Statistical inference for residual time quantiles in regression models for censored time-to-event data

Luis Crouch

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2013

Reading Committee:

Ying Qing Chen, Chair

Susanne May

Jim Hughes

Program Authorized to Offer Degree:
UW Department of Biostatistics

University of Washington

Abstract

Statistical inference for residual time quantiles in regression models for censored time-to-event data

Luis Crouch

Chair of the Supervisory Committee:
Professor Ying Qing Chen
Department of Biostatistics

In this dissertation, we set out to develop new methods for the analysis of time-to-event data. In particular, we are concerned with residual time, or the time remaining to an event after a certain amount of time has passed since time zero. We develop methods to estimate quantiles of residual time under a few different settings: the Cox proportional hazards model (with fixed and with external time-varying covariates) and the additive hazards model. In each setting, we consider point estimation, asymptotic properties, variance estimation, confidence interval construction, and inference. We also perform simulations to demonstrate our estimators' performance and provide examples of their application to sample data sets. We finish by discussing the many opportunities for future work and expansion of our methods to address limitations or allow application in a wider array of settings.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

First and foremost I would like to thank my committee members, particularly my adviser Ying Chen. I would also like to acknowledge the support I received from both the National Institutes of Health through the Cancer Epidemiology and Biostatistics Training Grant and the ARCS Foundation.

DEDICATION

This dissertation is dedicated to my wife, Grace.

Chapter 1

INTRODUCTION

1.1 Objective

While survival analysis is fundamentally about quantifying (differences in) time to a particular event, many analyses of time-to-event data are centered around fitting a proportional hazards model as described by Cox (1972). The proportional hazards model offers many advantages: the ability to handle censored data, a lack of parametric assumptions about the underlying shape of the hazard distribution, and output (in the form of hazard ratios) with good statistical properties. Unfortunately, hazard ratios are relatively unintuitive and may be difficult for physicians to communicate to patients in a clinical setting. Differences in survival times may be more easily understood. In particular, estimates of typical survival times specific to a patient's characteristics (i.e. covariate-specific) provide a compelling summary measure.

Survival times may be generalized to residual times, or the amount of survival time remaining at a given time t, assuming survival to that point. Residual times maintain the intuitive properties of survival times while being more flexible and informative. For example, many diseases have an initial high-risk (or -hazard) period resulting in low estimates of survival time starting at a sensible "time zero" such as time of infection or time of presentation of symptoms. Patients who survive this initial high-risk period, along with their physicians, will still be interested in prognosis. At this point, however, survival time calculated from time zero will not be appropriate; instead, one must consider residual time, which in this case may be much higher than the initial estimate of survival time.

The two most commonly used statistics to characterize typical values are the mean and median. While mean estimation is often simpler (at least in parametric settings), this is not

the case when data consist of censored survival times, as the censoring prevents straightforward estimation of the mean. We therefore investigate median (and other quantiles of) residual time, as censoring is of a lesser concern. Additionally, specific quantiles are generally more robust to outliers than is the mean. This is particularly important in survival analysis, as survival time distributions may be heavy-tailed.

We initially base our estimator for quantiles of residual time on the Cox proportional hazards model. This is useful for several reasons. Most importantly, it may allow for translation of results from fitting such models (hazard ratios and estimates of baseline hazard) to more intuitive quantities. It is also convenient for method development and applications. In terms of methods development, the formulation of our estimator and the characterization of its asymptotic properties can draw from the existing literature on proportional hazards models. In terms of applications, having an estimator based on such a commonly used model may reduce barriers to adoption.

Of course, not all data satisfy the proportional hazards assumption as in the Cox model. We therefore also develop methods for an estimator based on the additive hazards model. While not as commonly employed in survival analysis as the Cox proportional hazards model, the additive hazards model is one of the more frequently used alternatives and, importantly, relies of different assumptions.

The specific aims of our research are as follows.

- Develop point estimators for quantiles of residual time conditional on specific covariate values based on the Cox proportional hazards and additive hazards models
- Characterize the asymptotic properties of said estimators and provide variance estimators that can be used in finite sample settings
- Develop a way to quantify differences in median residual times (for different sets of covariates) and provide a mechanism for testing the statistical significance of said differences
- Generalize our estimator for time-dependent covariates in the Cox proportional haz-

ards setting

- Demonstrate the use of all our methods via simulation and application to real data

1.2 Background

In Dabrowska and Doksum (1987) estimators for mean and quantiles of survival time were developed for the Cox model. The application of such estimates may be clinically limited, however, by their beginning at "time zero." For example, an estimate of median survival time for a patient diagnosed with a disease with high initial hazard that drops over time may not be relevant when considering their outlook after the initial spike in hazard. It may therefore be useful to consider residual time, or the amount of survival time remaining at a given time t, assuming survival to t.

Quantification and analysis of survival (and residual) times has been approached in numerous ways. Perhaps the most fundamental difference in approaches is the choice of mean or median as the summary measure, both of which have their advantages and limitations. Other differences arise in estimation procedures: parametric, semiparametric, or nonparametric; inclusion or not of covariate effects; direct or indirect modeling of said effects; etc.

In an early paper on estimation of mean residual time, Yang (1978) approaches the problem with the motivation of calculating life expectancy. She proposes an estimator for mean residual time based on the empirical survival function and establishes its asymptotic properties. This estimator does not, however, allow for censoring or covariate effects, limiting its applicability.

An important development in estimation of mean residual time was the proposal and initial characterization of the proportional mean residual time model by Oakes and Dasu (1990). Maguluri and Zhang (1994) extended this work, noting that proportional mean residual time implies, for a fixed time, proportional hazards for the residual times. They also propose two ways to estimate the ratio of mean residual times: one based on the exponential regression model and one based on Cox's model, using their result about propor-

tional hazards for residual times. The asymptotic properties of both estimators are derived, and small sample performance is demonstrated through simulation. While both estimators proposed by Maguluri and Zhang allow for the modeling of a covariate effect, an important distinction from the earlier work of Yang (1978), they still do not explicitly account for censoring.

Yuen et al. (2003) extended the methods of Maguluri and Zhang (1994) by developing a goodness-of-fit test for the latter's (Cox based) mean residual time regression model. While the limiting distribution of the test statistic they introduce is "asymptotically intractable," Yuen et al. (2003) provide a pair of resampling methods for estimating its variance in finite samples: "random symmetrization" and the bootstrap.

A further extension of Maguluri and Zhang's (1994) work was carried out by Chen et al. (2005), who proposed a way to estimate covariate effects on mean residual times in the presence of censoring. Additionally, under further assumptions, they provide an estimator for baseline mean residual time, and therefore covariate-specific mean residual time, that allows for censoring. The wide prevalence of censored data, particularly in the biomedical setting, makes this a crucial development.

More recently this line of semiparametric models has been extended for use with general link functions via generalized estimating equation methods (Sun and Zhau, 2010) and to account for the possibility of covariate effects that vary across time (Sun et al., 2012). By allowing for different link functions, these methods allow for more flexible relationships between the covariates and the mean residual times than do those that assume proportional mean residual times.

Qin and Zhao (2007) approach mean residual time estimation in a different way. Instead of assuming a proportional mean residual time model they use empirical likelihood, basing part of their estimator on work by Kaplan and Meier (1958). While appealing in its allowance for censoring and reduced modeling assumptions, this method does not allow estimation of covariate effects.

For additional, more theoretical, development of mean residual time models, see Csörgő

and Zitikis (1996) and Nanda et al. (2006), among others.

While it took quite some time for mean residual time models to allow for censoring, median residual models allowed for it from the beginning, as discussed in a seminal paper by Schmittlein and Morrison (1981). In their paper, Schmittlein and Morrison propose median residual life as an alternative to mean residual life, characterize some of its properties, and provide an example of its application.

Many methods for estimating residual time quantiles are generalizations of those for estimating survival time quantiles. It is therefore useful to examine some key developments regarding the latter. Early estimators for survival time quantiles were based on empirical functions (Cheng, 1984), typically using the inverse of the product limit estimator for survival proposed by Kaplan and Meier (1958). Padgett (1986) suggested using a kernel-type estimator to allow for a smoother fit and showed it to have similar asymptotic properties to those based directly on empirical functions.

Continuing to base their work on the inverse of the product limit estimator, Wang and Hettmansperger (1990) developed methods for calculating one-sample confidence intervals for median survival time as well as two-sample tests and confidence intervals for the difference between two median survival times. Su and Wei (1993) provide a method for comparing median survival times that is slightly more general, allowing for inference on either the ratio or difference of two median survival times.

While still focusing on median survival times, Ying et al. (1995) and Yang (1999) proposed regression frameworks to allow for more general estimation of covariate effects, as opposed to simply comparing two groups' median survival times. Both modeled median survival time directly as a linear combination of covariates, though their estimation methods for the covariate effects differed. Portnoy (2003) presents a regression method that not only relaxes assumptions in Ying et al. (1995) and Yang (1999) but also allows estimation of general quantiles.

Methods for residual time quantiles have seen somewhat less development. An early paper by Csörgő and Csörgő (1987) laid out important theoretical results, including the

asymptotic distribution for the emprical quantile residual life function, though these assumed no censoring. Song and Cho (1995) provide some additional theory, giving conditions under which the underlying survival distribution is uniquely determined by the quantile residual life distributions.

Analogously to Padgett (1986), Alam and Kulasekera (1993) discuss a kernel-type and an empirical estimator for the residual time quantile function, showing that they are convergent. However, these methods are still limited by not allowing for censoring. More recently, Jeong et al. (2008) proposed nonparametric methods for estimating median residual time based on inverting Kaplan-Meier estimators, thereby accounting for censoring. They also provide a two-sample test which reduces to that of Su and Wei (1993) at $t = 0$. In a follow-up paper, Jung et al. (2009) provide a regression model that allows for modeling of covariate effects on general quantile residual time that "can be viewed as a generalization of the median regression model [of Ying et al. (1995)]." As in Ying et al.'s (1995) model, quantile residual time is modeled directly as a linear combination of covariates.

A different regression model is proposed by Ma and Yin (2010). In contrast to Jung et al.'s (2009) model, Ma and Yin's (2010) allows for estimation of quantiles of residual times, not just covariate effects on them. This is accomplished in a two-step process, where covariate effects are estimated based on all data (i.e. at $t = 0$, or using survival times rather than residual times), and these estimates are then used in estimation of quantile residual time at later time points. Another paper by Ma and Wei (2012) proposes regression methods that are more closely based on Jung et al.'s (2009) but allow for time-varying covariate effects modeled via spline smoothing.

1.3 Organization

This dissertation is organized as follows. In Chapter 2, we review some important results and developments in the survival analysis literature. In Chapter 3, we consider the Cox proportional hazards model: we develop an estimator for covariate-specific quantiles of residual time, characterize its behavior, determine its asymptotic properties, show how to

carry out comparisons with other covariate-specific quantiles of residual time, and demonstrate the estimator's performance via simulation. In Chapter 4, we extend this estimator to allow for time-varying covariates. In Chapter 5, we consider the additive hazards model, with similar organization to our consideration of the Cox proportional hazards model. Finally, in Chapter 6, we summarize how some of the limitations of our methods might be addressed in future work.

Chapter 2

LITERATURE REVIEW

In this Chapter we review some of the important theoretical developments in survival analysis, particularly as they pertain to our own methods. In Section 2.1, we review Martingale theory as it relates to survival analysis. We then discuss the models on which we'll base our methods, the Cox proportional hazards model and the additive hazards model, in Sections 2.2 and 2.3, respectively.

2.1 Martingale Theory for Counting Processes

For the entirety of this section we refer extensively to results presented in Fleming and Harrington (2005). We assume some familiarity with measure theoretic probability. Our main goal is to present the martingale central limit theorem. However, this requires a substantial amount of preparatory development of terminology and prerequisite results. For the sake of brevity, we comment only minimally.

Initially, we need definitions for counting processes and martingales. These can be found in Definitions 2.1.6 and 2.1.7, respectively. Before we introduce those, we provide several necessary intermediary definitions.

Assume that Ω is a space of outcomes of a random experiment, with each outcome denoted generically by ω, and $\mathcal{F}$ and P are, respectively, a σ-algebra of events and a probability measure on Ω.

Definition 2.1.1. A function Z from Ω to the real line $R = (-\infty, \infty)$ is called a *random variable* or is called *measurable* (relative to $\mathcal{F}$) if

$$\{Z \leq x\} = \{\omega : Z(\omega) \leq x\} \in \mathcal{F},$$

for all x. That is, Z is a measurable mapping from $(\Omega, \mathcal{F}, P)$ to the real line equipped with the Borel σ-algebra $\mathcal{B}$.

Definition 2.1.2. A (real-valued) *stochastic process* is a family of random variables $X = \{X(t) : t \in \Gamma\}$ indexed by a set Γ, all defined on the same probability space $(\Omega, \mathcal{F}, P)$.

Definition 2.1.3. A stochastic process Z is

1. *Integrable* if $\sup_{0 \le t < \infty} E|X(t)| < \infty$;

2. *Square integrable* if $\sup_{0 \le t < \infty} E\{X(t)\}^2 < \infty$;

3. *Bounded* if there exists a finite constant K such that

$$P\left\{\sup_{0 \le t < \infty} |X(t)| < K\right\} = 1.$$

Definition 2.1.4.

1. A family of sub-σ-algebras $\{\mathcal{F}_t : t \ge 0\}$ of a σ-algebra $\mathcal{F}$ is called *increasing* if $s \le t$ implies $\mathcal{F}_s \subset \mathcal{F}_t$. An increasing family of sub-σ-algebras is called a *filtration*.

2. When $\{\mathcal{F}_t : t \ge 0\}$ is a filtration, the σ-algebra $\cap_{h>0}\mathcal{F}_{t+h}$ is usually denoted by $\mathcal{F}_{t+}$. The corresponding limit from the left, $\mathcal{F}_{t-}$, is the smallest σ-algebra containing all the sets in $\cup_{h>0}\mathcal{F}_{t-h}$ and is written $\sigma\{\cup_{h>0}\mathcal{F}_{t-h}\}$ or $\vee_{h>0}\mathcal{F}_{t-h}$.

3. A filtration $\{\mathcal{F}_t : t \ge 0\}$ is *right-continuous* if, for any t, $\mathcal{F}_{t+} = \mathcal{F}_t$.

4. A *stochastic basis* is a probability space $(\Omega, \mathcal{F}, P)$ equipped with a right-continuous filtration $\{\mathcal{F}_t : t \ge 0\}$, and is denoted by $(\Omega, \mathcal{F}, \{\mathcal{F}_t : t \ge 0\}, P)$.

Definition 2.1.5. A stochastic process $X = \{X(t) : t \ge 0\}$ is *adapted* to a filtration if, for every $t \ge 0$, $X(t)$ is $\mathcal{F}_t$-measurable.

Definition 2.1.6. A *counting process* is a stochastic process $\{N(t) : t \geq 0\}$ adapted to a filtration $\{\mathcal{F}_t : t \geq 0\}$ with $N(0) = 0$ and $N(t) < \infty$ a.s., and whose paths are with probability one right-continuous, piecewise constant, and have only jump discontinuities, with jumps of size $+1$.

Definition 2.1.7. Let $X = \{X(t) : t \geq 0\}$ be a right-continuous stochastic process with left-hand limits and $\{\mathcal{F}_t : t \geq 0\}$ a filtration, defined on a common probability space. X is called a *martingale* with respect to $\{\mathcal{F}_t : t \geq 0\}$ if

1. X is adapted to $\{\mathcal{F}_t : t \geq 0\}$,

2. $E|X(t)| < \infty$ for all $t < \infty$

3. $E\{X(t+s)|\mathcal{F}_t\} = X(t)$ a.s. for all $s \geq 0, t \geq 0$.

X is called a *submartingale* if (3) is replaced by

$$E\{X(t+s)|\mathcal{F}_t\} \geq X(t) \text{ a.s.}$$

X is called a *supermartingale* if (3) is replaced by

$$E\{X(t+s)|\mathcal{F}_t\} \leq X(t) \text{ a.s.}$$

Having defined counting processes and martingales, we now proceed to define some properties they can exhibit which will be important in later theorems.

Definition 2.1.8. Let $(\Omega, \mathcal{F}, P)$ be a probability space with a filtration $\{\mathcal{F}_t : t \geq 0\}$. The σ-algebra on $[0, \infty) \times \Omega$ generated by all sets of the form

$$[0] \times A, \quad A \in \mathcal{F}_0$$

and

$$(a, b] \times A, \quad 0 \leq a < b < \infty, A \in \mathcal{F}_a$$

is called the *predictable* σ-algebra for the filtration $\{\mathcal{F}_t : t \geq 0\}$.

Definition 2.1.9. A process X is called *predictable* with respect to a filtration if, as a mapping from $[0, \infty) \times \Omega$ to R, it is measurable with respect to the predictable σ-algebra generated by that filtration. We call X an $\mathcal{F}_t$-predictable process.

Predictable processes play important roles in two key results. The process A in the Doob-Meyer Decomposition Theorem (2.1.1) is predictable, as is the process H in the martingale transform discussed in Theorem 2.1.9. The main use of predictability of a process Q is its $\mathcal{F}_{t-}$-measurability, implying

$$E\{Q(t)|\mathcal{F}_{t-}\} = Q(t) \text{ a.s.}$$

In other words, the value of a process at a fixed time is almost surely the expected value of that process at said fixed time given the history of the process up to that time. That is, given a process's history, we can accurately predict its current value.

Definition 2.1.10. A collection of random variables $\{X_t : t \in \tau\}$, where τ is an arbitrary index set, is *uniformly integrable* if

$$\lim_{n\to\infty} \sup_{t\in\tau} E(|X_t| I_{\{|X_t|>n\}}) = 0.$$

Theorem 2.1.1 (Doob-Meyer Decomposition)**.** *Let X be a right-continuous nonnegative submartingale with respect to a stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t : t \geq 0\}, P)$. Then there exists a right-continuous martingale M and an increasing right-continuous predictable process A such that $EA(t) < \infty$ and*

$$X(t) = M(t) + A(t) \ a.s.$$

for any $t \geq 0$. If $A(0) = 0$ a.s., and if $X = M' + A'$ is another such decomposition with $A'(0) = 0$, then for any $t \geq 0$,

$$P\{M'(t) \neq M(t)\} = 0 = P\{A'(t) \neq A(t)\}.$$

If in addition X is bounded, then M is uniformly integrable and A is integrable.

Corollary 2.1.1. *Let $\{N(t) : t \geq 0\}$ be a counting process adapted to a right-continuous filtration $\{\mathcal{F}_t : t \geq 0\}$ with $EN(t) < \infty$ for any t. Then there exists a unique increasing right-continuous $\mathcal{F}_t$-predictable process A such that $A(0) = 0$ a.s., $EA(t) < \infty$ for any t, and $\{M(t) = N(t) - A(t) : t \geq 0\}$ is a right-continuous $\mathcal{F}_t$-martingale.*

The process A in the Doob-Meyer decomposition is called the *compensator* for the submartingale X. Because the martingale approach to statistical models for counting processes is useful only in situations in which A is known, the Doob-Meyer Decomposition Theorem is a very important result. The uniqueness portion insures that, if the process A can be deduced and $N - A$ can be shown to be a martingale, then one need look no further.

We now present several results that follow from the Doob-Meyer Decomposition Theorem that will be useful later for establishing limiting distributions for processes.

Corollary 2.1.2. *Let M be a right-continuous martingale with respect to a right-continuous filtration $\{\mathcal{F}_t : t \geq 0\}$ and assume $EM^2(t) < \infty$ for any $t \geq 0$. Then there exists a unique increasing right-continuous predictable process $\langle M, M\rangle$, called the* predictable quadratic variation *of M, such that $\langle M, M\rangle(0) = 0$ a.s., $E\langle M, M\rangle(t) < \infty$ for each T, and $\{M^2(t) - \langle M, M\rangle(t) : t \geq 0\}$ is a right-continuous martingale.*

It can be shown that

$$d\langle M, M\rangle(s) = \mathrm{var}\{dM(s)|\mathcal{F}_{s-}\}.$$

Additionally, we can define a *predictable covariation process* $\langle M_1, M_2\rangle$ satisfying

$$d\langle M_1, M_2\rangle(s) = \mathrm{cov}\{dM_1(s), dM_2(s)|\mathcal{F}_{s-}\},$$

whose existence is proved in the following theorem.

Theorem 2.1.2. *Let M_1 and M_2 be two right-continuous martingales with respect to a stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t : t \geq 0\}, P)$, and assume $E\{M_i^2(t)\} < \infty$ for $t \geq 0$ and $i = 1, 2$. Then there exists a right-continuous process $\langle M_1, M_2\rangle$, called a* predictable covariation process*, with $\langle M_1, M_2\rangle(0) = 0$, $E\langle M_1, M_2\rangle(t) < \infty$, such that*

1. *$\langle M_1, M_2 \rangle$ is the difference of two increasing right-continuous predictable processes, and*

2. *$M_1 M_2 - \langle M_1, M_2 \rangle$ is a martingale.*

Corollary 2.1.3. *If M_1 and M_2 are two right-continuous $\mathcal{F}_t$-martingales with $E\{M_i^2(t)\} < \infty$ for any $t \geq 0$, then the right-continuous process $M_1 M_2$ is a martingale if and only if $\langle M_1, M_2 \rangle \equiv 0$. In this case, M_1 and M_2 are said to be* orthogonal.

The Doob-Meyer Decomposition Theorem can be extended somewhat, but we first need to establish some additional definitions.

Definition 2.1.11. Let $\{\mathcal{F}_t : t \geq 0\}$ be a filtration on a probability space. A nonnegative random variable τ is a *stopping time* with respect to $\{\mathcal{F}_t\}$ if $\{\tau \leq t\} \in \mathcal{F}_t$ for all $t \geq 0$.

Definition 2.1.12. An increasing sequence of random times $\tau_n, n = 1, 2, \ldots,$ is called *a localizing sequence with respect to a filtration* if the following hold true:

1. Each τ_n is a stopping time relative to the filtration, and

2. $\lim_{n \to \infty} \tau_n = \infty$ a.s.

Definition 2.1.13.

1. A stochastic process $M = \{M(t) : t \geq 0\}$ is a *local martingale (submartingale) with respect to a filtration* $\{\mathcal{F}_t : t \geq 0\}$ if there exists a localizing sequence $\{\tau_n\}$ such that, for each n, $M_n = \{M(t \wedge \tau_n) : 0 \leq t < \infty\}$ is an $\{\mathcal{F}_t\}$-martingale.

2. If M_n above is a martingale and a square integrable process, M_n is called a *square integrable martingale* and M is called a *local square integrable martingale*.

3. An adapted process $X = \{X(t) : t \geq 0\}$ is called *locally bounded* if, for a suitable localizing sequence $\{\tau_n\}$, $X_n = \{X(t \wedge \tau_n : t \geq 0\}$ is a bounded process for each n.

Theorem 2.1.3 (Extended Doob-Meyer Decomposition)**.** *Let $X = \{X(t) : t \geq 0\}$ be a nonnegative right-continuous $\mathcal{F}_t$-local submartingale with localizing sequence τ_n, where $\{\mathcal{F}_t : t \geq 0\}$ is a right-continuous filtration. Then there exists a unique increasing right-continuous predictable process A such that $A(0) = 0$ a.s., $P\{A(t) < \infty\} = 1$ for all $t > 0$ and $X - A$ is a right-continuous local martingale. At each t, $A(t)$ may be taken as the a.s. $\lim_{n\to\infty} A_n(t)$, where A_n is the compensator for the stopped submartingale $X(\cdot \wedge \tau_n)$.*

A few results follow from this extension that will be useful later on.

Theorem 2.1.4. *Let N be an arbitrary counting process.*

1. *Then there exists a unique right-continuous predictable increasing process A such that $A(0) = 0$ a.s., $A(t) < \infty$ a.s. for any t, and the process $M = N - A$ is a local martingale.*

2. *If A in (1) is locally bounded, M is a local square integrable martingale.*

Theorem 2.1.5. *Let N be a counting process and let A be its unique compensator in the Extended Doob-Meyer Decomposition Theorem. Then*

1. *A is a locally bounded process, and*

2. *$\Delta A(t) \equiv A(t) - \lim_{s\uparrow t} A(s) \leq 1$ a.s. for all $t \geq 0$.*

Lemma 2.1.1. *Suppose N is a counting process. Then $EN(t) = EA(t)$ for any $t \geq 0$, where A is the compensator for N. If $EA(t) < \infty$ for all t, then $M = N - A$ is a martingale.*

Theorem 2.1.6. *Suppose M is a right-continuous local square integrable martingale on $[0, \infty)$. Then there exists a unique predictable right-continuous increasing process $\langle M, M\rangle$ with $\langle M, M\rangle(0) = 0$ a.s. and $\langle M, M\rangle(t) < \infty$ a.s. for any T, such that $M^2 - \langle M, M\rangle$ is a right-continuous local martingale. If $\{\tau_n\}$ is a localizing sequence for M, then*

$$\langle M, M\rangle(t) = \lim_{n\to\infty} \langle M(\cdot \wedge \tau_n), M(\cdot \wedge \tau_n)\rangle(t).$$

Corollary 2.1.4. *Suppose M is a right-continuous local square integrable martingale on $[0, \infty)$, with $M(0) = 0$ a.s. Then*

$$\begin{aligned} EM^2(t) &\leq \lim_{n\to\infty} EM^2(t \wedge \tau_n) \\ &= E\langle M, M\rangle(t) \end{aligned}$$

Corollary 2.1.5. *If the local square integrable martingale M is a martingale with $M(0) = 0$ a.s., then for any $t \geq 0$*

$$EM^2(t) = E\langle M, M\rangle(t).$$

Theorem 2.1.7. *Suppose M_1 and M_2 are right-continuous local square integrable martingales on $[0, \infty)$. Then there exists a predictable right-continuous process $\langle M_1, M_2\rangle$ with $\langle M_1, M_2\rangle(0) = 0$ a.s. and $\langle M_1, M_2\rangle(t) < \infty$ a.s. for any t, and such that $M_1M_2 - \langle M_1, M_2\rangle$ is a right-continuous local martingale on $[0, \infty)$. In fact,*

$$\langle M_1, M_2\rangle = \frac{1}{2}\{\langle M_1 + M_2, M_1 + M_2\rangle - \langle M_1, M_1\rangle - \langle M_2, M_2\rangle\}$$

We can now begin preparation for the Martingale Central Limit Theorem in earnest. There are a number of important results to establish first, however, concerning the predictable variation and covariation processes for processes of the form $\int HdM$.

Theorem 2.1.8. *Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t : t \geq 0\}, P)$ be a stochastic basis with right-continuous filtration $\{\mathcal{F}_t : t \geq 0\}$, H a locally bounded $\{\mathcal{F}_t$-predictable process, and N a counting process. Let $M = N - A$ be the local square integrable $\{\mathcal{F}_t$-martingale whose existence is establish in Theorems 2.1.4 and 2.1.5. Then $\int HdM$ is a local square integrable martingale.*

Theorem 2.1.9. *Let N be a counting process with $EN(t) < \infty$ for any t. Let $\{\mathcal{F}_t : t \geq 0\}$ be a right-continuous filtration such that*

1. *$M = N - A$ is an $\mathcal{F}_t$-martingale, where $A = \{A(t) : t \geq 0\}$ is an increasing $\mathcal{F}_t$-predictable process with $A(0) = 0$;*

2. *H is bounded, $\mathcal{F}_t$-predictable process.*

Then the process L given by

$$L(t) = \int_0^t H(s)dM(s)$$

is an $\mathcal{F}_t$-martingale.

Theorem 2.1.10. *Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t : t \geq 0\}, P)$ be a stochastic basis, and assume:*

1. *H_i is a bounded (by Γ_i) $\mathcal{F}_t$-predictable process.*

2. *N_i is a bounded (by K_i) counting process.*

3. *The $\mathcal{F}_t$-martingale $M - i = N_i - A_i$ satisfies $EM_i^2(t) < \infty$ for any t,where A_i is the unique compensator arising in Corollary 2.1.1.*

Then

$$\left\langle \int H_1 dM_1, \int H_2 dM_2 \right\rangle = \int H_1 H_2 d\langle M_1, M_2 \rangle.$$

Theorem 2.1.11. *Assume that on a stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t : t \geq 0\}, P)$:*

1. *H_i is a locally bounded $\mathcal{F}_t$-predictable process;*

2. *N_i is a counting process.*

Then for the local martingales $M_i = N_i - A_i$,

$$\left\langle \int H_1 dM_1, \int H_2 dM_2 \right\rangle = \int H_1 H_2 d\langle M_1, M_2 \rangle;$$

that is, the process

$$\int H_1 dM_1 \int H_2 dM_2 - \int H_1 H_2 d\langle M_1, M_2 \rangle$$

is a local martingale over $[0, \infty)$

Theorem 2.1.12. *Let $H_i, M_i, i = 1, 2$, and $\{\mathcal{F}_t : t \geq 0\}$ be as defined in Theorem 2.1.11. Suppose $E\int_o^t H_i^2 d\langle M_i, M_i\rangle < \infty$ for $i = 1, 2$. Then*

1. *$\int H_i dM_i$ is a martingale over $[0, t]$;*
2. *$E\int_0^t H_i dM_i = 0$; and*
3. *For $i, j \in \{1, 2\}$,*

$$E\left(\int_0^t H_i dM_i \int_0^t H_j dM_j\right) = E\int_0^t H_i H_j d\langle M_i, M_j\rangle.$$

Theorem 2.1.13. *Consider two processes $U_\ell \equiv \sum_{i=1}^n \int H_{i,\ell} dM_{i,\ell}, \ell = 1, 2$. Let $\{\mathcal{F}_t : t \geq 0\}$ be a right-continuous filtration such that for each i and ℓ, $H_{i,\ell}$ is a locally bounded $\mathcal{F}_t$-predictable process, and $M_i = N_i - A_i$ is the local square integrable martingale corresponding to the arbitrary counting process N_i. Then*

1. *U_ℓ is a local square integrable martingale.*

IF $E\int_0^t H_{i,\ell}^2 d\langle M_i, M_i\rangle < \infty$ for any i, ℓ,

2. *$EU_\ell(t) = 0$;*
3. *$E\{U_1(t)U_2(t)\} = E\sum i = 1^n \sum_{\mathrm{j}=1}^n \int_0^t H_{i1} H_{j2} d\langle M_i, M_j\rangle$; and*
4. *U_ℓ is a martingale over $[0, t]$.*

Theorem 2.1.14. *Let N be a counting process on $[0, \infty)$ with $EN(t) < \infty$ for all $t \geq 0$, and let A denote its compensator. Assume that almost all paths of A are continuous, and that $EM^2(t) < \infty$ for all T, where $M = N - A$. Then $\langle M, M\rangle = A$, that is, $M^2 - A$ is a right-continuous martingale.*

Definition 2.1.14. A k-variate process $\{N_1, N_2, \ldots, N_k\}$ is called a *multivariate counting process* if

1. Each $N_j, j = 1, \ldots, k$, is a counting process, and

2. No two component processes jump at the same time.

Theorem 2.1.15. *Let $\{N_1, \ldots, N_k\}$ be a multivariate counting process and, for $j = 1, \ldots, k$, let A_j be the compensator of N_j. Assume that each A_j is a continuous process. Then*

1. *$\langle M_i, M_i \rangle = A_j$, that is, A_j is the unique predictable, right-continuous, increasing process with $A_j(0) = 0$ a.s. and $A_j(t) < \infty$ a.s. for any t, such that $M_j^2 - A_j$ is a local martingale, $j = 1, \ldots, k$.*

2. *If $i \neq j, \langle M_i, M_j \rangle(t) = 0$ a.s., that is, $M_i M_j$ is a local martingale.*

Theorem 2.1.16. *Let N be a counting process and A its compensator. If A is continuous then*

$$EM^2(t) \leq EA(t), \quad t \geq 0$$

If in addition, $EA(t) < \infty$ (or equivalently if $EN(t) < \infty$) for any t, then

$$EM^2(t) = EA(t), \quad t \geq 0$$

and $M^2 - A$ is a martingale.

2.1.1 Martingale Central Limit Theorem

In this section, we use the martingale structure of $U^{(n)} \equiv \sum_{i=1}^{n} \int H_i^{(n)} dM_i^{(n)}$. We consider the large-sample joint distribution of R statistics $(U_1^{(n)}, \ldots, U_r^{(n)})$. For each n, assume $\{N_{i,\ell}^{(n)} : i = 1, \ldots, n, \ell = 1, \ldots, r\}$ is a multivariate counting process with stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_t : t \geq 0\}, P)$ and the compensator $A_{i,\ell}^{(n)}$ of $N_{i,\ell}^{(n)}$ is continuous.

By Theorems 2.1.4 and 2.1.15, $M_{i,\ell}^{(n)} = N_{i,\ell}^{(n)} - A_{i,\ell}^{(n)}$ is a locally square integrable martingale, $\{M_{i,\ell}^{(n)}\}^2 - A_{i,\ell}^{(n)}$ is a local martingale, and $M_{i,\ell}^{(n)} M_{i',\ell'}^{(n)}$ is a local martingale if $i \neq i'$ or $\ell \neq \ell'$. For each i, ℓ we also assume $H_{i,\ell}^{(n)}$ is a locally bounded $\mathcal{F}_t$-predictable process.

Define

$$U_{i,\ell}^{(n)}(t) = \int_0^t H_{i,\ell}^{(n)}(s) dM_{i,\ell}^{(n)}(s),$$

$$U_{\ell}^{(n)}(t) = \sum_{i=1}^{n} \int_0^t H_{i,\ell}^{(n)}(s) dM_{i,\ell}^{(n)}(s),$$

and, for any $\epsilon > 0$,

$$U_{i,\ell,\epsilon}^{(n)}(t) = \int_0^t H_{i,\ell}^{(n)}(s) I_{\{|H_{i,\ell}^{(n)}(s)| \geq \epsilon\}} dM_{i,\ell}^{(n)}(s),$$

and

$$U_{\ell,\epsilon}^{(n)}(t) = \sum_{i=1}^{n} U_{i,\ell,\epsilon}^{(n)}(t).$$

By Theorem 2.1.8, $U_{i,\ell}^{(n)}$, $U_{\ell}^{(n)}(t)$, $U_{i,\ell,\epsilon}^{(n)}(t)$, and $U_{\ell,\epsilon}^{(n)}(t)$ are local square integrable martingales. The process $U_{\ell,\epsilon}^{(n)}(t)$ contains all jumps of the process $U_{\ell}^{(n)}(t)$ which are at least ϵ in size. By the linearity of $\langle \cdot, \cdot \rangle$ and by Theorem 2.1.11,

$$\langle U_{\ell}^{(n)}, U_{\ell}^{(n)} \rangle(t) = \sum_{i=1}^{n} \int_0^t \{H_{i,\ell}^{(n)}(s)\}^2 dA_{i,\ell}^{(n)}(s)$$

and

$$\langle U_{\ell,\epsilon}^{(n)}, U_{\ell\epsilon}^{(n)} \rangle(t) = \sum_{i=1}^{n} \int_0^t \{H_{i,\ell}^{(n)}(s)\}^2 I_{\{|H_{i,\ell}^{(n)}(s)| \geq \epsilon\}} dA_{i,\ell}^{(n)}(s).$$

Definition 2.1.15. The standard *Wiener* or *Brownian motion* process, W, is a stochastic process satisfying:

1. $W(0) = 0$, and $EW(t) = 0$ for any t;

2. W has independent increments, therefore, $W(t) - W(u)$ is independent of $W(u)$ for any $0 \leq u \leq t$;

3. $W(t)$ has variance t; and

4. W is a Gaussian process with continuous sample paths.

If f is a measurable nonnegative function and $\alpha(t) = \int_0^t f^2(s)ds$, then $\int f dW$ denotes the process satisfying (1), (2), and (4), but with

$$\text{var}\left\{\int_0^t f(s)dW(s) = \alpha(t)\right\}.$$

Theorem 2.1.17. *Let $W_1, \ldots, W_r$ be r independent Brownian motion processes and $f_1, \ldots, f_r$ be r measurable nonnegative functions such that $\alpha_\ell(t) = \int_0^t f_\ell^2(s)ds < \infty$ for all $t > 0$ and $\ell = 1, \ldots, r$.*

Suppose the conditions detailed earlier in this section hold. Assume for any $t \in [0, \tau]$ and for each $\ell = 1, \ldots, r$ that, as $n \to \infty$,

$$\langle U_\ell^{(n)}, U_\ell^{(n)}\rangle(t) \xrightarrow{P} \int_0^t f_\ell^2(s)ds,$$

and

$$\langle U_{\ell,\epsilon}^{(n)}, U_{\ell\epsilon}^{(n)}\rangle(t) \xrightarrow{P} 0$$

for any $\epsilon > 0$. Then

$$\left(U_1^{(n)}, \ldots, U_r^{(n)}\right) \Rightarrow \left(\int f_1 dW_1, \ldots, \int f_r dW_r\right) \text{ in } (D[0,\tau])^r \text{ as } n \to \infty.$$

This theorem is fundamental for obtaining limiting distributions for statistics using martingale theory. Knowing limiting distributions is critical for inference, as it allows for variance estimation, confidence interval and band construction, and hypothesis testing.

2.2 Cox Proportional Hazards Model

The Cox proportional hazards model was originally proposed in Cox (1972). For our purposes, however, we will summarize its key characteristics as discussed in Fleming and Harrington (2005) so that notation and presentation will be more consistent.

The proportional hazards model can be characterized by its hazard function:

$$\lambda(t, \boldsymbol{Z}) = \lambda_0(t) \exp(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{Z}).$$

The baseline cumulative hazard function can be estimated with the Breslow estimator

$$\widehat{\Lambda}_0(t) = \int_0^t \frac{\sum_{i=1}^n dN_i(s)}{\sum_{i=1}^n Y_i(s) \exp\left(\widehat{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{Z}_i(s)\right)},$$

and $\boldsymbol{\beta}$ can be estimated by solving the score equation

$$U(\boldsymbol{\beta}, \tau) = \sum_{i=1}^n \int_0^\infty \left\{ \boldsymbol{Z}_i - \frac{\boldsymbol{S}^{(1)}(\boldsymbol{\beta}, t)}{S^{(0)}(\boldsymbol{\beta}, t)} \right\} dN_i(t) = 0,$$

where τ is such that $\int_0^\tau \lambda_0(s)ds < \infty$, $S^{(0)}(\boldsymbol{\beta}, t) = n^{-1} \sum_{i=1}^n Y_i(t) e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{Z}_i(t)} \to_p s^{(0)}(\boldsymbol{\beta}, t)$, and $\boldsymbol{S}^{(1)}(\boldsymbol{\beta}, t) = n^{-1} \sum_{i=1}^n Y_i(t) \boldsymbol{Z}_i(t) e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{Z}_i(t)} \to_p \boldsymbol{s}^{(1)}(\boldsymbol{\beta}, t)$.

It can be shown that $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ converges in distribution as $n \to \infty$ to a mean zero p-variate Gaussian variable with covariance matrix $\{\boldsymbol{\Sigma}(\boldsymbol{\beta}, \tau)\}^{-1}$ with

$$\boldsymbol{\Sigma}(\boldsymbol{\beta}, \tau) = \int_0^\tau \mathbf{v}(\boldsymbol{\beta}, v) s^{(0)}(\boldsymbol{\beta}, v) \lambda_0(v) dv,$$

where $\boldsymbol{S}^{(2)}(\boldsymbol{\beta}, t) = n^{-1} \sum_{i=1}^n Y_i(t) \boldsymbol{Z}_i^{\otimes 2} e^{\boldsymbol{\beta}' \boldsymbol{Z}_i} \to_p \boldsymbol{s}^{(2)}(\boldsymbol{\beta}, t)$, $\boldsymbol{e}(\boldsymbol{\beta}, t) = \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}, t)}{s^{(0)}(\boldsymbol{\beta}, t)}$, and $\mathbf{v}(\boldsymbol{\beta}, t) = \frac{\boldsymbol{s}^{(2)}(\boldsymbol{\beta}, t)}{s^{(0)}(\boldsymbol{\beta}, t)} - \boldsymbol{e}(\boldsymbol{\beta}, t)^{\otimes 2}$. Additionally, the process $\sqrt{n}\{\widehat{\Lambda}_0(\cdot) - \Lambda(\cdot)\}$ converges weakly to a Gaussian process with mean zero, independent increments, and variance function

$$\int_0^t \frac{\lambda_0(s)}{s^{(0)}(\boldsymbol{\beta}, s)} ds + \boldsymbol{Q}^{\mathrm{T}}(\boldsymbol{\beta}, t) \boldsymbol{\Sigma}^{-1}(\boldsymbol{\beta}, \tau) \boldsymbol{Q}(\boldsymbol{\beta}, t),$$

where the vector function $\boldsymbol{Q}$ is given by

$$\boldsymbol{Q}(\boldsymbol{\beta}, t) = \int_0^t \boldsymbol{e}(\boldsymbol{\beta}, s) \lambda_0(s) ds.$$

This result in particular will be important in the development of our estimator for covariate-specific residual time under the proportional hazards model in Chapter 3.

2.2.1 Time-varying Covariates

All the above results hold even when time-varying covariates are included in the model. However, time-varying covariates require more care during analysis. In particular, it is worth considering whether a time-varying covariate is *external* or *internal* (Kalbfleisch and Prentice, 2002).

External covariates can fall into one of three categories. First, they can be fixed, such that $\boldsymbol{Z}(t) = \boldsymbol{Z}$ for all t. In other words, if they are time-invariate. Second, they can be defined such that their total path up to any time t, $\boldsymbol{Z}(t)$ is determined in advance for each individual under study. For example, under an intention-to-treat analysis, treatment in a cross-over study with fixed treatment times would be such a variable. Third, they can be ancillary in that the variable is the output of a stochastic process that is external to the individual under study and whose probability laws do not involve the parameters in the failure time model under study. For example, pollution levels in an asthma study would be considered ancillary.

On the other hand, an internal covariate is typically the output of a stochastic process that is generated by the individual under study and in many instances is observed only as long as the individual survives and is uncensored. This is an important distinction, as the observed value may carry information about the failure time. For a simple example, blood pressure in a cardiovascular study would be considered an internal covariate, as, if failure is death, blood pressure will be unmeasurable after failure.

While most of the statistical properties and developments apply to both external and internal covariates, there is the notable exception that is the value is known for an internal covariate we ordinarily know that the subject is alive. It is therefore very important to distinguish between the two when considering inference and application, particularly with regards to causality. We will be careful to consider this when extending our estimator for covariate-specific residual time to allow for time-varying covariates.

2.3 Additive Hazards Model

In this section we discuss the development of the additive hazards in model, reviewing the work of Lin and Ying (1994).

The additive hazard model can be characterized by its intensity function

$$Y_i(t)d\Lambda(t;\boldsymbol{Z}_i) = Y_i(t)\{d\Lambda_0(t) + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_i(t)dt\}$$

or, perhaps more simply, by its cumulative hazard function

$$\Lambda(t; \boldsymbol{Z}_i) = \Lambda_0(t) + \int_0^t \boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{Z}_i(t) dt\}.$$

The baseline cumulative hazard function can be estimated by

$$\widehat{\Lambda}_0(t) = \int_0^t \frac{\sum_{i=1}^n \{dN_i(s) - Y_i(s)\widehat{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{Z}_i ds\}}{\sum_{i=1}^n Y_i(s)},$$

and $\boldsymbol{\beta}$ can be estimated from the estimating function

$$U(\boldsymbol{\beta}) = \sum_{i=1}^n \int_0^\infty \{\boldsymbol{Z}_i(t) - \overline{\boldsymbol{Z}}(t)\}\{dN_i(t) - Y_i(t)\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{Z}_i(t) dt\},$$

where

$$\overline{\boldsymbol{Z}}(t) = \frac{\sum_{i=1}^n Y_i(t)\boldsymbol{Z}_i(t)}{\sum_{i=1}^n Y_i(t)}.$$

The resulting estimator takes the explicit form

$$\widehat{\boldsymbol{\beta}} = \frac{\sum_{i=1}^n \int_0^\infty Y_i(t)\{\boldsymbol{Z}_i(t) - \overline{\boldsymbol{Z}}(t)\}^{\otimes 2} dt}{\sum_{i=1}^n \int_0^\infty \{\boldsymbol{Z}_i(t) - \overline{\boldsymbol{Z}}(t)\} dN_i(t)}.$$

The estimating function can be rewritten as a martingale integral,

$$U(\boldsymbol{\beta}) = \sum_{i=1}^n \int_0^\infty \{\boldsymbol{Z}_i(t) - \overline{\boldsymbol{Z}}(t)\} dM_i(t),$$

allowing the application of counting process results as discussed in Section 2.1 and employed in Andersen and Gill (1982). We can conclude that the random vector $n^{-\frac{1}{2}} U(\boldsymbol{\beta})$ converges weakly to a p-variate normal with mean zero and a covariance matrix which can be consistently estimated by

$$\boldsymbol{B} = \frac{1}{n} \sum_{i=1}^n \int_0^\infty \{\boldsymbol{Z}_i - \overline{\boldsymbol{Z}}(t)\}^{\otimes 2} dN_i(t).$$

Furthermore, the random vector $\sqrt{n}(\widehat{\beta} - \beta)$ converges weakly to a p-variate normal with mean zero and with a covariance matrix which can be consistently estimated by $\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}$, where

$$\boldsymbol{A} = \frac{1}{n} \sum_{i=1}^n \int_0^\infty Y_i(t)\{\boldsymbol{Z}_i - \overline{\boldsymbol{Z}}(t)\}^{\otimes 2} dt.$$

Similar techniques can be applied to determine asymptotic results for the estimate of the baseline cumulative hazard. The process $\sqrt{n}\{\widehat{\Lambda}_0(\cdot) - \Lambda(\cdot)\}$ can be shown to converge weakly to a zero-mean Gaussian process whose variance at T can be consistently estimated by

$$\int_0^t \frac{n\sum_{i=1}^n dN_i(s)}{\{\sum_{i=1}^n Y_i(s)\}^2} - 2\boldsymbol{C}^{\mathrm{T}}(t)\boldsymbol{A}^{-1}\boldsymbol{D}(t) + \boldsymbol{C}^{\mathrm{T}}(t)\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{C}(t),$$

where

$$\boldsymbol{C}(t) = \int_0^t \overline{\boldsymbol{Z}}(s)ds$$

and

$$\boldsymbol{D}(t) = \int_0^t \frac{\sum_{i=1}^n \{\boldsymbol{Z}_i(s) - \overline{\boldsymbol{Z}}(s)\}dN_i(s)}{\sum_{i=1}^n Y_i(s)}.$$

This result in particular will be important in the development of our estimator for covariate-specific residual time under the additive hazards model in Chapter 5.

Chapter 3

RESIDUAL TIME QUANTILES IN THE PROPORTIONAL HAZARDS (PH) MODEL

3.1 Introduction

As previously mentioned, the Cox proportional hazards model is one of the most commonly used tools in the statistical analysis of time-to-event data. Because of this, it was natural to begin our development of an estimator for residual time quantiles by using the Cox proportional hazards framework and assumptions. Doing so ensured several things: that there would be a wealth of well-documented results about asymptotic properties of both the regression coefficients and the baseline hazard function; that there would be easy access to computational tools in statistical software, making implementation and simulation more feasible; and that our methods would be more approachable and intuitive to those familiar with the field of survival analysis.

In this chapter, we develop an estimator for quantiles of residual time using Cox model as a starting point. As shown in Section 3.2, the newly developed estimator allows estimation of covariate-specific residual time quantiles. We demonstrate our estimator's consistency, determine its limiting distribution, and provide a consistent estimator for its variance. Also included are discussions of methods for obtaining confidence intervals and bands that do not rely on direct estimation of the variance. We further develop our method in Section 3.3, determining the limiting distribution for a difference between two estimators of covariate-specific residual time quantiles and thereby allowing formal testing. In Section 3.4 we demonstrate our estimator's performance on simulated data, including figures showing confidence intervals and bands. Additionally, we apply our method to two real data sets: the VA lung cancer data as presented in Kalbfleisch and Prentice (2002) and the HIVNET

012 data as presented in Jackson et al. (2003). Finally, in Section 3.5 we discuss the limitations to our estimator.

3.2 Model-based Estimation of Residual Time Quantiles

Assume that T is a positive random variable, representing a subject's time-to-event. At a given time t, we first define the $(1-q)^{\text{th}}$ $(0 < q < 1)$ percentile residual time of a random variable T as the amount of additional time necessary for $(1-q) \times 100\%$ of the individuals still under observation at time t to fail. We denote this quantity as $\theta(t, q|\boldsymbol{Z})$, where $\boldsymbol{Z} \in \mathcal{R}^p$ is the associated p-dimensional vector of covariates. Let

$$S\{t + \theta(t, q|\boldsymbol{Z})|\boldsymbol{Z}\} = qS(t|\boldsymbol{Z}), \tag{3.1}$$

where $S(\cdot|\boldsymbol{Z})$ is the survival function. If expressed in terms of the cumulative hazard function, $\Lambda(\cdot|\boldsymbol{Z}) = -\log S(\cdot|\boldsymbol{Z})$, we then have $\Lambda\{t + \theta(t, q|\boldsymbol{Z})|\boldsymbol{Z}\} = \Lambda(t|\boldsymbol{Z}) - \log q$.

Now consider the Cox proportional hazards model that assumes

$$\Lambda(t|\boldsymbol{Z}) = \Lambda_0(t) \exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}), \tag{3.2}$$

where $\Lambda_0(\cdot)$ is the unspecified baseline cumulative hazard function, $\boldsymbol{\beta} \in \mathcal{B} \subset \mathcal{R}^p$ is the p-dimensional regression parameter, and $^{\mathrm{T}}$ denotes vector (matrix) transpose. Under model (3.2) we have

$$\Lambda_0\{t + \theta(t, q|\boldsymbol{Z})\} \quad = \quad \Lambda_0(t) - \log(q) \exp(-\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}).$$

As $\Lambda_0(\cdot)$ is monotonically increasing, solving this equation for $\theta(t, q|\boldsymbol{Z})$ yields

$$\theta(t, q|\boldsymbol{Z}) = \Lambda_0^{-1}\{\Lambda_0(t) - \log(q) \exp(-\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z})\} - t, \tag{3.3}$$

where $\Lambda_0^{-1}(s) = \inf\{t : \Lambda_0(t) \geq s\}$.

We can establish a few characteristics of $\theta(t, q|\boldsymbol{Z})$. It is decreasing both in q and in $\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}$ (i.e. for positive $\boldsymbol{\beta}$ it is decreasing in $\boldsymbol{Z}$ and vice versa). The latter is as we would expect based on the Cox model: individuals with greater values for covariates associated

with increased hazard will have lesser residual times. It is more difficult to characterize the dependence of $\theta(t, q|\boldsymbol{Z})$ on t. It will be constant with respect to t if survival is exponentially distributed. Beyond this, because of the basis on the cumulative hazard function, which is arbitrary under the Cox model, not much can be said.

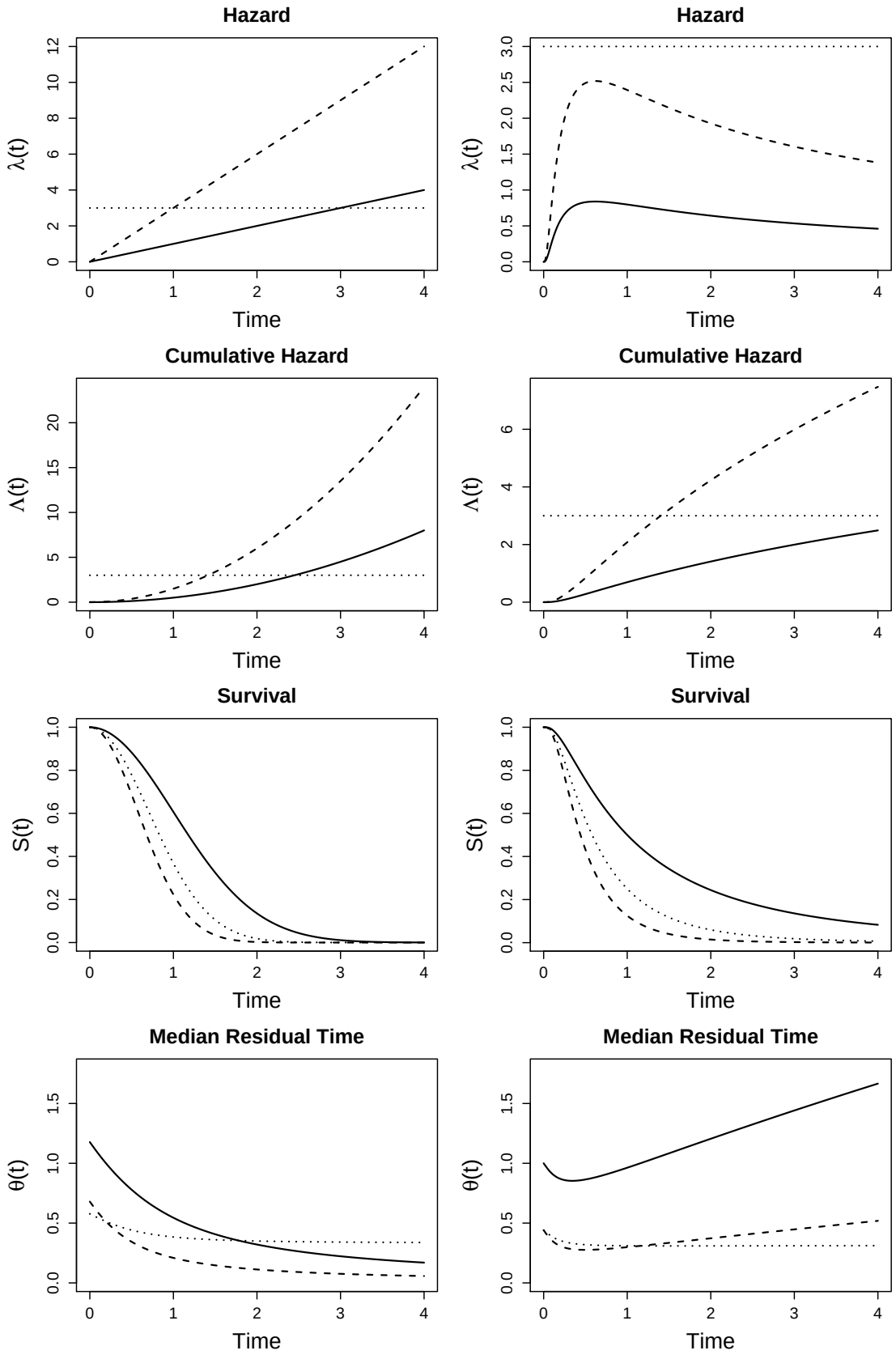
To provide more intuition regarding median residual time, we present Figure 3.1, displaying the hazard, cumulative hazard, survival, and median residual time curves for $z = 1$, $z = 0$, and the ratio of the former to the latter. Curves are based on either a Weibull or log-normal hazard, both under the proportional hazards assumption. For Weibull, hazard is specified as $\lambda(t|\boldsymbol{Z}) = t \exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z})$. For log-normal, hazard is specified as $\lambda(t|\boldsymbol{Z}) = \frac{\phi(\ln t)}{t\Phi(-\ln t)} \exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z})$, where ϕ and Φ are the standard normal density and distribution functions, respectively. In both cases, $\boldsymbol{\beta} = \log 3$ and $\boldsymbol{Z}$ is a single binary covariate.

Under Weibull hazard, for both covariate values, we see linearly increasing hazards, concave-up cumulative hazards, and concave-up decreasing median residual times. The ratio, by assumption, is constant between the hazards and cumulative hazards. The ratio between survival functions is decreasing, meaning that while they both converge to 0, one does so much more quickly. The ratio between median residual times is decreasing to approximately $\exp(-\boldsymbol{\beta}) = \frac{1}{3}$.

Under log-normal hazard, for both covariate values, hazards are concave-up and increasing initially before becoming concave-down and increasing, concave-down and decreasing, then finally concave-up and decreasing. Cumulative hazards both start out as concave-up and then become concave-down. Likewise for median residual times that initially decrease before becoming increasing. The ratio, by assumption, is constant between the hazards and cumulative hazards. The ratio between survival functions is decreasing, meaning that while they both converge to 0, one does so much more quickly. The ratio between median residual times is initially decreasing then increasing to approximately $\exp(-\boldsymbol{\beta}) = \frac{1}{3}$.

In order to estimate $\theta(t, q|\boldsymbol{Z})$ we need estimates for both $\boldsymbol{\beta}$ and $\Lambda_0(\cdot)$. Consider that data are collected in the form of $(X_i, \Delta_i, \boldsymbol{Z}_i)$ for $i = 1, \ldots, n$ with n being the sample size.

Figure 3.1: Plots of hazard, cumulative hazard, survival, and median residual time. For the left column, hazard is specified as $\lambda(t, q|\boldsymbol{Z}) = t \exp \boldsymbol{\beta Z}$. For the right column, hazard is specified as $\lambda(t, q|\boldsymbol{Z}) = \frac{\phi(\ln t)}{t\Phi(-\ln t)} \exp \boldsymbol{\beta Z}$. For both, $\boldsymbol{\beta} = \log 3$ and $\boldsymbol{Z}$ is binary. Solid curves correspond to $z = 0$, dashed curves correspond to $z = 1$, and dotted curves correspond to the ratio of the two (with that for $z = 1$ in the numerator).


Hazard
λ(t)
Time
Hazard
λ(t)
Time
Cumulative Hazard
Λ(t)
Time
Cumulative Hazard
Λ(t)
Time
Survival
S(t)
Time
Survival
S(t)
Time
Median Residual Time
θ(t)
Time
Median Residual Time
θ(t)
Time

For these data, $X_i = \min(T_i, C_i)$ where T_i is a failure time and C_i is a censoring time; $\Delta_i = I(T_i \leq C_i)$; and $\boldsymbol{Z}_i$ is a vector of covariates. Given $\boldsymbol{Z}_i$, T_i and C_i are assumed to be independent. Note that $N_i(t) = I(X_i \leq t, \delta_i = 1)$ and $Y_i(t) = I(X_i \geq t)$. Then under model (5.3), we can estimate $\boldsymbol{\beta}$ with $\widehat{\boldsymbol{\beta}}$, the standard estimator for Cox model regression coefficients (Fleming and Harrington, 2005), obtained by solving the partial score equations

$$\sum_{i=1}^{n} \int_0^{\infty} \left\{ \boldsymbol{Z}_i - \frac{\boldsymbol{S}^{(1)}(\boldsymbol{\beta}, t)}{S^{(0)}(\boldsymbol{\beta}, t)} \right\} dN_i(t) = 0,$$

where $S^{(0)}(\boldsymbol{\beta}, t) = n^{-1} \sum_{i=1}^{n} Y_i(t) e^{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{Z}_i} \rightarrow_p s^{(0)}(\boldsymbol{\beta}, t)$, and $\boldsymbol{S}^{(1)}(\boldsymbol{\beta}, t) = n^{-1} \sum_{i=1}^{n} Y_i(t) \times \boldsymbol{Z}_i e^{\boldsymbol{\beta}^{\mathrm{T}} Z_i} \rightarrow_p \boldsymbol{s}^{(1)}(\boldsymbol{\beta}, t)$. We can estimate the baseline cumulative hazard function $\Lambda_0(\cdot)$ with the Breslow estimator,

$$\widehat{\Lambda}_0(t) = \int_0^t \frac{\sum_{i=1}^{n} dN_i(s)}{\sum_{i=1}^{n} Y_i(s) \exp\left(\widehat{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{Z}_i\right)}.$$

Using the standard estimators $\widehat{\boldsymbol{\beta}}$ and $\widehat{\Lambda}_0(\cdot|\boldsymbol{Z})$, we can estimate $\theta(t, q|\boldsymbol{Z})$ with

$$\widehat{\theta}(t, q|\boldsymbol{Z}) = \widehat{\Lambda}_0^{-1}\{\widehat{\Lambda}_0(t) - \log(q) \exp(-\widehat{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{Z})\} - t, \tag{3.4}$$

where $\widehat{\Lambda}_0^{-1}(s) = \inf\{t : \widehat{\Lambda}_0(t) \geq s\}$. Note that $\widehat{\theta}(t, q|\boldsymbol{Z})$ is only defined when $\widehat{\Lambda}_0(t) - \log q \exp(-\widehat{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{Z}) \leq \widehat{\Lambda}_0(\tau)$ where τ is the largest observed failure time. For $\widehat{\theta}(t, q|\boldsymbol{Z})$, we have the following asymptotic properties, summarized in Theorem 3.2.1.

Theorem 3.2.1. *We assume conditions A-D (see below) of Andersen and Gill (1982) hold. If, as previously developed,* $\widehat{\theta}(t, q|\boldsymbol{Z}) = \widehat{\Lambda}_0^{-1}\{\widehat{\Lambda}_0(t) - \log(q) \exp(-\widehat{\boldsymbol{\beta}}^{T} \boldsymbol{Z})\} - t$, *then as* $n \rightarrow \infty$, *for a given* $\boldsymbol{Z}$,

(a) $\widehat{\theta}(t, q|\boldsymbol{Z}) \rightarrow_p \theta(t, q|\boldsymbol{Z})$;

(b) $\sqrt{n}\{\widehat{\theta}(t, q|\boldsymbol{Z}) - \theta(t, q|\boldsymbol{Z})\} \rightarrow_d N(0, V(t))$, *where*

$$V(t) = \frac{1}{\lambda_0\{\theta(t, q|\boldsymbol{Z}) + t\}^2} \left[\boldsymbol{A}^{T} \{\boldsymbol{\Sigma}(\boldsymbol{\beta}, \tau)\}^{-1} \boldsymbol{A} + \int_0^{\theta(t,q|\boldsymbol{Z})} \frac{\lambda_0(t+v)}{s^{(0)}(\boldsymbol{\beta}, t+v)} dv \right],$$

$$\boldsymbol{A} = \left[\log q \times \boldsymbol{Z} \exp\left(-\boldsymbol{\beta}^{T} \boldsymbol{Z}\right) + \int_0^{\theta(t,q|\boldsymbol{Z})} \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}, t+v)}{s^{(0)}(\boldsymbol{\beta}, t+v)} \lambda_0(t+v) d(t+v) \right]^{T},$$

and τ is as defined in condition A (see below): $\int_0^\tau \lambda(t)dt < \infty$.

(c) $\widehat{V}(t) \to_p V(t)$, where

$$\widehat{V}(t) = \frac{1}{\widehat{\lambda}_0(\widehat{\theta}(t,q|\boldsymbol{Z})+t)^2}\left[\widehat{\boldsymbol{A}}^T\{\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\beta}},\tau)\}^{-1}\widehat{\boldsymbol{A}} + \int_0^{\widehat{\theta}(t,q|\boldsymbol{Z})} \frac{d\widehat{\Lambda}_0(t+v)}{S^{(0)}(\widehat{\boldsymbol{\beta}},t+v)}\right]$$

and

$$\widehat{\boldsymbol{A}} = \left[\log q \times \boldsymbol{Z}\exp\left(-\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}\right) + \int_0^{\widehat{\theta}(t,q|\boldsymbol{Z})} \frac{\boldsymbol{S}^{(1)}(\widehat{\boldsymbol{\beta}},t+v)}{n\{S^{(0)}(\widehat{\boldsymbol{\beta}},t+v)\}^2} d\overline{N}(t+v)\right]^T.$$

Here, $\boldsymbol{S}^{(2)}(\boldsymbol{\beta},t) = n^{-1}\sum_{i=1}^n Y_i(t)\boldsymbol{Z}_i^{\otimes 2}e^{\boldsymbol{\beta}'\boldsymbol{Z}_i} \to_p \boldsymbol{s}^{(2)}(\boldsymbol{\beta},t)$, $\boldsymbol{e}(\boldsymbol{\beta},t) = \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta},t)}{s^{(0)}(\boldsymbol{\beta},t)}$, $\mathbf{v}(\boldsymbol{\beta},t) = \frac{\boldsymbol{s}^{(2)}(\boldsymbol{\beta},t)}{s^{(0)}(\boldsymbol{\beta},t)} - \boldsymbol{e}(\boldsymbol{\beta},t)^{\otimes 2}$, $\boldsymbol{\Sigma}(\boldsymbol{\beta},\tau) = \int_0^\tau \mathbf{v}(\boldsymbol{\beta},v)s^{(0)}(\boldsymbol{\beta},v)\lambda_0(v)dv$, $\boldsymbol{E}(\boldsymbol{\beta},t) = \frac{\boldsymbol{S}^{(1)}(\boldsymbol{\beta},t)}{S^{(0)}(\boldsymbol{\beta},t)}$, $\mathbf{V}(\boldsymbol{\beta},t) = \frac{\boldsymbol{S}^{(2)}(\boldsymbol{\beta},t)}{S^{(0)}(\boldsymbol{\beta},t)} - \left(\frac{\boldsymbol{S}^{(1)}(\boldsymbol{\beta},t)}{S^{(0)}(\boldsymbol{\beta},t)}\right)^{\otimes 2}$, and $\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\beta}},\tau) = \frac{1}{n}\int_0^\tau \sum_{i=1}^n \mathbf{V}(\widehat{\boldsymbol{\beta}},v)dN_i(v)$.

We restate conditions A-D of Andersen and Gill (1982) with minor modification as in Fleming and Harrington (2005):

A. The time τ is such that $\int_0^\tau \lambda(t)dt < \infty$.

B. For $\boldsymbol{S}^{(j)}, j = 0,1,2$, there exists a neighborhood $\mathcal{B}$ of $\boldsymbol{\beta}$ and scalar, vector, and matrix functions $s^{(0)}$, $\boldsymbol{s}^{(1)}$, and $\boldsymbol{s}^{(2)}$ defined on $\mathcal{B}\times[0,\tau]$ such that, for $j = 0,1,2$,

$$\sup_{t\in[0,\tau],\boldsymbol{\beta}\in\mathcal{B}} ||\boldsymbol{S}^{(j)}(\boldsymbol{\beta},t) - \boldsymbol{s}^{(j)}(\boldsymbol{\beta},t)|| \to 0$$

in probability as $n\to\infty$.

C. There exists $\delta > 0$ such that

$$\sqrt{n}\sup_{i,t} |\boldsymbol{Z}_i|Y_i(t)I(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_i > -\delta|\boldsymbol{Z}_i|) \to 0$$

in probability as $n\to\infty$.

D. Let $\mathcal{B}$ and $\boldsymbol{s}^{(j)}, j = 0,1,2$ be defined as in Condition B and $\boldsymbol{e}$ and $\mathbf{v}$ as in Section 5.2. Then for all $\boldsymbol{\beta}\in\mathcal{B}$ and $t\in[0,\tau]$:

$$\boldsymbol{s}^{(1)}(\boldsymbol{\beta},t) = \frac{\partial}{\partial\boldsymbol{\beta}}s^{(0)}(\boldsymbol{\beta},t), \quad \boldsymbol{s}^{(2)}(\boldsymbol{\beta},t) = \frac{\partial^2}{\partial\boldsymbol{\beta}^2}s^{(0)}(\boldsymbol{\beta},t),$$

$s^{(0)}(\cdot, t)$, $\boldsymbol{s}^{(1)}(\cdot, t)$, and $\boldsymbol{s}^{(2)}(\cdot, t)$ are continuous function of $\boldsymbol{\beta} \in \mathcal{B}$, uniformly in $t \in [0, \tau]$, $s^{(0)}$, $\boldsymbol{s}^{(1)}$, and $\boldsymbol{s}^{(2)}$ are bounded on $\mathcal{B} \times [0, \tau]$; $s^{(0)}$ is bounded away from zero on $\mathcal{B} \times [0, \tau]$, and the matrix $\boldsymbol{\Sigma}$ (as defined above) is positive definite.

Proof of Theorem 3.2.1. The proof for (a) follows directly from the established convergence of the individual estimators and the application of the continuous mapping theorem. Likewise for (c), where the consistency of the kernel-estimated baseline hazard has already been established (Ramlau-Hansen, 1983; Yandell, 1983). Therefore we need only prove (b) in detail.

We begin with two equations: one based on the true values,

$$\Lambda_0\{\theta(t, q|\boldsymbol{Z}) + t\} = \Lambda_0(t) - \log q \times \exp(-\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}),$$

and one based on the estimated values,

$$\widehat{\Lambda}_0\{\widehat{\theta}(t, q|\boldsymbol{Z}) + t\} = \widehat{\Lambda}_0(t) - \log q \times \exp(-\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{Z}).$$

Taking the differences between the left- and right-hand sides of both equations yields

$$\widehat{\Lambda}_0\{\widehat{\theta}(t, q|\boldsymbol{Z}) + t\} - \Lambda_0\{\theta(t, q|\boldsymbol{Z}) + t\} = \\ \widehat{\Lambda}_0(t) - \log q \times \exp(-\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{Z}) - \Lambda_0(t) + \log q \times \exp(-\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}).$$

Examining the left-hand side first, we have

$$\begin{aligned} \widehat{\Lambda}_0\{\widehat{\theta}(t, q|\boldsymbol{Z}) + t\} - \Lambda_0\{\theta(t, q|\boldsymbol{Z}) + t\} &= \widehat{\Lambda}_0\{\widehat{\theta}(t, q|\boldsymbol{Z}) + t\} - \widehat{\Lambda}_0\{\theta(t, q|\boldsymbol{Z}) + t\} \\ &\quad + \widehat{\Lambda}_0\{\theta(t, q|\boldsymbol{Z}) + t\} - \Lambda_0\{\theta(t, q|\boldsymbol{Z}) + t\} \\ &\approx \widehat{\lambda}_0\{\theta(t, q|\boldsymbol{Z}) + t\}\{\widehat{\theta}(t, q|\boldsymbol{Z}) - \theta(t, q|\boldsymbol{Z})\} \\ &\quad + \widehat{\Lambda}_0\{\theta(t, q|\boldsymbol{Z}) + t\} - \Lambda_0\{\theta(t, q|\boldsymbol{Z}) + t\}, \end{aligned}$$

where the approximation is due to Taylor's expansion. Examining the right-hand side, we have

$$\widehat{\Lambda}_0(t) - \Lambda_0(t) - \log q \times \{\exp(-\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{Z}) - \exp(-\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z})\} \approx \\ \widehat{\Lambda}_0(t) - \Lambda_0(t) + \log q \times \boldsymbol{Z} \exp(-\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{Z})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}),$$

again using Taylor's approximation. Combining everything gives us the expression

$$\widehat{\lambda}_0\{\theta(t,q|\boldsymbol{Z})+t\}\{\widehat{\theta}(t,q|\boldsymbol{Z})-\theta(t,q|\boldsymbol{Z})\}+\widehat{\Lambda}_0\{\theta(t,q|\boldsymbol{Z})+t\}-\Lambda_0\{\theta(t,q|\boldsymbol{Z})+t\}\approx$$
$$\widehat{\Lambda}_0(t)-\Lambda_0(t)+\log q\times\boldsymbol{Z}\exp(-\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{Z})(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}).$$

Rearranging yields the approximation

$$\begin{aligned}\widehat{\theta}(t,q|\boldsymbol{Z})-\theta(t,q|\boldsymbol{Z}) \;\approx\; & \frac{1}{\widehat{\lambda}_0\{\theta(t,q|\boldsymbol{Z})+t\}}\Bigg(\log q\times\boldsymbol{Z}\exp(-\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{Z})(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\\ & +\widehat{\Lambda}_0(t)-\Lambda_0(t)-\Big[\widehat{\Lambda}_0\{\theta(t,q|\boldsymbol{Z})+t\}-\Lambda_0\{\theta(t,q|\boldsymbol{Z})+t\}\Big]\Bigg).\end{aligned}$$

From Fleming and Harrington (2005) we know that

$$\widehat{\Lambda}_0(t)-\Lambda_0(t)\approx-\left[\int_0^t\frac{\boldsymbol{S}^{(1)}(\widehat{\boldsymbol{\beta}},u)}{n\{S^{(0)}(\widehat{\boldsymbol{\beta}},u)\}^2}d\overline{N}(u)\right]^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})+\int_0^t\frac{\sum_{i=1}^n dM_i(u)}{S^{(0)}(\widehat{\boldsymbol{\beta}},u)},$$

so we can rewrite the overall approximation as

$$\begin{aligned}\widehat{\theta}(t,q|\boldsymbol{Z})-\theta(t,q|\boldsymbol{Z})\approx\frac{1}{\widehat{\lambda}_0\{\theta(t,q|\boldsymbol{Z})+t\}}\Bigg\{&-\int_t^{\theta(t,q|\boldsymbol{Z})+t}\frac{\sum_{i=1}^n dM_i(u)}{S^{(0)}(\widehat{\boldsymbol{\beta}},u)}\\ &+\left[\log q\times\boldsymbol{Z}\exp(-\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{Z})+\int_t^{\theta(t,q|\boldsymbol{Z})+t}\frac{\boldsymbol{S}^{(1)}(\widehat{\boldsymbol{\beta}},u)}{n\{S^{(0)}(\widehat{\boldsymbol{\beta}},u)\}^2}d\overline{N}(u)\right]^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\Bigg\}.\end{aligned}$$

We can also perform a substitution, setting $u=t+v$, and integrating with respect to v. This yields the approximation

$$\begin{aligned}\widehat{\theta}(t,q|\boldsymbol{Z})-\theta(t,q|\boldsymbol{Z})\approx\frac{1}{\widehat{\lambda}_0\{\theta(t,q|\boldsymbol{Z})+t\}}\Bigg\{&-\int_0^{\theta(t,q|\boldsymbol{Z}))}\frac{\sum_{i=1}^n dM_i(t+v)}{S^{(0)}(\widehat{\boldsymbol{\beta}},t+v)}\\ &+\left[\log q\times\boldsymbol{Z}\exp(-\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{Z})+\int_0^{\theta(t,q|\boldsymbol{Z}))}\frac{\boldsymbol{S}^{(1)}(\widehat{\boldsymbol{\beta}},t+v)}{n\{S^{(0)}(\widehat{\boldsymbol{\beta}},t+v)\}^2}d\overline{N}(t+v)\right]^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\Bigg\}.\end{aligned}$$

Applying results from Fleming and Harrington (2005) about the asymptotic variance of $\widehat{\boldsymbol{\beta}}$ and martingales, we have:

$$\mathrm{Var}\{\widehat{\theta}(t,q|\boldsymbol{Z})-\theta(t,q|\boldsymbol{Z})\}\approx\frac{1}{\widehat{\lambda}_0\{\theta(t,q|\boldsymbol{Z})+t\}^2}\left\{\widetilde{\boldsymbol{A}}^{\mathrm{T}}\mathrm{Var}(\widehat{\boldsymbol{\beta}})\widetilde{\boldsymbol{A}}+\int_0^{\theta(t,q|\boldsymbol{Z})}\frac{d\widehat{\Lambda}_0(t+v)}{S^{(0)}(\widehat{\boldsymbol{\beta}},t+v)}\right\},$$

where

$$\widetilde{\boldsymbol{A}} = \left[\log q \times \boldsymbol{Z} \exp\left(-\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{Z}\right) + \int_0^{\theta(t,q|\boldsymbol{Z})} \frac{\boldsymbol{S}^{(1)}(\widehat{\boldsymbol{\beta}}, t+v)}{n\{S^{(0)}(\widehat{\boldsymbol{\beta}}, t+v)\}^2} d\overline{N}(t+v)\right]^{\mathrm{T}}.$$

Based on the consistency of the estimators used in this formulation, we have the asymptotic result $\sqrt{n}(\widehat{\theta}(t,q|\boldsymbol{Z}) - \theta(t,q|\boldsymbol{Z})) \to_d N(0, V(t))$, where

$$V(t) \approx \frac{1}{\lambda_0\{\theta(t,q|\boldsymbol{Z}) + t\}^2}\left[\boldsymbol{A}^{\mathrm{T}}\{\boldsymbol{\Sigma}(\boldsymbol{\beta},\tau)\}^{-1}\boldsymbol{A} + \int_0^{\theta(t,q|\boldsymbol{Z})} \frac{\lambda_0(t+v)}{s^{(0)}(\boldsymbol{\beta}, v)} dv\right]$$

and

$$\boldsymbol{A} = \left[\log q \times \boldsymbol{Z} \exp\left(-\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}\right) + \int_0^{\theta(t,q|\boldsymbol{Z})} \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}, v)}{s^{(0)}(\boldsymbol{\beta}, v)} \lambda_0(t+v) d(v)\right]^{\mathrm{T}}.$$

□

With Theorem 3.2.1, calculating pointwise confidence intervals is possible. We may simply take $\widehat{\theta}(t,q|\boldsymbol{Z}) \pm z^*_{1-\alpha/2}\sqrt{\widehat{V}(t)}$ for a $100(1-\alpha)\%$ confidence interval. Moreover, we can use asymptotic results to obtain confidence bands as well. Specifically, we know that

$$\sup_{0\le t\le\tau} \frac{\sqrt{n}\left|\widehat{\theta}(t,q|\boldsymbol{Z}) - \theta(t,q|\boldsymbol{Z})\right|}{V(\tau)} \to_p \sup_{0\le t\le 1} |W(t)|$$

where $W(t)$ is standard Brownian motion. We can therefore solve

$$\Pr\left(\sup_{0\le t\le 1}|W(t)| \le c\right) = \frac{4}{\pi}\sum_{k=0}^{\infty}\frac{(-1)^k}{2k+1} e^{-\frac{\pi^2(2k+1)^2}{8c^2}} \approx 1-\alpha$$

for c and compute $100(1-\alpha)\%$ confidence bands using $\widehat{\theta}(t,q|\boldsymbol{Z}) \pm c\sqrt{\widehat{V}(\tau)}$.

In addition, we can compute both confidence bands and intervals using either bootstrap or simulation methods. Using the bootstrap method, we simply find the 2.5%ile and 97.5%ile of the bootstrap sample estimates of $\widehat{\theta}^*(t,q|\boldsymbol{Z})$ at each time point in order to get pointwise intervals (Efron and Tibshirani, 1998). To obtain bands, we first calculate the maximum deviation within each bootstrap sample across time, $\sup_{0\le t\le\tau}|\widehat{\theta}^*(t,q|\boldsymbol{Z}) - \widehat{\theta}(t,q|\boldsymbol{Z})|$, and then find the 95%ile across samples, c. We can then construct bands using $\widehat{\theta}(t,q|\boldsymbol{Z}) \pm c$.

The simulation method is in fact another way of resampling (Parzen et al., 1994). In the formulation of $\widehat{\theta}(t, q|\boldsymbol{Z}) - \theta(t, q|\boldsymbol{Z})$, we replace $dM_i(u)$ with $G_i dN_i(u)$ where $G_i \sim N(0, 1)$. For each simulated sample, all G_i are randomly generated and we compute the estimated residual time quantile for that sample, $\widetilde{\theta}(t, q|\boldsymbol{Z})$. At each time point, the 2.5%ile and 97.5%ile of the deviations $\widetilde{\theta}(t, q|\boldsymbol{Z}) - \theta(t, q|\boldsymbol{Z})$ are calculated and added to the estimate $\widehat{\theta}(t, q|\boldsymbol{Z})$ to get lower and upper pointwise confidence intervals, respectively. Calculating bands is the same as with the bootstrap: we find the maximum deviation within each simulated sample across time, $\sup_{0\leq t\leq\tau} |\widetilde{\theta}(t, q|\boldsymbol{Z}) - \widehat{\theta}(t, q|\boldsymbol{Z})|$, and then find the 95%ile across samples, c. We can then construct bands using $\widehat{\theta}(t, q|\boldsymbol{Z}) \pm c$.

3.3 Comparing residual time quantiles

While being able to estimate covariate-specific residual time quantiles and their variance is useful, in most practical applications it is also important to be able to carry out comparisons between different covariate values and determine if any observed difference is statistically significant. We may also be interested in formally comparing residual times at different fixed time points or for different quantiles. All of these tasks require being able to estimate the covariance between two different residual time quantiles.

However, based on the formula for covariate-specific residual time quantiles,

$$\theta(t, q|\boldsymbol{Z}) = \Lambda_0^{-1}\{\Lambda_0(t) - \log(q)\exp(-\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z})\} - t,$$

it may appear that one can simply carry out the usual test of the significance of $\boldsymbol{\beta}$ to determine if any differences exist in the residual time quantiles (at least for fixed t and q). Indeed, we can outline certain relationships between $\theta(t, q|\boldsymbol{Z})$ and $\boldsymbol{\beta}$ as follows.

Lemma 3.3.1. *For* $\theta(t, q|\boldsymbol{Z}) = \Lambda_0^{-1}\{\Lambda_0(t) - \log(q)\exp(-\boldsymbol{\beta}^T\boldsymbol{Z})\} - t$ *and strictly increasing* $\Lambda(\cdot)$*, the following hold*

(a) $\theta(t, q|\boldsymbol{Z}_1) = \theta(t, q|\boldsymbol{Z}_2) \Leftrightarrow \boldsymbol{\beta}^T\boldsymbol{Z}_1 = \boldsymbol{\beta}^T\boldsymbol{Z}_2$

(b) $\theta(t, q|\boldsymbol{Z}_1) \neq \theta(t, q|\boldsymbol{Z}_2) \Leftrightarrow \boldsymbol{\beta}^T\boldsymbol{Z}_1 \neq \boldsymbol{\beta}^T\boldsymbol{Z}_2$.

For testing the effect of a single covariate this seems to imply that we can simply test the null hypothesis that $\boldsymbol{\beta} = 0$. This conclusion does not apply when considering the estimator, as outlined below.

Lemma 3.3.2. *For* $\widehat{\theta}(t, q|\boldsymbol{Z}) = \widehat{\Lambda}_0^{-1}\{\widehat{\Lambda}_0(t) - \log(q)\exp(-\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z})\} - t$, $\widehat{\Lambda}(\cdot)$ *cannot be assumed to be strictly increasing and therefore while the following hold*

(a) $\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}_1 = \widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}_2 \Rightarrow \widehat{\theta}(t, q|\boldsymbol{Z}_1) = \widehat{\theta}(t, q|\boldsymbol{Z}_2)$

(b) $\widehat{\theta}(t, q|\boldsymbol{Z}_1) \neq \widehat{\theta}(t, q|\boldsymbol{Z}_2) \Rightarrow \widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}_1 \neq \widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}_2$,

their converses do not.

Based on these considerations, one might conclude that testing $\boldsymbol{\beta}$ directly is always more powerful than testing for a difference in residual time quantiles. Unfortunately, this is not the case. We conducted simulations to demonstrate the occasional disagreement between conclusions based on the significance of $\widehat{\boldsymbol{\beta}}$ and the significance of the difference $\{\widehat{\theta}(t, q|\boldsymbol{Z}_1) - \widehat{\theta}(t, q|\boldsymbol{Z}_2)\}$. Survival times were generated with a baseline hazard specified as $\lambda(t|Z_i) = t\exp(\beta^{\mathrm{T}}Z_i)$, $\boldsymbol{\beta} = 0.25$, and $Z_i \sim \mathrm{Bernoulli}(0.5)$ for $i = 1, \ldots, n$. When generating censored data, censoring times followed an exponential distribution with rate parameter 0.17 ($\sim 10\%$ censoring) or 0.61 ($\sim 30\%$ censoring).

We used $q = 0.5$, $Z_1 = 0$ and $Z_2 = 1$ with evaluation at $t = 0.25$ and $t = 0.75$. Our simulation included 10,000 replications. We calculated the proportion of tests based on either quantity that would be rejected or failed to be rejected when compared to both the null value and to the true value.

Results highlighting the occasional disagreement between conclusions based on the significance of $\widehat{\beta}$ and the significance of the difference $\{\widehat{\theta}(t_1, q_1|Z_1) - \widehat{\theta}(t_2, q_2|Z_2)\}$ can be seen in Tables 3.1 and 3.2. Here we let the superscript ^{r} denote rejection and the subscript $_f$ denote failure of rejectection of the hypotheses about either β or θ. While tests based on the two different quantities agree in a large majority of cases, there are still a substantial number of disagreements (around 6-7% when comparing to the true values). That is, we may occasionally find a statistically significant difference between $\widehat{\theta}(t, q|Z_1)$ and $\widehat{\theta}(t, q|Z_2)$

Table 3.1: Percentages of simulations in which the null hypothesis (no difference) was rejected or failed to be rejected. The superscript ^{r} denotes rejection of the hypothesis while the subscript $_f$ denotes failure to reject the hypothesis about either β or θ.

		$t = 0.25$				$t = 0.75$			
Sample Size	% Censoring	$\beta^r\theta^r$	$\beta^r\theta_f$	$\beta_f\theta^r$	$\beta_f\theta_f$	$\beta^r\theta^r$	$\beta^r\theta_f$	$\beta_f\theta^r$	$\beta_f\theta_f$
$n = 200$	0	29.99	11.10	3.99	54.92	27.02	13.56	3.75	55.67
	10	24.73	10.77	3.52	60.98	22.91	13.17	3.71	60.21
	30	15.02	9.40	3.52	72.06	12.79	10.99	3.85	72.37
$n = 500$	0	70.37	8.53	1.97	19.13	68.15	11.17	2.20	18.48
	10	61.19	9.99	2.53	26.29	58.53	12.52	2.67	26.28
	30	38.70	11.97	3.52	45.81	35.90	15.87	3.64	44.59
$n = 1000$	0	96.06	1.72	0.30	1.92	95.07	2.58	0.38	1.97
	10	90.82	3.40	0.64	5.14	89.95	4.59	0.78	4.68
	30	72.32	8.23	1.89	17.56	69.77	12.08	1.81	16.34

Table 3.2: Percentages of simulations in which the truth (true difference) was rejected or failed to be rejected. The superscript ^{r} denotes rejection of the hypothesis while the subscript $_f$ denotes failure to reject the hypothesis about either β or θ.

		$t = 0.25$				$t = 0.75$			
Sample Size	% Censoring	$\beta^r\theta^r$	$\beta^r\theta_f$	$\beta_f\theta^r$	$\beta_f\theta_f$	$\beta^r\theta^r$	$\beta^r\theta_f$	$\beta_f\theta^r$	$\beta_f\theta_f$
$n = 200$	0	3.68	1.62	1.51	93.19	3.20	2.10	1.91	92.79
	10	3.43	1.83	1.63	93.11	2.88	2.28	1.70	93.14
	30	2.78	2.23	2.07	92.92	2.28	2.81	2.14	92.77
$n = 500$	0	3.67	1.34	1.57	93.42	2.94	1.64	1.91	93.51
	10	3.70	1.64	1.74	92.92	3.24	1.76	1.96	93.04
	30	3.52	1.38	1.62	93.48	3.10	1.96	1.78	93.16
$n = 1000$	0	3.07	1.54	1.84	93.55	3.01	1.76	2.25	92.98
	10	3.51	1.88	1.84	92.77	3.49	1.57	1.88	93.06
	30	3.40	1.66	1.69	93.25	2.99	2.26	1.90	92.85

while failing to reject that $\beta = 0$. It is therefore important in general to be able to calculate the covariance between $\widehat{\theta}(t, q|\boldsymbol{Z}_1)$ and $\widehat{\theta}(t, q|\boldsymbol{Z}_2)$ in order to construct confidence intervals and carry out hypothesis testing on the difference between the two. Furthermore, if we are not interested in keeping both t and q fixed, testing $\boldsymbol{\beta}$ alone will clearly be insufficient.

We can extend results from Theorem 3.2.1 to establish asymptotic properties for the differences between quantiles of residual time for different sets of covariate values, evaluation times, and quantiles. These properties are summarized in Theorem 3.3.1.

Theorem 3.3.1. *We assume conditions A-D (see above) of Andersen and Gill (1982) hold. If, as previously developed,* $\widehat{\theta}(t, q|\boldsymbol{Z}) = \widehat{\Lambda}_0^{-1}\{\widehat{\Lambda}_0(t) - \log(q)\exp(-\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z})\} - t$ *, then as* $n \to \infty$*, for a given* $\boldsymbol{Z}_1$ *and* $\boldsymbol{Z}_2$ *and fixed* t_1*,* t_2*,* q_1*, and* q_2*,*

(a) $\widehat{\theta}(t_1, q_1|\boldsymbol{Z}_1) - \widehat{\theta}(t_2, q_2|\boldsymbol{Z}_2) \to_p \theta(t_1, q_1|\boldsymbol{Z}_1) - \theta(t_2, q_2|\boldsymbol{Z}_2)$*;*

(b) $\sqrt{n}[\{\widehat{\theta}(t_1, q_1|\boldsymbol{Z}_1) - \widehat{\theta}(t_2, q_2|\boldsymbol{Z}_2)\} - \{\theta(t_1, q_1|\boldsymbol{Z}_1) - \theta(t_2, q_2|\boldsymbol{Z}_2)\}] \to_d N(0, W(t))$*,*

where

$$
\begin{aligned}
W(t) \;=\; & \frac{1}{\lambda_0(\theta(t_1, q_1|\boldsymbol{Z}_1) + t_1)^2}\left\{\boldsymbol{A}_1^T\{\boldsymbol{\Sigma}(\boldsymbol{\beta}, \tau)\}^{-1}\boldsymbol{A}_1 + \int_0^{\theta(t_1,q_1|\boldsymbol{Z}_1)} \frac{\lambda_0(t_1 + v)}{s^{(0)}(\boldsymbol{\beta}, t_1 + v)}dv\right\} \\
& + \frac{1}{\lambda_0(\theta(t_2, q_2|\boldsymbol{Z}_2) + t_2)^2}\left\{\boldsymbol{A}_2^T\{\boldsymbol{\Sigma}(\boldsymbol{\beta}, \tau)\}^{-1}\boldsymbol{A}_2 + \int_0^{\theta(t_2,q_2|\boldsymbol{Z}_2)} \frac{\lambda_0(t_2 + v)}{s^{(0)}(\boldsymbol{\beta}, t_2 + v)}dv\right\} \\
& - \frac{2}{\lambda_0(\theta(t_1, q_1|\boldsymbol{Z}_1) + t_1)\lambda_0(\theta(t_2, q_2|\boldsymbol{Z}_2) + t_2)} \\
& \times \left\{\int_0^{\eta_{\min} - t_{\max}} \frac{\lambda_0(t_{\max} + v)}{s^{(0)}(\boldsymbol{\beta}, t_{\max} + v)}dv + \boldsymbol{A}_1^T\{\boldsymbol{\Sigma}(\boldsymbol{\beta}, \tau)\}^{-1}\boldsymbol{A}_2\right\},
\end{aligned}
$$

$$
\boldsymbol{A}_j = \left[\log q_j \times \boldsymbol{Z}_j \exp\left(-\boldsymbol{\beta}^T\boldsymbol{Z}_j\right) + \int_0^{\theta(t_j,q_j|\boldsymbol{Z}_j)} \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}, t_j + v)}{s^{(0)}(\boldsymbol{\beta}, t_j + v)}\lambda_0(t_j + v)d(t_j + v)\right]^T
$$

for $j = 1, 2$*,* $\eta_{\min} = \min\{\theta(t_1, q_1|\boldsymbol{Z}_1) + t_1, \theta(t_2, q_2|\boldsymbol{Z}_2) + t_2\}$*,* $t_{\max} = \max\{t_1, t_2\}$*, and* τ *is as defined in condition A (see above):* $\int_0^\tau \lambda(t)dt < \infty$*.*

(c) $\widehat{W}(t) \to_p W(t)$*, where*

$$\begin{aligned}\widehat{W}(t) &= \frac{1}{\widehat{\lambda}_0(\widehat{\theta}(t_1, q_1|\boldsymbol{Z}_1) + t_1)^2}\left\{\widehat{\boldsymbol{A}}_1^T(\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\beta}}, \tau))^{-1}\widehat{\boldsymbol{A}}_1 + \int_0^{\widehat{\theta}(t_1,q_1|\boldsymbol{Z}_1)} \frac{d\widehat{\Lambda}_0(t_1+v)}{S^{(0)}(\widehat{\boldsymbol{\beta}}, t_1+v)}dv\right\} \\ &+\frac{1}{\widehat{\lambda}_0(\widehat{\theta}(t_2, q_2|\boldsymbol{Z}_2) + t_2)^2}\left\{\boldsymbol{A}_2^T(\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\beta}}, \tau))^{-1}\boldsymbol{A}_2 + \int_0^{\widehat{\theta}(t_2,q_2|\boldsymbol{Z}_2)} \frac{d\widehat{\Lambda}_0(t_2+v)}{S^{(0)}(\widehat{\boldsymbol{\beta}}, t_2+v)}dv\right\} \\ &-\frac{2}{\widehat{\lambda}_0(\widehat{\theta}(t_1, q_1|\boldsymbol{Z}_1) + t_1)\lambda_0(\widehat{\theta}(t_2, q_2|\boldsymbol{Z}_2) + t_2)} \\ &\times\left\{\int_0^{\widehat{\eta}_{\min}-t_{\max}} \frac{d\widehat{\Lambda}_0(t_{\max}+v)}{S^{(0)}(\widehat{\boldsymbol{\beta}}, t_{\max}+v)}dv + \widehat{\boldsymbol{A}}_1^T(\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\beta}}, \tau))^{-1}\widehat{\boldsymbol{A}}_2\right\},\end{aligned}$$

$$\widehat{\boldsymbol{A}}_j = \left[\log q_j \times \boldsymbol{Z}_j \exp\left(-\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}_j\right) + \int_0^{\widehat{\theta}(t_j,q_j|\boldsymbol{Z}_j)} \frac{\boldsymbol{S}^{(1)}(\widehat{\boldsymbol{\beta}}, t_j+v)}{n\{S^{(0)}(\widehat{\boldsymbol{\beta}}, t_j+v)\}^2}d\overline{N}(t_j+v)\right]^T$$

for $j = 1, 2$*, and* $\widehat{\eta}_{\min} = \min\{\widehat{\theta}(t_1, q_1|\boldsymbol{Z}_1) + t_1, \widehat{\theta}(t_2, q_2|\boldsymbol{Z}_2) + t_2\}$.

Proof of Theorem 3.3.1. The proof for (a) follows directly from the established convergence of the individual estimators and the application of the continuous mapping theorem. Likewise for (c), where the consistency of the kernel-estimated baseline hazard has already been established (Ramlau-Hansen, 1983; Yandell, 1983). Therefore we need only prove (b) in detail.

From the proof of Theorem 3.2.1, we have

$$\begin{aligned}\widehat{\theta}(t, q|\boldsymbol{Z}_j) - \theta(t_j, q_j|\boldsymbol{Z}_j) \approx \frac{1}{\widehat{\lambda}_0(\theta(t_j, q_j|\boldsymbol{Z}_j) + t)} \\ \times\left\{-\int_0^{\theta(t_j,q_j|\boldsymbol{Z}_j)} \frac{\sum_{i=1}^n dM_j(t_j+v)}{S^{(0)}(\widehat{\boldsymbol{\beta}}, t_j+v)} + \widetilde{\boldsymbol{A}}_j^{\mathrm{T}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right\}\end{aligned}$$

where

$$\widetilde{\boldsymbol{A}}_j = \left[\log q_j \times \boldsymbol{Z}_j \exp\left(-\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{Z}\right) + \int_0^{\theta(t_j,q_j|\boldsymbol{Z}_j)} \frac{\boldsymbol{S^{(1)}}(\widehat{\boldsymbol{\beta}}, t_j+v)}{n\{S^{(0)}(\widehat{\boldsymbol{\beta}}, t_j+v)\}^2}d\overline{N}(t_j+v)\right]^{\mathrm{T}}.$$

In order to calculate $\mathrm{Var}(\widehat{\theta}(t_1, q_1|\boldsymbol{Z}_1) - \widehat{\theta}(t_2, q_2|\boldsymbol{Z}_2))$, we note that

$$\begin{aligned}
&\mathrm{Var}(\widehat{\theta}(t_1, q_1|\boldsymbol{Z}_1) - \widehat{\theta}(t_2, q_2|\boldsymbol{Z}_2)) \\
&\qquad = \mathrm{Var}\left[\{\widehat{\theta}(t_1, q_1|\boldsymbol{Z}_1) - \theta(t_1, q_1|\boldsymbol{Z}_1)\} - \{\widehat{\theta}(t_2, q_2|\boldsymbol{Z}_2) - \theta(t_2, q_2|\boldsymbol{Z}_2)\}\right] \\
&\qquad = \mathrm{Var}\{\widehat{\theta}(t_1, q_1|\boldsymbol{Z}_1) - \theta(t_1, q_1|\boldsymbol{Z}_1)\} + \mathrm{Var}\{\widehat{\theta}(t_2, q_2|\boldsymbol{Z}_2) - \theta(t_2, q_2|\boldsymbol{Z}_2)\} \\
&\qquad\quad - 2 \times \mathrm{Cov}\{\widehat{\theta}(t_1, q_1|\boldsymbol{Z}_1) - \theta(t_1, q_1|\boldsymbol{Z}_1), \widehat{\theta}(t_2, q_2|\boldsymbol{Z}_2) - \theta(t_2, q_2|\boldsymbol{Z}_2)\}.
\end{aligned}$$

Now

$$\begin{aligned}
&\mathrm{Cov}\{\widehat{\theta}(t_1, q_1|\boldsymbol{Z}_1) - \theta(t_1, q_1|\boldsymbol{Z}_1), \widehat{\theta}(t_2, q_2|\boldsymbol{Z}_2) - \theta(t_2, q_2|\boldsymbol{Z}_2)\} \\
&\quad = \frac{1}{\widehat{\lambda}_0(\theta(t_1, q_1|\boldsymbol{Z}_1) + t_1)\widehat{\lambda}_0(\theta(t_2, q_2|\boldsymbol{Z}_2) + t_2)} \\
&\quad\quad \left[\mathrm{Cov}\left\{\int_0^{\theta(t_1,q_1|\boldsymbol{Z}_1)} \frac{\sum_{i=1}^n dM_i(t_1 + v)}{S^{(0)}(\widehat{\boldsymbol{\beta}}, t_1 + v)}, \int_0^{\theta(t_2,q_2|\boldsymbol{Z}_2)} \frac{\sum_{j=1}^n dM_j(t_2 + v)}{S^{(0)}(\widehat{\boldsymbol{\beta}}, t_2 + v)}\right\}\right. \\
&\quad\quad - \mathrm{Cov}\left\{\int_0^{\theta(t_1,q_1|\boldsymbol{Z}_1)} \frac{\sum_{i=1}^n dM_i(t_1 + v)}{S^{(0)}(\widehat{\boldsymbol{\beta}}, t_1 + v)}, \widetilde{\boldsymbol{A}}_2^{\mathrm{T}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right\} \\
&\quad\quad - \mathrm{Cov}\left\{\int_0^{\theta(t_2,q_2|\boldsymbol{Z}_2)} \frac{\sum_{j=1}^n dM_j(t_2 + v)}{S^{(0)}(\widehat{\boldsymbol{\beta}}, t_2 + v)}, \widetilde{\boldsymbol{A}}_1^{\mathrm{T}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right\} \\
&\quad\quad \left. + \mathrm{Cov}\left\{\widetilde{\boldsymbol{A}}_1^{\mathrm{T}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}), \widetilde{\boldsymbol{A}}_2^{\mathrm{T}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right\}\right] \\
&\quad = \frac{1}{\widehat{\lambda}_0(\theta(t_1, q_1|\boldsymbol{Z}_1) + t_1)\widehat{\lambda}_0(\theta(t_2, q_2|\boldsymbol{Z}_2) + t_2)} \\
&\quad\quad \left[\mathrm{Cov}\left\{\sum_{i=1}^n \int_0^{\eta_{\min} - t_{\max}} \frac{dM_i(t_{\max} + v)}{S^{(0)}(\widehat{\boldsymbol{\beta}}, t_{\max} + v)}, \sum_{j=1}^n \int_0^{\eta_{\min} - t_{\max}} \frac{dM_j(t_{\max} + v)}{S^{(0)}(\widehat{\boldsymbol{\beta}}, t_{\max} + v)}\right\}\right. \\
&\quad\quad \left. - 0 - 0 + \widetilde{\boldsymbol{A}}_1^{\mathrm{T}}\mathrm{Var}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\widetilde{\boldsymbol{A}}_2\right] \\
&\quad = \frac{1}{\widehat{\lambda}_0(\theta(t_1, q_1|\boldsymbol{Z}_1) + t_1)\widehat{\lambda}_0(\theta(t_2, q_2|\boldsymbol{Z}_2) + t_2)} \\
&\quad\quad \left[\sum_{i=j}^n \mathrm{Cov}\left\{\int_0^{\eta_{\min} - t_{\max}} \frac{dM_i(t_{\max} + v)}{S^{(0)}(\widehat{\boldsymbol{\beta}}, t_{\max} + v)}, \int_0^{\eta_{\min} - t_{\max}} \frac{dM_j(t_{\max} + v)}{S^{(0)}(\widehat{\boldsymbol{\beta}}, t_{\max} + v)}\right\}\right. \\
&\quad\quad \left. + \widetilde{\boldsymbol{A}}_1^{\mathrm{T}}\mathrm{Var}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\widetilde{\boldsymbol{A}}_2\right]
\end{aligned}$$

$$= \frac{1}{\widehat{\lambda}_0(\theta(t_1, q_1|\boldsymbol{Z}_1) + t_1)\widehat{\lambda}_0(\theta(t_2, q_2|\boldsymbol{Z}_2) + t_2)} \left[\mathrm{Var}\left\{\int_0^{\eta_{\min} - t_{\max}} \frac{\sum_{i=1}^n dM_i(t_{\max} + v)}{S^{(0)}(\widehat{\boldsymbol{\beta}}, t_{\max} + v)}\right\} + \widetilde{\boldsymbol{A}}_1^{\mathrm{T}} \mathrm{Var}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\widetilde{\boldsymbol{A}}_2\right],$$

where $\eta_{\min} = \min\{\theta(t_1, q_1|\boldsymbol{Z}_1) + t_1, \theta(t_2, q_2|\boldsymbol{Z}_2) + t_2\}$ and $t_{\max} = \max\{t_1, t_2\}$.

So, combining the above with results from Theorem 3.2.1 and taking into account the consistency of the estimators used in this formulation, we have the asymptotic result

$$\sqrt{n}[\{\widehat{\theta}(t_1, q_1|\boldsymbol{Z}_1) - \widehat{\theta}(t_2, q_2|\boldsymbol{Z}_2)\} - \{\theta(t_1, q_1|\boldsymbol{Z}_1) - \theta(t_2, q_2|\boldsymbol{Z}_2)\}] \to_d N(0, W(t)),$$

where

$$\begin{aligned} W(t) \;=\; & \frac{1}{\lambda_0(\theta(t_1, q_1|\boldsymbol{Z}_1) + t_1)^2}\left\{\boldsymbol{A}_1^{\mathrm{T}}(\boldsymbol{\Sigma}(\boldsymbol{\beta}, \tau))^{-1}\boldsymbol{A}_1 + \int_0^{\theta(t_1, q_1|\boldsymbol{Z}_1)} \frac{\lambda_0(t_1 + v)}{s^{(0)}(\boldsymbol{\beta}, t_1 + v)}dv\right\} \\ & + \frac{1}{\lambda_0(\theta(t_2, q_2|\boldsymbol{Z}_2) + t_2)^2}\left\{\boldsymbol{A}_2^{\mathrm{T}}(\boldsymbol{\Sigma}(\boldsymbol{\beta}, \tau))^{-1}\boldsymbol{A}_2 + \int_0^{\theta(t_2, q_2|\boldsymbol{Z}_2)} \frac{\lambda_0(t_2 + v)}{s^{(0)}(\boldsymbol{\beta}, t_2 + v)}dv\right\} \\ & - \frac{2}{\lambda_0(\theta(t_1, q_1|\boldsymbol{Z}_1) + t_1)\lambda_0(\theta(t_2, q_2|\boldsymbol{Z}_2) + t_2)} \\ & \times \left\{\int_0^{\eta_{\min} - t_{\max}} \frac{\lambda_0(t_{\max} + v)}{s^{(0)}(\boldsymbol{\beta}, t_{\max} + v)}dv + \boldsymbol{A}_1^{\mathrm{T}}(\boldsymbol{\Sigma}(\boldsymbol{\beta}, \tau))^{-1}\boldsymbol{A}_2\right\}, \end{aligned}$$

$$\boldsymbol{A}_j = \left[\log q_j \times \boldsymbol{Z}_j \exp\left(-\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_j\right) + \int_0^{\theta(t_j, q_j|\boldsymbol{Z}_j)} \frac{\boldsymbol{s}^{(1)}(\boldsymbol{\beta}, t_j + v)}{s^{(0)}(\boldsymbol{\beta}, t_j + v)}\lambda_0(t_j + v)d(t_j + v)\right]^{\mathrm{T}}.$$

□

With Theorem 3.3.1, calculating pointwise confidence intervals is possible. We may simply take $\{\widehat{\theta}(t_1, q_1|\boldsymbol{Z}_1) - \widehat{\theta}(t_2, q_2|\boldsymbol{Z}_2)\} \pm z^*_{1-\alpha/2}\sqrt{\widehat{W}(t)}$ for a $100(1-\alpha)\%$ confidence interval. Methods for calculating confidence bands and other methods for calculating confidence intervals are similar to those explained above.

3.4 Numerical Studies

3.4.1 Simulations

In order to demonstrate the asymptotic performance of our estimators $\widehat{\theta}(t, q|\boldsymbol{Z})$ (that it is consistent and asymptotically unbiased) and $\widehat{V}(t)$ (that it is consistent, asymptotically unbiased, and yields confidence intervals with the correct coverage probabilities) we conducted a simulation study. Survival times were generated under the assumption of proportional hazards with a baseline hazard of $\lambda(t|\boldsymbol{Z}_i) = t \exp(\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{Z}_i)$ with $\boldsymbol{\beta} = (1, 2)^{\mathrm{T}}$ and $\boldsymbol{Z}_i = (Z_{i1}, Z_{i2})^{\mathrm{T}}$, where $Z_{i1} \sim \mathrm{Bernoulli}(0.5)$ and $Z_{i2} \sim \mathrm{Uniform}(0, 1)$ for $i = 1, \ldots, n$. When generating censored data, censoring times followed an exponential distribution with rate parameter 0.17 ($\sim 10\%$ censoring) or 0.61 ($\sim 30\%$ censoring).

For the purposes of simulation, we considered median ($q = 0.5$) residual time (MRT) at $t = 0.25$ and $t = 0.75$. Covariate values of $Z = (0, 0.5)^{\mathrm{T}}$ were chosen. We used 1,000 replications in our simulation. Calculated quantities included the bias, sample standard error (i.e. the standard error of the MRT estimates), mean standard error (i.e. the mean of the standard error estimates), and the coverage probability for nominal 95% confidence intervals.

Simulation results can be seen in Table 3.3. Bias appears to be negligible. Sample standard errors (SSEs) and mean standard errors (MSEs) were very close to each other, regardless of sample size. Coverage probabilities closely matched nominal confidence levels. In order to demonstrate the performance of our estimator over time as well as the various methods of constructing confidence intervals and bands, we plotted results from a single replication of simulated data. As can be seen in Figure 3.2, all three methods give similar results, with similar widths for both confidence intervals and confidence bands.

We conducted similar simulations to evaluate the performance of $\widehat{W}(t)$. Survival times were again generated with a baseline hazard specified as $\lambda(t|Z_i) = t \exp(\beta^{\mathrm{T}} Z_i)$. We used $\beta = 1$ and $Z_i \sim \mathrm{Bernoulli}(0.5)$ for $i = 1, \ldots, n$. When generating censored data, censoring times followed an exponential distribution with rate parameter 0.17 ($\sim 10\%$ censoring) or

Table 3.3: Simulation results

		$t = 0.25$				$t = 0.75$			
Sample Size	% Censoring	Bias	SSE	MSE	95% CP	Bias	SSE	MSE	95% CP
$n = 200$	0	0.000	0.046	0.045	0.94	0.003	0.047	0.048	0.94
	10	-0.000	0.049	0.048	0.94	0.002	0.052	0.053	0.94
	30	-0.001	0.055	0.055	0.94	0.002	0.062	0.067	0.95
$n = 500$	0	-0.000	0.028	0.029	0.96	0.000	0.029	0.030	0.95
	10	0.001	0.030	0.030	0.95	0.001	0.031	0.033	0.95
	30	-0.000	0.034	0.035	0.95	0.000	0.042	0.041	0.94
$n = 1000$	0	0.001	0.020	0.020	0.95	0.001	0.021	0.021	0.96
	10	0.001	0.022	0.022	0.95	0.000	0.022	0.023	0.96
	30	-0.000	0.024	0.024	0.96	0.001	0.029	0.028	0.94

0.61 ($\sim 30\%$ censoring).

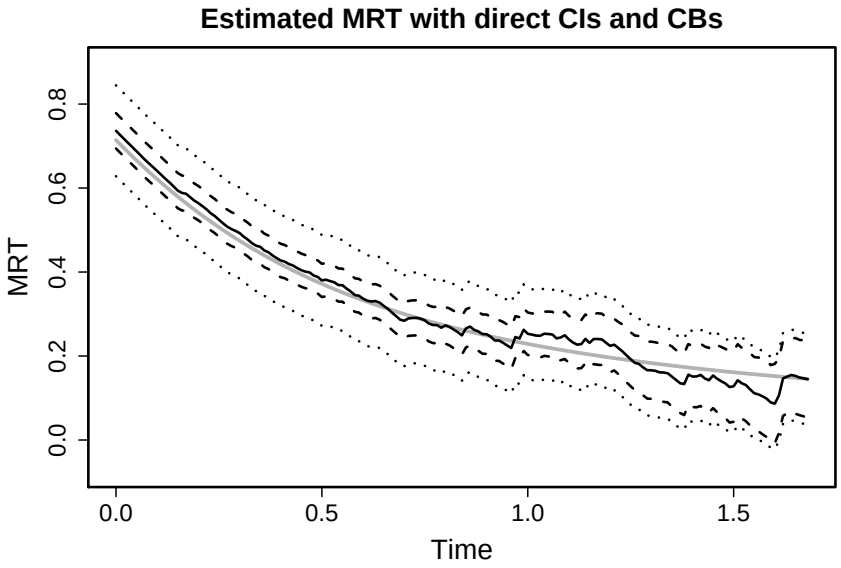
We considered the difference $\{\widehat{\theta}(t_1, q_1|Z_1) - \widehat{\theta}(t_2, q_2|Z_2)\}$ where $t_1 = 0.25$, $q_1 = 0.4$, $Z_1 = 0$, $t_2 = 0.75$, $q_2 = 0.6$, and $Z_2 = 1$. We used 10,000 replications in our simulation. We calculated bias, sample standard error, mean standard error, and the coverage probability for 95% confidence intervals associated with the difference $\{\widehat{\theta}(t_1, q_1|Z_1) - \widehat{\theta}(t_2, q_2|Z_2)\}$.

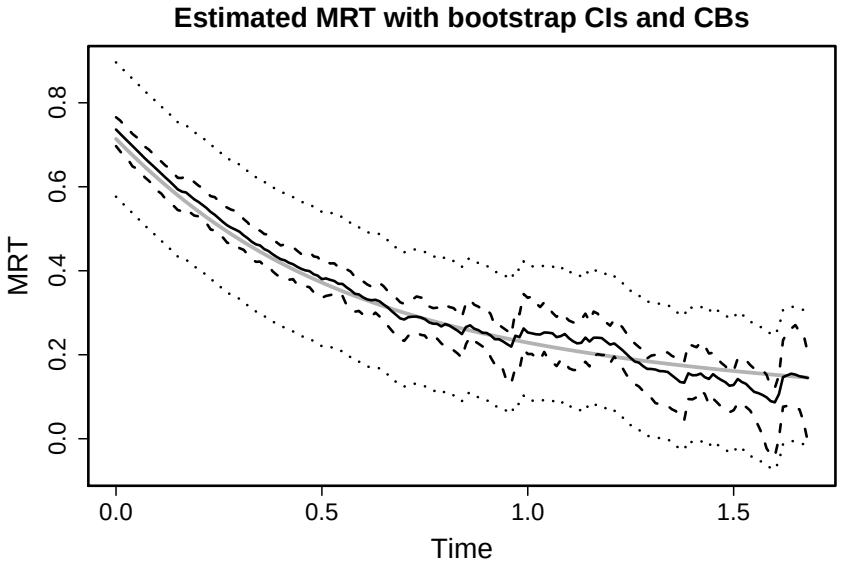
Results for the performance of $\widehat{W}(t)$ can be seen in Table 3.4. For most combinations of simulation parameters, bias appears to be negligible and sample standard errors (SSEs) and mean standard errors (MSEs) were very close to each other. Coverage probabilities were as expected.

3.4.2 Real data examples

We present two examples of analysis on existing data sets. For the first, we use data from a clinical trial in the treatment of carcinoma of the oropharynx as presented in Kalbfleisch and Prentice (2002). These data include 195 patients with squamous cell carcinoma of 3

Figure 3.2: Plots of MRT for simulated data with various methods for calculating confidence intervals and bands. The black solid curve represents estimated MRT, the gray solid curve represents true MRT, the dashed curve represents the 95% confidence interval, and the dotted curve represent the 95% confidence band.


Estimated MRT with direct CIs and CBs
MRT
0.0
0.2
0.4
0.6
0.8
0.0
0.5
1.0
1.5
Time


Estimated MRT with bootstrap CIs and CBs
MRT
0.0
0.2
0.4
0.6
0.8
0.0
0.5
1.0
1.5
Time

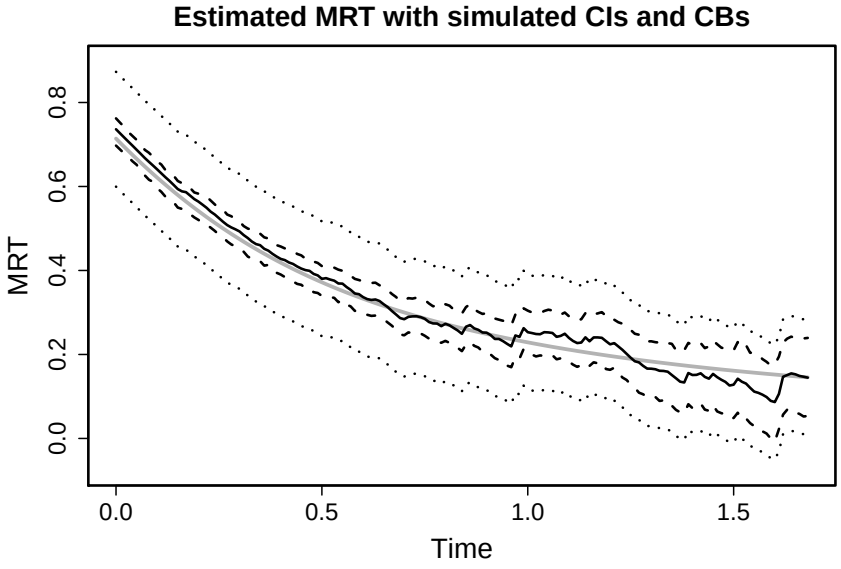

Estimated MRT with simulated CIs and CBs
MRT
0.0
0.2
0.4
0.6
0.8
0.0
0.5
1.0
1.5
Time

Table 3.4: Comparison simulation results

Sample Size	% Censoring	Bias	SSE	MSE	95% CP
$n = 200$	0	-0.003	0.091	0.092	0.95
	10	-0.002	0.098	0.100	0.94
	30	-0.006	0.125	0.120	0.94
$n = 500$	0	-0.002	0.057	0.058	0.94
	10	-0.001	0.062	0.061	0.95
	30	0.002	0.077	0.076	0.94
$n = 1000$	0	0.001	0.041	0.041	0.95
	10	0.000	0.044	0.044	0.95
	30	-0.002	0.055	0.054	0.95

sites in the oropharynx from the 6 largest of 16 total participating institutions. Patients were randomly assigned treatment with radiation therapy alone or radiation therapy with a chemotherapeutic agent. While many other characteristics were collected, the data set we examined included only: sex; treatment; tumor grade, site, T staging, and N staging; overall patient condition, date of entry, living status, and survival time. Of the 195 patients included, 142 were observed to fail.

We examined differences in MRTs to death across different T staging categories: I/II (35 patients), III (93 patients), or IV (67 patients). Regression results can be seen in Table 3.5 and reflect what we see when plotting MRTs (Figure 3.3): higher T stage is associated with increased hazard of death and lower MRT. The relationship between T stage and MRT is not statistically significant, however. As can be seen in Figure 3.4, 0 is well within the confidence bands for the difference between MRTs for any two values of T stage. It may be worth noting that for certain fixed time points, the difference in MRTs for T stages IV vs I/II and for IV vs III are significantly different from 0.

Nevertheless, our results highlight a benefit of examining MRTs: for patients with T

Table 3.5: Cox model results

Data Set	Variable	Hazard Ratio	95% Confidence Interval	P value
Carcinoma	T stage I/II	1.00	—	
	T stage III	1.24	0.77 to 2.02	0.007*
	T stage IV	2.01	1.22 to 3.31	
HIVNET 012	Nevirapine	1.00	—	—
	Zidovudine	0.92	0.70 to 1.21	0.550
	Maternal CD4 at pre-entry†	0.99	0.93 to 1.05	0.670
	Maternal HIV-1 RNA at pre-entry‡	1.41	1.15 to 1.72	0.001

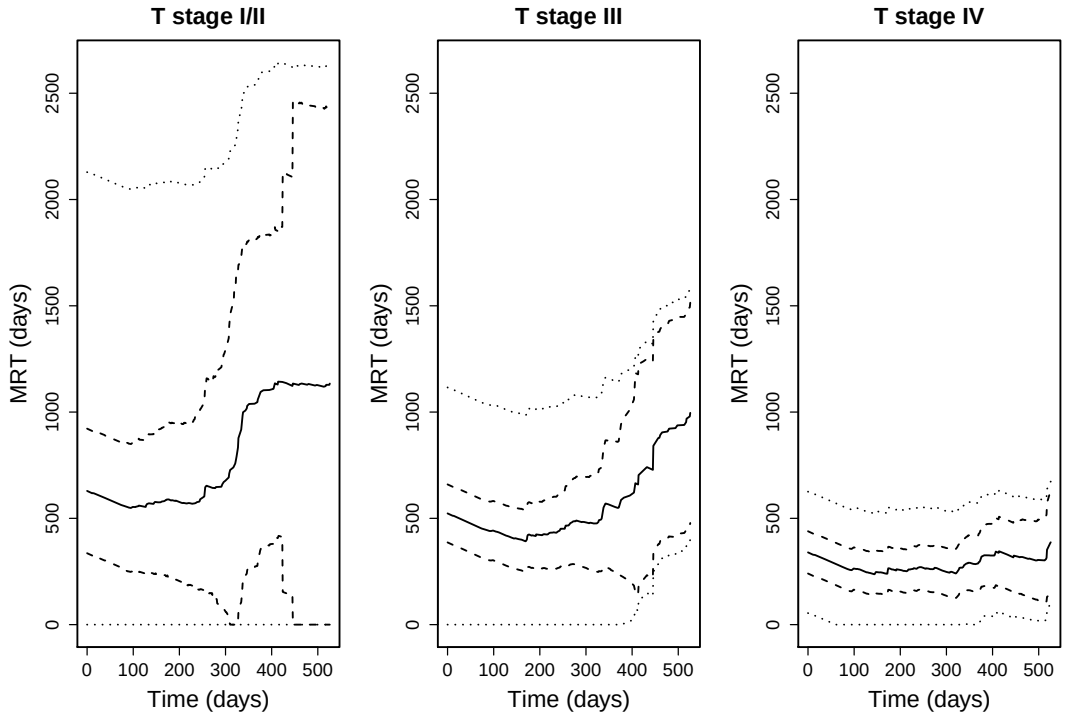
*From likelihood ratio test of complete versus null model. †Per 100 cells/μL. ‡ Per unit increase of $\log_{10}$ HIV-1 RNA copies/mL.

staging of I, II, or III, MRT changes markedly over time, increasing sharply after initially decreasing slightly. For example, patients with T stage III have a MRT of 422 days (95% CI 266-578 days) after having survived 200 days and a MRT of 855 days (95% CI 350-1360 days) after having survived 450 days. This would be important and welcome news to surviving patients and their caregivers alike.

Our second example uses data collected as part of the HIVNET 012 randomized trial (Jackson et al., 2003). This trial randomly assigned HIV-1 infected pregnant women in Kampala, Uganda to either a nevirapine- or zidovudine-based treatment. Their infants were followed and tested at pre-determined intervals for HIV-1. Data were also collected on adverse events through 6-8 weeks postpartum for mothers and 18 months for babies. The study enrolled 645 mothers: 313 assigned to nevirapine, 313 to zidovudine, and 19 to placebo. Within 18 months, 109 serious adverse events and 34 deaths were observed in the nevirapine group while 97 serious adverse events and 42 deaths were observed in the zidovudine group.

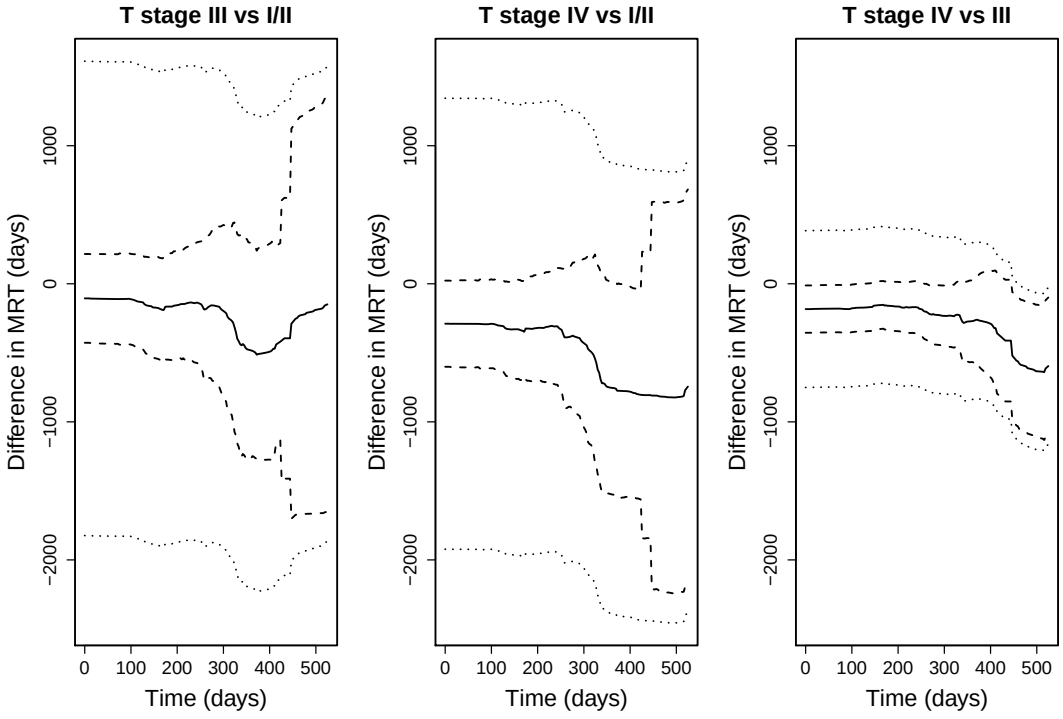
For our example we examined the relationship between the 10%ile of residual time to

Figure 3.3: Plots of MRTs for carcinoma data with pointwise confidence intervals and confidence bands. The solid curves represent estimated MRT, the dashed curves represent the pointwise 95% confidence intervals, and the dotted curves represent the 95% confidence bands.



death or serious adverse event and treatment group (nevirapine versus zidovudine), maternal CD4 at pre-entry, and maternal HIV-1 RNA at pre-entry. We present results from the Cox model in Table 3.5. In the multivariate model, only maternal HIV-1 RNA was significantly associated with time to death or serious adverse event, with a hazard ratio of 1.41 (95% CI 1.15-1.72) for a unit increase of $\log_{10}$ HIV-1 RNA copies/mL. A plot of 10%ile residual times for two different combinations of covariate values (a theoretical infant whose mother was treated with AZT, had a maternal CD4 of 600 cells/μL, and a $\log_{10}$ maternal viral load of 3.5 copies/mL and a theoretical infant whose mother was treated with NVP, had a maternal CD4 of 300 cells/μL, and a $\log_{10}$ maternal viral load of 5 copies/mL) can be seen in Figure 3.5, where the combinations were chosen to represent infants with relatively

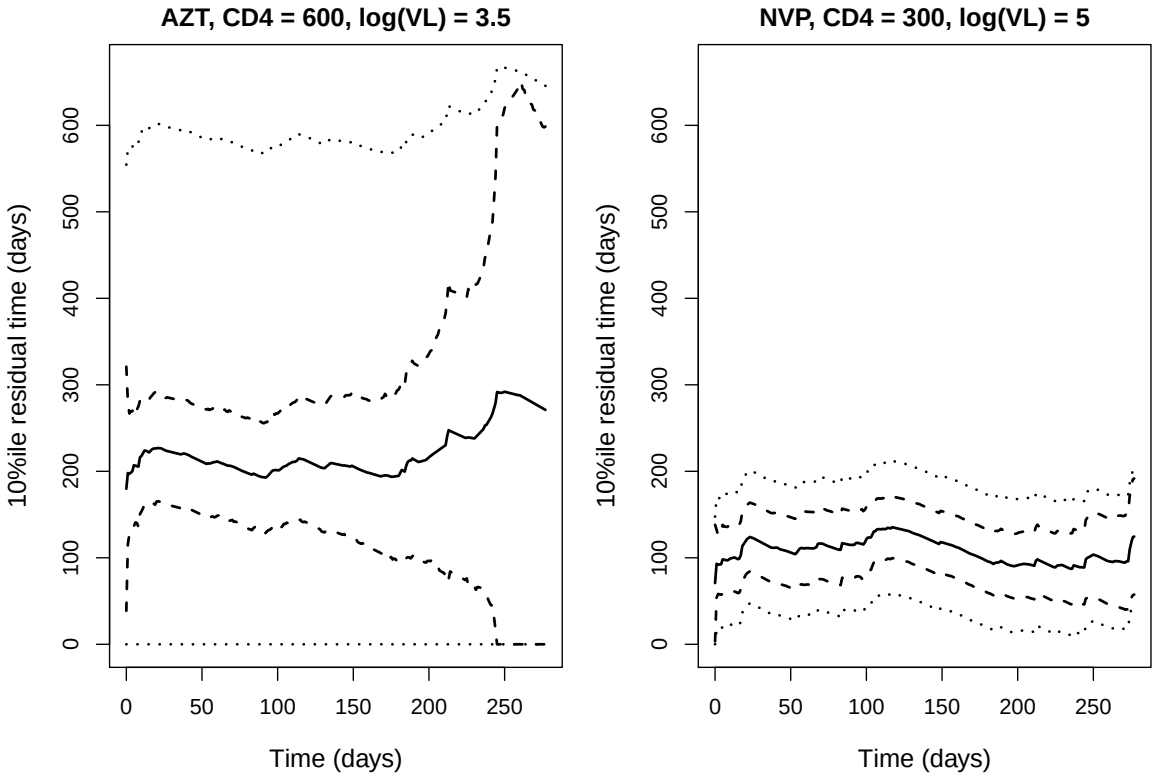
Figure 3.4: Plots of differences in MRTs for carcinoma data with pointwise confidence intervals and confidence bands. The solid curves represent estimated MRT, the dashed curves represent the pointwise 95% confidence intervals, and the dotted curves represent the 95% confidence bands.



protective or relatively non-protective characteristics. For both groups, the 10%ile residual times are fairly stable, particularly at early follow-up times. For any amount of time survived event-free less than 150 days, we expect 10% of infants to die or experience a serious adverse event within the next $\sim$ 200 days for the low-risk group and $\sim$ 100 days for the high-risk group.

When examining the difference in 10%ile residual times between the two different combinations of covariate values we conclude that it is not significant across all time as 0 is well within the limits of the confidence bands (see Figure 3.6). This is in spite of the fact that, for the majority of fixed time points, the difference is statistically significant. The lack of overall significance is driven to a large degree by the large increase in variance towards

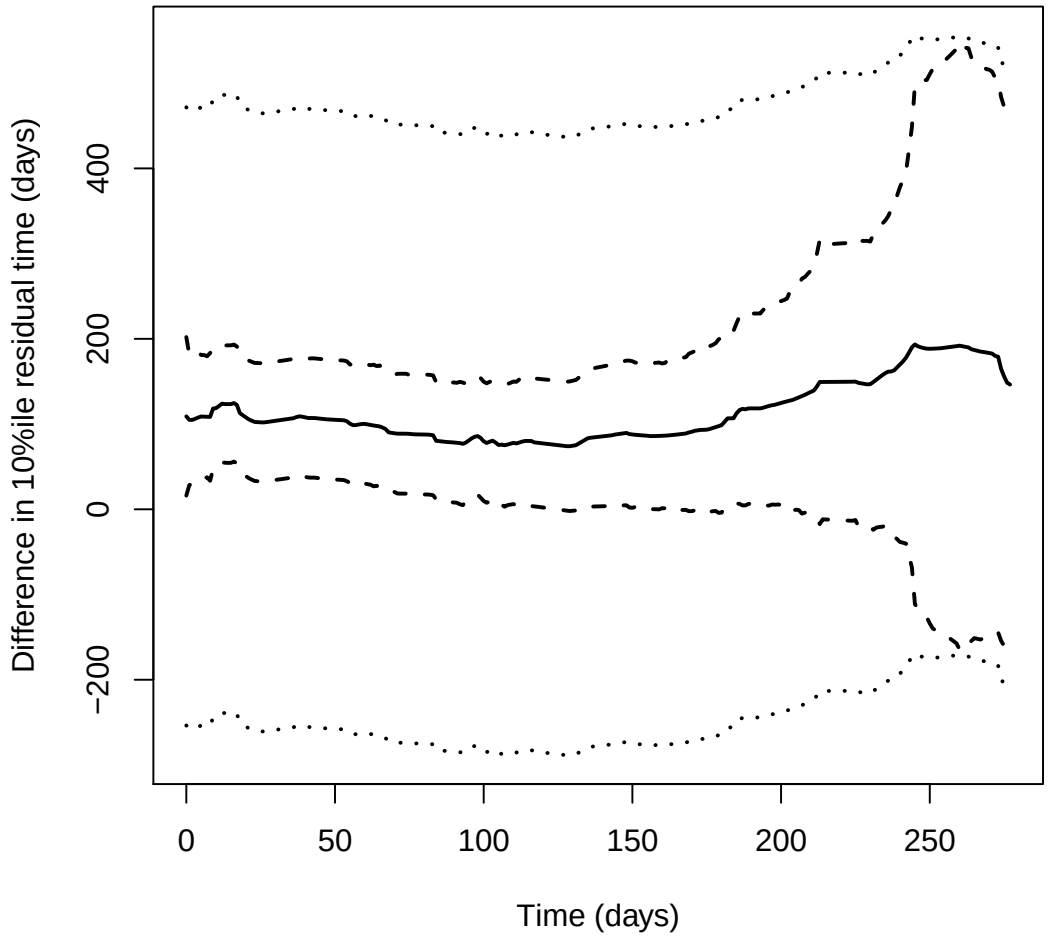
Figure 3.5: Plots of 10%ile residual times for HIVNET 012 data with pointwise confidence intervals and confidence bands. The solid curves represent estimated 10%ile residual time, the dashed curves represent the pointwise 95% confidence intervals, and the dotted curves represent the 95% confidence bands.



later times.

For both example data sets, we carried out model-checking to investigate the validity of the proportional hazards assumption. We used the test based on Schoenfeld residuals as detailed in Therneau and Grambsch (2000) with $g(t) = \log(t)$. For the carcinoma data, there was a slight suggestion of non-proportionality for the patient group with a T stage of IV, but this was not significant ($p = 0.09$). For the HIVNET 012 data, there was no indication of non-proportionality for any variable.

Figure 3.6: Plot of the difference in 10%ile residual times for HIVNET 012 data. The difference is between a theoretical infant whose mother was treated with AZT, had a maternal CD4 of 600 cells/μL, and a $\log_{10}$ maternal viral load of 3.5 copies/mL and a theoretical infant whose mother was treated with NVP, had a maternal CD4 of 300 cells/μL, and a $\log_{10}$ maternal viral load of 5 copies/mL. The solid curve represents estimated 10%ile residual time, the dashed curves represent the pointwise 95% confidence intervals, and the dotted curves represent the 95% confidence bands.



3.5 Discussion

In this chapter, we presented a model-based estimation method of residual time quantiles for censored time-to-event outcomes. As discussed earlier, in clinical practice knowing individuals' covariate-specific residual time quantiles shall assist both clinicians and patients to better understand how individually carried risk factors affect an event time such as cancer survival. The model-based estimation of residual time quantiles that we propose harnesses the popularity of the Cox proportional hazards model to provide a simple approach to estimating residual time quantiles with easy-to-compute variances.

Nonetheless, the Cox proportional hazards model that we used for methods development is not the only modeling tool for residual time quantiles estimation, as this model may not always fit the observed data. In particular, the proportionality assumption may be violated. One possible extension is to allow time-varying coefficients or time-varying covariates, such as covariate-time interaction, in the Cox model. This would allow for both relaxation of the assumption of proportionality and additional applications. For example, a patient undergoing a particular treatment regimen for cancer may be very interested in their median residual time to remission given their stage in the treatment process.

Alternatives to the Cox model, such as the accelerated failure time model or the additive hazard model, could also be used for estimating residual time quantiles, since the methods we propose in this chapter are general enough to be applied similarly to these models, although estimation and asymptotics will need to be developed separately.

There are some limits to our model-based estimation method. For one, it depends entirely on the correct specification of underlying models. While our method provides a different summary measure, it in no way escapes the limitations of the proportional hazards model, and is dependent on all the associated assumptions. Although the Cox model itself is semiparametric, which allows a great deal of flexibility, possible model mis-specification can still lead to biased estimate of residual time quantiles. Therefore, certain sensitivity analyses may be used to help detect such a bias and further provide guidance on alternative

models for more accurate and reliable model-based estimation.

Other limitations or concerns are inherited from the proportional hazards regression model. One such concern is the possibility of calculating residual time quantiles for (combinations of) covariates values that are not observed in the data. This is not an issue that is specific to our methods, but it bears reiteration. Any estimates for unobserved covariate values require some interpolation or extrapolation from the collected data and are subject to assumptions that individuals with said covariate values are well-represented by the underlying regression model.

Another concern that is common to any statistical analysis is that of multiple comparisons. In some ways, it is a more important concern for our methods than it is for inference on the regression parameters themselves. This is because our our ability to establish a covariate's effect is limited to comparing differences between two estimated residual time quantiles with different values for said covariate, instead of simply having a parameter that quantifies that covariate's effect. Furthermore, our method requires specification of a specific quantile and evaluation time as well as specific covariate values. While we have addressed the issue of comparisons at multiple time points by detailing how to compute confidence bands, the issue of comparisons for multiple covariate values or quantiles remains. Despite the increased concern about multiple comparisons, however, there are already many existing methods to adjust for them.

An additional limitation is that our estimator is only defined when

$$\widehat{\Lambda}_0(t) - \log(q)\exp(-\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{Z}) \leq \widehat{\Lambda}_0(\tau)$$

where τ is the largest observed failure time. Thus our estimator may not be calculable at later times, for smaller quantiles, for particularly low-risk patient characteristics, or some combination of the three. This is a bigger concern in studies with low event rates. Earlier censoring can offset this somewhat, though at the cost of increased variability. We plan to investigate extrapolation of the cumulative hazard estimate to allow for residual time quantile estimation even in studies with few events.

Despite these limitations, we feel that this method has a lot of potential as a tool in applied survival analysis. We also address some of these limitations in following chapters. Particularly, we extend our method to allow for time-varying covariates as well as develop an estimator based on the additive hazards model. We leave other considerations as avenues for future work.

Chapter 4

RESIDUAL TIME QUANTILES IN THE PRESENCE OF EXTERNAL TIME-VARYING COVARIATES IN THE PH MODEL

4.1 Introduction

As discussed in Chapter 3, there are several reasons why extending our estimator to allow for time-varying covariates is a useful exercise. First, the flexibility afforded by time-varying covariates may allow the Cox proportional hazards model to be applied to data where it would otherwise be a poor fit. In turn, this makes our own Cox model-based estimator more flexible and less reliant on the assumption of proportional hazards. Second, and perhaps more importantly, allowing for time-varying covariates greatly increases the applicability of our estimator. In particular, because we are fundamentally estimating times-to-event at different, progressive points in time, the assumption that all covariates values would remain the same, or at least not be subject to updated measurements, is very limiting. A patient visiting their caregiver for a follow-up appointment who is interested in their prognosis will certainly want said prognosis to reflect any changes in their environment or treatment since the initial diagnosis.

In this chapter, we extend our Cox model-based estimator for quantiles of residual time, developed in Chapter 3, to allow for time-varying covariates. While deriving asymptotic results analytically did not prove feasible, we discuss estimation or residual time quantiles in Section 4.2, along with variance estimation via bootstrap sampling. In Section 4.3 we demonstrate our estimator's performance on simulated data, including figures showing confidence intervals and bands. Additionally, we apply our method to real data from the UMARU impact study (UIS) as presented in Hosmer et al. (2008). Finally, in Section 4.4 we discuss the limitations to our method.

We should note that, in light of our discussion in Section 2.2.1, we assume any time-varying covariates to be external as defined by Kalbfleisch and Prentice (2002).

4.2 Model-based Estimation of Residual Time Quantiles

We are somewhat limited in estimating residual times when time-dependent variables are included in the model. This is because the residual time at a certain time t depends not only on the covariate values at that point in time $\boldsymbol{Z}(t)$ but also the covariate values for all future times preceding the residual time. Because we may be interested in residual time at any time point $t \geq 0$, it is in fact important to know the value of $\boldsymbol{Z}(t)$ for all time. We denote this complete knowledge as $\widetilde{\boldsymbol{Z}}$ as opposed to $\boldsymbol{Z}(t)$. So in practice we must know, or more plausibly assume, the entire trajectory of values for all covariates for all individuals. While this is certainly a limitation, it need not always be particularly fanciful. For example, a patient may be planned to continue a certain treatment for a set period of time before switching over to another, as in a cross-over trial.

Assume that T is a positive random variable, representing a subject's time-to-event. At a given time t, we first define the $(1-q)^{\text{th}}$ $(0 < q < 1)$ percentile residual time of a random variable T as the amount of additional time necessary for $(1-q) \times 100\%$ of the individuals still under observation at time t to fail. We denote this quantity as $\theta(t, q|\widetilde{\boldsymbol{Z}})$, where $\widetilde{\boldsymbol{Z}} \in \mathcal{R}^p$ is the associated p-dimensional vector of covariates. Let

$$S\{t + \theta(t, q|\widetilde{\boldsymbol{Z}})|\widetilde{\boldsymbol{Z}}\} = qS(t|\widetilde{\boldsymbol{Z}}), \tag{4.1}$$

where $S(\cdot|\widetilde{\boldsymbol{Z}})$ is the survival function. If expressed in terms of the cumulative hazard function, $\Lambda(\cdot|\widetilde{\boldsymbol{Z}}) = -\log S(\cdot|\widetilde{\boldsymbol{Z}})$, we then have $\Lambda\{t + \theta(t, q|\widetilde{\boldsymbol{Z}})|\widetilde{\boldsymbol{Z}}\} = \Lambda(t|\widetilde{\boldsymbol{Z}}) - \log q$.

Now consider the Cox proportional hazards model that assumes

$$\Lambda(t|\widetilde{\boldsymbol{Z}}) = \int_0^t \lambda_0(u) \exp\{\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{Z}(u)\} du, \tag{4.2}$$

where $\lambda_0(\cdot)$ is the unspecified baseline hazard function, $\boldsymbol{\beta} \in \mathcal{B} \subset \mathcal{R}^p$ is the p-dimensional

regression parameter, and $^{\mathrm{T}}$ denotes vector (matrix) transpose. Under model (4.2) we have

$$\begin{aligned}\int_0^{t+\theta(t,q|\widetilde{\boldsymbol{Z}})} \lambda_0(u)\exp\{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}(u)\}du &= \int_0^{t} \lambda_0(u)\exp\{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}(u)\}du - \log q \\ \int_t^{t+\theta(t,q|\widetilde{\boldsymbol{Z}})} \lambda_0(u)\exp\{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}(u)\}du &= -\log q. \end{aligned} \tag{4.3}$$

Unfortunately, we cannot obtain a closed-form solution for $\theta(t,q|\widetilde{\boldsymbol{Z}})$, though it can be solved for numerically. It is important to note that $\theta(t,q|\widetilde{\boldsymbol{Z}})$ only depends on covariate values between the evaluation time t and, somewhat recursively, the sum of the evaluation time and the quantile of residual time $t+\theta(t,q|\widetilde{\boldsymbol{Z}})$. It does not, however, depend on covariate values for any time prior to the evaluation time t. Therefore, if we wish to consider the historical values of time-varying covariates, they must be explicitly included in the Cox model (e.g. as cumulative exposures). As seen below, this has the important consequence that, conditional on survival to the point t, residual time quantiles at t for any individuals with equal covariate values for all points beyond t will themselves be equal.

In order to estimate $\theta(t,q|\widetilde{\boldsymbol{Z}})$ we need estimates for both $\boldsymbol{\beta}$ and $\Lambda_0(\cdot)$. Consider that data are collected in the form of $\{X_i, \Delta_i, \boldsymbol{Z}_i(t)\}$ for $i=1,\ldots,n$ with n being the sample size. For these data, $X_i = \min(T_i, C_i)$ where T_i is a failure time and C_i is a censoring time; $\Delta_i = I(T_i \le C_i)$; and $\boldsymbol{Z}_i(t)$ is a vector of covariates. Given $\boldsymbol{Z}_i(t)$, T_i and C_i are assumed to be independent. Note that $N_i(t) = I(X_i \le t, \delta_i = 1)$ and $Y_i(t) = I(X_i \ge t)$. Then under model (4.2), we can estimate $\boldsymbol{\beta}$ with $\widehat{\boldsymbol{\beta}}$, the standard estimator for Cox model regression coefficients (Fleming and Harrington, 2005), obtained by solving the partial score equations

$$\sum_{i=1}^{n}\int_0^{\infty}\left\{\boldsymbol{Z}_i(t) - \frac{\boldsymbol{S}^{(1)}(\boldsymbol{\beta},t)}{S^{(0)}(\boldsymbol{\beta},t)}\right\}dN_i(t) = 0,$$

where

$$S^{(0)}(\boldsymbol{\beta},t) = n^{-1}\sum_{i=1}^{n} Y_i(t)e^{\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_i(t)} \to_p s^{(0)}(\boldsymbol{\beta},t),$$

and

$$\boldsymbol{S}^{(1)}(\boldsymbol{\beta},t) = n^{-1}\sum_{i=1}^{n} Y_i(t)\boldsymbol{Z}_i(t)e^{\boldsymbol{\beta}^{\mathrm{T}}Z_i(t)} \to_p \boldsymbol{s}^{(1)}(\boldsymbol{\beta},t).$$

We can estimate the baseline cumulative hazard function $\Lambda_0(\cdot)$ with the estimator,

$$\widehat{\Lambda}_0(t) = \int_0^t \frac{\sum_{i=1}^n dN_i(s)}{\sum_{i=1}^n Y_i(s) \exp\{\widehat{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{Z}_i(s)\}}.$$

Substituting these standard estimators $\widehat{\boldsymbol{\beta}}$ and $\widehat{\Lambda}_0(\cdot)$ for their counterparts in equation (4.3), we can estimate $\theta(t, q|\widetilde{\boldsymbol{Z}})$ as the solution to

$$\int_t^{t+\widehat{\theta}(t,q|\widetilde{\boldsymbol{Z}})} \exp\{\widehat{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{Z}(u)\} d\widehat{\Lambda}_0(u) du = -\log q. \tag{4.4}$$

Because we cannot obtain a closed-form solution for $\widehat{\theta}(t, q|\widetilde{\boldsymbol{Z}})$, it is difficult to derive its asymptotic properties analytically. However, given the established consistency of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\Lambda}_0(\cdot)$, we can conclude that $\widehat{\theta}(t, q|\widetilde{\boldsymbol{Z}})$ will also be consistent. Nevertheless, we cannot estimate its variance. We are therefore limited to using bootstrap methods for estimation of confidence intervals and bands. This can be done in one of two ways. First, we can find the 2.5%ile and 97.5%ile of the bootstrap sample estimates of $\widehat{\theta}^*(t, q|\widetilde{\boldsymbol{Z}})$ at each time point in order to get pointwise intervals (Efron and Tibshirani, 1998). Second, we can calculate the bootstrap standard error associated with $\widehat{\theta}^*(t, q|\widetilde{\boldsymbol{Z}})$ and use it as an estimate for the standard error associated with $\widehat{\theta}(t, q|\widetilde{\boldsymbol{Z}})$, calculating intervals as usual for an asymptotically normal distribution. To obtain bands, we first calculate the maximum deviation within each bootstrap sample across time, $\sup_{0 \le t \le \tau} |\widehat{\theta}^*(t, q|\boldsymbol{Z}) - \widehat{\theta}(t, q|\boldsymbol{Z})|$, and then find the 95%ile across samples, c. We can then construct bands using $\widehat{\theta}(t, q|\boldsymbol{Z}) \pm c$.

4.3 Numerical Studies

4.3.1 Simulations

In order to demonstrate the asymptotic performance of our estimators $\widehat{\theta}(t, q|\widetilde{\boldsymbol{Z}})$ we conducted a simulation study. Survival times were generated under the assumption of proportional hazards with a baseline hazard of $\lambda\{t|Z_i(t)\} = t \exp\{\beta^{\mathrm{T}} Z_i(t)\}$ with $\beta = 1$ and $Z_i(t) = 1\{t > t_i^*\}$ or $1\{t \le t_i^*\}$ with equal probability, where $T_i^* \sim \Gamma(3, 5)$. We did not include censoring in our simulation.

For the purposes of simulation, we considered median ($q = 0.5$) residual time at $t = 0.5$. Covariate value was set to $Z(t) = 1\{t > 0.75\}$. Sample size was $n = 500$. We used 200 replications, each with 200 bootstrap samples of $n_{bs} = 500$, in our simulation. Calculated quantities included the bias, sample standard error (i.e. the standard error of the MRT estimates), mean standard error (i.e. the mean of the bootstrap standard error estimates), and the coverage probability for nominal 95% confidence intervals.

Bias was very small (0.002). Sample standard errors (SSEs) and mean standard errors (MSEs) were very close to each other.

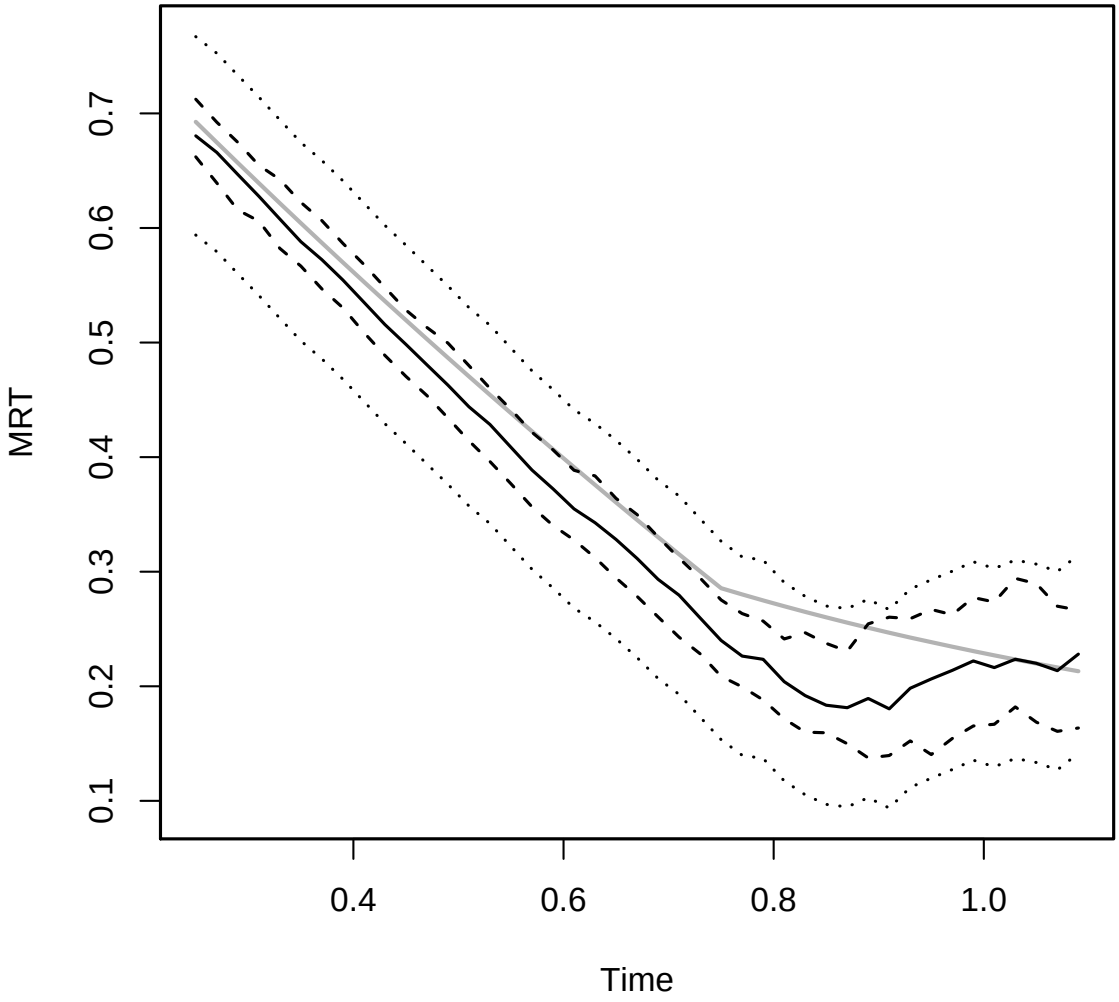
The coverage probabilities were equal for the standard error-based and quantile-based methods for generating intervals. Both were slightly lower (0.93) than the nominal level (0.95), suggesting that the bootstrap interval estimation is slightly anti-conservative. In order to demonstrate the performance of our estimator over time we plotted results from a single replication of simulated data. As can be seen in Figure 5.1, our estimator closely tracks the true median residual times.

4.3.2 Real data example

We present an example of analysis on existing data from the University of Massachusetts AIDS Research Unit (UMARU) IMPACT Study (UIS) as presented in Hosmer et al. (2008). These data include 628 subjects from a study conducted to compare treatment programs of different planned durations designed to reduce drug abuse and to prevent high-risk HIV behavior. In particular, the UIS sought to determine whether alternative residential treatment approaches are variable in effectiveness and whether efficacy depends on planned program duration. The data we used contained only a small subset of the variables recorded for the main study: age, Beck depression score, heroin/cocaine use prior to admission, IV drug use history at admission, number of prior drug treatments, race, treatment, site, length of treatment, and time to and indicator for return to drug use. Of the 628 subjects followed, 464 were observed to return to drug use.

We generated a time-varying covariate based on length of treatment to indicate, for a

Figure 4.1: Plots of MRTs for simulated data with bootstrap confidence intervals and bands. The black solid curve represents estimated MRT, the gray solid curve represents true MRT, the dashed curve represents the 95% confidence interval, and the dotted curve represent the 95% confidence band.



given time t, whether or not a particular patient was still undergoing treatment. This variable, which takes the value 1 when t is less than length of treatment and 0 otherwise, will be referred to as "ongoing treatment." We examined differences in MRTs to drug use for participants with different lengths of treatment or, in other words, differing value of ongoing treatment across time. For this application, it is somewhat difficult whether ongoing treatment should be classified as an external or internal time-varying covariate. Treatment length is determined *a priori* and therefore would typically be considered external. However, treatment is terminated in the event of a failure, so ongoing treatment does convey some information about failure status. Nevertheless, it is not as direct information as it would be with e.g. blood pressure where failure is death. For purposes of our example, we decided to consider ongoing treatment as an external time-varying covariate.

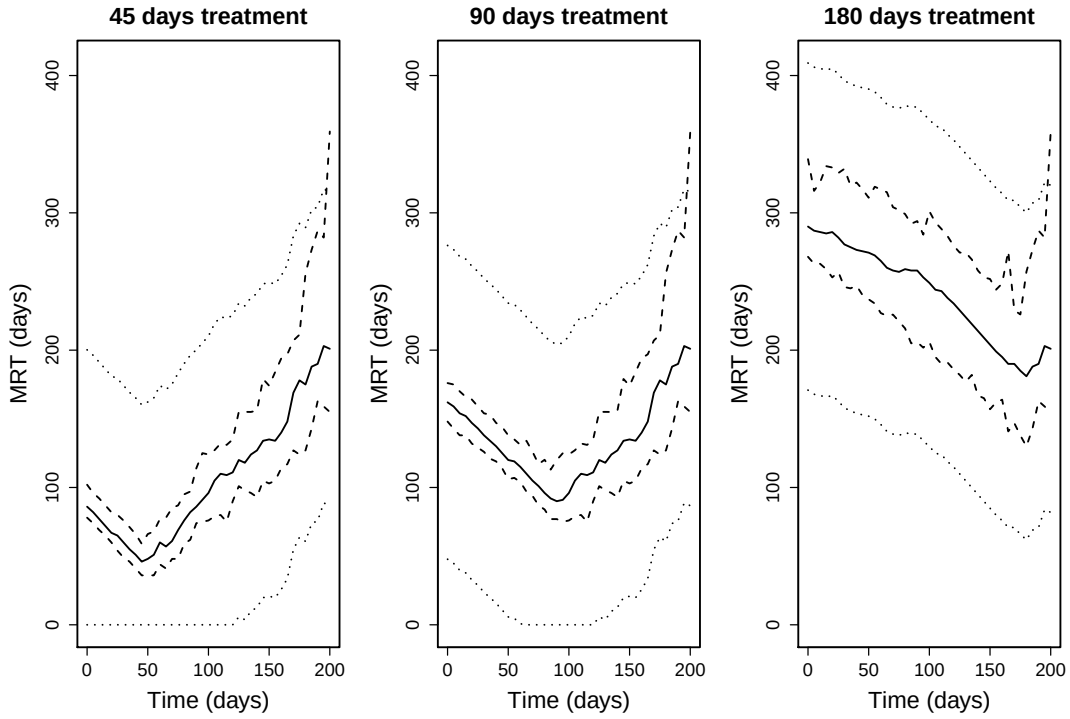
We found ongoing treatment to be associated with a hazard ratio of 0.08 (95% CI 0.06 to 0.11). Note that this analysis is only for illustrative purposes and we are ignoring all other covariates. In particular, we are not including the $apriori$ treatment group assigned to each patient. Furthermore, the nature of the study design is such that we would expect a rather extreme effect for the ongoing treatment covariate, so the hazard ratio of 0.08 is not as noteworthy as it might otherwise be.

We calculated MRTs for three different theoretical participants with varying ongoing treatment durations: 45 days, 90 days, or 180 days. Results can be seen in Figure 4.2. To some extent, they are what we expected. Theoretical participants with increased durations of ongoing treatment have much higher MRTs, at least initially. Necessarily, after ongoing treatment ends, the MRTs are the same all participants. While this may not be an unreasonable result, it does highlight the lack of ability for our method to take an individual's history into account without said history being explicitly included in the Cox regression model.

It is initially surprising that MRTs are decreasing for all participants while on treatment and increasing while off treatment. This may be due to one or both of two issues: first, that returning to drug use while on treatment is recorded as a premature end to treatment, meaning that what seems to be approaching end of treatment may actually be approaching

return to drug use; and second, that perhaps the end of treatment is in fact particularly hazardous in that many participants return to drug use very closely following the end up support provided on treatment. Regardless, we do see a very clear effect of the time-varying covariate on the estimated MRTs.

Figure 4.2: Plots of MRTs for UIS data with bootstrap confidence intervals and bands. The black solid curve represents estimated MRT, the dashed curves represents the 95% confidence interval, and the dotted curves represent the 95% confidence band.



4.4 Discussion

Given the importance of allowing for time-varying covariates in the estimation of covariate-specific quantiles of residual time, it is a bit disappointing that the development of such an estimator is not straightforward. While numerical solution for point estimation and bootstrap sampling for variance estimation does allow for application and inference in set-

tings with time-varying covariates, the implementation is not as elegant as it could be for a closed-form estimator with analytically established asymptotic properties. Furthermore, while these methods may be straightforward to apply for statisticians with advanced training, their relative complexity may limit adoption by academics in other fields and health care professionals.

Perhaps more problematic is the necessity of making assumptions about a patient's future values for all time-varying covariates. In some ways, this is not particularly limiting. For example, when discussing expected prognoses as they depend on length of treatment in a clinical trial it only makes sense to present results obtained from several different assumptions about the time-varying covariates' future trajectories. However, it would be nice to be able to present a somewhat more marginal estimation of residual time. That is, an estimate of residual time that is conditional on covariate values up until the evaluation time t but with no assumptions beyond then, instead relying on empirical sample or population averages. This would be particularly useful for covariates that vary more finely or are less easily controlled by the patient or caregiver. Examples might include the effect of pollution on the prognosis for a patient with a chronic respiratory illness.

Additionally, we have restricted consideration to external time-varying covariates. While difficult, extending our methods to allow for internal time-varying covariates as well would greatly expand their applicability.

Finally, while not specific to our methods, missing values for time-varying covariates can present a major challenge. In order to calculate our estimates we need covariate values for all individuals in the risk set at the time an event occurs. However, because events can occur at any time while measurement of time-varying covariates will likely only occur at certain fixed points in time, it is likely that covariate values will be missing for the majority of individuals. This will not always be the case. For example, in a cross-over trial current treatment will always be known (assuming intent-to-treat analysis). Contrarily, if we are modeling pollen level as a time-varying covariate it is likely that all or almost all individuals would be missing values at any given event time. This sort of missingness

necessitates imputation. There are many options for imputation: last observation carried forward, regression, stochastic, etc. As mentioned, this would be necessary for any time-to-event analysis involving time-varying covariates, but we emphasize it here as we have yet to explore the effects of imputation on our estimator's performance.

Chapter 5

RESIDUAL TIME QUANTILES IN THE ADDITIVE HAZARDS HAZARDS MODEL

5.1 Introduction

Not all time-to-event data can meet the assumptions of proportional hazards. While there was only a small suggestion on non-proportionality in the example data sets we considered in Section 3.4.2, for many data sets the proportionality assumption may be violated fairly severely. It is therefore important to show that an estimator similar to the one we discussed in Chapter 3 can be developed from other models. The additive hazards model, while not as common as the Cox proportional hazards model, is nevertheless relatively widely used for modeling and analyzing time-to-event data. The model formulation is also fairly simple and has well-documented asymptotic properties. Therefore, it presented an appealing alternative to the Cox proportional hazards model.

In this chapter, we develop an estimator for quantiles of residual time using the additive hazards model as a starting point. As shown in Section 5.2, the newly developed estimator allows estimation of covariate-specific residual time quantiles. We demonstrate our estimator's consistency, determine its limiting distribution, and provide a consistent estimator for its variance. Also included are discussions of methods for obtaining confidence intervals and bands that do not rely on direct estimation of the variance. We further develop our method in Section 5.3, determining the limiting distribution for a difference between two estimators of covariate-specific residual time quantiles and thereby allowing formal testing. In Section 5.4 we demonstrate our estimator's performance on simulated data, including figures showing confidence intervals and bands. Additionally, we apply our method to two real data sets: the VA lung cancer data as presented in Kalbfleisch and Prentice (2002) and

the HIVNET 012 data as presented in Jackson et al. (2003) and compare our results to those we obtained using the Cox model. Finally, in Section 5.5 we discuss the limitations to our estimator.

5.2 Model-based Estimation of Residual Time Quantiles

Assume that T is a positive random variable, representing a subject's time-to-event. At a given time t, we first define the $(1-q)^{\text{th}}$ $(0 < q < 1)$ percentile residual time of a random variable T as the amount of additional time necessary for $(1-q) \times 100\%$ of the individuals still under observation at time t to fail. We denote this quantity as $\theta(t, q|\boldsymbol{Z})$, where $\boldsymbol{Z} \in \mathcal{R}^p$ is the associated p-dimensional vector of covariates. Let

$$S\{t + \theta(t, q|\boldsymbol{Z})|\boldsymbol{Z}\} = qS(t|\boldsymbol{Z}), \tag{5.1}$$

where $S(\cdot|\boldsymbol{Z})$ is the survival function. If expressed in terms of the cumulative hazard function, $\Lambda(\cdot|\boldsymbol{Z}) = -\log S(\cdot|\boldsymbol{Z})$, we then have $\Lambda\{t + \theta(t, q|\boldsymbol{Z})|\boldsymbol{Z}\} = \Lambda(t|\boldsymbol{Z}) - \log q$.

Now consider the additive hazard model that assumes

$$\Lambda(t|\boldsymbol{Z}) = \Lambda_0(t) + t\boldsymbol{\beta}^{\text{T}}\boldsymbol{Z}, \tag{5.2}$$

where $\Lambda_0(\cdot)$ is the unspecified baseline cumulative hazard function, $\boldsymbol{\beta} \in \mathcal{B} \subset \mathcal{R}^p$ is the p-dimensional regression parameter, and $^{\text{T}}$ denotes vector (matrix) transpose. Under model (5.2) we have

$$\Lambda_0\{t + \theta(t, q|\boldsymbol{Z})\} + \theta(t, q|\boldsymbol{Z})\boldsymbol{\beta}^{\text{T}}\boldsymbol{Z} = \Lambda_0(t) - \log q.$$

Unfortunately we cannot obtain a closed-form solution for $\theta(t, q|\boldsymbol{Z})$, though it can be solved for numerically.

In order to estimate $\theta(t, q|\boldsymbol{Z})$ we need estimates for both $\boldsymbol{\beta}$ and $\Lambda_0(\cdot)$. Consider that data are collected in the form of $(X_i, \Delta_i, \boldsymbol{Z}_i)$ for $i = 1, \ldots, n$ with n being the sample size. For these data, $X_i = \min(T_i, C_i)$ where T_i is a failure time and C_i is a censoring time;

$\Delta_i = I(T_i \leq C_i)$; and $\boldsymbol{Z}_i$ is a vector of covariates. Given $\boldsymbol{Z}_i$, T_i and C_i are assumed to be independent. Note that $N_i(t) = I(X_i \leq t, \delta_i = 1)$ and $Y_i(t) = I(X_i \geq t)$.

We can estimate $\boldsymbol{\beta}$ with

$$\widehat{\boldsymbol{\beta}} = \frac{\sum_{i=1}^{n} \int_0^\infty Y_i(t)\{\boldsymbol{Z}_i - \overline{\boldsymbol{Z}}(t)\}^{\otimes 2} dt}{\sum_{i=1}^{n} \int_0^\infty \{\boldsymbol{Z}_i - \overline{\boldsymbol{Z}}(t)\} dN_i(t)},$$

the estimator proposed by Lin and Ying (1994), where $\overline{\boldsymbol{Z}}(t) = \frac{\sum_{i=1}^{n} Y_i(t)\boldsymbol{Z}_i(t)}{\sum_{i=1}^{n} Y_i(t)}$. Similarly, we can estimate the baseline cumulative hazard function $\Lambda_0(\cdot)$ with

$$\widehat{\Lambda}_0(t) = \int_0^t \frac{\sum_{i=1}^{n} \{dN_i(s) - Y_i(s)\widehat{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{Z}_i ds\}}{\sum_{i=1}^{n} Y_i(s)},$$

again proposed by Lin and Ying (1994).

Using these estimators $\widehat{\boldsymbol{\beta}}$ and $\widehat{\Lambda}_0(\cdot|\boldsymbol{Z})$, we can estimate $\theta(t, q|\boldsymbol{Z})$ as the solution to

$$\widehat{\Lambda}_0\{t + \widehat{\theta}(t, q|\boldsymbol{Z})\} + \widehat{\theta}(t, q|\boldsymbol{Z})\widehat{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{Z} = \widehat{\Lambda}_0(t) - \log q. \tag{5.3}$$

Note that $\widehat{\theta}(t, q|\boldsymbol{Z})$ is only defined when $\widehat{\Lambda}_0\{t + \widehat{\theta}(t, q|\boldsymbol{Z})\} + \widehat{\theta}(t, q|\boldsymbol{Z})\widehat{\boldsymbol{\beta}}^{\mathrm{T}} \boldsymbol{Z} + \log q \leq \widehat{\Lambda}_0(\tau)$ where τ is the largest observed failure time. For $\widehat{\theta}(t, q|\boldsymbol{Z})$, we have the following asymptotic properties, summarized in Theorem 5.2.1.

Theorem 5.2.1. *We assume conditions A-C (see below) as adapted from Chen et al. (2002) hold. If, as previously developed, $\widehat{\theta}(t, q|\boldsymbol{Z})$ is the solution to*

$$\widehat{\Lambda}_0\{t + \widehat{\theta}(t, q|\boldsymbol{Z})\} + \widehat{\theta}(t, q|\boldsymbol{Z})\widehat{\boldsymbol{\beta}}^{T} \boldsymbol{Z} = \widehat{\Lambda}_0(t) - \log q,$$

then as $n \to \infty$, for a given $\boldsymbol{Z}$, $\sqrt{n}\{\widehat{\theta}(t, q|\boldsymbol{Z}) - \theta(t, q|\boldsymbol{Z})\}$ converges weakly to a zero-mean Gaussian process whose variance function at t can be estimated consistently by

$$\frac{1}{\left[\widehat{\lambda}_0\{\widehat{\theta}(t, q|\boldsymbol{Z}) + t\} + \widehat{\boldsymbol{\beta}}^{T} \boldsymbol{Z}\right]^2} \left[\boldsymbol{C}^{\tau} \boldsymbol{A}^{-1} \boldsymbol{B} \boldsymbol{A}^{-1} \boldsymbol{C} - 2\boldsymbol{C}^{\tau} \boldsymbol{A}^{-1} \boldsymbol{D} + \int_0^{\widehat{\theta}(t,q|\boldsymbol{Z})} \frac{n \sum_{i=1}^{n} dN_i(t + v)}{\{\sum_{i=1}^{n} Y_i(t + v)\}^2}\right],$$

denoted as $\widehat{V}(t)$*, where*

$$\boldsymbol{A} = n^{-1}\sum_{i=1}^{n}\int_0^{\infty} Y_i(t)\{\boldsymbol{Z}_i - \overline{\boldsymbol{Z}}(t)\}^{\otimes 2}dt,$$
$$\boldsymbol{B} = n^{-1}\sum_{i=1}^{n}\int_0^{\infty} \{\boldsymbol{Z}_i - \overline{\boldsymbol{Z}}(t)\}^{\otimes 2}dN_i(t),$$
$$\boldsymbol{C} = \int_0^{\theta(t,q|\boldsymbol{Z})} \overline{\boldsymbol{Z}}(t+v)dv - \theta(t,q|\boldsymbol{Z})\boldsymbol{Z},$$

and

$$\boldsymbol{D} = \int_0^{\theta(t,q|\boldsymbol{Z})} \frac{\sum_{i=1}^{n}\{\boldsymbol{Z}_i - \overline{\boldsymbol{Z}}(t+v)\}dN_i(t+v)}{\sum_{i=1}^{n} Y_i(t+v)}.$$

We state conditions A-C as required by Theorem 5.2.1 and adapted from Chen et al. (2002):

A. There exists a time $t_0 > 0$ such that $\lim_{n\to\infty}\sum_{i=1}^{n} Y_i(t_0) > 0$.

B. There exists an integrable function $v(t)$ such that, for any $t \in [0, t_0]$,

$$n^{-1}\sum_{i=1}^{n} Y_i(t)(\boldsymbol{Z}_i - \overline{\boldsymbol{Z}})^{\otimes 2}\lambda_i(t|\boldsymbol{Z}_i) - v(t) \to 0,$$

in probability.

C. For any $\varepsilon > 0$,

$$n^{-1}\sum_{i=1}^{n}\int_0^{t_0} Y_i(s)\lambda(t|\boldsymbol{Z}_i)||\boldsymbol{Z}_i - \overline{\boldsymbol{Z}}||^2 I\{n^{-1}||\boldsymbol{Z}_i - \overline{\boldsymbol{Z}}|| > \varepsilon\}ds \to 0,$$

in probability.

Proof of Theorem 5.2.1. That $\sqrt{n}\{\widehat{\theta}(t,q|\boldsymbol{Z}) - \theta(t,q|\boldsymbol{Z})\}$ converges weakly to a zero-mean Gaussian process follows directly from the established convergence of the individual estimators and the application of the continuous mapping theorem. Therefore we need only consider the asymptotic variance function in detail.

We begin with two equations: one based on the true values,

$$\Lambda_0\{t+\theta(t,q|\boldsymbol{Z})\}+\theta(t,q|\boldsymbol{Z})\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}=\Lambda_0(t)-\log q,$$

and one based on the estimated values,

$$\widehat{\Lambda}_0\{t+\widehat{\theta}(t,q|\boldsymbol{Z})\}+\widehat{\theta}(t,q|\boldsymbol{Z})\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{Z}=\widehat{\Lambda}_0(t)-\log q.$$

Taking the differences between the left- and right-hand sides of both equations yields

$$\widehat{\Lambda}_0\{\widehat{\theta}(t,q|\boldsymbol{Z})+t\}-\Lambda_0\{\theta(t,q|\boldsymbol{Z})+t\}+\widehat{\theta}(t,q|\boldsymbol{Z})\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{Z}-\theta(t,q|\boldsymbol{Z})\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}=\widehat{\Lambda}_0(t)-\Lambda_0(t).$$

Examining the left-hand side first, we have

$$\begin{aligned}\widehat{\Lambda}_0\{\widehat{\theta}(t,q|\boldsymbol{Z})+t\}-\Lambda_0\{\theta(t,q|\boldsymbol{Z})+t\} &= \widehat{\Lambda}_0\{\widehat{\theta}(t,q|\boldsymbol{Z})+t\}-\widehat{\Lambda}_0\{\theta(t,q|\boldsymbol{Z})+t\}\\ &\quad+\widehat{\Lambda}_0\{\theta(t,q|\boldsymbol{Z})+t\}-\Lambda_0\{\theta(t,q|\boldsymbol{Z})+t\}\\ &\approx \widehat{\lambda}_0\{\theta(t,q|\boldsymbol{Z})+t\}\{\widehat{\theta}(t,q|\boldsymbol{Z})-\theta(t,q|\boldsymbol{Z})\}\\ &\quad+\widehat{\Lambda}_0\{\theta(t,q|\boldsymbol{Z})+t\}-\Lambda_0\{\theta(t,q|\boldsymbol{Z})+t\},\end{aligned}$$

where the approximation is due to Taylor's expansion, and

$$\begin{aligned}&\widehat{\theta}(t,q|\boldsymbol{Z})\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{Z}-\theta(t,q|\boldsymbol{Z})\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}\\ &\qquad=\widehat{\theta}(t,q|\boldsymbol{Z})\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{Z}-\widehat{\theta}(t,q|\boldsymbol{Z})\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}+\widehat{\theta}(t,q|\boldsymbol{Z})\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}-\theta(t,q|\boldsymbol{Z})\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}\\ &\qquad=\widehat{\theta}(t,q|\boldsymbol{Z})\boldsymbol{Z}^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})+\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}\{\widehat{\theta}(t,q|\boldsymbol{Z})-\theta(t,q|\boldsymbol{Z})\}.\end{aligned}$$

We therefore have the expression

$$\begin{aligned}\widehat{\Lambda}_0(t)-\Lambda_0(t) &\approx \widehat{\lambda}_0\{\theta(t,q|\boldsymbol{Z})+t\}\{\widehat{\theta}(t,q|\boldsymbol{Z})-\theta(t,q|\boldsymbol{Z})\}\\ &\quad+\widehat{\Lambda}_0\{\theta(t,q|\boldsymbol{Z})+t\}-\Lambda_0\{\theta(t,q|\boldsymbol{Z})+t\}\\ &\quad+\widehat{\theta}(t,q|\boldsymbol{Z})\boldsymbol{Z}^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})+\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}\{\widehat{\theta}(t,q|\boldsymbol{Z})-\theta(t,q|\boldsymbol{Z})\}.\end{aligned}$$

Rearranging yields the approximation

$$\begin{aligned}\widehat{\theta}(t,q|\boldsymbol{Z})-\theta(t,q|\boldsymbol{Z}) &\approx \frac{1}{\widehat{\lambda}_0\{\theta(t,q|\boldsymbol{Z})+t\}+\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}}\Bigg(\widehat{\Lambda}_0(t)-\Lambda_0(t)\\ &\quad-\Big[\widehat{\Lambda}_0\{\theta(t,q|\boldsymbol{Z})+t\}-\Lambda_0\{\theta(t,q|\boldsymbol{Z})+t\}\Big]-\widehat{\theta}(t,q|\boldsymbol{Z})\boldsymbol{Z}^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\Bigg).\end{aligned}$$

From Lee and Hyun (2011) we know that

$$\widehat{\Lambda}_0(t) - \Lambda_0(t) = -\left(\int_0^t \overline{\boldsymbol{Z}}(u)du\right)^{\mathrm{T}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + \int_0^t \frac{\sum_{i=1}^n dM_i(u)}{\sum_{i=1}^n Y_i(u)},$$

so we can rewrite the overall approximation as

$$\begin{aligned}\widehat{\theta}(t,q|\boldsymbol{Z}) - \theta(t,q|\boldsymbol{Z}) \;\approx\; & \frac{1}{\widehat{\lambda}_0\{\theta(t,q|\boldsymbol{Z}) + t\} + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}}\left\{-\int_t^{\theta(t,q|\boldsymbol{Z})+t}\frac{\sum_{i=1}^n dM_i(u)}{\sum_{i=1}^n Y_i(u)}\right. \\ & \left. + \left[-\widehat{\theta}(t,q|\boldsymbol{Z})\boldsymbol{Z} + \int_t^{\theta(t,q|\boldsymbol{Z})+t}\overline{\boldsymbol{Z}}(u)du\right]^{\mathrm{T}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right\}\end{aligned}$$

We can also perform a substitution, setting $u = t + v$, and integrating with respect to v. This yields the approximation

$$\begin{aligned}\widehat{\theta}(t,q|\boldsymbol{Z}) - \theta(t,q|\boldsymbol{Z}) \;\approx\; & \frac{1}{\widehat{\lambda}_0\{\theta(t,q|\boldsymbol{Z}) + t\} + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}}\left\{-\int_0^{\theta(t,q|\boldsymbol{Z})}\frac{\sum_{i=1}^n dM_i(t+v)}{\sum_{i=1}^n Y_i(t+v)}\right. \\ & \left. + \left[-\widehat{\theta}(t,q|\boldsymbol{Z})\boldsymbol{Z} + \int_0^{\theta(t,q|\boldsymbol{Z})}\overline{\boldsymbol{Z}}(t+v)du\right]^{\mathrm{T}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\right\}\end{aligned}$$

Applying the martingale central limit theorem as in Lin and Ying (1994) and Lee and Hyun (2011), it follows that the variance function can be consistently estimated by

$$\begin{aligned}\frac{1}{\left[\widehat{\lambda}_0\{\widehat{\theta}(t,q|\boldsymbol{Z}) + t\} + \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{Z}\right]^2}\Bigg[&\boldsymbol{C}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{C} - 2\boldsymbol{C}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{D} \\ &+ \int_0^{\widehat{\theta}(t,q|\boldsymbol{Z})}\frac{\sum_{i=1}^n dN_i(t+v)}{\{\sum_{i=1}^n Y_i(t+v)\}^2}dv\Bigg].\end{aligned}$$

□

With Theorem 5.2.1, calculating pointwise confidence intervals is possible. We may simply take $\widehat{\theta}(t,q|\boldsymbol{Z}) \pm z^*_{1-\alpha/2}\sqrt{\widehat{V}(t)}$ for a $100(1-\alpha)\%$ confidence interval. Moreover, we can use asymptotic results to obtain confidence bands as well. Specifically, we know that

$$\sup_{0\le t\le\tau}\frac{\sqrt{n}\left|\widehat{\theta}(t,q|\boldsymbol{Z}) - \theta(t,q|\boldsymbol{Z})\right|}{V(\tau)} \to_p \sup_{0\le t\le 1}|W(t)|$$

where $W(t)$ is standard Brownian motion. We can therefore solve

$$\Pr\left(\sup_{0\leq t\leq 1}|W(t)|\leq c\right)=\frac{4}{\pi}\sum_{k=0}^{\infty}\frac{(-1)^k}{2k+1}e^{-\frac{\pi^2(2k+1)^2}{8c^2}}\approx 1-\alpha$$

for c and compute $100(1-\alpha)\%$ confidence bands using $\widehat{\theta}(t,q|\boldsymbol{Z})\pm c\sqrt{\widehat{V}(\tau)}$.

In addition, we can compute both confidence bands and intervals using either bootstrap or simulation methods. Using the bootstrap method, we simply find the 2.5%ile and 97.5%ile of the bootstrap sample estimates of $\widehat{\theta}^*(t,q|\boldsymbol{Z})$ at each time point in order to get pointwise intervals (Efron and Tibshirani, 1998). To obtain bands, we first calculate the maximum deviation within each bootstrap sample across time, $\sup_{0\leq t\leq\tau}|\widehat{\theta}^*(t,q|\boldsymbol{Z})-\widehat{\theta}(t,q|\boldsymbol{Z})|$, and then find the 95%ile across samples, c. We can then construct bands using $\widehat{\theta}(t,q|\boldsymbol{Z})\pm c$.

The simulation method is in fact another way of resampling (Parzen et al., 1994). In the formulation of $\widehat{\theta}(t,q|\boldsymbol{Z})-\theta(t,q|\boldsymbol{Z})$, we replace $dM_i(u)$ with $G_i dN_i(u)$ where $G_i\sim N(0,1)$. For each simulated sample, all G_i are randomly generated and we compute the estimated residual time quantile for that sample, $\widetilde{\theta}(t,q|\boldsymbol{Z})$. At each time point, the 2.5%ile and 97.5%ile of the deviations $\widetilde{\theta}(t,q|\boldsymbol{Z})-\theta(t,q|\boldsymbol{Z})$ are calculated and added to the estimate $\widehat{\theta}(t,q|\boldsymbol{Z})$ to get lower and upper pointwise confidence intervals, respectively. Calculating bands is the same as with the bootstrap: we find the maximum deviation within each simulated sample across time, $\sup_{0\leq t\leq\tau}|\widetilde{\theta}(t,q|\boldsymbol{Z})-\widehat{\theta}(t,q|\boldsymbol{Z})|$, and then find the 95%ile across samples, c. We can then construct bands using $\widehat{\theta}(t,q|\boldsymbol{Z})\pm c$.

5.3 Comparing residual time quantiles

While being able to estimate covariate-specific residual time quantiles and their variance is useful, in most practical applications it is also important to be able to carry out comparisons between different covariate values and determine if any observed difference is statistically significant. We may also be interested in formally comparing residual times at different fixed time points or for different quantiles. All of these tasks require being able to estimate the covariance between two different residual time quantiles.

We can extend results from Theorem 5.2.1 to establish asymptotic properties for the differences between quantiles of residual time for different sets of covariate values, evaluation times, and quantiles. These properties are summarized in Theorem 5.3.1.

Theorem 5.3.1. *We assume conditions A-C (see above) hold. If, as previously developed,* $\widehat{\theta}(t, q|\boldsymbol{Z})$ *is the solution to* $\widehat{\Lambda}_0\{t + \widehat{\theta}(t, q|\boldsymbol{Z})\} + \widehat{\theta}(t, q|\boldsymbol{Z})\widehat{\boldsymbol{\beta}}^T\boldsymbol{Z} = \widehat{\Lambda}_0(t) - \log q$ *, then as* $n \to \infty$*, for a given* $\boldsymbol{Z}_1$ *and* $\boldsymbol{Z}_2$ *and fixed* t_1*,* t_2*,* q_1*, and* q_2*,* $\sqrt{n}\{\widehat{\theta}(t_1, q_1|\boldsymbol{Z}_1) - \widehat{\theta}(t_2, q_2|\boldsymbol{Z}_2)\}$ *converges weakly to a Gaussian process with mean* $\sqrt{n}\{\theta(t_1, q_1|\boldsymbol{Z}_1) - \theta(t_2, q_2|\boldsymbol{Z}_2)\}$ *and whose variance function at* t *can be estimated consistently by*

$$
\begin{aligned}
&\frac{1}{\left[\widehat{\lambda}_0\{\widehat{\theta}(t_1, q_1|\boldsymbol{Z}_1) + t_1\} + \widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}_1\right]^2} \\
&\times\left[\boldsymbol{C}_1^T\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{C}_1 - 2\boldsymbol{C}_1^T\boldsymbol{A}^{-1}\boldsymbol{D}_1 + \int_0^{\widehat{\theta}(t_1,q_1|\boldsymbol{Z}_1)} \frac{\sum_{i=1}^n dN_i(t_1+v)}{\{\sum_{i=1}^n Y_i(t_1+v)\}^2}dv\right] \\
&+\frac{1}{\left[\widehat{\lambda}_0\{\widehat{\theta}(t_2, q_2|\boldsymbol{Z}_2) + t_2\} + \widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}_2\right]^2} \\
&\times\left[\boldsymbol{C}_2^T\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{C}_2 - 2\boldsymbol{C}_2^T\boldsymbol{A}^{-1}\boldsymbol{D}_2 + \int_0^{\widehat{\theta}(t_2,q_2|\boldsymbol{Z}_2)} \frac{\sum_{i=1}^n dN_i(t_2+v)}{\{\sum_{i=1}^n Y_i(t_2+v)\}^2}dv\right] \\
&-\frac{2}{\left[\widehat{\lambda}_0\{\widehat{\theta}(t_1, q_1|\boldsymbol{Z}_1) + t_1\} + \widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}_1\right]\left[\widehat{\lambda}_0\{\widehat{\theta}(t_2, q_2|\boldsymbol{Z}_2) + t_2\} + \widehat{\boldsymbol{\beta}}^T\boldsymbol{Z}_2\right]} \\
&\times\Bigg[\boldsymbol{C}_1^T\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{C}_2 - \boldsymbol{C}_1^T\boldsymbol{A}^{-1}\boldsymbol{D}_2 - \boldsymbol{C}_2^T\boldsymbol{A}^{-1}\boldsymbol{D}_1 \\
&+\int_0^{\widehat{\eta}_{\min}-t_{\max}} \frac{\sum_{i=1}^n dN_i(t_{\max}+v)}{\{\sum_{i=1}^n Y_i(t_{\max}+v)\}^2}dv\Bigg]
\end{aligned}
$$

denoted as $\widehat{W}(t)$*, where*

$$
\begin{aligned}
\boldsymbol{A} &= n^{-1}\sum_{i=1}^n \int_0^\infty Y_i(t)\{\boldsymbol{Z}_i - \overline{\boldsymbol{Z}}(t)\}^{\otimes 2}dt, \\
\boldsymbol{B} &= n^{-1}\sum_{i=1}^n \int_0^\infty \{\boldsymbol{Z}_i - \overline{\boldsymbol{Z}}(t)\}^{\otimes 2}dN_i(t),
\end{aligned}
$$

$$\boldsymbol{C}_j = \int_0^{\theta(t_j,q_j|\boldsymbol{Z}_j)} \overline{\boldsymbol{Z}}(t_j+v)dv - \theta(t_j,q_j|\boldsymbol{Z}_j)\boldsymbol{Z}_j,$$
$$\boldsymbol{D}_j = \int_0^{\theta(t_j,q_j|\boldsymbol{Z}_j)} \frac{\sum_{i=1}^n \{\boldsymbol{Z}_i - \overline{\boldsymbol{Z}}(t_j+v)dN_i(t_j+v)\}}{\sum_{i=1}^n Y_i(t_j+v)},$$

and

$$\widehat{\eta}_{\min} = \min\{\widehat{\theta}(t_1,q_1|\boldsymbol{Z}_1)+t_1, \widehat{\theta}(t_2,q_2|\boldsymbol{Z}_2)+t_2\},$$

for $j = 1, 2$.

Proof of Theorem 5.3.1. That $\sqrt{n}\{\widehat{\theta}(t_1,q_1|\boldsymbol{Z}_1) - \widehat{\theta}(t_2,q_2|\boldsymbol{Z}_2)\}$ converges weakly to a Gaussian process with mean $\sqrt{n}\{\theta(t_1,q_1|\boldsymbol{Z}_1) - \theta(t_2,q_2|\boldsymbol{Z}_2)\}$ follows directly from the established convergence of the individual estimators and the application of the continuous mapping theorem. Therefore we need only consider the asymptotic variance function in detail.

From the proof of Theorem 5.2.1, we have

$$\widehat{\theta}(t_j,q_j|\boldsymbol{Z}_j) - \theta(t_j,q_j|\boldsymbol{Z}_j) \approx \frac{1}{\widehat{\lambda}_0\{\theta(t_j,q_j|\boldsymbol{Z}_j)+t_j\} + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_j} \times \left\{ -\int_0^{\theta(t_j,q_j|\boldsymbol{Z}_j)} \frac{\sum_{i=1}^n dM_i(t_j+v)}{\sum_{i=1}^n Y_i(t_j+v)} + \boldsymbol{C_j}^{\mathrm{T}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \right\}$$

In order to calculate $\mathrm{Var}(\widehat{\theta}(t_1,q_1|\boldsymbol{Z}_1) - \widehat{\theta}(t_2,q_2|\boldsymbol{Z}_2))$, we note that

$$\begin{aligned}
&\mathrm{Var}(\widehat{\theta}(t_1,q_1|\boldsymbol{Z}_1) - \widehat{\theta}(t_2,q_2|\boldsymbol{Z}_2)) \\
&\quad = \mathrm{Var}\left[\{\widehat{\theta}(t_1,q_1|\boldsymbol{Z}_1) - \theta(t_1,q_1|\boldsymbol{Z}_1)\} - \{\widehat{\theta}(t_2,q_2|\boldsymbol{Z}_2) - \theta(t_2,q_2|\boldsymbol{Z}_2)\}\right] \\
&\quad = \mathrm{Var}\{\widehat{\theta}(t_1,q_1|\boldsymbol{Z}_1) - \theta(t_1,q_1|\boldsymbol{Z}_1)\} + \mathrm{Var}\{\widehat{\theta}(t_2,q_2|\boldsymbol{Z}_2) - \theta(t_2,q_2|\boldsymbol{Z}_2)\} \\
&\qquad - 2 \times \mathrm{Cov}\{\widehat{\theta}(t_1,q_1|\boldsymbol{Z}_1) - \theta(t_1,q_1|\boldsymbol{Z}_1), \widehat{\theta}(t_2,q_2|\boldsymbol{Z}_2) - \theta(t_2,q_2|\boldsymbol{Z}_2)\}.
\end{aligned}$$

Now

$$
\begin{aligned}
& \mathrm{Cov}\{\widehat{\theta}(t_1,q_1|\boldsymbol{Z}_1)-\theta(t_1,q_1|\boldsymbol{Z}_1),\widehat{\theta}(t_2,q_2|\boldsymbol{Z}_2)-\theta(t_2,q_2|\boldsymbol{Z}_2)\} \\
= & \frac{1}{\left[\widehat{\lambda}_0\{\theta(t_1,q_1|\boldsymbol{Z}_1)+t_1\}+\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_1\right]\left[\widehat{\lambda}_0\{\theta(t_2,q_2|\boldsymbol{Z}_2)+t_2\}+\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_2\right]} \\
& \left[\mathrm{Cov}\left\{\int_0^{\theta(t_1,q_1|\boldsymbol{Z}_1)}\frac{\sum_{i=1}^n dM_i(t_1+v)}{\sum_{i=1}^n Y_i(t_1+v)},\int_0^{\theta(t_2,q_2|\boldsymbol{Z}_2)}\frac{\sum_{i=1}^n dM_i(t_2+v)}{\sum_{i=1}^n Y_i(t_2+v)}\right\}\right. \\
& -\mathrm{Cov}\left\{\int_0^{\theta(t_1,q_1|\boldsymbol{Z}_1)}\frac{\sum_{i=1}^n dM_i(t_1+v)}{\sum_{i=1}^n Y_i(t_1+v)},\boldsymbol{C}_2^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\right\} \\
& -\mathrm{Cov}\left\{\int_0^{\theta(t_2,q_2|\boldsymbol{Z}_2)}\frac{\sum_{i=1}^n dM_i(t_2+v)}{\sum_{i=1}^n Y_i(t_2+v)},\boldsymbol{C}_1^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\right\} \\
& \left.+\mathrm{Cov}\left\{\boldsymbol{C}_1^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}),\boldsymbol{C}_2^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\right\}\right] \\
= & \frac{1}{\left[\widehat{\lambda}_0\{\theta(t_1,q_1|\boldsymbol{Z}_1)+t_1\}+\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_1\right]\left[\widehat{\lambda}_0\{\theta(t_2,q_2|\boldsymbol{Z}_2)+t_2\}+\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_2\right]} \\
& \left[\mathrm{Cov}\left\{\sum_{i=1}^n\int_0^{\eta_{\min}-t_{\max}}\frac{dM_i(t_{\max}+v)}{\sum_{i=1}^n Y_i(t_{\max}+v)},\sum_{i=1}^n\int_0^{\eta_{\min}-t_{\max}}\frac{dM_i(t_{\max}+v)}{\sum_{i=1}^n Y_i(t_{\max}+v)}\right\}\right. \\
& \left.-\boldsymbol{C}_2^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{D}_1-\boldsymbol{C}_1^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{D}_2+\boldsymbol{C}_1^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{C}_2\right] \\
= & \frac{1}{\left[\widehat{\lambda}_0\{\theta(t_1,q_1|\boldsymbol{Z}_1)+t_1\}+\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_1\right]\left[\widehat{\lambda}_0\{\theta(t_2,q_2|\boldsymbol{Z}_2)+t_2\}+\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_2\right]} \\
& \left[\sum_{i=1}^n\mathrm{Cov}\left\{\int_0^{\eta_{\min}-t_{\max}}\frac{dM_i(t_{\max}+v)}{\sum_{i=1}^n Y_i(t_{\max}+v)},\int_0^{\eta_{\min}-t_{\max}}\frac{dM_i(t_{\max}+v)}{\sum_{i=1}^n Y_i(t_{\max}+v)}\right\}\right. \\
& \left.+\boldsymbol{C}_2^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{D}_1-\boldsymbol{C}_1^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{D}_2+\boldsymbol{C}_1^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{C}_2\right] \\
= & \frac{1}{\left[\widehat{\lambda}_0\{\theta(t_1,q_1|\boldsymbol{Z}_1)+t_1\}+\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_1\right]\left[\widehat{\lambda}_0\{\theta(t_2,q_2|\boldsymbol{Z}_2)+t_2\}+\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_2\right]} \\
& \left[\mathrm{Var}\left\{\int_0^{\eta_{\min}-t_{\max}}\frac{\sum_{i=1}^n dM_i(t_{\max}+v)}{\sum_{i=1}^n Y_i(t_{\max}+v)}\right\}\right. \\
& \left.+\boldsymbol{C}_2^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{D}_1-\boldsymbol{C}_1^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{D}_2+\boldsymbol{C}_1^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{C}_2\right],
\end{aligned}
$$

where $\eta_{\min} = \min\{\theta(t_1, q_1|\boldsymbol{Z}_1) + t_1, \theta(t_2, q_2|\boldsymbol{Z}_2) + t_2\}$ and $t_{\max} = \max\{t_1, t_2\}$.

So, combining the above with results from Theorem 5.2.1 and taking into account the consistency of the estimators used in this formulation, we can estimate the asymptotic variance with

$$\begin{aligned}
\widehat{W}(t) \;=\; & \frac{1}{\left[\widehat{\lambda}_0\{\widehat{\theta}(t_1, q_1|\boldsymbol{Z}_1) + t_1\} + \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{Z}_1\right]^2} \\
& \times \left[\boldsymbol{C}_1^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{C}_1 - 2\boldsymbol{C}_1^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{D}_1 + \int_0^{\widehat{\theta}(t_1, q_1|\boldsymbol{Z}_1)} \frac{\sum_{i=1}^n dN_i(t_1+v)}{\{\sum_{i=1}^n Y_i(t_1+v)\}^2} dv\right] \\
& + \frac{1}{\left[\widehat{\lambda}_0\{\widehat{\theta}(t_2, q_2|\boldsymbol{Z}_2) + t_2\} + \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{Z}_2\right]^2} \\
& \times \left[\boldsymbol{C}_2^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{C}_2 - 2\boldsymbol{C}_2^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{D}_2 + \int_0^{\widehat{\theta}(t_2, q_2|\boldsymbol{Z}_2)} \frac{\sum_{i=1}^n dN_i(t_2+v)}{\{\sum_{i=1}^n Y_i(t_2+v)\}^2} dv\right] \\
& - \frac{2}{\left[\widehat{\lambda}_0\{\widehat{\theta}(t_1, q_1|\boldsymbol{Z}_1) + t_1\} + \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{Z}_1\right]\left[\widehat{\lambda}_0\{\widehat{\theta}(t_2, q_2|\boldsymbol{Z}_2) + t_2\} + \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{Z}_2\right]} \\
& \times \left[\boldsymbol{C}_1^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{C}_2 - \boldsymbol{C}_1^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{D}_2 - \boldsymbol{C}_2^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{D}_1\right. \\
& \left. + \int_0^{\widehat{\eta}_{\min} - t_{\max}} \frac{\sum_{i=1}^n dN_i(t_{\max}+v)}{\{\sum_{i=1}^n Y_i(t_{\max}+v)\}^2} dv\right].
\end{aligned}$$

□

With Theorem 5.3.1, calculating pointwise confidence intervals is possible. We may simply take $\{\widehat{\theta}(t_1, q_1|\boldsymbol{Z}_1) - \widehat{\theta}(t_2, q_2|\boldsymbol{Z}_2)\} \pm z^*_{1-\alpha/2}\sqrt{\widehat{W}(t)}$ for a $100(1-\alpha)\%$ confidence interval. Methods for calculating confidence bands and other methods for calculating confidence intervals are similar to those explained above.

5.4 Numerical Studies

5.4.1 Simulations

In order to demonstrate the asymptotic performance of our estimators $\widehat{\theta}(t, q|\boldsymbol{Z})$ (that it is consistent and asymptotically unbiased) and $\widehat{V}(t)$ (that it is consistent, asymptotically unbi-

ased, and yields confidence intervals with the correct coverage probabilities) we conducted a simulation study. Survival times were generated under the assumption of additive hazards with a baseline hazard of $\lambda(t|\boldsymbol{Z}_i) = t\exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_i)$ with $\boldsymbol{\beta} = (1,2)^{\mathrm{T}}$ and $\boldsymbol{Z}_i = (Z_{i1}, Z_{i2})^{\mathrm{T}}$, where $Z_{i1} \sim \mathrm{Bernoulli}(0.5)$ and $Z_{i2} \sim \mathrm{Uniform}(0,1)$ for $i = 1,\ldots,n$. When generating censored data, censoring times followed an exponential distribution with rate parameter 0.195 ($\sim 10\%$ censoring) or 0.72 ($\sim 30\%$ censoring).

For the purposes of simulation, we considered median ($q = 0.5$) residual time at $t = 0.25$ and $t = 0.75$. Covariate values of $Z = (0, 0.5)^{\mathrm{T}}$ were chosen. We used 10,000 replications in our simulation. Calculated quantities included the bias, sample standard error (i.e. the standard error of the MRT estimates), mean standard error (i.e. the mean of the standard error estimates), and the coverage probability for nominal 95% confidence intervals.

Table 5.1: Simulation results

		$t = 0.25$				$t = 0.75$			
Sample Size	% Censoring	Bias	SSE	MSE	95% CP	Bias	SSE	MSE	95% CP
$n = 200$	0	0.000	0.060	0.062	0.95	-0.001	0.065	0.063	0.94
	10	-0.000	0.063	0.065	0.95	0.001	0.070	0.069	0.94
	30	0.000	0.073	0.073	0.94	0.003	0.090	0.089	0.95
$n = 500$	0	0.000	0.038	0.039	0.96	-0.000	0.041	0.040	0.95
	10	-0.001	0.040	0.041	0.95	0.000	0.045	0.044	0.94
	30	-0.000	0.045	0.047	0.95	0.001	0.058	0.056	0.94
$n = 1000$	0	-0.000	0.027	0.028	0.96	-0.000	0.029	0.029	0.95
	10	0.000	0.028	0.029	0.95	-0.000	0.032	0.031	0.94
	30	-0.000	0.032	0.033	0.96	0.000	0.040	0.040	0.95

Simulation results can be seen in Table 5.1. Bias appears to be negligible. Sample standard errors (SSEs) and mean standard errors (MSEs) were very close to each other, re-

gardless of sample size. Coverage probabilities closely matched nominal confidence levels. In order to demonstrate the performance of our estimator over time we plotted results from a single replication of simulated data (Figure 5.1). As can be seen in our estimated median residual time closely tracks the true median residual time, particularly for earlier evaluation times.

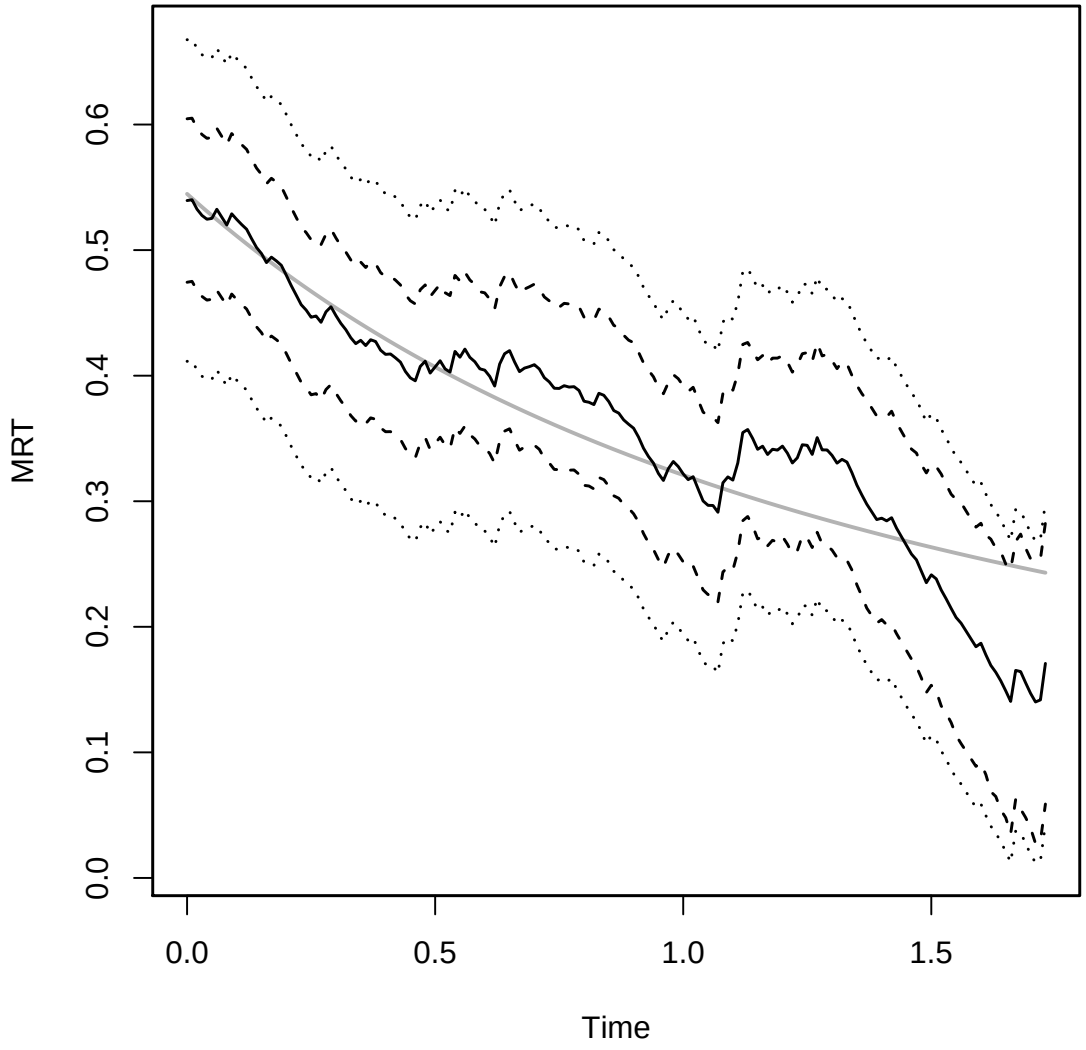
We conducted similar simulations to evaluate the performance of $\widehat{W}(t)$. Survival times were again generated under the assumption of additive hazards with a baseline hazard of $\lambda(t|\boldsymbol{Z}_i) = t\exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}_i)$ with $\boldsymbol{\beta} = (1,2)^{\mathrm{T}}$ and $\boldsymbol{Z}_i = (Z_{i1}, Z_{i2})^{\mathrm{T}}$, where $Z_{i1} \sim \mathrm{Bernoulli}(0.5)$ and $Z_{i2} \sim \mathrm{Uniform}(0,1)$ for $i = 1, \ldots, n$. When generating censored data, censoring times followed an exponential distribution with rate parameter 0.195 ($\sim 10\%$ censoring) or 0.72 ($\sim 30\%$ censoring).

We considered the difference $\{\widehat{\theta}(t_1, q_1|\boldsymbol{Z}_1) - \widehat{\theta}(t_2, q_2|\boldsymbol{Z}_2)\}$ where $t_1 = 0.25$, $q_1 = 0.4$, $\boldsymbol{Z}_1 = (0, 0.75)^{\mathrm{T}}$, $t_2 = 0.75$, $q_2 = 0.6$, and $\boldsymbol{Z}_2 = (1, 0.25)^{\mathrm{T}}$. We used 10,000 replications in our simulation. We calculated bias, sample standard error, mean standard error, and the coverage probability for 95% confidence intervals associated with the difference $\{\widehat{\theta}(t_1, q_1|\boldsymbol{Z}_1) - \widehat{\theta}(t_2, q_2|\boldsymbol{Z}_2)\}$.

Table 5.2: Comparison simulation results

Sample Size	% Censoring	Bias	SSE	MSE	95% CP
$n = 200$	0	-0.002	0.078	0.081	0.956
	10	-0.001	0.081	0.085	0.962
	30	-0.002	0.096	0.099	0.956
$n = 500$	0	-0.000	0.049	0.051	0.961
	10	-0.000	0.051	0.054	0.960
	30	-0.000	0.059	0.062	0.959
$n = 1000$	0	-0.000	0.034	0.036	0.960
	10	-0.000	0.036	0.038	0.960
	30	-0.000	0.042	0.044	0.958

Figure 5.1: Plots of MRT for simulated data with confidence intervals and bands. The black solid curve represents estimated MRT, the gray solid curve represents true MRT, the dashed curves represents the 95% confidence interval, and the dotted curves represent the 95% confidence band.



Results for the performance of $\widehat{W}(t)$ can be seen in Table 5.2. While bias was negligible, mean standard errors (MSEs) were larger than sample standard errors (SSEs) for all combinations of simulation parameters. This resulted in slightly conservative confidence

intervals, with coverage probabilities a bit larger than nominal levels.

5.4.2 Real data examples

We present two examples of analysis on existing data sets. For the first, we use data from a clinical trial in the treatment of carcinoma of the oropharynx as presented in Kalbfleisch and Prentice (2002). These data include 195 patients with squamous cell carcinoma of 3 sites in the oropharynx from the 6 largest of 16 total participating institutions. Patients were randomly assigned treatment with radiation therapy alone or radiation therapy with a chemotherapeutic agent. While many other characteristics were collected, the data set we examined included only: sex; treatment; tumor grade, site, T staging, and N staging; overall patient condition, date of entry, living status, and survival time. Of the 195 patients included, 142 were observed to fail.

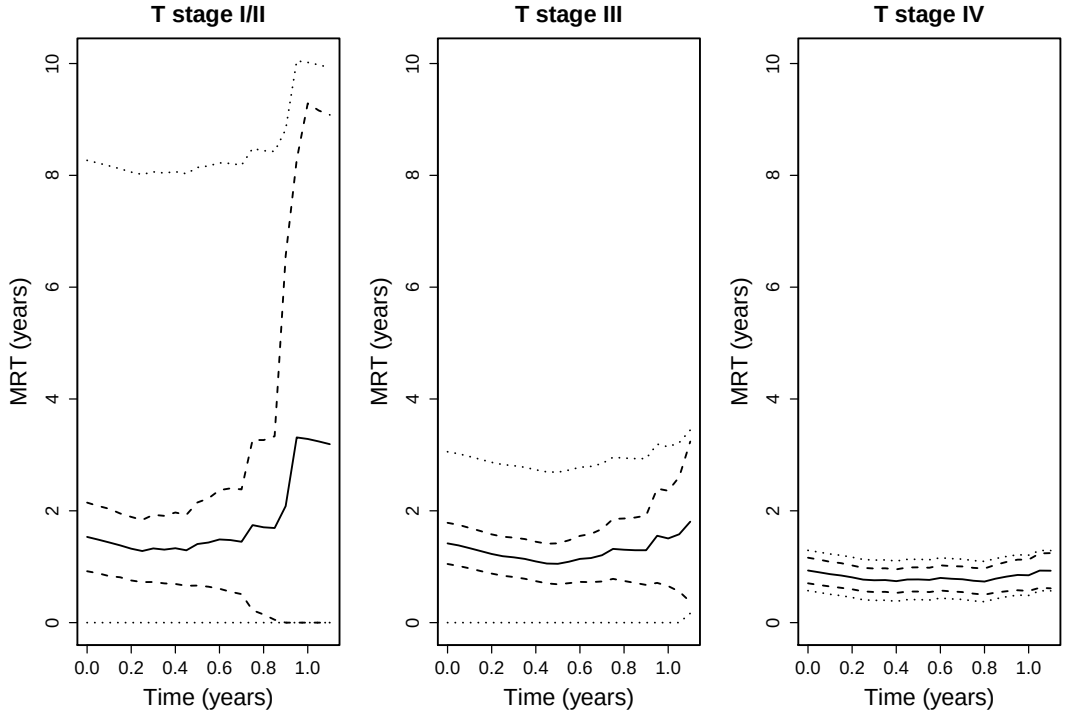
Table 5.3: Additive hazards model results

Data Set	Variable	Estimate	95% Confidence Interval	P value
Carcinoma	T stage I/II	0.00	—	
	T stage III	0.08	-0.09 to 0.25	0.011*
	T stage IV	0.35	0.12 to 0.59	
HIVNET 012	Nevirapine	0.00	—	—
	Zidovudine	-0.03	-0.11 to 0.06	0.511
	Maternal CD4 at pre-entry†	-0.00	-0.02 to 0.01	0.663
	Maternal HIV-1 RNA at pre-entry‡	0.10	0.04 to 0.17	0.001

*From Wald test of complete versus null model. †Per 100 cells/μL. ‡ Per unit increase of $\log_{10}$ HIV-1 RNA copies/mL.

We examined differences in MRTs to death across different T staging categories: I/II (35 patients), III (93 patients), or IV (67 patients). Regression results can be seen in Table 5.3. MRTs are plotted in Figure 5.2. It should be noted that the particularly wide confidence

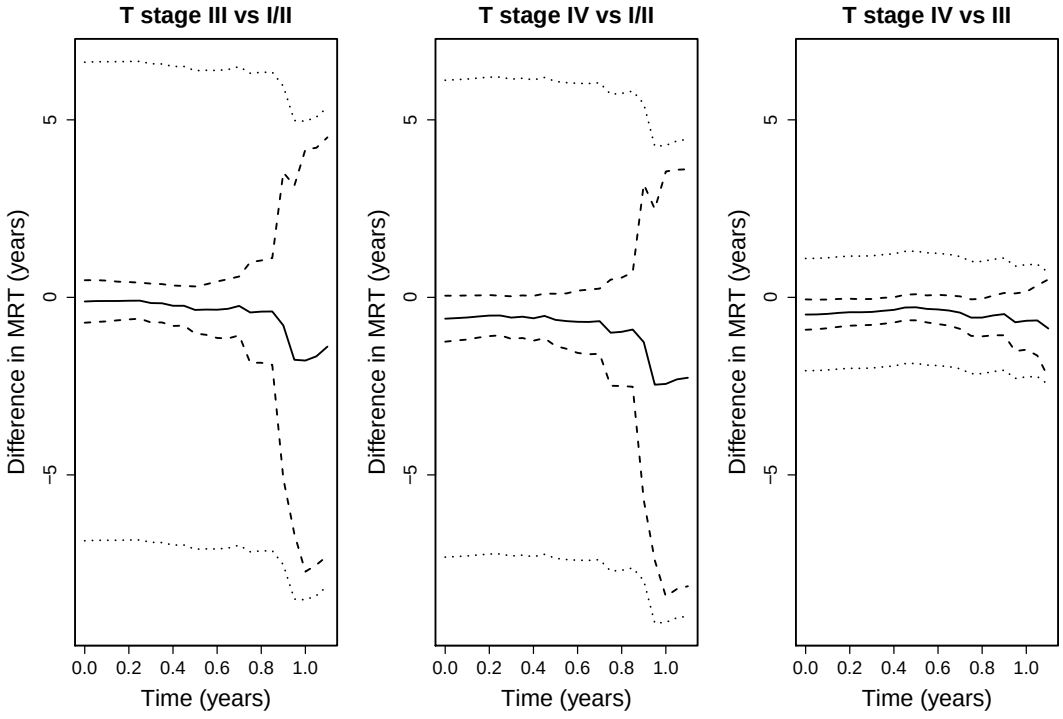
Figure 5.2: Plots of MRTs for carcinoma data with pointwise confidence intervals and confidence bands. The solid curves represent estimated MRT, the dashed curves represent the pointwise 95% confidence intervals, and the dotted curves represent the 95% confidence bands.



intervals and bands for patients with T stage I/II are likely due to the small number of patients in that category. The regression coefficients and plots indicate what we would expect: higher T stage is associated with increased hazard of death and lower MRT. This association is not statistically significant, however, as Figure 5.3 shows. The confidence bands (and usually the intervals) contain the null value 0 for the differences between any pair of T stages. Nevertheless, our results highlight a benefit of examining MRTs: for patients with T staging of I or II, MRT changes markedly over time, increasing sharply after initially decreasing slightly. For example, patients with T stage I or II have a MRT of 1.73 years (95% CI 0.30-3.15 years) after having survived 0.75 years and a MRT of 3.29 years (95% CI 0-9.29 years) after having survived 1 year. This would be important and

welcome news to surviving patients and their caregivers alike.

Figure 5.3: Plots of differences in MRTs for carcinoma data with pointwise confidence intervals and confidence bands. The solid curves represent estimated MRT, the dashed curves represent the pointwise 95% confidence intervals, and the dotted curves represent the 95% confidence bands.



Our second example uses data collected as part of the HIVNET 012 randomized trial (Jackson et al., 2003). This trial randomly assigned HIV-1 infected pregnant women in Kampala, Uganda to either a nevirapine- or zidovudine-based treatment. Their infants were followed and tested at pre-determined intervals for HIV-1. Data were also collected on adverse events through 6-8 weeks postpartum for mothers and 18 months for babies. The study enrolled 645 mothers: 313 assigned to nevirapine, 313 to zidovudine, and 19 to placebo. Within 18 months, 109 serious adverse events and 34 deaths were observed in the nevirapine group while 97 serious adverse events and 42 deaths were observed in the zidovudine group.

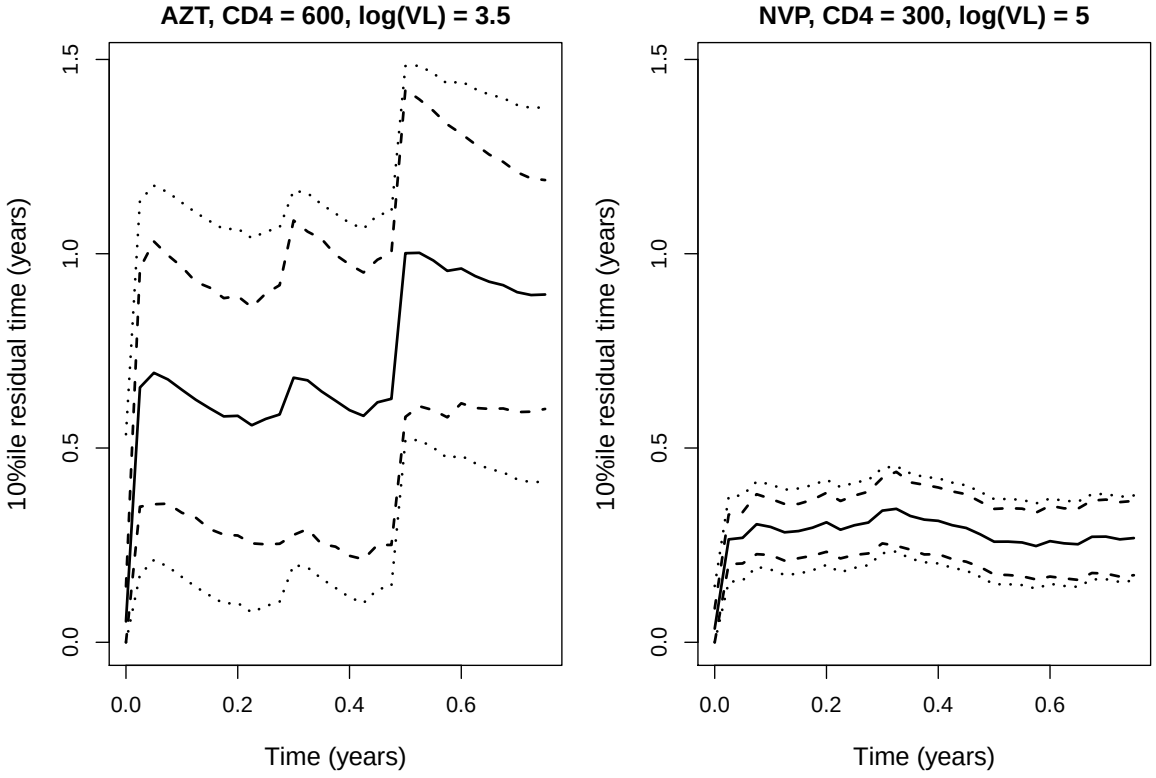
For our example we examined the relationship between the 10%ile of residual time to death or serious adverse event and treatment group (nevirapine versus zidovudine), maternal CD4 at pre-entry, and maternal HIV-1 RNA at pre-entry. We present results from the Cox model in Table 5.3. In the multivariate model, only maternal HIV-1 RNA was significantly associated with time to death or serious adverse event, with a coefficient of 0.10 (95% CI 0.04-0.17) for a unit increase of $\log_{10}$ HIV-1 RNA copies/mL. A plot of 10%ile residual times for two different combinations of covariate values (a theoretical infant whose mother was treated with AZT, had a maternal CD4 of 600 cells/μL, and a $\log_{10}$ maternal viral load of 3.5 copies/mL and a theoretical infant whose mother was treated with NVP, had a maternal CD4 of 300 cells/μL, and a $\log_{10}$ maternal viral load of 5 copies/mL) can be seen in Figure 5.4. These covariate combinations were chosen to represent infants with relatively protective or relatively non-protective characteristics. For both groups, the 10%ile residual times are fairly stable at early follow-up times (minus an immediate jump at the beginning of evaluation times). As expected, the 10%ile residual times for the infants with protective characteristics were consistently higher than those for the infants with non-protective characteristics, though they were also somewhat more variable.

When examining the difference in 10%ile residual times between the two different combinations of covariate values we conclude that it is not significant across all time as 0 is well within the limits of the confidence bands (see Figure 5.5). This is in spite of the fact that, for the majority of fixed time points, the difference is statistically significant. The lack of overall significance is driven to a large degree by the large increase in variance towards later times.

5.4.3 Comparison to Cox model results

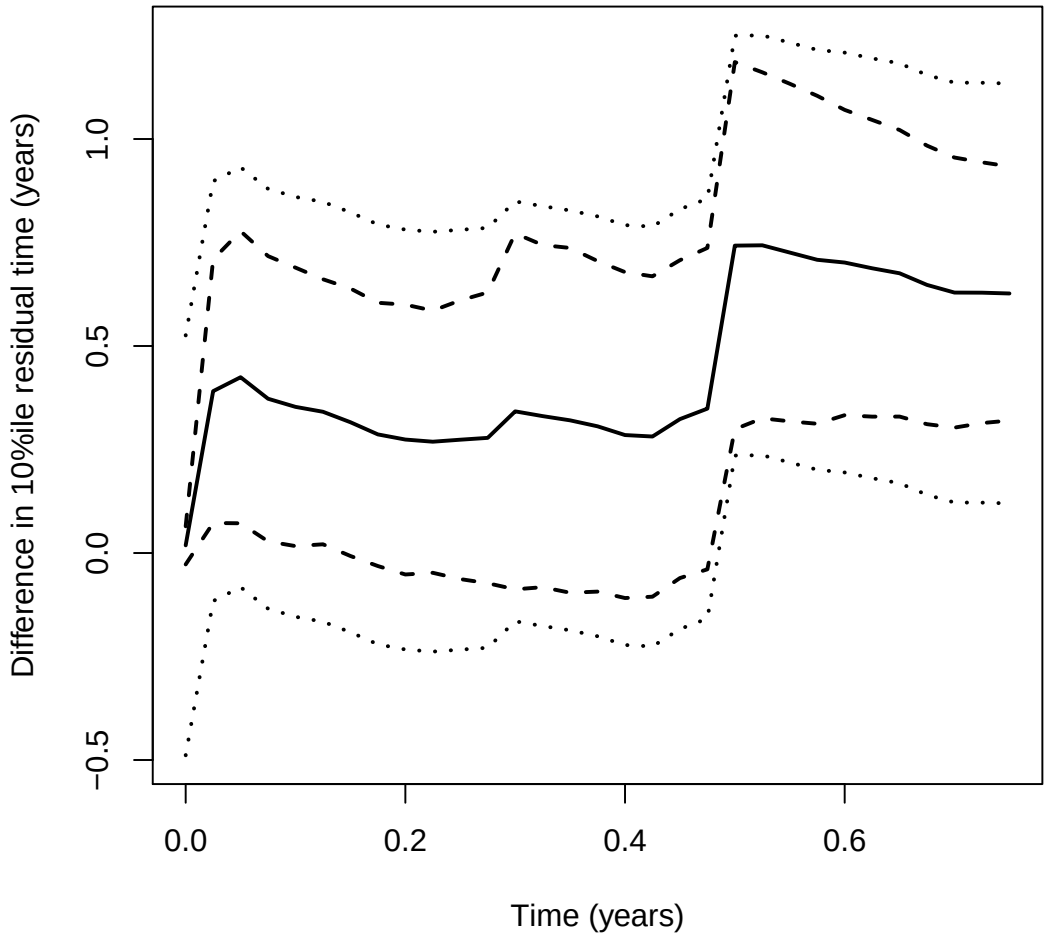
Because the underlying modeling assumptions differ substantially, the direct comparability of the two methods is limited. However, it may still be worth commenting on how their performance differs. As can be seen by comparing Tables 3.5 and 5.3, the results from the regressions match closely. The direction of the associations are the same and the signifi-

Figure 5.4: Plots of 10%ile residual times for HIVNET 012 data with pointwise confidence intervals. The solid curves represent estimated 10%ile residual times, the dashed curves represent the pointwise 95% confidence intervals, and the dotted curves represent the 95% confidence bands.



cance of each variable is similar. When considering the median residual times plotted for the carcinoma data, we notice that, while results are broadly similar, there does seem to be higher variance associated with estimating the MRT for those patients with T state I/II. On the other hand, there appears to be lower variance associated with estimating the MRT for those patients with T stage IV. The story is a bit different for the HIVNET 012 data. For the infants whose mothers have non-protective characteristics the results are quite similar. However for the infants whose mothers have protective characteristics the 10%ile residual times based on the additive model are more volatile, changing more dramatically over time,

Figure 5.5: Plots of the difference in 10%ile residual times for HIVNET 012 data. The difference is between a theoretical infant whose mother was treated with AZT, had a maternal CD4 of 600 cells/μL, and a $\log_{10}$ maternal viral load of 3.5 copies/mL and a theoretical infant whose mother was treated with NVP, had a maternal CD4 of 300 cells/μL, and a $\log_{10}$ maternal viral load of 5 copies/mL. The solid curves represent estimated 10%ile residual time, the dashed curves represent the pointwise 95% confidence intervals, and the dotted curves represent the 95% confidence bands.



particularly at later times. The estimate does seem to be sightly less variable at later times, though, resulting narrower confidence intervals and bands.

5.5 Discussion

One issue we encountered while working with the additive hazard model is that the chosen time scale is directly related to the magnitude of the point and variance estimates for the regression coefficients. For example, when using years as the basic time unit, the point estimate will be 365.25 times as large as it is when using days as the basic time unit. Because the variance changes similarly (in this example, by a factor of 365.25^2), the significance of any results does not differ. Additionally, changing the time scale had no effect on our point estimation of residual time quantiles. Unfortunately, this was not true for estimating the attendant variances. For example, when analyzing the carcinoma data, we found that using days as the time unit meant variance estimates were unable to be calculated or were nonsensical (negative) for earlier times. This issue did not arise when using either months or years as the time unit, so it appears there is some sort of threshold effect. Because the derivation of the variance estimator involves several approximations, it is our assumption that when the point and/or variance estimates in the additive hazard regression become sufficiently small (by using smaller units of time) the approximations become invalid. Since changing the time unit has no effect on the additive hazards regression itself, we therefore recommend that, if using our method to estimate residual time quantiles, one chooses as large a time unit as is reasonable.

Many of the limitations associated with our estimator are similar to those initially faced by our estimator based on the Cox model, though some are unique. As an example of the latter, we are limited to some degree by the inherent awkwardness of obtaining our estimate as a numerical solution rather than from a closed-form formula. While this does not limit the actual estimates we obtain, it does make application more cumbersome and thus less easily adopted.

As with the Cox model-based estimator, our additive hazards model-based estimator

would be greatly improved by being able to handle time-varying covariates, and this provides a good opportunity for future work for all the reasons outlined in previous chapters. Additionally, while different, the modeling assumptions for the additive hazards model are no more lenient or all-encompassing than those for the Cox proportional hazards model. Many data will not meet the assumptions of either model. As explained in previously, while our method provides a different summary measure, it in no way escapes the limitations of the additive hazards model, and is dependent on all the associated assumptions. While the additive hazards model, like the Cox model, is semiparametric and therefore allows a great deal of flexibility, possible model mis-specification can still lead to biased estimate of residual time quantiles. Therefore, certain sensitivity analyses may be used to help detect such a bias and further provide guidance on alternative models for more accurate and reliable model-based estimation. Unfortunately, said sensitivity analyses are much less developed for the additive hazards models than they are for the proportional hazards model. Nevertheless, if other models are indicated as being more appropriate, developing estimator based on said models should not be particularly difficult with the derivations under both the Cox and additive hazards models included in this dissertation to use as a guide.

Other concerns shared with our Cox model-based estimator include the calculation of residual time quantiles for unobserved covariate values, the issue of multiple comparisons, and being limited by the number (or proportion) of events that are actually observed, as our estimator is only defined when

$$\widehat{\Lambda}_0\{t + \widehat{\theta}(t, q|\boldsymbol{Z})\} + \widehat{\theta}(t, q|\boldsymbol{Z})\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{Z} + \log q \leq \widehat{\Lambda}_0(\tau)$$

where τ is the largest observed failure time.

Chapter 6
DISCUSSION

6.1 Summary

In this dissertation we have developed methods for estimating quantiles of residual time under the Cox proportional hazards model (with fixed and with external time-varying covariates) and under the additive hazards model. When restricting analysis to fixed covariates we were able to determine the asymptotic distributions of our estimators, allowing us to construct confidence intervals and carry out hypothesis testing. Additionally, we determined the asymptotic distributions for the difference between two estimates of residual time quantiles with varying evaluation times, covariate values, and quantiles. We demonstrated the performance of our methods both via simulation and application to example data sets.

We believe our methods to be a useful addition to the field of survival analysis, providing a different way of summarizing covariate effects on the time to an event. Our methods also allow for more exploration of trends across time than do summary measures such as hazard ratios. That being said, our methods have their limitations. Many have been covered in preceding chapters. Others are discussed below in Section 6.2 as topics for future work.

6.2 Future Work

There are substantial opportunities for future work building on the developments contained in this dissertation. Many of them are natural, if not necessarily straightforward, extensions or generalizations.

First of all, more thorough simulations could be run in order to address, or at least explore, some of the limitations or concerns with our methods. While we have shown our methods to work very well when applied to data that satisfy their assumptions (i.e. propor-

tional or additive hazards), it would be informative to demonstrate their performance when applied to data that do not satisfy their assumptions. For example, using our estimate based on the proportional hazards model with data generated using the additive hazards model or vice versa. We anticipate conducting and presenting the results of these simulations in a future manuscript. Additionally, it would be valuable to conduct simulations to explore the effect of missing values, or, more accurately, different methods of imputation to address missing values, on the performance of our estimators when including time-varying covariates in the model.

So far, we have developed covariate-specific estimators for residual time quantiles as based on the Cox proportional hazards and the additive hazards models. It would be useful to further develop estimators based on other models commonly used for analyzing time-to-event data. The accelerated failure time model, characterized by the cumulative hazard function

$$\Lambda(t|\boldsymbol{Z}) = \Lambda_0\{t \exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z})\},$$

sees fairly wide usage. We pursued this model as an initial alternative to the Cox proportional hazards model. While we were able to obtain a closed-form solution for residual time quantiles,

$$\theta(t, q|\boldsymbol{Z}) = \exp(-\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z})\Lambda_0^{-1}[\Lambda_0\{t \exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z})\} - \log q] - t,$$

we had difficulties implementing the estimator. In particular, we could not find good tools for estimating the baseline cumulative hazard function. While developing this from scratch would certainly be feasible, we opted to focus on the additive hazards model as a slightly more tractable model for full development. However, it still presents a good opportunity for future work, especially if more refined tools for the initial regression estimation become available. Unfortunately, the asymptotic results may be difficult to develop, as the terms for the regression coefficients and the baseline cumulative hazard are more mathematically entwined than they are in either the proportional hazards or additive hazards models. Another model, similar in name and construction, is the accelerated hazards model proposed

by Chen and Wang (2000). This model is characterized by the hazard function

$$\lambda(t|\boldsymbol{Z}) = \lambda_0\{t \exp(\boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z})\}.$$

We have not yet attempted to develop our residual time quantile estimator for this model, though it would likely share much in common with that for the accelerated failure time model. Both the accelerated failure time and accelerated hazards models are formulated above with $\Lambda_0(\cdot)$ as an unspecified function that could be estimated non-parametrically. One could instead suppose a parametric cumulative hazard function, and this may make estimation significantly more straight-forward. For the accelerated failure time model, the log-logistic and Weibull distributions are both commonly used. However, semi-parametric models are generally preferable for their flexibility and reduced necessity for assumptions about the underlying hazard function.

Another topic for future work would be further developing our estimator in the setting with time-varying covariates. As discussed in Chapter 4, our current estimator is limited by requiring a specific, assumed trajectory of future covariate values. While in some settings this is acceptable, in others it would be preferable to have a "future-agnostic" or marginal estimator based on expected or population averaged, rather than assumed, future covariate values. It would also make application easier if closed form versions of the point and variance estimators were developed, as numerical and bootstrap methods may be too complex for widespread use. The major hurdle seems to be the inability to mathematically separate the cumulative hazard into the baseline cumulative hazard for arbitrary covariates $\boldsymbol{Z}(t)$. This ability was fundamental to our other developments, particularly for asymptotic variance calculations, as it allowed us to apply established results about each component with minor modifications to suit our setting. It may be necessary, or at least easier, to attempt development using entirely different techniques than our own. While we relied on martingale theory, empirical process theory may in fact be more useful or more easily applied when considering time-varying covariates.

It would be valuable to address other limitations to our method. One of the most impor-

tant is the inability to calculate many quantiles of residual times when events are relatively rare. For example, events in the HIVNET 012 data were rare enough to limit our ability to estimate median residual times, much less higher percentiles such as the 75 or 90 percentiles. Our inability to estimate these quantiles is due to our estimators' basis on the cumulative hazard function which ceases to be estimable (for good reason) at times exceeding that of the last observed event. A fairly simple fix would be to simply extrapolate the estimated cumulative hazard past the final event time. However, this would obviously require strong and difficult to verify assumptions. Additionally, for any given method for extrapolation, substantial work would be necessary to determine effects on estimation, estimator performance, and asymptotic results. A more sophisticated solution would be desirable, though it would necessitate the same considerations. In lieu of such, there are many extrapolation methods to consider. These include nonparametric methods (Scholz, 1995) as well as parametric (Gelber et al., 1993) or regression methods (Jeong and Fine, 2007). While none of the aforementioned methods has been explicitly applied to cumulative hazard estimation, adapting them for such may not be too difficult, and would allow for residual time quantile estimation even when events are uncommon.

A minor but useful extension of our work would be to further characterize the relationship between our estimators for quantiles of residual time and the regression coefficients obtained from the statistical models upon which they are based (i.e. the Cox proportional hazards or the additive hazards regression models). In particular, relating the difference between two estimates of residual time to the relevant covariates coefficients may increase intuition in both directions. It may be that the ratio, rather than the difference, between two estimates of residual time is more directly relatable to regression coefficients, particularly in the proportional hazards model. This is suggested by what we saw in Figure 3.1: the ratio of the median residual times trending towards $\exp(-\boldsymbol{\beta})$. Indeed, whether or not this is the case, developing asymptotic properties for the ratio of residual time quantiles would be a useful endeavor, as this may be a more appealing summary measure. In our work, we only considered differences, which provided more direct interpretation as well as more

manageable determination of asymptotic properties. Considering the ratio of quantiles of residual time would require theoretically intensive derivation of the transformed process as opposed to simply the covariance between the two processes, unless one were able to appeal to a simplified approximation. However, the approximation we mentioned above, $\exp(-\boldsymbol{\beta})$, was only valid at later times and would not be generally applicable.

A compelling extension for applied settings would be to develop a sort of population-wide difference in residual times. That is, instead of considering covariate effects on the residual time of a single individual, consider them on the residual time for an entire population. This would be very useful in policy-making, as one could see a very direct estimation of e.g. person-years that could be gained by instituting a certain treatment as the standard of practice, or the difference in total person-years spent on an expensive treatment if it were applied earlier, rather than later, in a disease process.

We have mentioned the ability to explore temporal trends in residual time as an advantage of our methods. However, applying these types of results to a general population requires some additional assumptions. In particular, it is possible that temporal trends may be driven by the underlying frailty of certain individuals. That is, residual time may start off shorter and then increase as more frail individuals fail, leaving only the less frail individuals in the risk set. We must therefore assume that the general population has the same proportion of frail individuals as does the data we are analyzing. This may not be the case, particularly if residual time quantiles are being used as a prognostic tool. In a study, we will often include both incident and prevalent cases while in application incident cases will be substantially more common. If, as is plausible, prevalent cases are more frail, trends we see in the study data may be driven by them and thus not broadly applicable. While this is a somewhat broad limitation, it could be addressed by developing methods that allow for frailty in addition to conducting simulations to explore the affects of unmodeled frailty on the validity of our estimators.

Finally, we would like to explore and characterize the effects of including covariates that may be in the causal pathway. Typically we would want to avoid including any co-

variates in the model that are in the causal pathway from our covariate of interest and the outcome. However, given the nature of residual time, this may not be the case. Because we are interested in prognosis as time progresses and are therefore conditioning on, at very least, survival until the evaluation time, including a covariate in the causal pathway may actually provide additional information on an individual's status and therefore prognosis. It is difficult to lay out exactly how to explore this idea further. Certainly, we can run simulations with and without the inclusion of covariates in the causal pathway and compare estimator performance. This alone is likely insufficient, though, and must be balanced with more careful consideration of the inferential implications in addition to the numerical consequences.

All the aforementioned ideas for future work would be interesting to pursue and would make our methods more widely applicable, more robust, or both. It is our hope to continue work on this topic in order to increase the appeal of our methods for adoption.

BIBLIOGRAPHY

K. Alam and K. B. Kulasekera. Estimation of the quantile function of residual life time distribution. *Journal of statistical planning and inference*, 37(3):327–337, 1993.

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

N. E. Breslow and N. E. Day. *Statistical methods in cancer research. Volume II - The design and analysis of cohort studies*. International Agency for Research on Cancer, 1987.

Y. Q. Chen and M. C. Wang. Analysis of accelerated hazards models. *Journal of the American Statistical Association*, 95(450):608–618, 2000.

Y. Q. Chen, C. A. Rohde, and M. C. Wang. Additive hazards models with latent treatment effectiveness lag time. *Biometrika*, 89(4):917–931, 2002.

Y. Q. Chen, N. P. Jewell, X. Lei, and S. C. Cheng. Semiparametric estimation of proportional mean residual life model in presence of censoring. *Biometrics*, 61(1):170–178, 2005.

K. F. Cheng. On almost sure representation for quantiles of the product limit estimator with applications. *Sankhyā: The Indian Journal of Statistics, Series A*, 46(3):426–443, 1984.

D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 187–220, 1972.

M. Csörgő and S. Csörgő. Estimation of percentile residual life. *Operations research*, 35 (4):598–606, 1987.

M. Csörgő and R. Zitikis. Mean residual life processes. *The Annals of Statistics*, 24(4): 1717–1739, 1996.

D. M. Dabrowska and K. A. Doksum. Estimates and confidence intervals for median and mean life in the proportional hazard model. *Biometrika*, 74(4):799–807, 1987.

B. Efron and R. J. Tibshirani. *An introduction to the bootstrap*, volume 57. CRC Press LLC, 1998.

T.R. Fleming and D. P. Harrington. *Counting Processes and Survival Analysis*. John Wiley & Sons, Inc., Hoboken, New Jersey, 2005.

R. D. Gelber, A. Goldhirsch, and B. F. Cole. Parametric extrapolation of survival estimates with applications to quality of life evaluation of treatments. *Controlled Clinical Trials*, 14(6):485–499, 1993.

D. W. Hosmer, S. Lemeshow, and S. May. *Applied Survival Analysis: Regression Mdeling of Time-to-Event Data*. John Wiley & Sons, Inc., Hoboken, New Jersey, 2008.

J. B. Jackson, P. Musoke, T. Fleming, L. A. Guay, D. Bagenda, M. Allen, C. Nakabiito, J. Sherman, P. Bakaki, M. Owor, C. Ducar, M. Deseyve, A. Mwatha, L. Emel, C. Duefield, M. Mirochnick, M. G. Fowler, L. Mofenson, P. Miotti, M. Gigliotti, D. Bray, and F. Mmiro. Intrapartum and neonatal single-dose nevirapine compared with zidovudine for prevention of mother-to-child transmission of HIV-1 in Kampala, Uganda: 18-month follow-up of the HIVNET 012 randomised trial. *The Lancet*, 362:859–868, 2003.

J. H. Jeong and J. P. Fine. Parametric regression on cumulative incidence function. *Biostatistics*, 8(2):184–196, 2007.

J. H. Jeong, S. H. Jung, and J. P. Costantino. Nonparametric inference on median residual life function. *Biometrics*, 64(1):157–163, 2008.

S. H. Jung, J. H. Jeong, and H. Bandos. Regression on quantile residual life. *Biometrics*, 65(4):1203–1212, 2009.

J. D. Kalbfleisch and R. L. Prentice. *The Statistical Analysis of Failure Time Data*. John Wiley & Sons, Inc., Hoboken, New Jersey, 2002.

E. L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481, 1958.

J. Lee and S. Hyun. Confidence bands for the difference of two survival functions under the additive risk model. *Journal of Applied Statistics*, 38(4):785–797, 2011.

D. Y. Lin and Z. Ying. Semiparametric analysis of the additive risk model. *Biometrika*, 81 (1):61–71, 1994.

Y. Ma and Y. Wei. Analysis on censored quantile residual life model via spline smoothing. *Statisticca Sinica*, 22(1):47–68, 2012.

Y. Ma and G. Yin. Semiparametric median residual life model and inference. *Canadian Journal of Statistics*, 38(4):665–679, 2010.

G. Maguluri and C. H. Zhang. Estimation in the mean residual life regression model. *Journal of the Royal Statistical Society. Series B*, 56(3):477–489, 1994.

A. K. Nanda, S. Bhattacharjee, and S. S. Alam. Properties of proportional mean residual life model. *Statistics & probability letters*, 76(9):880–890, 2006.

D. Oakes and T. Dasu. A note on residual life. *Biometrika*, 77(2):409–410, 1990.

W. J. Padgett. A kernel-type estimator of a quantile function from right-censored data. *Journal of the American Statistical Association*, 81(393):215–222, 1986.

M. I. Parzen, L. J. Wei, and Z. Ying. A resampling method based on pivotal estimating functions. *Biometrika*, 81(2):341–350, 1994.

L. Peng and Y. Huang. Survival analysis with quantile regression models. *Journal of the American Statistical Association*, 103(482):637–649, 2008.

S. Portnoy. Censored regression quantiles. *Journal of the American Statistical Association*, 98(464):1001–1012, 2003.

G. Qin and Y. Zhao. Empirical likelihood inference for the mean residual life under random censorship. *Statistics & probability letters*, 77(5):549–557, 2007.

H. Ramlau-Hansen. Smoothing counting process intensities by means of kernel functions. *The Annals of Statistics*, 11(2):453–466, 1983.

D. C. Schmittlein and D. G. Morrison. The median residual lifetime: A characterization theorem and an application. *Operations Research*, 29(2):392–399, 1981.

F. W. Scholz. Nonparametric tail extrapolation. Technical report, 1995.

J. K. Song and G. Y. Cho. A note on percentile residual life. *Sankhyā: The Indian Journal of Statistics, Series A*, 57(2):333–335, 1995.

J. Q. Su and L. J. Wei. Nonparametric estimation for the difference or ratio of median failure times. *Biometrics*, 49(2):603–607, 1993.

L. Sun and Q. Zhau. A class of mean residual life regression models with censored survival data. *Journal of statistical planning and inference*, 140(2):3425–3441, 2010.

L. Sun, X. Song, and Z. Zhang. Mean residual life models with time-dependent coefficients under right censoring. *Biometrika*, 99(1):185–197, 2012.

T. M. Therneau and P. M. Grambsch. *Modeling Survival Data: Extending the Cox Model*. Springer, 2000.

J. L. Wang and T. P. Hettmansperger. Two-sample inference for median survival times based on one-sample procedures for censored survival data. *Journal of the American Statistical Association*, 85(410):529–536, 1990.

B. S. Yandell. Nonparametric inference for rates with censored survival data. *The Annals of Statistics*, 11(4):1119–1135, 1983.

G. L. Yang. Estimation of a biometric function. *The Annals of Statistics*, 6(1):112–116, 1978.

S. Yang. Censored median regression using weighted empirical survival and hazard functions. *Journal of the American Statistical Association*, 94(445):137–145, 1999.

Z. Ying, S. H. Jung, and L. J. Wei. Survival analysis with median regression models. *Journal of the American Statistical Association*, 90(429):178–184, 1995.

K. C. Yuen, L. X. Zhu, and N. Y. Tang. On the mean residual life regression model. *Journal of statistical planning and inference*, 113(2):685–698, 2003.

M. Zhou and J. H. Jeong. Empirical likelihood ratio test for median and mean residual lifetime. *Statistics in medicine*, 30(2):152–159, 2011.

Appendix A

CODE

In this Appendix we provide some example code to demonstrate the implementation of our estimators.

A.1 Proportional Hazards

Point estimation is fairly straightforward. Let `t` be the evaluation time t, `z` be the vector of covariates $\boldsymbol{Z}$, `cum.haz` be the values of the cumulative hazard at the observed event times `obs.times`, `est.beta` be the vector of estimated regression coefficients $\widehat{\boldsymbol{\beta}}$, and `q` be the quantile of interest q. $\widehat{\theta}(t, q|\boldsymbol{Z})$ can then be estimated with the following function.

```
estimated.rt <- function(t,z,cum.haz,obs.times,est.beta,q)
{
  approx(cum.haz, obs.times, approx(obs.times, cum.haz, t, rule=2)$y
    - log(q) * exp(-est.beta %*% z),rule=2)$y - t
}
```

Variance estimation requires substantially more work. Let `times` be the ordered list of event or censoring times and `z.mat` be the $n \times p$ matrix of covariate values. We start with functions for the two integrals in the variance formula from Theorem 3.2.1.

```
first.term.integrand <- function(x,est.beta,times)
{
  if ( sum(times >= x) > 0 )
  {
    (z.mat %*% t( (times >= x) * exp(t(est.beta) %*% z.mat) ) /
      ( (exp(t(est.beta) %*% z.mat) %*% (times >= x) )^2)
  }
```

```
  else
  {
    0
  }
}

second.term.integrand <- function(x,est.beta,times)
{
  if ( sum(times >= x) > 0 )
  {
    (exp(t(est.beta) %*% z.mat) %*% (times >= x) )^(-2)
  }
  else
  {
    0
  }
}
```

Next, we need functions for kernel estimation of baseline hazard. We used an Epanechnikov kernel, specifying bandwidth by `bw`. The vector of failure indicators was coded as `fail.ind`.

```
k <- function(x)
{
  0.75 * (1-x^2) * (abs(x)<=1)
}

k.haz <- function(t,est.beta,times,fail.ind,bw,n)
{
  summands <- rep(NA,n)
  for (i in 1:n)
  {
    if ( sum(times >= times[i]) > 0 )
    {
      summands[i] <- k( (t-times[i])/bw) * fail.ind[i] /
```

```
        ( sum( (times >= times[i]) * exp(t(est.beta) %*% z.mat) ) )
    }
    else
    {
      summands[i] <- 0
    }
  }
  sum(summands)/bw
}
```

The output of these functions can then be combined to obtain a variance estimate as shown below.

```
est.rt <- estimated.rt(t,z,cum.haz,obs.times,est.beta,q)

est.haz.rt <- k.haz(est.rt+t,est.beta,times,fail.ind,bw,n)

first.integral <- apply(sapply(times[(times>=t) & (times<=(t+est.rt) ) &
    fail.ind],function(x) first.term.integrand(x,est.beta,times),1,sum)

first.term <- (t(first.integral) %*% var.beta %*% first.integral) -
    (2 * -log(q) * t(z * exp(-z %*% est.beta) ) %*% var.beta %*%
    first.integral) + (log(q)^2 * t(z * exp(-z %*% est.beta)) %*%
    var.beta %*% (z * exp(-z %*% est.beta) ) )

second.integral <- sum(sapply(times[ (times>=t) & (times<=(t+est.rt) ) &
    fail.ind],function(x) second.term.integrand(x,est.beta,times) ) )

est.var <- (first.term + second.integral) / (est.haz.rt^2)
```

A.1.1 Time-varying covariates

When including time-varying covariates, point estimation is somewhat more complicated, and is carried out using numerical solution. We present a few requisite functions followed by the function for residual time quantile estimation. Let `z.init` be the vector

of initial covariate values $\boldsymbol{Z}(0)$, `t.change` be the times at which $\boldsymbol{Z}(t)$ changes value t^*, `fixed.z.init` be the initial covariate value for the individual being evaluated, and `fixed.t.change` be the time at which the covariate value changes for the individual being evaluated.

```
z <- function(t,t.change,z.init)
{
  if (z.init==0)
  {
    as.numeric(t>t.change)
  }
  else
  {
    1-as.numeric(t>t.change)
  }
}

cum.haz.integrand <- function(x,est.beta,times,z.init,t.change,
                              fixed.z.init,fixed.t.change)
{
  if ( sum(times >= x) > 0 )
  {
    z.current <- ifelse(x>t.change,1-z.init,z.init)
    fixed.z.current <- z(x,fixed.t.change,fixed.z.init)
    exp(est.beta * fixed.z.current) / (exp(t(est.beta) %*% z.current)
      %*% (times >= x) )
  }
  else
  {
    0
  }
}
```

```
cum.haz.estimate <- function(est.rt,t,est.beta,times,fail.ind,z.init,
                                t.change,fixed.z.init,fixed.t.change)
{
  sum(sapply(times[ (times>=t) & (times<=(t+est.rt) ) & fail.ind],
    function(x) cum.haz.integrand(x,est.beta,times,z.init,t.change,
    fixed.z.init,fixed.t.change) ) )
}

estimated.rt <- function(t,est.beta,times,fail.ind,z.init,t.change,
                                fixed.z.init,fixed.t.change)
{
  uniroot( function(y) cum.haz.estimate(y,t,est.beta,times,fail.ind,
    z.init,t.change,fixed.z.init,fixed.t.change) + log(q),
    c(0.1,as.numeric(quantile(times,0.99) ) ) )$root
}
```

Bootstrap estimation for variance is fairly standard, so is not detailed here.

A.2 Additive Hazards

Point estimation is somewhat more complicated in the additive hazards model. Using similar notation as above, with `rt` being an estimated residual time quantile, we have the following.

```
est.rt.function <- function(rt,t,q,est.beta,z,obs.times,cum.haz)
{
  approx(obs.times,cum.haz.points,rt + t,rule=2)$y + (est.beta %*% z)
     * rt - approx(obs.times,cum.haz,t,rule=2)$y + log(q)
}

est.rt <- uniroot(function(y) est.rt.function(y,t,q,est.beta,z,
                obs.times,cum.haz),c(times[n*0.03],times[n*0.97]))$root
```

Variance estimation again requires a few steps. Below, we suppose a model with two covariates, `z.rand.1` and `z.rand.2`.

```
c.fun <- function(n,times,z.rand.1,z.rand.2)
{
  sum.y <- n:1
  delta.t <- diff(c(0,times))
  cbind(cumsum( (rcumsum(z.rand.1) * delta.t) / sum.y),
          cumsum( (rcumsum(z.rand.2)*delta.t) / sum.y) )
}

int.fun <- function(n,fail.ind)
{
  sum.y <- n:1
  cumsum(fail.ind / (sum.y^2) )
}

d.fun <- function(n,times,z.rand.1,z.rand.2,fail.ind)
{
  sum.y <- n:1
  z.bar.1 <- rcumsum(z.rand.1) / sum.y
  z.bar.2 <- rcmsum(z.rand.2) / sum.y
  cbind(cumsum( ( (z.rand.1 - z.bar.1) * fail.ind) / sum.y),
          cumsum( ( (z.rand.2 - z.bar.2) * fail.ind) / sum.y) )
}
```

```
a.fun <- function(n,times,z.rand.1,z.rand.2)
{
  delta.t <- diff(c(0,times) )
  integrands <- array(rep(NA,4 * n),c(2,2,n) )
  for (i in 1:n)
  {
    time.i <- times[i]
    z.bar.t.1 <- sum(z.rand.1[times>=time.i] ) / sum(times>=time.i)
    z.bar.t.2 <- sum(z.rand.2[times>=time.i] ) / sum(times>=time.i)
    z.diff <- rbind( (z.rand.1-z.bar.t.1)[times>=time.i],
                     (z.rand.2-z.bar.t.2)[times>=time.i] )
    integrands[,,i] <- z.diff.sq <- z.diff %*% t(z.diff)
  }
  apply(integrands,c(1,2),sum)
}

k <- function(x)
{
  0.75 * (1-x^2) * (abs(x)<=1)
}

k.haz <- function(t,est.beta,times,z.rand,fail.ind,bw,n)
{
  summands <- diff(c(0,est.cum.haz(est.beta,times,z.rand,fail.ind) ) )
    * k( (t-times)/bw)
  sum(summands) / bw
}
```

Using these functions, we can then estimate the variance as follows.

```
est.haz.rt <- k.haz(est.rt+t,est.beta,times,z.rand,fail.ind,bw,n)

int <- int.fun(n,fail.ind)
int <- max(int[times<=est.rt + t] ) - min(int[times>=t] )
```

```
max.ind <- max( (1:n)[times<=est.rt + t] )
min.ind <- min( (1:n)[times>=t] )

c  <- c.fun(n,times,z.rand.1,z.rand.2)
c  <- c[max.ind,] - c[min.ind,]

d  <- d.fun(n,times,z.rand.1,z.rand.2,fail.ind)
d  <- d[max.ind,] - d[min.ind,]

a <- a.fun(n,times,z.rand.1,z.rand.2)

new.c <- -z * est.rt + c

var.rt <- ( int + new.c %*% var.beta %*% new.c - 2 * new.c %*% solve(a)
              %*% d ) / ( (est.haz.rt + est.beta %*% z)^2)
```