T_{n-1}}$ -measurable (i.e. $\mathcal{F}_{T_{n-1}} = \tilde{\mathcal{F}}_{T_{n-1}}$) in any of these two cases (1.5) coincides with the second term in(1.4). Moreover, our Markov assumption makes

$\Lambda_{hj}(dt) = [X_{n-1} = h]\alpha_{hj}(t)dt$, $t \in (T_{n-1}, T_n]$. In fact taking $A = \{h, j, (0, \infty]\}$ we get

$\Lambda_{hj}(dt) = E[\Lambda(dt, A) \mid \mathcal{F}_{t-}]$. Also $\tilde{\mathcal{F}}_t$ is the minimal σ-field generated by $\mathcal{F}_t$ and $\sigma\{C_{hj}(s)\Delta N_{hj}(s) : s \le t,\ h \ne j,\ h, j \in \mathrm{E}\}$. The first term in (1.4) is the distribution of the cost C_n conditional on the past $\tilde{\mathcal{F}}_{T_{n-1}}$, the destination state $X_n = j$ and the transition time $T_n = t$.

With $A = \{h, j, [y, y + dy)\}$ the cost incurred in $[t, t + dt)$ is

$$\int_0^\infty yN(dt, A) = C_{hj}(t)\Delta N_{hj}(t)$$

and therefore the present value of costs incurred in $[0, t]$ due to transitions of the type $h \to j$ is

$$C_{hj}^{(1)}(t) = \int_0^t e^{-rs} \int_0^\infty yN(ds, A).$$

Ignoring discounting $C_{hj}^{(1)}(t) = \int_0^t \int_0^\infty yN(ds, A)$.

Let Z_0 a fixed covariate profile. Then using the martingale property of $N(t, A)$, the conditional expectation of this net present value given $X_0 = i, Z_0$ is

$$E(C_{hj}^{(1)}(t) \mid X_0 = i, Z_0) = E(\int_0^t e^{-rs} \int_0^\infty yN(ds, A) \mid X_0 = i, Z_0) =$$

$$= E(\int_0^t e^{-rs} \int_0^\infty y\tilde{\Lambda}(ds, A) \mid X_0 = i, Z_0).$$

Using (1.4), $E(C_{hj}^{(1)}(t) \mid X_0 = i, Z_0)$ becomes

$$E(\int_0^t e^{-rs} \int_0^\infty yP[C_n \in dy \mid V_n = u, X_n = j, \tilde{\mathcal{F}}_{T_{n-1}}][X_{n-1} = h]\alpha_{hj}(s)ds \mid X_0 = i, Z_0) =$$

$$= \int_0^t e^{-rs} E(\int_0^\infty yP[C_n \in dy \mid V_n = u, X_n = j, \tilde{\mathcal{F}}_{T_{n-1}}][X_{n-1} = h] \mid X_0 = i, Z_0)\alpha_{hj}(s \mid Z_0)ds.$$

With our previous notations, $C_{hj}(s) = [X_{n-1} = h] \int_0^\infty yP[C_n \in dy \mid V_n = u, X_n = j, \tilde{\mathcal{F}}_{T_{n-1}}]$.

Then

$E(C_{hj}^{(1)}(t) \mid X_0 = i, Z_0) =$

$\int_0^t e^{-rs} E(C_{hj}(s) \mid X(s-) = h, X_0 = i, Z_0) P_{ih}(0, s \mid Z_0) \alpha_{hj}(s \mid Z_0) ds$.

By **A2**, $E(\int_0^t \int_0^\infty yN(ds, A) \mid X(0) = i) = \int_0^t e^{-rs} c_{hj}(s \mid Z_0) P_{ih}(0, s \mid Z_0) \alpha_{hj}(s \mid Z_0) ds$ which is the same formula as (1.1), where $c_{hj}(t) = E(C_{hj}(t) \mid X(t-) = h)$.

We will estimate $A_{hj}(t \mid Z_0)$ and $P_{hj}(s, t \mid Z_0)$ from a Cox proportional hazards model for multiple states and $c_{hj}(s \mid Z_0)$ from a random-effects model. Putting everything together we are able to compute $\hat{NPV}_{hj}^{(1)}(t \mid i, Z_0)$ an estimator of $NPV_{hj}^{(1)}(t \mid i, Z_0)$ and show asymptotic properties for the derived estimator.

1.3 Estimation of transition probabilities

We now turn to the estimation of (1.1) and (1.2), focusing on the former. Andersen *et al* (1993)[25] pioneered an elegant asymptotic theory for estimators of β, $A_{hj}(t \mid Z_0)$ and $P_{hj}(s, t \mid Z_0)$. For each of n patients in a study we observe processes of the type described. For the ith patient the basic covariate vector is $\mathbf{z}_i(t)$, the initial

state $X_i(0)$, the state indicator $Y_{hi}(t) = [X_i(t-) = h, U_i \geq t]$ and

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

Conditionally on $\{ \mathbf{z}_i, X_i(0) : i = 1,\ldots,n \}$ assume the processes $\{ X_i(t) : t \in \mathcal{T} \}$ are independent and that the Cox regression model $\alpha_{hj}(t \mid \mathbf{z}(t)) = \alpha_{hj0}(t)\exp(\beta' \mathbf{z}_{hj}(t))$ described in Section 1.1 holds for each individual with the same baseline intensities, i.e.

$$\alpha_{hji}(t \mid \mathbf{z}_i(t)) = \alpha_{hj0}(t)\exp(\beta' \mathbf{z}_{hji}(t))$$

for all $i = 1,\ldots,n$. From now on denote by $N_{hj}(t)$ and $Y_h(t)$ respectively, the aggregated processes $\sum_{i=1}^{n} N_{hji}(t)$ and $\sum_{i=1}^{n} Y_{hi}(t)$.

The following standard notation will be used.

For any h

- $J_h(t) = [Y_h(t) \neq 0]$, $1 - J_h(t) = [Y_h(t) = 0] = [\sum_{i=1}^{n} Y_{hi}(t) = 0] = [Y_{hi}(t) = 0, \forall i]$

For $h \neq j$:

- $\mathbf{z}_{hji}(t)^{\otimes m} = \mathbf{z}_{hji}(t)\mathbf{z}'_{hji}(t)$, if $m = 2$;
- $\mathbf{z}_{hji}(t)^{\otimes m} = \mathbf{z}_{hji}(t)$ if $m = 1$ and $\mathbf{z}_{hji}(t)^{\otimes m} = 1$ if $m = 0$;
- $S_{hj}^{(m)}(t,\beta) = \sum_{i=1}^{n} Y_{hi}(t)\mathbf{z}_{hji}(t)^{\otimes m}\exp(\beta' \mathbf{z}_{hji}(t))$, $m \in \{0,1,2\}$;
- $E_{hj}(t,\beta) = S_{hj}^{(1)}(t,\beta) / S_{hj}^{(0)}(t,\beta)$;
- $V_{hj}(t,\beta) = S_{hj}^{(2)}(t,\beta) / S_{hj}^{(0)}(t,\beta) - E_{hj}(t,\beta)^{\otimes 2}$;
- $I(t,\beta) = \sum_{h \neq j} \int_0^t V_{hj}(u,\beta) dN_{hj}(u)$ with $N_{hj} = \sum_{i=1}^{n} N_{hji}$;

- $s_{hj}^{(m)}(t,\beta) = E[Y_{h1}(t)\mathbf{z}_{hj1}(t)^{\otimes m}\exp(\beta'\mathbf{z}_{hj1}(t))], \quad m \in \{0,1,2\}$;
- $e_{hj}(t,\beta) = s_{hj}^{(1)}(t,\beta)/s_{hj}^{(0)}(t,\beta)$;
- $v_{hj}(t,\beta) = s_{hj}^{(2)}(t,\beta)/s_{hj}^{(0)}(t,\beta) - e_{hj}^{\otimes 2}$
- $\hat{A}_{hj}(t \mid Z_0) = \hat{A}_{hj0}(t,\hat{\beta})\exp(\hat{\beta}'Z_{hj0})$ where the maximum likelihood estimator $\hat{\beta}$ is defined as the solution of the equation $U(\tau,\beta) = 0$ where

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

- $\Sigma(t,\beta) = \sum_{h\neq j}\int_0^t v_{hj}(u,\beta)s_{hj}^{(0)}(u,\beta)\alpha_{hj0}(u)du$, $p\times p$ nonrandom matrix does not epend on n.

The following assumptions will be adopted throughout this chapter. Although not all conditions are needed for every result, we state them all to avoid too many technical distractions in the theorems. We denote by $\| \cdot \|$ the supremum norm of a vector or a matrix, e.g. the norm of a vector $\mathbf{a} = (a_i)$ or a matrix $\mathbf{A} = (a_{ij})$ is $\| a \| = \sup_i | a_i |$ and $\| \mathbf{A} \| = \sup_{i,j} | a_{ij} |$, respectively. Convergence in probability and weak convergence are always as n tends to infinity.

Model Assumptions and Conditions:

A5 Conditional on $\mathbf{z}_i(\cdot)$, U_i is independent of $X_i(.)$;

A6 $\left(N_i(.), Y_i(.), \mathbf{z}_i(t)\right), 1 \le i \le n$ are independent identically distributed;

For $h \neq j$:

A7 $A_{hj0}(\tau) = \int_0^\tau \alpha_{hj0}(t)dt < \infty$;

A8 $\Sigma_\tau \underset{notation}{\overset{by}{=}} \Sigma(\tau,\beta) = \sum_{h\neq j} \int_0^\tau v_{hj}(u,\beta) s_{hj}^{(0)}(u,\beta)\alpha_{hj0}(u)du$ is positive definite.

There exist a compact neighborhood $\mathcal{B}$ of β, with $\beta \in \overset{o}{\mathcal{B}}$ (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 $[0,\tau]\times\mathcal{B}$ such that for $m \in \{0,1,2\}$ and $h, j \in \{1,...,k\}$, $h \neq j$:

A9 $\sup_{(t,\tilde{\beta})\in[0,\tau]\times\mathcal{B}} \left\| \frac{1}{n} S_{hj}^{(m)}(t,\tilde{\beta}) - s_{hj}^{(m)}(t,\tilde{\beta}) \right\| \xrightarrow{P} 0$ (Asymptotic Stability);

A10 $s_{hj}^{(m)}(.,.)$ are continuous functions of $\beta \in \mathcal{B}$ uniformly in $t \in [0,\tau]$ and bounded on $[0,\tau]\times\mathcal{B}$; $s_{hj}^{(0)}(.,\beta)$ is bounded away from zero on $[0,\tau]$ and

$s_{hj}^{(1)}(t,\beta) = \frac{\partial}{\partial\beta} s_{hj}^{(0)}(t,\beta), s_{hj}^{(2)}(t,\beta) = \frac{\partial}{\partial\beta} s_{hj}^{(1)}(t,\beta)$ (Asymptotic Regularity Conditions);

A11 There exists $\delta > 0$ such that

$n^{-1/2} \sup_{h\neq j,i,t} |\mathbf{z}_{hji}(t)| Y_{hi}(t)[\beta'\mathbf{z}_{hji}(t) > -\delta |\mathbf{z}_{hji}(t)|] \xrightarrow{P} 0$ (Lindeberg Condition)

where $[a > b] = 1$ if $a > b$ and zero otherwise;

Conditions **A7-A11** are implied in the independent identically distributed case (i.e. under assumptions **A5-A6**) by more general conditions as proven by Andersen *et al* (1982)[57] Theorem 4.1. These conditions are:

1. $\mathbf{z}_{hji}(\cdot)$ and $Y_{hj}(\cdot)$ are left continuous processes with right hand limits processes
2. $A_{hj0}(\tau) = \int_0^\tau \alpha_{hj0}(t)dt < \infty$

3. Σ_τ is positive definite

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

5. $E[\sup_{t,\tilde{\beta}} Y_{h1}(t) \,|\, \mathbf{z}_{hj1}(t)|^2 \exp(\tilde{\beta}'\mathbf{z}_{hj1}(t))] < \infty$ where the supremum is over

$t \in [0,\tau]$ and $\tilde{\beta} \in B(\beta)$ with $B(\beta)$ being some neighborhood of the true parameter β. Although not absolutely necessary, hereafter $\mathbf{z}_{hji}(.)$ are considered to be bounded predictable processes.

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.* (1983)[25]) is:

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

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 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 $N_{hj} = \sum_{i=1}^{n} N_{hji}$, $J_h(u) = [Y_h(u) > 0]$, $Y_h = \sum_{i=1}^{n} Y_{hi}$. 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

$A_0(t) = \left(A_{hj0}(t), h, j \in \{1,...,k\}\right)$ is estimated by

$\hat{A}_0(t, \hat{\beta}) = \left(\hat{A}_{hj0}(t, \hat{\beta}), h, j \in \{1,...,k\}\right)$.We define for a fixed covariate profile Z_0, (and corresponding type-specific covariate Z_{hj0}) $\hat{A}_{hj}(t \mid Z_0) = \hat{A}_{hj0}(t) \exp(\hat{\beta}' Z_{hj0})$, $h \neq j$,

$\hat{A}_{hh}(t \mid Z_0) = -\sum_{j \neq h} \hat{A}_{hj}(t \mid Z_0)$.

Central to all our proofs is the derivation of asymptotically equivalent representations of $\sqrt{n}(\hat{\beta} - \beta)$ and $\sqrt{n}(\hat{\mathbf{A}}(t \mid Z_0) - \mathbf{A}(t \mid Z_0))$ in terms of iid random variables. By asymptotically equivalence of two quantities is meant convergence in probability to zero of their difference (and where appropriate, uniformly for $t \in [0, \tau]$). We use the next theorem from Andersen *et al* (1993)[25].

Theorem 1 (Theorem VII.2.1, p497, Andersen *et al* (1993)[25])

Under **A5-A11**, the probability that the equation $U(\tau, \beta) = 0$ has a unique solution $\hat{\beta}$ tends to one and $\hat{\beta} \xrightarrow{P} \beta$ as $n \to \infty$. □

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

Theorem 2 (Theorem VII.2.2, p498, Andersen *et al* (1993)[25])

Assume **A5-A11**. Then $n^{1/2}(\hat{\beta}-\beta)$ converges in distribution to a zero mean normal p-dimensional random vector with covariance matrix Σ_τ^{-1} and

$\sup_{t\in[0,\tau]}\left\|n^{-1}I(t,\hat{\beta})-\Sigma(t,\beta)\right\| \xrightarrow{P} 0$. In particular $\hat{\Sigma}_\tau \underset{notation}{\overset{by}{=}} n^{-1}I(\tau,\hat{\beta}) \xrightarrow{P} \Sigma_\tau$. □

Following the proof of Theorem VII.2.2 of Andersen *et al* (1993)[25] we can show that the (h,j) element, $h \neq j$, of the $k\times k$ matrix $\sqrt{n}(\hat{\mathbf{A}}(t\,|\,Z_0)-\mathbf{A}(t\,|\,Z_0))$ is asymptotically equivalent to $\exp(\beta' Z_{hj0})\{X_{1hj}^{(n)}(t)+X_{2hj}^{(n)}(t)\}$ where

$$X_{1hj}^{(n)}(t)=\sqrt{n}(\hat{\beta}-\beta)'\int_0^t (Z_{hj0}-e_{hj}(u,\beta))\alpha_{hj0}(u)du$$

$$X_{2hj}^{(n)}(t)=\sqrt{n}\int_0^t \frac{J_h(u)}{S_{hj}^{(0)}(u,\beta)}dM_{hj}(u).$$

Here $M_{hj}(t)=N_{hj}(t)-\int_0^t S_{hj}^{(0)}(u,\beta)dA_{hj0}(u)$ and $M_{hj}(t)=\sum_{i=1}^{n} M_{hji}(t)$, where

$$M_{hji}(t)=N_{hji}(t)-\int_0^t Y_{hi}(u)\exp(\beta'\mathbf{z}_{hji}(u))dA_{hj0}(u).$$

Let $\mathbf{b}_{hj}(t)=\int_0^t (Z_{hj0}-e_{hj}(u,\beta))\alpha_{hj0}(u)du$ ($p\times 1$ vector).

Then $X_{1hj}^{(n)}(t)=\sqrt{n}(\hat{\beta}-\beta)'\mathbf{b}_{hj}(t)$.

We expand both $X_{1hj}^{(n)}$ and $X_{2hj}^{(n)}$ as sums of iid random variables.

We consider $X_{2hj}^{(n)}$ first. Indeed $X_{2hj}^{(n)}(t)=\sqrt{n}\int_0^t \frac{J_h(u)}{S_{hj}^{(0)}(u,\beta)}dM_{hj}(u)=$

$$= \frac{1}{\sqrt{n}} \int_0^t J_h(u)[\frac{n}{S_{hj}^{(0)}(u,\beta)} - \frac{1}{s_{hj}^{(0)}(u,\beta)}]dM_{hj}(u) +$$

$$+ \frac{1}{\sqrt{n}} \int_0^t \frac{dM_{hj}(u)}{s_{hj}^{(0)}(u,\beta)} - \frac{1}{\sqrt{n}} \int_0^t (1 - J_h(u)) \frac{dM_{hj}(u)}{s_{hj}^{(0)}(u,\beta)}.$$

For easiness of notations we will denote the three terms of the above sum as T_1, T_2, T_3,

where $T_1 = \frac{1}{\sqrt{n}} \int_0^t J_h(u)[\frac{n}{S_{hj}^{(0)}(u,\beta)} - \frac{1}{s_{hj}^{(0)}(u,\beta)}]dM_{hj}(u)$, $T_2 = \frac{1}{\sqrt{n}} \int_0^t \frac{dM_{hj}(u)}{s_{hj}^{(0)}(u,\beta)}$ and

$T_3 = \frac{1}{\sqrt{n}} \int_0^t (1 - J_h(u)) \frac{dM_{hj}(u)}{s_{hj}^{(0)}(u,\beta)}$.

To reduce the representation of $X_{2hj}^{(n)}$ as sum of iid variables, under the assumptions **A5 - A11** we prove that the first term, T_1 converges in probability to 0 and the third term, T_3 is null. Indeed, if we consider the first term

$$|T_1| = |\frac{1}{\sqrt{n}} \int_0^t J_h(u)[\frac{n}{S_{hj}^{(0)}(u,\beta)} - \frac{1}{s_{hj}^{(0)}(u,\beta)}]dM_{hj}(u)| \leq$$

$$\leq \frac{1}{\sqrt{n}} | \int_0^t J_h(u) \frac{s_{hj}^{(0)}(u,\beta) - \frac{S_{hj}^{(0)}(u,\beta)}{n}}{\frac{S_{hj}^{(0)}(u,\beta)}{n} s_{hj}^{(0)}(u,\beta)} dM_{hj}(u)| \leq$$

$$\leq \frac{1}{\sqrt{n}} | \int_0^t J_h(u) dM_{hj}(u)| \sup_{u \leq t} | s_{hj}^{(0)}(u,\beta) - \frac{S_{hj}^{(0)}(u,\beta)}{n}| \sup_{u \leq t} | \frac{1}{s_{hj}^{(0)}(u,\beta) \frac{S_{hj}^{(0)}(u,\beta)}{n}}|.$$

Using Central Limit Theorem for $\frac{1}{\sqrt{n}}\int_0^t J_h(u)dM_{hj}(u) = \frac{1}{\sqrt{n}}(\sum_{i=1}^{n}\int_0^t J_h(u)dM_{hji}(u))$,

$\sup_{u\le t}|s_{hj}^{(0)}(u,\beta) - \frac{S_{hj}^{(0)}(u,\beta)}{n}| \xrightarrow{P} 0$ (by Assumption **A9**) and $\sup_{u\le t}|s_{hj}^{(0)}(u,\beta)| < B < \infty$

(by Assumption **A10**) it follows that $|T_1| \xrightarrow{P} 0$.

The third term, T_3, is null:

$$T_3 = \frac{1}{\sqrt{n}}\int_0^t (1-J_h(u))\frac{dM_{hj}(u)}{s_{hj}^{(0)}(u,\beta)} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^t [Y_{hi}(u)=0,\forall i]\frac{dM_{hji}(u)}{s_{hj}^{(0)}(u,\beta)} =$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^t [Y_{hi}(u)=0,\forall i]\frac{dN_{hji}(u)}{s_{hj}^{(0)}(u,\beta)} -$$

$$-\sum_{i=1}^{n}\int_0^t [Y_{hi}(u)=0,\forall i]\frac{Y_{hi}(u)\exp(\beta' \mathbf{z}_{hji}(u))dA_{hj0}(u)}{s_{hj}^{(0)}(u,\beta)} = 0$$

Using $M_{hj}(u) = \sum_{i=1}^{n} M_{hji}(u)$, the second term, T_2 can be expanded as sum of iid random

variables as $T_2 = \frac{1}{\sqrt{n}}\int_0^t \frac{dM_{hj}(u)}{s_{hj}^{(0)}(u,\beta)} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^t \frac{dM_{hji}(u)}{s_{hj}^{(0)}(u,\beta)}$. Therefore we have showed that

$X_{2hj}^{(n)}(t) = \sqrt{n}\int_0^t \frac{J_h(u)}{S_{hj}^{(0)}(u,\beta)} dM_{hj}(u)$ is asymptotically equivalent to the sum of iid

random variables: $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^t \frac{dM_{hji}(u)}{s_{hj}^{(0)}(u,\beta)}$.

To obtain the derivation of $\sqrt{n}(\hat{\beta}-\beta)$, we use Theorems 1 and 2 stated above.

Consider now $X_{1hj}^{(n)}(t) = \sqrt{n}(\hat{\beta}-\beta)'\mathbf{b}_{hj}(t)$, where

$\mathbf{b}_{hj}(t) = \int_0^t (Z_{hj0} - e_{hj}(u,\beta))\alpha_{hj0}(u)du$ is a $p\times 1$ vector. We have

$\sqrt{n}(\hat{\beta}-\beta) \overset{a.e.}{=} \frac{1}{\sqrt{n}}\Sigma(\tau,\beta)^{-1}U(\tau,\beta)$, where by $a_n \overset{a.e.}{=} b_n$ we mean a_n is asymptotically equivalent to b_n. We can write $U(\tau,\beta)$ as a sum of independent identically distributed random variables as in the proof of Theorem 2 (see proof of Theorem VII.2.2 p498, Andersen *et al.* (1993)[25]):

$\frac{1}{\sqrt{n}}U(\tau,\beta)=\sum_{i=1}^{n}\sum_{\substack{h,j=1\\ h\neq j}}^{k}\int_0^\tau H_{hji}(u)dM_{hji}(u)$ where $H_{hji}(u)=\frac{1}{\sqrt{n}}(\mathbf{z}_{hji}(u)-E_{hj}(u,\beta))$ are predictable, locally bounded processes (p496 Andersen *et al* (1993) [25]).

Consequently,

$$\sqrt{n}(\hat{A}_{hj}(t\,|\,Z_0)-A_{hj}(t\,|\,Z_0)\overset{a.e.}{=}\exp(\beta' Z_{hj0})\{X_{1hj}^{(n)}(t)+X_{2hj}^{(n)}(t)\}=$$

$$=\exp(\beta' Z_{hj0})\{\sum_{i=1}^{n}(\sum_{\substack{h,j=1\\ h\neq j}}^{k}\int_0^\tau H_{hji}(u)dM_{hji}(u))'\Sigma(\tau,\beta)'^{-1}\mathbf{b}_{hj}(t)+$$

$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^t\frac{dM_{hji}(u)}{s_{hj}^{(0)}(u,\beta)}\} \quad (1.6)$$

We use:

$$\int_0^\tau H_{hji}(u)dM_{hji}(u)=\int_0^\tau\frac{1}{\sqrt{n}}(\mathbf{z}_{hji}(u)-E_{hj}(u,\beta))dM_{hji}(u)=$$

$$=\int_0^\tau\frac{1}{\sqrt{n}}(\mathbf{z}_{hji}(u)-e_{hj}(u,\beta))dM_{hji}(u)+\int_0^\tau\frac{1}{\sqrt{n}}(e_{hj}(u,\beta)-E_{hj}(u,\beta))dM_{hji}(u).$$

Since $\sup_{(u,\beta)\in[0,\tau]x\mathfrak{A}}|\frac{S_{hj}^{(m)}(u,\beta)}{n}-s_{hj}^{(m)}(u,\beta)|\xrightarrow{P}0$ and $s_{hj}(.,.)$ are uniformly continuous bounded functions on $[0,\tau]x\mathfrak{A}$ it follows

that $\sup_{(u,\beta)\in[0,\tau]\times\Omega} |e_{hj}(u,\beta)-E_{hj}(u,\beta)| \xrightarrow{P} 0$.

Therefore $\int_0^\tau \frac{1}{\sqrt{n}}(e_{hj}(u,\beta)-E_{hj}(u,\beta))dM_{hji}(u) \xrightarrow{P} 0$ and

$\sum_{i=1}^n \int_0^\tau H_{hji}(u)dM_{hji}(u) \overset{a.e.}{=} \sum_{i=1}^n \int_0^\tau \frac{1}{\sqrt{n}}(\mathbf{z}_{hji}(u)-e_{hj}(u,\beta))dM_{hji}(u)$. Therefore for $h \neq j$

we have proved the next lemma

Lemma 1

Under the assumptions **A5-A11**, $\sqrt{n}(\hat{A}_{hj}(t\,|\,Z_0)-A_{hj}(t\,|\,Z_0)$ is asymptotically

equivalent to

$$\exp(\beta' Z_{hj0})\{\sum_{i=1}^n(\sum_{\substack{h,j=1\\h\neq j}}^k \int_0^\tau \frac{1}{\sqrt{n}}(\mathbf{z}_{hji}(u)-e_{hj}(u,\beta))dM_{hji}(u))'\Sigma(\tau,\beta)^{-1}\mathbf{b}_{hj}(t)+$$

$$+\frac{1}{\sqrt{n}}\sum_{i=1}^n \int_0^t \frac{dM_{hji}(u)}{s_{hj}^{(0)}(u,\beta)}\}.$$

If we denote

$$S_{n,hj}^{(1)}(t,\tau)=\sum_{i=1}^n(\sum_{\substack{h,j=1\\h\neq j}}^k \int_0^\tau \frac{1}{\sqrt{n}}(\mathbf{z}_{hji}(u)-e_{hj}(u,\beta))dM_{hji}(u))'\Sigma(\tau,\beta)^{-1}\mathbf{b}_{hj}(t) \text{ and}$$

$$S_{n,hj}^{(2)}(t)=\frac{1}{\sqrt{n}}\sum_{i=1}^n \int_0^t \frac{dM_{hji}(u)}{s_{hj}^{(0)}(u,\beta)},$$

then

$$\sqrt{n}(\hat{A}_{hj}(t\,|\,Z_0)-A_{hj}(t\,|\,Z_0) \overset{a.e.}{=} \exp(\beta' Z_{hj0})\{S_{n,hj}^{(1)}(t)+S_{n,hj}^{(2)}(t)\}.\square$$

Comments

The k^2 components of $\mathbf{S}_n^{(2)}(.) = \left(S_{n,hj}^{(2)}(.), h, j \in \{1,...,k\}\right)$ are processes in $D[0,\tau]$, the space of real-valued right continuous functions with left-end limits on $[0,\tau]$. Equip $D[0,\tau]$ with the supremum norm and the σ – filed generated by the collection of open balls, and for product spaces their usual extensions. Regarded as multivariate processes on $(D[0,\tau])^{k^2}$, we can establish their weak convergence in the Skorohod topology. However, because the limiting processes have almost surely continuous paths, the convergence is true in the supremum norm as well.

Under assumptions **A5-A11**, the processes $\{S_{n,hj}^{(2)}(t), (h,j) \in E\}$ converge weakly to a process $U_2^*(t)$ (see proof of Theorem VII.2.3, p503, Andersen et al (1993)[25]). The limiting distribution of the $k \times k$ matrix-valued process $\mathbf{S}_n^{(2)}(.) = \left(S_{n,hj}^{(2)}(.), h, j \in \{1,...,k\}\right)$ is that of a $k \times k$ matrix-valued process $\mathbf{U}_2^*(.) = \left(U_{2hj}^*(.), h, j \in \{1,...,k\}\right)$, where $U_{2hh}^* = -\sum_{j \neq h} U_{2hj}^*$ and $\left\{U_{2hj}^*(.), (h,j) \in \{1,...,k\}\right\}$ is a continuous Gaussian vector martingale, with

i) $U_{2hj}^*(0) = 0$,

ii) $\left\langle U_{2hj}^*, U_{2mr}^* \right\rangle = 0$ for $(h,j) \neq (m,r), (h,j), (m,r) \in E$

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

iv) $Cov(U_{2hj}^*(s), U_{2hj}^*(t)) = \omega_{hj}^2(s \wedge t)$

The (h, j)-th element of the limit process $\mathbf{U}_1^*(t)$ of $\mathbf{S}_n^{(1)}(t,\tau)$ can be expressed as $\xi_0' \mathbf{b}_{hj}(t)$, where ξ_0 is a p-dimensional normal random variable (its distribution does not depend on t) with zero mean and covariance matrix $\Sigma(\tau,\beta)^{-1}$ and $\mathbf{b}_{hj}(t) = \int_0^t (Z_{hj0} - e_{hj}(u,\beta))\alpha_{hj0}(u)du$ is a p-dimensional vector depending on t. The processes U_{1hj}^* and U_{2hj}^* are asymptotically independent and the limiting distribution of $\sqrt{n}(\hat{A}_{hj}(t \mid Z_0) - A_{hj}(t \mid Z_0))$,

$$\exp(\beta' Z_{hj0})(U_{1hj}^*(t) + U_{2hj}^*(t)) \tag{1.7}$$

has mean zero and variance

$$\exp(2\beta' Z_{hj0})\{\omega_{hj}^2(t) + \mathbf{b}_{hj}(t)'\Sigma(\tau,\beta)^{-1}\mathbf{b}_{hj}(t)\} \tag{1.8}$$

The covariance matrix of $(S_{n,hj}^{(1)}(t,\tau), S_{n,hj}^{(2)}(t,\tau))$ can also be computed directly, using the definitions of $s_{hj}^{(m)}(u,\beta), m = 0,1,2$ and the independence of the processes $\{X_i(\cdot), i = 1,\ldots,n\}$. Then

$$n \operatorname{var}(S_{n,hj}^{(1)}(t,\tau)) = Var(\sum_{i=1}^{n}(\sum_{\substack{h,j=1 \\ h\neq j}}^{k}\int_0^\tau (\mathbf{z}_{hji}(u) - e_{hj}(u,\beta))dM_{hji}(u))'\Sigma(\tau,\beta)^{-1}\mathbf{b}_{hj}(t)) =$$

$$= \mathbf{b}_{hj}(t)'\Sigma(\tau,\beta)^{-1} Var(\sum_{i=1}^{n}(\sum_{\substack{h,j=1 \\ h\neq j}}^{k}\int_0^\tau (\mathbf{z}_{hji}(u) - e_{hj}(u,\beta))dM_{hji}(u))')\Sigma(\tau,\beta)^{-1}\mathbf{b}_{hj}(t) =$$

$$= \mathbf{b}_{hj}(t)'\Sigma(\tau,\beta)^{-1}\sum_{i=1}^{n} V\operatorname{ar}(\sum_{\substack{h,j=1 \\ h\neq j}}^{k}\int_0^\tau (\mathbf{z}_{hji}(u) - e_{hj}(u,\beta))dM_{hji}(u)))'\Sigma(\tau,\beta)^{-1}\mathbf{b}_{hj}(t) =$$

$$= \mathbf{b}_{hj}(t)'\Sigma(\tau,\beta)^{-1}\sum_{i=1}^{n}\sum_{\substack{h,j=1\\ h\neq j}}^{k} E\Big(\int_0^{\tau}(\mathbf{z}_{hji}(u)-e_{hj}(u,\beta))^{\otimes 2}Y_{hi}(u)\exp(\beta'\mathbf{z}_{hji}(u))\alpha_{hj0}(u)du\Big)'\cdot$$

$$\cdot\Sigma(\tau,\beta)^{-1}\mathbf{b}_{hj}(t) =$$

$$= n\mathbf{b}_{hj}(t)'\Sigma(\tau,\beta)^{-1}\Big(\sum_{h\neq j}\int_0^{\tau}(s_{hj}^{(2)}(u,\beta)-e_{hj}(u,\beta)s_{hj}^{(1)'}(u,\beta)-s_{hj}^{(1)}(u,\beta)e_{hj}'(u,\beta)+$$

$$+e_{hj}^{\otimes 2}(u,\beta)s_{hj}^{(0)}(u,\beta))\alpha_{hj0}(u)du\Big)\Sigma(\tau,\beta)^{-1}\mathbf{b}_{hj}(t) =$$

$$= n\mathbf{b}_{hj}(t)'\Sigma(\tau,\beta)^{-1}\Sigma(\tau,\beta)'\Sigma(\tau,\beta)^{-1}\mathbf{b}_{hj}(t) = n\mathbf{b}_{hj}(t)'\Sigma(\tau,\beta)^{-1}\mathbf{b}_{hj}(t).$$

For the last equality we use

$$s_{hj}^{(2)}(u,\beta)-e_{hj}(u,\beta)s_{hj}^{(1)'}(u,\beta)-s_{hj}^{(1)}(u,\beta)e_{hj}'(u,\beta)+e_{hj}^{\otimes 2}(u,\beta)s_{hj}^{(0)}(u,\beta) =$$

$$= s_{hj}^{(2)}(u,\beta)-\frac{s_{hj}^{(1)}(u,\beta)}{s_{hj}^{(0)}(u,\beta)}s_{hj}^{(1)'}(u,\beta)-s_{hj}^{(1)}(u,\beta)\frac{s_{hj}^{(1)'}(u,\beta)}{s_{hj}^{(0)}(u,\beta)}+$$

$$+\frac{s_{hj}^{(1)}(u,\beta)s_{hj}^{(1)'}(u,\beta)}{s_{hj}^{(0)2}(u,\beta)}s_{hj}^{(0)}(u,\beta) = v_{hj}(u,\beta)s_{hj}^{(0)}(u,\beta)$$

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

Also $n\,\mathrm{var}(S_{n,hj}^{(2)}(t,\tau)) = Var\Big(\sum_{i=1}^{n}\int_0^{t}\frac{dM_{hji}(u)}{s_{hj}^{(0)}(u,\beta)}\Big)=\sum_{i=1}^{n}Var\Big(\int_0^{t}\frac{dM_{hji}(u)}{s_{hj}^{(0)}(u,\beta)}\Big)=$

$$=\sum_{i=1}^{n}E\Big(\int_0^{t}\frac{1}{s_{hj}^{(0)}(u,\beta)^2}d<M_{hji},M_{hji}>(u)\Big)=$$

$$=\sum_{i=1}^{n}\int_0^{t}\frac{1}{s_{hj}^{(0)}(u,\beta)^2}E(Y_{hi}(u)\exp(\beta'\mathbf{z}_{hji}(u)))\alpha_{hj}(u)du=$$

$$= \int_0^t \frac{1}{s_{hj}^{(0)}(u,\beta)^2} \alpha_{hj}(u) E(\sum_{i=1}^{n} Y_{hi}(u) \exp(\beta' \mathbf{z}_{hji}(u))) du =$$

$$= n \int_0^t \frac{1}{s_{hj}^{(0)}(u,\beta)^2} \alpha_{hj}(u) s_{hj}^{(0)}(u,\beta) du = n \int_0^t \frac{1}{s_{hj}^{(0)}(u,\beta)} \alpha_{hj}(u) du = n\omega_{hj}^2(t).$$

Also $S_{n,hj}^{(1)}(t,\tau)$ and $S_{n,hj}^{(2)}(t)$ are asymptotically uncorrelated because the martingales M_{hji} and M_{hlk} are orthogonal for $i \neq k$ (by independence over subjects) and for $j \neq l$ (by continuity of the functions $A_{hj0}(t)$). Indeed

$$Cov(S_{n,hj}^{(1)}(t,\tau), S_{n,hj}^{(2)}(t)) =$$

$$= E\{[\sum_{i=1}^{n} (\sum_{\substack{h,j=1 \\ h \neq j}}^{n} \int_0^\tau \frac{1}{\sqrt{n}} (\mathbf{z}_{hji}(u) - e_{hj}(u,\beta)) dM_{hji}(u))' \Sigma(\tau,\beta)^{-1} \mathbf{b}_{hj}(t)] \cdot$$

$$\cdot [\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^t \frac{dM_{hji}(u)}{s_{hj}^{(0)}(u,\beta)}]\} =$$

$$= \frac{1}{n} \sum_{i=1}^{n} E\{[\sum_{\substack{h,j=1 \\ h \neq j}}^{k} \int_0^\tau (\mathbf{z}_{hji}(u) - e_{hj}(u,\beta)) dM_{hji}(u))' \Sigma(\tau,\beta)^{-1} \mathbf{b}_{hj}(t)][\int_0^t \frac{dM_{hji}(u)}{s_{hj}^{(0)}(u,\beta)}]\} =$$

$$= \frac{1}{n} \sum_{i=1}^{n} E\{[\int_0^\tau (\mathbf{z}_{hji}(u) - e_{hj}(u,\beta)) dM_{hji}(u))' \Sigma(\tau,\beta)^{-1} \mathbf{b}_{hj}(t)][\int_0^t \frac{dM_{hji}(u)}{s_{hj}^{(0)}(u,\beta)}]\} =$$

$$= \frac{1}{n} \sum_{i=1}^{n} E\{E[\int_0^\tau (\mathbf{z}_{hji}(u) - e_{hj}(u,\beta)) dM_{hji}(u))' \Sigma(\tau,\beta)^{-1} \mathbf{b}_{hj}(t) \int_0^t \frac{dM_{hji}(u)}{s_{hj}^{(0)}(u,\beta)}) | \Im_t]\} =$$

$$= \frac{1}{n} \sum_{i=1}^{n} E\{[\int_0^t (\mathbf{z}_{hji}(u) - e_{hj}(u,\beta)) dM_{hji}(u))' \Sigma(\tau,\beta)^{-1} \mathbf{b}_{hj}(t)][\int_0^t \frac{dM_{hji}(u)}{s_{hj}^{(0)}(u,\beta)}]\}.$$

Using $d < M_{hji}, M_{hji} > (u) = Y_{hi}(u) \exp(\beta' \mathbf{z}_{hji}(u)) \alpha_{hj}(u) du$ and Fubini Theorem, the covariance of the two sums becomes

$$Cov(S^{(1)}_{n,hj}(t,\tau), S^{(2)}_{n,hj}(t)) =$$

$$= \frac{1}{n}\sum_{i=1}^{n} E\{ \int_0^t \frac{1}{s^{(0)}_{hj}(u,\beta)} (\mathbf{z}_{hji}(u) - e_{hj}(u,\beta)))' d < M_{hji}, M_{hji} > (u) \Sigma(\tau,\beta)^{-1} \mathbf{b}_{hj}(t)\} =$$

$$= \frac{1}{n} E\{ (\int_0^t \frac{1}{s^{(0)}_{hj}(u,\beta)} [\sum_{i=1}^{n} Y_{hi}(u) \exp(\beta' \mathbf{z}_{hji}(u)) (\mathbf{z}_{hji}(u) - e_{hj}(u,\beta))'] \alpha_{hj}(u) du) \cdot$$

$$\cdot \Sigma(\tau,\beta)^{-1} \mathbf{b}_{hj}(t)\} =$$

$$= \frac{1}{n} \int_0^t \frac{1}{s^{(0)}_{hj}(u,\beta)} (s^{(1)}_{hj}(u,\beta) - e_{hj}(u,\beta) s^{(0)}_{hj}(u,\beta)) \Sigma(\tau,\beta)^{-1} \mathbf{b}_{hj}(t) = 0.$$

The last integral is null since, by definition, $e_{hj}(t,\beta) = s^{(1)}_{hj}(t,\beta) / s^{(0)}_{hj}(t,\beta)$.

The variance of $A_{hh} = -\sum_{j \neq h} A_{hj}$ can be calculated using formula above. Indeed,

$$\sqrt{n}(\hat{A}_{hh}(t \mid Z_0) - A_{hh}(t \mid Z_0) = -\sum_{j \neq h} \sqrt{n}(\hat{A}_{hj}(t \mid Z_0) - A_{hj}(t \mid Z_0) \overset{a.e.}{=}$$

$$\overset{a.e.}{=} -\sum_{j \neq h} \exp(\beta' Z_{hj0}) \{ S^{(1)}_{n,hj}(t,\tau) + S^{(2)}_{n,hj}(t) \}.$$

Using definitions of $S^{(i)}_{n,hj}(t,\tau), i = 1,2$ and same arguments, the variance of the last sum is

$$Var\{ -\sum_{j \neq h} \exp(\beta' Z_{hj0}) [\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\sum_{\substack{h,j=1 \\ h \neq j}}^{k} \int_0^\tau \frac{1}{\sqrt{n}} (\mathbf{z}_{hji}(u) - e_{hj}(u,\beta)) dM_{hji}(u))' \Sigma(\tau,\beta)^{-1} \mathbf{b}_{hj}(t) +$$

$$+ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_0^t \frac{dM_{hji}(u)}{s^{(0)}_{hj}(u,\beta)}] \} =$$

$$= Var(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\sum_{j\neq h}\exp(\beta' Z_{hj0})\cdot\{\sum_{\substack{h,j=1\\ h\neq j}}^{k}\int_0^{\tau}\frac{1}{\sqrt{n}}(\mathbf{z}_{hji}(u)-e_{hj}(u,\beta))dM_{hji}(u))'\}\Sigma(\tau,\beta)^{-1}\mathbf{b}_{hj}(t)+$$

$$+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^{\tau}\frac{dM_{hji}(u)}{s_{hj}^{(0)}(u,\beta)}\}=$$

$$=\{\sum_{j\neq h}\exp(\beta' Z_{hj0})\mathbf{b}_{hj}(t)\}'\Sigma(\tau,\beta)^{-1}\{\sum_{j\neq h}\exp(\beta' Z_{hj0})\mathbf{b}_{hj}(t)\}+\sum_{j\neq h}\exp(2\beta' Z_{hj0})\omega_{hj}^2(t).$$

Using these results one also obtains the convergence of

$\sqrt{n}(\mathbf{P}(s,t\,|\,Z_0)-\mathbf{P}(s,t\,|\,Z_0))$. The Aalen -Johansen-estimator of the transition probabilities is given by

$$\hat{\mathbf{P}}(s,t\,|\,Z_0)=\prod_{(s,t]}(\mathbf{I}+d\hat{\mathbf{A}}(u\,|\,Z_0)),$$

this estimate being meaningful as long as $\Delta\hat{A}_{hh}(u\,|\,Z_0)\geq -1$ on $(s,t]$. Here $\hat{\mathbf{A}}(t\,|\,Z_0)$ is the $k\times k$ matrix of elements $\hat{A}_{hj}(t\,|\,Z_0)$ with $\hat{A}_{hh}(t\,|\,Z_0)=-\sum_{j\neq h}\hat{A}_{hj}(t\,|\,Z_0)$. If a transition occurs at time u, then $\mathbf{I}+d\hat{\mathbf{A}}(u\,|\,\mathbf{z})$ is the matrix whose (h,j)-th element is $\Delta N_{hj}(u)/Y_h(u)$ if $h\neq j$ and equal to $1-\sum_{j\neq h}\Delta N_{hj}(u)/Y_h(u)$ if $h=j$. The properties of $\hat{\mathbf{P}}(s,t)$ follow from those of $\hat{\mathbf{A}}$ as will see below.

It can be shown that

$$\sqrt{n}(\mathbf{P}(s,t\,|\,Z_0)-\mathbf{P}(s,t\,|\,Z_0))=\sqrt{n}(\prod_{s<u\leq t}(\mathbf{I}+d\hat{\mathbf{A}}(u\,|\,Z_0)-\prod_{s<u\leq t}(\mathbf{I}+d\mathbf{A}(u\,|\,Z_0))$$

is asymptotically equivalent to $\sqrt{n}\int_s^t\mathbf{P}(s,u\,|\,Z_0)d(\hat{\mathbf{A}}-\mathbf{A})(u)\mathbf{P}(u,t\,|\,Z_0)$.

Therefore

$$\sqrt{n}(\hat{\mathbf{P}}(s,t\mid Z_0)-\mathbf{P}(s,t\mid Z_0)) \overset{a.e.}{=} \sqrt{n}\int_s^t \mathbf{P}(s,u\mid Z_0)d(\mathbf{S}_n^{(1)}(u)+\mathbf{S}_n^{(2)}(u))\mathbf{P}(u,t\mid Z_0) \quad (1.9)$$

where we define the matrices $\mathbf{S}_n^{(1)}(t)=\{S_{n,hj}^{(1)}(t,\tau)\}_{h,j}$ and $\mathbf{S}_n^{(2)}(t)=\{S_{n,hj}^{(2)}(t)\}_{h,j}$.

It follows that

$$\sqrt{n}(\hat{P}_{ih}(s,t\mid Z_0)-P_{ih}(s,t\mid Z_0)) \overset{a.e.}{=}$$

$$=\sqrt{n}\{\sum_{g=1}^{k}\sum_{l\neq g}^{k}\int_s^t P_{ig}(s,u\mid Z_0)d(S_{n,gl}^{(1)}(u)+S_{n,gl}^{(2)}(u))P_{lh}(u,t\mid Z_0)+$$

$$+\sum_{g=1}^{k}\int_s^t P_{ig}(s,u\mid Z_0)d(S_{n,gg}^{(1)}(u)+S_{n,gg}^{(2)}(u))P_{gh}(u,t\mid Z_0)\}. \quad (1.10)$$

Since $S_{n,gg}^{(1)}(u)+S_{n,gg}^{(2)}(u)=-(\sum_{l\neq g}S_{n,gl}^{(1)}(u)+S_{n,gl}^{(2)}(u))$ we have

$$\sqrt{n}(\hat{P}_{ih}(s,t\mid Z_0)-P_{ih}(s,t\mid Z_0)) \overset{a.e.}{=}$$

$$=\sqrt{n}\sum_{g=1}^{k}\sum_{l\neq g}^{k}\int_s^t P_{ig}(s,u\mid Z_0)d(S_{n,gl}^{(1)}(u)+S_{n,gl}^{(2)}(u))(P_{lh}(u,t\mid Z_0)-P_{gh}(u,t\mid Z_0)) \quad (1.11)$$

If we replace $S_{n,gl}^{(1)}(u)$ and $S_{n,gl}^{(2)}(u)$ by sums of independent identically distributed variables we will get an independent identically distributed representation for $\sqrt{n}(\hat{P}_{ih}(s,t\mid Z_0)-P_{ih}(s,t\mid Z_0))$. Here we use

$$dS_{n,hj}^{(1)}(t,\tau)=\sum_{i=1}^{n}(\sum_{\substack{h,j=1\\ h\neq j}}^{k}\int_0^\tau \frac{1}{\sqrt{n}}(\mathbf{z}_{hji}(u)-e_{hj}(u,\beta))dM_{hji}(u))'\Sigma(\tau,\beta)'^{-1}d\mathbf{b}_{hj}(t) \text{ and}$$

$dS^{(2)}_{n,hj}(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{dM_{hji}(t)}{s^{(0)}_{hj}(t,\beta)}$. Using Theorem VII.2.3, p503, Andersen et al (1993)[25] and (1.11), the limiting distribution of $\sqrt{n}(\hat{P}_{ih}(s,t \mid Z_0) - P_{ih}(s,t \mid Z_0))$ is given by the next theorem.

Theorem 3

Under **A5 - A11**, $\sqrt{n}(\hat{P}_{ih}(s,t \mid Z_0) - P_{ih}(s,t \mid Z_0))$ converges weakly to $U_{1ih}(s,t \mid Z_0) + U_{2ih}(s,t \mid Z_0)$, where

$$U_{1ih}(s,t \mid Z_0) = \sqrt{n} \sum_{g=1}^{k} \sum_{l \neq g}^{k} \int_s^t P_{ig}(s,u \mid Z_0)(P_{lh}(u,t \mid Z_0) - P_{gh}(u,t \mid Z_0)) dU^{(*)}_{1gl}(u) \text{ and}$$

$$U_{2ih}(s,t \mid Z_0) = \sqrt{n} \sum_{g=1}^{k} \sum_{l \neq g}^{k} \int_s^t P_{ig}(s,u \mid Z_0)(P_{lh}(u,t \mid Z_0) - P_{gh}(u,t \mid Z_0)) dU^{(*)}_{2gl}(u).$$

The processes $U_{1ih}(s,t \mid Z_0)$, $U_{2ih}(s,t \mid Z_0)$ are independent. □

Next we calculate the asymptotic covariance function of $n^{1/2}(\hat{\mathbf{P}}(s,t \mid Z_0) - \mathbf{P}(s,t \mid Z_0))$. Because $\mathbf{S}^1_{(n)}(t)$ and $\mathbf{S}^2_{(n)}(t)$ are asymptotically uncorrelated the $k^2 \times k^2$ covariance matrix of (1.9) is the sum of two terms. From (1.6) the (i, h)-th element of the first term in (1.9) is

$$n^{1/2}(\hat{\beta} - \beta)' \sum_{g,l,g \neq l} \exp(\beta' Z_{gl0}) \int_s^t P_{ig}(s,u \mid Z_0) d\mathbf{b}_{gl}(u) P_{lh}(u,t \mid Z_0) = n^{1/2}(\hat{\beta} - \beta)' \mathbf{a}_{ih}$$

where $\mathbf{b}_{gl}(t) = \int_0^t (Z_{gl0} - e_{gl}(u,\beta)) \alpha_{gl0}(u) du$ is a p-dimensional vector depending on t,

and $\mathbf{a}_{ih} = \sum_{g,l,g\neq l} \exp(\beta' Z_{gl0}) \int_s^t P_{ig}(s,u \mid Z_0) d\mathbf{b}_{gl}(u) P_{lh}(u,t \mid Z_0)$. Therefore the asymptotic covariance of the (i, h)-th and (q, r)-th elements is $\mathbf{a}'_{ih}\Sigma(\beta,\tau)^{-1}\mathbf{a}_{qr}$. In order to estimate this covariance we replace $\Sigma(\beta,\tau)$ by $n^{-1}I(\hat{\beta},\tau)$ and $\mathbf{a}_{ih}$ by

$\sum_{g,l} \exp(\hat{\beta}' Z_{gl0}) \int_s^t \hat{P}_{ig}(s,u \mid Z_0) d\hat{\mathbf{b}}_{gl}(u) \hat{P}_{lh}(u,t \mid Z_0)$, where

$\hat{\mathbf{b}}_{gl}(t) = \int_0^t \{Z_{gl0} - E_{gl}(u,\hat{\beta})\} d\hat{A}_{gl0}(u)$.

For the second term in (1.9) the asymptotic covariance has the form

$\int_s^t \mathbf{P}'(u,t \mid Z_0) \otimes \mathbf{P}(s,u \mid Z_0) Cov(vec(d\mathbf{S}^2_{(n)}(u))) \mathbf{P}(u,t \mid Z_0) \otimes \mathbf{P}'(s,u \mid Z_0)$. The inner covariance matrix is expressible as $\sum_l \sum_{g\neq l} \exp(2\beta' Z_{gl0}) \{vec\, C_{gl}\} \{vec\, C_{gl}\}' \int_s^t \frac{dA_{gl0}(u)}{s^{(0)}_{gl}(\beta,u)}$

where C_{gl} is a $m \times m$ matrix with (g, l)-th element equal to 1, (g, g)-th element equal to -1, and all other elements zero to zero. Combined with the previous expression we get a compact form for the covariance of the first term in (1.9), namely

$$\sum_l \sum_{g\neq l} \exp(2\beta' Z_{gl0}) \int_s^t \{vec\, \mathbf{P}(s,u \mid Z_0) C_{gl} \mathbf{P}(u,t \mid Z_0)\} \cdot$$

$$\cdot \{vec\, \mathbf{P}(s,u \mid Z_0) C_{gl} \mathbf{P}(u,t \mid Z_0)\}' \frac{dA_{gl0}(u)}{s^{(0)}_{gl}(\beta,u)}$$

Therefore the asymptotic covariance of $n^{1/2}(\hat{\mathbf{P}}(s,t \mid Z_0) - \mathbf{P}(s,t \mid Z_0))$ is the sum of this expression and the covariance matrix of terms $\mathbf{a}'_{ih}\Sigma(\beta,\tau)^{-1}\mathbf{a}_{qr}$. Chapter VII, pages 514-515, Andersen *et al* (1993) gives formulas for estimated covariance of $\hat{\mathbf{P}}_{ih}(s,t \mid Z_0)$ and $\hat{\mathbf{P}}_{qr}(s,t \mid Z_0)$.

1.4 Estimation of average transition costs

We are now left with the estimation of $c_{hj}(t \mid Z_0)$ in (1.1). Suppose costs can potentially be incurred up to earliest of a fixed time τ or τ_k, the time to absorption in state *k*. If all observation ends at τ then we are restricted to transitions and their associated costs that are observed by τ. However, allowing for censoring of follow up at time *U*, the period of observation actually does not exceed $U \wedge \tau \wedge \tau_k$. We now introduce some new notations.

Let y_{ij} be the transition costs in the *i*th patient at the chronologically ordered transition times t_{ij}, $j = 1,...,n_i$, where $n_i = \max\{j: t_{ij} \le U_i \wedge \tau\}$ is the number of observed transitions. The cost y_{ij} is observed provided $s_{ij} = 1$ where $s_{ij} = [t_{ij} \le U_i \wedge \tau]$, that is, if the transition time t_{ij} occurs by time τ and is not censored by time U_i. We denote by $\mathbf{y}_i$ the $n_i \times 1$ vector of costs and by $\mathbf{X}_i$ a $n_i \times p$ matrix of covariates. The matrix $\mathbf{X}_i$ includes variables that are constant in time, but vary at the patient level, variables that vary with time but are constant across individuals (e.g., prices of resources), and variables that vary both between and within patients. For example, the *j*th row of $\mathbf{X}_i$ will typically contain time-constant factors such as age at entry, gender, baseline comorbidity, and variables for modeling time such as t_{ij}, t_{ij}^2 and interactions between time and time-constant variables. We would also include dummies for the transition types $h \rightarrow j$.

Consider a random effects (RE) model[34] given by

$$\mathbf{y}_i = \mathbf{X}_i \beta + a_i \mathbf{1}_i + \mathbf{u}_i \qquad (1.12)$$

where β is an unknown $p\times 1$ parameter, $\mathbf{1}_i$ the $n_i \times 1$ vector with all elements equal to 1, a_i an unobserved patient-specific heterogeneity and $\mathbf{u}_i$ is the $n_i \times 1$ vector of idiosyncratic errors. The composite error in is $\mathbf{v}_i = a_i \mathbf{1}_i + \mathbf{u}_i$. where β is an unknown $p\times 1$ parameter, $\mathbf{1}_i$ the $n_i \times 1$ vector with all elements equal to1, a_i an unobserved patient-specific heterogeneity and $\mathbf{u}_i$ is the $n_i \times 1$ vector of idiosyncratic errors. The composite error in (1.12) is $\mathbf{v}_i = a_i \mathbf{1}_i + \mathbf{u}_i$. Assume $\mathbf{\Omega}_i = E(\mathbf{v}_i \mathbf{v}_i')$ is positive definite. Note that the parameter β in (1.12) is unrelated to the regression parameter in the Cox model for the transition intensities. In this section our notation conforms to standard usage in RE models.

RE model Assumptions:

RE1 (i) $E(\mathbf{u}_i \mid \mathbf{X}_i, a_i) = \mathbf{0}$ and (ii) $E(a_i \mid \mathbf{X}_i) = 0$.

RE2 *rank* $E(\mathbf{X}_i' \mathbf{\Omega}_i^{-1} \mathbf{X}_i) = p$.

RE3 $E(\mathbf{u}_i \mathbf{u}_i' \mid \mathbf{X}_i, a_i) = \sigma_u^2 \mathbf{I}_i$ and $E(a_i^2 \mid \mathbf{X}_i) = \sigma_a^2$,

where σ_u^2 and σ_a^2 are constants and $\mathbf{I}_i$ is the $n_i \times n_i$ identity matrix.

Under **RE1** and **RE3**, $E(\mathbf{v}_i) = \mathbf{0}$ and $E(\mathbf{v}_i \mathbf{v}_i') = \mathbf{\Omega}_i = \sigma_u^2 \mathbf{I}_i + \sigma_a^2 \mathbf{J}_i$ where $\mathbf{J}_i$ is the $n_i \times n_i$ matrix with all elements equal to 1. Since $E(\mathbf{y}_i \mid \mathbf{X}_i) = \mathbf{X}_i \beta$ we see that estimates of $c_{hj}(t \mid Z_0)$ for specified transitions (h, j) and Z_0 are derivable from the estimates of β and specification of the **X**-matrix.

In the RE model an estimate of β is obtained by minimizing (with respect to β) the objective function $n^{-1}\sum_{i=1}^{n} q(\mathbf{y}_i, \mathbf{X}_i)$ where $q(\mathbf{y}_i, \mathbf{X}_i) = \frac{1}{2}(\mathbf{y}_i - \mathbf{X}_i\beta)'\boldsymbol{\Omega}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\beta)$.

Condition **RE1** suffices to ensure $E(\mathbf{X}_i'\boldsymbol{\Omega}_i^{-1}\mathbf{v}_i) = 0$. The feasible generalized least squares (GLS) estimator of β is

$$\hat{\beta}_{RE} = (\sum_{i=1}^{n} \mathbf{X}_i'\hat{\boldsymbol{\Omega}}_i^{-1}\mathbf{X}_i)^{-1}(\sum_{i=1}^{n} \mathbf{X}_i'\hat{\boldsymbol{\Omega}}_i^{-1}\mathbf{y}_i) \tag{1.13}$$

where $\hat{\boldsymbol{\Omega}}_i = \hat{\sigma}_u^2 \mathbf{I}_i + \hat{\sigma}_a^2 \mathbf{J}_i$ is a consistent estimator of $\boldsymbol{\Omega}_i$ derived from suitable consistent estimators $\hat{\sigma}_u^2, \hat{\sigma}_a^2$ of σ_u^2, σ_a^2 respectively. However, (1.13) is implicitly conditional upon the availability of the sample $\{(\mathbf{y}_i, \mathbf{X}_i): 1 \le i \le n\}$. The cost y_{ij} is observed conditional on $t_{ij} \le \tau$. The number of transitions n_i is also random and depends on the length of the observational period and censoring time. Our notation implicitly assumes conditioning on the n_i.

1.4.1 Modification for censoring

As noted earlier, time censoring might lead to incomplete observation of transition costs. Since y_{ij} and a portion of the jth row of $\mathbf{X}_i$ are observed only if $s_{ij} = 1$, the estimator in (1.13) needs to be modified to account for this selection. Let $\mathbf{s}_i$ denote the diagonal matrix with jth entry s_{ij} and $\mathbf{t}_i = (t_{i1}, ..., t_{in_i})'$. A covariate vector $\mathbf{z}_i$ is observed initially which might include some covariates contained in $\mathbf{X}_i$ that do not depend on the

transition times. Consider the observable data $\{(\mathbf{s}_i, \mathbf{z}_i)\colon 1 \le i \le n\}$ and assume that given $\mathbf{z}_i$, the censoring time U_i is independent of $(\mathbf{t}_i, \mathbf{y}_i, \mathbf{X}_i)$, that is, censoring is independent of the transition times and costs. This is a natural assumption with administrative censoring, in which case the independence is related to the distribution of entry times of patients in the study.

Since $E(s_{ij} \mid \mathbf{z}_i, \mathbf{y}_i, \mathbf{X}_i) = P[U_i \ge t_{ij}, t_{ij} \le \tau \mid \mathbf{z}_i, \mathbf{y}_i, \mathbf{X}_i] = P[U_i \ge t_{ij} \mid \mathbf{z}_i]$ provided $t_{ij} \le \tau$, we define weights $w_{ij} = s_{ij} / p(t_{ij}, \mathbf{z}_i)$ where $p(t, \mathbf{z}_i) = P[U_i \ge t \mid \mathbf{z}_i]$. If G is the survival distribution of the censoring time U_i, then $p(t_{ij}, \mathbf{z}_i) = G(t_{ij} - \mid \mathbf{z}_i)$ on $t_{ij} \le \tau$ and therefore under the mild assumption that $G(\tau \mid \mathbf{z}_i) > 0$ with probability 1, we ensure $p(t_{ij}, \mathbf{z}_i) > 0$ whenever $s_{ij} = 1$. Hence $E(w_{ij} \mid \mathbf{z}_i, \mathbf{y}_i, \mathbf{X}_i) = E(w_{ij} \mid \mathbf{z}_i) = 1$ if $t_{ij} \le \tau$ and 0 otherwise.

Let $\mathbf{w}_i$ be the diagonal matrix with jth entry w_{ij}. Then the applicable modification of the previously mentioned objective function $q(\mathbf{y}_i, \mathbf{X}_i)$ is

$$\tilde{q}(\mathbf{y}_i, \mathbf{w}_i, \mathbf{X}_i) = \tfrac{1}{2}\{\mathbf{w}_i^{1/2}(\mathbf{L}_i')^{-1}(\mathbf{y}_i - \mathbf{X}_i\beta)\}'\{\mathbf{w}_i^{1/2}\mathbf{L}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\beta)\}. \quad (1.14)$$

Here $\mathbf{L}_i$ denotes the unique lower triangular matrix with positive diagonal elements such that $\mathbf{\Omega}_i = \mathbf{L}_i\mathbf{L}_i'$. This exists since $\mathbf{\Omega}_i$ is positive definite. Our assumptions ensure the diagonal matrix $E(\mathbf{w}_i \mid \mathbf{z}_i, \mathbf{y}_i, \mathbf{X}_i) = \mathbf{I}_i$ provided $t_{in_i} \le \tau$. Then

$$E[\tilde{q}(\mathbf{y}_i, \mathbf{w}_i, \mathbf{X}_i)] = E[E[\tilde{q}(\mathbf{y}_i, \mathbf{w}_i, \mathbf{X}_i) \mid \mathbf{y}_i, \mathbf{z}_i, \mathbf{X}_i]] =$$

$$= E(\tfrac{1}{2}\{(\mathbf{L}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\beta)\}'E(\mathbf{w}_i \mid \mathbf{z}_i, \mathbf{y}_i, \mathbf{X}_i)\mathbf{L}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\beta)) = E[q(\mathbf{y}_i, \mathbf{X}_i)]. \quad (1.15)$$

Therefore a modification of (1.13) is obtained by minimizing with respect to β the objective function $n^{-1}\sum_{i=1}^{n}\tilde{q}(\mathbf{y}_i,\mathbf{w}_i,\mathbf{X}_i)$. Let $\tilde{\mathbf{X}}_i=\mathbf{w}_i^{1/2}\mathbf{L}_i^{-1}\mathbf{X}_i$ and $\tilde{\mathbf{y}}_i=\mathbf{w}_i^{1/2}\mathbf{L}_i^{-1}\mathbf{y}_i$. Our estimator of β is given by

$$\hat{\beta}_w=(\sum_{i=1}^{n}\tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i)^{-1}(\sum_{i=1}^{n}\tilde{\mathbf{X}}_i'\tilde{\mathbf{y}}_i). \tag{1.16}$$

Remark 1.4.1.1

1. Because $\mathbf{\Omega}_i$ and $p(t,\mathbf{z}_i)$ are generally unknown, to make (1.16) operational we need to replace them by consistent estimators. Assume for now that these are known.

2. Use of the transformation $(\mathbf{y}_i,\mathbf{X}_i)\rightarrow(\tilde{\mathbf{y}}_i,\tilde{\mathbf{X}}_i)$ preserves the time ordering of costs because $\tilde{y}_{ij}$ depends only on $\{\ y_{ik},k\le j\ \}$. This is also true of the rows of $\tilde{\mathbf{X}}_i$ if only covariates ascertained at prior times are involved.

3. Model (1.12) can be generalized to include additional random effects, through a $q\times1$ random vector $\boldsymbol{\alpha}_i$. The model is then $\mathbf{y}_i=\mathbf{X}_i\beta+\mathbf{Z}_i\boldsymbol{\alpha}_i+\mathbf{u}_i$, with $\mathbf{Z}_i$ containing a subset of the covariates in $\mathbf{X}_i$. For example $\mathbf{Z}_i=[\mathbf{1}_i\,|\,\mathbf{t}_i\,|\,\mathbf{t}_i^2]$ might be used to model the time dependence at the individual level. Under an obvious modification of **A3,** the composite error $\mathbf{v}_i=\mathbf{Z}_i\boldsymbol{\alpha}_i+\mathbf{u}_i$ satisfies $E(\mathbf{v}_i\,|\,\mathbf{X}_i)=\mathbf{0}$ and $E(\mathbf{v}_i\mathbf{v}_i'\,|\,\mathbf{X}_i)=\mathbf{\Omega}_i=\sigma_u^2\,\mathbf{I}_i+\mathbf{Z}_i\mathbf{G}\mathbf{Z}_i'$ where $\mathbf{G}=E(\boldsymbol{\alpha}_i\boldsymbol{\alpha}_i'\,|\,\mathbf{X}_i)$. Even if we assume the $q\times q$ matrix $\mathbf{G}$ to be constant, there is conditional heteroscedasticity in $\mathbf{\Omega}_i$. However, the same arguments leading to (1.15) and (1.16) will obtain.

1.4.2 Consistency and asymptotic normality of $\hat{\beta}_w$

Since $E(\tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i) = E[\mathbf{X}_i'\mathbf{L}_i^{-1}E(\mathbf{w}_i \mid \mathbf{X}_i, \mathbf{z}_i)\mathbf{L}_i^{-1}\mathbf{X}_i] = E(\mathbf{X}_i'\boldsymbol{\Omega}_i^{-1}\mathbf{X}_i)$, assumptions **RE1, RE2** ensure that $(n^{-1}\sum_{i=1}^{n}\tilde{\mathbf{X}}_i'\tilde{\mathbf{X}}_i)^{-1}$ converges in probability to $(E(\mathbf{X}_i'\boldsymbol{\Omega}_i^{-1}\mathbf{X}_i))^{-1}$ by the weak law of large numbers (WLLN). Also writing $\tilde{\mathbf{v}}_i = \mathbf{w}_i^{1/2}\mathbf{L}_i^{-1}\mathbf{v}_i$ we get $E(\tilde{\mathbf{X}}_i'\tilde{\mathbf{v}}_i) = E[\mathbf{X}_i'(\mathbf{L}_i')^{-1}E(\mathbf{w}_i \mid \mathbf{y}_i, \mathbf{X}_i, \mathbf{z}_i)\mathbf{L}_i^{-1}\mathbf{v}_i] = E(\mathbf{X}_i'\boldsymbol{\Omega}_i^{-1}\mathbf{v}_i) = 0$ under **RE1.** Hence, by the WLLN we have consistency of $\hat{\beta}_w$. Also from (1.16)

$$n^{1/2}(\hat{\beta}_w - \beta) = (\mathbf{A}_w^{-1} + o_p(1))(n^{-1/2}\sum_{i=1}^{n}\tilde{\mathbf{X}}_i'\tilde{\mathbf{v}}_i) \tag{1.17}$$

where $\mathbf{A}_w = E(\mathbf{X}_i'\boldsymbol{\Omega}_i^{-1}\mathbf{X}_i)$. Application of the central limit theorem gives the normality of $\hat{\beta}_w$,

$$n^{1/2}(\hat{\beta}_w - \beta) \xrightarrow{\mathcal{D}} N(0, \mathbf{A}_w^{-1}\mathbf{B}_w\mathbf{A}_w^{-1}) \tag{1.18}$$

where $\mathbf{B}_w = E[\tilde{\mathbf{X}}_i'\tilde{\mathbf{v}}_i\tilde{\mathbf{v}}_i'\tilde{\mathbf{X}}_i]$. Another form of $\mathbf{B}_w$ is derived under additional assumptions as follows. Now

$$\begin{aligned}\mathbf{B}_w &= E[\tilde{\mathbf{X}}_i'\tilde{\mathbf{v}}_i\tilde{\mathbf{v}}_i'\tilde{\mathbf{X}}_i] = E[\mathbf{X}_i'(\mathbf{L}_i')^{-1}\mathbf{w}_i\mathbf{L}_i^{-1}\mathbf{v}_i\mathbf{v}_i'(\mathbf{L}_i')^{-1}\mathbf{w}_i\mathbf{L}_i^{-1}\mathbf{X}_i] \\ &= E[\mathbf{X}_i'(\mathbf{L}_i')^{-1}\mathbf{w}_i\mathbf{L}_i^{-1}E[\mathbf{v}_i\mathbf{v}_i' \mid \mathcal{F}](\mathbf{L}_i')^{-1}\mathbf{w}_i\mathbf{L}_i^{-1}\mathbf{X}_i],\end{aligned} \tag{1.19}$$

where $\mathcal{F}$ is the set $(U_i, \mathbf{t}_i, \mathbf{X}_i, \mathbf{z}_i)$ and recalling that $\mathbf{w}_i$ is a function of $(U_i, \mathbf{t}_i, \mathbf{z}_i)$. Since U_i is independent of $(\mathbf{t}_i, \mathbf{y}_i, \mathbf{X}_i)$ given $\mathbf{z}_i$ we get $E[\mathbf{v}_i\mathbf{v}_i' \mid \mathcal{F}] = E[\mathbf{v}_i\mathbf{v}_i' \mid \mathbf{X}_i, \mathbf{z}_i]$. Our matrix $\mathbf{X}_i$ includes $t_{ij}, 1 \le j \le n_i$, but $\mathbf{z}_i$ would have variables not in $\mathbf{X}_i$. From **RE1** and **RE3** we get only $E[\mathbf{v}_i \mid \mathbf{X}_i] = 0$ and $E[\mathbf{v}_i\mathbf{v}_i' \mid \mathbf{X}_i] = \boldsymbol{\Omega}_i$. Strengthening these to

$E[\mathbf{v}_i \mid \mathbf{X}_i, \mathbf{z}_i] = 0$, $E[\mathbf{v}_i \mathbf{v}_i' \mid \mathbf{X}_i, \mathbf{z}_i] = \boldsymbol{\Omega}_i$ preserves the previous assumptions and gives $\mathbf{B}_w = E[\mathbf{X}_i'(\mathbf{L}_i')^{-1} \mathbf{w}_i \mathbf{w}_i \mathbf{L}_i^{-1} \mathbf{X}_i]$. However, without these additional assumptions on $\mathbf{v}_i$ we have to be content with $\mathbf{B}_w$ given in (1.19).

1.4.3 Computation of $\mathbf{A}_w$ and $\mathbf{B}_w$

To implement the result (1.18) we will need estimates of $\mathbf{A}_w$ and $\mathbf{B}_w$. From the discussion leading to (1.19) consistent estimators are obtained as

$$\hat{\mathbf{A}}_w = n^{-1} \sum_{i=1}^{n} \tilde{\mathbf{X}}_i' \tilde{\mathbf{X}}_i , \qquad \hat{\mathbf{B}}_w = n^{-1} \sum_{i=1}^{n} \tilde{\mathbf{X}}_i' \hat{\tilde{\mathbf{v}}}_i \hat{\tilde{\mathbf{v}}}_i' \tilde{\mathbf{X}}_i . \tag{1.20}$$

Here $\hat{\tilde{\mathbf{v}}}_i = \tilde{\mathbf{y}}_i - \tilde{\mathbf{X}}_i \hat{\beta}_w$ are the residuals obtained after estimation of β. A consistent estimator of $\boldsymbol{\Omega}_i$ may be obtained by OLS estimation in the model transformed by $\mathbf{w}_i^{1/2}$, that is, the model

$$\mathbf{y}_i^* = \mathbf{X}_i^* \beta + \mathbf{v}_i^*$$

where $\mathbf{y}_i^* = \mathbf{w}_i^{1/2} \mathbf{y}_i$ and the other asterisked symbols are similarly defined. Let $\hat{\beta}_{OLS}$ denote the OLS estimator from this model and $\hat{\mathbf{v}}_i = \mathbf{w}_i^{-1/2}(\mathbf{y}_i^* - \mathbf{X}_i^* \hat{\beta}_{OLS})$ the corresponding transformed residuals. Consistency of $\hat{\beta}_{OLS}$ follows from **RE1** and the mild condition that $E(\mathbf{X}_i' \mathbf{X}_i)$ has full rank. A general consistent estimator of $\boldsymbol{\Omega}_i$ is $\hat{\hat{\boldsymbol{\Omega}}}_i = n^{-1} \sum_{i=1}^{n} \hat{\mathbf{v}}_i \hat{\mathbf{v}}_i'$. However, under **RE3**, $\boldsymbol{\Omega}_i$ has a special structure involving just two

unknown variance components σ_u^2 and σ_c^2. We would then estimate these components from $\hat{\sigma}_u^2 + \hat{\sigma}_a^2 = (\sum_{i=1}^{n} n_i - p)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \hat{v}_{ij}^2$ and $\hat{\sigma}_a^2 = (\sum_{i=1}^{n} \tfrac{1}{2} n_i (n_i - 1) - p)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n_i - 1} \sum_{k>j}^{n_i} \hat{v}_{ij} \hat{v}_{ik}$.

1.4.4 Estimation of $G(\cdot \mid \mathbf{z})$

To completely specify the weights we are left with estimation of $p(t, \mathbf{z})$ which is given in terms of the censoring distribution $G(\cdot \mid \mathbf{z})$. In our observational scheme follow-up of the *i*th patient starting in transient state $X_i(0)$ at time $t_{i0} = 0$ may lead to one of three distinct scenarios. (1) Observation ends at the *j*th transition time $t_{ij} \leq \tau$ in the absorbing state *k*. Then $U_i \geq t_{ij}$ and $X_i(t_{ij}) = k$. (2) Observation continues beyond t_{ij} but ends at τ without the occurrence of another transition or censoring. Then $U_i > \tau$, $t_{ij+1} > \tau$ and $X_i(t_{ij}) \neq k$. (3) Observation continues beyond t_{ij} but ends at U_i before τ or the next transition time. Here $t_{ij} < U_i < t_{ij+1} \wedge \tau$ and $X_i(t_{ij}) \neq k$. The likelihood function involving $G(\cdot \mid \mathbf{z})$ and its density $g(\cdot \mid \mathbf{z})$ can be expressed as

$$\prod_{i=1}^{n} \prod_{j \geq 0} \{G(t_{ij} \mid \mathbf{z}_i)\}^{s_{ij} \delta_{ij}} \{G(\tau \mid \mathbf{z}_i)\}^{(1-\delta_{ij})[U_i \wedge t_{ij+1} > \tau > t_{ij}]} \{g(U_i \mid \mathbf{z}_i)\}^{(1-\delta_{ij})[t_{ij} < U_i < t_{ij+1} \wedge \tau]} \qquad (1.21)$$

where $\delta_{ij} = [X_i(t_{ij}) = k]$. If G is known except for an unknown q dimensional parameter θ then maximizing (1.21) yields an estimator $\hat{\theta}$ of θ giving an estimator $G(\cdot \mid \mathbf{z}, \hat{\theta})$ of

$G(\cdot|\mathbf{z})$. Assuming all regularity conditions for maximum likelihood estimation hold, $\hat{\theta}$ is a solution to $\sum_{i=1}^{n}\sum_{j\geq 0} d_{ij}(\theta)=0$ where

$$d_{ij}(\theta)=s_{ij}\delta_{ij}\frac{\nabla_\theta G_\theta(t_{ij}|\mathbf{z}_i)}{G_\theta(t_{ij}|\mathbf{z}_i)}+(1-\delta_{ij})[U_i\wedge t_{ij+1}>\tau>t_{ij}]\frac{\nabla_\theta G_\theta(\tau|\mathbf{z}_i)}{G_\theta(\tau|\mathbf{z}_i)}$$
$$+(1-\delta_{ij})[t_{ij}<U_i<t_{ij+1}\wedge\tau]\frac{\nabla_\theta g_\theta(U_i|\mathbf{z}_i)}{g_\theta(U_i|\mathbf{z}_i)}.$$

Estimation of $G(\cdot|\mathbf{z})$ changes the weights w_{ij} to $\hat{w}_{ij}=s_{ij}/G(t_{ij}|\mathbf{z}_i,\hat{\theta})$ and would modify the arguments leading to the asymptotic normality result (1.18). In the sequel call θ_0 the true parameter and let $\dot{d}_{ij}(\theta)$ denote the derivative of $d_{ij}(\theta)$ with respect to θ. Using standard arguments we get

$n^{-1}\sum_{i=1}^{n}\sum_{j\geq 0}\dot{d}_{ij}(\theta_0)\rightarrow E(\sum_{j\geq 0}\dot{d}_{ij}(\theta_0))=J(\theta_0)$ in probability. The $q\times q$ matrix $J(\theta_0)$ is assumed to be positive definite. Also $n^{-1/2}\sum_{i=1}^{n}\sum_{j\geq 0} d_{ij}(\theta_0)=O_p(1)$. Then using an expansion of $\sum_{i=1}^{n}\sum_{j\geq 0} d_{ij}(\hat{\theta})$ at θ_0 we obtain

$$n^{1/2}(\hat{\theta}-\theta_0)=-J^{-1}(\theta_0)(n^{-1/2}\sum_{i=1}^{n}\sum_{j\geq 0} d_{ij}(\theta_0))+o_p(1) \qquad (1.22)$$

The same steps leading to (1.18) gives

$$n^{1/2}(\hat{\beta}_w-\beta)=(n^{-1}\sum_{i=1}^{n}\sum_{j}\frac{s_{ij}\mathbf{x}^*_{ij}\mathbf{x}^{*'}_{ij}}{G(t_{ij}|\mathbf{z}_i,\hat{\theta})})^{-1}(n^{-1/2}\sum_{i=1}^{n}\sum_{j}\frac{s_{ij}\mathbf{x}^*_{ij}v^*_{ij}}{G(t_{ij}|\mathbf{z}_i,\hat{\theta})}) \qquad (1.23)$$

where $\mathbf{X}^*_i=\mathbf{L}_i^{-1}\mathbf{X}_i$ and $\mathbf{v}^*_i=\mathbf{L}_i^{-1}\mathbf{v}_i$. Since $G(\tau|\mathbf{z}_i)>0$ with probability 1 and $\hat{\theta}\rightarrow\theta_0$ in probability, by application of the uniform WLLN the first term in (1.23) converges in

probability to $\mathbf{A}_w$. Its estimator is $\hat{\mathbf{A}}_w = n^{-1}\sum_{i=1}^{n}\mathbf{X}_i^{*'}\hat{\mathbf{w}}_i\mathbf{X}_i^{*}$ where $\hat{\mathbf{w}}_i$ is diagonal with elements $\hat{w}_{ij}$. For the second term in (1.23) we have the expansion

$$n^{-1/2}\sum_{i=1}^{n}\sum_{j}\frac{s_{ij}\mathbf{x}_{ij}^{*}v_{ij}^{*}}{G(t_{ij}\,|\,\mathbf{z}_i,\theta_0)}-(n^{-1}\sum_{i=1}^{n}\sum_{j}\frac{s_{ij}\mathbf{x}_{ij}^{*}v_{ij}^{*}}{\{G(t_{ij}\,|\,\mathbf{z}_i,\tilde{\theta})\}^2}\nabla'_{\theta}G(t_{ij}\,|\,\mathbf{z}_i,\tilde{\theta}))n^{1/2}(\hat{\theta}-\theta_0)$$

where $\tilde{\theta}$ is between $\hat{\theta}$ and θ_0. Again, by standard arguments the bracketed expression above converges in probability to $\mathbf{D}(\theta_0)=E(\sum_{j}\frac{s_{ij}\mathbf{x}_{ij}^{*}v_{ij}^{*}}{\{G(t_{ij}\,|\,\mathbf{z}_i,\theta_0)\}^2}\nabla'_{\theta}G(t_{ij}\,|\,\mathbf{z}_i,\theta_0))$ a $p\times q$ matrix. Note that $\mathbf{D}(\theta_0)$ is the same as $E(\sum_{j}\frac{[t_{ij}\le\tau]\mathbf{x}_{ij}^{*}v_{ij}^{*}}{G(t_{ij}\,|\,\mathbf{z}_i,\theta_0)}\nabla'_{\theta}G(t_{ij}\,|\,\mathbf{z}_i,\theta_0))$. Combining these results together with (1.22) into (1.23) we obtain

$$n^{1/2}(\hat{\beta}_w-\beta)=\mathbf{A}_w^{-1}(n^{-1/2}\sum_{i=1}^{n}\sum_{j}\{\frac{s_{ij}\mathbf{x}_{ij}^{*}v_{ij}^{*}}{G(t_{ij}\,|\,\mathbf{z}_i,\theta_0)}-\mathbf{D}(\theta_0)J^{-1}(\theta_0)d_{ij}(\theta_0)\})+o_p(1)$$

$$=\mathbf{A}_w^{-1}(n^{-1/2}\sum_{i=1}^{n}\{\mathbf{k}_i-\mathbf{D}(\theta_0)J^{-1}(\theta_0)\mathbf{d}_i(\theta_0)\mathbf{1}_i\})+o_p(1)$$

where $\mathbf{k}_i=\sum_{j=1}^{n_i}\frac{s_{ij}\mathbf{x}_{ij}^{*}v_{ij}^{*}}{G(t_{ij}\,|\,\mathbf{z}_i,\theta_0)}$, $\mathbf{d}_i(\theta_0)$ the $q\times(n_i+1)$ matrix of the $d_{ij}(\theta_0)$ and $\mathbf{1}_i$ a $(n_i+1)\times 1$ vector with elements all 1. A direct calculation shows that

$$n^{1/2}(\hat{\beta}_w-\beta)\to N(0,\mathbf{A}_w^{-1}\tilde{\mathbf{B}}_w\mathbf{A}_w^{-1})$$

where $\tilde{\mathbf{B}}_w=E(\mathbf{k}_i\mathbf{k}_i')-\mathbf{D}(\theta_0)J^{-1}(\theta_0)\mathbf{D}(\theta_0)$. Compared to (1.18) we notice that $E(\mathbf{k}_i\mathbf{k}_i')$ is the matrix $\mathbf{B}_w$ in (1.19) when G was assumed known. Thus estimation of G here gives

an asymptotic covariance that is no larger than if G was assumed known. This is in the sense that $\mathbf{A}_w^{-1}(\tilde{\mathbf{B}}_w - \mathbf{B}_w)\mathbf{A}_w^{-1}$ is negative definite.

1.5 Asymptotic distribution of the mean transition cost

From our model (1.20) for all transition costs we obtain estimates of $c_{hj}(t \mid Z_0)$ for a covariate profile Z_0 by specifying the appropriate vector X_0 of covariates corresponding to a column positions in **X.** Suppose for example in the model for y_{ij} the row vector $\mathbf{x}'_{ij}$ of **X** contains the fixed covariates $\mathbf{x}_i$, dummies for transitions types, terms of modeling the transition times such as t_{ij}, t_{ij}^2 and perhaps interactions between these times and $\mathbf{x}_i$. Our special X_0 will contain the desired Z_0, interactions between Z_0, t and t^2, indicator variables with value 1 for transition type $h \rightarrow j$, and value 0 for all other transition types. Denoting this covariate profile by $X_{hj0}(t)$ then $c_{hj}(t \mid Z_0) = X'_{hj0}(t)\beta$ and from (1.16) we obtain the estimator

$$\hat{c}_{hj}(t \mid Z_0) = X'_{hj0}(t)\hat{\beta}_w. \tag{1.24}$$

Although the consistency of $\hat{c}_{hj}(t \mid Z_0)$ might seem immediate from (1.16) the final form of the computable $\hat{\beta}_w$ involves the estimated $\mathbf{\Omega}_i$ and weights $\mathbf{w}_i$, the latter through the censoring distribution G.

Now recall from (1.1), the expected net present value $NPV_{hj}^{(1)}(t \mid i, Z_0) =$

$E(C_{hj}^{(1)}(\tau)\,|\,X_0 = i, Z_0)$ has the form $\int_0^\tau e^{-rs} c_{hj}(s\,|\,Z_0) P_{ih}(0, s\,|\,Z_0) dA_{hj}(s\,|\,Z_0)$. From (1.24) and Section 1.3 our estimator is

$$\hat{NPV}_{hj}^{(1)}(t\,|\,i, Z_0) = \int_0^\tau e^{-rs} \hat{c}_{hj}(s\,|\,Z_0) \hat{P}_{ih}(0, s-\,|\,Z_0) d\hat{A}_{hj}(s\,|\,Z_0) \qquad (1.25)$$

which also leads to the estimate of the first term for the expression for NPV in (1.3).

Remark 1.5.1

1. Our model for *all* transition costs $\mathbf{y}_i$ uses the entire sample of patients to estimate simultaneously all $c_{hj}(\cdot\,|\,Z_0), h \neq j$. This has the advantage of drawing strength from other parts of the data set when patients do not have observed transitions of every type. Its limitation lies in discerning a simple but adequate specification in the model for the times t_{ij}. Based on our previous experience we suggested using terms t_{ij}, t_{ij}^2 and their interactions with other covariates in the rows of $\mathbf{X}_i$. Although our estimation strategy does not depend on this specification, as a practical matter a small number of terms involving t_{ij} is desirable and should be adequate in capturing the dynamics of time. Also note that the first term in the NPV formula is actually a sum over the transition times because $\hat{A}_{hj}(\cdot\,|\,Z_0)$ can jump only at these times. Therefore, unless some smoothing of $\hat{A}_{hj}(\cdot\,|\,Z_0)$ is done we will need estimates of the $c_{hj}(\cdot\,|\,Z_0)$ only at the transition times.

2. Since inadequate follow up in some patients might result in incomplete cost observation, we used inverse probability of censoring weights (IPCW) to obtain consistent estimates of the regression parameter β. Many investigators[4, 31] have applied

this technique in a standard survival model with a single cost determination for the whole follow up period.

3. The fixed covariate profile $X_{hj0}(t)$ is assumed to be continuous in t. The asymptotic properties of the estimator $\hat{\beta}$ in (1.18) imply that

$$\sup_{t\in[0,\tau]} |e^{-rt}(\hat{c}_{hj}(t\,|\,Z_0) - c_{hj}(t\,|\,Z_0))| \xrightarrow{P} 0.$$

Also $c_{hj}(.\,|\,Z_0)$ is assumed to be bounded. We have uniform consistency of $\hat{P}$, i.e. $\sup_{t\in[0,\tau]} |\hat{P}_{ih}(0,t\,|\,Z_0) - P_{ih}(0,t\,|\,Z_0)| \xrightarrow{P} 0$. Based on these results and Theorems 1 and 2, the net present value estimator is uniformly consistent, that is

$$\sup_{t\in[0,\tau]} |\hat{NPV}_{hj}^{(1)}(t\,|\,i,Z_0) - NPV_{hj}^{(1)}(t\,|\,i,Z_0)| \xrightarrow{P} 0.$$

This result is proved in Polverejan (2001)[48].

Conditional on the initial state i, given the vector Z_0 of basic covariates, we will assess the asymptotic normality of the net present value of all expenditures associated with the h to j transitions in $(0,t]$, i.e. the asymptotic distribution of $n^{1/2}\left(\hat{NPV}_{hj}^{(1)}(t\,|\,i,Z_0) - NPV_{hj}^{(1)}(t\,|\,i,Z_0)\right)$ using the Functional Delta Method.

Consider the trivariate process $\mathbf{Z}^{(n)}(t) = (Z_1^{(n)}(t), Z_2^{(n)}(t), Z_3^{(n)}(t))$, with components

$$Z_1^{(n)}(t) = \sqrt{n}e^{-rt}(\hat{c}_{hj}(t\,|\,Z_0) - c_{hj}(t\,|\,Z_0)),$$

$$Z_2^{(n)}(t) = \sqrt{n}(\hat{P}_{ih}(s,t\,|\,Z_0) - P_{ih}(s,t\,|\,Z_0)),$$

$$Z_3^{(n)}(t) = \sqrt{n}(\hat{A}_{hj}(t\,|\,Z_0) - A_{hj}(t\,|\,Z_0).$$

We recall our convergence results from the previous sections.

From (1.17), (1.18) and (1.24) we have

$Z_1^{(n)}(t) \overset{a.e.}{=} X'_{hj0}(t)e^{-\pi}\mathbf{A}_w^{-1}(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{X}'_i(\mathbf{L}'_i)^{-1}\mathbf{w}_i\mathbf{L}_i^{-1}\mathbf{v}_i)$ and the weak convergence result

$Z_1^{(n)}(t) \xrightarrow{\mathcal{D}} e^{-\pi}c_{hj0}(t \mid Z_0)$ where $c_{hj0}(t \mid Z_0) = X'_{hj0}(t)\xi_1$ and ξ_1 is a multivariate normal with mean zero and covariance matrix $\mathbf{A}_w^{-1}\mathbf{B}_w\mathbf{A}_w^{-1}$.

Using (1.7) and Theorem 3 we obtain the weakly convergences

$$Z_2^{(n)}(t) = \sqrt{n}(\hat{P}_{ih}(s,t \mid Z_0) - P_{ih}(s,t \mid Z_0)) \overset{a.e.}{=}$$

$$= \sum_{g=1}^{k}\sum_{l\neq g}^{k}\int_s^t P_{ig}(s,u- \mid Z_0)d(S_{n,gl}^{(1)}(u) + S_{n,gl}^{(2)}(u))(P_{lh}(u,t \mid Z_0) - P_{gh}(u,t \mid Z_0))$$

$$\xrightarrow{\mathcal{D}} U_{1ih}(s,t \mid Z_0) + U_{2ih}(s,t \mid Z_0).$$

where

$$U_{1ih}(s,t \mid Z_0) = \sum_{g=1}^{k}\sum_{l\neq g}^{k}\int_s^t P_{ig}(s,u- \mid Z_0)(P_{lh}(u,t \mid Z_0) - P_{gh}(u,t \mid Z_0))dU_{1gl}^{(*)}(u),$$

$$U_{2ih}(s,t \mid Z_0) = \sum_{g=1}^{k}\sum_{l\neq g}^{k}\int_s^t P_{ig}(s,u- \mid Z_0)(P_{lh}(u,t \mid Z_0) - P_{gh}(u,t \mid Z_0))dU_{2gl}^{(*)}(u).$$

and $U_{1ih}(s,t \mid Z_0)$, $U_{2ih}(s,t \mid Z_0)$ are independent; and also

$$Z_3^{(n)}(t) = \sqrt{n}(\hat{A}_{hj}(t \mid Z_0) - A_{hj}(t \mid Z_0) \overset{a.e.}{=} \exp(\beta' Z_{hj0})\{S_{n,hj}^{(1)}(t,\tau) + S_{n,hj}^{(2)}(t)\}$$

$$\xrightarrow{\mathcal{D}} \exp(\beta' Z_{hj0})(U_{1hj}^{*}(t) + U_{2hj}^{*}(t)),\ h \neq j$$

where

$$S_{n,hj}^{(1)}(t,\tau) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\sum_{\substack{h,j=1\\ h\neq j}}^{k}\int_0^\tau \frac{1}{\sqrt{n}}(\mathbf{z}_{hji}(u) - e_{hj}(u,\beta))dM_{hji}(u))'\Sigma(\tau,\beta)'^{-1}\mathbf{b}_{hj}(t),$$

$$S^{(2)}_{n,hj}(t)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^t \frac{dM_{hji}(u)}{s^{(0)}_{hj}(u,\beta)},$$

and the (h,j)-th element of the limit process $\mathbf{U}_1^*(t)$ of $\mathbf{S}_n^{(1)}(t,\tau)$ can be expressed as $\xi_0'\mathbf{b}_{hj}(t)$, where ξ_0 is a p-dimensional normal random variable (its distribution does not depend on t) with zero mean and covariance matrix $\Sigma(\tau,\beta)^{-1}$,

$\mathbf{b}_{hj}(t)=\int_0^t (Z_{hj0}-e_{hj}(u,\beta))\alpha_{hj0}(u)du$ is a p-dimensional vector depending on t; also $U^*_{1hj}(t)$ and $U^*_{2hj}(t)$ are independent.

Theorem 4

Under assumptions **A1** - **A11** and **RE1-RE3** for a fixed time t and $h \neq j$:

$$\sqrt{n}\begin{pmatrix} \hat{A}_{hj}(t\,|\,Z_0)-A_{hj}(t\,|\,Z_0) \\ \hat{P}_{ih}(0,t\,|\,Z_0)-P_{ih}(0,t\,|\,Z_0) \\ \hat{c}_{hj}(t\,|\,Z_0)-c_{hj}(t\,|\,Z_0)\end{pmatrix} \xrightarrow{\mathcal{D}} \begin{pmatrix} U^*_{hj}(t\,|\,Z_0) \\ U_{ih}(t\,|\,Z_0) \\ c_{hj0}(t\,|\,Z_0)\end{pmatrix}$$

where the elements of the covariance matrix Σ are

$$\Sigma(1,1)=\exp(2\beta'Z_{hj0})\{\omega^2_{hj}(t)+\mathbf{b}'_{hj}(t)\Sigma(\tau,\beta)^{-1}\mathbf{b}_{hj}(t)\},$$

$$\Sigma(2,2)=\sum_{g=1}^{k}\sum_{l\neq g}^{k}\int_s^t P^2_{ig}(s,u\,|\,Z_0)(P_{lh}(u,t\,|\,Z_0)-P_{gh}(u,t\,|\,Z_0))^2\cdot$$

$$\cdot\exp(\beta'Z_{gl0})d(\omega^2_{gl}(t)+\mathbf{b}'_{gl}(u)\Sigma(\tau,\beta)^{-1}\mathbf{b}_{gl}(u)),$$

$$\Sigma(3,3)=e^{-2\pi}X'_{hj0}(t)\mathbf{A}_w^{-1}\mathbf{B}_w\mathbf{A}_w^{-1}X_{hj0}(t),$$

$$\Sigma(1,2)=\exp(\beta'Z_{hj0})\mathbf{b}'_{hj}(t)\Sigma(\tau,\beta)^{-1}\mathbf{F}_{ih}(t,\beta)+$$

$$+\exp(2\beta'Z_{hj0})\int_0^t P_{ih}(0,u\,|\,Z_0)(P_{jh}(u,t\,|\,Z_0)-P_{hh}(u,t\,|\,Z_0))d\omega^2_{hj}(u))$$

where

$$F_{ih}(t)=\sum_{g=1}^{k}\sum_{l\neq g}^{k}\int_0^t P_{ig}(0,u\mid Z_0)(P_{lh}(u,t\mid Z_0)-P_{gh}(u,t\mid Z_0))\exp(\beta' Z_{gl0})d\mathbf{b}_{gl}(u)),$$

$\Sigma(1,3)=\Sigma(2,3)=0$.□

Proof:

The expressions above for the diagonal elements of Σ have been previously proved in (1.8), Section 1.3 and (1.18). We prove that $\Sigma(1,3)=\Sigma(2,3)=0$, which verifies that

$$Z_1^{(n)}(t) \text{ and } (Z_2^{(n)}(t),Z_3^{(n)}(t)) \text{ are asymptotically independent.}$$

Let us remember that

$$\Sigma(1,3)=Cov[X'_{hj0}(t)\mathbf{A}_w^{-1}\mathbf{X}_i'(\mathbf{L}_i')^{-1}\mathbf{w}_i\mathbf{L}_i^{-1}\mathbf{v}_i, F(t,\tau,M_i,z_i)]$$

where

$$F(t,\tau,M_i,z_i)=$$
$$\{\sum_{\substack{h,j=1\\h\neq j}}^{k}\int_0^\tau\frac{1}{\sqrt{n}}(z_{hji}(u)-e_{hj}(u,\beta))dM_{hji}(u)\}'\Sigma(\tau,\beta)'^{-1}\mathbf{b}_{hj}(t)+\int_0^t\frac{dM_{hji}(u)}{s_{hj}^{(0)}(u,\beta)}.$$

The expression $\mathbf{X}_i'(\mathbf{L}_i')^{-1}\mathbf{w}_i\mathbf{L}_i^{-1}\mathbf{v}_i$ is a function of all transition times $\mathbf{t}_i=(t_{i1},t_{i2},\ldots t_{in_i})$ observed in [0, τ] as well of $(U_i,\mathbf{z}_i,\mathbf{y}_i,\mathbf{X}_i)$. We will impose the assumption $E(\mathbf{v}_i\mid\mathbf{X}_i,\mathbf{z}_i)=0$. The function $F(t,\tau,M_i,z_i)$ depends on $M_{hj}=\sum_i M_{hji}$ where

$$M_{hji}(t)=N_{hji}(t)-\int_0^t Y_{hi}(u)\exp(\beta'\mathbf{z}_{hji}(u))dA_{hj0}(u)$$

depends on a subset of each of $\mathbf{t}_i, \mathbf{z}_i, \mathbf{X}_i$ and U_i. Since $\mathbf{w}_i$ is a function of $(U_i, \mathbf{t}_i, \mathbf{z}_i)$ and U_i is independent of $(\mathbf{t}_i, \mathbf{y}_i, \mathbf{X}_i)$ given $\mathbf{z}_i$ we get $E[\mathbf{v}_i \mid \mathcal{F}] = E[\mathbf{v}_i \mid \mathbf{X}_i, \mathbf{z}_i]$ where $\mathcal{F}$ is the conditioning set $(U_i, \mathbf{t}_i, \mathbf{X}_i, \mathbf{z}_i)$. Then

$$E(\tilde{\mathbf{X}}_i' \tilde{\mathbf{v}}_i M_{hji}(t) \mid \mathcal{F}) = \mathbf{X}_i'(\mathbf{L}_i')^{-1} \mathbf{w}_i \mathbf{L}_i^{-1} E(\mathbf{v}_i \mid \mathbf{X}_i, \mathbf{z}_i) M_{hji}(t)$$

and so under the assumption $E(\mathbf{v}_i \mid \mathbf{X}_i, \mathbf{z}_i) = 0$ we get $\Sigma(1,3) = 0$. Similar considerations lead to $\Sigma(2,3) = 0$.

The covariance of $\exp(\beta' Z_{hj0})(U_{1hj}^*(t) + U_{2hj}^*(t))$ and $U_{1ih}(0,t \mid Z_0) + U_{2ih}(0,t \mid Z_0)$ can be calculated using results previously stated in Section 1.3.

$$\Sigma(1,2) = E(\exp(\beta' Z_{hj0})(U_{1hj}^*(t) + U_{2hj}^*(t))\ (U_{1ih}(0,t \mid Z_0) + U_{1ih}(0,t \mid Z_0))) =$$

$$= E(\exp(\beta' Z_{hj0}) U_{1hj}^*(t) \sum_{g=1}^{k} \sum_{l \neq g}^{k} \int_0^t P_{ig}(0,u \mid Z_0)(P_{lh}(u,t \mid Z_0) - P_{gh}(u,t \mid Z_0)) \cdot$$

$$\cdot \exp(\beta' Z_{gl0}) dU_{1gl}^{(*)}(u)) +$$

$$+ E(\exp(\beta' Z_{hj0}) U_{2hj}^*(t) \sum_{g=1}^{k} \sum_{l \neq g}^{k} \int_0^t P_{ig}(0,u \mid Z_0)(P_{lh}(u,t \mid Z_0) - P_{gh}(u,t \mid Z_0)) \cdot$$

$$\cdot \exp(\beta' Z_{gl0}) dU_{2gl}^{(*)}(u)) = \mathbf{I} + \mathbf{II}.$$

Using $U_{1hj}^*(t) = \xi_0' \mathbf{b}_{hj}(t)$, the first term of the sum can be calculated as

$$\mathbf{I} = E(\exp(\beta' Z_{hj0}) U_{1hj}^*(t) \sum_{g=1}^{k} \sum_{l \neq g}^{k} \int_0^t P_{ig}(0,u \mid Z_0)(P_{lh}(u,t \mid Z_0) - P_{gh}(u,t \mid Z_0)) \cdot$$

$$\cdot \exp(\beta' Z_{gl0}) dU_{1gl}^{(*)}(u)) =$$

$$= E(\exp(\beta' Z_{hj0})\xi_0' \mathbf{b}_{hj}(t) \sum_{g=1}^{k}\sum_{l\neq g}^{k} \int_0^t P_{ig}(0,u \mid Z_0)(P_{lh}(u,t \mid Z_0) - P_{gh}(u,t \mid Z_0))\cdot$$

$$\cdot \exp(\beta' Z_{gl0}) d\xi_0' \mathbf{b}_{gl}(t).$$

Therefore **I**= $\exp(\beta' Z_{hj0}) E(\xi_0' \mathbf{b}_{hj}(t) \xi_0' \mathbf{F}_{hj}(t,\beta))$ where

$$\mathbf{F}_{ih}(t,\beta) = \sum_{g=1}^{k}\sum_{l\neq g}^{k} \int_0^t P_{ig}(0,u \mid Z_0)(P_{lh}(u,t \mid Z_0) - P_{gh}(u,t \mid Z_0)) \exp(\beta' Z_{gl0}) d\mathbf{b}_{gl}(u)).$$

Hence,

$$\mathbf{I} = \exp(\beta' Z_{hj0}) E(\mathbf{b}_{hj}'(t) \xi_0 \xi_0' \mathbf{F}_{hj}(t,\beta)) = \exp(\beta' Z_{hj0}) \mathbf{b}_{hj}'(t) \Sigma(\tau,\beta)^{-1} \mathbf{F}_{ih}(t,\beta).$$

The second term is:

$$\mathbf{II} = E(\exp(\beta' Z_{hj0}) U_{2hj}^*(t) \sum_{g=1}^{k} \sum_{l\neq g}^{k} \int_0^t P_{ig}(0,u \mid Z_0)(P_{lh}(u,t \mid Z_0) - P_{gh}(u,t \mid Z_0))\cdot$$

$$\cdot \exp(\beta' Z_{gl0}) dU_{2gl}^{(*)}(u)) =$$

$$= E(\exp(\beta' Z_{hj0}) \int_0^t dU_{2hj}^*(u) \int_0^t (\sum_{g=1}^{k} \sum_{l\neq g}^{k} P_{ig}(0,u \mid Z_0)(P_{lh}(u,t \mid Z_0) - P_{gh}(u,t \mid Z_0))\cdot$$

$$\cdot \exp(\beta' Z_{gl0}) dU_{2gl}^{(*)}(u)) =$$

Using $\langle U_{2hj}^*, U_{2mr}^* \rangle = 0$ for $(h,j) \neq (m,r),(h,j),(m,r) \in E^*$ (so the processes

$\{U_{2hj}^*(.), (h,j) \in E^*\}$ are independent) and

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

we calculate

$$\mathbf{II} = E(\exp(\beta' Z_{hj0}) \int_0^t dU_{2hj}^*(u) \int_0^t P_{ih}(0,u \mid Z_0)(P_{jh}(u,t \mid Z_0) - P_{hh}(u,t \mid Z_0))\cdot$$

$$\cdot \exp(\beta' Z_{hj0}) dU_{2hj}^{(*)}(u)) =$$

$$= E(\exp(2\beta' Z_{hj0}) \int_0^t P_{ih}(0,u \mid Z_0)(P_{jh}(u,t \mid Z_0) - P_{hh}(u,t \mid Z_0))d < U_{2hj}^{(*)} > (u))$$

$$= \exp(2\beta' Z_{hj0}) \int_0^t P_{ih}(0,u \mid Z_0)(P_{jh}(u,t \mid Z_0) - P_{hh}(u,t \mid Z_0))d\omega_{hj}^2(u)).$$

Putting **I** and **II** together,

$$\Sigma(1,2) = \exp(\beta' Z_{hj0})\mathbf{b}'_{hj}(t)\Sigma(\tau,\beta)^{-1}\mathbf{F}_{ih}(t,\beta) +$$

$$+ \exp(2\beta' Z_{hj0}) \int_0^t P_{ih}(0,u- \mid Z_0)(P_{jh}(u,t \mid Z_0) - P_{hh}(u,t \mid Z_0))d\omega_{hj}^2(u)).$$

Fix the time t and consider the functional

$$\varphi_t : E^* \to \mathbb{R},\ \varphi_t(z_1,z_2,z_3) = \int_0^t z_1(s)z_2(s)dz_3(s),$$

where E^* is a subset of $D[0,\tau]^3$ such that φ_t is well defined. Notice we can write $NPV_{hj}^{(1)}(t \mid i, Z_0)$ as

$$\varphi_t(z_1,z_2,z_3) = \varphi_t(e^{-r\cdot}c_{hj}(.\mid Z_0), P_{ih}(0,. \mid Z_0), A_{hj}(.\mid Z_0)).$$

If φ_t has an extension to $D[0,\tau]^3$ that is Hadamard differentiable in (z_1,z_2,z_3) then, under some extra-conditions, we can apply the Functional Delta Method.

Lemma 2

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 $\varphi_t(z_1,z_2,z_3) = \int_0^t z_1(s)z_2(s)dz_3(s)$.

Let (z_1, z_2, z_3) be a fixed point of E^* such that $\int_0^\tau |d(z_1 z_2)| < \infty$. Then φ_t can be extended to the space $D[0,\tau]^3$ so as to be Hadamard differentiable at (z_1, z_2, z_3), with derivative

$$d\varphi_t(z_1, z_2, z_3).(Z_1, Z_2, Z_3) = \int_0^t Z_1 z_2 dz_3 + \int_0^t z_1 Z_2 dz_3 + \int_0^t z_1 z_2 dZ_3 , \qquad (1.26)$$

where the integral with respect to Z_3 is defined by the integration by parts formula if Z_3 is not of finite variation.

For technical reasons the following extra-assumption is considered: $c_{hj}(.|Z_0)$ is of finite variation over $[0,\tau]$. We write $\int_0^\tau |dc_{hj}(s|Z_0)| < \infty$. Lemma 2 is proved in Polverejan (2001)[48] and more details on regularity conditions can be found in Gill (1989)[58] and Praestgaard (1991)[26].

Theorem 5

Under the assumptions **A1 – A11** and **RE1 – RE3** for a fixed time *t*:

$$n^{1/2}\left(N\hat{P}V_{hj}^{(1)}(t|i,Z_0) - NPV_{hj}^{(1)}(t|i,Z_0)\right) =$$

$$= n^{1/2}[\varphi_t\left(e^{-r.}\hat{c}_{hj}(.|Z_0), \hat{P}_{ih}(0,.|Z_0), \hat{A}_{hj}(.|Z_0)\right) - \varphi_t\left(e^{-r.}c_{hj}(.|Z_0), P_{ih}(0,.|Z_0), A_{hj}(.|Z_0)\right)]$$

$$\xrightarrow{\mathcal{D}} d\varphi_t\left(e^{-r.}c_{hj}(.|Z_0), P_{ih}(0,.|Z_0), A_{hj}(.|Z_0)\right).\left(e^{-r.}c_{hj0}(.|Z_0), U_{ih}(0,.|Z_0), U_{hj}^*(.|Z_0)\right) =$$

$$= \int_0^t e^{-rs}c_{hj0}(s|Z_0)P_{ih}(0,s|Z_0)dA_{hj}(s|Z_0) + \int_0^t e^{-rs}c_{hj}(s|Z_0)U_{ih}(0,s|Z_0)dA_{hj}(s|Z_0) +$$

$$+ \int_0^t e^{-rs}c_{hj}(s|Z_0)P_{ih}(0,s|Z_0)dU_{hj}^*(s|Z_0) = P(t)$$

where $U^*_{hj}(t) = \exp(\beta' Z_{hj0})(U^*_{1hj}(t) + U^*_{2hj}(t))$, $U_{ih}(t) = U_{1ih}(t) + U_{2ih}(t)$ and

$c_{hj0}(t \mid Z_0) = X'_{hj0}(t)\xi_1$. □

Proof:

The mapping $\varphi_t : E^* \to \mathbb{R}$ is defined by $\varphi_t(z_1, z_2, z_3) = \int_0^t z_1(s) z_2(s) dz_3(s)$,

where $E^* = \{(z_1, z_2, z_3) \in D[0,\tau]^3 : \int_0^\tau |dz_3| \leq C\}$ and $C = A_{hj}(\tau \mid Z_0) + 1 < \infty$. Let

$z_1(s) = e^{-rs} c_{hj}(s \mid Z_0), z_2(s) = P_{ih}(0, s \mid Z_0), z_3(s) = A_{hj}(s \mid Z_0)$. All $z_1, z_2, z_3 \in D[0,\tau]$.

Therefore $(z_1, z_2, z_3) \in E^*$. We have that $\int_0^\tau |d(z_1 z_2)| < \infty$.

The previous Lemma implies that φ_t can be extended to $D[0,\tau]^3$ so as to be

Hadamard differentiable at (z_1, z_2, z_3), with derivative

$$d\varphi_t(z_1, z_2, z_3).(Z_1, Z_2, Z_3) = \int_0^t Z_1 z_2 dz_3 + \int_0^t z_1 Z_2 dz_3 + \int_0^t z_1 z_2 dZ_3 ,$$

where the integral with respect to Z_3 is defined by the integration by parts formula if Z_3

is not of finite variation. Denote by $\varphi_t^{E^*}$ the extension of φ_t to $D[0,\tau]^3$.

Define $\hat{z}_1(s) = e^{-rs}\hat{c}_{hj}(s \mid Z_0), \hat{z}_2(s) = \hat{P}_{ih}(0, s \mid Z_0), \hat{z}_3(s) = \hat{A}_{hj}(s \mid Z_0), \ s \in [0,\tau]$.

We have $(\hat{z}_1, \hat{z}_2, \hat{z}_3) \in D[0,\tau]^3$ for every n and

$$P(\int_0^\tau d|z_3| < C) = P(\int_0^\tau d\,|\hat{A}_{hj0}(s,\hat{\beta})| < C) = P(\hat{A}_{hj}(\tau \mid Z_0) < C) \to 1 \text{ as } n \to \infty$$

because $\hat{A}_{hj}(\tau \mid Z_0) \xrightarrow{P} A_{hj}(\tau \mid Z_0) < C$. Thus $(\hat{z}_1, \hat{z}_2, \hat{z}_3) \in E^*$ with probability tending

to one.

We have

$$n^{1/2}\left(\hat{z}_1(\cdot)-z_1(\cdot)\right)\xrightarrow{\mathcal{D}}Z_1(\cdot)=e^{-r\cdot}c_{hj0}(\cdot\,|\,Z_0),$$

$$n^{1/2}\left(\hat{z}_2(\cdot)-z_2(\cdot)\right)\xrightarrow{\mathcal{D}}Z_2(\cdot)=U_{ih}(0,\cdot\,|\,Z_0),$$

$$n^{1/2}\left(\hat{z}_3(\cdot)-z_3(\cdot)\right)\xrightarrow{\mathcal{D}}Z_3(\cdot)=U^*_{hj}(\cdot\,|\,Z_0).$$

The processes Z_1, Z_2, Z_3 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 (Z_1, Z_2, Z_3) has separable support.

Because $\hat{\mathbf{P}}(0,.\,|\,Z_0)=\prod_{(0.,]}(\mathbf{I}+d\hat{\mathbf{A}}(.\,|\,Z_0))$ and $\mathbf{P}(0,.\,|\,Z_0)=\prod_{(0,.]}(\mathbf{I}+d\mathbf{A}(.\,|\,Z_0))$,

the matrices $\hat{\mathbf{P}}(0,.\,|\,Z_0), \mathbf{P}(0,.\,|\,Z_0)$ are functionals of $\hat{\mathbf{A}}(.\,|\,Z_0), \mathbf{A}(.\,|\,Z_0)$, respectively. Also by Theorem 4,

$$\sqrt{n}\begin{pmatrix}\hat{A}_{hj}(t\,|\,Z_0)-A_{hj}(t\,|\,Z_0)\\ \hat{P}_{ih}(0,t\,|\,Z_0)-P_{ih}(0,t\,|\,Z_0)\\ e^{-rt}(\hat{c}_{hj}(t\,|\,Z_0)-c_{hj}(t\,|\,Z_0))\end{pmatrix}\xrightarrow{\mathcal{D}}\begin{pmatrix}U^*_{hj}(t\,|\,Z_0)\\ U_{ih}(t\,|\,Z_0)\\ e^{-rt}c_{hj0}(t\,|\,Z_0)\end{pmatrix}$$

Therefore by Functional Delta Method,

$n^{1/2}\left(\varphi_t(\hat{z}_1,\hat{z}_2,\hat{z}_3)-\varphi_t(z_1,z_2,z_3)\right)\xrightarrow{\mathcal{D}}d\varphi_t^{E^*}(z_1,z_2,z_3).(Z_1,Z_2,Z_3)=P(t)$ defined in the theorem statement.

Theorem 6 (Variance of $P(t)$)

Under the assumptions in Theorem 5, the variance of

$$P(t)=\int_0^t e^{-rs}c_{hj0}(s\,|\,Z_0)P_{ih}(0,s\,|\,Z_0)dA_{hj}(s\,|\,Z_0)+\int_0^t e^{-rs}c_{hj}(s\,|\,Z_0)U_{ih}(0,s\,|\,Z_0)dA_{hj}(s\,|\,Z_0)+$$

$$+\int_0^t e^{-rs}c_{hj}(s\,|\,Z_0)P_{ih}(0,s\,|\,Z_0)dU^*_{hj}(s\,|\,Z_0)\text{ is}$$

$$Var(P(t)) = T_{1hj}(t)\mathbf{A}_w^{-1}\mathbf{B}_w\mathbf{A}_w^{-1}T'_{1hj}(t) + T_{2hj}(t)'\Sigma(\tau,\beta)^{-1}T_{2hj}(t) +$$

$$+\exp(2\beta' Z_{hj0})(2EB_1 + 2EB_2 + \int_0^\tau \{e^{-rt}c_{hj}(t \mid Z_0)P_{ih}(0,t \mid Z_0)\}^2 \frac{dA_{hj0}(t)}{s_{hj}^{(0)}(\beta,u)})$$

where $T_{1hj}(t) = \int_0^t e^{-rs}X'_{hj0}(s)P_{ih}(0,s \mid Z_0)dA_{hj}(s \mid Z_0)$,

$$T_{2hj}(t) = \exp(\beta' Z_{hj0})\int_0^t e^{-rs}c_{hj}(s \mid Z_0)\{P_{ih}(0,s \mid Z_0)d\mathbf{b}_{hj}(s) + \mathbf{F}_{ih}(s)dA_{hj0}(s \mid Z_0)\},$$

$$EB_1 = \exp(\beta' Z_{hj0})\int_0^\tau e^{-rt}c_{hj}(t \mid Z_0)dA_{hj0}(t)\cdot$$

$$\cdot(\int_0^t e^{-rs}c_{hj}(s \mid Z_0)P_{ih}^2(0,s \mid Z_0)\{P_{jh}(s,t \mid Z_0) - P_{hh}(s,t \mid Z_0)\}\frac{dA_{hj0}(s)}{s_{hj}^{(0)}(\beta,u)})$$

and $EB_2 = \int_0^\tau e^{-rt}c_{hj}(t \mid Z_0)dA_{hj0}(t)\cdot(\int_0^t e^{-rs}c_{hj}(s \mid Z_0)H_{ih}(s,t)dA_{hj0}(s))$.

Proof:

The form of the variance can further be elaborated with the following notation. Let

$$\mathbf{F}_{ih}(t) = \sum_{g=1}^{m}\sum_{l\neq g}\exp(\beta' Z_{gl0})\int_0^t P_{ig}(0,u \mid Z_0)(P_{lh}(u,t \mid Z_0) - P_{gh}(u,t \mid Z_0))d\mathbf{b}_{gl}(u)$$

$$V_{ih}(t) = \sum_{g=1}^{m}\sum_{l\neq g}\exp(\beta' Z_{gl0})\int_0^t P_{ig}(0,u \mid Z_0)(P_{lh}(u,t \mid Z_0) - P_{gh}(u,t \mid Z_0))dU_{2gl}^*(u).$$

While $\mathbf{F}_{ih}(t)$ is a non-random p-dimensional vector, $V_{ih}(t)$ is an example of an Ito process. Here $\mathbf{U}_2^*(.) = \left(U_{2hj}^*(.), h, j \in \{1,...,k\}\right)$ is a $k \times k$ matrix-valued process, where $U_{2hh}^* = -\sum_{j\neq h}U_{2hj}^*$ and $\left\{U_{2hj}^*(.),(h,j)\in\{1,...,k\}\right\}$ is a continuous Gaussian vector

martingale, properties of this process are described in Section 1.3. It follows that V_{ih} is Gaussian and by Fubini's theorem $E(V_{ih}) = 0$.

Let us consider the three terms of $P(t)$:

$$P(t) = \int_0^t e^{-rs} c_{hj0}(s \mid Z_0) P_{ih}(0, s \mid Z_0) dA_{hj}(s \mid Z_0) +$$
$$+ \int_0^t e^{-rs} c_{hj}(s \mid Z_0) U_{ih}(0, s \mid Z_0) dA_{hj}(s \mid Z_0) +$$
$$+ \int_0^t e^{-rs} c_{hj}(s \mid Z_0) P_{ih}(0, s \mid Z_0) dU^*_{hj}(s \mid Z_0) = I + II + III.$$

Since the first term $\int_0^t e^{-rs} c_{hj0}(s \mid Z_0) P_{ih}(0, s \mid Z_0) dA_{hj}(s \mid Z_0)$ is uncorrelated with the other two we calculate its variance separately. Indeed

$$Var(I) = Var(\int_0^t e^{-rs} c_{hj0}(s \mid Z_0) P_{ih}(0, s \mid Z_0) dA_{hj}(s \mid Z_0)) =$$

$$= Var(\xi_1 \int_0^t e^{-rs} X'_{hj0}(s) P_{ih}(0, s \mid Z_0) dA_{hj}(s \mid Z_0))$$

For easiness of notation let $T_{1hj}(t) = \int_0^t e^{-rs} X'_{hj0}(s) P_{ih}(0, s \mid Z_0) dA_{hj}(s \mid Z_0)$.

Then $Var(I) = Var(\xi_1 T_{1hj}(t)) = T_{1hj}(t) Var(\xi_1) T'_{1hj}(t) = T_{1hj}(t) \mathbf{A}_w^{-1} \mathbf{B}_w \mathbf{A}_w^{-1} T'_{1hj}(t)$

The second and third term are correlated and we look at them together

$$Var(II + III) = Var(\int_0^t e^{-rs} c_{hj}(s \mid Z_0) U_{ih}(0, s \mid Z_0) dA_{hj}(s \mid Z_0) +$$
$$+ \int_0^t e^{-rs} c_{hj}(s \mid Z_0) P_{ih}(0, s \mid Z_0) dU^*_{hj}(s \mid Z_0)).$$

Some algebraic calculations lead to

$$Var(II + III) = Var(\int_0^t e^{-rs} c_{hj}(s \mid Z_0)(U_{1ih}(0, s \mid Z_0) + U_{2ih}(0, s \mid Z_0)) dA_{hj}(s \mid Z_0) +$$
$$+ \int_0^t e^{-rs} c_{hj}(s \mid Z_0) P_{ih}(0, s \mid Z_0) \exp(\beta' Z_{hj0}) dU^*_{hj}(s \mid Z_0)) =$$

$$= Var(\exp(\beta' Z_{hj0}) \int_0^t e^{-rs} c_{hj}(s \mid Z_0) \xi_0' \mathbf{F}_{ih}(s) dA_{hj0}(s \mid Z_0) +$$

$$+ \exp(\beta' Z_{hj0}) \int_0^t e^{-rs} c_{hj}(s \mid Z_0) V_{ih}(s) dA_{hj0}(s \mid Z_0) +$$

$$+ \exp(\beta' Z_{hj0}) \int_0^t e^{-rs} c_{hj}(s \mid Z_0) P_{ih}(0, s \mid Z_0) dU_{2hj}^*(s \mid Z_0) +$$

$$+ \exp(\beta' Z_{hj0}) \int_0^t e^{-rs} c_{hj}(s \mid Z_0) P_{ih}(0, s \mid Z_0) \xi_0' d\mathbf{b}_{hj}(s)) =$$

$$= Var(\exp(\beta' Z_{hj0}) \xi_0' \int_0^t e^{-rs} c_{hj}(s \mid Z_0) \{ P_{ih}(0, s \mid Z_0) d\mathbf{b}_{hj}(s) + \mathbf{F}_{ih}(s) dA_{hj0}(s \mid Z_0) \} +$$

$$+ \exp(\beta' Z_{hj0}) \int_0^t e^{-rs} c_{hj}(s \mid Z_0) \{ P_{ih}(0, s \mid Z_0) dU_{2hj}^*(s \mid Z_0) + V_{ih}(s) dA_{hj0}(s \mid Z_0) \}.$$

The last two terms are independent, and therefore

$$Var(II + III) =$$

$$Var(\exp(\beta' Z_{hj0}) \xi_0' \int_0^t e^{-rs} c_{hj}(s \mid Z_0) \{ P_{ih}(0, s \mid Z_0) d\mathbf{b}_{hj}(s) + \mathbf{F}_{ih}(s) dA_{hj0}(s \mid Z_0) \} +$$

$$+ Var(\exp(\beta' Z_{hj0}) \int_0^t e^{-rs} c_{hj}(s \mid Z_0) \{ P_{ih}(0, s \mid Z_0) dU_{2hj}^*(s \mid Z_0) + V_{ih}(s) dA_{hj0}(s \mid Z_0) \}).$$

The first variance does not pose a challenge:

$$Var(\exp(\beta' Z_{hj0}) \xi_0' \int_0^t e^{-rs} c_{hj}(s \mid Z_0) \{ P_{ih}(0, s \mid Z_0) d\mathbf{b}_{hj}(s) + \mathbf{F}_{ih}(s) dA_{hj0}(s \mid Z_0) \} =$$

$$= Var(\xi_0' T_{2hj}(t)) = T_{2hj}(t)' \Sigma(\tau, \beta)^{-1} T_{2hj}(t),$$

where $T_{2hj}(t) = \exp(\beta' Z_{hj0}) \int_0^t e^{-rs} c_{hj}(s \mid Z_0) \{ P_{ih}(0, s \mid Z_0) d\mathbf{b}_{hj}(s) + \mathbf{F}_{ih}(s) dA_{hj0}(s \mid Z_0) \}$.

The second term

$$J(t) = \exp(\beta' Z_{hj0}) \int_0^t e^{-rs} c_{hj}(s \mid Z_0) \{ P_{ih}(0, s \mid Z_0) dU_{2hj}^*(s \mid Z_0) + V_{ih}(s) dA_{hj0}(s \mid Z_0) \}) \quad (1.27)$$

is an Ito process (here the dependence on *i*, *h*, *j* and Z_0 has been suppressed) and it follows that $J(t)$ is Gaussian. The first term in (1.27) is the integral of a deterministic function with respect to a Brownian motion, and hence the mean is zero. The same result applies to the second term, as seen by Fubini's theorem.

Also since $C_{hj}(.)$ are bounded, non-negative processes over $\mathcal{T}$ and $P_{ih}(0,s\,|\,Z_0)$ are probabilities (<1), $P\{\exp(\beta' Z_{hj0})\int_0^\tau(\int_0^t e^{-rs}\,|\,c_{hj}(s\,|\,Z_0)P_{ih}(0,s\,|\,Z_0)ds)^2<\infty\}=1$.

For the second term we need to check that

$$P\{\exp(\beta' Z_{hj0})\int_0^\tau|\int_0^t e^{-rs}c_{hj}(s\,|\,Z_0)V_{ih}(s)dA_{hj0}(s\,|\,Z_0)|<\infty\}=1.$$

This follows directly under the assumptions **A7:** $A_{hj0}(\tau)=\int_0^\tau\alpha_{hj0}(t)dt<\infty$ and **A10:** $s_{hj}^{(m)}(.,.)$ are uniformly continuous bounded functions of $(t,\beta)\in[0,\tau]\times\mathcal{B}$ and boundedness of $C_{hj}(.)$. If we apply Ito's formula for the function $f(t,J(t))=J^2(t)$ then $J^2(t)$ has also a stochastic differential form

$$J^2(t)=2\int_0^t J(s)dJ(s)+\int_0^t\{e^{-rs}c_{hj}(s\,|\,Z_0)P_{ih}(0,s\,|\,Z_0)\}^2\frac{dA_{hj0}(s)}{s_{hj}^{(0)}(\beta,s)}.$$

Split $\int_0^t J(s)dJ(s)=$

$$=\int_0^t J(s)\{e^{-rs}c_{hj}(s\,|\,Z_0)P_{ih}(0,s\,|\,Z_0)\}dU_{2hj}^{*}(s)+\int_0^t J(s)\{e^{-rs}c_{hj}(s\,|\,Z_0)V_{ih}(s)\}dA_{hj0}(s),$$

and note that the expectation of the first term is zero. Hence to evaluate $E(J^2(\tau))$ we need examine the second term in the expression above. Inserting $J(s)$ into the second term gives

$$\int_0^\tau J(t)\{e^{-\eta}c_{hj}(t\,|\,Z_0)V_{ih}(t)\}dA_{hj0}(t)=$$

$$\int_0^\tau\left(\int_0^t e^{-\eta s}c_{hj}(s\,|\,Z_0)P_{ih}(0,s\,|\,Z_0)dU^*_{2hj}(s)\right)\cdot e^{-\eta}c_{hj}(t\,|\,Z_0)V_{ih}(t)dA_{hj0}(t)+$$

$$\int_0^\tau\left(\int_0^t e^{-\eta s}c_{hj}(s\,|\,Z_0)V_{ih}(s)dA_{hj0}(s)\right)\cdot e^{-\eta}c_{hj}(t\,|\,Z_0)V_{ih}(t)dA_{hj0}(t)=B_1+B_2.$$

We have for $s<t$,

$$H_{ih}(s,t)=E\{V_{ih}(s)V_{ih}(t)\}=$$

$$=\sum_{g=1}^{m}\sum_{l\neq g}\exp(2\beta'Z_{gl0})\int_0^t P^2_{ig}(0,u\,|\,Z_0)(P_{lh}(u,t\,|\,Z_0)-P_{gh}(u,t\,|\,Z_0))\cdot$$

$$\cdot(P_{lh}(u,s\,|\,Z_0)-P_{gh}(u,s\,|\,Z_0))\frac{dA_{gl0}(u)}{s^{(0)}_{gl}(\beta,u)}$$

and therefore we get

$$EB_2=\int_0^\tau e^{-\eta}c_{hj}(t\,|\,Z_0)dA_{hj0}(t)\cdot\left(\int_0^t e^{-\eta s}c_{hj}(s\,|\,Z_0)H_{ih}(s,t)dA_{hj0}(s)\right).$$

Finally evaluate EB_1 by noting that the processes U^*_{2hj}, U^*_{2gl} are independent for all (g, l) $\neq (h, j)$. We get a single term

$$EB_1=\exp(\beta'Z_{hj0})\int_0^\tau e^{-\eta}c_{hj}(t\,|\,Z_0)dA_{hj0}(t)\cdot$$

$$\cdot\left(\int_0^t e^{-\eta s}c_{hj}(s\,|\,Z_0)P^2_{ih}(0,s\,|\,Z_0)\{P_{jh}(s,t\,|\,Z_0)-P_{hh}(s,t\,|\,Z_0)\}\frac{dA_{hj0}(s)}{s^{(0)}_{hj}(\beta,u)}\right)$$

Therefore

$$Var(II+III)=T_{2hj}(t)'\Sigma(\tau,\beta)^{-1}T_{2hj}(t)+$$

$$+\exp(2\beta'Z_{hj0})\left(2EB_1+2EB_2+\int_0^\tau\{e^{-\eta}c_{hj}(t\,|\,Z_0)P_{ih}(0,t\,|\,Z_0)\}^2\frac{dA_{hj0}(t)}{s^{(0)}_{hj}(\beta,u)}\right)$$

The proof of the theorem is now complete.

CHAPTER 2

ESTIMATING MEAN COST FOR THE TWO-STATE CASE

With the rapid escalation of costs of medical treatment, interest in accurately determining the cost of medical care has increased. Estimates from cost studies are needed to determine the economic burden of disease, to predict the economic consequences of new medical interventions, and for comparative purposes such as cost-effectiveness analyses.

Health care studies are typically designed with a period of recruitment in which patients enter the study, and an additional period of follow up in which health outcomes are recorded. At study termination however, some patients would have not reached their end-point of interest which leads to right censoring of their time-to-event response. The medical cost associated with the follow up period in these patients is also right censored in the sense that the total cost at the time of censoring is less than the cost that would have accrued if their follow up continued until their end-point was reached. Due to this analogy with censoring and event times, it is tempting to use standard techniques from survival analysis for the analysis of medical costs. The survival analysis approach to costs seems also appealing because of its simplicity, its nonparametric nature and its apparent

robustness in the presence of censoring. However, although the approach is apparently free from distributional assumptions, it is not entirely assumption-free. Some of the assumptions that underlie such analyses, for example, the independence of event and censoring times, are not tenable with censored costs[4, 30, 59]. Generally, cost at the time of censoring and cost at the event time are correlated. A simple sample average of the observed costs in the patient sample would underestimate the true expected medical cost for the treatment under study, and using the average in the sub-sample of patients with complete costs would be inefficient.

The current methods for estimation of the population mean cost are both nonparametric and semi-parametric. Key references are Lin *et al* (1997, 2000, 2003)[4,5,30], Bang & Tsiatis (2000)[31], Strawderman (2000)[32], Willan *et al* (2002, 2003)[2, 3], Huang & Louis (1998)[60], Arijas & Haara (1984)[56] and Wooldridge (2003)[33].

2.1 Introduction and Background

Suppose costs can potentially be accrued over a fixed time period $[0, \tau]$ with expenditure terminating at some event time T so that complete cost observation occurs if a patient is followed through time $T^* = \min(T, \tau)$. Suppose $y(t)$ is a right-continuous process that represents the cumulative cost up to time t (including time t) for a typical patient in the population under study. If lifetime cost is of interest then T denotes survival time. Since costs do not accumulate after T, $y(t) = y(T)$ for all $t \geq T$. The cumulative cost $y(\tau)$ at time τ is the principle random variable of interest, so inference will focus

on the mean cost, $\mu = E(y(\tau)) = E(y(T^*))$. With lifetime medical cost, the cumulative cost is $y(T)$ and estimating the average $E(y(T))$ is of interest.

Censoring at time U might preclude observation of T^*. Let $X = \min(T,U)$, $\delta = [T \le U]$, $X^* = \min(T^*,U)$ and $\delta^* = [T^* \le U]$, where $[A]$ denotes the indicator function of the displayed event A. The observed cost $y = y(T^*)$ is uncensored if, and only if $\delta^* = 1$, and the event occurring at T is observed if, and only if $\delta = 1$. We assume $\{(T_i^*, U_i, y_i = y_i(T_i^*)); i = 1,...,n\}$ are independent identically distributed copies of (T^*, U, y), with the observed data given by

$$\{X_i^* = \min(T_i^*, U_i), \delta_i^* = [T_i^* \le U_i], y_i, i = 1,...,n\}.$$

We assume U is independent of $(y(\cdot), T)$. Then $P[\delta^* = 1 \mid y, T] = G(T^*)$, where $G(t) = P[U \ge t]$. For $i = 1,...,n$ the variables corresponding to the ith subject are indexed by the subscript i. Hereafter, unless otherwise specified, the distributions of the event time T and censoring time U are considered continuous.

In estimating $E(y(T^*))$ the available data could be minimal, in the sense that only $\{X_i^*, \delta_i^*, y_i, i = 1,...,n\}$ is recorded. However when patients are followed overtime, we may have costs observed in multiple intervals, that is the cost history is available.

Lin *et al* (1997)[30] proposed two different estimators of $\mu = E(y(\tau))$, for these two cases. Both estimators are proved to be consistent and asymptotically normal under certain conditions. Suppose that the interval $[0,\tau)$ is divided into G intervals

$[0,\tau) = \bigcup_g [a_{g-1}, a_g)$ with $0 = a_0 < a_1 < \ldots < a_G = \tau$. If we use intervals closed at the left, so that the gth interval is $[a_{g-1}, a_g)$, to be consistent with Lin *et al* (1997, 2000) and Willan *et al* (2002, 2003)[2-4, 30] the survivor function is defined as $S(t) = P(T \geq t)$. Keeping with the more classical definition, $S(t) = P(T > t)$, we would need to partition $(0,\tau]$ as $\bigcup_g (a_{g-1}, a_g]$. Become S continuous, we need not be concerned with this distinction. In both methods, either with minimal cost data or using cost histories, the Kaplan-Meier estimator of S is estimated based on $\{(X_i, \delta_i), i = 1,\ldots,n\}$.

Lin's first estimator refers to the case where we observe the cumulative cost at time X_i and at each of the observation points $a_{g-1} \leq X_i$. Let $\Delta y_{ig} = y_i(a_g) - y_i(a_{g-1})$ be the incremental cost for the ith patient in interval $[a_{g-1}, a_g)$. Since the ith patient may be censored during the gth interval we define

$$\tilde{\Delta} y_{ig} = \begin{cases} y_i(a_g) - y_i(a_{g-1}) & \text{if} \quad X_i > a_g \\ y_i(X_i) - y_i(a_{g-1}) & \text{if} \quad a_{g-1} \leq X_i < a_g \end{cases}$$

as the cost incurred over the interval $[a_{g-1}, a_g)$. Observe that $\tilde{\Delta} y_{ig} \neq \Delta y_{ig}$ only if the ith patient is censored during the interval $[a_{g-1}, a_g)$.

Lin's first estimator of $\mu = E(y(\tau))$ is

$$\mu_{L1} = \sum_{g=1}^{G} \hat{S}(a_{g-1}) \hat{E}_{g-1} \tag{2.1}$$

where $\hat{E}_{g-1} = \dfrac{\sum_{i=1}^{n} [X_i \geq a_{g-1}] \tilde{\Delta} y_{ig}}{\sum_{i=1}^{n} [X_i \geq a_{g-1}]}$, $g = 1,\ldots,G$, and $\hat{S}(t)$ is the Kaplan-Meier estimator

of $S(t)$. Equation (2.1) arises from the identity

$$\mu = E(y(\tau)) = E(\sum_{g=1}^{G}(y(a_g) - y(a_{g-1})) = \sum_{g=1}^{G} E(y(a_g) - y(a_{g-1}) \mid T \geq a_{g-1}) S(a_{g-1})$$

and assumes no cost at time 0. If there is cost at time 0 then it should be added to the expression of (2.1).

Lin's second estimator is appropriate when we do not have cost histories, so only the accumulated cost at time X_i is observed. If we suppose $T \leq \tau$ a.s., this estimator is based on the identity

$$\mu = E(y(T^*)) = E(y(T)[T \leq \tau]) = \sum_{g=1}^{G} E(y(T) \mid a_{g-1} \leq T < a_g)(S(a_{g-1}) - S(a_g)),$$

consistent with the definition of S as $S(t) = P(T \geq t)$.

This suggests:

$$\hat{\mu}_{L2} = \sum_{g=1}^{G} \hat{A}_{g-1}(\hat{S}(a_{g-1}) - \hat{S}(a_g)) \tag{2.2}$$

where $\hat{A}_{g-1} = \dfrac{\sum_{i=1}^{n}[a_{g-1} \leq X_i < a_g, \delta_i = 1] y_i(T_i)}{\sum_{i=1}^{n}[a_{g-1} \leq X_i < a_g, \delta_i = 1]}$ averages the observed final costs for all patients observed to die in the gth interval. The observed costs of patients who are censored before τ are not involved in any calculations so they need not be recorded when using this estimator. For G=1, i.e. $0 = a_0 < a_1 = \tau$, and $T \leq \tau$ a.s, (2.2) becomes

$$\hat{\mu}_{L2} = (1 - \hat{S}(\tau)) \frac{\sum_{i=1}^{n} \delta_i y_i(T_i)}{\sum_{i=1}^{n} \delta_i}.$$

Bang and Tsiatis (2000)[31] proposed a basic unbiased weighted estimator for the mean cost that they call the weighted complete-case estimator. It is applicable only when for each patient i we observe the cumulative cost $y_i(T_i)$. They assume all times T_i to be bounded by τ, so $X_i = X_i^*$ and $\delta_i = \delta_i^*$ for all $i = 1,...,n$. The estimated mean cost is

$$\hat{\mu}_{BT} = \sum_{i=1}^{n} w_i y_i(T_i) \tag{2.3}$$

with $w_i = \delta_i / G(T_i)$, where $G(\cdot)$ is the survival distribution function for censoring time $G(t) = P(U \geq t)$. The heuristic argument leading to this estimator is as follows. A patient who is observed to die at time $X_i = T_i$ has a probability $G(T_i)$ of not being censored. Hence we can think of this patient as representing, on average, $G(T_i)^{-1}$ individuals who might have been censored. They propose to estimate $G(T_i)$ using $\hat{G}(T_i)$, where $\hat{G}$ denotes the Kaplan-Meier estimator of the G based on the data $\{(X_i, \overline{\delta_i}) : i = 1,...,n\}$, where $\overline{\delta_i} = 1 - \delta_i$. Formally, if $N^c(u) = \sum_{i=1}^{n} [X_i \leq u, \delta_i = 0]$ then

$$\hat{G}(t) = \prod_{u \leq t} \left(1 - \frac{\Delta N^c(u)}{Y(u)}\right) \tag{2.4}$$

for any $t \leq X_{(n)} = \max\{X_i : i = 1,...,n\}$. In (2.3) w_i is then $\delta_i / \hat{G}(T_i)$.

They show that $\hat{\mu}_{BT}$ is unbiased, consistent and asymptotically normal and the variance can be estimated by:

$$\frac{1}{n^2} \sum_{i=1}^{n} \frac{\delta_i (y_i - \hat{\mu}_{BT})^2}{\hat{G}(T_i)} + \frac{1}{n^2} \int_0^{\tau} \frac{dN^c(u)}{\hat{G}^2(u)} \{\hat{V}(y^2, u) - \hat{V}^2(y, u)\}$$

where $\hat{V}(y,u) = \frac{1}{n\hat{S}(u)} \sum_{i=1}^{n} \frac{\delta_i y_i [T_i \geq u]}{\hat{G}(T_i)}$.

Lin (2000)[4] considers a regression setup $y_i = \mathbf{z}_i \beta + \varepsilon_i$, where $\mathbf{z}_i$ is a $1 \times p$ vector of covariates. He estimates β using a weighted least-squares method and proposes two different estimators, for the cases of minimal cost data and interval cost data. When no covariates are present, his estimator reduces to $(\sum_{i=1}^{n} w_i y_i)/(\sum_{i=1}^{n} w_i)$, where the weights are $w_i = \delta_i^* / G(T_i^*)$ for all $i = 1,\ldots,n$. For minimal cost data these weights are estimated by $\hat{w}_i = \delta_i^* / \hat{G}(T_i^*)$, where $\hat{G}$ is an appropriate estimator of G and the mean estimated cost is

$$\hat{\mu}_{L3} = (\sum_{i=1}^{n} \frac{\delta_i^* y_i(T_i^*)}{\hat{G}(T_i^*)})/(\sum_{i=1}^{n} \frac{\delta_i^*}{\hat{G}(T_i^*)}) \tag{2.5}$$

For the interval cost data consider a model

$$y_{ig} = \mathbf{z}_i \beta_g + \varepsilon_{ig} \tag{2.6}$$

for each of the G intervals, where y_{ig} denotes the cost incurred over the time interval $[a_{g-1}, a_g)$, β_g, $g = 1,\ldots,G$ are $p \times 1$ vectors of unknown regression parameters, and the error terms ε_{ig} are assumed to be independent among different subjects but allowed to be correlated within the same subject. The initial cost is supposed to be null. Two different censoring types might arise:

A. Time censoring: $\delta_i^* = [T_i^* \leq U_i] = 0$ if $T_i^* > U_i$, and

B. Cost censoring: $\delta_{ig}^* = [T_{ig}^* \leq U_i] = 0$ if $T_{ig}^* > U_i$.

where $T_{ig}^{*} = T_i \wedge a_{g+1}$.

By summing both sides of (2.6) over g we obtain $y_i = \mathbf{z}_i\beta + u_i$, where

$$y_i = \sum_{g=1}^{G} y_{ig}\ ,\ \beta = \sum_{g=1}^{G} \beta_g\ ,\ u_i = \sum_{g=1}^{G} u_{ig}\ .$$

Lin's estimator is then

$$\hat{\mu}_{L4} = \sum_{g=1}^{G}\left(\sum_{i=1}^{n} \frac{\delta_{ig}^{*} y_{ig}}{\hat{G}(T_{ig}^{*})}\right) / \left(\sum_{i=1}^{n} \frac{\delta_{ig}^{*}}{\hat{G}(T_{ig}^{*})}\right) \tag{2.7}$$

We compare (2.3) and (2.5). If $T < \tau$ a.s. then $X_i = X_i^{*}$ and $\delta_i = \delta_i^{*}$. Assuming no ties in the data set and $\hat{G}$ derived as in (2.4), we have

$$\frac{1}{n}\sum_{i=1}^{n} w_i = \frac{1}{n}\sum_{i=1}^{n} \delta_i / \hat{G}(X_i) = \frac{1}{n}\sum_{i=1}^{n} \delta_i / \hat{G}(T_i) = 1 - \hat{S}(T_{(n)}) \tag{2.8}$$

where $T_{(n)}$ is the largest uncensored observation. To justify (2.8) we define $\hat{S}(t) = \prod_{u \le t}\left(1 - \frac{\Delta N(u)}{Y(u)}\right)$, where $N(u) = \sum_{i=1}^{n}[X_i \le u, \delta_i = 1]$, $Y(t) = \sum_{i=1}^{n}[X_i \ge t]$ from where we deduce $\Delta\hat{S}(t) = -\hat{S}(t-)\frac{\Delta N(t)}{Y(t)}$. Also if no ties among failure and censoring times, $\hat{S}(t-)\hat{G}(t-) = n^{-1}Y(t)$. If not ties among failure times $\frac{\delta}{n\hat{G}(t-)} = \frac{\Delta N(t)}{n\hat{G}(t-)}$, therefore $\frac{\delta}{n\hat{G}(t-)} = \hat{S}(t-)\frac{\Delta N(t)}{Y(t)} = -\Delta\hat{S}(t)$, and the equality (2.8) follows. If, in addition, $\hat{S}(T_{(n)}) = 0$ then (2.3) and (2.5) coincide.

2.2 Applying general transition model methods to the 2-state model

We will prove that our transition model described in Chapter 1 also captures costs under the simpler two state survival model with a single transition time and sojourn. In this case, as described in Section 2.1, several investigators have developed techniques for regression analysis of medical costs with the focus being on estimation in the presence of time censoring that might result in incomplete costs data on some patients. Our transition model can be therefore viewed as an extension of this methodology to multiple transition times and sojourns. We still use the same inverse-probability of censoring-weighted (IPCW) technique to derive consistent and asymptotically normal estimators of regression parameters and for the net present values.

Let us specialize our multiple transition model methods from Chapter 1 to a two state model with patients starting in state '0' (alive) and followed until they reach a terminal state '1' (death) at time T. The total cost for a patient i can be interpreted as a sojourn cost that ends at time T or τ whichever occurs first or as a transition cost at time T if the patient dies in the interval of time $[0,\tau]$. While in the first case we will be able to estimate $E[y(T^*)]$, in the second case we estimate $E[y(T)[T \le \tau])$. Throughout this chapter we assume that both S and G, the survival distributions for time and censoring time are continuous, unless otherwise specified. Suppose $\mathbf{z}_i$ is a baseline covariate vector, containing time-constant factors such as age at entry, gender, baseline comorbidity. The covariate vector $\mathbf{z}_i$ will be used for the estimation of weights $w_i = s_i / G(T_i - | \mathbf{z}_i)$. We

also introduce the vector $\mathbf{x}_i$ which typically contains time-constant factors, including $\mathbf{z}_i$, variables for modeling time such as T_i, T_i^2 and interactions between time and time-constant variables. Let Z_0 be a fixed profile covariate. Since we only have two states, we will denote the net present value $NPV_{01}^{(1)}(t \mid Z_0)$ as $NPV^{(1)}(t \mid Z_0)$ and $NPV^{(1)}(Z_0) = NPV_{01}^{(1)}(\tau \mid Z_0)$. Throughout the rest of this chapter we assume the discount rate $r = 0$.

2.2.1 Mean cost in the minimal cost data case

NPV for transition costs

Since the only transition is $0 \to 1$, the NPV for transition costs at time t, is

$$NPV^{(1)}(t \mid Z_0) = \int_0^t c_{01}(s \mid Z_0) P_{00}(0, s \mid Z_0) \alpha_{01}(s \mid Z_0) ds, \tag{2.9}$$

where $c_{01}(s \mid Z_0) = E(C_{01}(s) \mid X(s-) = 0, Z_0)$ and $P_{00}(0, t \mid Z_0) = P(T_1 > t \mid Z_0) = S(t \mid Z_0)$.

The quantity $C_{01}(t)$ was described in Chapter 1, and represents the amount incurred just after time t if the patient dies at time t. Note that we use now $S(t) = P(T > t)$ consistent with our notations in Chapter 1. Since $-dS(t \mid Z_0) =$ $= S(t- \mid Z_0) \alpha_{01}(t \mid Z_0)$, (2.9) reduces to

$$NPV^{(1)}(Z_0) = -\int_0^\tau c_{01}(s \mid Z_0) dS(t \mid Z_0) \tag{2.10}$$

In the two-state model a cost $y_i = y(T_i)$ is incurred in the transition from the initial state "0" at time t =0 to state "1" at time T_i, provided that $T_i \le \tau$. The probability of still being in the initial state just prior to time t is $P[T \ge t] = S(t-)$, and the (conditional) probability of transition to state "1" in $(t, t+dt)$ is $dA(t) = \alpha(t)dt$, where $\alpha(t)$ is the hazard function. This differs slightly from the set up described in section 2.1, where costs are incurred through time $T \wedge \tau$ (and not at time T). Naturally, while in Lin's set-up (Section 2.1) the censoring indicator for cost is $\delta_i^* = [T_i^* \le U_i]$, in our set-up y_i is observed provided s_i=1 where $s_i = [T_i \le U_i \wedge \tau]$, that is, if the transition time T_i occurs by time τ and is not censored by time U_i. Assuming U_i is independent of (y_i, T_i) conditional on $\mathbf{z}_i$ we get

$$P(s_i = 1 | y_i, T_i, \mathbf{z}_i) = P(T_i \le U_i \wedge \tau | y_i, T_i, \mathbf{z}_i) = [T_i \le \tau] G(T_i - | \mathbf{z}_i).$$

In a regression set-up, let y_i be the transition costs in the *i*th patient at the transition time T_i. Consider now a regression model

$$y_i = \mathbf{x}_i \beta + \varepsilon_i \tag{2.11}$$

where $\mathbf{x}_i$ is a $1 \times p$ vector of covariates as described at the beginning of the section. Our model (2.11) differs from Lin's model (2000)[4] by the fact that we include in (2.11) time-varying covariates. This is critical as we shall see later. We assume:

$$\mathbf{A1}: \quad E(\varepsilon_i | T_i, \mathbf{z}_i) = 0, E(\varepsilon_i^2 | T_i, \mathbf{z}_i) = \sigma_\varepsilon^2 \tag{2.12}$$

$$\mathbf{A2}: \quad rank\ E(\mathbf{x}_i' \mathbf{x}_i) = p \tag{2.13}$$

Since $E(s_i | \mathbf{z}_i, y_i, \mathbf{x}_i) = P[U_i \ge T_i, T_i \le \tau | \mathbf{z}_i, y_i, \mathbf{x}_i] = P[U_i \ge T_i | \mathbf{z}_i]$ provided $T_i \le \tau$, we define weights $w_i = s_i / G(T_i - | \mathbf{z}_i)$ where G is the survival

distribution of the censoring time U_i. We assume that $G(\tau \mid \mathbf{z}_i) > 0$ with probability 1, to ensure $w_i > 0$ whenever s_i=1. Hence $E(w_i \mid \mathbf{z}_i, y_i, \mathbf{x}_i) = [T_i \le \tau]$. Now $q(y_i, w_i, \mathbf{x}_i) = \frac{1}{2}\sigma_\varepsilon^{-2} w_i (y_i - \mathbf{x}_i\beta)^2$, and minimizing the objective function $\frac{1}{n}\sum_{i=1}^{n} q(y_i, w_i, \mathbf{x}_i)$ with respect to β yields the estimator

$$\hat{\beta}_w = (\sum_{i=1}^{n} w_i \mathbf{x}_i' \mathbf{x}_i)^{-1} (\sum_{i=1}^{n} w_i \mathbf{x}_i' y_i) \tag{2.14}$$

We estimate $c_{01}(t \mid Z_0)$ by $\mathbf{x}_0(t)\hat{\beta}_w$, where $\mathbf{x}_0(t)$ denotes the covariate profile at time t in this model containing t, t^2 and Z_0, and (2.10) by

$$N\hat{P}V^{(1)}(Z_0) = \hat{\beta}_w' \int_0^\tau \mathbf{x}_0(t) d(-\hat{S}(t \mid Z_0)).$$

Remark 2.2.1.1

1. If covariates depending on time are not included in the regression model (2.11), (2.14) yields the same estimator as Lin (2000)[4], except for a slight difference in weights as noted above.

2. If no covariates are present, the regression model (2.11) has only one intercept (and p=1). If all costs are observed before τ, this intercept is estimated by $\bar{y} = n^{-1}\sum_{i=1}^{n} y_i$.

From (2.14) we get $\hat{\beta}_w = (\sum_{i=1}^{n} w_i)^{-1} (\sum_{i=1}^{n} w_i y_i)$ and so

$$\hat{\beta}_w \xrightarrow{P} (1 - S(\tau))^{-1} E(y_i [T_i \le \tau]),$$

$N\hat{P}V^{(1)} \xrightarrow{P} E(y_i[T_i \le \tau])$. Hence if $T \le \tau$ a.s. the natural estimator of $N\hat{P}V^{(1)}$ is simply $\overline{y}$. In practice the costs of some subjects would be censored. Then it is natural to estimate (2.10) by

$$\hat{\mu}_{WT} = -\int_0^{\tau} \hat{c}_{01}(t) d\hat{S}(t) \tag{2.15}$$

where $\hat{S}(t)$ is the Kaplan-Meier estimator of $S(t)$ based on the data $\{(X_i, \delta_i) : i = 1,...,n\}$ and $\hat{c}_{01}(t)$ is an appropriate estimator of $c_{01}(t)$ which we will define later. Note that $d\hat{S}(t) = -\hat{S}(t-)d\hat{A}(t) = -\hat{S}(t-)\frac{dN(t)}{Y(t)}$ and $\hat{A}(t) = \int_0^t \frac{dN(s)}{Y(s)}$ is the usual Nelson-Aalen estimator of the integrated hazard function (Andersen *et al* (1993)[25]). Since costs are realized only at times T_i we estimate $\hat{c}_{01}(t)$ at time $t = T_i$ by $\overline{y}(T_i)$, the average costs observed at T_i. Substitution in (2.15) gives

$$\hat{\mu}_{WT1} = \sum_{i=1}^{n} \delta_i \overline{y}(T_i) \frac{\hat{S}(T_i-)}{Y(T_i)} [T_i \le \tau] \tag{2.16}$$

Some simplification is possible if we assume there are no ties in the data set between event times and censoring times as well as among event times. Let $\hat{G}$ denote the Kaplan-Meier estimator of the G based on the data $\{(X_i, \overline{\delta}_i) : i = 1,...,n\}$, where $\overline{\delta}_i = 1 - \delta_i$ as defined in (2.4). Since the processes N and N^c do not have jumps in common, $\hat{S}(t-)\hat{G}(t-) = n^{-1}Y(t)$. Also if there are no ties among the event times $\overline{y}(T_i) = y_i(T_i)$ the observed cost in the ith subject. Hence (2.16) reduces to

$$\hat{\mu}_{WT2} = n^{-1} \sum_{i=1}^{n} \delta_i \frac{y_i(T_i)}{\hat{G}(T_i-)} [T_i \le \tau] \tag{2.17}$$

Observe that (2.17) can be also written as $\hat{\mu}_{WT2} = n^{-1}\sum_{i=1}^{n}\delta_i \frac{y_i(T_i)}{\hat{G}(T_i)}[T_i \le \tau]$ since does not jumps at those times T_i for which $\delta_i = 1$, so $\frac{\delta_i}{\hat{G}(T_i)} = \frac{\delta_i}{\hat{G}(T_i-)}$. Relation (2.17) produces the same estimator as described by Bang and Tsiatis (2000)[31]. If there are ties among the event times and $t_1^* < \ldots < t_k^* \le \tau$ are the distinct event times observed in $[0, \tau]$, then $\overline{y}(t_j^*) = \overline{y}_j^*$, the mean of the observed costs at time t_j^*. This simplifies further the aforementioned estimator to $n^{-1}\sum_{j:t_j^* \le \tau} \overline{y}_j^* d_j / \hat{G}(t_j^*-)$ where d_j is the multiplicity of t_j^*.

If some event and censoring times are tied in the data set in (0, t), $\hat{S}(t-)\hat{G}(t-)$ will not equal $n^{-1}Y(t)$, see Van der Vaart (1998)[61]. In fact

$$\hat{S}(t-)\hat{G}(t-) = \prod_{u<t}(1-\frac{\Delta N(u)}{Y(u)})\prod_{u<t}(1-\frac{\Delta N^c(u)}{Y(u)}) > \prod_{u<t}(1-\frac{\Delta N^0(u)}{Y(u)}) = n^{-1}Y(t),$$

where $N^0 = N + N^c$ which means that $\hat{\mu}_{WT1} > \hat{\mu}_{WT2}$. Since our concern is with $\hat{S}(t)$ and $\hat{G}(t)$ for $t \in [0, \tau)$ we will show (under some assumptions) that

$\sup_{t<\tau} | n^{-1}Y(t) - \hat{S}(t-)\hat{G}(t-) | \to 0$ in probability.

Proof:

By the Glivenko-Cantelli theorem for iid random variables,

$\sup_{t<\tau} | n^{-1}Y(t) - H(t-) | \to 0$ in probability, provided $H(\tau-) > 0$, where $H(t) = S(t)G(t)$.

Under this same condition the Kaplan–Meier estimators $\hat{S}$ and $\hat{G}$ are uniformly

consistent on $[0,\tau]$. Hence the assertion follows from the inequality

$$|n^{-1}Y(t)-\hat{S}(t-)\hat{G}(t-)|\leq|n^{-1}Y(t)-H(t-)|+|\hat{S}(t-)-S(t-)|+|\hat{G}(t-)-G(t-)|.\square$$

Remark 2.2.1.2

We estimate $c_{01}(s\,|\,Z_0)$ from a weighted regression model using weights $w_i = s_i / G(T_i-|\,\mathbf{z}_i)$. There are several avenues for estimating the survival distribution for censoring time, G, we summarize some ideas in the following table:

Table 2.1 Estimating the survival distribution for censoring time

Model	Comments
1. Nonparametric	Use $\hat{G}$, the Kaplan-Meier estimator based on the data $\{(X_i,\overline{\delta}_i): i=1,...,n\}$, where $\overline{\delta}_i = 1-\delta_i$. Then the weights can be estimated as $w_i = s_i/\hat{G}(T_i)$ or $w_i = s_i\hat{S}(T_i-)/nY(T_i)$.
2. Semi-parametric	Estimate $G(T_i-|\,\mathbf{z}_i)$ from a Cox proportional hazards model
3. Parametric	Assume G has a parametric form $G(t,\theta)=P[U>t\,|\,\theta]$. We assume that the functional form of G is known except for an unknown q-dimensional parameter θ.

In model 1 choosing to estimate $G(t)=P[U>t]$ by $\hat{G}^*(t)$ for all $t\leq\tau$, where $\hat{G}^*$ is the Kaplan-Meier estimator from the data $\{(X_i^*,1-\delta_i^*): i=1,...,n\}$ will not produce changes

in the formula for the weights since $\hat{G}^*(t) = \hat{G}(t)$ for all $t \le \tau$. We will come back to the problem of estimating the weights later on in this chapter.

NPV for sojourn costs

The NPV for sojourn costs in state 0='Alive' is

$$NPV^{(2)}(Z_0) = \int_0^\tau b_0(t \mid Z_0) P_{00}(0, t- \mid Z_0) dt .$$

The quantity $b_0(t \mid Z_0)$ is the expected mean rate of expenditure at time t while sojourning in state 0='Alive'. In practice it will not be observable unless discrete information is available. Instead we will only know the total cost of the sojourn. Since a sojourn ends at a transition time or at τ, the total cost is observed if censoring has not occurred before time $T^* = \min(T, \tau)$. The set-up now is the same as described in Chapter 2.1 in the introduction and background section. The observed cost $y = y(T^*)$ is uncensored if, and only if $\delta^* = [T^* \le U] = 1$, and the event occurring at T is observed if, and only if $\delta = 1$. In this case the weights are $w_i^* = \delta_i^* / G(T_i^* - \mid \mathbf{z}_i)$ and assuming that the censoring time is conditionally independent of survival time and cost we get $E(w_i^* \mid T_i, y_i, \mathbf{z}_i) = 1$.

The NPV of interest is

$$\int_0^\tau b_0(t \mid Z_0) S(t- \mid Z_0) dt = \int_0^\tau S(t- \mid Z_0) dm(t \mid Z_0),$$

where $m(t \mid Z_0) = \int_0^t b_0(u \mid Z_0) du$. By an integration-by-parts we get $NPV(Z_0) = E(m(T^*) \mid Z_0)$. From a standard Cox model we obtain an estimator $\hat{S}(t \mid Z_0)$ of $S(t \mid Z_0)$.

Consider a regression model of the observed costs y_i on time T_i^*, noting that only observations for which $\delta_i^* = 1$ will be used. In addition to the observed (fixed time) covariates $\mathbf{z}_i$ our model would include terms for modeling T_i^*, for example $\mathbf{x}_i = (\mathbf{z}_i, T_i^*, T_i^{*2})$.

We use a model

$$y_i = \mathbf{x}_i \beta + \varepsilon_i \tag{2.18}$$

analogous to (2.11) and the same scheme to obtain the estimator $\hat{\beta}_w = (\sum_{i=1}^{n} w_i^* \mathbf{x}_i' \mathbf{x}_i)^{-1} (\sum_{i=1}^{n} w_i^* \mathbf{x}_i y_i)$ of β as in (2.14). All of this is exactly the same except for the new weights w_i^*. Now let $\mathbf{x}_0(t)$ denote the covariate profile at time t in this model and derive the estimator of $m(t \mid Z_0)$ as $\hat{m}(t \mid Z_0) = \hat{\beta}_w' \int_0^t \dot{\mathbf{x}}_0(u) du$, where the dot denotes differentiation with respect to time. This gives our NPV estimator as $\hat{NPV}(Z_0) = \hat{\beta}_w' \int_0^\tau \hat{S}(t- \mid Z_0) \dot{\mathbf{x}}_0(t) dt$.

Remark 2.2.1.3

Suppose now that we include neither time or covariates in the regression model (2.18). The net present value is $\int_0^\tau b(t) S(t-) dt = \int_0^\tau S(t-) dm(t)$ where S is the survival distribution of T, $b(t)$ is the expected mean rate of expenditure at time t while sojourning in state 0 and $m(t) = \int_0^t b(s) ds$. By an integration-by-parts we get $\int_0^\tau S(t-) dm(t) = \int_0^\tau m(t) dS(t) + S(\tau-) m(\tau)$. Here we have assumed that $m(0) = 0$, if

this is not the case an initial cost should be added to the final computation of NPV. The estimation of $m(t)$, $t<\tau$ and $m(\tau)$ should be done differently in the following cases:

a) patient dies before τ, i.e. $T<\tau$ and

b) patient does not die and is not censored before τ, i.e. $T \geq \tau, U \geq \tau$.

Our goal is to estimate $E(y(T^*)) = E(y(T)[T<\tau]) + E(y(\tau)[T \geq \tau])$. When viewed as a transition model we are only able to estimate the first part. If we look at the problem as a sojourn model we are able to estimate $E(y(T^*))$. Indeed,

$E(y(T^*)) = E(\int_0^{\tau} E(y(T^*)\,|\,T^*=t)(-dS^*(t)))$ since $T^* \leq \tau$, where S^* is the survival distribution for T^*. On the set $t \leq \tau$, note that $\{t \leq T^*\} = \{t \leq T\}$ and $S(t-) = S^*(t-)$. Also $S^*(t) = P[T^* > t] = 0$ for any $t \geq \tau$ and S^* has a discontinuity point at τ.
Let us denote $m^*(t) = E(y(T^*)\,|\,T^*=t)$. Then

$$E(y(T^*)) = \int_0^{\tau} m^*(t)(-dS^*(t)) = \int_0^{\tau-} m^*(t)(-dS(t)) + m^*(\tau)(S^*(\tau-) - S^*(\tau))$$

$$= \int_0^{\tau-} m^*(t)(-dS(t)) + S(\tau-)m^*(\tau)$$

Also $m^*(t) = E(y(T^*)\,|\,T^*=t) = E(y(T)\,|\,T=t) = m(t)$ for all $t<\tau$ and

$m^*(\tau) = E(y(\tau)\,|\,T^*=\tau) = E(y(\tau)\,|\,T \geq \tau)$.

Therefore

$$E(y(T^*)) = \int_0^{\tau-} m(t)(-dS(t)) + S(\tau-)E(y(\tau)\,|\,T \geq \tau) \qquad (2.19)$$

Naturally for the estimation of the first term of the expression we use the weighted mean of all observations with $X_i^* < \tau$ and $\delta_i^* = 1$. For the estimation of the second term we use all observations with $\{X_i^* = \tau, \delta_i^* = 1\} = \{X_i \geq \tau\}$. If there are no

observations such that $X_i^* = \tau, \delta_i^* = 1$ then the estimation of the second term is not possible. Assuming that this is not the case here, we estimate the expression (2.19) as

$$\hat{E}(y(T^*)) = (1-\hat{S}(\tau-))\hat{\beta}_w + \hat{S}(\tau-)\hat{E}(y(\tau)|T \geq \tau) \tag{2.20}$$

where $\hat{S}(t)$ is the Kaplan-Meier estimate of S, $\hat{\beta}_w = \frac{\sum_{\{i:X_i^* < \tau\}} w_i^* y_i}{\sum_{\{i:X_i^* < \tau\}} w_i^*}$ and

$\hat{E}(y(\tau)|T \geq \tau) = \frac{\sum_{\{i:X_i^* = \tau\}} w_i^* y_i}{\sum_{\{i:X_i^* = \tau\}} w_i^*}$, where $w_i^* = \delta_i^* / \hat{G}(T_i^* -)$ and $\hat{G}$ denote the Kaplan-Meier estimator of the G based on the data $\{(X_i, \overline{\delta}_i) : i = 1,...,n\}$.

However assuming no ties among event and censoring times that are strictly smaller than τ, and no ties among event times strictly smaller than τ,

$$\sum_{\{i:X_i^* < \tau\}} w_i^* = \sum_{\{i:X_i^* < \tau\}} \frac{\delta_i^*}{\hat{G}(T_i^*)} = \sum_{\{i:X_i^* < \tau\}} n(\hat{S}(T_i^* -) - \hat{S}(T_i^*)) = n(1-\hat{S}(\tau-)),$$

$$\sum_{\{i:X_i^* = \tau\}} w_i^* = \frac{\Delta N(\tau)}{\hat{G}(\tau)} = \frac{\Delta N(\tau)}{(1-\frac{\Delta N^c(\tau)}{Y(\tau)})\hat{G}(\tau-)} = n\hat{S}(\tau-)$$

and $\sum_i w_i^* = n(1-\hat{S}(\tau-)) + n\hat{S}(\tau-) = n$. (2.21)

Expression (2.20) reduces to

$$\hat{E}(y(T^*)) = (1-\hat{S}(\tau-))\hat{\beta}_w + \hat{S}(\tau-)\hat{E}(y(\tau)|T \geq \tau) =$$

$$= (1-\hat{S}(\tau-))\frac{\sum_{\{i:X_i^*<\tau\}} w_i^* y_i}{\sum_{\{i:X_i^*<\tau\}} w_i^*} + \hat{S}(\tau-)\frac{\sum_{\{i:X_i^*=\tau\}} w_i^* y_i}{\sum_{\{i:X_i^*=\tau\}} w_i^*} =$$

$$= (1-\hat{S}(\tau-))\frac{\sum_{\{i:X_i^*<\tau\}} w_i^* y_i}{n(1-\hat{S}(\tau-))} + \hat{S}(\tau-)\frac{\sum_{\{i:X_i^*=\tau\}} w_i^* y_i}{n\hat{S}(\tau-)} = \frac{\sum_{\{i\le n\}} w_i^* y_i}{n} = \frac{\sum_{\{i\le n\}} w_i^* y_i}{\sum_{\{i\le n\}} w_i^*}.$$

Therefore if times and covariates are not included in (2.18) we can estimate the net present value by $\frac{\sum_{\{i\le n\}} w_i^* y_i}{\sum_{\{i\le n\}} w_i^*}$ which is exactly (2.5), i.e. Lin's estimator (2000)[4].

2.2.2 Mean cost in the interval cost data case

Suppose that the interval $[0,\tau]$ is divided into G intervals with $0 = a_0 < a_1 < \ldots < a_G = \tau$. Let T, U denote the survival time and censoring time respectively. Let $S(t) = P(T > t)$ be the survival function for T. Patients are followed through time τ, where τ indicates the end of the study, i.e. all observation is ceased at τ.

Consider costs incurred in interval $I_g = [a_{g-1}, a_g)$. If death or censoring preceded a_{g-1} then the cost in the gth interval is either 0 or unknown. We observe a *non-zero* cost in the gth interval for the ith patient in two possible situations:

(1) If the patient is observed throughout $I_g = [a_{g-1}, a_g)$, then

$X_i = \min(T_i, U_i) \ge a_g$. Let y_{ig} be the associated cost *at* time a_g.

(2) If the patient dies in $I_g = [a_{g-1}, a_g)$, then $a_{g-1} \le X_i < a_g$ and $\delta_i = [T_i \le U_i] = 1$. In this case the cost y_{ig} can be regard as a transition cost at time X_i.

We will regard (1) as a sojourn cost and (2) as a transition cost. Let $\delta_{ig}^a = [X_i \ge a_g]$, $\delta_{ig}^d = [a_{g-1} \le X_i < a_g, \delta_i = 1]$ be the censoring indicators for cases (1) and (2) respectively. For the interval I_g, the sojourn cost is observed if and only if $\delta_{ig}^a = 1$. If $\delta_{ig}^a = 0$, but $\delta_{ig}^d = 1$ then the transition cost y_{ig} is observed.

Using similar methods to Section 2.2.1, taking together all sojourn costs y_{ig}, when $\delta_{ig}^a = 1$, the average sojourn cost in the gth interval is estimated by $\hat{b}_0(a_g) = \dfrac{\sum_{i=1}^{n} \delta_{ig}^a y_{ig}}{\sum_{i=1}^{n} \delta_{ig}^a}$.

Therefore the NPV for sojourn costs is

$$\int_{[a_{g-1}, a_g)} \hat{b}_0(t)\hat{S}(t)dt = \hat{b}_0(a_g)\hat{S}(a_g) \tag{2.22}$$

With transition cost in interval I_g, we have $P(\delta_{ig}^d = 1 \mid T_i) = [a_{g-1} \le T_i < a_g]G(T_i)$.

Taking together all transition costs when $\delta_{ig}^d = 1$, the average transition cost in the gth interval is estimated $\hat{c}_{01g} = (\sum_{i=1}^{n} y_{gi} \dfrac{\delta_{ig}^d}{\hat{G}(X_i)}) / (\sum_{i=1}^{n} \dfrac{\delta_{ig}^d}{\hat{G}(X_i)})$. Therefore the NPV for transition costs is

$$\int_{[a_{g-1}, a_g)} \hat{c}_{01}(s)d\hat{S}(s) = \hat{c}_{01g}(\hat{S}(a_{g-1}) - \hat{S}(a_g)) \tag{2.23}$$

Therefore for each interval $I_g = [a_{g-1}, a_g)$, $g = 1,...,G$, using (2.22) and (2.23), we have estimated the mean cost given by

$$\hat{\mu} = \sum_{g=1}^{G} \{ \frac{\sum_{i=1}^{n} \delta_{ig}^{a} y_{ig}}{\sum_{i=1}^{n} \delta_{ig}^{a}} \hat{S}(a_g) + [(\sum_{i=1}^{n} y_{ig} \frac{\delta_{ig}^{d}}{\hat{G}(X_i)}) / (\sum_{i=1}^{n} \frac{\delta_{ig}^{d}}{\hat{G}(X_i)})][\hat{S}(a_{g-1}) - \hat{S}(a_g))] \} \quad (2.24)$$

Throughout we have assumed that both S and G are continuous and $S(t) = P(T > t)$, $G(t) = P(U > t)$. If this was not the case here, then keeping all previous notations we must use left-hand limits $\hat{S}(t-)$ and $\hat{G}(t-)$ throughout. Alternatively we could define $S(t) = P(T \geq t)$ and $G(t) = P(U \geq t)$, i.e. left continuous versions of the previously defined (right-continuous) survival distributions. When S and G are continuous this distinction is unnecessary, however when it comes to estimation, and $\hat{S}$ and $\hat{G}$ denote Kaplan-Meier estimates we need to make the distinction. If there is an initial cost at time 0 then it needs to be added to the final mean cost expression, however for now, as previously specified, we assume initial cost to be null. To ensure all weights are defined properly we assume $\hat{G}(\tau) > 0$.

Using general properties of the Kaplan-Meier estimator $\hat{S}(t)$ of $S(t) = P(T > t)$, $\Delta\hat{S}(t) = -\hat{S}(t-)\frac{\Delta N(t)}{Y(t)}$, and under the assumptions of no ties among failure times we have $\Delta\hat{S}(T_i) = \frac{-\delta_i \hat{S}(T_i-)}{Y(T_i)}$. If we assume there are no ties among failure and censoring time then $\hat{S}(t-)\hat{G}(t-) = Y(t)/n$. Both assumptions hold under the case that both S and G are continuous. Then

$$\sum_{i=1}^{n}\frac{\delta_{ig}^{d}}{\hat{G}(X_i-)}=\sum_{i=1}^{n}\frac{[a_{g-1}\le T_i<a_g,\delta_i=1]}{\hat{G}(T_i-)}=n\sum_{i=1}^{n}\frac{\delta_i\,\hat{S}(T_i-)[a_{g-1}\le T_i<a_g]}{Y(T_i)}=$$

$$=n\sum_{i=1}^{n}(\hat{S}(T_i-)-\hat{S}(T_i))[a_{g-1}\le T_i<a_g]=n(\hat{S}(a_{g-1}-)-\hat{S}(a_g-)) \quad (2.25)$$

Also

$$\sum_{i=1}^{n}\delta_{ig}^{a}=\sum_{i=1}^{n}[X_i\ge a_g]=Y(a_g)=n\hat{S}(a_g-)\hat{G}(a_g-) \quad (2.26)$$

Using (2.25) and (2.26) in the two terms of (2.24),

$$\sum_{g=1}^{G}\left(\frac{\sum_{i=1}^{n}\delta_{ig}^{a}y_{ig}}{\sum_{i=1}^{n}\delta_{ig}^{a}}\right)\hat{S}(a_g-)=\sum_{g=1}^{G}\left(\frac{\sum_{i=1}^{n}\delta_{ig}^{a}y_{ig}}{n\hat{S}(a_g-)\hat{G}(a_g-)}\right)\hat{S}(a_g-)=\sum_{g=1}^{G}\sum_{i=1}^{n}\delta_{ig}^{a}y_{ig}/n\hat{G}(a_g-),$$

$$\sum_{g=1}^{G}\left[\left(\sum_{i=1}^{n}y_{ig}\frac{\delta_{ig}^{d}}{\hat{G}(X_i-)}\right)/\left(\sum_{i=1}^{n}\frac{\delta_{ig}^{d}}{\hat{G}(X_i-)}\right)\right](\hat{S}(a_{g-1}-)-\hat{S}(a_g-))=$$

$$=\sum_{g=1}^{G}\sum_{i=1}^{n}y_{ig}\frac{\delta_{ig}^{d}}{n\hat{G}(X_i-)}$$

allows simplification of the mean cost:

$$\hat{\mu}=\sum_{g=1}^{G}\left\{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\delta_{ig}^{d}y_{ig}}{\hat{G}(X_i-)}+\frac{\delta_{ig}^{a}y_{ig}}{\hat{G}(a_g-)}\right)\right\}=\sum_{g=1}^{G}\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_{ig}y_{ig}}{\hat{G}(X_{ig}^{*}-)}\right\}=\sum_{g=1}^{G}\left\{\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_{ig}y_{ig}}{\hat{G}(T_{ig}^{*}-)}\right\}$$

where $\delta_{ig}=\delta_{ig}^{a}+\delta_{ig}^{d}=[X_i\ge a_g]+[a_{g-1}\le X_i<a_g,\delta_i=1]$,

$X_{ig}^{*}=\min(X_i,a_g)=\min(X_i^{*},a_g)$ and T_{ig}^{*} the same as in (2.7).

This is the same as Lin's estimator (2000)[4] (2.7) for the interval case data. Indeed note that

$$\delta_{ig}^{*} = [T_i \wedge a_g \le U_i] = [X_i \ge a_g] + [X_i < a_g, \delta_i = 1] =$$

$$= [X_i \ge a_g] + [a_{g-1} \le X_i < a_g, \delta_i = 1] + [X_i < a_{g-1}, \delta_i = 1] =$$

$$= \delta_{ig} + [X_i < a_{g-1}, \delta_i = 1].$$

Also under the no ties assumptions above and using the last equality in (2.21), with the role of T_i^{*} played by T_{ig}^{*}, $\sum_{i=1}^{n} \frac{\delta_{ig}^{*}}{\hat{G}(X_{ig}^{*}-)} = \sum_{i=1}^{n} \frac{[T_{ig}^{*} \le U_i]}{\hat{G}(T_{ig}^{*}-)} = n$ for all $g = 1,\ldots,G$.

Therefore, the mean cost is

$$\frac{1}{n}\sum_{g=1}^{G}\sum_{i=1}^{n} \frac{\delta_{ig}}{\hat{G}(X_{ig}^{*})} y_{ig} = \frac{1}{n}\sum_{g=1}^{G}\sum_{i=1}^{n} \frac{\delta_{ig}^{*}}{\hat{G}(X_{ig}^{*})} y_{ig} + \frac{1}{n}\sum_{g=1}^{G}\sum_{i=1}^{n} \frac{[X_i < a_{g-1}, \delta_i = 1]}{\hat{G}(X_{ig}^{*})} y_{ig} =$$

$$= \frac{1}{n}\sum_{g=1}^{G}\sum_{i=1}^{n} \frac{\delta_{ig}^{*}}{\hat{G}(X_{ig}^{*})} y_{ig} = \sum_{g=1}^{G} \left(\sum_{i=1}^{n} \frac{\delta_{ig}^{*}}{\hat{G}(X_{ig}^{*})}\right)^{-1} \left(\sum_{i=1}^{n} \frac{\delta_{ig}^{*}}{\hat{G}(X_{ig}^{*})} y_{ig}\right)$$

which is exactly (2.7).

2.3 Asymptotic normality of the mean cost

We are interested in calculating the variance of the estimate $\int_0^{\tau} \hat{c}_{01}(t \mid Z_0)(-d\hat{S}(t \mid Z_0))$ of the mean for transition costs, $\int_0^{\tau} c_{01}(t \mid Z_0)(-dS(t \mid Z_0))$. There are various possibilities for estimating $S(t \mid Z_0)$. In Chapter 1, we have already developed estimation methods, proved asymptotic properties and calculated the asymptotic variance of the estimators of $\int_0^{\tau} c_{01}(t \mid Z_0)(-dS(t \mid Z_0))$ when $S(t \mid Z_0)$ is

estimated from a Cox proportional hazards model and $c_{01}(t \mid Z_0)$ is estimated from a regression model.

If the assumption of proportional hazards is not satisfied we may consider estimating the survival function for time using a stratified Kaplan-Meier estimate. We will ignore covariates in the estimation of $S(t)$. Let $\hat{S}(t)$ be the Kaplan Meier estimate of $S(t)$ and suppose $\hat{c}(s) = \hat{c}_{01}(s \mid Z_0)$ is estimated from the regression model (2.11). Then $\hat{\mu}(Z_0) = -\int_0^{\tau} \hat{c}(s) d\hat{S}(t)$. For ease of notations we assume that $T_i \le \tau$ for all patients in the data set. Then $s_i = [T_i \le U_i \wedge \tau] = [T_i \le U_i] = \delta_i$. This assumption is implicit in Bang and Tstiatis (2000)[31].

Consider the regression model (2.11):

$$y_i = \mathbf{x}_i \beta + \varepsilon_i$$

where $\mathbf{x}_i$ is a $1 \times p$ vector of covariates with first component 1 so that the first component of β corresponds to an intercept. Under model assumptions **A1** and **A2**, see (2.12), (2.13), a consistent estimator for β is given in (2.14)

$$\hat{\beta}_w = (\sum_{i=1}^{n} w_i \mathbf{x}_i' \mathbf{x}_i)^{-1} (\sum_{i=1}^{n} w_i \mathbf{x}_i' y_i)$$

where $w_i = \dfrac{\delta_i}{G(T_i -)}$. Therefore $\hat{\beta}_w - \beta = (\sum_{i=1}^{n} w_i \mathbf{x}_i' \mathbf{x}_i)^{-1} (\sum_{i=1}^{n} w_i \mathbf{x}_i' \varepsilon_i)$.

Condition **A1** can be weakened to **A'1**: $E(\mathbf{x}_i \varepsilon_i) = 0$, $E(\varepsilon_i^2 \mid \mathbf{x}_i) = \sigma_\varepsilon^2$. In fact under **A'1** and **A2**, the IPW estimator is always consistent. Indeed

$\frac{1}{n} \sum_{i=1}^{n} w_i \mathbf{x}_i' \mathbf{x}_i \xrightarrow{P} E(\mathbf{x}_i' \mathbf{x}_i)$, since $E(w_i \mathbf{x}_i' \mathbf{x}_i) = E(E(w_i \mid \mathbf{x}_i) \mathbf{x}_i' \mathbf{x}_i) = E(\mathbf{x}_i' \mathbf{x}_i) \overset{def}{=} \mathbf{A}$, and

$\frac{1}{n}\sum_{i=1}^{n} w_i \mathbf{x}_i' \varepsilon_i \xrightarrow{P} 0$, since $E(w_i \mid \mathbf{x}_i, y_i, T_i) = P(U_i \geq T_i \mid \mathbf{x}_i, y_i, T_i)/G(T_i -) = 1$ and under

A'1, $E(w_i \mathbf{x}_i' \varepsilon_i) = E(E(w_i \mathbf{x}_i' \varepsilon_i \mid \mathbf{x}_i, y_i)) = E(\mathbf{x}_i' \varepsilon_i E(w_i \mid \mathbf{x}_i, y_i)) = E(\mathbf{x}_i' \varepsilon_i) = 0$.

We need **A1** and **A2** for the consistency of the (unweighted) estimator

$\hat{\beta}_u = (\sum_{i=1}^{n} \delta_i \mathbf{x}_i' \mathbf{x}_i)^{-1} (\sum_{i=1}^{n} \delta_i \mathbf{x}_i' y_i)$. We have

$$\sqrt{n}(\hat{\beta}_w - \beta) = (\frac{1}{n}\sum_{i=1}^{n} w_i \mathbf{x}_i' \mathbf{x}_i)^{-1} (\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_i \mathbf{x}_i' \varepsilon_i) = (\mathbf{A}^{-1} + o_p(1))(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_i \mathbf{x}_i' \varepsilon_i).$$

By the ordinary CLT, $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_i \mathbf{x}_i' \varepsilon_i \xrightarrow{\mathcal{D}} N(\underline{0}, \mathbf{B})$, where $\mathbf{B} = E(w_i^2 \mathbf{x}_i' \mathbf{x}_i \varepsilon_i^2)$. Using **A1,**

the form of $\mathbf{B}$ can be reduced further. Since $E(w_i^2 \mid \mathbf{x}_i, y_i, T_i) =$

$P(U_i \geq T_i \mid \mathbf{x}_i, y_i, T_i)/G^2(T_i -) = 1/G(T_i -)$, it follows that $\mathbf{B} = E(w_i^2 \mathbf{x}_i' \mathbf{x}_i \varepsilon_i^2) =$

$E(E(w_i^2 \mid \mathbf{x}_i, y_i, T_i)\mathbf{x}_i' \mathbf{x}_i \varepsilon_i^2) = E((1/G(T_i -))\mathbf{x}_i' \mathbf{x}_i \varepsilon_i^2) = E((1/G(T_i -))\mathbf{x}_i' \mathbf{x}_i E(\varepsilon_i^2 \mid \mathbf{x}_i)) =$

$= \sigma_\varepsilon^2 E(\frac{\mathbf{x}_i' \mathbf{x}_i}{G(T_i -)})$. Therefore

$$\sqrt{n}(\hat{\beta}_w - \beta) \xrightarrow{\mathcal{D}} N(0, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}),$$

where $\mathbf{A} = E(\mathbf{x}_i' \mathbf{x}_i)$ and $\mathbf{B} = \sigma_\varepsilon^2 E(\frac{\mathbf{x}_i' \mathbf{x}_i}{G(T_i -)})$.

Then $\hat{c}(t) = \mathbf{x}_0(t)\hat{\beta}_w$ and for a fixed t, $\sqrt{n}(\hat{c}(t) - c(t)) = \mathbf{x}_0(t)(\sqrt{n}(\hat{\beta}_w - \beta))$. It

follows that:

$$\sqrt{n}(\hat{c}(t) - c(t)) = \mathbf{x}_0(t)\mathbf{A}^{-1}(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_i \mathbf{x}_i' \varepsilon_i) + o_p(1) \qquad (2.27)$$

We will denote $v_0(t) = x_0(t)E(\mathbf{x}'\mathbf{x})^{-1}$ so $\sqrt{n}(\hat{c}(t) - c(t)) = v_0(t)(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_i \mathbf{x}_i' \varepsilon_i) + o_p(1)$.

We now turn to the estimation of the survival function $S(t) = P(T > t)$. Let $\hat{S}(t)$ denote the Kaplan-Meier estimator of based on the data $\{ (X_i, \delta_i) : i = 1,...,n \}$. Note that $d\hat{S}(t) = -\hat{S}(t-)d\hat{A}(t) = -\hat{S}(t-)\frac{dN(t)}{Y(t)}$, where $N(t) = \sum_{i=1}^{n}[T_i \le t, \delta_i = 1]$, $Y(t) = \sum_{i=1}^{n}[X_i \ge t]$ and $\hat{A}(t) = \int_0^t \frac{J(s)dN(s)}{Y(s)}$ is the usual Nelson-Aalen estimator of the integrated hazard function (Andersen *et al* (1993)[25]), where $J(s) = [Y(s) > 0]$. We will assume both $S(t)$ and $G(t) = P(U > t)$ are continuous and $S(\tau) > 0$.

Notations:

$\pi(t) = P(X \ge t) = H(t) = S(t)G(t)$

$\sigma^2(t) = \int_0^t \frac{\alpha(u)}{\pi(u)}du$

Model Assumptions (Andersen *et al* (1993), p190[25]):

ABGK1: For each $s \in [0,\tau]$, $n\int_0^\tau \frac{J(s)}{Y(s)}\alpha(s)ds \xrightarrow{P} \sigma^2(s)$ as $n \to \infty$

ABGK2: For all $\varepsilon > 0$, $n\int_0^t \frac{J(s)}{Y(s)}[|\sqrt{n}\frac{J(s)}{Y(s)}| > \varepsilon]\alpha(s)ds \xrightarrow{P} 0$ as $n \to \infty$

ABGK3: $\sqrt{n}\int_0^\tau (1 - J(s))\alpha(s)ds \xrightarrow{P} 0$ as $n \to \infty$

Under conditions **ABGK1-3**, following steps in theorems IV.1.1, IV.1.2(Andersen *et al* (1993), p190[25]) and our proofs in Chapter 1, Section 1.3, $\sqrt{n}(\hat{A}(t) - A(t))$ is asymptotically equivalent to $\frac{1}{\sqrt{n}}\int_0^t \frac{dM(s)}{\pi(s)} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_0^t \frac{dM_i(s)}{\pi(s)}$,

where $M(t) = \sum_{i=1}^{n} M_i(t)$, $M_i(t) = N_i(t) - \int_0^t Y_i(t)dA(s)$. Therefore

$\int_0^t \frac{dM_i(s)}{\pi(s)} = \frac{\delta_i[X_i \le t]}{\pi(X_i)} - \int_0^t \frac{Y_i(s)dA(s)}{\pi(s)}$. Let $\sigma^2(t) = \int_0^t \frac{dA(s)}{\pi(s)}$. Then $\sqrt{n}(\hat{A}(t) - A(t))$ is

asymptotically equivalent to

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_i[X_i \le t]}{\pi(X_i)} - \int_0^t \frac{Y_i(s)dA(s)}{\pi(s)} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\frac{\delta_i[X_i \le t]}{\pi(X_i)} - \sigma^2(X_i \wedge t) \right).$$

Using $S(t) = \exp(-A(t))$ and the delta method we conclude that $\sqrt{n}(\hat{S}(t) - S(t))$

is asymptotically equivalent to $-S(t)(\sqrt{n}(\hat{A}(t) - A(t)))$. Therefore

$$\sqrt{n}(\hat{S}(t) - S(t)) \overset{a.e.}{=} -\frac{1}{\sqrt{n}} S(t) \sum_{i=1}^{n} \left\{ \frac{\delta_i[X_i \le t]}{\pi(X_i)} - \sigma^2(X_i \wedge t) \right\} \tag{2.28}$$

where we use notation $a_n \overset{a.e.}{=} b_n$ to imply asymptotic equivalence. Using (2.27) and (2.28)

we obtain

$$\begin{pmatrix} \sqrt{n}(\hat{S}(t) - S(t)) \\ \sqrt{n}(\hat{c}_{01}(t) - c_{01}(t)) \end{pmatrix} \overset{a.e.}{=} \begin{pmatrix} -S(t) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \frac{\delta_i[X_i \le t]}{\pi(X_i)} - \sigma^2(X_i \wedge t) \right\} \\ v_0(t) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_i \mathbf{x}_i' \varepsilon_i \end{pmatrix} \tag{2.29}$$

Under the assumptions **A1-A2** and **ABGK1-3** we have, by the ordinary CLT:

$$\begin{pmatrix} \sqrt{n}(\hat{S}(t) - S(t)) \\ \sqrt{n}(\hat{c}_{01}(t) - c_{01}(t)) \end{pmatrix} \xrightarrow{\mathcal{D}} N(0_2, \Sigma_{2x2}(t)).$$

where the elements of the asymptotic variance are

$$\Sigma_{2x2}(t) = \begin{pmatrix} \sigma^2(t) S^2(t) & 0 \\ 0 & v_0(t) \sigma_\varepsilon^2 E\left(\frac{\mathbf{x}_i' \mathbf{x}_i}{G(T_i -)} \right) v_0(t)' \end{pmatrix}$$

The components of $\sum_{2x2}(t)$ are derived as follows. The first diagonal component is

$S^2(t)Var\{\frac{\delta_i[X_i \le t]}{\pi(X_i)} - \sigma^2(X_i \wedge t)\}$. Using the independence of U_i and T_i,

$$E(\frac{\delta_i[X_i \le t]}{\pi^2(X_i)}) = E(E(\frac{\delta_i[X_i \le t]}{\pi^2(X_i)} | T_i)) =$$

$$= E(\frac{G(T_i)}{\pi^2(T_i)}[T_i \le t]) = E(\frac{1}{S(T_i)\pi(T_i)}[T_i \le t]) = \sigma^2(t),$$

$$E(\frac{\delta_i[X_i \le t]}{\pi(X_i)}\sigma^2(X_i \wedge t)) = E(E(\frac{\delta_i[X_i \le t]}{\pi(X_i)}\sigma^2(X_i \wedge t) | T_i)) =$$

$$= E(\frac{[T_i \le t]}{S(T_i)}\sigma^2(T_i \wedge t)) = \int_0^t \alpha(u)\sigma^2(u)du, \text{ and}$$

$$E(\sigma^4(X_i \wedge t)) = E(\sigma^4(X_i)[X_i \le t]) + \sigma^4(t)P(X_i > t) =$$

$$= \int_0^t \sigma^4(u)(-d\pi(u)) + \sigma^4(t)\pi(t) = \int_0^t \pi(u)2\sigma^2(u)d\sigma^2(u) = 2\int_0^t \sigma^2(u)\alpha(u)du$$

The first diagonal component is:

$$\sum_{11}(t) = Var(-S(t)\frac{\delta_i[X_i \le t]}{\pi(X_i)} - \sigma^2(X_i \wedge t))) = S^2(t)E((\frac{\delta_i[X_i \le t]}{\pi(X_i)} - \sigma^2(X_i \wedge t))^2) =$$

$$= S^2(t)E(\frac{\delta_i[X_i \le t]}{\pi^2(X_i)} - 2\frac{\delta_i[X_i \le t]}{\pi(X_i)}\sigma^2(X_i \wedge t) + \sigma^4(X_i \wedge t)) \qquad (2.30)$$

Therefore $\sum_{11}(t) = S^2(t)\sigma^2(t)$. This result can be also be derived using the martingale representation of $\sqrt{n}(\hat{A}(t) - A(t))$. Also $\sum_{12}(t) = E((\frac{\delta_i[X_i \le t]}{\pi(X_i)} - \sigma^2(X_i \wedge t))w_i\mathbf{x}_i'\varepsilon_i) =$

$$= E(E((\frac{\delta_i[X_i \le t]}{\pi(X_i)} - \sigma^2(X_i \wedge t))\frac{\delta_i}{G(T_i)}\mathbf{x}_i'\varepsilon_i \mid \mathbf{x}_i, y_i)) =$$

$$= E((\frac{[T_i \le t]}{\pi(T_i)} - \sigma^2(T_i \wedge t))\mathbf{x}_i'\varepsilon_i) = E(E((\frac{[T_i \le t]}{\pi(T_i)} - \sigma^2(T_i \wedge t))\mathbf{x}_i'\varepsilon_i \mid \mathbf{x}_i)) =$$

$= E((\frac{[T_i \le t]}{\pi(T_i)} - \sigma^2(T_i \wedge t))\mathbf{x}_i' E(\varepsilon_i \mid \mathbf{x}_i)) = 0$ since by assumption **A1**, $E(\varepsilon_i \mid \mathbf{x}_i) = 0$. The second diagonal component is $\Sigma_{22}(t) = Var(v_0(t)w_i\mathbf{x}_i'\varepsilon_i) = v_0(t)\sigma_\varepsilon^2 E(\frac{\mathbf{x}_i'\mathbf{x}_i}{G(T_i-)})v_0(t)'$.

Next we use Funtional Delta Method and a similar version of Lemma 2 in Chapter 1 to prove the asymptotic normality of mean cost.

Consider the functional

$$\varphi: E^* \to \mathbb{R},\ \varphi(x,y)(t) = \int_0^t x(s)dy(s),$$

where E^* is a subset of $D[0,\tau]^2$, $E^* = \left\{(x,y) \in D[0,\tau]^2 : \int_0^\tau |dy| \le C\right\}$, where $0 < C < \infty$.

Let (x_0, y_0) be a fixed point of E^* such that $\int_0^\tau |d(x_0)| < \infty$. Then φ_t can be extended to the space $D[0,\tau]^2$ so as to be Hadamard differentiable at (x_0, y_0), with derivative

$$d\phi(x,y)(h,k)(t) = \int_0^t x(s)dk(s) + \int_0^t h(s)dy(s) \qquad (2.31)$$

where the integral with respect to k is defined by the integration by parts formula if k is not of finite variation.

Identifying $x(t)$ with $c(t)$, and $y(t)$ with $S(t)$, since $c(t), S(t) \in D[0,\tau]$, $\int_0^\tau |dS| = 1 - S(\tau) < C = 1$ and under the extra assumption

EA1: $c(t)$ is of finite variation on $[0,\tau]$,

by (2.31) and functional delta method stated in the Appendix,

$$\sqrt{n}\{\int_0^\tau \hat{c}(t)(-d\hat{S}(t)) - \int_0^\tau c(t)(-dS(t))\} =$$

$$= \sqrt{n}(\phi(\hat{c},-\hat{S})(\tau) - \phi(c,-S)(\tau)) \xrightarrow{\mathcal{D}} d\phi(c,-S)(Z_1,Z_2)(\tau)$$

where $Z_1(t) = a_0(t)\xi$, $a_0^2(t) = v_0(t)\sigma_\varepsilon^2 E(\frac{\mathbf{x}_i'\mathbf{x}_i}{G(T_i-)})v_0(t)'$, $v_0(t) = x_0(t)E(\mathbf{x}'\mathbf{x})^{-1}$,

$Z_2(t) = -S(t)U(t)$, $U(t)$ is a gaussian martingale independent of ξ, which is a standard normal random variable. Also $E(U(t)) = 0$ and $E(U^2(t)) = \sigma^2(t)$.

Using (2.31) we get

$$\sqrt{n}\{\int_0^\tau \hat{c}(t)(-d\hat{S}(t)) - \int_0^\tau c(t)(-dS(t))\} \xrightarrow{\mathcal{D}} \int_0^\tau c(s)dZ_2(s) + \int_0^\tau Z_1(s)dS(s).$$

To compute $Var(\int_0^\tau c(s)dZ_2(s)) = E((\int_0^\tau c(s)dZ_2(s))^2) =$

$= E((\int_0^\tau c(s)d(-S(s)U(s))^2)$ we just simplify the integrand

$$\int_0^\tau c(s)U(s)S(s)dA(s) + \int_0^\tau c(s)(-S(s))dU(s) = \int_0^\tau U(s)d\bar{A}(s) + \int_0^\tau c(s)(-S(s))dU(s),$$

where $\bar{A}(t) = \int_t^\tau c(s)S(s)dA(s)$. By an integration by parts for the second term

$\int_0^\tau c(s)d(-S(s)U(s)) = -\int_0^\tau (c(s)S(s) - \bar{A}(s))dU(s)$. Since $U(t)$ is a zero-mean gaussian martingale, with $<U(t)>=\sigma^2(t)$, we get immediately

$Var(\int_0^\tau c(s)dZ_2(s)) = \int_0^\tau (c(s)S(s) - \int_s^\tau c(u)S(u)dA(u))^2 d\sigma^2(s)$. The second term $\int_0^\tau Z_1(s)dS(s)$ has variance $(\int_0^\tau a_0(s)d\dot{S}(s))^2$ and the cross product has mean zero, since $U(t)$ and ξ are independent. Hence we have proved the next theorem:

Theorem 1

Under the assumptions **A1-A2, EA1** and **ABGK1-3** we have

$$\sqrt{n}\left(\int_0^\tau \hat{c}(t)(-d\hat{S}(t)) - \int_0^\tau c(t)(-dS(t))\right) \xrightarrow{\mathcal{D}}$$

$$N\left(0,\left\{\int_0^\tau (c(s)S(s)-\bar{A}(s))^2 d\sigma^2(s)\right)\right) + \left(\int_0^\tau a_0(s)dS(s))^2\right\}). \square \quad (2.32)$$

Remark 2.3.1

1. Suppose costs are not considered random, but fixed. Take $c(t)=1$.

Then $\sqrt{n}\left(\int_0^\tau \hat{c}(t)(-d\hat{S}(t)) - \int_0^\tau c(t)(-dS(t))\right) = -\sqrt{n}(\hat{S}(t)-S(t))$. The asymptotic variance in the above theorem has only the first term,

$\bar{A}(t) = \int_t^\tau c(t)S(t)dA(t) = S(\tau)-S(t)$ and so,

$\int_0^\tau (c(s)S(s)-\bar{A}(s))^2 d\sigma^2(s))) = S^2(\tau)\sigma^2(\tau)$. This verifies the well known result,

$\sqrt{n}(\hat{S}(t)-S(t)) \xrightarrow{\mathcal{D}} N(0, S^2(\tau)\sigma^2(\tau))$.

2. Let cost be fixed and set $c(t)=e^{-rt}$. Then $\int_0^t e^{-rt}(-dS(t)) = \int_0^t e^{-rt}S(t)dA(t)$ is the actuarial value discussed in Andersen *et al* (1993), page 284[25]. Then (2.32) captures the results stated on page 284, namely

$$\sqrt{n}\left(\int_0^t e^{-rt}\hat{S}(t)d\hat{A}(t) - \int_0^t e^{-rt}S(t)dA(t)\right) \xrightarrow{\mathcal{D}}$$

$$N\left(0,\left\{\int_0^\tau (e^{-rs}S(s) - \int_s^\tau e^{-ru}S(u)dA(u))^2 d\sigma^2(s)\right\}\right).$$

3. If no fixed or time-varying covariates are included then the cost $c(t)$ is constant on t and can be estimated as $\hat{c}(t) = \hat{\beta}_0 = \dfrac{\sum_{i=1}^{n} \dfrac{\delta_i}{G(T_i)} y_i}{\sum_{i=1}^{n} \dfrac{\delta_i}{G(T_i)}}$ for all $t \leq \tau$. Then

$\hat{\mu} = \hat{\beta}_0 (1 - \hat{S}(\tau))$ and from (2.32) the estimated asymptotic variance is $\hat{\beta}_0^2 \hat{S}^2(\tau)\hat{\sigma}^2(\tau) + (1 - \hat{S}(\tau))^2 \text{var}(\hat{\beta}_0)$.

2.3.1 Regression model for log-cost

Suppose all costs in our data set are not null and consider the regression model with log cost as dependent variable:

$$\log(y_i) = \mathbf{x}_i \beta + \varepsilon_i \tag{2.33}$$

where we keep all our previous notations. In this section, we consider the case of an error term with unknown distribution under homoscedasticity. More precisely we assume:

A3: ε_i are i.i.d. with $\varepsilon_i = v_i \sqrt{s(\mathbf{x}_i, \alpha)}$ (the function $s(\cdot)$ allows for heteroscedasticity conditional on $\mathbf{x}$ and α is a vector of unknown parameters)

A4: v_i is independent of $\mathbf{x}_i$ and has zero mean and unit variance.

A5: *rank* $E(\mathbf{x}_i' \mathbf{x}_i) = p$

A6: $E(e^{\varepsilon} \mathbf{x}) < \infty$

Our focus is the estimation of $c(t) = c_{01}(t \mid Z_0)$ as in the previous section. In view of (2.33) we identify $c(t)$ with $E(y_i \mid \mathbf{x}_0(t)) = e^{\mathbf{x}_0(t)\beta} E(e^{\varepsilon})$, where $\mathbf{x}_0(t)$ is a fixed covariate profile generated from Z_0. Following Ai and Norton (2000)[20] estimation of $c(t)$ is accomplished by estimating β in (2.33) and the smearing factor $a = E(e^{\varepsilon})$. Define $\varphi(t,\beta,a) = e^{x_0(t)\beta} a$ and identify $c(t)$ with $\varphi(t,\beta,a)$. Then $\hat{c}(t) = \varphi(t,\hat{\beta},\hat{a})$ and $\hat{a} = \frac{1}{n}\sum_{i=1}^{n} e^{\hat{\varepsilon}_i}$ and $\hat{\beta}$ is an estimator of β which will be obtained later. Here $\hat{\varepsilon}_i = \log(y_i) - \mathbf{x}_i\beta$. Then $\sqrt{n}(\hat{c}(t) - c(t)) = \sqrt{n}(e^{x_0(t)\hat{\beta}}\hat{a} - e^{x_0(t)\beta}a)$.

By Taylor expansion, retaining only the first order terms, we get:

$$\sqrt{n}(\hat{c}(t) - c(t)) = \sqrt{n}(\varphi(t,\hat{\beta},\hat{a}) - \varphi(t,\beta,a)) \sim$$

$$= \sqrt{n}\left(\frac{\partial\varphi(t,\beta,a)}{\partial\beta}(\hat{\beta} - \beta) + \frac{\partial\varphi(t,\beta,a)}{\partial a}(\hat{a} - a)\right)$$

$$= \sqrt{n}(c(t)\mathbf{x}_0(t)(\hat{\beta} - \beta) + e^{\mathbf{x}_0(t)\beta}(\hat{a} - a)) =$$

$$= c(t)\mathbf{x}_0(t)\sqrt{n}(\hat{\beta} - \beta) + e^{\mathbf{x}_0(t)\beta}\sqrt{n}(\hat{a} - a). \qquad (2.34)$$

Consider first the estimation of β in (2.33). We follow exactly the same steps as in the previous section 2.3 using $\eta_i = \log(y_i)$ instead of y_i. Then

$\hat{\beta} = (\sum_{i=1}^{n} w_i \mathbf{x}_i' \mathbf{x}_i)^{-1} (\sum_{i=1}^{n} w_i \mathbf{x}_i' \eta_i)$, where $w_i = \frac{\delta_i}{G(T_i)}$. Therefore

$$\sqrt{n}(\hat{\beta} - \beta) = \left(\frac{1}{n}\sum_{i=1}^{n} w_i \mathbf{x}_i' \mathbf{x}_i\right)^{-1} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_i \mathbf{x}_i' \varepsilon_i\right) \qquad (2.35)$$

From **A5**, $\frac{1}{n}\sum_{i=1}^{n} w_i \mathbf{x}_i' \mathbf{x}_i \xrightarrow{P} E(w\mathbf{x}'\mathbf{x}) = E(\mathbf{x}'\mathbf{x}) = \mathbf{A}$. From **A3/A4**, $E(w_i \mathbf{x}_i' \varepsilon_i) =$

$$= E(E(w_i \mathbf{x}_i' \varepsilon_i \mid \mathbf{x}_i, y_i)) = E(\mathbf{x}_i' \varepsilon_i E(w_i \mid \mathbf{x}_i, y_i)) = E(\mathbf{x}_i' \varepsilon_i) == E(\mathbf{x}_i' v_i \sqrt{s(\mathbf{x}_i, \alpha)}) =$$

$$= E(\mathbf{x}_i' \sqrt{s(\mathbf{x}_i, \alpha)} E(v_i \mid \mathbf{x}_i)) = 0.$$

Hence $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_i \mathbf{x}_i' \varepsilon_i = O_p(1)$. Then (2.35) yields

$$\sqrt{n}(\hat{\beta} - \beta) = \mathbf{A}^{-1} (\frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_i \mathbf{x}_i' \varepsilon_i) + o_p(1) \tag{2.36}$$

Next, consider the estimation of a by $\hat{a}$. Then

$$\sqrt{n}(\hat{a} - a) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (e^{\hat{\varepsilon}_i} - e^{\varepsilon_i}) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (e^{\varepsilon_i} - E(e^{\varepsilon_i}))$$

Use

$$e^{\hat{\varepsilon}_i} - e^{\varepsilon_i} = e^{\varepsilon_i} (e^{\hat{\varepsilon}_i - \varepsilon_i} - 1) = e^{\varepsilon_i} (\exp((\eta_i - \mathbf{x}_i \hat{\beta}) - (\eta_i - \mathbf{x}_i \beta)) - 1) = e^{\varepsilon_i} \exp(-\mathbf{x}_i (\hat{\beta} - \beta) - 1)$$

to obtain $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (e^{\hat{\varepsilon}_i} - e^{\varepsilon_i}) = -(\frac{1}{n} \sum_{i=1}^{n} e^{\varepsilon_i} \mathbf{x}_i) \sqrt{n}(\hat{\beta} - \beta) + o_p(1)$. By **A6** we get

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (e^{\hat{\varepsilon}_i} - e^{\varepsilon_i}) = -E(e^{\varepsilon} \mathbf{x}) \sqrt{n}(\hat{\beta} - \beta) + o_p(1) \tag{2.37}$$

Using formulas (2.34), (2.36) and (2.37) we obtain

$$\sqrt{n}(\hat{c}(t) - c(t)) = c(t) \mathbf{x}_0(t) \sqrt{n}(\hat{\beta} - \beta) + e^{\mathbf{x}_0(t)\beta} \sqrt{n}(\hat{a} - a) =$$

$$= c(t) \mathbf{x}_0(t) (\mathbf{A}^{-1} (\frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_i \mathbf{x}_i' \varepsilon_i) + o_p(1)) +$$

$$+ e^{\mathbf{x}_0(t)\beta} (-E(e^{\varepsilon} \mathbf{x}) (\mathbf{A}^{-1} (\frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_i \mathbf{x}_i' \varepsilon_i) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (e^{\varepsilon_i} - E(e^{\varepsilon_i})) + o_p(1))) + o_p(1)) =$$

$$= (c(t) \mathbf{x}_0(t) - e^{\mathbf{x}_0(t)\beta} E(e^{\varepsilon} \mathbf{x})) \mathbf{A}^{-1} \{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} w_i \mathbf{x}_i' \varepsilon_i\} + e^{\mathbf{x}_0(t)\beta} \{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (e^{\varepsilon_i} - E(e^{\varepsilon_i}))\} + o_p(1).$$

If we denote $\vartheta_1(t) = (c(t)\mathbf{x}_0(t) - e^{\mathbf{x}_0(t)\beta}E(e^{\varepsilon}\mathbf{x}))\mathbf{A}^{-1}$ and $\vartheta_2(t) = e^{\mathbf{x}_0(t)\beta}$ then

$$\sqrt{n}(\hat{c}(t) - c(t)) = \vartheta_1(t)\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} w_i\mathbf{x}_i'\varepsilon_i\} + \vartheta_2(t)\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(e^{\varepsilon_i} - E(e^{\varepsilon_i}))\} + o_p(1).$$

If $s(\mathbf{x},\alpha) = \alpha$ (i.e. the error term does not depend on $\mathbf{x}$, so the error term is homoscedastic) as in Duan (1983)[19] then following the same steps as in the previous section $\Sigma_{12}(t) = E[(\frac{\delta_i[X_i \le t]}{\pi(X_i)} - \sigma^2(X_i \wedge t))\{\vartheta_1(t)w_i\mathbf{x}_i'\varepsilon_i + \vartheta_2(t)(e^{\varepsilon_i} - E(e^{\varepsilon_i}))\}] =$

$$= E[(\frac{[T_i \le t]}{\pi(T_i)} - \sigma^2(T_i \wedge t))\vartheta_1(t)\mathbf{x}_i'E(\varepsilon_i \mid \mathbf{x}_i)] +$$

$$+ E[(\frac{[T_i \le t]}{\pi(T_i)} - \sigma^2(T_i \wedge t))\vartheta_2(t)(e^{\varepsilon_i} - E(e^{\varepsilon_i}))] = 0$$

since by **A3/A4**, $E(\varepsilon_i \mid \mathbf{x}_i) = E(v_i\sqrt{\alpha} \mid \mathbf{x}_i) = E(v_i\sqrt{\alpha}) = 0$ and

$$E(e^{\varepsilon_i} - E(e^{\varepsilon_i}) \mid \mathbf{x}_i) = E(e^{v_i\sqrt{\alpha}} \mid \mathbf{x}_i) - E(e^{\varepsilon_i}) = E(e^{v_i\sqrt{\alpha}}) - E(e^{\varepsilon_i}) = 0.$$

The second diagonal component of $\sum_{2x2}(t)$ is

$$\Sigma_{22}(t) = \vartheta_1(t)Var(w_i\mathbf{x}_i'\varepsilon_i)\vartheta_1'(t) + \vartheta_2^2(t)Var(e^{\varepsilon_i}) + 2E[\vartheta_1(t)w_i\mathbf{x}_i'\varepsilon_i\vartheta_2(t)(e^{\varepsilon_i} - E(e^{\varepsilon_i}))] =$$

$$= \sigma_\varepsilon^2\vartheta_1(t)E(\frac{\mathbf{x}_i'\mathbf{x}_i}{G(T_i-)})\vartheta_1'(t) + \sigma_{e^\varepsilon}^2\vartheta_2^2(t) + 2\vartheta_1(t)E[\mathbf{x}_i'\varepsilon_i(e^{\varepsilon_i} - E(e^{\varepsilon_i}))]\vartheta_2(t)$$

where $\sigma_{e^\varepsilon}^2 = Var(e^{\varepsilon_i})$. Obviously the first diagonal component of $\sum_{2x2}(t)$ remains unchanged and the proof is now identical to the proof in the previous section.

2.4 Parametric estimation of the survival distribution for censoring time

2.4.1 Minimal cost data

Suppose costs can potentially be accrued over a fixed time period $[0,\tau]$ with expenditure terminating at some event time T so that complete cost observation occurs if a patient is followed through time $T^* = \min(T,\tau)$. Further accumulation of medical cost is not possible since there is no cost after death. Censoring at time U might preclude observation of T^*. Let $X = \min(T,U)$, $\delta = [T \le U]$, where $[A]$ denoted the indicator function of the displayed event A. The observed cost y is uncensored if, and only if $\delta^* = 1$, where $\delta^* = [T^* \le U]$, and the event occurring at T is observed if, and only if $\delta = 1$.

Suppose that we have a random sample of size n. The variables corresponding to the ith subject are indexed by the subscript i. Let $\{(X_i^*, \delta_i^*) : 1 \le i \le n\}$ denote the random sample with $X_i^* = \min(T_i^*, U_i)$.

Consider a linear model for the cost y_i observed at T_i^* in the ith subject given by $y_i = \mathbf{x}_i\beta + \varepsilon_i$ where $\mathbf{x}_i$ is a vector of $p \times 1$ covariates, ε_i is an unobserved error term. In estimation of β we use the data $\{(y_i, \mathbf{x}_i), i = 1,\ldots,n\}$. However since cost will be

incomplete if $\delta_i^* = 0$, we use the weighted sample $\{(y_i, \mathbf{x}_i, w_i), i = 1, \ldots, n\}$

where $w_i = \dfrac{\delta_i^*}{G(T_i^*)}$. Formally $\hat{\beta}$ is obtained by minimizing (with respect to β).

$$\min \sum_{i=1}^{n} w_i (y_i - \mathbf{x}_i \beta)^2 \tag{2.38}$$

The objective function (2.38) weights each observation by the inverse probability of being uncensored. This gives $\hat{\beta}_w = (\sum_{i=1}^{n} w_i \mathbf{x}_i' \mathbf{x}_i)^{-1} (\sum_{i=1}^{n} w_i \mathbf{x}_i' y_i)$. We also assume censoring to be non-informative in the sense that given fixed time covariates, U is independent of cost and event times. Note that under this assumption $E(w_i \mid \mathbf{x}_i, T_i) = P(\delta_i^* = 1 \mid \mathbf{x}_i, T_i) / G(T_i^* -) = 1$, where $G(t) = P(U > t)$.

The weights w_i are unspecified since G is unknown, however in order to be able to use $\hat{\beta}_w$ we need to have a suitable estimator for G. So far we have suggested its estimation from a Cox proportional hazards model or non-parametric model. Assume that the distribution, G, of the censoring time U_i has a parametric form. We assume that the functional form of G is known except for an unknown q-dimensional parameter θ. Then $P[\delta^* = 1 \mid y, T] = G(T^* -, \theta)$, where $G(t, \theta) = P[U > t \mid \theta]$. Let $g(., \theta)$ and f be the density functions of U and T respectively, and let $S(t) = P(T > t)$ be the survival function of the event time T.

The estimation of θ will be accomplished via maximum likelihood. In order to write down the appropriate likelihood, we must be precise about what is actually observed.

1. Observation scheme I.

We only observe $X = \min(T,U)$ and $\delta = [T \le U]$. This is the usual random censorship model. The contribution of a single observation, (X_i, δ_i) to the likelihood is then

$$l_i(\theta) = \{G(T_i,\theta) f(T_i)\}^{\delta_i} \{S(U_i) g(U_i,\theta)\}^{1-\delta_i}$$

Then keeping only terms involving θ, the relevant part of the log-likelihood for estimation of θ is then

$$\psi_i(\theta) = \delta_i \frac{\nabla_\theta G(T_i,\theta)}{G(T_i,\theta)} + (1-\delta_i)\frac{\nabla_\theta g(U_i,\theta)}{g(U_i,\theta)} \tag{2.39}$$

Then $\hat{\theta}$ can be estimated as a solution to $\sum_{i=1}^{n} \psi_i(\theta) = 0$.

2. Observation scheme II.

Suppose observation does not go past time τ (a fixed time). If T is observed, then necessarily $T \le \tau$ and, of course $T \le U$. If on the other hand U is observed, we must have just the opposite: $U \le \tau$ and $T > U$. It is also possible that neither T nor U are observed, in which case $T > \tau$ and $U > \tau$. The likelihood would now consist of three parts:

(1) $\{G(T) f(T)\}^{\delta[T<\tau]}$

(2) $\{S(U) g(U)\}^{(1-\delta)[U<\tau]}$

(3) $\{G(\tau) S(\tau)\}^{[T\ge\tau, U\ge\tau]}$

The contribution of a single observation to the likelihood is then

$$l_i(\theta) = \{G(T_i,\theta) f(T_i)\}^{\delta_i[T_i<\tau]} \{S(U_i) g(U_i,\theta)\}^{(1-\delta_i)[U_i<\tau]} \{G(\tau,\theta) S(\tau)\}^{[T\ge\tau, U\ge\tau]}$$

The relevant part of the log-likelihood for estimation of θ is now

$$\psi(\theta) = [T < \tau]\delta \frac{\nabla_\theta G(T,\theta)}{G(T,\theta)} + [U < \tau](1-\delta)\frac{\nabla_\theta g(U,\theta)}{g(U,\theta)} + [T \geq \tau, U \geq \tau]\frac{\nabla_\theta G(\tau,\theta)}{G(\tau,\theta)} \tag{2.40}$$

Combining the first and third terms,

$$[T < \tau]\delta \frac{\nabla_\theta G(T,\theta)}{G(T,\theta)} + [T \geq \tau, U \geq \tau]\frac{\nabla_\theta G(\tau,\theta)}{G(\tau,\theta)} =$$

$$= \frac{\nabla_\theta G(T^*,\theta)}{G(T^*,\theta)}\{[T < \tau][T^* < U] + [T \geq \tau][U \geq T^*] = \delta^* \frac{\nabla_\theta G(T^*,\theta)}{G(T^*,\theta)}.$$

Therefore (2.40) can be re-written using (X^*, δ^*) as:

$$\psi(\theta) = \delta^* \frac{\nabla_\theta G(T^*,\theta)}{G(T^*,\theta)} + (1-\delta^*)\frac{\nabla_\theta g(U,\theta)}{g(U,\theta)} \tag{2.41}$$

We will focus on estimation of θ under the observation scheme II.

Estimation of θ

Suppose Θ is a compact set in $\mathbf{R}^q$ and assume that θ_0 is the unique solution of the problem

$$\max_{\theta \in \Theta} Q_0(\theta) = \max_{\theta \in \Theta} E(l(\theta)) \tag{2.42}$$

Under general conditions, the sample analogue of the expression (2.42),

$Q_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} l_i(\theta)$ which we denote by $\hat{\theta}$, is consistent and asymptotically normal , see Newey, McFadden (1994), Theorems 2.5 & 3.3[62] and Wooldridge (2002, 2003)[33, 34].

Hereafter θ_0 is the true underlined parameter and all convergences are under θ_0.

Based upon (2.41), an estimator $\hat{\theta}$ of θ_0 is obtained as a solution to

$\sum_i \psi_i(\theta) = 0$. Now expand $\sum_i \psi_i(\hat{\theta})$ at the true parameter θ_0 to get

$$0 = \sum_i \psi_i(\hat{\theta}) = \sum_i \psi_i(\theta_0) + \sum_i \nabla_\theta \psi_i(\tilde{\theta})(\hat{\theta} - \theta_0) \tag{2.43}$$

where $\tilde{\theta}$ lies between $\hat{\theta}$ and θ_0 and $\nabla_\theta \psi_i$ denotes the $q \times q$ matrix of the derivative (with respect to θ) of $\psi_i(\theta)$. Then under standard assumptions

$n^{-1} \sum_i \nabla_\theta \psi_i(\tilde{\theta}) \xrightarrow{P} E[\nabla_\theta \psi_i(\theta)]$. From (2.43) we get

$$\sqrt{n}(\hat{\theta} - \theta_0) = J^{-1}(\theta_0)\left(\frac{1}{\sqrt{n}} \sum_i \psi_i(\theta_0)\right) + o_p(1) \tag{2.44}$$

where $J(\theta_0) = -E(\nabla_\theta \psi_i(\theta_0))$. To verify that (2.44) would give the asymptotic distribution of $\hat{\theta}$ we will now check that $E(\psi_i(\theta)) = 0$. Examine each term in (2.41):

$$E[\psi(\theta)] = E[\delta^* \frac{\nabla_\theta G(T^*, \theta)}{G(T^*, \theta)} + (1 - \delta^*) \frac{\nabla_\theta g(U, \theta)}{g(U, \theta)}] =$$

$$= E[\frac{\nabla_\theta G(T^*, \theta)}{G(T^*, \theta)} [T \wedge \tau \le U]] + E[\frac{\nabla_\theta g(U, \theta)}{g(U, \theta)} [T \wedge \tau > U]]$$

The first term is $E[\frac{\nabla_\theta G(T^*, \theta)}{G(T^*, \theta)} [T \wedge \tau \le U]] =$

$$= E[\frac{\nabla_\theta G(T, \theta)}{G(T, \theta)} [T \le \tau][T \le U]] + E[\frac{\nabla_\theta G(\tau, \theta)}{G(\tau, \theta)} [\tau \le T][\tau \le U]] =$$

$$= E[\nabla_\theta G(T, \theta)[T \le \tau]] + \nabla_\theta G(\tau, \theta) S(\tau) =$$

$$= -\nabla_\theta \int_0^\tau G(u, \theta) dS(u) + \nabla_\theta G(\tau, \theta) S(\tau). \tag{2.45}$$

The second term is

$$E[\frac{\nabla_\theta g(U,\theta)}{g(U,\theta)}[T \wedge \tau > U]] = E[\frac{\nabla_\theta g(U,\theta)}{g(U,\theta)} S(U)[\tau > U]] =$$

$$= -\nabla_\theta \int_0^\tau S(u) dG(u,\theta). \tag{2.46}$$

Combining (2.45) and (2.46) we will get

$$E\{\psi(\theta)\} = -\nabla_\theta \int_0^\tau G(u,\theta) dS(u) + \nabla_\theta G(\tau,\theta) S(\tau) - \nabla_\theta \int_0^\tau S(u) dG(u,\theta) =$$

$$= -\nabla_\theta G(u,\theta) S(u)\Big|_0^\tau + \nabla_\theta G(\tau,\theta) S(\tau) = 0.$$

An application of the Central Limit Theorem to (2.44) gives

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{\mathcal{D}} N(0, J^{-1}(\theta_0)).$$

where $J^{-1}(\theta_0) E(\psi(\theta_0)\psi'(\theta_0)) J^{-1}(\theta_0) = J^{-1}(\theta_0) J(\theta_0) J^{-1}(\theta_0) = J^{-1}(\theta_0)$. We can replace θ_0 by the consistent estimator $\hat{\theta}$, and use $\hat{p}_i = P[U_i \geq T_i^* \mid T_i] = G(T_i^*, \hat{\theta})$ for the unknown $p_i = P[U_i \geq T_i^* \mid T_i]$. For this adjusted estimator we have

$$\hat{\beta}_{wp} = \left(\sum_{i=1}^n \hat{w}_i \mathbf{x}_i' \mathbf{x}_i \right)^{-1} \left(\sum_{i=1}^n \hat{w}_i \mathbf{x}_i' y_i \right)$$

which is a two-step estimator, first step would be estimating $\hat{\theta}$, and then estimating β in (2.38) after we have replaced $w_i = \delta_i^* / P[U_i \geq T_i^* \mid T_i]$ by $\hat{w}_i = \delta_i^* / G(T_i^*, \hat{\theta})$. In other words the two-step estimator solves the problem $\min_{\beta \in \Theta} \sum_{i=1}^n \hat{w}_i (y_i - \mathbf{x}_i \beta)^2$.

Therefore

$$\sqrt{n}(\hat{\beta}_{wp} - \beta) = \left(\frac{1}{n} \sum_{i=1}^n \hat{w}_i \mathbf{x}_i' \mathbf{x}_i \right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{w}_i \mathbf{x}_i' \varepsilon_i \right) \tag{2.47}$$

Following closely Chapter 12.4 of Wooldridge (2001)[63], we expand the second term on the right hand side of the equation above to get

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i^*\mathbf{x}_i'\varepsilon_i}{\hat{p}_i}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i^*\mathbf{x}_i'\varepsilon_i}{p_i}-\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i^*\mathbf{x}_i'\varepsilon_i}{(G(T_i^*,\tilde{\theta}))^2}\nabla_\theta G(T_i^*,\tilde{\theta})\right)\sqrt{n}(\hat{\theta}-\theta_0)$$

where $\tilde{\theta}$ is between $\hat{\theta}$ and θ_0. Again use the standard arguments to claim

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i^*\mathbf{x}_i'\varepsilon_i}{(G(T_i^*,\tilde{\theta}))^2}(\nabla_\theta G(T_i^*,\tilde{\theta})')\xrightarrow{P}E\left(\frac{\delta_i^*\mathbf{x}_i'\varepsilon_i}{(G(T_i^*,\theta_0))^2}(\nabla_\theta G(T_i^*,\theta_0))'\right)=$$

$$=E\left(\frac{\mathbf{x}_i'\varepsilon_i}{G(T_i^*,\theta_0)}(\nabla_\theta G(T_i^*,\theta_0))'\right).$$

Denote this limit by $D(\theta_0)$, a $K\times q$ matrix.

Combining these results we have the established the expansion

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\delta_i^*\mathbf{x}_i'\varepsilon_i}{\hat{p}_i}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\frac{\delta_i^*\mathbf{x}_i'\varepsilon_i}{p_i}-D(\theta_0)J^{-1}(\theta_0)\psi_i(\theta_0)\}+o_p(1)$$

Let $\mathbf{k}_i=\frac{\delta_i^*\mathbf{x}_i\varepsilon_i}{G(T_i^*,\theta_0)}=\frac{\delta_i^*\mathbf{x}_i\varepsilon_i}{p_i}$. Then

$$\psi_i(\theta_0)\mathbf{k}_i'=\delta_i^*\frac{\nabla_\theta G(T_i^*,\theta_0)}{G(T_i^*,\theta_0)}\frac{\mathbf{x}_i'\varepsilon_i}{G(T_i,\theta_0)}=$$

$$=\delta_i[T_i<\tau]\frac{\nabla_\theta G(T_i,\theta_0)}{G(T_i,\theta_0)}\frac{\mathbf{x}_i'\varepsilon_i}{G(T_i,\theta_0)}+\delta_i[T_i\wedge U_i\geq\tau]\frac{\nabla_\theta G(\tau,\theta_0)}{G(\tau,\theta_0)}\frac{\mathbf{x}_i'\varepsilon_i}{G(\tau,\theta_0)}.$$

Taking the expectation yields,

$$E(\psi_i(\theta_0)\mathbf{k}_i')=E\left([T_i<\tau]\frac{\nabla_\theta G(T_i,\theta_0)}{G(T_i,\theta_0)}\mathbf{x}_i'\varepsilon_i+[T_i\geq\tau]\frac{\nabla_\theta G(\tau,\theta_0)}{G(\tau,\theta_0)}\mathbf{x}_i'\varepsilon_i\right).$$

Observe that is the same as $D'(\theta_0)=E\left(\frac{(\nabla_\theta G(T_i^*,\theta_0))}{G(T_i^*,\theta_0)}\mathbf{x}_i'\varepsilon_i\right)$.

Then the asymptotic variance V of $\sqrt{n}(\hat{\beta}_{wp}-\beta)$ is

$$V=(E(\mathbf{x}_i'\mathbf{x}_i)^{-1})E\left(\mathbf{k}_i-D(\theta_0)J^{-1}(\theta_0)\psi_i(\theta_0)\right)\left(\mathbf{k}_i-D(\theta_0)J^{-1}(\theta_0)\psi_i(\theta_0)\right)'(E(\mathbf{x}_i'\mathbf{x}_i)^{-1})$$

$$=(E(\mathbf{x}_i'\mathbf{x}_i)^{-1})\{E(\mathbf{k}_i\mathbf{k}_i')+D(\theta_0)J^{-1}(\theta_0)E\left(\psi_i(\theta_0)\psi_i'(\theta_0)\right)J^{-1}(\theta_0)D'(\theta_0)(E(\mathbf{x}_i'\mathbf{x}_i)^{-1})$$

$$-D(\theta_0)J^{-1}(\theta_0)E\left(\psi_i(\theta_0)\mathbf{k}_i'\right)-E\left(\mathbf{k}_i\psi_i'(\theta_0)\right)J^{-1}(\theta_0)D'(\theta_0)\}.$$

Notice that $E\left(\psi_i(\theta_0)\psi_i'(\theta_0)\right)=J(\theta_0)$ which makes the second term above

$D(\theta_0)J^{-1}(\theta_0)D'(\theta_0)$.

So the variance V is

$$V=(E(\mathbf{x}_i'\mathbf{x}_i)^{-1})\{E(\mathbf{k}_i\mathbf{k}_i')-D(\theta_0)J^{-1}(\theta_0)D'(\theta_0)\}(E(\mathbf{x}_i'\mathbf{x}_i)^{-1}).$$

Under the assumption $E\left(\varepsilon_i\mid\mathbf{x}_i\right)=0$, $D'(\theta_0)=E\left(\frac{(\nabla_\theta G(T_i^*,\theta_0))}{G(T_i^*,\theta_0)}\mathbf{x}_i'\varepsilon_i\right)=0$ and

$V=(E(\mathbf{x}_i'\mathbf{x}_i)^{-1})E(\mathbf{k}_i\mathbf{k}_i')(E(\mathbf{x}_i'\mathbf{x}_i)^{-1})$. In general the asymptotic variance of $\sqrt{n}(\hat{\beta}_{wp}-\beta)$ is smaller when we use an estimate $\hat{\theta}$ for θ_0 in the expression of $\hat{\beta}_{wp}$ as in (2.47).

2.4.2 Interval cost data

Suppose that the interval $[0,\tau)$ is divided into G intervals $[0,\tau)=\bigcup_g[a_{g-1},a_g)$ as in Lin (2000), and Willan *et al* (2002, 2003)[2-4] with $0=a_0<a_1<\ldots<a_G=\tau$. For the interval cost data first define $T_{ig}^*=T_i\wedge a_g$ and $\delta_{ig}^*=[T_{ig}^*\le U_i]$, also denote

$T_i^* = T_{iG}^* = T_i \wedge a_G = T_i \wedge \tau$. For the ith patient we denote y_{ig} the cost in gth interval at time T_{ig}^*. We will have two different censoring types

A. Time censoring: $\delta_i^* = [T_i^* \le U_i] = 0$ if $T_i^* > U_i$, and

B. Cost censoring: $\delta_{ig}^* = [T_{ig}^* \le U_i] = 0$ if $T_{ig}^* > U_i$.

Consider the model

$$y_{ig} = \mathbf{x}_{ig}\beta + u_i$$

for all $g = 1,\ldots,G$ or $\mathbf{y}_i = \mathbf{X}_i\beta + \mathbf{u}_i$ where

$\mathbf{y}_i = \begin{pmatrix} y_{i1} \\ \cdot \\ y_{iG} \end{pmatrix}$, $\mathbf{u}_i = \begin{pmatrix} u_{i1} \\ \cdot \\ u_{iG} \end{pmatrix}$ and $\mathbf{X}_i = \begin{pmatrix} \mathbf{x}_{i1} \\ \cdot \\ \mathbf{x}_{iG} \end{pmatrix}$. Here $\mathbf{y}_i$, $\mathbf{u}_i$ are $G\times 1$ and $\mathbf{X}_i$ is $G\times p$.

The following are assumed to be true:

AG1 $E(\mathbf{X}_i'\mathbf{u}_i) = 0$

AG2 *rank* $E(\mathbf{X}_i'\mathbf{X}_i) = p$

The IPW-POLS estimator of β ($p\times 1$ vector) in the model above is given by

$$\hat{\beta}_{wp} = (\sum_{i=1}^{n} \hat{\mathbf{X}}_i'\hat{\mathbf{X}}_i)^{-1}(\sum_{i=1}^{n} \hat{\mathbf{X}}_i'\hat{\mathbf{y}}_i) \qquad (2.48)$$

where $\hat{\mathbf{y}}_i = \mathbf{S}_i\mathbf{P}_i^{-1}\mathbf{y}_i$, and $\hat{\mathbf{X}}_i$, $\hat{\mathbf{u}}_i$ are similarly defined. Here $\mathbf{S}_i$ is the diagonal matrix with elements $\delta_{ig}^* = [T_{ig}^* \le U_i]$, $g = 1,\ldots,G$ in the main diagonal (will allow costs to be zero in some intervals), and $\mathbf{P}_i$ is the diagonal matrix with elements, $\sqrt{p_{ig}}$ in the main diagonal, where $p_{ig} = P[U_i \ge T_{ig}^* \mid T_i]$. Under the assumptions above we have

$P[\delta_{ig}^* = 1 \mid \mathbf{x}_{ig}, y_{ig}, T_i] = P[U_i \ge T_{ig}^* \mid T_i] = p_{ig}$.

Assume the distribution of the censoring time U_i has a parametric form, $P[U_i > t] = p(t,\theta)$, where the function p is known except for the unknown q-dimensional parameter θ. Then $p_{ig} = p(T_{ig}^* -,\theta)$. Now replace θ by a consistent estimator $\hat{\theta}$, and use $\hat{p}_{ig} = p(T_{ig}^* -,\hat{\theta})$ in (2.48) for the unknown p_{ig}. For this adjusted estimator we have

$$\sqrt{n}(\hat{\beta}_{wp} - \beta) = (\frac{1}{n}\sum_{i=1}^{n}\sum_{g=1}^{G}\frac{\delta_{ig}^* \mathbf{x}_{ig}\mathbf{x}_{ig}'}{\hat{p}_{ig}})^{-1}(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{g=1}^{G}\frac{\delta_{ig}^* \mathbf{x}_{ig}u_{ig}}{\hat{p}_{ig}}) \qquad (2.49)$$

Estimation of θ

The observable data on the censoring times is restricted to $X_i^* = \min(T_i^*, U_i)$ and $\delta_i^* = [T_i^* \le U_i]$. Here $a_G \equiv \tau$ is the upper limit of observation. For a patient i, $\{X_i^* = \tau\} = \{T_i \ge \tau, U_i \ge \tau\}$. This event has probability $S(\tau)p(\tau,\theta)$, assuming both S and p are continuous. Now consider our adopted panel framework in which the observed X_i^* falls into some interval $[a_{g-1}, a_g)$, $g = 1,\ldots,G$. The part of the likelihood of (X_i^*, δ_i^*) that is relevant for estimation of θ has the form $\{p(X_i^*,\theta)\}^{\delta_i^*}\{g(X_i^*,\theta)\}^{1-\delta_i^*}$ where $g(t,\theta)$ is a density for U_i. We will assume that $p(a_G,\theta) > 0$ and that $\theta \to g(\cdot,\theta)$ fulfills all regularity conditions needed for maximum likelihood estimation of θ.

Note that $X_i^* \in [a_{g-1}, a_g)$ and $\delta_i^* = 1$ is equivalent to $[U_i \ge T_{ig}^*][a_{g-1} \le T_{ig}^* < a_g]$ $= \delta_{ig}^* I_g(T_{ig}^*) = 1$, whereas $X_i^* \in [a_{g-1}, a_g)$ and $\delta_i^* = 0$ is equivalent to $[a_g \le U_i \le T_{ig}^*]$ $= (1-\delta_{ig}^*)I_g(U_i) = 1$, where $I_g(t) = [a_{g-1} \le t < a_g]$. To include the interval $t \ge a_G$,

define the indicator $I_G(t) = [t \geq a_G]$. Then the derivative with respect to θ of the aforementioned log-likelihood can be written

$$\sum_{g=1}^{G}\{\delta_{ig}^{*} I_g(T_{ig}^{*})\frac{\nabla_\theta p(T_{ig}^{*},\theta)}{p(T_{ig}^{*},\theta)} + (1-\delta_{ig}^{*})I_g(U_i)\frac{\nabla_\theta g(U_i,\theta)}{g(U_i,\theta)} +$$

$$+(1/G)I_{G+1}(T_i \wedge U_i)\frac{\nabla_\theta p(\tau,\theta)}{p(\tau,\theta)}\} = \sum_{g=1}^{G} d_{ig}(\theta).$$

The estimator $\hat{\theta}$ is a solution to $\sum_{i=1}^{n}\sum_{g=1}^{G} d_{ig}(\theta) = 0$. Consistency of $\hat{\theta}$ follows from the standard regularity conditions on the functions $\theta \to g(\cdot,\theta)$ and $\theta \to p(\cdot,\theta)$ for maximum likelihood estimation of θ. In the sequel we will call θ_0 the true parameter.

Note that $d_{ig}(\theta_0)$ is a $q \times 1$ vector. Using a Taylor expansion of $\sum_{i=1}^{n}\sum_{g=1}^{G} d_{ig}(\hat{\theta})$ at θ_0 yields

$$\sqrt{n}(\hat{\theta} - \theta_0) = -(\frac{1}{n}\sum_{i=1}^{n}\sum_{g=1}^{G}\dot{d}_{ig}(\tilde{\theta}))^{-1}(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{g=1}^{G} d_{ig}(\theta_0)),$$

where $\tilde{\theta}$ is between $\hat{\theta}$ and the true θ_0. Also, $\dot{d}_{ig}(\theta)$ is the derivative of $d_{ig}(\theta)$ with respect to θ. Note that $\dot{d}_{ig}(\theta)$ is a $q \times q$ matrix. Now use the standard arguments to claim

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{g=1}^{G}\dot{d}_{ig}(\tilde{\theta}) \to E(\sum_{g=1}^{G}\dot{d}_{ig}(\theta_0)) \text{ in probability, and } \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{g=1}^{G} d_{ig}(\theta_0) = O_p(1).$$

The last claim is simply a consequence of the central limit theorem.

Now $E(\sum_{g=1}^{G}\dot{d}_{ig}(\theta_0)) = -E((\sum_{g=1}^{G} d_{ig}(\theta_0))(\sum_{g=1}^{G} d_{ig}'(\theta_0))) = J(\theta_0)$. Hence we have

$$\sqrt{n}(\hat{\theta}-\theta_0)=(J(\theta_0))^{-1}(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{g=1}^{G}d_{ig}(\theta_0))+o_p(1).$$

We can expand the sum $\sum_{i=1}^{n}\sum_{g=1}^{G}\frac{\delta_{ig}^{*}\mathbf{x}_{ig}u_{ig}}{\hat{p}_{ig}}$ to get

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{g=1}^{G}\frac{\delta_{ig}^{*}\mathbf{x}_{ig}u_{ig}}{\hat{p}_{ig}}=$$

$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{g=1}^{G}\frac{\delta_{ig}^{*}\mathbf{x}_{ig}u_{ig}}{p_{ig}}-(\frac{1}{n}\sum_{i=1}^{n}\sum_{g=1}^{G}\frac{\delta_{ig}^{*}\mathbf{x}_{ig}u_{ig}}{(p(T_{ig}^{*},\tilde{\theta}))^{2}}(\nabla_{\theta}p(T_{ig}^{*},\tilde{\theta}))')\sqrt{n}(\hat{\theta}-\theta_0)$$

where $\tilde{\theta}$ is between $\hat{\theta}$ and θ_0. Again use the standard arguments to claim

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{g=1}^{G}\frac{\delta_{ig}^{*}\mathbf{x}_{ig}u_{ig}}{(p(T_{ig}^{*},\tilde{\theta}))^{2}}(\nabla_{\theta}p(T_{ig}^{*},\tilde{\theta}))'\rightarrow E(\sum_{g=1}^{G}\frac{\delta_{ig}^{*}\mathbf{x}_{ig}u_{ig}}{(p(T_{ig}^{*},\theta_0))^{2}}(\nabla_{\theta}p(T_{ig}^{*},\theta_0))')$$

in probability. Denote this limit by $D(\theta_0)$, a $K\times q$ matrix. Note that $D(\theta_0)$ is also the same as $E(\sum_{g=1}^{G}\frac{\mathbf{x}_{ig}u_{ig}}{p(T_{ig}^{*},\theta_0)}(\nabla_{\theta}p(T_{ig}^{*},\theta_0))')$.

Combining these results we have the established the expansion

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{g=1}^{G}\frac{\delta_{ig}^{*}\mathbf{x}_{ig}u_{ig}}{\hat{p}_{ig}}=n^{-1/2}\sum_{i=1}^{n}\sum_{g=1}^{G}\{\frac{\delta_{ig}^{*}\mathbf{x}_{ig}u_{ig}}{p_{ig}}-D(\theta_0)J^{-1}(\theta_0)d_{ig}(\theta_0)\}+o_p(1)=$$

$$=n^{-1/2}\sum_{i=1}^{n}\left\{\mathbf{X}_i'\mathbf{S}_i\mathbf{P}_i^{-1}\mathbf{P}_i^{-1}\mathbf{u}_i-D(\theta_0)J^{-1}(\theta_0)\mathbf{d}_i(\theta_0)\mathbf{j}_G\right\}+o_p(1),$$

where $\mathbf{d}_i(\theta_0)=[d_{i1}(\theta_0),\ldots,d_{iG}(\theta_0)]$ is a $q\times G$ matrix, and $\mathbf{j}_G$ is a $G\times 1$ vector of 1's. Now apply the central limit theorem to show that it converges in distribution to a G-variate normal, mean vector zero and variance matrix V.

The assumption **AG1**: $E(\mathbf{X}_i'\mathbf{u}_i)=0$ means that $E(\sum_{g=1}^{G}\mathbf{x}_{ig}u_{ig})=0$. If

$E(\mathbf{x}_{ig}u_{ig})=0$ or, more strongly, $E(u_{ig}\mid\mathbf{x}_{ig})=0$ for all g, assumption **AG1** holds.

Without additional assumptions on u_{ig} we can not conclude that $D(\theta_0)=0$. Suppose we impose $E(u_{ig}\mid\mathbf{x}_{ig},T_i)=0$ for all $g=1,\ldots,G$. Then

$$E(\sum_{g=1}^{G}\frac{\delta_{ig}^{*}\mathbf{x}_{ig}u_{ig}}{(p(T_{ig}^{*},\theta_0))^2}(\nabla_\theta p(T_{ig}^{*},\theta_0))')=\sum_{g=1}^{G}E(u_{ig}\mid\mathbf{x}_{ig},T_i)\frac{\mathbf{x}_{ig}(\nabla_\theta p(T_{ig}^{*},\theta_0))'}{(p(T_{ig}^{*},\theta_0))^2}=0.$$

Computing *V*

Let $\mathbf{k}_i=\mathbf{X}_i'\mathbf{S}_i\mathbf{P}_i^{-1}\mathbf{P}_i^{-1}\mathbf{u}_i=\sum_{g=1}^{G}\frac{\delta_{ig}^{*}\mathbf{x}_{ig}u_{ig}}{p(T_{ig}^{*},\theta_0)}$. Then the aforementioned variance V is

$$V=E\left(\mathbf{k}_i-D(\theta_0)J^{-1}(\theta_0)\mathbf{d}_i(\theta_0)\mathbf{j}_G\right)\left(\mathbf{k}_i-D(\theta_0)J^{-1}(\theta_0)\mathbf{d}_i(\theta_0)\mathbf{j}_G\right)'=$$

$$=E(\mathbf{k}_i\mathbf{k}_i')+D(\theta_0)J^{-1}(\theta_0)E\left(\mathbf{d}_i(\theta_0)\mathbf{j}_G\mathbf{j}_G'\mathbf{d}_i'(\theta_0)\right)J^{-1}(\theta_0)D'(\theta_0)-$$

$$-D(\theta_0)J^{-1}(\theta_0)E\left(\mathbf{d}_i(\theta_0)\mathbf{j}_G\mathbf{k}_i'\right)-E\left(\mathbf{k}_i\mathbf{j}_G'\mathbf{d}_i'(\theta_0)\right)J^{-1}(\theta_0)D'(\theta_0)$$

Notice that $E\left(\mathbf{d}_i(\theta_0)\mathbf{j}_G\mathbf{j}_G'\mathbf{d}_i'(\theta_0)\right)=J(\theta_0)$ which makes the second term above $D(\theta_0)J^{-1}(\theta_0)D'(\theta_0)$.

Now consider $E\left(\mathbf{d}_i(\theta_0)\mathbf{j}_G\mathbf{k}_i'\right)=E((\sum_{g=1}^{G}d_{ig}(\theta_0))(\sum_{g=1}^{G}\frac{\delta_{ig}^{*}\mathbf{x}_{ig}'u_{ig}}{p(T_{ig}^{*},\theta_0)}))$. We will prove that this would reduce to $D(\theta_0)=E(\sum_{g=1}^{G}\frac{\mathbf{x}_{ig}u_{ig}}{p(T_{ig}^{*},\theta_0)}(\nabla_\theta p(T_{ig}^{*},\theta_0))')$ so that V can be expressed as $E(\mathbf{k}_i\mathbf{k}_i')-D(\theta_0)J^{-1}(\theta_0)D'(\theta_0)$, a similar formula to case G=1.

Recall that

$$d_{ig}(\theta_0) = \delta^*_{ig} I_g(T_i) \frac{\nabla_\theta p(T_i,\theta_0)}{p(T_i,\theta_0)} + (1-\delta^*_{ig}) I_g(U_i) \frac{\nabla_\theta g(U_i,\theta_0)}{g(U_i,\theta_0)} +$$

$$+(1/G) I_{G+1}(T_i \wedge U_i) \frac{\nabla_\theta p(\tau,\theta_0)}{p(\tau,\theta_0)},$$

and let $k_{ig} = \frac{\delta^*_{ig} \mathbf{x}_{ig} u_{ig}}{p(T^*_{ig},\theta_0)}$. To prove the equality

$$E((\sum_{g=1}^{G} d_{ig}(\theta_0))(\sum_{g=1}^{G} \frac{\delta^*_{ig} \mathbf{x}'_{ig} u_{ig}}{p(T^*_{ig},\theta_0)})) = E(\sum_{g=1}^{G} \frac{\nabla_\theta p(T^*_{ig},\theta_0)}{p(T^*_{ig},\theta_0)} \mathbf{x}'_{ig} u_{ig})$$

we first prove that for any $h = 1,\ldots,G$ we have

$$E((\sum_{g=1}^{G} d_{ig}(\theta_0)) \frac{\delta^*_{ih} \mathbf{x}'_{ih} u_{ih}}{p(T^*_{ih},\theta_0)}) = E(\frac{\nabla_\theta p(T^*_{ih},\theta_0)}{p(T^*_{ih},\theta_0)} \mathbf{x}'_{ih} u_{ih}).$$

We prove the equality above for h=1. If h>1 the proof is the same. So,

$$A = E((\sum_{g=1}^{G} d_{ig}(\theta_0)) \frac{\delta^*_{i1} \mathbf{x}'_{i1} u_{i1}}{p(T^*_{i1},\theta_0)}) = = E(\{\sum_{g=1}^{n} \delta^*_{ig} I_g(T_i) \frac{\nabla_\theta p(T_i,\theta_0)}{p(T_i,\theta_0)} \frac{\delta^*_{i1} \mathbf{x}'_{i1} u_{i1}}{p(T^*_{i1},\theta_0)}\}) +$$

$$+E(\sum_{g=1}^{n} (1-\delta^*_{ig}) I_g(U_i) \frac{\nabla_\theta g(U_i,\theta_0)}{g(U_i,\theta_0)} \frac{\delta^*_{i1} \mathbf{x}'_{i1} u_{i1}}{p(T^*_{i1},\theta_0)}) +$$

$$+E(\sum_{g=1}^{n} (1/G) I_{G+1}(T_i \wedge U_i) \frac{\nabla_\theta p(\tau,\theta_0)}{p(\tau,\theta_0)} \frac{\delta^*_{i1} \mathbf{x}'_{i1} u_{i1}}{p(T^*_{i1},\theta_0)}).$$

Dropping orthogonal terms,

$$A = E(\{\delta^*_{i1} I_1(T_i) \frac{\nabla_\theta p(T_i,\theta_0)}{p(T_i,\theta_0)} \frac{\delta^*_{i1} \mathbf{x}'_{i1} u_{i1}}{p(T^*_{i1},\theta_0)} + \sum_{g\geq 2} \delta^*_{ig} I_g(T_i) \frac{\nabla_\theta p(T_i,\theta_0)}{p(T_i,\theta_0)} \frac{\delta^*_{i1} \mathbf{x}'_{i1} u_{i1}}{p(T^*_{i1},\theta_0)}\}) +$$

$$+E(\{\sum_{g\geq 2}(1-\delta_{ig}^{*})I_g(U_i)\frac{\nabla_\theta g(U_i,\theta_0)}{g(U_i,\theta_0)}\frac{\delta_{i1}^{*}\mathbf{x}_{i1}'u_{i1}}{p(T_{i1}^{*},\theta_0)}\})+$$

$$E(\{[T_i\wedge U_i\geq\tau]\frac{\nabla_\theta p(\tau,\theta_0)}{p(\tau,\theta_0)}\frac{\delta_{i1}^{*}\mathbf{x}_{i1}'u_{i1}}{p(T_{i1}^{*},\theta_0)}\}).$$

The first term of the sum above is

$$E(\delta_{i1}^{*}I_1(T_i)\frac{\nabla_\theta p(T_i,\theta_0)}{p(T_i,\theta_0)}\frac{\delta_{i1}^{*}\mathbf{x}_{i1}'u_{i1}}{p(T_{i1}^{*},\theta_0)})=E([T_i<U_i]I_1(T_i)\frac{\nabla_\theta p(T_i,\theta_0)}{p^2(T_i,\theta_0)}\mathbf{x}_{i1}'u_{i1})=$$

$$=E(I_1(T_i)\frac{\nabla_\theta p(T_i,\theta_0)}{p(T_i,\theta_0)}\mathbf{x}_{i1}'u_{i1}).$$

The second term is:

$$E(\sum_{g\geq 2}\delta_{i1}^{*}\delta_{ig}^{*}I_g(T_i)\frac{\nabla_\theta p(T_i,\theta_0)}{p(T_i,\theta_0)}\frac{\mathbf{x}_{i1}'u_{i1}}{p(T_{i1}^{*},\theta_0)})=$$

$$=E(\sum_{g\geq 2}[T_i\leq U_i]I_g(T_i)\frac{\nabla_\theta p(T_i,\theta_0)}{p(T_i,\theta_0)}\frac{\mathbf{x}_{i1}'u_{i1}}{p(a_1,\theta_0)})=$$

$$=E(\sum_{g\geq 2}I_g(T_i)\nabla_\theta p(T_i,\theta_0)\frac{\mathbf{x}_{i1}'u_{i1}}{p(a_1,\theta_0)})=E([\tau>T_i\geq a_1]\nabla_\theta p(T_i,\theta_0)\frac{\mathbf{x}_{i1}'u_{i1}}{p(a_1,\theta_0)}).$$

The third term is

$$E(\{\sum_{g\geq 2}\delta_{i1}^{*}(1-\delta_{ig}^{*})I_g(U_i)\frac{\nabla_\theta g(U_i,\theta_0)}{g(U_i,\theta_0)}\frac{\mathbf{x}_{i1}'u_{i1}}{p(T_{i1}^{*},\theta_0)}\})=$$

$$=E(\{\sum_{g\geq 2}[a_g\leq U_i<T_{ig}^{*}]\frac{\nabla_\theta g(U_i,\theta_0)}{g(U_i,\theta_0)}\frac{\mathbf{x}_{i1}'u_{i1}}{p(a_1,\theta_0)}\})=$$

$$=E(\{[a_1\leq U_i<T_i^{*}]\frac{\nabla_\theta g(U_i,\theta_0)}{g(U_i,\theta_0)}\frac{\mathbf{x}_{i1}'u_{i1}}{p(a_2,\theta_0)}\}).$$

The third term becomes

$$E(\{[a_1\leq U_i<T_i^{*}]\frac{\nabla_\theta g(U_i,\theta_0)}{g(U_i,\theta_0)}\frac{\mathbf{x}_{i1}'u_{i1}}{p(a_1,\theta_0)}\})=$$

$$= E((\nabla_\theta \int_{a_1}^{T_i^*} g(u,\theta_0))\frac{\mathbf{x}_{i1}' u_{i1}}{p(a_1,\theta_0)}) = E([T_i^* > a_1](\nabla_\theta p(a_1,\theta_0) - \nabla_\theta p(T_i^*,\theta_0))\frac{\mathbf{x}_{i1}' u_{i1}}{p(a_1,\theta_0)}).$$

The last term is

$$E(\{\delta_{i1}^*[T_i \wedge U_i \geq \tau]\frac{\nabla_\theta p(\tau,\theta_0)}{p(\tau,\theta_0)}\frac{\mathbf{x}_{i1}' u_{i1}}{p(T_{i1}^*,\theta_0)}\}) = E(\{[T_i \geq \tau]\nabla_\theta p(\tau,\theta_0)\frac{\mathbf{x}_{i1}' u_{i1}}{p(a_1,\theta_0)}\})$$

Combining the above formulas we get

$$E((\sum_{g=1}^{G} d_{ig}(\theta_0))\frac{\delta_{i1}^* \mathbf{x}_{i1}' u_{i1}}{p(T_{i1}^*,\theta_0)}) = E(I_1(T_i)\frac{\nabla_\theta p(T_i,\theta_0)}{p(T_i,\theta_0)}\mathbf{x}_{i1}' u_{i1}) +$$

$$+ E([\tau > T_i \geq a_2]\nabla_\theta p(T_i,\theta_0)\frac{\mathbf{x}_{i1}' u_{i1}}{p(a_1,\theta_0)}) +$$

$$+ E([T_i^* > a_1](\nabla_\theta p(a_1,\theta_0) - \nabla_\theta p(T_i^*,\theta_0))\frac{\mathbf{x}_{i1}' u_{i1}}{p(a_1,\theta_0)}) +$$

$$+ E(\{[T_i \geq \tau]\nabla_\theta p(\tau,\theta_0)\frac{\mathbf{x}_{i1}' u_{i1}}{p(a_1,\theta_0)}\}).$$

The second and fourth terms combined give

$$E([\tau > T_i \geq a_1]\nabla_\theta p(T_i,\theta_0)\frac{\mathbf{x}_{i1}' u_{i1}}{p(a_1,\theta_0)}) + E(\{[T_i \geq \tau]\nabla_\theta p(\tau,\theta_0)\frac{\mathbf{x}_{i1}' u_{i1}}{p(a_1,\theta_0)}\}) =$$

$$= E([T_i^* \geq a_1]\nabla_\theta p(T_i^*,\theta_0)\frac{\mathbf{x}_{i1}' u_{i1}}{p(a_1,\theta_0)}).$$

Therefore $E((\sum_{g=1}^{G} d_{ig}(\theta_0))\frac{\delta_{i1}^* \mathbf{x}_{i1}' u_{i1}}{p(T_{i1}^*,\theta_0)}) = E(I_1(T_i)\frac{\nabla_\theta p(T_i,\theta_0)}{p(T_i,\theta_0)}\mathbf{x}_{i1}' u_{i1}) +$

$$+ E([T_i \geq a_2]\nabla_\theta p(a_2,\theta_0)\frac{\mathbf{x}_{i1}' u_{i1}}{p(a_2,\theta_0)}) = E(\frac{\nabla_\theta p(T_{i1}^*,\theta_0)}{p(T_{i1}^*,\theta_0)}\mathbf{x}_{i1}' u_{i1}).$$

In conclusion V can be expressed as $E(\mathbf{k}_i \mathbf{k}_i') - D(\theta_0)J^{-1}(\theta_0)D'(\theta_0)$.

CHAPTER 3

ESTIMATING HOSPITAL COST FOR AMI PATIENTS IN THE NIS 2000

3.1 Description of the data set and background

The Nationwide Inpatient Sample (NIS)

The Nationwide Inpatient Sample (NIS) is the largest all-payer inpatient care database that is publicly available in the United States, containing data from 5 to 8 million hospital stays from about 1000 hospitals sampled to approximate a 20-percent stratified sample of U.S. community hospitals. It is part of the Healthcare Cost and Utilization Project (HCUP), sponsored by the Agency for Healthcare Research and Quality (AHRQ, formerly known as the Agency for Health Care Policy and Research). Currently data is available for a 13-year time period, from 1988 to 2000, allowing analysis of trends over time. Researchers and policymakers use the NIS data to identify, track, and analyze national trends in health care utilization, access, charges, quality, and outcomes.

NIS is the only national hospital database with charge information on all patients, regardless of payer, including persons covered by Medicare, Medicaid, private insurance,

and the uninsured. NIS's large sample size enables analyses of rare conditions, such as congenital anomalies, uncommon treatments, such as organ transplantation, and special patient populations, such as children.

Inpatient stay records in the NIS include clinical and resource use information typically available from discharge abstracts. Hospital and discharge weights are provided for producing national estimates. The NIS can be linked to hospital-level data from the American Hospital Association's Annual Survey of Hospitals and county-level data from the Bureau of Health Professions' Area Resource File, except in those states that do not allow the release of hospital identifiers. Beginning in 1998, the NIS differs from previous NIS releases: some data elements were dropped, some were added, for some data elements the coding was changed, and the sampling and weighting strategy was revised to improve the representativeness of the data.

There is a growing literature on use of the NIS in health services research that we will use for guidance [42-47, 64]. For 2000, the NIS contains over 7.4 million discharges from 28 states. Sixty strata are defined by a combination of region (Northeast, South, Midwest and West), location (urban, rural), ownership (public, private), teaching status, and bed size (small, median, large). The first stage samples approximately 20% of hospitals within each stratum. Then all discharges from the sampled hospitals are included in the database. Patient demographics in the NIS include: age at admission, gender and race.

Total charge and length of stay (LOS) are the main healthcare utilization variables in the NIS for each hospital stay. Discharge status identifies whether it was routine, or resulted in death, or the patient was disposed to other care facilities. The NIS database contains a number of variables describing the hospital stay. Up to 15 diagnoses and 15

procedures are coded based on the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM). They can be used to select samples of hospital discharges for specific diagnoses and procedures. A broader categorization of ICD codes may be used for this selection based on HCUP's Clinical Classification Software (CCS), or by diagnosis-related groups (DRG) that combine information on patient age, sex, diagnoses and procedures accounting for relationships among them and to inpatient resource use.

The main limitation of the NIS is its inability to track individual patients. Each record in the NIS is a separate hospital discharge, and thus we cannot identify multiple admissions by the same individual within a year (or across years). Also, the NIS does not include all preoperative and postoperative treatments rendered in an outpatient setting, although services in ambulatory surgery centers may be included. We will use the NIS to examine length of stay and hospital charges associated with admissions for heart disease. In particular we concentrate on treatments undergone for acute myocardial infarction (AMI).

Coronary heart disease and treatment procedures

The heart is a muscle that works 24 hours a day. To perform well, it needs a constant supply of oxygen and nutrients, which is delivered to the myocardium (heart muscle tissue) by the blood through the coronary arteries. The blood flow to the heart can be reduced by a process called atherosclerosis, in which plaques of fatty substances build up inside the walls of blood vessels. The plaques attract blood components, which stick to the inside surface of the vessel walls. Atherosclerosis can affect any blood vessels and

causes them to narrow their lumina and harden their walls. This process develops over many years and can begin early, even in childhood.

Coronary heart disease (CHD) is the most common form of heart disease, the leading cause of death for Americans. About 12.6 million Americans suffer from CHD, which often results in a heart attack. About 1.1 million Americans suffer a heart attack each year, and about 515,000 of these heart attacks are fatal[65] that is about 2,600 every day; one person every 33 seconds. It takes more lives than cancer-in fact, more than cancer, accidents, and the next five leading causes of death in the United States combined[66].

In CHD, atherosclerosis affects the coronary arteries. The fatty buildup, or plaque, can break open and lead to the formation of a blood clot. The clot covers the site of the rupture, also reducing blood flow. Eventually, the clot becomes firm. The process of fatty buildup, plaque rupture, and clot formation recurs, progressively narrowing the arteries. Ever less blood reaches the heart muscle, and thus fewer quantities of oxygen and nutrients reach the myocardium, leading to ischemia (oxygen starvation of the heart), clinically translated into chest pain. Depending on the degree of pain, the level of physical activity that pain occurs at, and the degree of coronary obstruction, coronary artery diseases can be classified into asymptomatic or mild angina, angina class II-IV or unstable angina, acute myocardial infarction, ischemia after Coronary Artery Bypass Graft (CABG)[67].

Following previous analyses of charges and LOS of AMI patients in the MICH study[48-50], we will focus on patients admitted in the hospital with AMI, a common high-mortality condition whose outcomes are affected by the process of care. We only focus

on patients that underwent as their primary procedure Coronary Artery Bypass Grafting (CABG), Cardiac Catheterization (CATH) or Percutaneous Transluminal Coronary Angioplasty (PTCA) or patients with AMI who underwent no procedures at all.

Cardiac catheterization

In cardiac catheterization (abbreviated "CATH"), a diagnostic procedure in which a very small catheter (hollow tube) is advanced from a blood vessel in the groin through the aorta into the heart. Once the catheter is in place, several diagnostic techniques may be used. The tip of the catheter can be placed into various parts of the heart to measure the pressure within the chambers. The catheter can be advanced into the coronary arteries and a dye injected into the arteries (coronary angiography or arteriography). With the use of fluoroscopy (a special type of X-ray), the physician can tell where any blockages in the coronary arteries are located as the dye moves through the arteries.

Percutaneous transluminal coronary angioplasty

Percutaneous transluminal coronary angioplasty, also known as PTCA, is an established, effective therapy for some patients with coronary artery disease. PTCA is used to dilate (widen) narrowed arteries. A doctor inserts and advances a catheter with a deflated balloon at its tip into the narrowed part of an artery. Then the balloon is inflated, compressing the plaque and enlarging the inner diameter of the blood vessel so blood can flow more easily. Then the balloon is deflated and the catheter removed. PTCA is a less traumatic and less expensive alternative to bypass surgery for some patients with coronary artery disease. In about 40 percent of patients who've had PTCA, the dilated segment of the artery narrows again within six months after the procedure. They may require either another PTCA or coronary artery bypass surgery[68].

Improvements in technologies for PTCA include use of stents and more-recently drug-eluting stents (DESs). A stent is a surgical stainless steel coil that is inserted into the blocked artery via a catheter. The stent serves as a scaffold, supporting the artery walls, and reducing the risk of the artery re-closing (restenosis) over time. Bare-metal stenting after angioplasty has more durable effects than angioplasty alone, but researchers are still seeking ways to reduce restenosis, which occurs about 20% of the time in the clinical setting. The most promising method to lower restenosis rates appears to be using drug-eluting stents (DESs). Some researchers are hailing DESs as one of the greatest interventional advances in the past decade. But others warn that while DESs appear to offer benefit over bare-metal stents, the hype may be outrunning the science. In his article 'Drug-Eluting Stents Show Promise', Mitka (2004)[69] presents the beneficial effects of DESs, but also talks about hospital concerns that reimbursements for such devices may be too low, leading to financial losses for these institutions. NIS database precedes the widespread use of stents and therefore the analyses that we describe would refer to PTCA.

Coronary artery bypass graft

Also known as "bypass surgery" or CABG, coronary artery bypass graft operation uses a piece of vein taken from the leg, or of an artery taken from the chest or wrist. This piece is attached to the heart artery above and below the narrowed area, thus making a bypass around the blockage. Sometimes, more than one bypass is needed. Bypass surgery may be needed due to various reasons, such as an angioplasty that did not sufficiently widen the blood vessel, or blockages that cannot be reached by, or are too long or hard for, angioplasty. In certain cases, bypass surgery may be preferred to angioplasty. For

instance, it may be used for persons who have both CHD and diabetes. A bypass also can close after a period of time. This happens in about 10 percent of bypass surgeries, usually after 10 or more years.

Variations in hospital costs/charges

As suppliers of the most expensive type of health service, hospitals have been particularly vulnerable to the pressures of competition. The health economics literature contains a number of recent studies on hospital costs. A common technique that has been proven to be well-suited to examining the effects of provider institutions on patient costs in a managed care environment is multilevel modeling. This framework is used to analyze data that fall naturally into hierarchical structures consisting of multiple 'micro' units nested within 'macro' units. Multilevel models analyze variability arising at distinct levels within data by extending the more traditional statistical techniques and introducing a degree of realism often absent from single-level models such as multiple regressions. A good description of why charges may differ between facilities is presented in *Health Care Data Report, 2000* (PHC 5320) from Department of Health and Family Services, Division of Health Care Financing, Bureau of Health Information[70].

New technology - The equipment facilities use to provide services differs in age, sophistication, and utilization. Facilities with the latest technology may have higher charges than those with older, less sophisticated equipment.

Staffing costs - Salary scales may differ regionally and are typically higher in urban than rural areas. Furthermore, competition for nurses and other skilled personnel may result in higher staffing costs and, therefore, higher charges.

Intensity of care - Facilities differ in the severity of illness of patients (i.e., some facilities care for more severely ill patients than others). Patients within the same diagnosis or procedure classification may need very different levels of service and staff.

Efficiency of operation - Facilities vary in the utilization and efficiency of services they provide. Infrequently used services may cost more per patient than services that are used more frequently.

Differences in coding - Facilities vary in their coding systems and personnel, and in the number of billing codes they put on a billing form. The use of additional appropriate codes may result in a patient being assigned to a diagnosis or procedure classification with greater reimbursement or may otherwise justify higher charges. Facilities with better-trained personnel or more sophisticated coding software are more likely to place these additional codes on their billing forms and, therefore, may have higher charges than facilities with less expertise.

Discounts - Facilities negotiate and offer volume discounts to Health Maintenance Organizations (HMOs), Preferred Provider Organizations (PPOs), and other large-volume purchasers of health care services. The number of these organizations has grown considerably in recent years. Full charges are paid for only a very small proportion of patients.

Percentage of government pay – Government payers generally reimburse facilities at rates below their full charges, similar to the discounts offered to commercial payers. Therefore, facilities with a large percentage of patients whose charges are paid either by government programs or discounted commercial payers may report large gaps

between what they bill and what they actually receive. This may result in higher charges, including those for non-discounted patients.

Facility price structures - Some facilities spread the cost of services and equipment over all patients. Others bill the full cost of a service to those patients actually using the service. Furthermore, facilities may provide some services at a loss while allowing other facility operations to subsidize the losses. Any of these practices can result in significantly different charges for a given diagnosis or procedure classification.

Range of services provided - Facilities differ in the range of services they provide to patients. Some may provide the full range of services required for diagnosis and treatment during the stay. Others may stabilize patients and then transfer them to another facility for more specialized or rehabilitative care.

Data-related issues - Facilities differ in the number of cases served, the case-mix and illness severity of patients, and the comparability of patients within a given diagnosis or procedure classification. For example, a single case can greatly affect a facility's average charge if the facility reported only a few cases.

Capital expenses - Facilities differ in the amount of debt and depreciation they must cover in their rate structure. A facility with a heavy debt load, a new building, or a major renovation to amortize may have higher charges than a facility not facing such expenses. Furthermore, facilities may choose to lease or purchase equipment or facilities. The choices made about financing of capital projects may affect charges in different ways.

Rice *et al* (1997)[51], Carey (2000, 2002) [52, 53] and Goldstein (2002)[54] insist on the usefulness of multilevel methods in studies where data on cost are collected over multiple

sites (hospitals in our data). This chapter takes a multilevel modeling approach to estimate costs in patients hospitalized for AMI. Patients nested within hospitals form a natural hierarchical structure suited to analysis that models each level simultaneously. The existence of a non-zero intra-hospital correlation, resulting from the presence of more than one residual term in the model, means that traditional estimation procedures such as OLS, which is used for example in multiple regression, are inapplicable. Application of OLS techniques leads in this case to incorrect inferences, although when the intra-class correlations are small we can expect reasonably good agreement between estimates from the multilevel and the simpler OLS approaches. One can also go one step further and extend our model to include higher levels such as hospitals nested in states.

Correlates of total charges available in the data set

Possible correlates of total charges available in the data set are: procedure, gender, age, number of procedures, hospital characteristics (location, teaching status, bed size, region), length of stay (LOS), Charlson Comorbidity Index (CCI).

Procedure: Clinical Classifications Software (CCS), developed by the Agency for Healthcare Research and Quality (AHRQ), is a tool for clustering patient diagnoses and procedures into a manageable number of clinically meaningful categories. CCS is used for grouping conditions and procedures without having to wade through thousands of codes. This "clinical grouper" makes it easier to quickly understand patterns of diagnoses and procedures so that health plans, policymakers, and researchers can analyze costs, utilization, and outcomes associated with particular illnesses and procedures.

CCS collapses diagnosis and procedure codes from the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM), which contains over 12,000 diagnosis codes and 3,500 procedure codes. Without CCS, the large number of ICD-9-CM codes poses difficulties in statistical analysis and reporting.

CCS consists of two related classification systems, single level and multi-level, which are designed to meet different needs. The multi-level CCS groups single-level CCS categories into broader body systems or condition categories (e.g., "Diseases of the Circulatory System," "Mental Disorders," and "Injury"). *Multi-level* CCS is most useful when evaluating larger aggregations of conditions and procedures or exploring them in greater detail. *Single-level* CCS is most useful for ranking of diagnoses and procedures. The single-level diagnosis CCS aggregates illnesses and conditions into 259 mutually exclusive categories. We consider discharges that have either CABG (CCS=44), CATH (CCS=47), PTCA (CCS=45) or no procedure (CCS=.) as a primary procedure.

Demographics Variables: We use gender and age of the patient. Patient age is recorded as age in years at admission. Other demographic variables in the NIS are race and household income by zip code, but they are not uniformly recorded for all states.

Number of procedures (NPR): NPR counts the number of ICD-9-CM procedures coded on the discharge record. The principal procedure is included in this count. A value of 0 means that the patient underwent no procedures on record, a value of 1 means that only the primary procedure is recorded, secondary procedures are left blank, etc. A maximum of 15 procedures have been retained on a NIS inpatient record. States

that provide fewer than 15 procedures have had the procedure vector padded with blank values. For example, if a state supplied 5 procedures, PR6 through PR15 are blank (" ") on all records from that state. States that provide more than 15 procedures may have information truncated. All states have provided at least 6 procedures.

CCI (Charlson Comorbidity Index[71]) is used to assess comorbidity. CCI is a weighted sum of the presence of 15 specified medical conditions at admission. There are two ICD-9-CM adaptations, Deyo (1992)[72] and Dartmouth-Manitoba, Romano *et al* (1993)[73] of the Charlson comorbidity index, as well as other various searches for improved clinical comorbidity indices[74-76]. We use the Dartmouth-Manitoba version of the index. The conditions and associated weights (shown in parentheses) are:

CHF = 'Congestive Heart Failure' (1)

PVD = 'Peripheral Vascular Disease' (1)

CVD = 'Cerebrovascular Disease' (1)

DEM = 'Dementia' (1)

COPD = 'Chronic Obstructive Pulmonary Disease' (1)

ULCD = 'Ulcer Disease' (1)

MLIVD = 'Mild Liver Disease' (1)

DIAB = 'Diabetes without Complication' (1)

HEPL = 'Hemiplegia' (2)

REND = 'Renal Disease' (2)

DIABCC = 'Diabetes with Complications' (2)

MALIG = 'Any Malignancy' (2)

SLIVD = 'Moderate or Severe Liver Disease' (3)

CANCER = 'Metastatic Solid Tumor' (6)

AIDS = 'AIDS' (6)

NIS_STRATUM is a four-digit stratum identifier used to post-stratify hospitals for the calculation of universe and frame weights. The hospital's census region, ownership/control, location/teaching, and bed size were obtained from the American Hospital Association (AHA) Annual Survey of Hospitals.

Region: States were grouped in 4 regions

Northeast (CT, MA, ME, NY, NJ, MA)

Midwest (IA, IL, KS, MO, WI)

South (FL, GA, KY, MD, NC, SC, TN, TX, VA, WV)

West (AZ, CA, CO, HI, OR, UT, WA)

Location/ Teaching: A metropolitan statistical area is considered urban, and a non-metro statistical area is rural. Teaching hospitals have an AMA-approved residency program, are a member of the Council of Teaching Hospitals (COTH) or have a ratio of full-time equivalent interns and residents to beds of .25 or higher. All hospital in the rural area were classified as non-teaching.

Bedsize categorizes the number of short-term acute beds in a hospital into 'small', 'medium', 'large'. A hospital's bed size category depends upon region, location and teaching status. For example, urban teaching hospitals in the Northeast were classified as having 'large' bed size if the number of hospital beds exceeds 425. In the western region for the same location/teaching status, the 'large' bed size category is defined if the count

exceeds 325. This categorization was created to have approximately 1/3 of hospitals in each bed size category in a given region, location, teaching status combination.

Ownership/control includes categories for government nonfederal (public), private not-for-profit (voluntary) and private investor-owned (proprietary). However when the sample size was sufficiently large, hospitals were stratified as public (Ownership=1), voluntary (Ownership=2) and proprietary (Ownership=3). This stratification was used for southern rural, southern urban non-teaching, and western urban non-teaching hospitals. For smaller strata – the midwest and western rural hospitals –a collapsed stratification of public versus private was used, with the voluntary and proprietary hospitals combined to form a single 'private' category (Ownership=4). For all other combinations of region, location and teaching status, no stratification based on control was advisable given the number of hospitals in these cells (Ownership=0). Although the CONTROL variable was used to define the strata, AHRQ collapsed some categories. In consequence, because of the overlapping categories, this variable will not be used further.

Total Charges

Total charges in NIS data are the amount the hospital charged or billed for the entire hospital stay. They do not necessarily reflect reimbursements or costs. Charge data were present for 98 percent of all discharges. Charges are generally higher than costs. Generally, total charges (TOTCHG) do not include professional fees and non-covered charges. If the source provides total charges with professional fees, then the professional fees are removed from the charge during HCUP processing. In a small number of cases,

professional fees were not be removed from total charges because the data source did not provide the necessary information. Two states, MA and WI, will be excluded from the working data set (see exclusion criterion #8 in the next section), because they may have included professional fees. Emergency department charges incurred prior to admission to the hospital may be included in total charges.

3.2 Creating a working subset data set

From the NIS 2000 Core file of 7.4 million discharges we extracted all records with a primary diagnosis of AMI. This is based on DXCCS1=100 or ICD9-CM codes 410.xx and it yields n=157,263 discharges. All discharges fall within Major Diagnostic Code MDC=5 (Circulatory System). There are 165 possible primary procedures for these discharges. Table 3.1 shows the distribution of primary procedures among discharges, excluding procedures that are present in <1% of discharges. We will consider only 121,264 discharges that have either no procedure, Coronary Artery Bypass Grafting (CABG), diagnostic cardiac catheterization or coronary artheriography (CATH) or Percutanerous Transluminal Coronary Angioplasty (PTCA) as their primary procedure.

Guided by several published analyses from the NIS, the following **exclusion criteria** will be applied (all n's are out of the 121264 patients):

1. We exclude admissions for AMI that were not the first episode of care for a newly diagnosed AMI (ICD9-CM code: 410.x1), or if the location of the infarction was unspecified (410.x0)[64] (n=983).
2. We exclude discharges in hospitals that performed ≤5 cases since these cases are most likely coding errors[42, 43, 47] (n= 256).
3. We exclude patients <18 years or >85 years at admission. The pathophysiology of disease in patients <18 years is likely to be different than for adults (n=22). Also patients >85 years are less likely to be treated aggressively than younger patients[43, 47, 64] (n=11,238).
4. Because of concerns that patients with diagnosis code 'AMI' may have included 'rule out' AMIs, all discharges of < 2 days are excluded (n=16,682).
5. Patients with LOS >60 days are eliminated because they are probably long term care patients [52] (n=57).
6. Newborn admission types are also excluded (n=7).
7. For hospital costs within the US health-care system one common approach is to assign a cost to each hospitalization based on the basis of its associated Diagnostic-Related Group[77]. In order to make sure that we only have patients with AMI in the working data set, we exclude patients with the following DRG in effect on discharge date (DRG=104, 105, 113, 119,124, 125,144, 145, 468, 477,478,479,483) (n=1,611).

8. Excluding states: States MA and WI may have included physician fees in the total hospital charges. Some hospitals in TX did not report total charges until July 2000. Because identification of these hospitals was not available, we exclude all discharges from TX, MA, WI (n=20,560).
9. One hospital in Arizona is also excluded because 44.3% of its total charges data is coded C=inconsistent (either excessively low or high) (n=149).
10. One can not link information from the discharging and receiving hospitals. Therefore, if we want to capture LOS and total hospital charge for AMI admissions, then we must restrict to ED admissions and all routine referrals from physicians, clinics, and HMOs (n=23,244).
11. We drop records missing an essential field such as patient age, gender, length of stay, total charges, DRG, admission source or discharge status (n=10,511).
12. We also eliminate patients who left the hospital against medical advice since we do not know whether they received subsequent care (n=703).

The exclusion criteria above are not of course mutual exclusive, a discharge may be excluded for more than one reason. After all exclusions we are left with n=58,469 discharges.

Censoring

The variable DISPUNIFORM indicates the discharge status assigned to each record. We have already excluded missing, invalid records and patients who left the hospital against medical advice. DISPUNIFORM=1 (routine) or DISPUNIFORM=20

(died in hospital) cover a complete hospital episode. In these two cases we will regard TOTCHG and LOS as completely observed. Patients transferred to a short-term hospital, transfers to skilled nursing facilities (SNF), intermediate care, home health care have incomplete TOTCHG and LOS. They will be treated as censored. Of the 58,469 discharges in our working data set 18,753 (32%) of the discharges are censored. Figure 3.1 summarizes the process of creating our data set from the initial 157,263 discharges for AMI.

Characteristics of the patients

Table 3.2 shows the characteristics of the patients. As already mentioned we only consider discharges that have either CABG (CCS code: 44; N=7,369, 12.6%), CATH (CCS code: 47; N=14,264, 24.4%), PTCA (CCS code: 45; N=17,901, 30.6%) or no procedure (CCS code: '.'; N=18,935, 32.4%) as a primary procedure. There are N= 35,977 (61.5%) males and mean age is 65.72 (STD=12.7). Since the higher the number of procedures, the higher the cost, we categorized patients in 3 possible groups: no procedure (N=18,935, 32.4%), 1-4 procedures (16,489, 28.2%), 5 or more procedures (23,045, 39.4%). We will form two comorbidity groups based on CCI score: 0 (no comorbidities, N=23294, 39.8%) and 1 or more (at least one comorbidity). The biggest percent of discharges in the data set comes from the South region (43.1%), followed by Midwest (19.9%), Northeast (19.6%) and West (17.4%). The hospitals are categorized as: rural (13.5%), urban/teaching (42.1%), urban/non-teaching. Hospitals are grouped as: Small (8.9%), Medium (25.5%) or Large.

3.3 Estimation methods

We will follow methods similar to those described in Chapter 2. Total hospital charge (TOTCHG) and length of stay (LOS) are two primary outcome variables. We identify the average total charge over a specified duration τ with the net present value. Given a covariate profile Z_0 and ignoring discounting the average cost restricted to τ is $\mu(\tau) = -\int_0^\tau c_{01}(s \mid Z_0) dS(t \mid Z_0)$. We first describe a multilevel model to estimate $c_{01}(t \mid Z_0)$, including a method for estimating weights. Then we will describe methods to estimate the LOS distribution. Combining these two estimation steps we arrive at an estimate of $\mu(\tau)$.

ESTIMATING $c_{01}(t \mid Z_0)$

A traditional approach for estimating patient costs involves ordinary least squares (OLS) models containing a response variable at the individual level and correlates at both individual and higher levels of analysis. This disregards correlations structures in the data emanating from common influences operating within groups. As argued in the previous section, hospital characteristics, such as quality of service, managerial performance or treatment patterns administered by physicians may impose distinct effects on the costs of treating patients[52, 53].

The working data set consists of 58,469 discharges for AMI patients. Of these, 39,716 are uncensored for total charge and LOS. There are 617 hospitals with discharges between 6 to 1181 per hospital. Because charge data is skewed we consider a log-

transformation to mitigate the effects of skewness of TOTCHG. Figure 3.2 shows the histogram of log-transformed charges in all 58,469 discharges.

3.3.1 Unconditional means model

A one-way random effects model is fitted first. We express the outcome, $Y_{ij} = \log(TOTCHG)$ associated with the i^{th} discharge in the j^{th} hospital, as a linear combination of a grand error mean β_0, a series of deviations from the grand mean u_{0j} and a random error ε_{ij}:

Model 1: $$Y_{ij} = \beta_0 + u_{0j} + \varepsilon_{ij}$$

where $u_{0j} \sim iid\ N(0, \sigma^2_{u0})$, $\varepsilon_{ij} \sim iid\ N(0, \sigma^2_{\varepsilon})$ and u_{0j}, ε_{ij} are assumed independent.

We exclude censored observations from this model by using weights δ_i, where δ_i is the censoring indicator. Note that since u_{0j} and ε_{ij} are assumed independent, we have $Var(Y_{ij}) = Var(\beta_0 + u_{0j} + \varepsilon_{ij}) = \sigma^2_{u0} + \sigma^2_{\varepsilon}$, and for two different discharges $i \neq i'$ within the same hospital $Cov(Y_{ij}, Y_{i'j}) = \sigma^2_{\varepsilon}$.

Using estimates of the variance components, the intra-class correlation $\rho = \dfrac{\sigma^2_{u0}}{\sigma^2_{u0} + \sigma^2_{\varepsilon}}$ is estimated to be $\dfrac{.3954}{.3954 + .3466} = .533$. The most variation occurs at the hospital level and the value of ρ suggests that there is considerable clustering effect within hospitals, invalidating the traditional OLS results.

3.3.2 Including effects of hospital level (level 2) correlates

The unconditional means model, Model 1, provides a baseline against which we can compare more complex models. We include hospital characteristics which are level 2 variables: Region, Location, Bedsize. These categorical variables have more than 2 levels. For example, region has 4 categories (Northeast, Midwest, South, West). To keep our notation simple we use one name for each categorical variable indicating the region, location and bed size of each hospital. The model is:

Model 2: $Y_{ij} = \beta_{0j} + \varepsilon_{ij}$ and

$$\beta_{0j} = \beta_0 + \beta_{01} Region_j + \beta_{02} Location_j + \beta_{03} Bedsize_j + u_{0j}$$

where $u_{0j} \sim iid\ N(0, \sigma^2_{u0})$, $\varepsilon_{ij} \sim iid\ N(0, \sigma^2_{\varepsilon})$ and u_{0j}, ε_{ij} are assumed independent.

Substituting the level 2 equation into the level 1 equation yields:

$$Y_{ij} = [\beta_0 + \beta_{01} Region_j + \beta_{02} Location_j + \beta_{03} Bedsize_j] + [u_{0j} + \varepsilon_{ij}].$$

The terms in the first bracket represent the fixed part, while the terms in the second bracket represent the random part. The residual (error) variance within hospital σ^2_{ε}, remains almost unchanged (going from *.3466 to .3465*). However the variance component representing variance between hospitals, σ^2_{u0}, is much lower (going from *.3954 to .1682*). Therefore the three variables, Region, Location, Bedsize explain a large portion of the hospital-to-hospital variation in log total charges. About $\frac{.3954 - .1682}{.3954} = .57$ of the explainable variation in hospital log total charges is explained by the three variables. We can estimate once again the intra-class correlation as

$\frac{.1682}{.1682+.3465} = .33$ which remains high.

3.3.3 Including effects of discharge level (level 1) correlates

We illustrate the effect of including discharge level correlates by initially examining a model with only level 1 correlates, excluding level 2 correlates. After reviewing the necessary steps for including level 1 correlates, we fit a combined model. Suppose first that all correlates: Procedure, LOS, Age, Female, CCI, NPR are fixed effects. Since NPR, the number of procedures, is in strict dependence with Procedure (for example a patient who receives a procedure, CABG, PTCA or CATH will have NPR ≥ 1, a patient who has no procedures will have NPR=0), we use the cross-term NPR*Procedure to assure a unique interpretability of the coefficients in the model. Based on residual analyses we also decide to include LOSSQ = LOS^2 and Agesq=Age^2 in the model.

Then the model can be written as:

Model 3a: $$Y_{ij} = \beta_{0j} + \sum_{k=1}^{K} \beta_k x_{(k)ij} + \varepsilon_{ij} \ , \ \beta_{0j} = \beta_0 + u_{0j}$$

where $u_{0j} \sim iid\ N(0, \sigma^2_{u0})$, $\varepsilon_{ij} \sim iid\ N(0, \sigma^2_{\varepsilon})$ and u_{0j}, ε_{ij} are assumed independent.

The covariates $x_{(k)}$ include LOS, LOSSQ, Age, Agesq all as continuous variables; and indicator variables for procedure type (CABG, PTCA, CATH), Gender (female), CCI (0), NPR (1-4), and two dummies for CABG*NPR (CABG, 1-4 procedures) and PTCA*NPR

(PTCA, 1-4 procedures). As expected these fixed effects variables at the discharge level considerably reduce the within hospital error variance, σ_ε^2, going from .3466 in Model 1 to .0965.

Since LOS is a very strong correlate of the outcome and the variance of Y_{ij} varies with LOS as seen in Figure 3.3, we consider having a random coefficient for LOS. This will allow for the relationship between LOS and patient total charges to vary across hospitals. The model can be written as:

Model 3b: $$Y_{ij} = \beta_{0j} + \beta_{1j} LOS_{ij} + \sum_{k=1}^{K} \beta_k x_{(k)ij} + \varepsilon_{ij} ,$$

$$\beta_{0j} = \beta_0 + u_{0j}, \beta_{1j} = \beta_1 + u_{1j},$$

where $\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim iid\ N(O_2, \Sigma_{2x2})$, $\varepsilon_{ij} \sim iid\ N(0, \sigma_\varepsilon^2)$ and (u_{0j}, u_{1j}), ε_{ij} are assumed independent. Our unstructured covariance matrix $\Sigma = \begin{pmatrix} \sigma_{u0}^2 & \sigma_{u01} \\ \sigma_{u01} & \sigma_{u1}^2 \end{pmatrix}$ is chosen by considering the AIC, BIC criteria. Model 3b becomes

$Y_{ij} = \beta_0 + \sum_{k=1}^{K} \beta_k' x_{(k)ij} + u_{0j} + u_{1j} LOS_{ij} + \varepsilon_{ij}$. Within a hospital, the variance of Y_{ij} is

$Var(Y_{ij}) = Var(u_{0j} + u_{1j} LOS_{ij} + \varepsilon_{ij}) = \sigma_{u0}^2 + 2\sigma_{u01} LOS_{ij} + \sigma_{u1}^2 LOS_{ij}^2 + \sigma_\varepsilon^2$. The estimated covariance parameter estimates are $\sigma_{u0}^2 = .1958$, $\sigma_{u1}^2 = .00052$, $\sigma_{u01}^2 = -.00449$. Also $\sigma_\varepsilon^2 = .09352$. We compare the two models above, Model 3a and Model 3b. When fitting this models we find that the change in -2LL when using REML method to estimate parameters is ~728.6. An approximate test of the null hypothesis that this change is 0 is

obtained when we compare the difference in -2LL's to a χ^2 distribution with 2 degrees of freedom. The p-value is <.0001. Therefore we will keep random slopes for LOS.

Should LOSSQ, along with LOS, be included as random effects at the hospital level? If we do so, there is an additional variance component that comes from the LOSSQ term. The error term is $\begin{pmatrix} u_{0j} \\ u_{1j} \\ u_{2j} \end{pmatrix}$. Figure 3.3 suggests that LOSSQ is not needed. Fitting a general covariance matrix $\Sigma = \begin{pmatrix} \sigma_{u0}^2 & \sigma_{u01} & \sigma_{u02} \\ \sigma_{u01} & \sigma_{u1}^2 & \sigma_{u12} \\ \sigma_{u02} & \sigma_{u12} & \sigma_{u2}^2 \end{pmatrix}$ results in an estimate $\sigma_{u2}^2 \approx 0$, i.e. the estimate is on the boundary of the parameter space. Another reason for this is seen in

$$Var(Y_{ij}) = Var(u_{0j} + u_{1j} LOS_{ij} + u_{2j} LOS_{ij}^2 + \varepsilon_{ij}) =$$

$$= \sigma_{u0}^2 + \sigma_{u1}^2 LOS_{ij}^2 + \sigma_{u2}^2 LOS_{ij}^4 + \sigma_{\varepsilon}^2 + 2\sigma_{u01} LOS_{ij} + 2\sigma_{u02} LOS_{ij}^2 + 2\sigma_{u12} LOS_{ij}^3$$

which has a 4th-order term in LOS. If $\sigma_{u2}^2 \approx 0$ so are σ_{u02} and σ_{u12}.

3.3.4 Including both effects of discharge level (level 1) and hospital level (level 2) correlates

Having considered models with either just 1 level correlates or level 2 correlates, we can now consider models which contain variables of both types. Consider the following model:

Model 4: $$Y_{ij} = \beta_{0j} + \beta_{1j} LOS_{ij} + \sum_{k=1}^{K} \beta_k x_{(k)ij} + \varepsilon_{ij} \ ,$$

where $\beta_{0j} = \beta_0 + \beta_{01} Region_j + \beta_{02} Location_j + \beta_{03} Bedsize_j + u_{0j}$,

$$\beta_{1j} = \beta_1 + \beta_{11} Region_j + \beta_{12} Location_j + \beta_{13} Bedsize_j + u_{1j}.$$

Here $\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim iid\ N(O_2, \Sigma_{2x2})$ where $\Sigma = \begin{pmatrix} \sigma_{u0}^2 & \sigma_{u01} \\ \sigma_{u01} & \sigma_{u1}^2 \end{pmatrix}$, $\varepsilon_{ij} \sim iid\ N(0, \sigma_\varepsilon^2)$ and (u_{0j}, u_{1j}), ε_{ij} are assumed independent. Combining the two equations above yields the single two level equation model:

$$Y_{ij} = \beta_0 + \sum_{h=1}^{3} \beta_{0h} z_{hj} + \sum_{h=1}^{3} \beta_{1h} z_{hj} * LOS_{ij} + \sum_{k=1}^{K} \beta_k' x_{(k)ij} + (u_{0j} + u_{1j} LOS_{ij} + \varepsilon_{ij})$$

where $z_{1j} = Region_j$, $z_{2j} = Location_j$, $z_{3j} = Bedsize_j$.

In addition to the cross-terms required by the model: Region*LOS, Location*LOS, Bedsize*LOS, and NPR*Procedure, significant cross-terms procedure*LOS, Procedure*LOSSQ, NPR*LOS, CCI*LOS, CCI*LOSSQ are included in the model. Within a hospital, the conditional variance of Y_{ij} is $Var(Y_{ij}) = \sigma_{u0}^2 + 2\sigma_{u01} LOS_{ij} + \sigma_{u1}^2 LOS_{ij}^2 + \sigma_\varepsilon^2$ and the covariance between two different patients: $Cov(Y_{ij}, Y_{lj}) = CoVar(u_{0j} + u_{1j} LOS_{ij} + \varepsilon_{ij}, u_{0j} + u_{1j} LOS_{lj} + \varepsilon_{lj}) =$

$$= \sigma_{u0}^2 + \sigma_{u1}^2 LOS_{ij} LOS_{lj} + \sigma_{u01} (LOS_{ij} + LOS_{lj}).$$

These elements form higher level block matrices V_j represented as $V_j = Z\Omega_2 Z' + \Omega_1$, where Z contains the *j*th hospital values for the known design matrix of explanatory variables for the random portion of the model, etc. The covariance parameter estimates are $\sigma_{u0}^2 = .1189$, $\sigma_{u1}^2 = .000202$, $\sigma_{u01} = -.00137$ and the estimated level one

variance is $\sigma_\varepsilon^2 = .08866$. This can be compared with the level one variance obtained from the null model, Model 1, or total within hospital variance, .3466. The ratio $\frac{.3466 - .08866}{.3466} = .74$ represents the proportion of within hospital variance that is explained by correlates (comorbidity, procedure, demographic variables). The between hospital covariance is 0.

3.3.5 Inverse probability weighting

Before estimating $c_{01}(t \mid Z_0)$ for specific covariate profiles, we need to account for the uncensored observations in the analysis. As discussed, previously in Chapter 2, we will weight each observation by $w_i = \delta_i [T_i \leq \tau] / G(T_i)$, where δ_i is the censoring indicator and $G(\cdot)$ is the survival function for the censoring distribution.

To be consistent with our notations above we will use *j* for hospital and *i* for discharges, although we are well aware that these are not the standard notations when one would consider hospital as the primary units and discharges within hospitals as subunits. Consider the Random-Effects model

$$y_j = X_j \beta + Z_j \alpha_j + u_j \tag{3.1}$$

where the dimensions of y_j and u_j are $n_j \times 1$, β is $p \times 1$, X_j is $n_j \times p$, Z_j is $n_j \times q$ and α_j is $q \times 1$, where p is the number of fixed effects, q is the number of random effects

and n_j is the number of discharges in the jth hospital. We assume the usual conditions of a Random-Effects model[34]

$$E(u_j \mid X_j, Z_j, \alpha_j) = 0,\ E(\alpha_j \mid X_j, Z_j) = 0.$$

Let $v_j = Z_j\alpha_j + u_j$ be the composite error. Under these assumptions $E(v_j) = 0$ and $Var(v_j) = Z_j G Z_j' + R_j$, where $Var(\alpha_j) = E(\alpha_j\alpha_j' \mid X_j, Z_j) = G$ and $Var(u_j) = E(u_j u_j' \mid X_j, Z_j, \alpha_j) = R_j$. The form of G is specified by the TYPE option in the RANDOM statement, while the form of R_j is specified through the TYPE option in the REPEATED statement. To estimate β we use generalized least squares (GLS) estimation, so an estimate $\hat{\beta}$ of β that minimizes the expression $\sum_j (y_j - X_j\beta)' V_j^{-1} (y_j - X_j\beta)$ is given by

$$\hat{\beta} = (\sum_j X_j' V_j^{-1} X_j)^{-1} (\sum_j X_j' V_j^{-1} y_j).$$

Weighting observations

Let $W_j = diag\{w_{ij}, i = 1, \ldots n_j\}$ denote a diagonal matrix of weights w_{ij} for discharges within the jth hospital. Weighting observations is more than simply transforming model (3.1) by $W_j^{1/2}$. If we do so, and write the transformed model as

$$\tilde{y}_j = \tilde{X}_j\beta + \tilde{Z}_j\alpha_j + \tilde{u}_j \qquad (3.2)$$

(where $\tilde{y}_j = W_j^{1/2} y_j$ and the other quantities are similarly defined), then it is easily seen that GLS applied to (3.2) does not change our estimator $\hat{\beta}$, indeed if

$\tilde{V}_j = Var(W_j^{1/2} v_j) = W_j^{1/2}(Z_j G Z_j' + R_j) W_j^{1/2}$ then

$$(\sum_j \tilde{X}_j' \tilde{V}_j^{-1} \tilde{X}_j)^{-1} (\sum_j \tilde{X}_j' \tilde{V}_j^{-1} \tilde{y}_j) = (\sum_j X_j' V_j^{-1} X_j)^{-1} (\sum_j X_j' V_j^{-1} y_j).$$

Consider first the standard linear regression model $y_i = x_i \beta + u_i$ (where we now return to standard subscript 'i' for subject). By weighting the observations by w_i, we actually estimate β by minimizing $\sum_i w_i (y_i - x_i \beta)^2$. This gives

$$\hat{\beta} = (\sum_i w_i x_i' x_i)^{-1} (\sum_i w_i x_i' y_i) \text{ and } Var(\hat{\beta}) = (E(w_i x_i' x_i))^{-1} E(w_i^2 u_i^2 x_i' x_i)(E(w_i x_i' x_i))^{-1}.$$

In the general RE model (3.1), by weighting we mean transforming $y_j \to \tilde{y}_j$, $Z_j \to \tilde{Z}_j$, $X_j \to \tilde{X}_j$, but preserving the original variance form $V_j = Z_j G Z_j' + R_j$. Then $\hat{\beta}$ is obtained by minimizing $\sum_j (y_j - X_j \beta)' W_j^{1/2} \tilde{V}_j^{-1} W_j^{1/2} (y_j - X_j \beta)$. Now consider $\tilde{V}_j = \tilde{Z}_j G \tilde{Z}_j' + R_j$.

1. If $G = 0$, then $W_j^{1/2} \tilde{V}_j^{-1} W_j^{1/2} = W_j^{1/2} R_j W_j^{1/2}$. In effect therefore, the original R_j is replaced by $W_j^{-1/2} R_j W_j^{-1/2}$ to get

$$\hat{\beta} = (\sum_j X_j' W_j^{1/2} R_j^{-1} W_j^{1/2} X_j)^{-1} (\sum_j X_j' W_j^{1/2} R_j^{-1} W_j^{1/2} y_j).$$

This is the practice adopted in SAS PROC MIXED (Chapter 18, Version 8.2 p633). If in particular $R_j = \sigma_\varepsilon^2 I$ then $\hat{\beta} = (\sum_j X_j' W_j X_j)^{-1} (\sum_j X_j' W_j y_j)$.

2. If $G \neq 0$, then the form of $\tilde{V}_j^{-1} = (\tilde{Z}_j G \tilde{Z}_j' + R_j)^{-1}$ is more complicated. Indeed

$\tilde{V}_j^{-1} = R_j^{-1}[I - \tilde{Z}_j(G^{-1} + \tilde{Z}_j' R_j^{-1}\tilde{Z}_j)^{-1}\tilde{Z}_j' R_j^{-1}]$. Then

$$W_j^{1/2}\tilde{V}_j^{-1}W_j^{1/2} = W_j^{1/2}R_j^{-1}[I - \tilde{Z}_j(G^{-1} + \tilde{Z}_j' R_j^{-1}\tilde{Z}_j)^{-1}\tilde{Z}_j' R_j^{-1}]W_j^{1/2} =$$

$$= (W_j^{-1/2}R_jW_j^{-1/2})^{-1}[I - Z_j(G^{-1} + Z_j'(W_j^{-1/2}R_jW_j^{-1/2})^{-1}Z_j)^{-1}Z_j'(W_j^{-1/2}R_jW_j^{-1/2})^{-1}].$$

This shows that the effect of weighting is to change R_j to $W_j^{-1/2}R_jW_j^{-1/2}$.

The final form of the GLS estimator of β under weighting is then

$$\hat{\beta} = (\sum_j X_j'\tilde{\tilde{V}}_j^{-1}X_j)^{-1}(\sum_j X_j'\tilde{\tilde{V}}_j^{-1}y_j), \tag{3.3}$$

where $\tilde{\tilde{V}}_j = W_j^{-1/2}\tilde{V}_jW_j^{-1/2}$ and the variance of $\hat{\beta}$ is

$$Var(\hat{\beta}) = (\sum_j X_j'\tilde{\tilde{V}}_j^{-1}X_j)^{-1}(\sum_j X_j'\tilde{\tilde{V}}_j^{-1}\hat{v}_j\hat{v}_j'\tilde{\tilde{V}}_j^{-1}X_j)(\sum_j X_j'\tilde{\tilde{V}}_j^{-1}X_j)^{-1},$$

where $\hat{v}_j$ are the residuals $y_j - X_j\hat{\beta}$. This can be obtained using the EMPIRICAL option in the PROC MIXED statement. This estimator has been described in White (1980)[78], Liang and Zeger (1986)[79], and Diggle, Liang, and Zeger (1994)[80] and is commonly referred to as the "sandwich" estimator. We need to specify this in accordance with our methods described in Chapter 2 of the thesis. For this option to take effect we must include the RANDOM or REPEATED statements.

If $R_j = \sigma_\varepsilon^2 I$, then

$$\tilde{\tilde{V}}_j^{-1} = W_j^{1/2}\tilde{V}_j^{-1}W_j^{1/2} = W_j^{1/2}(W_j^{1/2}Z_jGZ_j'W_j^{1/2} + \sigma_\varepsilon^2 I)^{-1}W_j^{1/2} = (Z_jGZ_j' + \sigma_\varepsilon^2 W_j^{-1})^{-1}$$

and (3.3) gives $\hat{\beta} = (\sum_j X_j'(Z_jGZ_j' + \sigma_\varepsilon^2 W_j^{-1})^{-1}X_j)^{-1}(\sum_j X_j'(Z_jGZ_j' + \sigma_\varepsilon^2 W_j^{-1})^{-1}y_j)$.

3.4 Results of estimation

In this section we will concentrate on obtaining estimates of median LOS and total charges at median LOS, at specified covariate profiles. For example, we vary the type of procedure, hospital region and the number of comorbidities, but we keep fixed age (= 65), gender (= male), hospital location (= rural), hospital bed size (= medium) and the number of procedures between 1 and 4, if there is a primary procedure. All analyses were carried out using SAS. Various estimation methods for G and S, the survival distribution functions for censoring time, U and LOS (= T), are discussed in Chapter 2, Remark 2.2.1.2. Here, we choose parametric methods for estimating both G and S.

Estimation of G

Given a smaller number of covariates z (procedure, Charlson Comorbidity Index, region and number of procedures) an estimate of $G(t \mid z)$ may be obtained from a parametric model $\log(U) = z\beta_u + \sigma_u \varepsilon$, where β_u and σ_u are parameters and ε is an unknown error with a specified parametric distribution. PROC LIFEREG is used for the estimation of β_u. Various choices for the distribution of ε were considered. Based on the AIC we chose the log-normal and gamma distributions.

Graphs of the Kaplan-Meier estimates of the cumulative hazard versus the regression model estimates of the cumulative hazard from both log-normal and gamma models are shown in Figure 3.4. The plotted pairs of points for the gamma regression model initially fall on the referent line and then fall mostly above the referent line. Based on this plots we decide that the log-normal distribution provides a better fit to the data.

The estimated weights are $w_i = \delta_i / \hat{G}(T_i \mid z_i)$ where $\hat{G}(t \mid z) = 1 - \Phi(\frac{\log(t) - z\hat{\beta}_u}{\hat{\sigma}_u})$,

where Φ is the cumulative distribution function for the normal distribution.

Estimation of $c_{01}(t \mid Z_0)$

To estimate $c_{01}(t \mid Z_0)$, we use the weighted regression model described in the previous section. Parameter estimates derived using PROC MIXED are summarized in Table 3.4. Although coefficient estimates for continuous variables must be interpreted with care, all estimates seem to be in the right direction. For example, charges increase with the complexity of the cardiac procedure, with CABG having the highest charge followed by PTCA and diagnostic CATH. Patients who underwent none of these procedures had the lowest cost. For a given covariate profile Z_0, which contains t and t^2,

$\hat{c}_{01}(t \mid Z_0) = \exp(Z_0\hat{\beta})\exp(½(\hat{\sigma}^2_{u0} + 2\hat{\sigma}_{u01}t + \hat{\sigma}^2_{u1}t^2 + \hat{\sigma}^2_{\varepsilon}))$.

Estimation of $S(t \mid Z_0)$

Similar to estimating *G*, we estimate $S(t \mid Z_0)$ using PROC LIFEREG in SAS, from a parametric model for uncensored observation (i.e. $\delta_i = 1$), $\log(T) = X\beta_s + \sigma_s\varepsilon$, where ε is a vector of errors. The log–logistic and gamma distributions exhibit the lowest AIC. Graphs of the Kaplan-Meier estimates of the cumulative hazard versus the regression model estimates of the cumulative hazard from both log-logistic and gamma models are shown in Figure 3.5.

Based on these graphs, the AIC criterion and previous experiences with the analysis of LOS[49, 50] we use the log-logistic distribution. Estimates of $dS(t \mid Z_0)$ are obtained as $d\hat{S}(t \mid Z_0) = \frac{\exp(w)}{\hat{\sigma}_s (1+\exp(w))^2} dt$, where $w = (\log(t) - Z_0\hat{\beta}_s)/\hat{\sigma}_s$.

Finally, cost estimates at any time t for fixed time covariate profile Z_0 are obtained as $\int_0^t \hat{c}_{01}(u \mid Z_0) d\hat{S}(u \mid Z_0) =$

$$= \int_0^t \exp(Z_0(u)\hat{\beta}) \exp(½(\hat{\sigma}_{u0}^2 + 2\hat{\sigma}_{u01}u + \hat{\sigma}_{u1}^2 u^2 + \hat{\sigma}_\varepsilon^2)) \frac{u^{1/\hat{\sigma}_s} \exp(-Z_0\hat{\beta}_s/\hat{\sigma}_s)}{\hat{\sigma}_s (1+u^{1/\hat{\sigma}_s} \exp(-Z_0\hat{\beta}_s/\hat{\sigma}_s))^2} du$$

where we denote $Z_0(u) = (Z_0, u, u^2)$. The expression of $\int_0^t \hat{c}_{01}(u \mid Z_0) d\hat{S}(u \mid Z_0)$ can be viewed as a function $\rho_t(\hat{\beta}, \hat{\sigma}, \hat{\beta}_s, \hat{\sigma}_s)$ where $\hat{\sigma} = (\hat{\sigma}_{u0}^2, \hat{\sigma}_{u01}, \hat{\sigma}_{u1}^2, \hat{\sigma}_\varepsilon^2)$. For a fixed t, using the Delta Method we compute the variance of the mean cost. Estimates of median LOS, and total charges at median LOS, together with confidence intervals, for specified covariate profiles as well as graphic representations are presented in Table 3.4, and Figures 3.6, and 3.7.

Discussion

We have applied methods discussed in Chapter 2 of this thesis to estimate mean charges for hospital stays of specified duration for patients hospitalized for AMI. Our method accounts for the existence of a non-zero intra-hospital correlation, censoring and skewness of total charges.

Estimates derived in our model are quantitatively similar to those reported in other studies of health care utilization in patients undergoing CABG surgery or PTCA.

Although 2000 costs might seem more relevant for comparison purposes, costs inflated to 2000 may not fully reflect current costs because the efficiency of intervention procedures and cardiac surgery services probably has increased during the last years; Weintraub (1995)[81] argues that costs may best be used for comparison between procedures. Also, direct comparisons with our estimates is not possible because of different patient demographic and clinical characteristics, use of charges as a proxy for costs and different time periods.

We now discuss results of previous studies. Three randomized clinical trials have evaluated the relative costs of angioplasty versus bypass surgery[82]. These studies included all costs to the patient, including hospitalization, procedures, medications and follow-up care. The Bypass Angioplasty Revascularization Investigation (BARI) Study of Economics and Quality of Life (SEQOL) (n =1829 patients) was collected over a three-year period from August 1988 to August 1991. Using BARI data, Hlatky *et al* (1997)[83] report that the mean initial hospital cost of angioplasty was 65% of that of CABG, while same percent in our data is 67%. The Randomised Intervention Treatment of Angina (RITA) trial (UK) reported that an angioplasty had only 52 percent of the cost of a bypass operation following the initial procedure[84]. The Emory Angioplasty Versus Surgery (EAST) trial found the same percent to be 63%[81].

Using data from Michigan Inter-institutional Collaborative Heart Study (MICH), collected on 360 patients who underwent cardiac procedures in two large urban medical centers in eastern Michigan during January 1994 through April 1995, Polverejan *et al* (2003)[49] report mean charges of 23,545$ and 26,213 (inflated 2000 costs: 26,080$, 28,435$) for patients who underwent CABG with 1 or 2 comorbidities respectively. From

a large national sample of Medicare patients who underwent CABG surgery in 1990, Cowper *et al* (1997)[85] report a mean hospital cost of 26,976$ (inflated to 2000 dollars: 33,202$) for patients with AMI undergoing bypass surgery.

We observe higher length of stay in NE then in any other region. Similar findings from Centers for Disease Control and Prevention, National Center for Health Statistics[86-88] are reported: in most years, the average length of stay in the Northeast was significantly longer than the average stays in the other three regions.

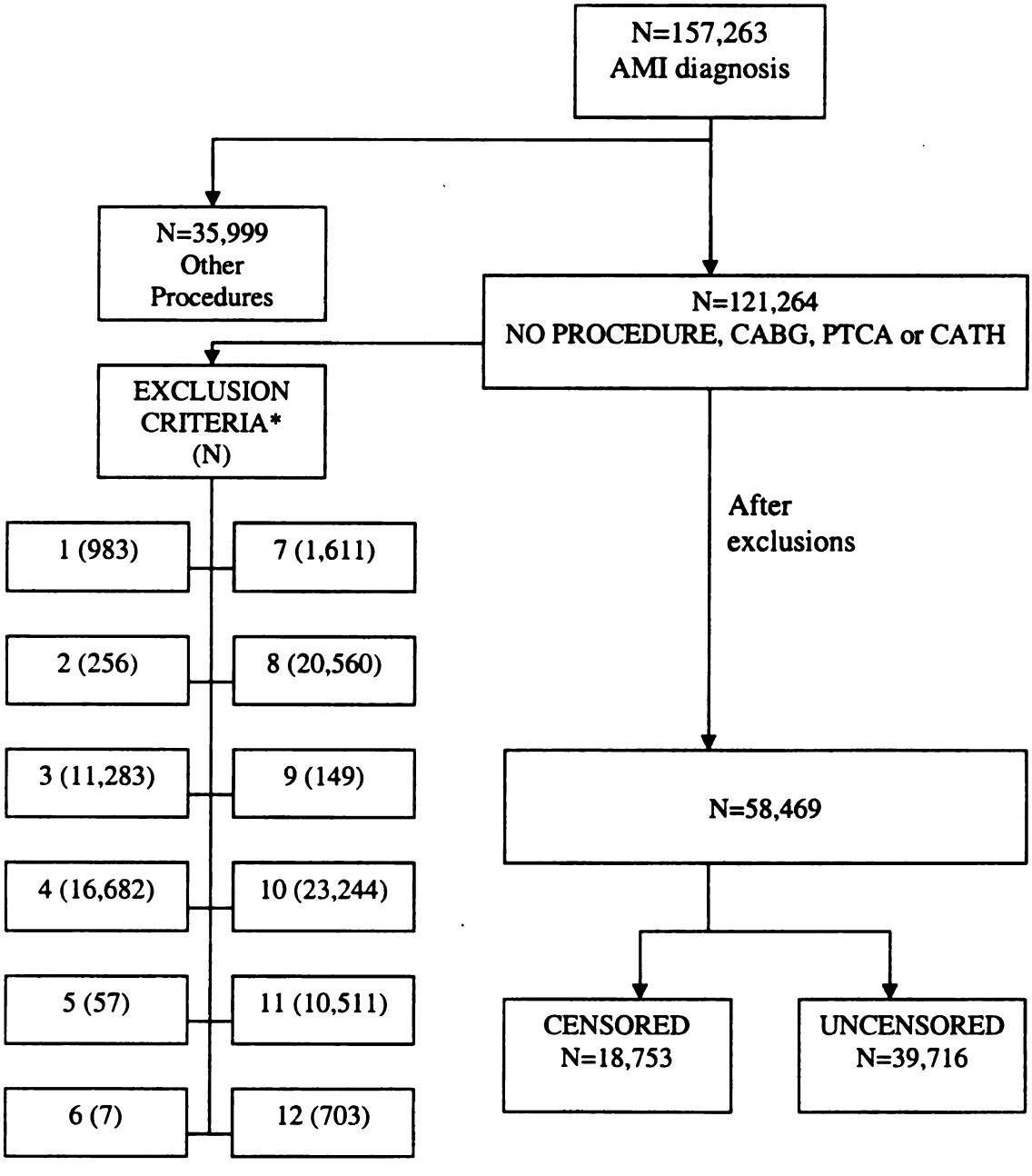
NIS data has both strengths and limitations. Because NIS data represent hospital discharges and not individual persons, the main limitation is that there exists the possibility that these discharges will show up as complete cases in the data set as another record. However, restricting our admissions to ER admissions and all routine referrals would decrease this chance considerably. This issue is impossible to resolve completely because of the inability in the NIS to track individual patients. Another limitation of the data set is that it lacks clinical detail (e.g., stage of disease, vital statistics) and laboratory and pharmacy data. The main strength of the NIS is that it is the largest collection of all-payer, uniform, state based inpatient data.

TABLE 3.1 Primary procedures

Primary procedure* **CCS code: Description**	**N**	**PERCENT**
.: No PR code	44,003	28.0
44: Coronary artery bypass graft (CABG)	15,290	9.7
45: Percutaneous coronary angioplasty (PTCA)	3,7249	23.7
47: Diagnostic cardiac catheterization, coronary artheriography (CATH)	24,722	15.7
48: Insertion, revision, replacement, removal of cardiac pacemaker or cardioverter/defibrillator	2,026	1.3
49: Other O.R. heart procedures	2,908	1.8
63: Other non-O.R. therapeutic cardiovascular procedures	1,714	1.1
193: Diagnostic ultrasound of heart (echocardiogram)	2,549	1.6
216: Respiratory intubation and mechanical ventilation	5,562	3.5
222: Blood transfusion	1,826	1.2
231: Other therapeutic procedures	6,350	4.0

*All AMI admissions (N=157,263). Only primary procedures that are present in more than 1% of discharges are shown

Figure 3.1 Creating a working data set



*Exclusion criteria above are not mutual exclusive, a discharge may be excluded for more than one reason

TABLE 3.2 Descriptive statistics (N=58469)

		N=58,469 (all discharges in the working data set)		N=39,716 (only discharges with complete TOTCHG)	
Variable	**Subgroup**	**Frequency**	**Percent**	**Frequency**	**Percent**
Procedure	CABG	7,369	12.6	4,586	11.6
	PTCA	17,901	30.6	16,540	41.7
	CATH	14,264	24.4	9,420	23.7
	No procedure	18,935	32.4	9,170	23.1
Gender	Male	35,977	61.5	25,466	64.1
	Female	22,492	38.5	14,250	35.9
NPR (number of procedures)	None	18,935	32.4	9,170	23.1
	1 - 4	16,489	28.2	11,915	30.0
	≥ 5	23,045	39.4	18,631	46.9
CCI	0	23,294	39.8	17,505	44.1
	1+	35,175	60.2	22,211	55.9
Region	Northeast	11,458	19.6	6,432	16.2
	Midwest	11,628	19.9	8,441	21.3
	South	25,219	43.1	17,344	43.7
	West	10,164	17.4	7,499	18.9
Location	Rural	7,900	13.5	4,210	10.6
	Urban Teaching	25,981	44.4	16,358	41.2
	Urban Non-Teaching	24,588	42.1	19,148	48.2
Bedsize	Small	5,174	8.9	2,743	6.9
	Medium	14,904	25.5	8,992	22.6
	Large	38,391	65.7	27,981	70.5
LOS ≥ 2	Continuous	Mean = 5.44 25th Percentile = 3 Median = 4 75th Percentile = 7	STD = 4.1	Mean = 5.06 25th Percentile=3 Median = 4 75th Percentile =6	STD = 3.4
18 ≤ AGE ≤ 85	Continuous	Mean = 65.73 25th Percentile = 56 Median = 67 75th Percentile = 76	STD = 12.7	Mean = 64.03 25th Percentile =54 Median = 65 75th Percentile =75	STD = 12.8

Figure 3.2 . Histogram of Log (TOTCHG)

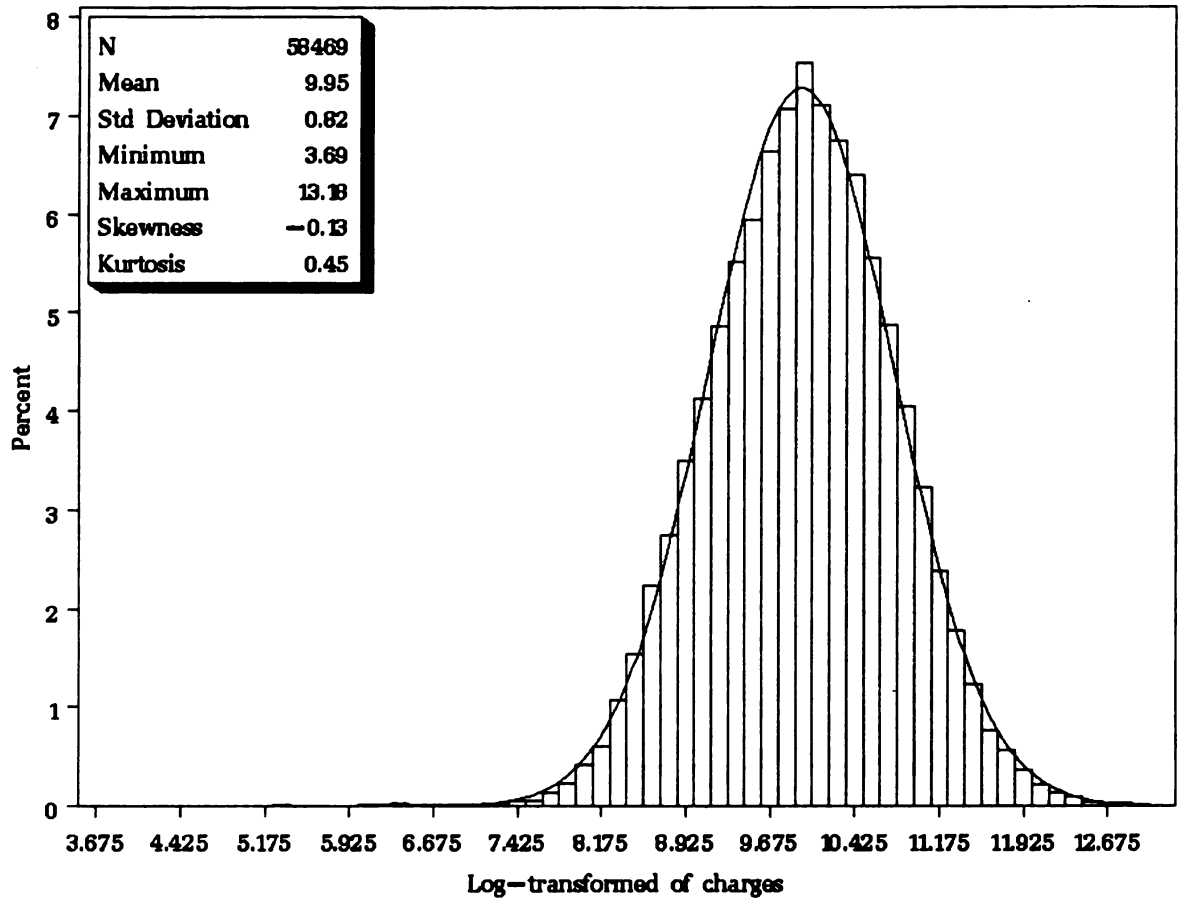


Figure 3.3 Graph of Log (TOTCHG) versus LOS

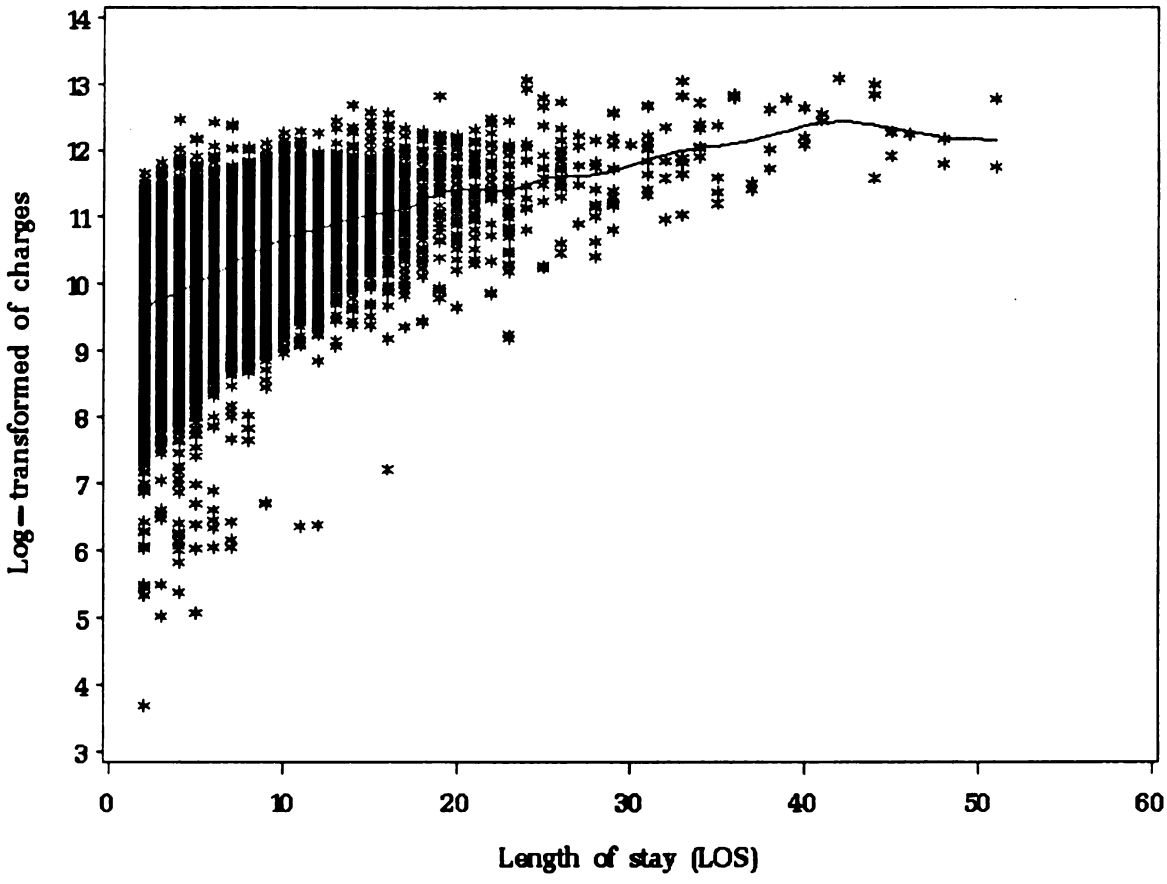


Figure 3.4 Estimation of G: Graphs of the Kaplan-Meier estimates of the cumulative hazard versus the regression model estimates of the cumulative hazard from both log-normal and gamma models

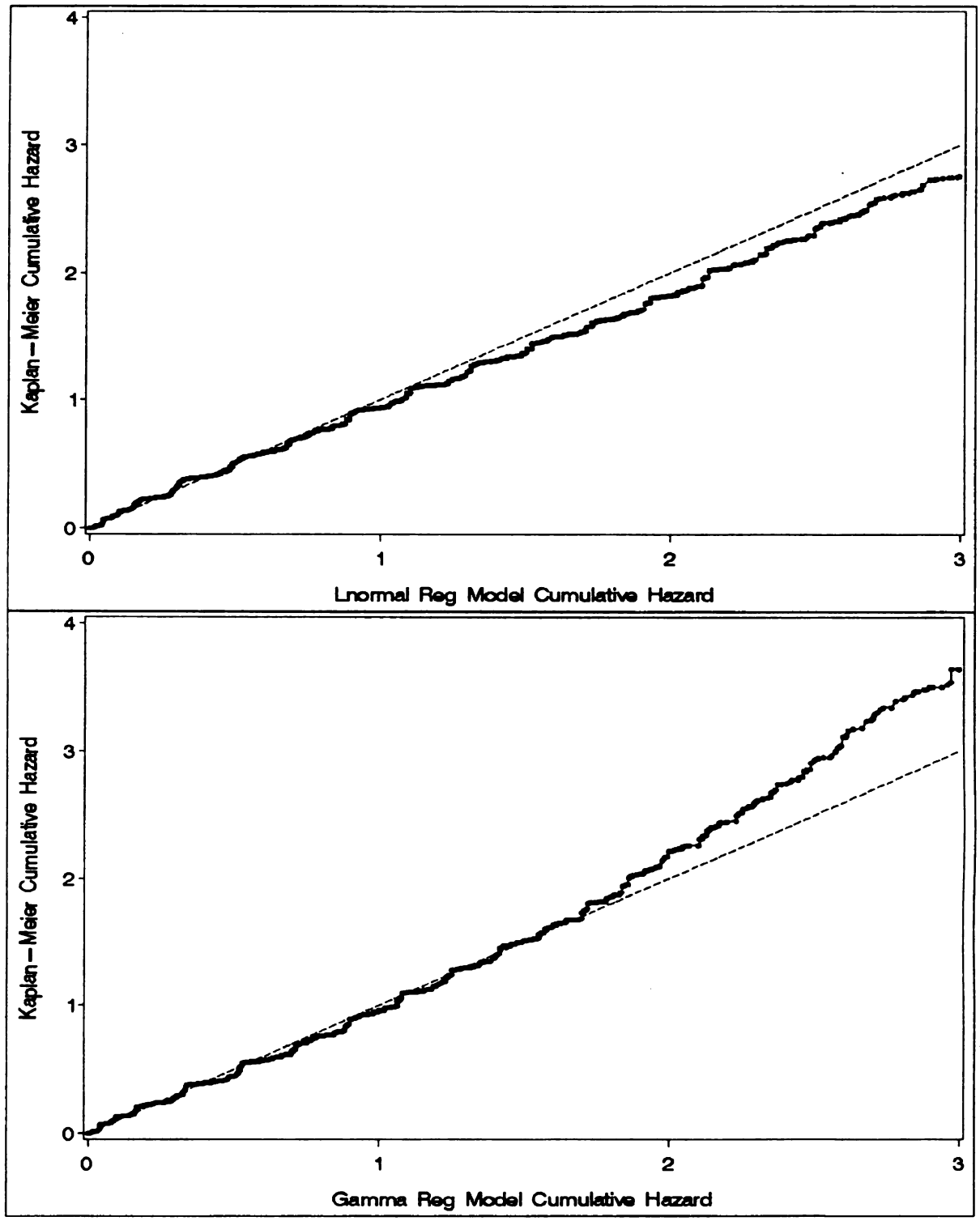


TABLE 3.3 Parameter estimates in multilevel regression

Effect	Groups	Estimate	StdErr	tValue	*P*-value
Intercept	_	8.400	0.0790	106.36	<.0001
Procedure	CABG	2.022	0.0553	36.56	<.0001
	PTCA	1.620	0.0424	38.21	<.0001
	CATH	1.031	0.0419	24.59	<.0001
LOS	_	0.206	0.0060	34.18	<.0001
LOSsq	_	-0.004	0.0003	-14.05	<.0001
CCI	No comorbidities	-0.059	0.0126	-4.68	<.0001
Male	_	0.022	0.0044	4.96	<.0001
AGE	_	0.005	0.0016	2.90	0.0037
Agesq	_	0.000	0.0000	-4.18	<.0001
Number of procedures	0 or ≥5 procedures	0.395	0.0317	12.47	<.0001
Procedure*NPR	CABG , ≥5procedures	-0.316	0.0571	-5.53	<.0001
	PTCA , ≥5procedures	-0.292	0.0432	-6.75	<.0001
Region	Northeast	-0.256	0.0405	-6.33	<.0001
	Midwest	-0.363	0.0346	-10.50	<.0001
	South	-0.172	0.0424	-4.06	<.0001
Location	Rural	-0.107	0.0358	-2.97	0.0031
Bedsize	Small	-0.126	0.0092	-13.71	<.0001
	Medium	-0.097	0.0067	-14.51	<.0001
LOS*Procedure	CABG	-0.074	0.0059	-12.61	<.0001
	PTCA	0.003	0.0004	9.30	<.0001
	CATH	0.003	0.0003	9.61	<.0001
LOSsq*Procedure	CABG	0.002	0.0003	7.85	<.0001
	PTCA	-0.141	0.0347	-4.06	<.0001
	CATH	-0.203	0.0310	-6.53	<.0001
LOS*NPR	0 or ≥5 procedures	-0.010	0.0022	-4.55	<.0001
LOS*CCI	No comorbidities	0.013	0.0032	4.03	<.0001
LOSsq*CCI	No comorbidities	0.000	0.0001	-3.53	0.0004
LOS*Region	Northeast	-0.008	0.0040	-1.98	0.0472
	Midwest	-0.007	0.0031	-2.37	0.0179
	South	-0.012	0.0030	-4.07	<.0001

Effect	Groups	Estimate	StdErr	tValue	*P*-value
LOS*Location	Rural	-0.001	0.0036	-0.28	0.7803
LOS*Bedsize	Small	0.001	0.0040	0.32	0.7481
	Medium	0.002	0.0029	0.70	0.4839
Covariance Parameters	**Subject**	**Estimate**	**StdErr**	**ZValue**	***P*-value**
UN(1,1)	HOSPID	0.1235	0.00837	14.75	<.0001
UN(2,1)	HOSPID	-0.0024	0.00049	-4.93	<.0001
UN(2,2)	HOSPID	0.0004	0.00004	9.57	<.0001
Residual		0.1384	0.00100	139.03	<.0001

Figure 3.5 Estimation of S: Graphs of the Kaplan-Meier estimates of the cumulative hazard versus the regression model estimates of the cumulative hazard from both log-logistic and gamma models

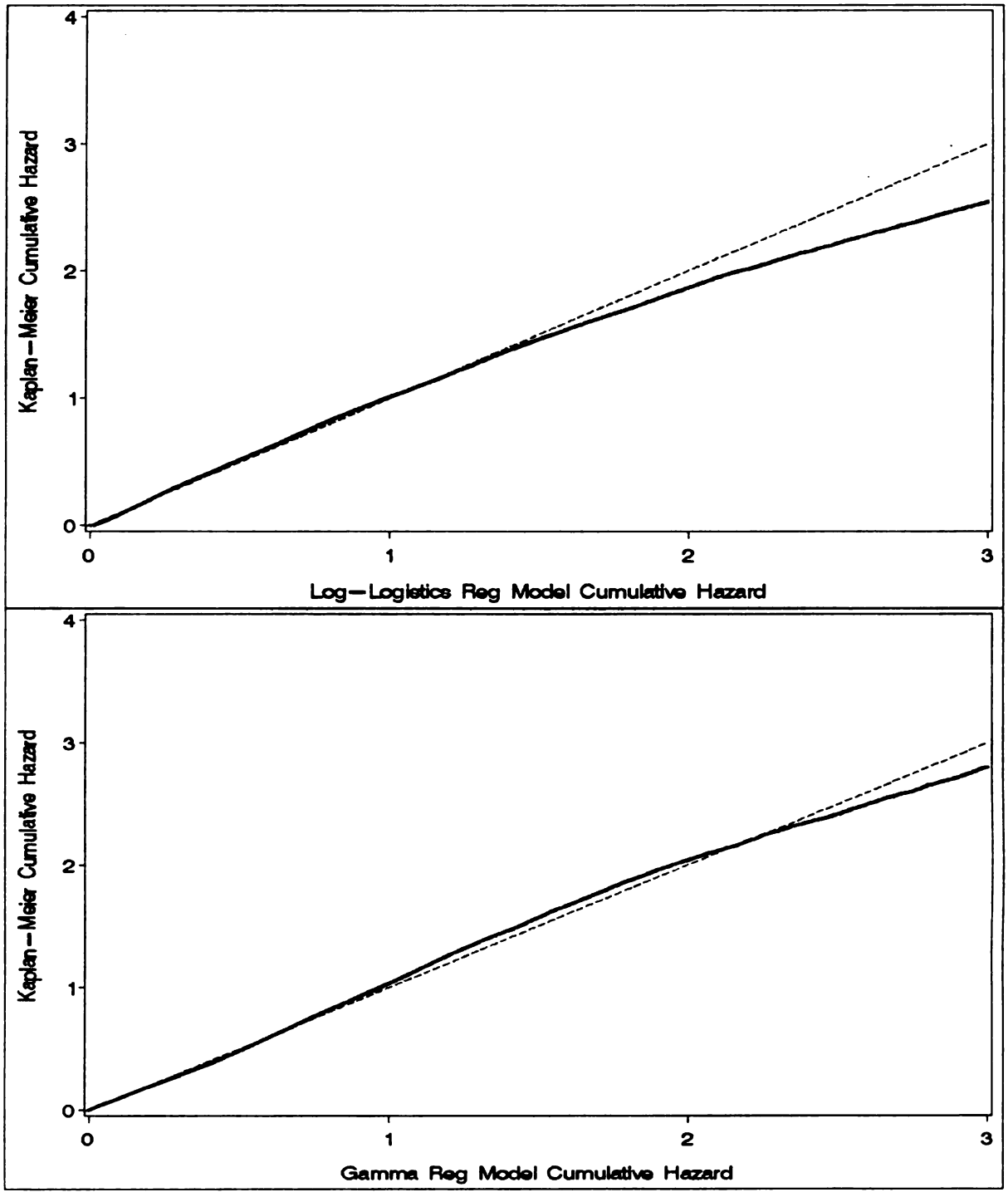
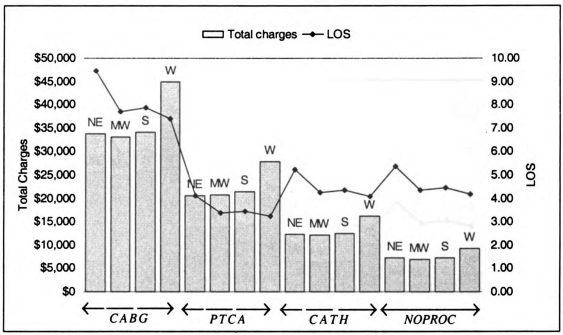


Table 3.4 Estimates of total charges at median LOS by procedure types, comorbidity levels (0, 1+) and regions*.

No comorbidity	CABG	NE	$33,705.24 (± $4,931.65) (LOS=9.45)	CATH	NE	$12,247.12 (± $1,554.89) (LOS=5.22)
		MW	$33,192.80 (± $3,997.98) (LOS=7.71)		MW	$12,111.16 (± $1,214.83) (LOS=4.26)
		S	$34,098.87 (± $3,924.04) (LOS=7.87)		S	$12,507.77 (± $1,173.06) (LOS=4.35)
		W	$44,882.55 (± $5,757.71) (LOS=7.41)		W	$16,278.16 (± $1,786.27) (LOS=4.10)
	PTCA	NE	$20,657.89 (± $2,646.69) (LOS=4.13)	NONE	NE	$7,214.51 (± $896.86) (LOS=5.36)
		MW	$20,703.55 (± $2,075.27) (LOS=3.37)		MW	$6,973.11 (± $672.76) (LOS=4.37)
		S	$21,386.47 (± $2,020.57) (LOS=3.44)		S	$7,205.49 (± $641.46) (LOS=4.46)
		W	$27,796.50 (± $3,071.51) (LOS=3.24)		W	$9,339.37 (± $982.97) (LOS=4.20)
At least one comorbidity	CABG	NE	$36,544.54 (± $5,350.00) (LOS=11.78)	CATH	NE	$13,356.57 (± $1,687.48) (LOS=6.51)
		MW	$35,770.72 (± $4,332.80) (LOS=9.61)		MW	$13,124.59 (± $1,316.81) (LOS=5.31)
		S	$36,682.38 (± $4,245.21) (LOS=9.82)		S	$13,532.62 (± $1,267.17) (LOS=5.42)
		W	$48,570.85 (± $6,267.93) (LOS=9.24)		W	$17,668.42 (± $1,943.16) (LOS=5.11)
	PTCA	NE	$22,232.91 (± $2,858.71) (LOS=5.15)	NONE	NE	$8,071.18 (± $994.88) (LOS=6.68)
		MW	$22,202.56 (± $2,260.66) (LOS=4.20)		MW	$7,721.79 (± $740.39) (LOS=5.45)
		S	$22,923.91 (± $2,199.08) (LOS=4.29)		S	$7,965.56 (± $702.93) (LOS=5.56)
		W	$29,858.35 (± $3,349.71) (LOS=4.04)		W	$10,360.54 (± $1,086.92) (LOS=5.24)

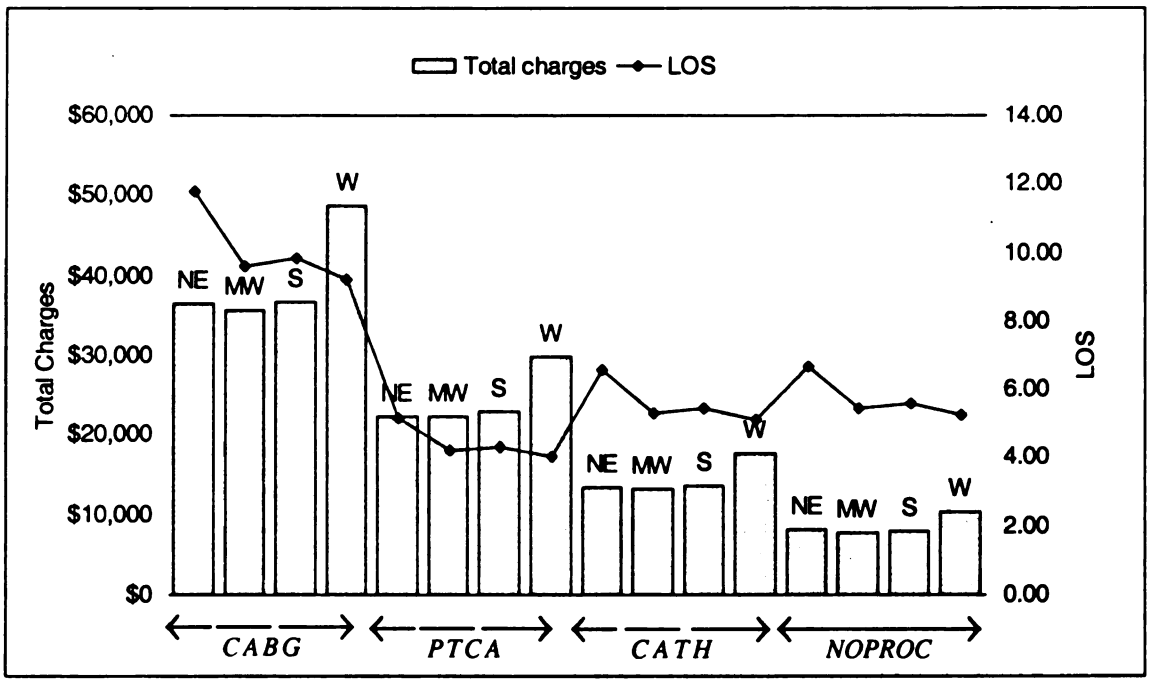
*for fixed covariate profile: age=65, gender=male, hospital location = rural, hospital bed size = medium, and 1-4 procedures (if there is a primary procedure)

Figure 3.6 Estimates of total charges at median LOS – no comorbidities*



*Estimates corresponding at median LOS, for covariate profile:
age=65, male, rural location, medium bed size, between 1-4 procedures

Figure 3.7 Estimates of total charges at median LOS – at least one comorbidity*



*Estimates corresponding at median LOS, for covariate profile:
age=65, male, rural location, medium bed size, between 1-4 procedures

CHAPTER 4

CONCLUSIONS AND FUTURE WORK

Evidence of rising health care expenditures is widespread. The resurgence in health care spending is fueled by fundamental forces that will prove difficult to resist[89]. One analyst has predicted that if current trends continue, health care will consume 25 percent of the GDP by the year 2030[90]. The enormous investment in biomedical research will probably accelerate the rate of technological development in medicine, with effects on overall expenses[89]. Also the proportion of the population 65 years of age or older and the proportion over 80 will increase by 33 percent and 14 percent, respectively, between 2000 and 2020[91]. No matter how healthy elderly persons are in the future, as compared with earlier generations, their sheer numbers are almost certain to result in increased expenditures for health care[89]. Yet another important influence will be increasing national prosperity. The more income people have, the more of it they tend to spend on health care. One of the strongest predictors of the proportion of the GDP that Western nations spend on health care is the GDP itself[89]. Therefore, substantial increases in health care spending over the next 5 to 10 years are virtually inevitable.

The United States spends 3.4 percent more on health care as a percentage of the GDP than any other Western country[92], but there is no conclusive evidence that health outcomes are better in the United States than in other industrialized nations[89, 93, 94].

Because of concerns on the availability of resources to pay for different health care interventions, medical technologies and treatments, CEA has evolved to become an important discipline in the pursuit of maximizing health benefits from a specified expenditure, or in finding the lowest-cost strategy for a specified health effect.

Although several advances have been made over the past decade, complexities in analyses of health care cost and outcomes still pose many methodological challenges. Guided by other research in this field [2-9, 11, 15, 30, 33, 34, 48, 49, 52, 53, 59, 60, 95-103], in this thesis we have addressed several statistical issues in the analysis of medical cost data.

The objective of Chapter 1 was to demonstrate how costs could be incorporated into a longitudinal model in which a patient's event history unfolds as sojourns in health states and as transitions between health states. Using a Markov model the counting process methodology was used to estimate the transition probabilities and integrated transition intensities with patient heterogeneity modeled through Cox regression on the underlying transition intensities. There are extensive uses of this model in biomedical applications. Incorporating costs into this framework is new, although the basic idea of calculating present value is ubiquitous in the insurance and finance literature where costs are manifest as fixed amounts paid at random times, and as fixed payment streams over random durations. However, the fundamental difference in our context is the need to estimate both the transition costs and the sojourn costs. Our approach uses a mixed random effects model to estimate costs conditional on time, and then the dynamics of time through transition probabilities and integrated intensities. In this way we obtain estimates of net present for expenditures incurred over a finite time horizon. The asymptotic distribution of these estimates is shown to depend on three stochastic

components: the regression parameter in the random effects model for costs, the regression parameter in the Cox regression model for transition intensities and the estimator of the baseline integrated intensity.

In Chapter 2, we showed that our transition model described in Chapter 1 captures costs under the simpler two state survival model with a single transition time and sojourn. We discuss parametric methods for estimating the survival distribution for censoring time, and therefore the weights in the IPCW technique, as well as methods of estimation for the survival distribution for event time.

In Chapter 3 we provided an application of our method using length of stay (LOS) and inpatient costs for patients hospitalized for AMI from the Nationwide Inpatient Sample 2000 (NIS 2000) of Healthcare Cost and Utilization Project. Our analysis addresses a number of characteristics of these utilization data including censoring of costs, skewness in cost distributions, heterogeneity in LOS and costs across patients. Our method is based on a hierarchical model in which patients are nested within hospitals.

Future work

To obtain cost-effectiveness measures such the cost-effectiveness ratio (CER) and net health cost (NHC) for competing health care interventions or treatments, we need both accurate estimates of costs and health benefits. In this thesis we have addressed several issues on estimating costs. The three basic measures of health outcomes are survival probability, life-expectancy (LE) and quality adjusted survival years (QALY).

To assess the health benefit of an intervention relative to another, we may also use any clinically meaningful measure such as improvement in life expectancy, deaths averted, or number of toxic side effects prevented. For example, in our multi-state description, life expectancy translates to the average time from the time origin until death or some other absorbing state is reached. Since the goal of any health care intervention is much broader than simply treating the disease condition or preventing death, the use of quality-adjusted life years (QALYs) to quantify health outcomes has been advocated[15]. QALYs also provide a common metric to gauge health benefit across disease types. For example, a decision maker facing resource allocation can compare the cost-effectiveness of coronary artery bypass surgery versus a percutaneous coronary intervention, and the cost-effectiveness of different lipid lowering therapies for the prevention of cardiovascular disease, and the cost-effectiveness of different regimens of screening women for their susceptibility to breast cancer.

For each unit of time spent in a health state, a quality weight is the relative value placed on that health state against the state of perfect health. Perfect health has a quality weight of one, while death, or states judged equivalent to death, get a quality weight of zero. All other health states receive a quality weight between zero and one. Having determined quality-weights for various health states, QALYs adjusts the length of life for the quality of life during those years. With multiple health states, each state is associated with its sojourn. For example, suppose we have estimated a total life-expectancy of 30 years, which consists of 10 years in a state with quality-weight 0.5, 10 years in a health state with quality weight 0.65, and 10 years in perfect health. This results in an expected 21.5 QALYs (=10×.5+10×.65+10×1). It is worth noting that the two main elements in

calculating QALYs, time spent in each health state and its associated utility, can vary across patients. Accounting for the random variation is an important task in any CEA.

Suppose the state space E is labeled with 'k' being an absorbing state and the remaining states being transient. For example, in cancer treatment studies[104] transient states are relapse and remission and death is an absorbing state. Survival time is then the time to absorption defined by $\tau_k = \inf\{t > 0 : X(t) = k\}$, with survival distribution conditional on the initial state,

$$S_{ik}(t) = P[\tau_k > t \mid X_0 = i] = 1 - P_{ik}(0,t).$$

If $\tau < \infty$ is the pre-specified time horizon of the analysis, we define *life-expectancy* (*LE*) restricted to τ and discounted at a constant rate r by

$$LE = \int_0^\tau e^{-rt} S_{ik}(t)dt,$$

conditional of the initial state. The total sojourn in state h in $[0,\tau]$ is $\int_0^\tau e^{-rt} Y_h(t)dt$ with expectation $\int_0^\tau e^{-rt} P_{ih}(0,t)dt$, the sum of which over all states $h \neq k$ is *LE*.

Let $q(h,t)$ denote the quality weight for the health state $h = X(t)$ occupied at time t, with the state 'dead' having quality weight equal to zero. The total quality adjusted time in $[0,\tau]$ is $\sum_{h \in E} \int_0^\tau e^{-rt} q(h,t) Y_h(t)dt$, and its expectation conditional on $X_0 = i$ defines *expected quality adjusted life years*, $QALY(i) = \sum_{h \in E} \int_0^\tau e^{-rt} q(h,t) P_{ih}(0,t)dt$. If $q(h,t) = q(h)$ are constant in time then $QALY(i)$ becomes $\sum_{h \in E} q(h) \int_0^\tau e^{-rt} P_{ih}(0,t)dt$, which is the sum over states of the discounted average time spent in each state weighted

by its quality. This is the standard definition for QALY[105]. Finally, an unconditional version of $QALY = \sum_{i \in E} \pi_i(0) QALY(i)$, where $\pi_i(0) = P(X(0) = i)$.

In the simplest multi-state model with two states, 0=alive (transient) and 1=dead (absorbing), the only transition is 0⟶1 which has hazard $\alpha_{00} = -\alpha_{01}$, and the initial distribution is degenerate, $\pi_0(0) = 1$. Also, the survival distribution is $S(t) = P[\tau_1 > t]$, τ_1 being the survival time. Without discounting (r =0), LE reduces to restricted mean survival, $E(\tau_1 \wedge \tau) = \int_0^\tau S(t)dt$. This entity is often used in economic evaluations when survival time is the principal endpoint. Finally, if we write $q(t) = q(0,t)$ we get $QALY = \int_0^\tau e^{-rt} q(t) S(t) dt$.

In comparing two competing health care interventions, a test intervention versus a standard, the *cost-effectiveness ratio* (CER) is the ratio of the incremental cost relative to the incremental benefit. In the patient population for which these interventions are intended, let μ_{tc} and μ_{te} denote respectively, the cost and effectiveness of the test intervention, and, μ_{sc} and μ_{se} be the corresponding measures for the standard intervention. Then the CER is the parameter $\theta = (\mu_{tc} - \mu_{sc})/(\mu_{te} - \mu_{se})$, and having established a maximum value of the CER θ_0, a value society is willing to pay to gain one unit of effectiveness by adopting the test intervention, the *net health cost* is given by $\mathrm{NHC}(\theta_0) = (\mu_{tc} - \mu_{sc}) - \theta_0 (\mu_{te} - \mu_{se})$. We equate μ_{tc} and μ_{sc} by $-\int_0^\tau c_{01}(s \mid Z_0) dS(t \mid Z_0)$, where $Z_0 = 1$ for the new intervention and $Z_0 = 0$ for placebo. We may define effectiveness, μ_{te} and μ_{se}, in terms of the restricted mean survival as $\int_0^\tau S(t \mid Z_0) dt$. Then

$$NHC(\theta_0) = \int_0^{\tau} c_{01}(s \mid Z_0 = 0) dS(t \mid Z_0 = 0) - \int_0^{\tau} c_{01}(s \mid Z_0 = 1) dS(t \mid Z_0 = 1) -$$

$$- \theta_0 \int_0^{\tau} (S(t \mid Z_0 = 1) - S(t \mid Z_0 = 0)) dt .$$

The methods developed in Chapters 1 and 2 can be applied to the estimation of QALY and NHC. Because our models incorporate covariates, their practical importance lies in the fact that both QALY and NHC can be estimated along specified covariate profiles. However, this requires data on patient health histories and costs over time. With the increased focus being given to controlling health care expenditures, such longitudinal data sets are likely to be compiled from administrative databases, large scale chronological and epidemiological studies. Many recent clinical trials have included economic sub-studies. For example the Antiarrhythmics Versus Implantable Defibrillators (AVID)[106] and Canadian Implantable Defibrillator Study (CIDS)[107] trials yielded useful information on both the clinical effectiveness of the implantable defibrillator as well as its cost-effectiveness. Many researchers have examined issues of statistical power and sample size assessments for cost-effectiveness studies[99,100,103,108-113]. They show that the requirements on sample size demonstrating cost-effectiveness can be many times larger than that needed to establish effectiveness alone. Therefore the investigators face the ethical dilemma of continuing trial and enlarging them for economic evaluation studies. It is here where our regression techniques could be most informative. Indeed, Simon Thompson recently editorialized in *Statistical Methods for Medical Research*[114], 'the challenge for the future will be development of regression models for costs and cost-effectiveness that can assess the impact of patient characteristics and cost elements that drive total costs'. Determining in which subgroups

of patients an intervention is more cost effective from health care utilization studies and clinical trials would be an active field of research in the coming years.

APPENDIX

MATHEMATICAL BACKGROUND

Concepts of stochastic processes, stochastic integration, functional delta method and results on Ito integration, all used in this thesis, will be briefly reviewed in this appendix. This appendix uses results from Andersen et al (1993)[25], Gill (1989)[58], Harrison (1985)[115], Oskendal (1995)[116] and Polverejan (2001)[48].

Stochastic processes

Let $(\Omega, \mathcal{F}, P)$ be a probability space. A filtration $(\mathcal{F}_t, t \in \mathcal{T})$ is an increasing right-continuous family of sub-σ-algebras of $\mathcal{F}$. A stochastic process X is just a time-index collection of random variables $\{X(t) : t \in \mathcal{T}\}$. The process X is called adapted to the filtration $\mathcal{F}_t$ if $X(t)$ is $\mathcal{F}_t$-measurable for each t. We write $X(t, \omega)$ for the realized value of $X(t)$ at the point $\omega \in \Omega$.

The process X is called cadlag if its sample paths $\{X(t, \omega) : t \in \mathcal{T}\}$, for almost all ω, are right-continuous with left-hand limits. The set of cadlag functions is denoted by $D(\mathcal{T})$, the Skorohod space of weak convergence theory.

For two stochastic processes X and Y, $\int XdY$ denotes the stochastic process $t \to \int_0^t X(s)dY(s)$ defined for each ω and t such that $\int_0^t |X(s)||dY(s)| < \infty$. Here Y is

assumed a cadlag process with paths of locally bounded variation, i.e. $\int_0^t |dY(s)| < \infty$ for all $t \in \mathcal{T}$, for almost all $\omega \in \Omega$. We call such a process Y a finite variation process, and the process $t \to \int_0^t |dY(s)|$ is called its total variation process.

To a cadlag process X, we associate its left-continuous modification X^-, defined by $X^-(t) = X(t-)$ and its jump process defined by $\Delta X(t) = X(t) - X(t-)$.

Martingale, predictable processes, compensators

A martingale is a cadlag adapted process M which is integrable, i.e. $E(|M(t)|) < \infty$ for all $t \in \mathcal{T}$ and satisfies the martingale property $E(M(t) | \mathcal{F}_s) = M(s)$ for all $s \le t$. The process is called sub(super)-martingale if the equality is replaced by $E(M(t) | \mathcal{F}_s) \ge M(s)$ ($E(M(t) | \mathcal{F}_s) \le M(s)$ respectively). A martingale is called square integrable if $\sup_{t \in \mathcal{T}} E(M(t)^2) < \infty$. A local martingale is a process M such that an increasing sequence of stopping times T_n exists, $P(T_n \ge t) \to 1$ as $n \to \infty$ for all $t \in \mathcal{T}$, such that the stopped process $[T_n > 0]M^{T_n}$ are martingales for each n (here $X^T(t) = X(T \wedge t)$). A local square integrable martingale is a process M as above such that the localizing sequence can be chosen making $[T_n > 0]M^{T_n}$ a square integrable martingale.

A class of processes complementary to martingales is the class of predictable processes. A stochastic process H is called predictable if, as a function of $(t, \omega) \in \mathcal{T} \times \Omega$,

it is measurable with respect to the σ – algebra on $\mathcal{T} \times \Omega$ generated by the left-continuous adapted processes. Any left-continuous adapted process is predictable.

There is an important orthogonality between martingales and finite variation predictable processes. This is due to the fact that if a process is at the same time both a local martingale and a predictable finite variation process, then it is trivial or constant process. Suppose that X is a cadlag adapted process. We say that $\tilde{X}$ is the compensator of X if $\tilde{X}$ is a predictable, cadlag, and finite variation process such that $X - \tilde{X}$ is a local martingale, zero at time 0. If a compensator exists it is unique, by Doob-Meyer decomposition theorem. A semi-martingale is the sum of a local martingale and a cadlag adapted finite variation process. All (local) sub-martingales and super-martingales have compensators. In particular, non-decreasing, non-negative locally integrable cadlag processes have compensators, because such processes are sub-martingales.

Suppose M, N are local square integrable martingales. The square of a local square integrable martingale is a local sub-martingale and also has a non-decreasing compensator denoted by $< M >$. Also the product, $MN = \frac{1}{4}\{(M+N)^2 - (M-N)^2\}$, is the difference of two local sub-martingales, therefore it has a compensator denoted by $\langle M, N \rangle$. By definition, $< M >$ and $\langle M, N \rangle$ are the unique finite variation cadlag predictable process such that, $M^2 - < M >$ and $MN - \langle M, N \rangle$ are local martingales, zero at time 0. The process $< M >$ is called the predictable variation process of M, and $\langle M, N \rangle$ the predictable covariation process of M and N.

Theorem (see Theorem II.3.1, Andersen et al (1993)[25])

Suppose M is a finite variation local square integrable martingale, H is a predictable process, and $\int H^2 d < M >$ is just locally finite (automatically true of H is locally bounded). Then $\int HdM$ is a local square integrable martingale, and $< \int HdM >= \int H^2 d < M >$.

The predictable process H is locally bounded if it is left continuous and also if it is right-continuous.

Formulas for predictable and optional covariation processes of stochastic integrals follow the same form. We have $< \int HdM , \int KdN >= \int HKd < M , N >$.

Fubini-type Theorem

Let $(s,u) \rightarrow H(s,u)$, $(s,u) \in (\mathcal{T} \times \mathcal{T})$ be a bounded, $\mathcal{B} \times \mathcal{B}$ measurable function, where $\mathcal{B}$ is the set of Borelians on $\mathcal{T}$. Let μ be a finite measure on the space $(\mathcal{T}, \mathcal{B})$. Then for every $t \in \mathcal{T}$, $A \in \mathcal{B}$:

$$\int_A [\int_0^t H(s,u) dM(u)] \mu(ds) = \int_0^t [\int_A H(s,u) d\mu(ds)] dM(u) \text{ for almost all } \omega \in \Omega .$$

For more general versions of this theorem and their proofs see Protter (1990)[117].

Functional Delta-Method

A concept of differentiability that allows generalization of the usual Delta Method is the one of Hadamard or compact differentiability. Let B_1 and B_2 two normed vector spaces.

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 $h_n \to h \in B_1$,

$$a_n(\varphi(\theta + a_n^{-1}h_n) - \varphi(\theta)) \to d\varphi(\theta).h \text{ as } n \to \infty.$$

Here $d\varphi(\theta)$ is called the derivative of φ at the point θ.

Next we define the concept of weak convergence in normed vector spaces. Let $(B, \|.\|)$ 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 $\mathcal{B}' = \mathcal{B}''$.

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 $X_n \xrightarrow{\mathcal{D}} X$ if and only if $E(f(X_n)) \xrightarrow{\mathcal{D}} Ef(X)$ for all bounded, norm-continuous, $\mathcal{B}$-measurable $f: B \to \mathbf{R}$.

Theorem (see Theorem 3, Gill (1989)[58])

Suppose $\varphi: B_1 \to B_2$ is compactly or Hadamard 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 elements of B_1 such that $Z_n = \sqrt{n}(X_n - \mu) \xrightarrow{\mathcal{D}} Z$ 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. Then

(1) $[\sqrt{n}(X_n - \mu), \sqrt{n}(\varphi(X_n) - \varphi(\mu)) - d\varphi(\mu).\sqrt{n}(X_n - \mu)] \xrightarrow{\mathcal{D}} (Z,0)$ in $B_1 \times B_2$ and consequently

(2) $\sqrt{n}(\varphi(X_n) - \varphi(\mu)) - d\varphi(\mu).\sqrt{n}(X_n - \mu) \xrightarrow{P} 0$ and

(3) $\sqrt{n}(\varphi(X_n) - \varphi(\mu)) \xrightarrow{\mathcal{D}} d\varphi(\mu).Z$.

Measurability of $d\varphi: B_1 \to B_2$ can often be shown to follow from measurability of φ (see Lemmas 4.4.3 and 4.4.4, Van der Vaart (1988)[61]).

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)[58])

Consider $x \in E \subset B_1$ and $\varphi: B_1 \to B_2$. Suppose there exists a continuous linear map $d\varphi(x): B_1 \to B_2$ such that for all $t_n \to 0$ ($t_n \in \mathbf{R}$) and $h_n \to h \in B_1$ such that $x_n = x + t_n h_n \in E$ for all n, we have:

$$t_n^{-1}(\varphi(x + t_n h_n) - \varphi(x)) \to d\varphi(x).h \text{ as } n \to \infty.$$

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.

Suppose we endow $D[0,\tau]$ with $\|.\|_\infty$, the supremum norm, and $(D[0,\tau])^p$ with the maximum-supremum norm (i.e. for $x \in (D[0,\tau])^p$, $\| x \| = \max(\| x_1 \|_\infty, \ldots \| x_p \|_\infty)$). Under this norm $(D[0,\tau])^p$ is Banach, but not separable. However under the Skorohod topology, these spaces are Banach and separable. 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.

Results on Ito integration

Let $(\Omega, \mathcal{F}, P)$ be a probability space with a filtration $(\mathcal{F}_t, t \in \mathcal{T})$, where $\mathcal{F}_0$ contains all null sets of $\mathcal{F}$. A continuous *k*-dimensional vector martingale $M = (M(t), t \in \mathcal{T})$, $\mathcal{T} = [0,\tau)$, $\tau \in \bar{\mathbf{R}}$ is called Gaussian if:

i) $< M >$ is a continuous deterministic $k \times k$ positive semidefinite matrix valued function on $\mathcal{T}$, with positive definite increments, zero at time 0.

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

Let $\mathcal{H}^2$ be the set of all adapted processes X on $((\Omega, \mathcal{F}, P), (\mathcal{F}_t, t \in \mathcal{T}))$ satisfying $E(\int_0^t X^2(s) < M(ds) >) < \infty$ for all $t \in \mathcal{T}$ and let M be a continuous Gaussian martingale on this space. For any fixed t, one can define the Ito integral $I_t(X) = \int_0^t X dM$ for $X \in \mathcal{H}^2$.

Properties of the Ito integral:

Let $X, Y \in \mathcal{H}^2$ and let $0 \le s < u < t$, $s,u,t \in \mathcal{T}$. Then

i) $\int_s^t X dM = \int_s^u X dM + \int_u^t X dM$ for a.a. ω;

ii) $\int_s^t (cX + Y) dM = c \int_s^t X dM + \int_s^t Y dM$ for a.a. ω;

iii) $\int_s^t X dM$ is $\mathcal{F}_t$-measurable;

iv) If $X(t,\omega) = X(t)$ only depends on t (so $X(t)$ is deterministic) then

$I_t(X) = \int_0^t X dM$ is normally distributed, with mean 0 and variance $\int_0^t X^2(s) < M(ds) >$.

BIBLIOGRAPHY

1. Heffler S, Smith S, Keehan S, Clemens M, Zezza M, Truffer C. Health Spending Projections Through 2013. *Health Affairs.* 2004.

2. Willan AR, Chen EB, Cook RJ, Lin DY. Incremental net benefit in randomized clinical trials with quality-adjusted survival. *Statistics in Medicine.* Feb 15 2003;22(3):353-362.

3. Willan AR, Lin DY, Cook RJ, Chen EB. Using inverse-weighting in cost-effectiveness analysis with censored data. *Statistical Methods in Medical Research.* Dec 2002;11(6):539-551.

4. Lin DY. Linear regression of censored medical costs. *Biostatistics.* 2000;1:35-47.

5. Lin DY. Regression analysis of incomplete medical cost data. *Statistics in Medicine.* Apr 15 2003;22(7):1181-1200.

6. Gardiner JC, Bradley CJ, Huebner M. The cost-effectiveness ratio in the analysis of health care programs. In: Rao CR, Sen PK, eds. *Handbook of Statistics, Bioenvironmental and Public Health Statistics.* Vol 18. New York: North-Holland; 2000:841-869.

7. Gardiner JC, Indurkhya A, Luo Z. The performance of estimation procedures for cost-effectiveness ratios. In: Balakrishnan N, ed. *Advances on Methodological and Applied Aspects of Probability and Statistics.* New York: Gordan Breach; 2001:547-559.

8. 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.

9. Baser O, Bradley CJ, Gardiner JC. Estimation from censored medical costs. *In Review.* 2002.

10. Baser O, Bradley CJ, Gardiner JC, Given CC. Testing and correcting for non-random selection bias due to censoring: An application to medical costs. *Health Services and Outcomes Research Methodology.* 2003;4(2):93-107.

11. Baser O, Gardiner JC, Bradley CJ, Yuce H, Given CC. Longitudinal analysis of censored medical costs data. *In Review.* 2002.

12. Commonwealth of Australia: Guidelines for the pharmaceutical industry on preparation of submissions to the Pharmaceutical Benefits Advisory Committee: including major submissions involving economic analyses. 1995. Located at: Canberra: Australian Government Publishing Service.

13. National Institute for Clinical Excellence. Technology appraisals work programme, new guidance documents. 2001.

14. Canadian Coordinating Office for Health Technology Assessment. Guidelines for economic evaluation of pharmaceuticals:. *Canada. 2nd ed. Ottawa: Canadian Coordinating Office for Health Technology Assessment (CCOHTA).* 1997.

15. Gold MR, Siegel JE, Russell LB, Weinstein MC, eds. *Cost-Effectiveness in Health and Medicine*. New York: Oxford University Press; 1996.

16. Stinnett AA, Mullahy J. Net Health Benefits: A new framework for the analysis of uncertainty in cost-effectiveness analysis. *Medical Decision Making.* 1998;18(Supplemental):S68-S80.

17. Manning WG, Mullahy J. Estimating log models: to transfrom or not to trasform. *Journal of Health Economics.* 2001;20:461-494.

18. Manning WG. The logged dependent variable, heteroscedasticity, and the retransformation problem. *Journal of Health Economics.* 1998;17:283-295.

19. Duan N. Smearing estimate: A nonparametric retransformation method. *Journal of the American Statistical Association.* 1983;78(383):605-610.

20. Ai C, Norton EC. Standard errors for the retransformation problem with heteroscedasticity. *Journal of Health Economics.* 2000;19:697-718.

21. Mullahy J. Much ado about two: Reconsidering retransformation and the two-part model in health econometrics. *Journal of Health Economics.* 1998;17:247-281.

22. Lucas JW SJ, Benson V. *Summary health statistics for U.S. Adults: National Health Interview Survey 2001.* Washington, DC: Centers for Disease Control and Prevention, National Center for Health Statistics; 2004. Library of Congress Catalog Number 362.

23. Hay JW, Olsen RJ. Let them eat cake: A note comparing alternative models of the demand for medical care. *Journal of Business & Statistics.* 1984;2(3):279-282.

24. 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.

25. Andersen PK, Borgan O, Gill RD, Keiding N. *Statistical Models Based on Counting Processes.* New York: Springer-Verlag; 1993.

26. Praestgaard J. Nonparametric estimation of actuarial values. *Scandinavian Actuarial Journal.* 1991;2:129-143.

27. Norberg R. Ruin problems with assets and liabilities of diffusion type. *Stochastic Processes and Their Applications.* Jun 1 1999;81(2):255-269.

28. Norberg R. Differential-Equations for Moments of Present Values in Life-Insurance. *Insurance Mathematics & Economics.* Oct 1995;17(2):171-180.

29. Norberg R. A Time-Continuous Markov-Chain Interest Model with Applications to Insurance. *Applied Stochastic Models and Data Analysis.* Sep 1995;11(3):245-256.

30. Lin DY, Feuer EJ, Etzioni R, Wax Y. Estimating medical costs from incomplete follow-up data. *Biometrics.* 1997;53:419-434.

31. Bang H, Tsiatis AA. Estimating medical costs with censored data. *Biometrika.* 2000;87(2):329-343.

32. Strawderman RL. Estimating the mean of an increasing stochastic process at a censored stopping time. *Journal of the American Statistical Association.* Dec 2000;95(452):1192-1208.

33. Wooldridge JM. Inverse Probability Weighted Estimation for General Missing Data Problems. *Working Paper Michigan State University.* April 2003.

34. Wooldridge JM. *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA: MIT Press; 2002.

35. Zhou XH, Tu WZ. Interval estimation for the ratio in means of log-normally distributed medical costs with zero values. *Computational Statistics & Data Analysis.* Dec 28 2000;35(2):201-210.

36. Zhou XH, Tu WZ. Confidence intervals for the mean of diagnostic test charge data containing zeros. *Biometrics.* Dec 2000;56(4):1118-1125.

37. Zhou XH, Stroupe KT, Tierney WM. Regression analysis of health care charges with heteroscedasticity. *Journal of the Royal Statistical Society Series C-Applied Statistics.* 2001;50:303-312.

38. Zhou XH, Melfi CA, Hui SL. Methods for comparison of cost data. *Annals of Internal Medicine.* 1997;127(8):752-756.

39. Zhou XH, Li CM, Gao SJ, Tierney WM. Methods for testing equality of means of health care costs in a paired design study. *Statistics in Medicine.* Jun 15 2001;20(11):1703-1720.

40. Zhou XH. Inferences about population means of health care costs. *Statistical Methods in Medical Research.* Aug 2002;11(4):327-339.

41. Zhou XH, Gao SJ. One-sided confidence intervals for means of positively skewed distributions. *American Statistician.* May 2000;54(2):100-104.

42. Maynard C, Chapko MK, Every NR, Martin DC, Ritchie JL, Department of Medicine DoVASW. Coronary angioplasty outcomes in the Healthcare Cost and Utilization Project, 1993-1994. *The American journal of cardiology.* Apr 1 1998;81(7):848-852.

43. Ritchie JL, Maynard C, Chapko MK, Every NR, Martin DC, Department of Medicine DoVASWUSAruwe. Association between percutaneous transluminal coronary angioplasty volumes and outcomes in the Healthcare Cost and Utilization Project 1993-1994. *The American journal of cardiology.* Feb 15 1999;83(4):493-497.

44. Brooks JM, McClellan M, Wong HS. The marginal benefits of invasive treatments for acute myocardial infarction: does insurance coverage matter? *Inquiry : a journal of medical care organization, provision and financing.* Spring 2000;37(1):75-90.

45. Brooks SE, Ahn J, Mullins CD, Baquet CR, D'Andrea A. Health care cost and utilization project analysis of comorbid illness and complications for patients undergoing hysterectomy for endometrial carcinoma. *Cancer.* Aug 15 2001;92(4):950-958.

46. Howard TS, Hoffman LH, Stang PE, Simoes EA, Galt Associates SVUSA. Respiratory syncytial virus pneumonia in the hospital setting: length of stay, charges, and mortality. *The Journal of pediatrics.* Aug 2000;137(2):227-232.

47. Watanabe CT, Maynard C, Ritchie JL, Department of Medicine UoWSWUSA. Short-term outcomes after percutaneous coronary intervention: effects of stenting and institutional volume shifts. *American heart journal.* Aug 2002;144(2):310-314.

48. Polverejan E. *Regression Models for Analysis of Medical Costs* [PhD]. East Lansing: Statistics & Probability, Michigan State University; 2001.

49. Polverejan E, Gardiner JC, Bradley CJ, Holmes-Rovner M, Rovner D, Department of Epidemiology MSUELMIUSA. Estimating mean hospital cost as a function of length of stay and patient characteristics. *Health economics.* Nov 2003;12(11):935-947.

50. Gardiner JC, Luo Z, Bradley CJ, Polverejan E, Holmes-Rovner M, Rovner D. Longitudinal assessment of Cost in Health Care Interventions. *Health Services &Outcomes Research Methodology.* 2002;3:149-168.

51. Rice N, Jones A. Multilevel models and health economics. *Health Economics.* Nov-Dec 1997;6(6):561-575.

52. Carey K. Hospital Length of Stay and Cost: A Multilevel Modelling Analysis. *Health Services &Outcomes Research Methodology.* 2002;3:41-56.

53. Carey K. A multilevel modelling approach to analysis of patient costs under managed care. *Health Economics.* Jul 2000;9(5):435-446.

54. Goldstein H, Browne W, Rasbash J. Multilevel modelling of medical data. *Statistics in medicine.* Nov 15 2002;21(21):3291-3315.

55. Jacobsen M. *Statistical Analysis of Counting Processes.* New York: Springer-Verlag; 1982.

56. Arijas E, Haara P. A marked point process approach to censored failure time data with complicated covariates. *Scandinavian Journal of Statistics.* 1984;11:193-209.

57. Andersen PK, Gill RD. Cox's regression model for counting processes: A large sample study. *Annals of Statistics.* 1982;10(4):1100-1120.

58. Gill RD. Non-and semi parametric maximum likelihood estimators and the von Mises method (Part I). *Scandinavian Journal of Statistics.* 1989;16:97-128.

59. 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.

60. Huang YJ, Louis TA. Nonparametric estimation of the joint distribution of survival time and mark variables. *Biometrika.* Dec 1998;85(4):785-798.

61. Vaart AWvd. Asymptotic Statistics. *Cambridge University Press.* December 1998 1998, 460 pages.

62. Newey WK, McFadden D. Large sample estimation and hypothesis. *Handbook of econometrics.* 1994;IV:2111-2245.

63. Wooldridge JM. *Econometric Analysis of Cross Section and Panel Data.* Cambridge, MA: MIT Press; 2001.

64. Volpp KG, Williams SV, Waldfogel J, et al. Market reform in New Jersey and the effect on mortality from acute myocardial infarction. *Health services research.* Apr 2003;38(2):515-533.

65. National Heart Lung and Blood Institute. Facts About Coronary Heart Disease. http://www.nhlbi.nih.gov/health/public/heart/other/chdfacts2.htm.

66. American Heart Association. *Heart Disease and Stroke Statistics -- 2004 Update.* Dallas, Texas 2004.

67. Smith SCJ, Dove JT, Jacobs AK, et al. ACC/AHA guidelines of percutaneous coronary interventions (revision of the 1993 PTCA guidelines)--executive summary. A report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (committee to revise the 1993 guidelines for percutaneous transluminal coronary angioplasty). *J Am Coll Cardiol.* 2001;37(8):2215-2239.

68. American Heart Association. Angioplasty, Percutaneous Transluminal Coronary (PTCA). http://www.americanheart.org/presenter.jhtml?identifier=4454.

69. Mitka M. Drug-Eluting Stents Show Promise. *JAMA.* Feb 11, 2004 2004;291(6):682.

70. Wisconsin Department of Health and Family Services. *Why Charges May Differ Between Facilities*: Division of Health Care Financing, Bureau of Health Information; 2001. Health Care Data Report, 2000 (PHC 5320).

71. 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.

72. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative data. *Journal of Clinical Epidemiology.* 1992;45:613-619.

73. Romano PS, Ross LL, Jollis JG. Adapting a clinical comorbidity index for use with ICD-9-CM administrative data: differing perspectives. *Journal of Clinical Epidemiology.* 1993;46:1075-1079.

74. Ghali WA, Hall RE, Rosen AK, Ash AS, Moskowitz MA. Searching for an improved clinical comorbidity index for use with ICD-9-CM administrative data. *J Clin Epidemiol.* Mar 1996;49(3):273-278.

75. Southern DA, Quan H, Ghali WA. Comparison of the Elixhauser and Charlson/Deyo methods of comorbidity measurement in administrative data. *Med Care.* Apr 2004;42(4):355-360.

76. Quan H, Parsons GA, Ghali WA. Validity of information on comorbidity derived rom ICD-9-CCM administrative data. *Med Care.* Aug 2002;40(8):675-685.

77. Taira DA, Seto TB, Siegrist R, Cosgrove R, Berezin R, Cohen DJ. Comparison of analytic approaches for the economic evaluation of new technologies alongside multicenter clinical trials. *American Heart Journal.* 2003/3 2003;145(3):452-458.

78. White H. A Heteroskedasticity-Consistent Covariance-Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica.* 1980;48(4):817-838.

79. Liang KY, Zeger SL. Longitudinal Data-Analysis Using Generalized Linear-Models. *Biometrika.* Apr 1986;73(1):13-22.

80. Diggle PJ, Liang KY, Zeger SL. *The Analysis of Longitudinal Data.* Oxford, UK: Oxford University Press; 1994.

81. Weintraub WS, Mauldin PD, Becker E, Kosinski AS, King SB, III. A Comparison of the Costs of and Quality of Life After Coronary Angioplasty or Coronary Surgery for Multivessel Coronary Artery Disease : Results From the Emory

Angioplasty Versus Surgery Trial (EAST). *Circulation.* November 15, 1995 1995;92(10):2831-2840.

82. Faxon D. Myocardial Revascularization in 1997: Angioplasty Versus Bypass Surgery. *American Family Physician.* 1997;56(5):1409-1420.

83. Hlatky MA, Rogers WJ, Johnstone I, et al. Medical Care Costs and Quality of Life after Randomization to Coronary Angioplasty or Coronary Bypass Surgery. *N Engl J Med.* January 9, 1997 1997;336(2):92-99.

84. Sculpher MJ, Buxton MJ, Seed P, et al. Health service costs of coronary angioplasty and coronary artery bypass surgery: the Randomised Intervention Treatment of Angina (RITA) trial. *The Lancet.* 1994/10/1 1994;344(8927):927-930.

85. Cowper PA, DeLong ER, Peterson ED, et al. Geographic variation in resource use for coronary artery bypass surgery. *Medical Care.* 1997;35(4):320-333.

86. Kozak L, Hall M, Owings M. National Hospital Discharge Survey: 2000 Annual Summary with detailed diagnosis and procedure data. *National Center for Health Statistics. Vital Health Statistics.* 2002;13(153).

87. Pokras R, Kozak L, McCarthy E, Graves E. Trends in Hospital Utilization: United States 1965-1986. *National Center for Health Statistics, Vital Health Statistics.* 1989;13(101).

88. Gillum BS GE, Kozak LJ. Trends in Hospital Utilization: United States, 1988–92. *National Center for Health Statistics, Vital Health Statistics.* 1996;13(124).

89. Blumenthal D. Controlling Health Care Expenditures. *N Engl J Med.* March 8, 2001 2001;344(10):766-769.

90. Morris CR. *Too much of a good thing? Why health care spending won't make us sick.* New York: Century Foundation Press; 2000.

91. Anderson GF, Hussey PS. Population aging: a comparison among industrialized countries. *Health Aff (Millwood).* May-Jun 2000;19(3):191-203.

92. Organisation for Economic Co-operation and Development. OECD Health Data 2004. 2004.

93. Anderson GF, Hurst J, Hussey PS, Jee-Hughes M. Health spending and outcomes: Trends in OECD countries, 1960-1998. *Health Affairs.* May-Jun 2000;19(3):150-157.

94. World Health Organization. *Health systems: improving performance.* Geneva 2000.

95. 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.

96. Gardiner J, Holmes-Rovner M, Rovner D, Goddeeris J, Kupersmith J. Cost-effectiveness ratios in multi-state Markov models. Paper presented at: Biometrics Society, 1996; Richmond, VA.

97. Gardiner JC. The asymptotic distribution of mortality rates in competing risks analyses. *Scandinavian Jouranl of Statistics.* 1982;9:31-36.

98. Gardiner JC. Properties of occurrence-exposure rates in multiple decrement theory. *Journal of Statistical Planning & Inference.* 1983;8:301-314.

99. Gardiner JC, Huebner M, Jetton J, Bradley CJ. Power assessments and confidence intervals for cost-effectiveness analyses. *Medical Decision Making.* 1998;18(4):465.

100. Gardiner JC, Huebner M, Jetton J, Bradley CJ. Power and sample size assessments for tests of hypotheses on cost-effectiveness ratios. *Health Economics.* 2000.

101. Lin DY. Proportional means regression for censored medical costs. *Biometrics.* 2000;56:775-778.

102. Willan AR, Lin DY. Incremental net benefit in randomized clinical trials. *Statistics in Medicine.* Jun 15 2001;20(11):1563-1574.

103. Willan AR. Analysis, sample size, and fewer for estimating incremental net health benefit from clinical trial data. *Controlled Clinical Trials.* Jun 2001;22(3):228-237.

104. 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.

105. Glasziou PP, Simes RJ, Gelber RD. Quality adjusted survival analysis. *Statistics in Medicine.* 1990;9:1259-1276.

106. Antiarrhythmics Versus Implantable Defibrillators (AVID)--rationale, design, and methods. *Am J Cardiol.* Mar 1 1995;75(7):470-475.

107. Connolly S, Gent M, Roberts R. Canadian Implantable Defibrillator Study (CIDS). A randomized trial of the implantable cardioverter defibrillator against amiodarone. *Circulation.* 2000;101:1297-1302.

108. Gardiner CJ, Sirbu MC, Rahbar M. Update on statistical power and sample size assessments for cost-effectiveness studies (Review). *Expert Rev. Pharmacoeconomics Outcomes Res.* 2004;4(1):89-98.

109. Briggs AH, Gray AM. Power and sample size calculations for stochastic cost-effectiveness analysis. *Medical Decision Making.* 1998;18(supplemental):S81-S92.

110. Sacristán JA, Day SJ, Navarro O, Ramos J, Hernández JM. Use of confidence intervals and sample size calculations in health economic studies. *The Annals of Pharmacotherapy.* 1995;29:719-725.

111. Laska EM, Meisner M, Siegel C. Power and sample size in cost-effectiveness analysis. *Med Decis Making.* 1999;19:339-343.

112. Willan AR, Obrien BJ. Confidence intervals for cost-effectiveness ratios: An application of Fieller's theorem. *Health Economics.* Jul-aug 1996;5(4):297-305.

113. Willan AR, O'Brien BJ. Sample size and power issues in estimating incremental cost-effectiveness ratios from clinical trials data. *Health Economics.* 1999;8:203-211.

114. Thompson S. Statistical issues in cost-effectiveness analyses. *Stat Methods Med Res.* Dec 2002;11(6):453-454.

115. Harrison JM. *Brownian motion and stochastic flow systems.* New York: John Wiley; 1985.

116. Oskendal B. *Stochastic differential equations.* New York: Springer-Verlag; 1995.

117. Protter P. *Stochastic integration and differential equations.* New-York: Springer-Verlag; 1990.