University of Windsor
Scholarship at UWindsor

Electronic Theses and Dissertations Theses, Dissertations, and Major Papers

2012

Shrinkage Estimation for Aalen's Additive Model

Katrina Tomanelli
University of Windsor

Follow this and additional works at: https://scholar.uwindsor.ca/etd

Recommended Citation

Tomanelli, Katrina, "Shrinkage Estimation for Aalen's Additive Model" (2012). *Electronic Theses and Dissertations*. 353.
https://scholar.uwindsor.ca/etd/353

This online database contains the full-text of PhD dissertations and Masters' theses of University of Windsor students from 1954 forward. These documents are made available for personal study and research purposes only, in accordance with the Canadian Copyright Act and the Creative Commons license—CC BY-NC-ND (Attribution, Non-Commercial, No Derivative Works). Under this license, works must always be attributed to the copyright holder (original author), cannot be used for any commercial purposes, and may not be altered. Any other use would require the permission of the copyright holder. Students may inquire about withdrawing their dissertation and/or thesis from this database. For additional inquiries, please contact the repository administrator via email (scholarship@uwindsor.ca) or by telephone at 519-253-3000ext. 3208.

Shrinkage estimation for Aalen's additive model

by

Katrina Tomanelli

A Thesis

Submitted to the Faculty of Graduate Studies

through the Department of Mathematics and Statistics

in Partial Fulfillment of the Requirements for

the Degree of Master of Science at the

University of Windsor

Windsor, Ontario, Canada

2012

© 2012 Katrina Tomanelli

Shrinkage estimation for Aalen's additive model

by

Katrina Tomanelli

APPROVED BY:

Sévérien Nkurunziza, Advisor
Department of Mathematics and Statistics

A. Hussein, Advisor
Department of Mathematics and Statistics

Tim Traynor, Internal Reader
Department of Mathematics and Statistics

May 14, 2012

Author's Declaration of Originality

I hereby certify that I am the sole author of this thesis and that no part of this thesis has been published or submitted for publication.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such materials in my major paper and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my committee and the Graduate Studies office, and that this major paper has not been submitted for a higher degree to any other University or Institution.

Abstract

Survival analysis is a branch of statistics which deals with the analysis of time to event (or in general event history). In particular, regression models that relate event occurrence rates to predictor variables are quite common in the medical field. One such regression model is the Aalen's nonparametric additive model in which the regression coefficients are assumed to be unspecified functions of time. In this project we consider estimation of Aalen's nonparametric regression coefficients when some uncertain prior information is available about these coefficients. More precisely, we combine unrestricted estimators and estimators that are restricted by a linear hypothesis (prior information) and produce James-Stein-type of shrinkage estimators. We develop the asymptotic joint distribution of such restricted and unrestricted estimators and use it for studying the relative performance of the proposed estimators via their asymptotic distributional biases and risks. We conduct Monte Carlo simulations to examine relative performance of the estimators in small samples and we illustrate the methodology by using a real data on the survival of primary billiary cirrhosis patients.

This thesis is dedicated to Joshua, Alysia, Maya and Isabella.

Look, Auntie wrote you a book! Don't worry, I won't make you read it.

Acknowledgments

I would like to express my sincerest gratitude to my advisor, Dr. Nkurunziza, for having answers to all of my questions and for showing me what I can accomplish when I'm willing to work hard for it. I am also grateful to my co-advisor, Dr. Hussein, for introducing me to the wonderful world of simulations and data analysis; it turns out that all this theory is actually applicable in the real world!

I would also like to thank my family and friends, especially my mother for taking care of my stress-induced upset stomach, my father for reminding me that Tomanelli's can handle anything, and the rest of my family for pretending to be mildly interested in my research.

Without these people in my life, I would not have been able to accomplish this task. At the very least, I would have lost my mind in the attempt.

Contents

List of Figures

Chapter 1

Introduction

1.1 Preliminaries

Survival analysis is a branch of statistics which deals with the analysis of time-to-event (or in general event history). Applications of event history analysis methodologies are numerous in the medical field, but are also found in economics, engineering and sociology. Usually, the investigator would collect data on the occurrence times of a certain event of interest (outcome variable) along with a set of independent predictor variables (covariates) such gender, age, social status, biomarkers of diseases and similar variables. The investigator would then desire to know if and how such covariates influence the occurrence rates (intensity) of the event of interest. The Cox's proportional hazards (PH) model is one of the earliest and perhaps the most used statistical model which attempts to address such questions. Cox's original model was later formulated in terms of counting process theory and named multiplicative intensity model (Andersen and Gill (1999)). The multiplicative intensity model (MI) has extended the PH model in the sense that it allowed multiple events and time varying covariate processes. The PH model and its variants assume that the intensity function of the counting process defined by the events of interest is made up of the product of a baseline nonparametric intensity function of time, and a parametric part consisting of a function of a linear combination of the independent variables. The effect of the covariates is measured through the unknown coefficients of the linear combination (regression parameters) which do not depend on time. This entails that the hazard ratio of two individuals differing just by the level of a given covariate would be constant over time. This property, which is known as the proportional hazards assumption, gives an attractive interpretation in terms of risk ratio and is mathematically tractable. However, such an assumption is sometimes not satisfied by the data at hand and hence, in these situations, time-varying regression coefficients are required in order to quantify the effect of the covariates on the intensity function. As an alternative to MI models, Aalen (1993) proposed an additive regression model, whereby the intensity function is governed by the covariates as well as by past events through a linear regression with time-varying coefficients. Estimation of Aalen's nonparametric time-varying regression coefficients is performed via weighted least squares and their asymptotic properties are studied by using the martingale theory for counting processes (Martinussen & Scheike (2006)).

Sometimes, there may be prior non-sample information that is available about the covariates of interest, whereby a subset of the covariates is known to be irrelevant to the event occurrence rate. That is, in general, the investigator may have prior uncertain information in the form of a linear hypothesis which restricts the regression coefficients to a subspace. If such information is correct, then restricted estimators which incorporate the hypothesis should be more efficient than the unrestricted least squares estimators. However, when the linear hypothesis is incorrect, the restricted estimators perform quite poorly and are largely dominated by the unrestricted estimators. A safe way, in between these two strategies, is to use James-Stein-type shrinkage estimators, which are linear combinations of the restricted and unrestricted estimators. In the classical linear regression models, the shrinkage estimators are known to dominate the unrestricted estimators over the whole parameter space and dominate the restricted estimators everywhere except in a small neighborhood of the linear restriction.

1.2 Problem Statement and Objectives

In this project, we propose shrinkage estimators for the nonparametric regression coefficients in Aalen's additive model under a general linear hypothesis involving the coefficients. We study the asymptotic joint distribution of the restricted and unrestricted estimators of such coefficients via martingale central limit theorems. Consequently, we define integrated distributional quadratic risks and biases of the proposed shrinkage estimators and compare them analytically with those of the restricted and unrestricted estimators. We then take on the task of comparing the performance of the estimators in small samples via a Monte Carlo simulation study. The methodology is then illustrated by using a data set on the survival times of patients with primary billiary cirrhosis.

This project is organized as follows. In the following few sections, we introduce the existing literature on Aalen's additive hazard regression model. We define the unrestricted estimators of the cumulative regression coefficients and re-iterate some of the results on their asymptotic normality through martingale theory. To render this project self-contained, a brief introduction to counting processes and related martingale theory is provided in Appendix A. In chapter 2, we define a general linear hypothesis about the regression coefficients and provide restricted estimators of the cumulative coefficients. We also study the joint asymptotic normality of the restricted and unrestricted estimators. We then define James-Stein-type shrinkage estimators of the coefficients under the prior uncertain information given in the form of the linear hypothesis. We define and study the distributional quadratic biases and risks of the proposed shrinkage estimators and compare them asymptotically to those of the restricted and unrestricted estimators. In chapter 3, we conduct Monte Carlo simulations examining the small sample performance of the estimators and provide an application of the methods to a data set on the survival times of patients with primary billiary cirrhosis.

1.3 Aalen's Additive Hazards Regression Model

Survival data usually come in complex form where a censored time-to-event variable, along with some past history, is recorded. The history consists of the past of the time-to-event process itself as well as an accompanying covariates process. The objective is to study the relationship between event occurrence rates and the history variables provided with it. The analysis of such data sets is easily formulated in terms of counting processes, so that existing martingale theory can be employed to study properties of estimators thereof. For the i^{th} individual, we record their event or censoring time, say T_i, as well as some important additional information. Therefore, let $[N_i(t), X_i(t), Y_i(t)]$ for $i = 1, ..., n$ be the random sample from the i^{th} individual, where

$$\begin{aligned} N_i(t) &= \text{number of events up to time } t, \\ X_i(t) &= \text{the } i^{th} \text{ covariate of a locally bounded vector of } k \text{ covariates} \\ \text{and} \quad Y_i(t) &= \text{risk indicator at time } t, \end{aligned}$$

i.e.

$$Y_i(t) = \begin{cases} 1, & \text{if } i^{th} \text{ individual is at risk at time } t, \\ 0, & \text{otherwise .} \end{cases}$$

Further, let $\{\mathcal{F}_{it}, t \geqslant 0\}$, be the family of $\mathfrak{S}$-fields generated by the process $\{N_i(t); t \geq 0\}$. $\mathcal{F}_t$ is the $\mathfrak{S}$-field generated by $\bigcup_{i=1}^{n} \mathcal{F}_{it}$. The natural filtration of any counting process is right-continuous, so we can state that $\{\mathcal{F}_t, t \geq 0\}$ is a right-continuous filtration. In this paper, we consider that $\{(X_i(t), Y_i(t)); t \geq 0\}$ is a counting process adapted to the filtration $\{\mathcal{F}_t; t \geq 0\}$.

Following Martinussen & Scheike (2006), Aalen's nonparametric additive regression model is defined through the intensity function of the counting process, $N(t)$, as follows

$$\lambda(t) = Y(t)h(t|X) = Y(t)X'(t)\beta(t) \tag{1.1}$$

with *cumulative intensity function* given by

$$\Lambda(t) = \int_0^t \lambda(s)ds.$$

The nonparametric functions, $\beta_j(t)$ for $j = 1, ..., k$, are assumed to be locally integrable, i.e.

$$\int_0^t |\beta(s)|ds < \infty, \tag{1.2}$$

for all $0 \leq t \leq \tau$, where τ is the maximum time which the investigators consider (often end-of-study time).

In general, a major objective is to estimate the cumulative regression coefficients defined as

$$B_j(t) = \int_0^t \beta_j(u)du \qquad \text{for } \; j = 1, ..., k.$$

As proposed in Aalen (1993), this is easily achieved by using the least squares technique. For simplicity of notation, let us redefine the vector of covariates to incorporate the risk indicator functions, and organize them in a design matrix,

$$X(t) = (Y_1(t)X_1(t), ..., Y_n(t)X_n(t))',$$

so that $\lambda(t) = X'(t)\beta(t)$.

Recall from Section 1.1 that $N_i(t)$ is the number of events up to time t, where $\{N_i(t); t \geq 0\}$ is a counting process. Then, by equation (1.1),

$$\Lambda(t) = \int_0^t X'(s)\beta(s)ds,$$

and by Doob-Meyer decomposition, we get $N(t) - \Lambda(t) = M(t)$ where $M(t)$ is an $\mathcal{F}_t - martingale$ (for brief introduction to martingales, please see Appendix A). Then we can write $N(t) = M(t) + \Lambda(t)$

$$\begin{aligned} &\Rightarrow\ dN(t) = dM(t) + d\left[\int_0^t X'(s)\beta(s)ds\right] \\ &\Rightarrow\ dN(t) = dM(t) + X'(t)\beta(t)dt \qquad (1.3) \\ &\Rightarrow\ dN(t) = dM(t) + X'(t)dB(t). \end{aligned}$$

Now, since M is a martingale, we have $\mathrm{E}[dM(t)] = 0$. Therefore, heuristically and in parallel with the usual regression models, one could propose least squares estimators for $dB(t)$. Now assume that $X'(t)W(t)X(t)$ is a full rank matrix, for some predictable diagonal $k \times k$ weight matrix $W(t)$. In the case where $X(t)$ is not of full rank, it suffices to replace the usual matrix inverse by a generalized matrix inverse. So, let

$$X^{-1}(t) = (X'(t)W(t)X(t))^{-1}X'(t)W(t). \qquad (1.4)$$

From equation 1.3, we have

$$dN(t) = 0 + X'(t)d\widehat{B}(t)$$

$$\Rightarrow d\widehat{B}(t) = X^{-1}(t)dN(t).$$

This leads to the estimators

$$\Rightarrow \widehat{B}(t) = \int_0^t X^{-1}(s)dN(s). \tag{1.5}$$

Obviously, (1.3) and (1.5) imply that

$$\widehat{B}(t) = \int_0^t X^{-1}(s)[dM(s) + X'(t)dB(s)],$$

and by (1.4)

$$\widehat{B}(t) = \int_0^t X^{-1}(s)dM(s) + B(t).$$

Since

$$\int_0^t X^{-1}(s)dM(s)$$

is an $\mathcal{F}_t$-martingale with mean 0, we get

$$\mathrm{E}[\widehat{B}(t)] = 0 + 1 \times B(t) = B(t),$$

which shows that $\hat{B}(t)$ is an unbiased estimator.

The derivation of the least squares estimator can also be done slightly more rigorously by noticing that the model in (1.3) can be considered a regression model of the form $\tilde{Y}(t) = X'(t)b(t) + \varepsilon(t)$ where $\tilde{Y}(t) = dN(t)$, $b(t) = dB(t)$ is the vector of coefficients, with the matrix of covariates $X(t)$, and $\varepsilon(t) = dM(t)$ being the model's error term. Therefore we need to minimize the squared residual function

$$L(b(t)) = (\tilde{Y}(t) - X'(t)b(t))W(t)(\tilde{Y}(t) - X'(t)b(t))'$$

with respect to $b(t)$, where $W(t)$ is a weight matrix. Then, taking the derivative with respect to $b(t)$, we get

$$\frac{\partial L}{\partial b(t)} = -2X(t)W(t)\tilde{Y}(t) + 2X(t)W(t)X'(t)b(t) \overset{set}{=} 0.$$

Further,

$$\frac{\partial^2 L}{\partial (b(t))^2} = -2X(t)W(t)X'(t),$$

which is a positive-definite matrix by part (iv) of Condition $\mathcal{A}$, stated in the next section. Therefore, the least squares estimator for $b(t)$ is

$$\hat{b}_{LS}(t) = [X(t)W(t)X'(t)]^{-1}X(t)W(t)dN(t).$$

By using the relationship between $b(t)$ and $B(t)$, the resulting estimator for $B(t)$ is

$$\widehat{B}_{LS}(t) = \int_0^t [X(s)W(s)X'(s)]^{-1}X(s)W(s)dN(s)$$

for $0 \leq t \leq \tau$.

1.4 Asymptotic normality of $\hat{B}(t)$

In this section, we state a result on the asymptotic normality of Aalen's least squares estimator for the cumulative regression coefficients. To this end, we need some further notation and regularity conditions. Let,

$$\Gamma_{2js}(t) = \sum_{i=1}^{n} W_i(t)X_{ij}(t)X_{is}(t)$$

and

$$\Gamma_{3jsl}(t) = \sum_{i=1}^{n} W_i^2(t)X_{ij}(t)X_{is}(t)X_{il}(t)$$

for $j, s, l = 1, ..., k$. Let also,

$$\Omega^{-1}(t) = \mathrm{E}[W_1(t)X_1'(t)X_1(t)].$$

Some results given in throughout this paper use operations such as supremums and infimums of an uncountable collection of random variables. Thus, in order to guarantee that the resulting quantities are random variables, we assume that all processes under consideration are separable. It should be noted that this restriction is without loss of generality, and this is just for technical consideration. Indeed, it is well known that every stochastic process is equivalent to a separable process (Doob (1953, p. 51-57), Burrill (1972, p. 436-443), Billingsley (1995, p. 531)).

The following condition will be referred to throughout this paper in order to establish most of the asymptotic results of the estimators given.

- **Condition $\mathcal{A}$**

i) $\{(X_i(t), N_i(t)),\ 0 \le t \le \tau\}$ for $i = 1, ..., n$ are i.i.d

ii) $\mathrm{E}\left[\sup_{0\le t\le\tau} \left|W_1^2(t)X_{1j}(t)X_{1s}(t)X_{1l}(t)\right|\right] < \infty$ for $j, s, l = 1, ..., k$

iii) $\mathrm{E}\left[\sup_{0\le t\le\tau} |W_1(t)X_{1j}(t)X_{1s}(t)|\right] < \infty$ for $j, s = 1, ..., k$

iv) $X'(t)W(t)X(t)$ is a positive-definite matrix,$\forall\ t \in [0, \tau]$

v) $\mathrm{E}[W_1(t)X_1'(t)X_1(t)]$ is non-singular $\forall\ t \in [0, \tau]$

$$vi) \quad \int_0^t |\beta(s)|ds < \infty, \ \forall \ t \ \epsilon \ [0, \tau].$$

Recall that a metric space-valued function is called càdlàg if it is right-continuous with left limits. Let $\mathcal{D}([0, \ \tau], \mathbb{R}^p)$ denote the space that consists of càdlàg functions on $[0, \tau]$ into $\mathbb{R}^p$ and endowed with the sup-norm Skorohod topology (see Billingsley, 1995). For the sake of simplicity, we let $\mathcal{D}([0, \tau])$ stand for $\mathcal{D}([0, \tau], \mathbb{R}^p)$.

According to Theorem 5.1.1 of Martinussen and Scheike, if Condition $\mathcal{A}$ holds, then

$$\sqrt{n}(\widehat{B}(t) - B(t)) \xrightarrow{\mathcal{D}} U$$

on $\mathcal{D}([0, \tau])$, where U is a Gaussian martingale with covariance function

$$\Phi(t) = \int_0^t \Omega(s) \mathrm{E}[W_1^2(s) X_1'(t) X_1(t) X_1'(t) \beta(s)] \Omega(s) ds.$$

1.5 MLE of Aalen's regression coefficients

The argument used thus far to derive estimators of the cumulative regression coefficients in Aalen's additive model was through least squares analogy. However, it is also possible to derive the same estimator through a likelihood argument. To this end, we notice that log-likelihood function for $\beta(t)$ is simply

$$l(\beta(t)|X(t)) = \sum_{i=1}^{n} \int ln(X_i'(t)\beta(t)) dN_i(t) - \int X_i'(t)\beta(t) dt.$$

To find the maximum likelihood estimate for $\beta(t)$, one could solve the system of score equations, $\dfrac{\partial l(\beta(t)|X(t))}{\partial \beta(t)} = 0$. Martinussen and Scheike (2006, p. 114) choose to take this derivative heuristically, leaving us with the score functions:

$$X'(t) \operatorname{diag}\left(\frac{Y_i(t)}{\lambda_i(t)}\right)(dN(t) - X(t) dB(t)). \tag{1.6}$$

Setting this equal to 0,

$$X'(t)W(t)dN(t) = X'(t)W(t)X(t)dB(t)$$

$$\Rightarrow \; dB(t) = [X'(t)W(t)X(t)]^{-1}X'(t)W(t)dN(t)$$

where $W(t) = diag\left(\frac{Y_i(t)}{\lambda_i(t)}\right)$. Thus,

$$\tilde{B}(t) = \int [X'(t)W(t)X(t)]^{-1}X'(t)W(t)dN(t).$$

This is obviously same as the estimator obtained through least squares method. In the next chapters, we shall use this estimator as the basis for building shrinkage estimators of the regression coefficients. The least squares estimator will be called the unrestricted estimator of the cumulative regression coefficients in the sense that it does not assume or incorporate any prior information provided in the form of linear hypothesis. We then develop the restricted estimators of the cumulative regression coefficients and study their asymptotic properties as well as their joint asymptotic normality with the unrestricted least squares estimators described earlier.

In passing, we notice that given the definition of $B(t)$, we can extract an estimate for $\beta(t)$ from $\tilde{B}(t)$ through kernel smoothing technique. The reader is referred to Martinussen & Sheike (2006, p. 114-115) for further discussion on this topic.

Chapter 2

The proposed shrinkage estimators

The estimators of $B(t)$ defined in the previous chapter are known as *Unrestricted Estimators (UE's)* because they are not subject to any constraints. There are many other methods of finding unrestricted estimators - when discussing estimators, we tend to assume that they are unrestricted unless a restriction is specified. From now on, we will use UE to stand for unrestricted estimators.

A *Restricted Estimator (RE)* is simply an estimator which is computed in the same way as its UE, under some specified condition. It may be the case that we know some information about the parameter β or we only wish to deal with it under certain restrictions. In fact, when the restriction holds, the RE is known to be more efficient than its related UE. We will use the subscript $_R$ to denote RE's from this point on. In this chapter, we derive restricted estimators of $B(t)$ under a linear hypothesis on the regression coefficients and study their asymptotic properties. We also examine the asymptotic joint normality of the proposed restricted estimator and the unrestricted estimators derived in Chapter 1. We then proceed and propose James-Stein-type shrinkage estimators of $B(t)$ and study their performance in comparison with the restricted and unrestricted estimators via a concept known as asymptotic distributional risk and biases.

2.0.1 Restricted Estimators of $B(t)$

In this subsection, we will derive an estimator for $dB(t)$, restricted to some constraints based on prior knowledge. This will be used in deriving the restricted estimator for $B(t)$, which is what we are actually interested in.

Suppose we have some prior information about the parameter of interest. In particular, the parameter $b(t) = dB(t)$ is supposed to satisfy the following restriction:

$$Rb(t) = r_1(t) \quad \text{for} \quad 0 \leq t \leq \tau$$

where R is a known $q \times k$ matrix of rank $q < k$ and $r_1(t)$ is a known, q-column vector which is Riemman integrable on every compact subset of $\mathbb{R}$.

Again, we start with equation (1.3) :

$$dN(t) = X'(t)dB(t) + dM(t).$$

In order to derive the restricted estimator, $\tilde{b}_R(t)$, we will use Lagrange multipliers. The Lagrange function of $b(t)$ and its Lagrange multipliers, λ, is:

$$L(b(t), \lambda) = (\tilde{Y}(t) - X'(t)b(t))W(t)(\tilde{Y}(t) - X'(t)b(t))' - 2\lambda\left(Rb(t) - r_1(t)\right).$$

Then, taking the derivative with respect to $b(t)$ gives us k linear equations, which we set equal to 0:

$$\frac{\partial L}{\partial b(t)} = -2X(t)W(t)\tilde{Y}(t) + 2X(t)W(t)X'(t)b(t) - 2R'\hat{\lambda} \overset{set}{=} 0. \tag{2.1}$$

Now, taking the derivative with respect to the vector of Lagrange multipliers λ and evaluating them at $(\tilde{b}_R(t), \hat{\lambda})$ gives us q linear equations, which we also set equal to 0:

$$\frac{\partial L}{\partial \lambda} = 2R\tilde{b}_R(t) - 2r_1(t) \quad \overset{set}{=} \quad 0. \tag{2.2}$$

This leaves us with a system of $(k+q)$ linear equations with $(k+q)$ unknowns.

Equation (2.2) implies that

$$R\tilde{b}_R(t) = r_1(t),$$

and equation (2.1) implies that

$$X(t)W(t)\tilde{Y}(t) - X(t)W(t)X'(t)\tilde{\beta}_R(t) + R'\hat{\lambda} = 0.$$

Doing some algebraic manipulations of this equation, we can solve for $\hat{\lambda}$. We get

$$X(t)W(t)X'(t)\tilde{b}_R(t) = X(t)W(t)\tilde{Y}(t) + R'\hat{\lambda}$$

$$\Rightarrow \tilde{b}_R(t) = [X(t)W(t)X'(t)]^{-1}X(t)W(t)\tilde{Y}(t) + [X(t)W(t)X'(t)]^{-1}R'\hat{\lambda}$$

$$\Rightarrow \tilde{b}_R(t) = \hat{\beta}_{LS}(t) + [X(t)W(t)X'(t)]^{-1}R'\hat{\lambda} \tag{2.3}$$

where $\hat{b}_{LS}(t)$ is the familiar least squares regression estimator for $b(t)$, i.e. $\hat{b}_{LS}(t) = [X(t)W(t)X'(t)]^{-1}X(t)W(t)\tilde{Y}(t)$.

Then, $R\tilde{b}_R(t) = R\hat{b}_{LS}(t) + R[X(t)W(t)X'(t)]^{-1}R'\hat{\lambda}$.

Hence, since $R[X(t)W(t)X'(t)]^{-1}R'$ is positive-definite (see Proposition A.1.2 in the appendix), we get

$$[R(X(t)W(t)X'(t)R']^{-1}R\tilde{b}_R(t) = [R[X(t)W(t)X'(t)]^{-1}R']^{-1}R\hat{b}_{LS}(t) + \hat{\lambda}$$

$$\Rightarrow \hat{\lambda} = [R[X(t)W(t)X'(t)]^{-1}R']^{-1}(R\tilde{b}_R(t) - R\hat{b}_{LS}(t)) \ .$$

So from (2.2),

$$\hat{\lambda} = [R[X(t)W(t)X'(t)]^{-1}R']^{-1}\left(r_1(t) - R\hat{b}_{LS}(t)\right).$$

Plugging this back into (2.3), we get

$$\begin{aligned}\tilde{b}_R(t) &= \hat{b}_{LS}(t) + [X(t)W(t)X'(t)]^{-1}R'[R[X(t)W(t)X'(t)]^{-1}R']^{-1}\left(r(t) - R\hat{b}_{LS}\right)\\ &= (I_k - [X(t)W(t)X(t)]^{-1}R'[R[X(t)W(t)X'(t)]^{-1}R']^{-1}R)\hat{b}_{LS}(t)\\ &\qquad + [X(t)W(t)X'(t)]^{-1}R'[R[X(t)W(t)X'(t)]^{-1}R']^{-1}r_1(t)\\ &= (I_k - A_n(t)R)\hat{b}_{LS}(t) + A_n(t)r_1(t)\end{aligned}$$

where

$$A_n(t) = [X(t)W(t)X'(t)]^{-1}R'[R[X(t)W(t)X'(t)]^{-1}R']^{-1}. \tag{2.4}$$

Recall that the restricted estimator for $B(t)$ is our main focus here, and that

$$B(t) = \int_0^t dB(s)\ ,\ 0 \le t \le \tau.$$

Accordingly, the RE for $B(t)$ is given by

$$\widetilde{B}_R(t) = \int_0^t d\widetilde{B}_R(s) = \int_0^t \tilde{b}_R(s)ds.$$

Formally, the following proposition gives the RE for $B(t)$, for $0 \le t \le \tau$.

Proposition 2.0.1. *The restricted estimator for $B(t)$ can be written in the following way:*

$$\widetilde{B}_R(t) = \int_0^t (I_k - A_n(s)R)\hat{b}_{LS}(s)ds\ +\ \int_0^t A_n(s)r_1(s)ds, \tag{2.5}$$

for $0 \le t \le \tau$.

The proof of this proposition is given in Appendix B.

2.0.2 Asymptotic Normality of the restricted estimators

In this subsection, we derive the asymptotic normality of the restricted and unrestricted estimators. At this point, we have a restricted estimator for $B(t)$ and it is important that we know it's aymptotic distribution. This is particularly useful in studying the asymptotic optimality of the proposed estimators.

As a preliminary step, below we recall the asymptotic properties of $\widehat{B}_{LS}(t)$.

Proposition 2.0.2. *If Condition $\mathcal{A}$ holds, then*

$$n^{\frac{1}{2}}(\widehat{B}_{LS} - B)\ \xrightarrow[n\to\infty]{D} U \tag{2.6}$$

on $\mathcal{D}([0,\,\tau])$*, where* U *is a Gaussian martingale with mean* 0 *and covariance function*

$$\int_0^t \Omega(s)\mathrm{E}[W_1^2(s)X_1(s)X_1'(s)X_1(s)\beta(s)]\Omega(s)ds \tag{2.7}$$

for $0 \le t \le \tau$ *.*

The proof is outlined in Martinussen & Scheike (2006, p. 110 - 112). Also, for completeness, we provide a proof with further details in the appendix.

Note that, following Proposition 2.0.1, $\widetilde{B}_R(t)$ has the following decomposition:

$$\begin{aligned}\widetilde{B}_R(t) &= \left[\int_0^t \hat{b}_{LS}(s)ds\right] \\ &\quad - \int_0^t [X(s)W(s)X'(s)]^{-1}R'[R[X(s)W(s)X'(s)]^{-1}R']^{-1}(R\hat{b}_{LS}(s) - r_1(s))ds \\ &= \widehat{B}_{LS}(t) - \int_0^t A_n(s)(R\hat{b}_{LS}(s) - r_1(s))ds\end{aligned}$$

where $W(s)$ is a weight matrix.

We want to study the asymptotic behaviour of $\sqrt{n}(\widetilde{B}_R(t) - B(t))$, and we can do this by recognizing its relationship with $\sqrt{n}(\widehat{B}_{LS}(t) - B(t))$, which has asymptotic behaviour that we are already familiar with.

Also, in the sequel, we consider the following sequence of *local alternatives:*

$$H_{1,n} : Rb(t) = r_1(t) + \frac{\delta_1(t)}{\sqrt{n}},\ n = 1, 2, 3, ... \tag{2.8}$$

where R is a known $q \times k$ full-rank matrix, and $r_1(t)$ and $\delta_1(t)$ are known integrable $q \times 1$ vectors, Riemman-integrable on every compact subset of $\mathbb{R}$.

Notice that

$$\begin{aligned}
&\sqrt{n}(\widetilde{B}_R(t) - B(t)) \\
= \quad & \sqrt{n}(\widehat{B}_{LS}(t) - B(t)) \;-\; \sqrt{n}\left(\int_0^t A_n(s)(R\hat{b}_{LS}(s) - r_1(s))ds\right) \\
= \quad & \sqrt{n}(\widehat{B}_{LS}(t) - B(t)) - \sqrt{n}\int_0^t A_n(s)(RdB(s) - r_1(s))ds \\
& -\int_0^t A_n(s)Rd[\sqrt{n}(\widehat{B}_{LS}(s) - B(s))].
\end{aligned}$$

So, under the sequence of local alternatives in (2.8), and following the results in Martinussen & Scheike (2006, p. 110),

$$\begin{aligned}
\sqrt{n}(\widetilde{B}_R(t) - B(t)) \;=\; & \sqrt{n}(\widehat{B}_{LS}(t) - B(t)) - \int_0^t A_n(s)\delta_1(s)ds \\
& -\int_0^t A_n(s)Rd[\sqrt{n}(\widehat{B}_{LS}(s) - B(s))]
\end{aligned}$$

$$\begin{aligned}
= \quad & n^{-\frac{1}{2}}\int_0^t \Omega(s)X(s)W(s)dM(s) + n^{-\frac{1}{2}}\int_0^t \left\{(n^{-1}\Gamma(s))^{-1} - \Omega(s)\right\} X(s)W(s)dM(s) \\
& -\int_0^t A_n(s)\delta_1(s)ds \\
& -\int_0^t A_n(s)R \\
& \quad \times \left(n^{-\frac{1}{2}}\Omega(s)X(s)W(s)dM(s) + n^{-\frac{1}{2}}\left\{(n^{-1}\Gamma(s))^{-1} - \Omega(s)\right\} X(s)W(s)dM(s)\right).
\end{aligned}$$

Further, note that the sequence of local alternatives in 2.8 implies the following sequence of local alternatives,

$$H_{2,n} : RB(t) = r_2(t) + \frac{\delta_2(t)}{\sqrt{n}}, \tag{2.9}$$

where R is the same known $q \times k$ full-rank matrix at (2.8),

$$r_2(t) = \int_0^t r_1(s)ds,$$

and

$$\delta_2(t) = \int_0^t \delta_1(s)ds.$$

Then we have the following very useful decomposition:

$$\sqrt{n}(\widetilde{B}_R(t) - B(t)) = P_{1,n}(t) + P_{2,n}(t) - P_{3,n}(t)$$

where

$$\begin{aligned} P_{1,n}(t) &= n^{-\frac{1}{2}} \int_0^t [I_k - A_n(s)R]\Omega(s)X(s)W(s)dM(s), &(2.10)\\ P_{2,n}(t) &= n^{-\frac{1}{2}} \int_0^t [I_k - A_n(s)R]\left\{(n^{-1}\Gamma(s))^{-1} - \Omega(s)\right\}X(s)W(s)dM(s), &(2.11)\\ P_{3,n}(t) &= \int_0^t A_n(s)\delta_1(s)ds. &(2.12)\end{aligned}$$

In order to show the convergence of $\sqrt{n}(\widetilde{B}_R(t) - B(t))$, we can study the convergence of the 3 seperate parts of this decomposition.

To simplify notation, let

$$A(t) = \Omega(t)R'[R\Omega(t)R']^{-1}$$

for $0 \leq t \leq \tau$.

We begin with the convergence of $P_{1,n}(t)$ to a Gaussian martingale.

Proposition 2.0.3. *Assume that Condition $\mathcal{A}$ and (2.9) hold, and let $P_{1,n}(t)$ be the random quantity given in (2.10). Then, $\{P_{1,n}(t), t \geq 0\}$ converges in distribution to a Gaussian martingale on $\mathcal{D}([0,\tau])$, with covariance function*

$$\Phi^*(t) = \int_0^t [I - A(s)R]\Omega(s)\mathrm{E}[W_1^2(s)[X_1(s)X_1'(s)]X_1(s)\beta(s)]\Omega(s)[I - R'A'(s)]ds.$$

The proof of this proposition is given in Appendix B. Further, in the following proposition, we show that $P_{2,n}(t)$ converges in probability to 0, uniformly on $[0,\tau]$.

Proposition 2.0.4. *Suppose that Condition $\mathcal{A}$ and (2.9) hold, and let $P_{2,n}(t)$ be the random quantity given in (2.11). Then,*

$$P_{2,n}(t) \xrightarrow[n\to\infty]{P} 0$$

uniformly over $[0,\tau]$.

The proof of this proposition is given in Appendix B. Concerning the quantity $P_{3,n}(t)$, the following proposition shows that it converges in probability to a non-random matrix.

Proposition 2.0.5. *Suppose that Condition $\mathcal{A}$ and (2.9) hold, and let $P_{3,n}(t)$ be the random quantity given in (2.12). Then,*

$$P_{3,n}(t) \xrightarrow[n\to\infty]{P} \int_0^t A(s)\delta_1(s)ds$$

uniformly over $[0,\tau]$.

The proof of this proposition is given in Appendix B.

Note that, from the convergence of each part of the decomposition, we can draw a conclusion about the convergence of their sum. In particular, since $\sqrt{n}(\widetilde{B}_R(t) - B(t)) = P_{1,n}(t) + P_{2,n}(t) - P_{3,n}(t)$, we conclude that

$$\sqrt{n}(\widetilde{B}_R - B) \xrightarrow[n\to\infty]{D} U^* \tag{2.13}$$

on $\mathcal{D}([0, \tau])$, where U^* is a Gaussian martingale such that

$$U^*(t) \sim \mathcal{N}_k\left(-\int_0^t \Omega(s)R'[R\Omega(s)R']^{-1}\delta_1(s)ds\ , \Phi^*(t)\right)$$

with

$$\begin{aligned}\Phi^*(t) = &\int_0^t [I - A(s)R]\Omega(s)\mathrm{E}[W_1^2(s)[X_1(s)X_1'(s)]X_1(s)\beta(s)]\Omega(s) \\ &\times [I - R'A'(s)]ds\end{aligned}$$

for $0 \leq t \leq \tau$.

Now we know the asymptotic behaviour of the restricted estimator, as well as that of the unrestricted estimator. More generally, we derive below the joint asymptotic normality of the UE and RE. To simplify notation, let

$$\xi_n(t) = \sqrt{n}(\widetilde{B}_R(t) - \widehat{B}_{LS}(t))$$

and

$$\eta_n(t) = \sqrt{n}(\widetilde{B}_R(t) - B(t)).$$

In addition, let

$$C(s) = \Omega(s)\mathrm{E}[W_1^2(s)[X_1(s)X_1'(s)]X_1(s)\beta(s)]\Omega(s).$$

Proposition 2.0.6. *Suppose that Condition $\mathcal{A}$ and* (2.9) *hold. Then*

$$(\xi_n'(t), \eta_n'(t))' \xrightarrow[n\to\infty]{D} (\xi_0'(t), \eta_0'(t))'$$

on $\mathcal{D}([0,\, \tau])$, where $\{(\xi_0'(t), \eta_0'(t))', t \geq 0\}$ is the Gaussian martingale with

$$(\xi_0'(t), \eta_0'(t))' \;\sim\; N_{2k}\left[\int_0^t \begin{pmatrix} I_k \\ I_k \end{pmatrix} A(s)\delta_1(s)ds,\;\; \Phi^{**}(t)\right]$$

with

$$\Phi^{**}(t) = \int_0^t \begin{pmatrix} A_{11}(s) & A_{12}(s) \\ A_{21}(s) & A_{22}(s) \end{pmatrix} ds,$$

where

$$\begin{aligned} A_{11}(s) &= A(s)RC(s)R'A'(s), \\ A_{12}(s) &= -A(s)RC(s)[I_k - R'A'(s)], \\ A_{21}(s) &= [I_k - A(s)R]C(s)R'A'(s), \\ A_{22}(s) &= [I_k - A(s)R]C(s)[I_k - R'A'(s)]. \end{aligned}$$

By using Proposition 2.0.6, we establish below Corollary 2.1. Briefly, Corollary 2.1 gives the asymptotic normality of $(\xi'_n(t), \eta'_n(t))'$ with a more tractable expression of the variance-covariance matrix, Φ^{**}.

Consider an additional condition, this time on the weight matrix, $W(t)$:

- **Condition $\mathcal{B}$**

The weight matrix, $W(t)$ satisfies

$$W_i(t) = \frac{1}{X_i(t)\beta(t)}, \quad i = 1, ..., n$$

for $0 \leq t \leq \tau$.

Note that the chosen $W_i(t)$ depends on the parameter $\beta(s)$. In practice, this parameter is replaced by the least-squares estimator $[X(t)X'(t)]^{-1}X'(t)dN(t)$. Further, as mentioned above, under Condition $\mathcal{B}$ we can further simplify the variance-covariance in Proposition 2.0.6. This is done in the following corollary:

Corollary 2.1. *Suppose that Condition $\mathcal{A}$, Condition $\mathcal{B}$ and* (2.9) *hold. Then*

$$(\xi'_n(t), \eta'_n(t))' \xrightarrow[n\to\infty]{D} (\xi'(t), \eta'(t))'$$

on $\mathcal{D}([0,\ \tau])$, where $\{(\xi'(t), \eta'(t))', t \geq 0\}$ is the Gaussian martingale with

$$(\xi'(t), \eta'(t))' \sim N_{2k}\left[\int_0^t \begin{pmatrix} I_k \\ I_k \end{pmatrix} A(s)\delta_1(s)ds\ ,\ \Phi^{***}(t)\right]$$

where

$$\Phi^{***} = \int_0^t \begin{pmatrix} A(s)R\Omega(s) & 0 \\ 0 & \Omega(s) - A(s)R\Omega'(s) \end{pmatrix} ds,$$

for $0 \leq t \leq \tau$.

2.2 Shrinkage Estimators

The *Shrinkage Estimator* was developed by Charles Stein (1956), and improved by Stein in collaboration with Willard James 5 years later.

In the following subsection, we let

$$\begin{aligned}
J(t) &= R'\left(\int_0^t R\Omega(s)R'ds\right)^{-1} R \\
\delta^*(t) &= \int_0^t \Omega(s)R'[R\Omega(s)R']^{-1}\delta(s)ds, \\
\Sigma_{11}(t) &= \int_0^t \Omega(s)R'[R\Omega(s)R']^{-1}R\Omega(s)ds \\
\Delta(t) &= \left(\int_0^t \delta^{*\prime}(s)ds\right) \times R'\left(\int_0^t R\Omega(s)R'ds\right)^{-1} R \times \left(\int_0^t \delta^*(s)ds\right) \\
q &= \text{rank}(\Sigma_{11}(t)J(t)\Sigma_{11}(t)) \\
\phi_n(t) &= \xi_n'(t)J(t)\xi_n(t) \xrightarrow[n\to\infty]{D} \phi(t) \sim \chi_q^2(\Delta(t)),
\end{aligned}$$

and

$$\varphi_n(t) = \xi_n'(t)\widehat{J}(t)\xi_n(t), \tag{2.14}$$

where $\widehat{J}(t)$ is a consistent estimator for $J(t)$, uniformly on $[0, \tau]$.

In addition, we have the following proposition.

Proposition 2.2.1. *Suppose that Conditions $\mathcal{A}$, $\mathcal{B}$ and 2.9 hold. Then*

$$i)\ \phi_n(t) \xrightarrow[n\to\infty]{D} \xi'(t)J(t)\xi(t) \sim\ \chi^2_q(\Delta(t)),$$

on $\mathcal{D}\left([0,\ \tau]\right)$, *and*

$$ii)\ \varphi_n(t) \xrightarrow[n\to\infty]{D} \xi'(t)J(t)\xi(t) \sim\ \chi^2_q(\Delta(t)), \quad on\ \mathcal{D}\left([0,\ \tau]\right)$$

The proof of this proposition can be found in Appendix B.

2.2.1 Shrinkage Estimator

The shrinkage estimator can be defined by:

$$\widehat{B}^S(t) = \widetilde{B}_R(t) + (1 - c\varphi_n^{-1}(t))(\widehat{B}_{LS}(t) - \widetilde{B}_R(t)) \tag{2.15}$$

where c, which is known as the *shrinkage constant*, is chosen in an interval such that $\widehat{B}^S(t)$ dominates $\widehat{B}_{LS}(t)$, and $\varphi_n(t)$ is defined in (2.14). In the sequel, we will consider $c = q - 2$.

The shrinkage estimator tends to overshrink the estimator, especially when $\varphi_n(t)$ is very small in comparison with c. To remedy this issue, the *Positive-Part Shrinkage (PS) Estimator* was developed by truncating the shrinkage estimator in the following way:

$$\begin{aligned}\widehat{B}^{S+}(t) &= \widetilde{B}_R(t) + max(0, 1 - c\varphi_n^{-1}(t))(\widehat{B}(t) - \widetilde{B}_R(t)) \\ &= \widetilde{B}_R(t) + \left(1 - \frac{q-2}{\varphi_n(t)}\right)^+ (\widehat{B}(t) - \widetilde{B}_R(t)), \end{aligned} \tag{2.16}$$

where

$$(f(t))^+ = max\,(0, f(t))\,.$$

2.3 Asymptotic Distributional Risk and Bias

In Ahmed & Nicol (1999), a paper on shrinkage estimators in non-linear regression, the authors state that because the test based on $\varphi_n(t)$ is consistent against fixed alternatives, the shrinkage estimators become asymptotically isomorphic to $\widehat{B}_{LS}(t)$ as $n \to \infty$.

2.3.1 Asymptotic Distributional Risk

In this section, we provide the asymptotic distributional risk for each of the estimators of $B(t)$. In order to evaluate the performance of the proposed estimators, we calculate their risks and compare them. The risk of an estimator is the expected value of its loss function, so let us consider the weighted quadratic *Loss Function* for an estimator $\hat{\theta}$ of θ:

$$L(\hat{\theta}, \theta; W^*) = n \int_0^{\tau} (\hat{\theta}(s) - \theta(s))' W^*(s)(\hat{\theta}(s) - \theta(s))ds. \tag{2.17}$$

Then the *Risk* is given by

$$R(\hat{\theta},\theta;W^*) = \mathrm{E}[L(\hat{\theta},\theta;W^*)] = \mathrm{E}\left[n\int_0^\tau (\hat{\theta}(s)-\theta(s))'W^*(s)(\hat{\theta}(s)-\theta(s))ds\right].$$

Now consider the asymptotic distribution of the loss function. We choose to look at the *asymptotic* risk since it is difficult (or, rather, impossible) to find the small-sample distribution of $(\hat{\theta}(s)-\theta(s))'W^*(s)(\hat{\theta}(s)-\theta(s))$.

Suppose that the distribution of the loss function converges to an integrable random variable, Ψ. Then the *Asymptotic Distributional Risk* (ADR) is defined by

$$ADR(\hat{\theta},\theta;W^*) = \mathrm{E}[\Psi].$$

In this paper we consider the loss function given in equation (2.17).

For the purpose of simplification, define

$$\overline{W}^*(s) = \int_s^\tau W^*(t)dt. \tag{2.18}$$

We will start with the ADR of the unrestricted estimator.

Proposition 2.3.1. *Suppose that Conditions $\mathcal{A}$, $\mathcal{B}$, and the sequence of local alternatives in* (2.9) *hold. Then,*

$$ADR\left(\widehat{B}_{LS},B;W^*\right) = \int_0^\tau tr\left(\Omega(s)\overline{W}^*(s)\right)ds.$$

The proof of this proposition can be found in Appendix B.

The ADR of the restricted estimator is given in the following proposition.

Proposition 2.3.2. *Suppose that Condition $\mathcal{A}$, Condition $\mathcal{B}$ and* (2.9) *hold. Then,*

$$\begin{aligned} ADR(\widetilde{B}_R, B; W^*) &= \int_0^\tau tr\left(\Omega(s)\overline{W}^*(s)\right) ds - \int_0^\tau tr\left(\Sigma(s)\overline{W}^*(s)\right) ds \\ &+ \int_0^\tau \delta^{*\prime}(s) W^*(s) \delta^*(s) ds. \end{aligned}$$

The proof of this proposition can be found in Appendix B.

Notice that, by Propositions 2.3.1 and 2.3.2,

$$\begin{aligned} ADR(\widetilde{B}_R, B; W^*) &= ADR\left(\widehat{B}_{LS}, B; W^*\right) \\ &- \int_0^\tau tr\left(\Sigma(s)\overline{W}^*(s)\right) ds + \int_0^\tau \delta^{*\prime}(s) W^*(s) \delta^*(s) ds. \end{aligned}$$

The following proposition provides the ADR of the shrinkage estimator.

Proposition 2.3.3. *Suppose that Condition $\mathcal{A}$, Condition $\mathcal{B}$ and* (2.9) *hold, and that $\widehat{B}^S(t)$ is the shrinkage estimator given in* (2.15)*. Then,*

$$\begin{aligned} ADR\left(\widehat{B}^S, B; W^*\right) &= ADR(\widehat{B}_{LS}, B; W^*) \\ &- \int_0^\tau tr\left(\Sigma(s)\overline{W}^*(s)\right) ds + \int_0^\tau \delta^{*\prime}(t) W^*(t) \delta^*(t) dt \\ &-2 \int_0^\tau E\left[1 - \frac{q-2}{D_1(t)}\right] \delta^{*\prime}(t) W^*(t) \delta^*(t) dt \\ &+ \int_0^\tau E\left[\left(1 - \frac{q-2}{D_1(t)}\right)^2\right] \text{trace}(W^*(t)\Sigma_{11}(t)) dt \end{aligned}$$

$$+\int_0^\tau E\left[\left(1-\frac{q-2}{D_2(t)}\right)^2\right]\ \delta^*(t)'W^*(t)\delta^*(t)dt,$$

where

$$D_1(t)=\chi^2_{q+2}(\ \delta^*(t)'W^*(t)\delta^*(t)) \tag{2.19}$$

and

$$D_2(t)=\chi^2_{q+4}(\ \delta^*(t)'W^*(t)\delta^*(t)). \tag{2.20}$$

The proof of this proposition can be found in Appendix B.

It is clear from the last proposition that, in order for our shrinkage estimator to be better than the unrestricted estimator (with respect to the risk),

$$\begin{aligned}
&-\int_0^\tau tr\left(\Sigma(s)\overline{W}^*(s)\right)ds+\int_0^\tau \delta^{*\prime}(t)W^*(t)\delta^*(t)dt\\
&-2\int_0^\tau E\left[1-\frac{q-2}{D_1(t)}\right]\delta^{*\prime}(t)W^*(t)\delta^*(t)dt\\
&+\int_0^\tau E\left[\left(1-\frac{q-2}{D_1(t)}\right)^2\right]\ \mathrm{trace}(W^*(t)\Sigma_{11}(t))dt\\
&+\int_0^\tau E\left[\left(1-\frac{q-2}{D_2(t)}\right)^2\right]\ \delta^*(t)'W^*(t)\delta^*(t)dt,
\end{aligned}$$

must be negative. We use this fact to prove the following proposition. Let $ch_{min}(A)$ and $ch_{max}(A)$ be the smallest and largest eigenvalues of the matrix A, respectively.

Proposition 2.3.4. *Suppose that Condition $\mathcal{A}$, Condition $\mathcal{B}$ and* (2.9) *hold. Also suppose that the weight matrix, $W^*(t)$, is an element of the set*

$$\left\{W^*(t):0\leq(q+2)ch_{max}\left(W^*(t)\Sigma_{11}(t)\right)\leq 2\mathrm{trace}\left(W^*(t)\Sigma_{11}(t)\right),\ t\geqslant 0\right\}. \tag{2.21}$$

Then,

$$ADR\left(\widehat{B}^{S}, B; W^{*}\right) \leq ADR\left(\widehat{B}_{LS}, B; W^{*}\right).$$

The proof of this proposition can be found in Appendix B.

The ADR of the positive-part shrinkage estimator is given in the following proposition, and follows similar steps to those in Proposition 2.3.3.

Proposition 2.3.5. *Suppose that Condition $\mathcal{A}$, Condition $\mathcal{B}$ and* (2.9) *hold. Then,*

$$\begin{aligned} ADR\left(\widehat{B}^{S^{+}}, B; W^{*}\right) &= \int_{0}^{\tau} tr\left(\left[\Omega_{11}(t) - \Sigma_{11}(t)\right] W^{*}(t)\right) dt + \int_{0}^{\tau} \delta^{*\prime}(t) W^{*}(t) \delta^{*}(t) dt \\ &\quad -2\int_{0}^{\tau} E\left[\left(1 - \frac{q-2}{D_{1}(t)}\right)^{+}\right] \delta^{*\prime}(t) W^{*}(t) \delta^{*}(t) dt \\ &\quad +\int_{0}^{\tau} E\left[\left(\left[1 - \frac{q-2}{D_{1}(t)}\right]^{+}\right)^{2}\right] \operatorname{trace}(W^{*}(t)\Sigma_{11}(t)) dt \\ &\quad +\int_{0}^{\tau} E\left[\left(\left[1 - \frac{q-2}{D_{2}(t)}\right]^{+}\right)^{2}\right] \delta^{*}(t)' W^{*}(t) \delta^{*}(t) dt, \end{aligned}$$

where $D_1(t)$ and $D_2(t)$ are defined as in equations (2.19) *and* (2.20)*, respectively.*

This result follows directly from the proof of Proposition 2.3.3, found in Appendix B.

It is possible to write $ADR\left(\widehat{B}^{S^{+}}, B; W^{*}\right)$ as a function of $ADR\left(\widehat{B}^{S}, B; W^{*}\right)$, which is illustrated in the following proposition. This will be useful for comparison purposes.

Proposition 2.3.6. *Suppose that the conditions in Proposition 2.3.5 hold. Then*

$$ADR\left(\widehat{B}^{S^{+}}, B; W^{*}\right)$$

$$
\begin{aligned}
= \; & ADR\left(\widehat{B}^{S}, B; W^{*}\right) \\
& +2\int_0^{\tau} E\left[\mathbf{I}_{(-\infty,0)}\left\{1-\frac{q-2}{D_1(t)}\right\}\left(1-\frac{q-2}{D_1(t)}\right)\right]\delta^{*\prime}(t)W^{*}(t)\delta^{*}(t)dt \\
& +\int_0^{\tau} E\left[\mathbf{I}_{(-\infty,0)}\left\{1-\frac{q-2}{D_1(t)}\right\}\left(1-\frac{q-2}{D_1(t)}\right)^2\right]\operatorname{trace}(W^{*}(t)\Sigma_{11}(t))dt \\
& +\int_0^{\tau} E\left[\mathbf{I}_{(-\infty,0)}\left\{1-\frac{q-2}{D_1(t)}\right\}\left(1-\frac{q-2}{D_1(t)}\right)^2\right]\delta^{*}(t)'W^{*}(t)\delta^{*}(t)dt.
\end{aligned}
$$

The proof of this proposition can be found in Appendix B.

While the shrinkage estimator has a lower risk than the unrestricted estimator, the positive-part shrinkage estimator is an even bigger improvement.

Proposition 2.3.7. *Suppose that Condition $\mathcal{A}$, Condition $\mathcal{B}$ and* (2.9) *hold. Also, suppose that the weight matrix, $W^{*}(t)$, is chosen to satisfy* (2.21) *Then,*

$$ADR\left(\widehat{B}^{S+}, B; W^{*}\right) \leq ADR\left(\widehat{B}^{S}, B; W^{*}\right).$$

The proof of this proposition follows directly from the proof of Proposition 2.3.4.

2.3.2 Asymptotic Distributional Bias

In addition to minimizing risk, we are also interested in minimizing bias. In this section, we will provide the asymptotic distributional bias (ADB) for each of the estimators. Using results from George Judge and Mary Bock, Ahmed & Nicol showed that the *asymptotic distributional bias* of an estimator $\hat{\theta}$ of θ is defined by:

$$ADB(\hat{\theta}, \theta) = \lim_{n\to\infty} \mathrm{E}[n^{\frac{1}{2}}(\hat{\theta}-\theta)].$$

We will start with the ADB of the unrestricted estimator.

Proposition 2.3.8. *Suppose that Conditions $\mathcal{A}$, $\mathcal{B}$, and the sequence of local alternatives in* (2.9) *hold. Then,*

$$ADB\left(\widehat{B}_{LS}, B\right) = 0.$$

The proof of this proposition follows directly from Proposition 2.0.2.

The ADB of the restricted estimator is given in the following proposition.

Proposition 2.3.9. *Suppose that Condition $\mathcal{A}$, Condition $\mathcal{B}$ and* (2.9) *hold. Then,*

$$ADB(\widetilde{B}_R, B) = \int_0^\tau \int_0^t A(s)\delta_1(s)dsdt.$$

The proof of this proposition follows directly from relation (2.13).

We will now provide the ADB of the shrinkage estimator.

Proposition 2.3.10. *Suppose that Condition $\mathcal{A}$, Condition $\mathcal{B}$ and* (2.9) *hold. Then,*

$$ADB(\widehat{B}^S, B) \quad = \quad \int_0^\tau \delta^*(t)dt + \int_0^\tau E\left[1 - \frac{q-2}{D_1(t)}\right]\delta^*(t)dt.$$

The proof of this proposition can be found in Appendix B.

We provide the ADB of the positive-part shrinkage estimator, $\widehat{B}^{S^+}$, in the following proposition. The proof follows directly from the proof of Proposition 2.3.10.

Proposition 2.3.11. *Suppose that Condition $\mathcal{A}$, Condition $\mathcal{B}$ and* (2.9) *hold. Then,*

$$ADB(\widehat{B}^{S^+}, B) = \int_0^\tau \delta^*(t)dt + \int_0^\tau E\left[\left(1 - \frac{q-2}{D_1(t)}\right)^+\right]\delta^*(t)dt.$$

The proof of this proposition can be found in Appendix B.

Chapter 3

Empirical Studies

3.1 Simulation study

In this section we study the performance of the shrinkage estimators by using Monte Carlo simulations. To this end, we consider a simple survival model whose intensity function for the ith individual is given by $\lambda(t|X_i(t)) = Y_i(t)X_i\beta(t)$, where we set $\beta(t) = (\beta_0 t, \beta_2 t, ..., \beta_4 t)$, where β_q for $q = 0, ..., 4$ are unknown but constant and the covariate process is time-independent. Thus, we assumed the time-dependent regression coefficients to be linear. This leads to a cumulative intensity function given by $\Lambda(t) = t^2 Y_i(t)X_i\beta$, where $\beta = (\beta_0, ..., \beta_4)$ and since we only allow one-event for each subject, $\int_0^t Y_i(s)ds = Y_i(t)$. This enables us to generate random times, T, via $Uniform(0, 1)$ numbers, U, by inverting the relationship $1 - F(T) = \exp\{-\Lambda(T)\}$. We generated the covariates $X_{i1}, ..., X_{i4}$ from $Uniform(0, 20)$ and we set $\beta = c(2, 0, 0, 0, 0)$, under the null hypothesis and $\beta = c(2, 0, 0, 0, 0 + \delta)$ under the alternative hypothesis, where

δ varied from zero to one with steps of .05. Our restriction is given by $R\beta t = r$, where

$$R = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

and

$$r = (0\ 0\ 0\ 0)'.$$

Thus, generated random survival times were then censored by using independent random variates generated from $Uniform(0, 3)$. This setting led to censoring rates varying from 5-15%. Each scenario was simulated 1000 times for sample sizes of $n = 250, 500, 750$ and 1000. In each scenario, we computed the empirical mean squared errors (MSE) of the all the estimators (shrinkage, positive shrinkage, restricted and unrestricted) and, by taking the unrestricted estimator as benchmark, we reported the ratios of these MSEs relative to the benchmark. The results are summarised in Figures 1-4.

It is clearly visible that the proposed shrinkage estimators outperform the usual restricted and unrestricted estimators on almost all of the parameter space. When the null hypothesis is true (in other words, $\delta = 0$), we see that the best estimator is the restricted estimator as foreseen from the analytic derivations of the chapter 2, while its performance deteriorates substantially when we get away from the null space. On the other hand, the positive shrinkage estimator dominates the unrestricted estimator throughout the null and alternative space and converges to it in terms of MSE for all of the sample sizes considered. However, the shrinkage estimator seems to be worse than the unrestricted at the null hypothesis for sample sizes that are smaller than $n = 1000$. This may indicate that the asymptotic distributional risk dominance of the shrinkage estimator requires quite large samples to kick in. In summary, this simulation study agrees with our theoretical results about the asymptotic distribution risk, which were developed in the previous section.

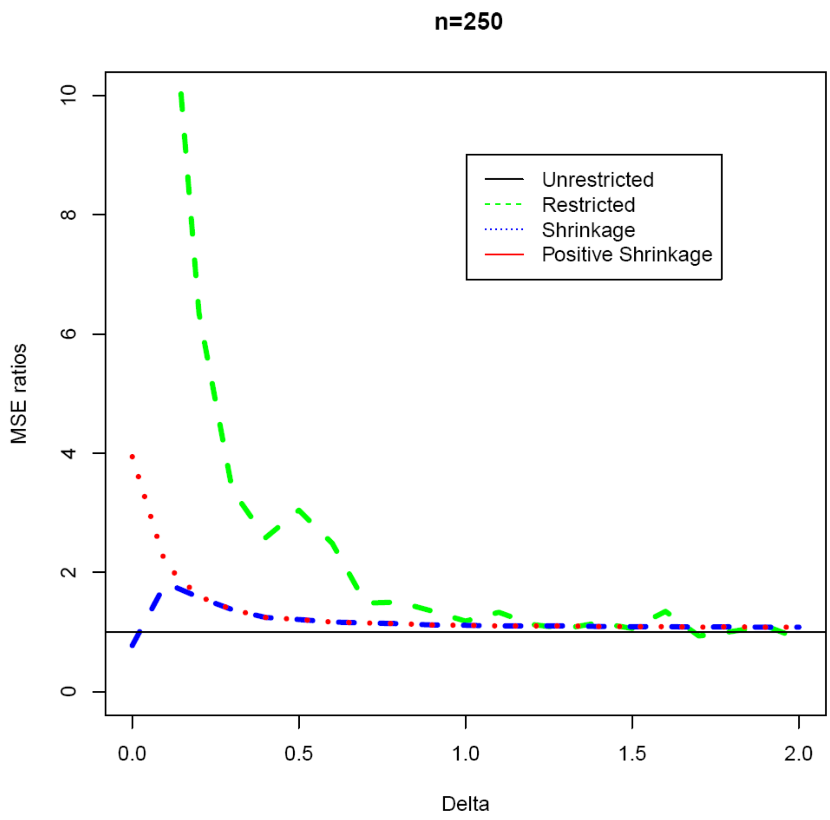


Figure 3.1: Ratios of MSE for all estimators relative to the unrestricted estimator with delta varying over the parameter space and sample size n=250

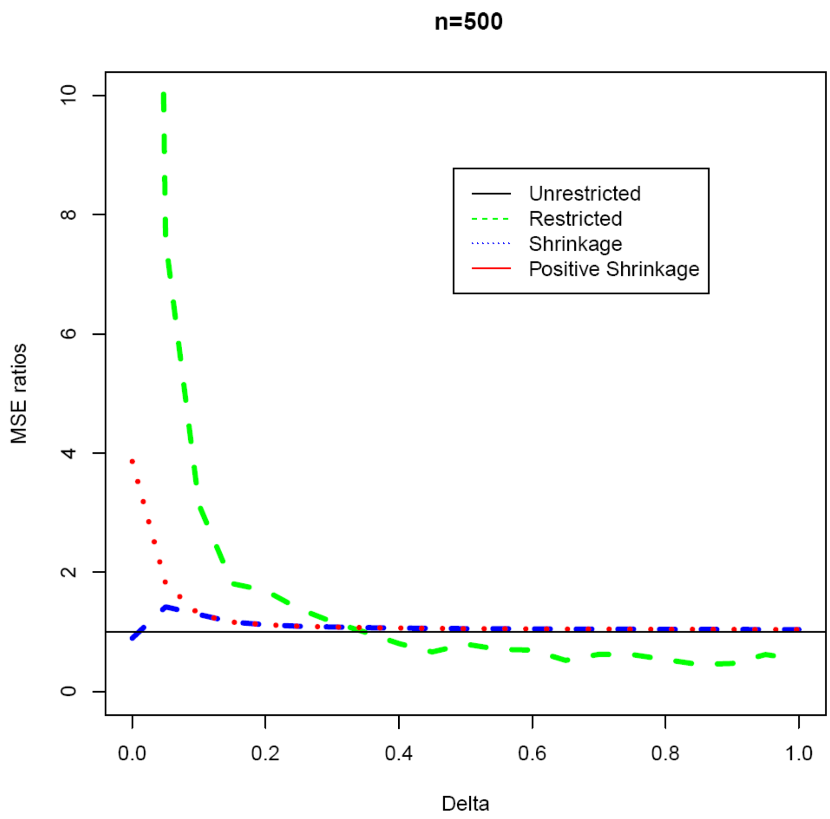


Figure 3.2: Ratios of MSE for all estimators relative to the unrestricted estimator with delta varying over the parameter space and sample size n=500

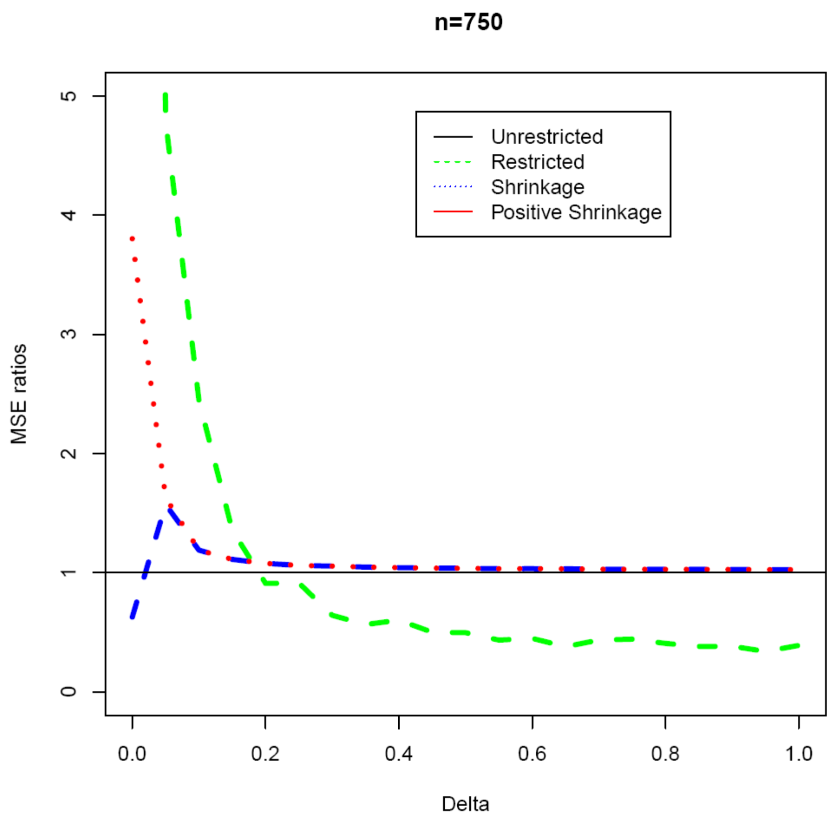


Figure 3.3: Ratios of MSE for all estimators relative to the unrestricted estimator with delta varying over the parameter space and sample size n=750

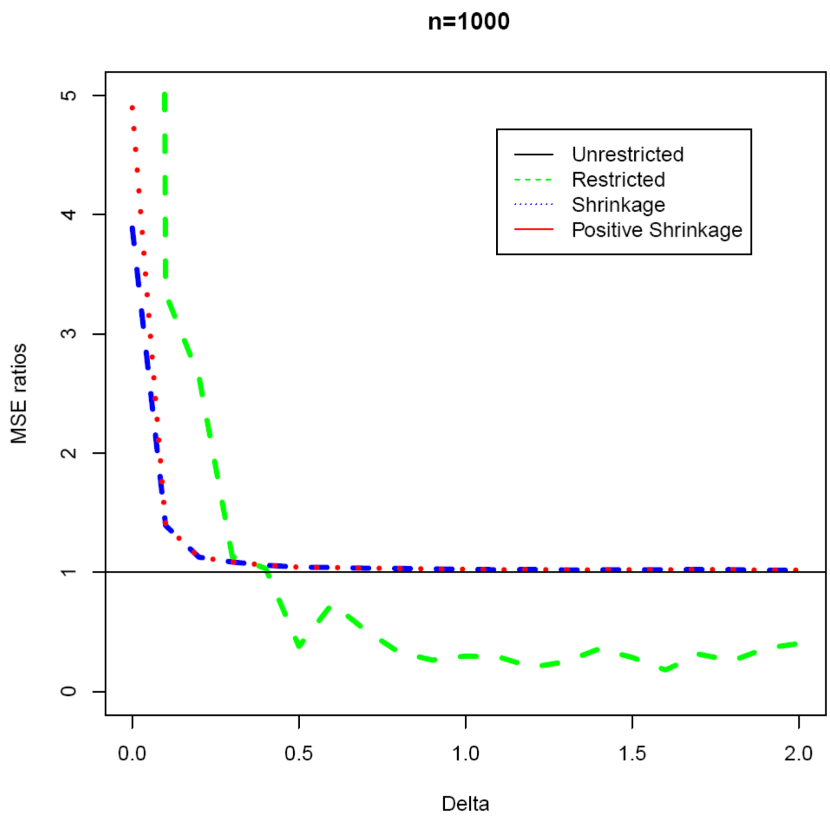


Figure 3.4: Ratios of MSE for all estimators relative to the unrestricted estimator with delta varying over the parameter space and sample size n=1000

3.1.1 Application to Real Data Sets

In 1974, the Mayo Clinic began a ten-year trial in primary biliary cirrhosis (PBC) of the liver. This was a randomized placebo-controlled trial of the drug D-penicillamine, and there were 424 patients in total at the clinic who were eligible to participate. The data is mostly complete for the first 312 patients, however the last 112 did not participate in the clinical trial, consenting only to have measurements recorded and followed for survival. In addition, six of those cases were lost. This data set is fairly well-known and a nearly identical set appears in appendix D of Fleming and Harrington (1991).

While Cox's proportional model has been used for this data, the covariate values were all determined at the time when they entered the study. It is, however, quite possible that the values of these covariates would change over the duration of the trial. For this reason, it may be a good idea to consider a model which is time-dependent, such as Aalen's model.

There were seventeen covariates collected from the patients who participated in the clinical trial, including age, sex, blood clotting time and treatment, to name a few. Here, we fit Aalen's additive regression model to the PBC data using the R function *aalen*, found in the *timereg* package. We will fit the full model and use the supremum-test to determine which covariates are significant. Our restriction is given by

$R\beta(t) = r$, where R is the identity matrix, and $r = (0, ..., 0)'$.

Table 1: Supremum-test of significance

Covariate	P-Value ($H_0 : B(t) = 0$)
trt	0.298
age	0.037
sex	0.623
ascites	0.178
hepato	0.375
spiders	0.667
edema	0.190
log(bili)	0.000
log(albumin)	0.004
log(copper)	0.068
log(alk.phos)	0.374
log(ast)	0.315
log(trig)	0.536
log(platelet)	0.602
protime	0.195
stage	0.573

We can see that the drug D-penicillamine (covariate trt) did not have a significant impact on time of death in this trial. Next, we use the Kolmogorov-Smirnov test to determine which effects appear to be time-invariant.

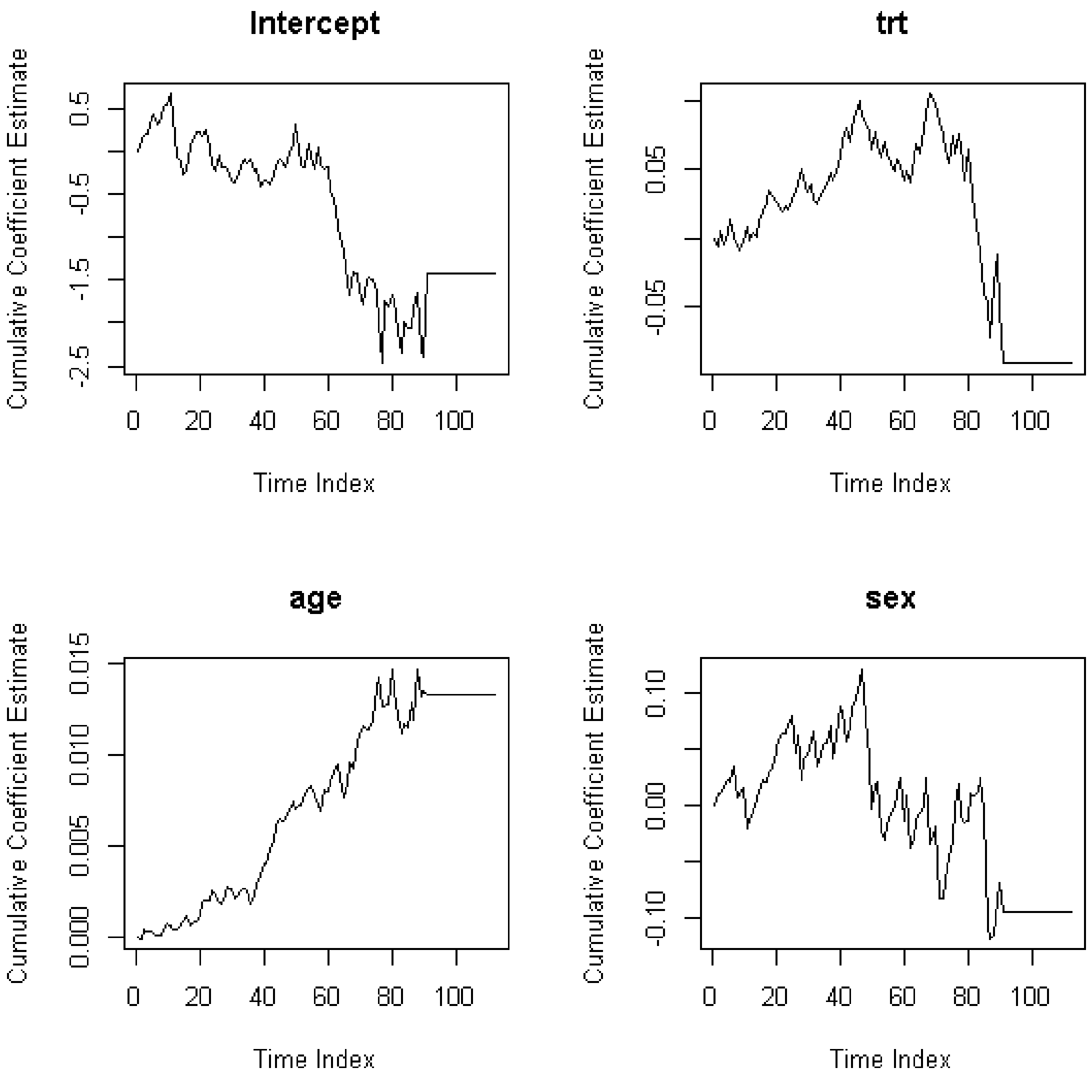
Table 2: Kolmogorov-Smirnov test

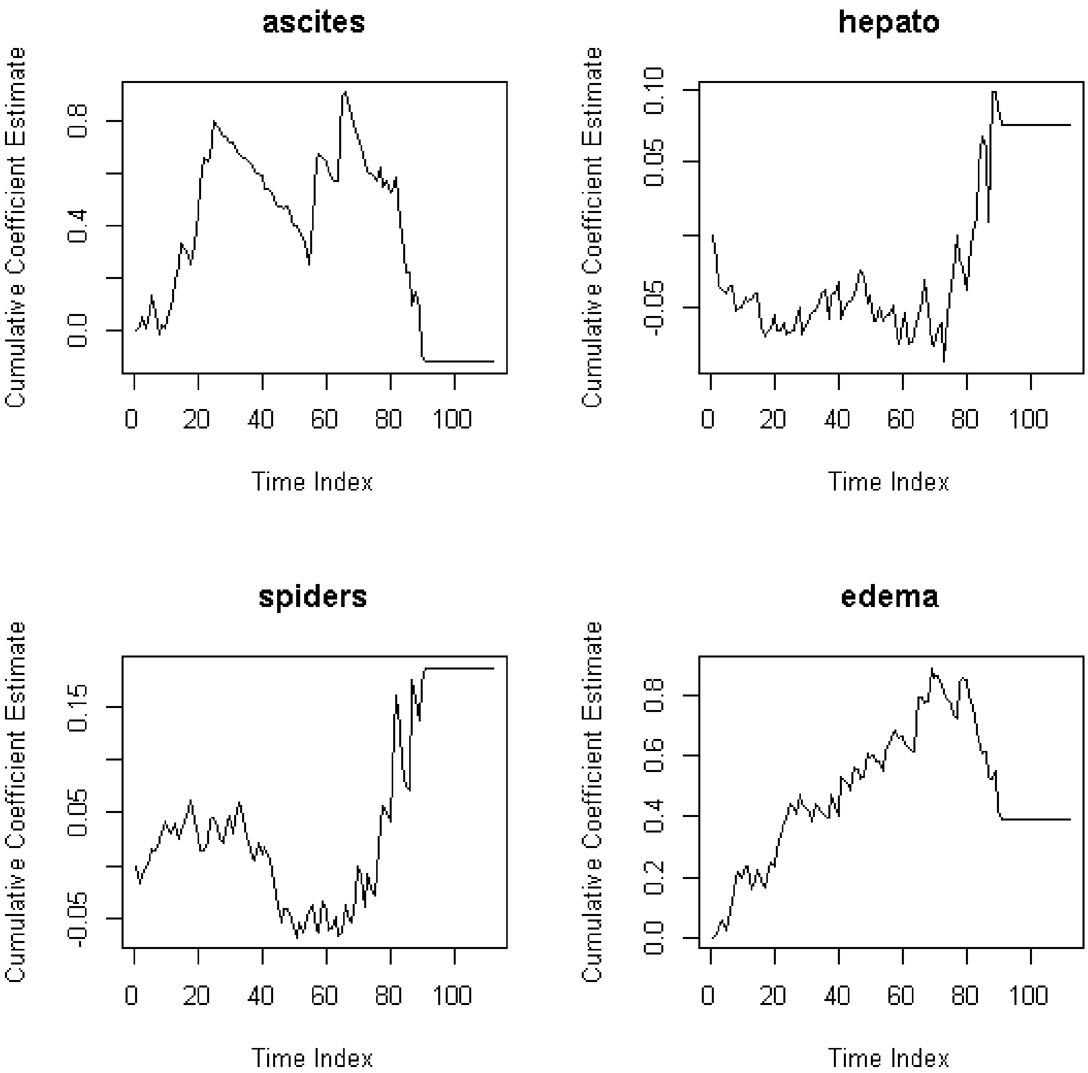
Covariate	P-Value ($H_0 : B(t) =$ constant effect)
trt	0.071
age	0.045
sex	0.324
ascites	0.110
hepato	0.302
spiders	0.673
edema	0.040
log(bili)	0.122
log(albumin)	0.253
log(copper)	0.012
log(alk.phos)	0.062
log(ast)	0.340
log(trig)	0.455
log(platelet)	0.138
protime	0.271
stage	0.381

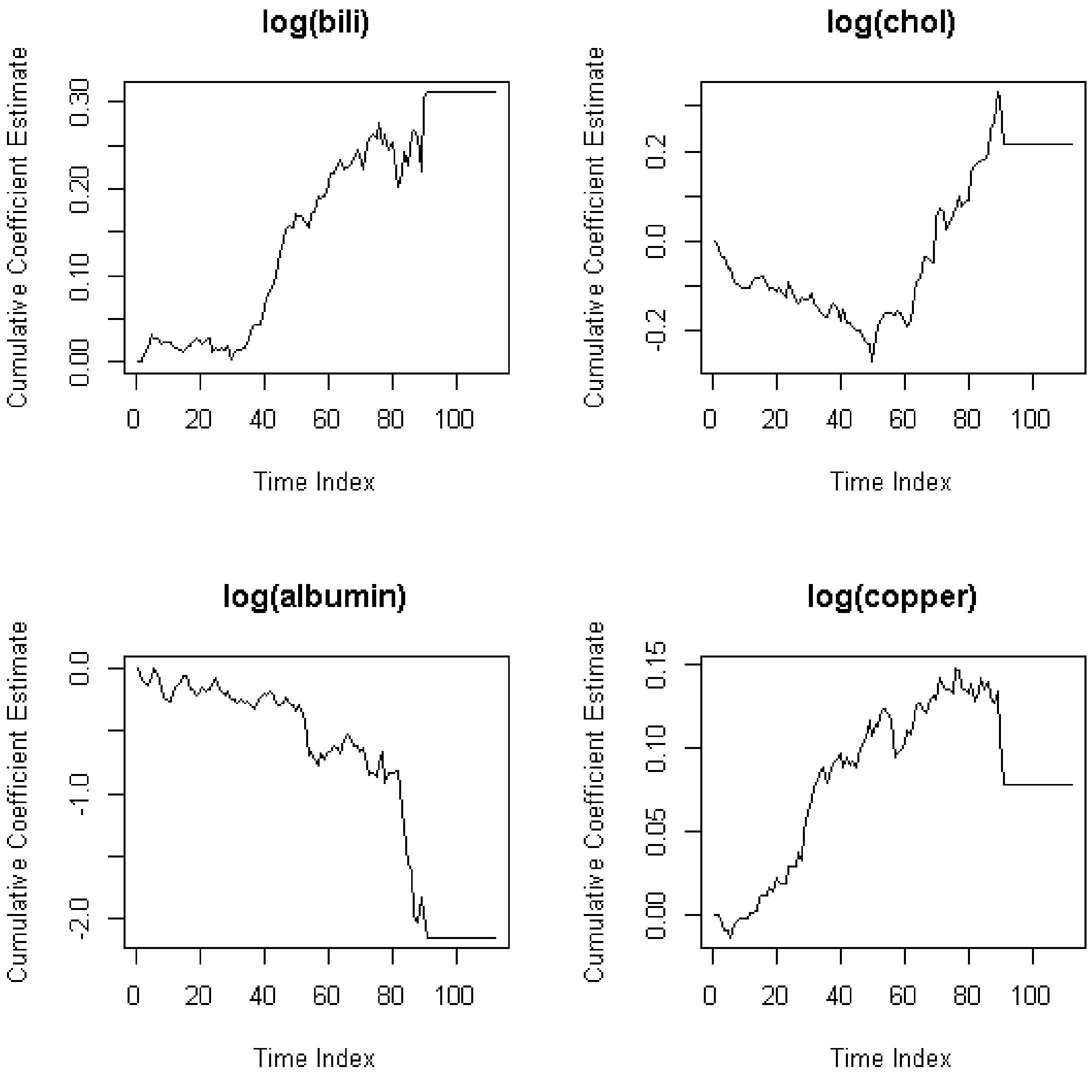
Note that very few of the p-values are significant. Therefore, in this model of the

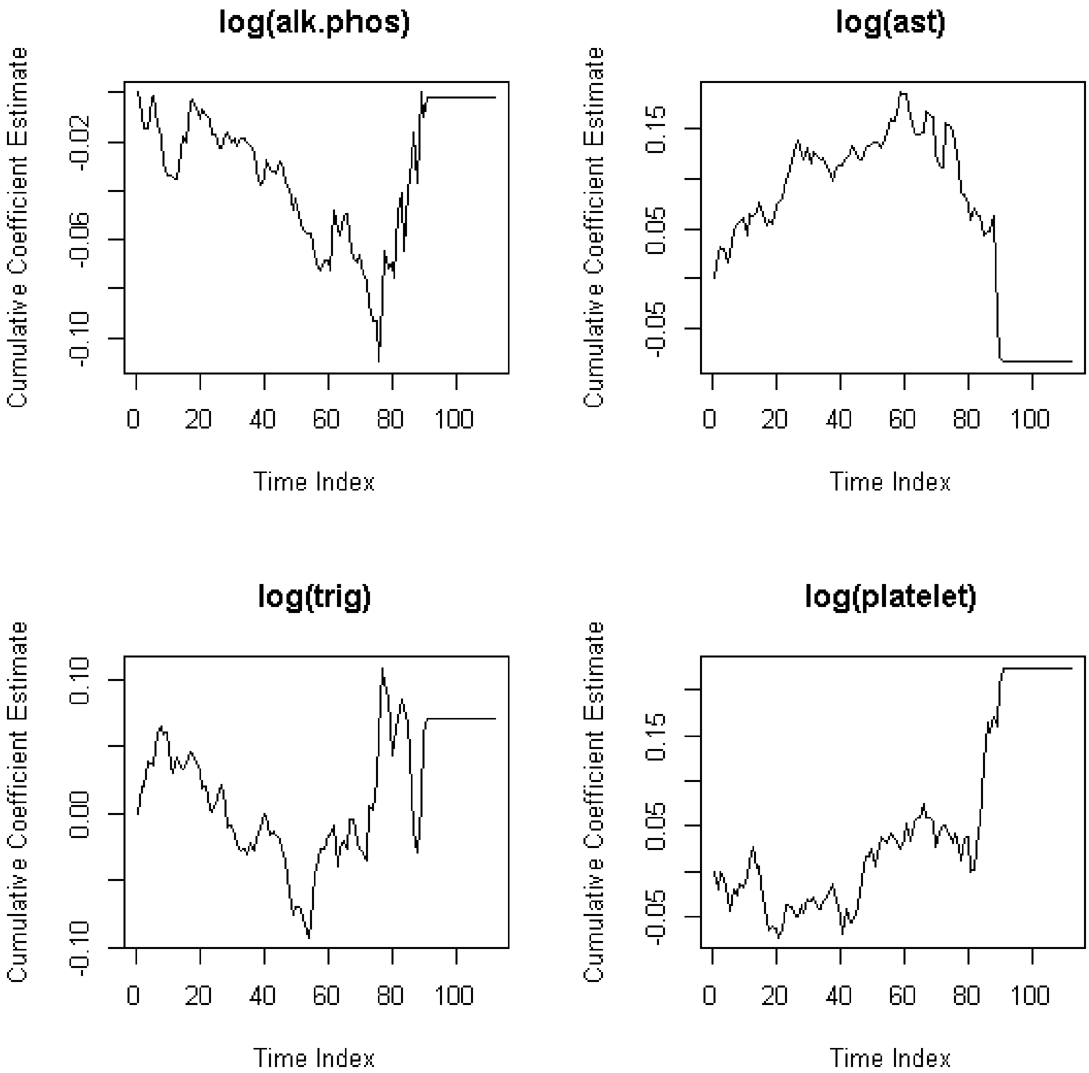
PBC data, most of the $B(t)$ appear to be time-independent. The covariates that *do* appear to be time-dependent are: age, edema and copper.

In addition, R will estimate the cumulative coefficient estimates for each of the covariates in our model, at 112 different points in time. These are given in the following plots.




ascites
hepato
spiders
edema
Cumulative Coefficient Estimate
Time Index
0 20 40 60 80 100
0.0 0.4 0.8
-0.05 0.05 0.10
-0.05 0.05 0.15
0.0 0.2 0.4 0.6 0.8


log(bili)
Cumulative Coefficient Estimate
0.00
0.10
0.20
0.30
0
20
40
60
80
100
Time Index
log(chol)
Cumulative Coefficient Estimate
-0.2
0.0
0.2
0
20
40
60
80
100
Time Index
log(albumin)
Cumulative Coefficient Estimate
-2.0
-1.0
0.0
0
20
40
60
80
100
Time Index
log(copper)
Cumulative Coefficient Estimate
0.00
0.05
0.10
0.15
0
20
40
60
80
100
Time Index



Using Table 3.1.1, we can reduce our model to include only the significant covariates, i.e. age, log(bili) and log(albumin).

To calculate the estimators, we considered the full model. We computed the UE, RE, shrinkage and positive-part shrinkage estimators for the coefficient of each covariate. Here we provide the plots of all estimators for only the three significant covariates, and the intercept.

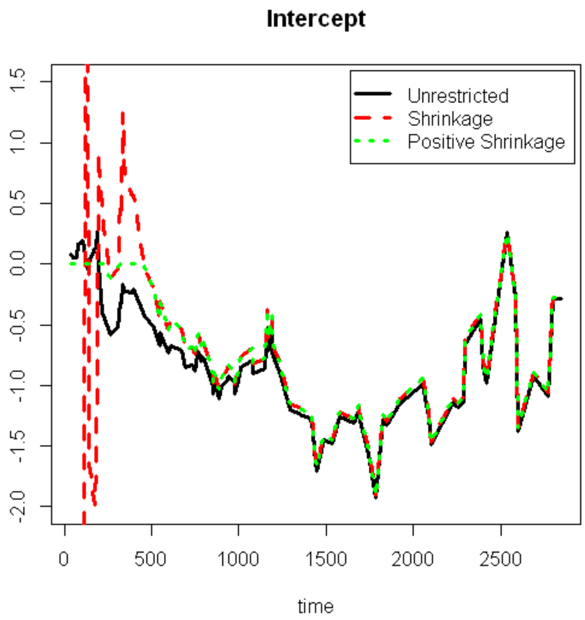


Figure 3.5: Intercept estimates for all estimators over the 95 unique event times

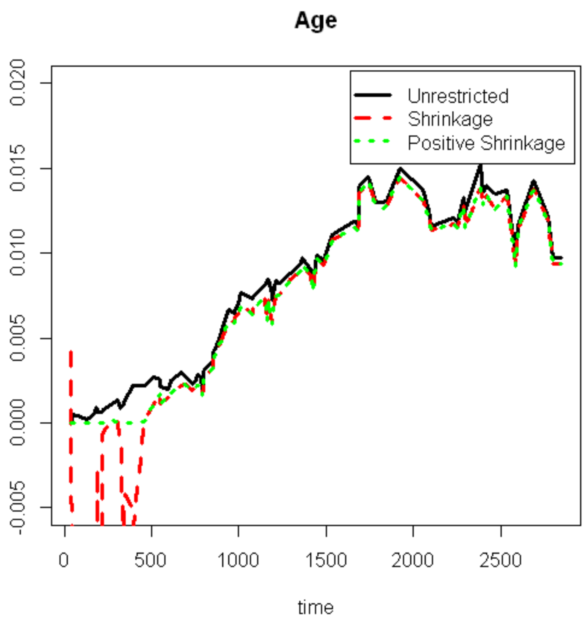


Figure 3.6: Coefficient estimates for the age covariate for all estimators over the 95 unique event times

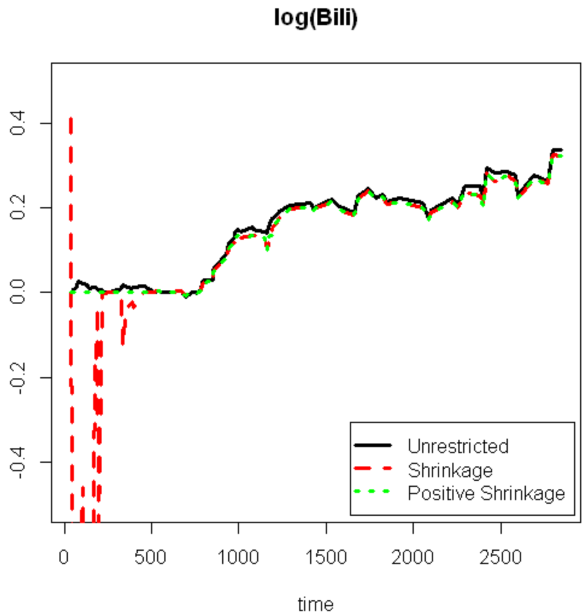


Figure 3.7: Coefficient estimates for the log(bili) covariate for all estimators over the 95 unique event times

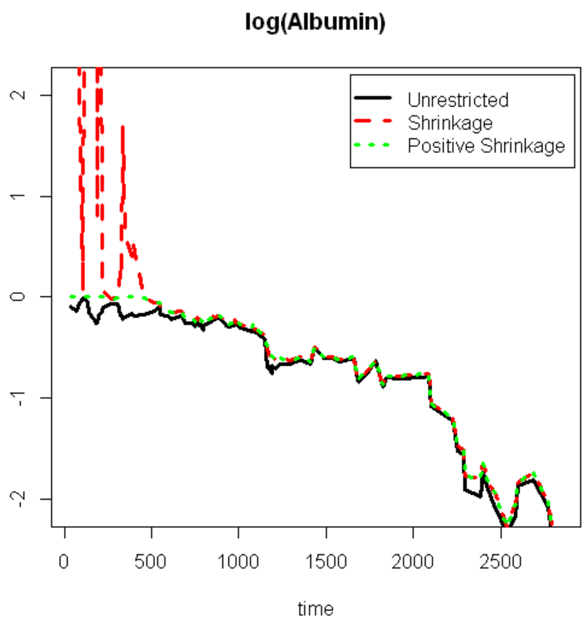


Figure 3.8: Coefficient estimates for the log(albumin) covariate for all estimators over the 95 unique event times

Chapter 4

Conclusion

This paper proposes a shrinkage estimator and positive-part shrinkage estimator for the nonparametric regression coefficients in Aalen's additive model, under a general linear hypothesis involving the coefficients. Using some martingale theory, we have shown that the unrestricted and restricted estimators for the cumulative regression coefficients are asymptotically normal. We then compared the estimators analytically, by calculating their risks and biases, considering a quadratic loss function.

To conclude the paper, we evaluated their performance in a Monte Carlo simulation study, and then applied the model to a data set on the survival times of patients with primary billiary cirrhosis. It was clear from the simulation that the proposed shrinkage estimators outperformed the usual restricted and unrestricted estimators on the parameter space, provided that the sample size was large ($n \geq 1000$).

Appendix A

A.1 Martingale Theory

This appendix provides a bit of a background on the martingale and counting process theory that is used in the estimation of risk coefficients for the additive hazard models. We begin with a few elementary definitions and propositions.

The "norm" of a matrix can be defined a number of different ways. Wherever the "norm" of a matrix appears throughout *this* paper, we refer to the following definition.

Definition A.1.1. *For a matrix, A, let the "norm" of A be defined by*

$$|A| = \sqrt{\lambda(AA')}$$

where λ is the largest eigenvalue of the matrix (AA') .

The following proposition is used to show that the matrix $R[X(t)W(t)X'(t)]^{-1}R'$ is invertible, in order to derive the RE in section 2.0.1 .

Proposition A.1.2. *Let the $n \times n$ matrix A be positive definite and let B be an $m \times n$ matrix. If B is full-rank, then the matrix BAB' is positive definite.*

Proof. Let A be an $n \times n$ positive definite matrix, and B be any full-rank, $m \times n$ matrix.

Since A is positive definite, $z'Az > 0 \;\; \forall \; z \in \mathbb{R}^n \;\; \ni \;\; z \neq 0$.

Then $z'BAB'z = (B'z)'A(B'z) \geq 0$.

Also, $(B'z)'A(B'z) = 0 \;\Leftrightarrow\; B'z = 0$. Thus $z = 0$, since B is full-rank.

Therefore BAB' is an $m \times m$ matrix such that, $z'(BAB')z = w'Aw > 0$ for all $z \in \mathbb{R}^m$ such that $z \neq 0$)

i.e. BAB' is positive-definite. □

Martingale theory is an important part of the foundation behind the estimators developed in this paper. The ability to write the cumulative coefficients in terms of martingales, follows from the Doob-Meyer Decomposition Theorem, which uses the notion of integrable and uniformly integrable processes. For this reason, we will define these concepts. For further details about these concepts, the reader is referred to Fleming & Harrington (1991, Chapters 1 & 2), Rao (1969) and Lipster & Shiryayev (1989, Chapter 5).

Definition A.1.3. *Let C be a non-empty set of elements, and let A be a collection of subsets of C. A is a "σ-field" (or σ-algebra) if*

1.) If $E \in A$, then $E^c \in A$ (A is closed under complements), and

2.) If the sequence of sets $\{E_1, E_2, ...\}$ is in A, then $\bigcup_{j=1}^{\infty} E_j \, \epsilon \, A$ (A is closed under countable unions).

This definition can be found in Fleming & Harrington (1991, p. 322).

When introducing a martingale, it is important to state that it is a martingale with respect to its filtration. We define this concept below.

Definition A.1.4. *(Fleming & Harrington, 1991, p. 17)*
A "filtration" is a family, $\mathcal{F} = \{\mathcal{F}_t, t \in T\}$, of sub-$\sigma$-fields of C such that $\forall s, t \, \epsilon \, T$

$$s < t \Rightarrow \; \mathcal{F}_s \subset \mathcal{F}_t.$$

▷ *Notation* : Let $\mathcal{F}_{t^-}$ denote the filtration at an instant before t. This is the smallest σ-field containing $\bigcup_{s>t} \mathcal{F}_s$

▷ *Remark* : The filtration (or *history*) of a counting process is a right-continuous filtration. That is, $\bigcap_{s>t} \mathcal{F}_s = \mathcal{F}_t$. This property is important for many of the martingale theory proofs.

Definition A.1.5. *(Martinussen & Scheike, 2006, p. 19)*
A "stochastic process" is a family of random variables $X = \{X(t), \; t \geq 0\}$ indexed by time, all defined on the same probablility space $(\Omega, \mathcal{F}_t, P)$.

Definition A.1.6. *(Kalbfleisch & Prentice, 1980, p. 157)*
A stochastic process $X = \{X_t\ ,\ t \geq 0\}$ is "adapted to the filtration" $\mathcal{F}_t$ if

$$\forall\ t \geq 0,\ X_t \text{ is } \mathcal{F}_t - measurable.$$

In the first section of this paper, we impose a set of conditions, Condition $\mathcal{A}$, on our model. One condition in this set is that $\{X_i'(t)N_i(t),\ 0 \leq t \leq \tau\}$ is a separable process for $i = 1, ..., n$. To better explain this condition, we provide the following definition of a "separable process".

Definition A.1.7. *(Doob, 1953, p.51)*
Let $\{X_t, t \in T\}$ be a real stochastic process with linear parameter set T. Let $\mathcal{S}$ be a system of linear Borel sets. Then $\{X_t, t \in T\}$ is "separable" relative to $\mathcal{S}$ if there is a sequence $\{t_j\}$ of parameter values and an ω set, ω_1, of probability 0 such that if $S \in \mathcal{S}$ and I is any open interval, the ω sets

$$\{X_t(\omega), t \in IT\}\,,\ \left\{X_{t_j}(\omega) \in S, t_j \in IT\right\}$$

differ by, at most, a subset of ω_1.

The stochastic process $\{N_t, t \geq 0\}$ introduced in Chapter 1 is a *counting process.* This concept is defined below.

Definition A.1.8. *(Fleming & Harrington, 1991, p. 18)*
A "counting process" is a stochastic process $\{N(t), t \geq 0\}$ adapted to a filtration $\{\mathcal{F}_t : t \geq 0\}$ whose paths are right-continuous (almost surely), piecewise constant and whose discontinuities are jumps of magnitude 1, such that

(i) $N(0) = 0$

(ii) $P[N(t) < \infty] = 1$.

The process $\{M_t, t \geq 0\}$ that appears throughout this paper is a *martingale*. It is a counting process with some special properties which are useful in the derivation of our estimators, and it is defined in greater detail below.

Definition A.1.9. *(Kalbfleisch & Prentice, 1980, p. 157)*
Let $\{\mathcal{F}_t, t \geq 0\}$ be a filtration, and let $\{M_t, t \geq 0\}$ be a [continuous] real-valued stochastic process adapted to the filtration $\{\mathcal{F}_t, t \geq 0\}$.
$\{M_t, t \geq 0\}$ is a "martingale" with respect to $\{\mathcal{F}_t, t \geq 0\}$ if the following are satisfied:

(i) $\mathrm{E}[\ |M_t|\ ] < \infty\ ,\ t \geq 0$

(ii) $\mathrm{E}[M_t|\mathcal{F}_s] = M_s\ ,\ s \leq t$

or, equivalently, $\mathrm{E}[dM(t)|\mathcal{F}_{t^-}] = 0\ ,\ \forall\ t\ \in\ (0, \tau]$.

▷ <u>*Note*</u> : Given a process $\{M_t, t = 0, 1, 2, ...\}$, one can choose $\mathcal{F}_t = \sigma\{M_s, s = 0, 1, ..., t\}$, the σ -field generated by $\{M_s, s = 0, 1, ..., t\}$. This is called the "natural filtration" of the process $\{M_s, s = 0, 1, 2, ...\}$. In continuous time,

the natural filtration is $\mathcal{F}_t = \sigma\{M_s, s \leq t\}$.

For our purposes, let $N_i(t)$ be a counting process such that

$$N_i(t) = \begin{cases} 1, & \text{if } T_i \leq t \text{ and an event occurred at } T_i \ , \\ 0, & \text{otherwise} \ , \end{cases}$$

and let $N(t) = [N_1(t), ..., N_n(t)]'$.

We redefined $X(t)$ as an $n \times k$ matrix with the i^{th} row defined by:

$$[Y_i(t), Y_i(t)X_{i1}(t), ..., Y_i(t)X_{ik}(t)] \quad \text{for } i = 1, ..., n.$$

Following Aalen's additive model, the intensity function for $N_i(t)$ is

$$\begin{aligned} \lambda_i(t) &= [Y_i(t), Y_i(t)X_{i1}(t), ..., Y_i(t)X_{ik}(t)]\beta(t) \\ &= [Y_i(t), Y_i(t)X_{i1}(t), ..., Y_i(t)X_{ik}(t)][\beta_0(t), ..., \beta_k(t)]' \end{aligned}$$

Further, let $\Lambda(t)$ denote the cumulative intensity function for $N_i(t)$,

i.e.

$$\Lambda_i(t) = \int_0^t \lambda_i(s)ds,$$

for $0 \leq t \leq \tau$.

The following are definitions of types of processes that are used throughout this paper.

Definition A.1.10. *(Flemming & Harrington, 1991, p. 17)*
An "increasing process", $A(t)$, is a right-continuous process with non-decreasing sample paths such that $P[A(0) = 0] = 1$.

Definition A.1.11. *(Flemming & Harrington, 1991, p. 17)*
An increasing process, $A(t)$, is an "integrable process" if and only if

$$\sup\{\mathrm{E}[|A(t)|]\} < \infty.$$

Definition A.1.12. *(Klebaner, 2005, p. 185)*
An integrable process, $A(t)$ is "uniformly integrable" if

$$\sup \mathrm{E}[|A(t)|I_{\{|A(t)|>M\}}] \to 0 \quad \text{as} \quad M \to \infty.$$

Definition A.1.13. *The $\mathcal{F}_t$-process, $\{M_t\}_{t\geq 0}$, is a "local martingale" if there exists an increasing sequence of Markov stopping times (T_n) such that*

$$(i)\ T_n \xrightarrow[n\to\infty]{P} \infty$$

and $\quad (ii)$ For each $n \geq 1$, the stopped process $M_{T_n} I_{\{t\geq 0\}}$ is an $\mathcal{F}_t$-martingale.

This definition of a local martingale can be found in Borovskikh & Semenovi(1997, p. 6).

▷ <u>*Note*</u> *:* The sequence (T_n) is called a "*localizing sequence*".

Definition A.1.14. *An $\mathcal{F}_t$-martingale, $\{M_t\}_{t\geq 0}$, is "square integrable" (or to have "finite variance") if*

$$\sup_{t\geq 0} \mathrm{E}[M(t)^2] < \infty.$$

For further details, we refer the reader to Kalbfleisch & Prentice (1980, p. 158).

Definition A.1.15. *A stochastic process* $\{M_t, t \geq 0\}$ *is said to be a "local square integrable martingale" if it satisfies both definitions* $A.1.13$ *and* $A.1.14$*.*

Now we can proceed to analyze the properties of our cumulative intensity process $\Lambda(t)$. First, note that, as given in the following proposition, M is in fact a martingale.

Proposition A.1.16. *The stochastic process* $M(t) = N(t) - \Lambda(t)$ *is a martingale.*

This result is established in Fleming & Harrington (1991, Theorem 1.4.1, p. 37).

Proposition A.1.17. *The cumulative intensity process,* $\Lambda(t)$ *has the following property:*

$$\mathrm{E}[N(t)|\mathcal{F}_{t^-}] = \mathrm{E}[\Lambda(t)|\mathcal{F}_{t^-}] = \Lambda(t).$$

The proof of this proposition is given in Fleming & Harrington (1991, p. 62).

In the study of the optimality of the proposed estimators, we use large-sample theory. Indeed, the exact [finite-sample] distributions are impossible to obtain in this context. More precisely, the established results involve three different notions of convergence: *convergence in mean*, *convergence in probability*, *convergence in distribution* . For more details about these concepts, we refer the reader to Billingsley (1995), Taylor (1973, p. 166-182) and Whittle (1976, Chapter 16).

Further, for the convenience of the reader, we give below very brief definitions of these concepts.

Consider the sequence of real-valued random variables $\{X_n; n = 1, 2, ...\}$ and let X be a real-valued random variable.

Definition A.1.18. *(Billingsley, 1995, p. 70)*
X_n is said to "converge almost surely" (or with probability 1) to X iff

$$P\left[|X_n - X| \geq \varepsilon, \text{ infinitely often }\right] = 0.$$

This is denoted by $X_n \xrightarrow[n\to\infty]{a.s.} X$.

A well-known example of this mode of convergence is the Strong Law of Large Numbers (SLLN).

- **Example:**

Suppose $\{X_1, X_2, ..., X_n\}$ is a sequence of independent random variables, with

$$\mathrm{E}[X_1] = \mu, \text{ and } \mathrm{E}[|X_1|^4] < \infty.$$

Then $\overline{X}_n \xrightarrow[n\to\infty]{a.s.} \mu$.

This is Borel's version of the SLLN, and it can be found in Athreya & Lahiri (2006, p.40).

Definition A.1.19. *(Billingsley, 1995, p. 243)*

X_n is said to "converge in p^{th} mean" (or in $\mathcal{L}^p$ norm) to X if

$$\lim_{n\to\infty} \mathrm{E}[|X_n - X|^p] = 0.$$

In particular, if $p = 1$, X_n converges in mean to X if

$$\lim_{n\to\infty} \mathrm{E}[|X_n - X|] = 0.$$

- **Example:** Suppose $\{X_1, X_2, ..., X_n\}$ is a sequence of i.i.d. random variables, with

$$\mathrm{E}[X_1] = \mu, \text{ and } var[X_1] = \sigma^2 < \infty.$$

Then $\overline{X}_n$ converges in mean to μ, as n tends to ∞.

Definition A.1.20. *(Billingsley, 1995, p. 70)*

X_n is said to "converge in probability" to X if

$$\forall\ \varepsilon > 0\ , \quad \lim_{n\to\infty} P[|X_n - X| > \varepsilon] = 0.$$

This is denoted by $X_n \xrightarrow[n\to\infty]{P} X$.

A well-known example of convergence in probability is the Weak Law of Large Numbers.

Definition A.1.21. *(Billingsley, 1995, p. 329)*

Let $F_n(t) = P[X_n \leq t]$ and $F(t) = P[X \leq t]$.

Then X_n is said to "converge in distribution" iff

$$\lim_{n\to\infty} F_n(t) = F(t)$$

at all continuity points t of F. This is denoted by $X_n \xrightarrow{D} X$.
One can also say that F_n converges to F weakly.

The notion of the "predictable variation" of a stochastic process is defined below.

Definition A.1.22. *The "predictable variation process" of the $\mathcal{F}_t$-martingale $M(t)$ is a compensator for the process $M^2(t)$ and is denoted $\langle M\rangle(t)$.*

In particular, the predictable variation of a square-integrable $\mathcal{F}_t$-martingale M is given by

$$\langle M\rangle(t) = \int_0^t \mathrm{var}[dM(u)|\mathcal{F}_{u^-}].$$

For more details and applications of this concept, we refer the reader to Kalbfleisch & Prentice (1980, p. 158).

We also have the notion of "optional variation". This appears in the martingale central limit theorem, however we will not use this type of variation in any of our proofs.

Definition A.1.23. *The "optional variation process" of the $\mathcal{F}_t$-martingale $M(t)$ is a compensator of the process $M^2(t)$ and is denoted $[M](t)$.*
It is given by

$$[M] = \sum_{i=1}^{n} \{M(t_{i+1}) - M(t_i)\}^2$$

over increasingly fine partitions of the interva $[0, \tau]$.

This definition can be found in Therneau & Grambsch (2000, p.21).
The *martingale central limit theorem* will be the foundation of any proof regarding the asymptotic behaviour of martingales. There are various different versions and generalizations of this theorem, and we will be using that which closely resembles that given by Rebolledo.

Theorem A.1.24. *Martingale Central Limit Theorem*
Let $M^{(n)} = (M_1^{(n)}, ..., M_k^{(n)})$ *be a vector of local square integrable* $\mathcal{F}_t$*-martingales.*
Also, let $\mathcal{T}_0 \subseteq \mathcal{T}$ *and consider the conditions*
i) $\langle M^{(n)} \rangle(t) \xrightarrow{P} V(t)$ *for all* $t \in \mathcal{T}_0$ *as* $n \to \infty$
ii) $[M^{(n)}](t) \xrightarrow{P} V(t)$ *for all* $t \in \mathcal{T}_0$ *as* $n \to \infty$
iii) $\langle M_{\varepsilon h}^{(n)} \rangle(t) \xrightarrow{P} 0$ *for all* $t \in \mathcal{T}_0$*,* h *and* $\varepsilon > 0$ *as* $n \to \infty$*, where* $M_\varepsilon^{(n)}$ *is a vector of local square integrable* $\mathcal{F}_t$*-martingales which contains all the jumps of components of* $M^{(n)}$ *of absolute value greater than* ε*.*

If iii) and either i) or ii) are true, then

$(M^{(n)}(t_1), ..., M^{(n)}(t_l)) \xrightarrow{\mathcal{D}} (M^{(\infty)}(t_1), ..., M^{(\infty)}(t_l))$ *as* $n \to \infty$ *for all* $t_1, ..., t_l \in \mathcal{T}_0$
where $M^{(\infty)}$ *is a continuous Gaussian vector martingale with* $\langle M^{(\infty)} \rangle = [M^{(\infty)}] = V$ *for a continuous deterministic* $k \times k$ *positive semi-definite matrix-valued function on* $\mathcal{T}$*.*

In addition, if $\mathcal{T}_0$ *is dense in* $\mathcal{T}$ *(i.e. if the closure of* $\mathcal{T}_0$ *is* $\mathcal{T}$ *)*
and contains τ *if* $\tau \in \mathcal{T}$ *then, under the same conditions, then*

$M^{(n)} \xrightarrow{\mathcal{D}} M^{(\infty)}$ *in* $D(\mathcal{T})$ *as* $n \to \infty$.

This theorem is established in Andersen et. al. (1993, p. 83), as well as in Borovskikh & Semenovi (1997, Chapter 5).

It is well known that convergence in mean implies convergence in probability, and convergence in probability implies convergence in distribution. The converse is not, in general true. However, under certain conditions the converse *does* hold, and this is outlined in Lebesgue's Dominated Convergence Theorem below.

Theorem A.1.25. ***Lebesgue's Dominated Convergence Theorem (LDCT)*** *Let* $\{f_n\}$ *be a sequence of measurable functions such that* $f_n \to f$ *pointwise almost everywhere as* $n \to \infty$ *and, for an integrable function* g, $|f_n| \leq g$ *for all* n.
Then f *is integrable and*

$$\int f d\mu = \lim_{n\to\infty} \int f_n d\mu.$$

(Williams, 1991, p. 54)

Recalling the earlier definition of predictable variation, let us now consider the variance of such a process.

Proposition A.1.26. *Let* $M(t)$ *be an* $\mathcal{F}_t$*-martingale. Then the variance of* $dM(t)$*, conditional upon its history just prior to t, is*

$$var(dM(t)|\mathcal{F}_{t^-}) = d\langle M\rangle(t).$$

The proof of this proposition is given in Fleming & Harrington (1991, p. 40).

Appendix B

B.1 Lemmas, Proofs and Definitions

This appendix contains the lemmas and propositions that are referred to throughout this paper, as well as some necessary definitions. More importantly, this appendix contains proofs of the main propositions used in this paper. We will begin with Gill's Lemma (Gill, 1980, p. 98).

Lemma B.1.1. (Gill's Lemma)

Suppose that

1. $f_n(t) \xrightarrow[n\to\infty]{P} f(t)$ *for almost all* $t \in [0, \tau]$ *and* $\displaystyle\int_0^\tau |f(t)|dt < \infty.$

2. $\forall\ \delta > 0,\ \exists\ k_\delta(t)\ \geq\ 0$ with $\displaystyle\int_0^\tau k_\delta(t)dt < \infty$

such that

$$\liminf P[\ |f_n(t)| \leq k_\delta(t),\ \forall t\ \epsilon\ \ [0,\tau]\ ] \geq\ 1-\delta.$$

Here, each bound, $k_\delta(t)$*, is non-random. Then*

$$\int_0^t f_n(s)ds \ \xrightarrow[n\to\infty]{P} \int_0^t f(s)ds,$$

uniformly on $[0, \tau]$.

The proof of Gill's Lemma is found in Gill (1980, p. 98).

By using Gill's Lemma above, we establish the following lemma, which is useful in deriving the joint asymptotic normality of the UE and RE in Chapter 2.

Lemma B.1.2. *Suppose that*

$$\sup_{0 \leq t \leq \tau} |f_n(t) - f(t)| \xrightarrow[n \to \infty]{P} 0 \ , \quad \tau < \infty,$$

where

$$\int_0^\tau |f(t)| dt < \infty.$$

Then

$$\int_0^t f_n(s) ds \ \xrightarrow[n \to \infty]{P} \int_0^t f(s) ds$$

uniformly on $[0, \tau]$.

Proof. First, note that condition 1 of Gill's Lemma (Lemma B.1.1) is satisfied. Further, since $f_n(t) \xrightarrow[n \to \infty]{P} f(t)$ uniformly on $[0, \tau]$ we have, for all $\varepsilon > 0$,

$$\forall \ \delta > 0, \exists \ N_0 \in \mathbb{N} \ni (\ \forall \ n \geq N_0, \ P[\ |f_n(t) - f(t)| < \varepsilon] < \delta \ ) \, .$$

Then

$$\forall \ \varepsilon > 0, \forall \ \delta > 0, \exists \ N_0 \in \mathbb{N} \ni (\ \forall \ n \geq N_0, \ P[\ |f_n(t) - f(t)| < \varepsilon] \geq 1 - \delta \ ) \, ,$$

and thus

$$\forall \ \varepsilon > 0, \forall \ \delta > 0, \exists \ N_0 \in \mathbb{N} \ni (\ \forall \ n \geq N_0, \ P[\ \forall t \in [0, \tau], |f_n(t) - f(t)| < \varepsilon] \geq 1 - \delta \ ) \, .$$

We will now use this to show that condition 2 of Gill's Lemma is satisfied. Notice that the set

$$\begin{aligned}\{\forall\ t\epsilon[0,\tau], |f_n(t) - f(t)| < \varepsilon\} &\subset \{\forall\ t\epsilon[0,\tau]\ ||f_n(t)| - |f(t)|| < \varepsilon\} \\ &\subset \{|f_n(t)| < |f(t)| + \varepsilon, \forall\ t\epsilon[0,\tau]\}\,.\end{aligned}$$

This is because

$$\begin{aligned}\forall \varepsilon > 0, |f_n(t) - f(t)| < \varepsilon &\Rightarrow \forall \varepsilon > 0, ||f_n(t)| - |f(t)|| < \varepsilon \\ &\Rightarrow \forall \varepsilon > 0, |f_n(t)| < |f(t)| + \varepsilon.\end{aligned}$$

It follows that

$$\liminf_n P[|f_n(t)| < |f(t)| + \varepsilon, \forall\ t \in [0,\tau]\ ] \geq 1 - \delta,$$

with

$$\int_0^\tau (|f(t)| + \varepsilon)\ dt = \int_0^\tau |f(t)|dt\ +\ \varepsilon\tau\ < \infty,$$

since $\tau < \infty$.

Therefore, if we let $k_\delta(t) = |f(t)| + \varepsilon$, we have

$$\liminf_n P[|f_n(t)| < k_\delta(t), \forall\ t \in [0,\tau]\ ] \geq 1 - \delta$$

with

$$\int_0^\tau k_\delta(t)dt < \infty.$$

Thus, condition 2 of Lemma B.1.1 is satisfied. Therefore we can conclude that

$$\int_0^\tau f_n(t)dt \xrightarrow[n\to\infty]{P} \int_0^\tau f(t)dt$$

uniformly on $[0,\tau]$. $\square$

Lenglart's Inequality will be useful in the proofs of the joint asymptotic normality of the UE and the RE. It is outlined in the following proposition. For more details about this inequality, including the proof of this result, the reader is referred to Lenglart (1977).

Proposition B.1.3. *(Lenglart's Inequality) If M is a local square integrable $\mathcal{F}_t$-martingale, then*

$$P\left[\sup_{0\leq t\leq\tau}|M|>\eta\right]\leq\frac{\delta}{\eta^2}+P\left[\langle M\rangle(\tau)>\delta\right]$$

for any $\eta>0$ and $\delta>0$.

Before we can move on to prove Proposition B.1.5, we need the following lemma, which uses a stronger condition (and therefore a stronger result) than that given in Martinussen & Scheike (2006, p. 110).

Lemma B.1.4. *If Condition $\mathcal{A}$ holds, then there exist continuous functions $r_{2jp}(t)$ and $r_{3jpl}(t)$ such that*

$$\sup_{0\leq t\leq\tau}|n^{-1}\Gamma_{2jp}(t)-r_{2jp}(t)|\ \xrightarrow[n\to\infty]{a.s.}\ 0$$

and

$$\sup_{0\leq t\leq\tau}|n^{-1}\Gamma_{3jpl}(t)-r_{3jpl}(t)|\ \xrightarrow[n\to\infty]{a.s.}\ 0$$

for $j,p,l=1,...,k$.

Note that, as mentioned above, the stated result has stronger conclusion than that given in Martinussen & Scheike (2006, p. 110), for which only the convergence in probability is established. However, it should be noted that our Condition $\mathcal{A}$ is stronger than that in Martinussen & Scheike (2006, p. 110). The proof follows from the uniform strong law of large numbers. Also, for a similar proof, we refer the reader to Andersen & Gill (1982).

Proposition B.1.5. *Suppose that Condition $\mathcal{A}$ is satisfied. Then*

$$\int_0^t \sum_{l=1}^k \sum_{h=1}^k \left\{(n^{-1}\Gamma(s))^{-1} - \Omega(s)\right\}_{hi}^2 \times \left(n^{-1} \sum_{i=1}^n X_{il}^2(s) X_{ih}(s) W_i^2(s)\right) \beta_h(s) ds \xrightarrow[n\to\infty]{P} 0,$$

uniformly on $[0, \tau]$.

Proof. First, notice that, from Lemma B.1.4,

$$n^{-1}\Gamma(t) \xrightarrow[n\to\infty]{P} \Omega^{-1}(t)$$

uniformly on $[0, \tau]$. Then, using the fact that inversion of a matrix is a continuous operation,

$$(n^{-1}\Gamma(t))^{-1} \xrightarrow[n\to\infty]{P} \Omega(t)$$

uniformly on $[0, \tau]$.

Notice that

$$\begin{aligned} & \sum_{l=1}^k \sum_{h=1}^k \left\{(n^{-1}\Gamma(t))^{-1} - \Omega(t)\right\}_{hi}^2 \times \left(n^{-1} \sum_{i=1}^n X_{il}^2(t) X_{ih}(t) W_i^2(t)\right) \beta_h(t) \\ = & \sum_{l=1}^k \sum_{h=1}^k \left\{(n^{-1}\Gamma(t))^{-1} - \Omega(t)\right\}_{hi}^2 (n^{-1}\Gamma_{3llh}(t))\beta_h(t). \end{aligned}$$

Therefore, since $n^{-1}\Gamma_{3llh}(t)$ converges to $r_{3llh}(t)$ uniformly on $[0, \tau]$ by Lemma B.1.4,

$$\sum_{l=1}^k \sum_{h=1}^k \left\{(n^{-1}\Gamma(t))^{-1} - \Omega(t)\right\}_{hi}^2 (n^{-1}\Gamma_{3llh}(t))\beta_h(t) \xrightarrow[n\to\infty]{P} 0$$

uniformly on $[0,\tau]$,

i.e.

$$\sup_{0\le t\le\tau}\left|\sum_{l=1}^{k}\sum_{h=1}^{k}\left\{(n^{-1}\Gamma(t))^{-1}-\Omega(t)\right\}_{hi}^{2}(n^{-1}\Gamma_{3llh}(t))\beta_h(t)\right| \xrightarrow[n\to\infty]{P} 0.$$

Therefore, by Lemma B.1.2,

$$\int_0^t\sum_{l=1}^{k}\sum_{h=1}^{k}\left\{(n^{-1}\Gamma(s))^{-1}-\Omega(s)\right\}_{hi}^{2}\times\left(n^{-1}\sum_{i=1}^{n}X_{il}^2(s)X_{ih}(s)W_i^2(s)\right)\beta_h(s)ds \xrightarrow[n\to\infty]{P} 0,$$

uniformly on $[0,\tau]$, which is the desired result. □

Lemma B.1.6. *If the sequence* $\{H_n\}_{n=1}^{\infty}$ *is a non-negative r.v. and*

$$H_n \xrightarrow[n\to\infty]{P} 0$$

then, $\forall\ \varepsilon>0$,

$$nH_n\ I_{\{H_n>\varepsilon\}} \xrightarrow[n\to\infty]{P} 0.$$

Proof. First, since

$$H_n \xrightarrow[n\to\infty]{P} 0,$$

we have $\forall\ \varepsilon>0$,

$$\lim_{n\to\infty}P[\ |H_n|>\varepsilon\ ]=0.$$

Let $\delta>0$. Now,

$$\begin{aligned}P[\ |nH_nI_{\{H_n>\varepsilon\}}|>\delta\ ] &= P[\ |nH_nI_{\{H_n>\varepsilon\}}|>\delta\ ,H_n>\varepsilon\ ]\\ &\quad+ P[\ |nH_nI_{\{H_n>\varepsilon\}}|>\delta\ ,H_n\le\varepsilon\ ]\\ &= P[\ |nH_n|>\delta\ ,\ H_n>\varepsilon\ ]+0\end{aligned}$$

$$= \quad P[\ nH_n > \delta\ ,\ H_n > \varepsilon\ ]$$

$$\leq \quad P[\ H_n > \varepsilon\ ].$$

Thus

$$\lim_{n\to\infty} P[\ |nH_n I_{\{H_n>\varepsilon\}}| > \delta\ ] \leq \lim_{n\to\infty} P[\ H_n > \varepsilon\ ] = 0.$$

Therefore, by definition of convergence in probability,

$$nH_n I_{\{H_n>\varepsilon\}} \xrightarrow[n\to\infty]{P} 0,$$

and this completes the proof. $\square$

Let $\{\mathcal{F}_t\ ,\ t \geq 0\}$ be the σ-field generated by the process $\{M(s), 0 \leq s \leq t\}$. Note that $\{\mathcal{F}_t\ ,\ t \geq 0\}$ is a filtration.

Also, let

$$Z_1(t) = n^{-\frac{1}{2}} \int_0^t \left(\left[\frac{\Gamma(s)}{n} \right]^{-1} - \Omega(s) \right) X(s)W(s)dM(s).$$

Combining Proposition B.1.3 and Proposition B.1.5, we prove the following proposition. The established result is useful in deriving the asymptotic normality of the UE. Also, the provided proof gives more detail than that given in Martinussen & Scheike (2006, p. 111).

Lemma B.1.7. *Suppose that Condition $\mathcal{A}$ is satisfied.*

Then $Z_1(t)$ is a local $\mathcal{F}_t$-martingale and

$$Z_1(t) \xrightarrow[n\to\infty]{P} 0$$

uniformly on $[0, \tau]$.

Proof. The j^{th} component of $Z_1(t)$ is given by:

$$Z_{1j}(t) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^t V_{ji}(s) dM_i(s)$$

where

$$V_{ji}(t) = \sum_{l=1}^{k} \left\{ (n^{-1}\Gamma(t))^{-1} - \Omega(t) \right\}_{jl} X_{il}(t) W_i(t). \tag{B.1}$$

Also, the integral, $Z_1(t)$, has predictable variation given by

$$\langle Z_1 \rangle(\tau) = n^{-1} \int_0^\tau V(t) \; diag(\lambda(t)) V'(t) dt$$

where $V(t)$ is the matrix with $(j,i)^{th}$ component given in equation (B.1) . Then the predictable variation of the j^{th} component of Z_1 is

$$\langle Z_{1j} \rangle(\tau) = n^{-1} \sum_{i=1}^{n} \int_0^\tau V_{ji}^2(t) \lambda_i(t) dt.$$

Also, since $M(t)$ is a local square integrable $\mathcal{F}_t$-martingale, $Z_{1j}(t)$ is also locally square integrable because

$$Z_{1j}(t) = n^{-\frac{1}{2}} \sum_{i=1}^{n} \int_0^t H_i(s) dM_i(s)$$

where $H_i(t) = V_{ji}(t)$, which is locally bounded.

Further, by Cauchy-Shwarz inequality,

$$\langle Z_{1j} \rangle(\tau) = n^{-1} \sum_{i=1}^{n} \int_0^\tau V_{ji}^2(t) \lambda_i(t) dt \le \int_0^\tau G_j(t) dt,$$

where

$$G_j(t) = \sum_{l=1}^{k} \sum_{h=1}^{k} \left\{ (n^{-1}\Gamma(t))^{-1} - \Omega(t) \right\}_{jl}^2 \times \left(n^{-1} \sum_{i=1}^{n} X_{il}^2(t) X_{ih}(t) W_i^2(t) \right) \beta_h(t).$$

Now, by Proposition B.1.5,

$$\int_0^\tau G_j(t)dt \xrightarrow[n\to\infty]{P} 0.$$

Then $\langle Z_{1j}\rangle \xrightarrow[n\to\infty]{P} 0$, and thus, by Lenglart's inequality (Proposition B.1.3),

$$\sup_{0\le t\le\tau} |Z_{1j}(t)| \xrightarrow[n\to\infty]{P} 0,$$

i.e. $Z_{1j}(t)$ converges in probability to 0, uniformly on $[0,\tau]$. □

Proposition B.1.8. *Let $A_n(t)$ be defined as in* (2.4)*, let $A(t)$ be defined as in* (B.7)*, and suppose that the conditions in Lemma B.1.4 hold. Then*

$$A_n(t) \xrightarrow[n\to\infty]{P} A(t)$$

uniformly on $[0,\tau]$.

Proof. We have $A_n(t) = [X(t)W(t)X'(t)]^{-1}R'[R[X(t)W(t)X'(t)]^{-1}R']^{-1}$, so

$$\begin{aligned} A_n &= \left[\frac{X(t)W(t)X'(t)}{n}\right]^{-1} \frac{R'[R[X(t)W(t)X'(t)]^{-1}R']^{-1}}{n} \\ &= \left[\left(\frac{\Gamma(t)}{n}\right)^{-1} R' \left[R\left(\frac{\Gamma(t)}{n}\right)^{-1} R'\right]^{-1}\right]. \end{aligned}$$

By Lemma B.1.4,

$$\frac{\Gamma(t)}{n} \xrightarrow[n\to\infty]{P} \Omega(t)\ ,\ \textit{uniformly on } [0,\tau].$$

Further, applying Slutsky's Theorem, we get

$$A_n(t) \xrightarrow[n\to\infty]{P} \Omega(t)R'[R\Omega(t)R']^{-1}. \tag{B.2}$$

Since this convergence follows directly from the fact that R is a full rank matrix, along with the fact that

$$\left(\frac{X(t)W(t)X'(t)}{n}\right)^{-1} \xrightarrow[n\to\infty]{P} \Omega(t)$$

uniformly on $[0,\tau]$, we conclude that the convergence in (B.2) is uniform on $[0,\tau]$, i.e.

$$\lim_{n\to\infty} A_n(t) \xrightarrow[n\to\infty]{P} A(t)$$

uniformly on $[0,\tau]$. □

Lemma B.1.9. *Let* $\tilde{Q}_{ji}(t) = \sum_{l=1}^{k} \{(I - A_n(t)R)\Omega(t)\}_{jl} W_i(t)X_{il}(t)$ *and assume that Condition* $\mathcal{A}$ *holds. Then*

$$\sup_{0\le t\le\tau} |n^{-\frac{1}{2}}\tilde{Q}_{ji}(t)| \le \tilde{H}_n$$

where $\tilde{H}_n$ *converges in probability to* 0 *as* n *tends to* ∞*. In particular,*

$$\sup_{0\le t\le\tau} \left|n^{-\frac{1}{2}}\tilde{Q}_{ji}(t)\right| \xrightarrow[n\to\infty]{P} 0.$$

Proof.

$$\begin{aligned}
\tilde{Q}_{ji}(t) &= \sum_{l=1}^{k} \{(I - A_n(t)R)\Omega(t)\}_{jl} W_i(t)X_{il}(t) \\
&= \sum_{l=1}^{k}\sum_{h=1}^{k} \{I - A_n(t)R\}_{jh} \{\Omega(t)\}_{hl} W_i(t)X_{il}(t) \\
&= \sum_{h=1}^{k} \{I - A_n(t)R\}_{jh} \left(\sum_{l=1}^{k} \{\Omega(t)\}_{hl} W_i(t)X_{il}(t)\right) \\
&= \sum_{h=1}^{k} \{I - A_n(t)R\}_{jh} \left(\tilde{V}_{hi}(t)\right)
\end{aligned}$$

where

$$\tilde{V}_{hi}(t) = \sum_{l=1}^{k} \{\Omega(t)\}_{hl} W_i(t)X_{il}(t).$$

Then

$$\begin{aligned}\sup_{0\le t\le\tau}|n^{-\frac{1}{2}}\tilde{Q}_{ji}(t)| &= \sup_{0\le t\le\tau}\left|\sum_{h=1}^{k}\{I-A_n(t)R\}_{jh}\,(n^{-\frac{1}{2}}\tilde{V}_{hi}(t))\right| \\ &\le \sup_{0\le t\le\tau}\left|\sum_{h=1}^{k}\{I-A_n(t)R\}_{jh}\right|\sup_{0\le t\le\tau}\left|(n^{-\frac{1}{2}}\tilde{V}_{hi}(t))\right| .\end{aligned}$$

Then, using the result in Martinussen & Scheike (2006, p. 112),

$$\sup_{0\le t\le\tau}\left|n^{-\frac{1}{2}}\tilde{Q}_{ji}(t)\right| \le \sup_{0\le t\le\tau}\left|\sum_{h=1}^{k}\{I-A_n(t)R\}_{jh}\right| \times H_n, \tag{B.3}$$

where

$$H_n \equiv k\cdot\sup_{t,h,l}|(\Omega^{-1}(t))^{-1}_{hl}|\; n^{-\frac{1}{2}}\sup_{t,i,l}|W_i(t)X_{il}(t)|.$$

Following the result in Martinussen & Scheike (2006, p. 112), we have

$$H_n \xrightarrow[n\to\infty]{P} 0. \tag{B.4}$$

Therefore, using Proposition B.1.8 and relation (B.4), we get

$$\tilde{H}_n \xrightarrow[n\to\infty]{P} 0 \tag{B.5}$$

with

$$\tilde{H}_n = \sum_{h=1}^{k}\sup_{0\le t\le\tau}\left|\{I-A_n(t)R\}_{jh}\right| \times H_n. \tag{B.6}$$

Therefore, combining relations (B.3) and (B.5), we get

$$\sup_{0\le t\le\tau}\left|n^{-\frac{1}{2}}\tilde{Q}_{ji}(t)\right| \xrightarrow[n\to\infty]{P} 0.$$

□

Proposition B.1.10. *Under Condition $\mathcal{A}$ and the local alternatives in* (2.9)*,* $P_{1,n}(t)$ *is a locally square integrable martingale.*

Proof. Recall that

$$P_{1,n}(t) = n^{-\frac{1}{2}} \int_0^t [I_k - A_n(s)R]\Omega(s)X(s)W(s)dM(s).$$

Then $P_1(t)$ is of the form

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

where

$$H(s) = n^{-\frac{1}{2}}[I_k - A_n(s)R]\Omega(s)X(s)W(s).$$

By Theorem 2.2.2 of Martinussen & Scheike (2006), since $H(s)$ is locally bounded and $\mathcal{F}_{s^{-1}}$-predictable, $P_1(t)$ is a local square integrable martingale.
Therefore $P_1(t)$ *is* a locally square integrable martingale. □

Proof of Proposition 2.0.1 . Recall that we developed the following restricted estimator for $b(t) = dB(t)$:

$$\hat{b}_R(t) = (I_k - A_n(t)R)\hat{b}_{LS}(t) + \ A_n(t)r(t),$$

where

$$A_n(s) = [X(s)W(s)X'(s)]^{-1}R'[R[X(s)W(s)X'(s)]^{-1}R']^{-1}.$$

Then a restricted estimator for $B(t)$ is simply

$$\widetilde{B}_R(t) = \int_0^t (I_k - A_n(s)R)\hat{b}_{LS}(s)ds + \int_0^t A_n(s)r(s)ds.$$

Therefore,

$$\widetilde{B}_R(t) = \int_0^t [(I_k - A_n(s)R)\hat{b}_{LS}(s)ds \;+\; A_n(s)r(s)]ds.$$

□

Proof of Proposition 2.0.3 . Using Proposition B.1.10 , we conclude that the predictable variation of $P_{1,n}(t)$ is :

$$\begin{aligned}
\langle P_{1,n}\rangle(\tau) &= n^{-1}\int_0^\tau [I - A_n(t)R]\Omega(t)X(t)W(t)diag(\lambda(t))W'(t)X'(t)\Omega[I - R'A'_n(t)]dt \\
&= \int_0^\tau [I - A_n(t)R]\Omega(t)\left(\frac{1}{n}\sum_{i=1}^n W_i^2(t)[X_i(t)X_i'(t)]\lambda_i(t)\right)\Omega[I - R'A'_n(t)]dt \\
&= \int_0^\tau \alpha_n(t)m_n(t)\alpha_n'(t)dt,
\end{aligned}$$

where

$$\alpha_n(t) = I - A_n(t)R$$

and

$$m_n(t) = \Omega(t)\left(\frac{1}{n}\sum_{i=1}^n W_i^2(t)[X_i(t)X_i'(t)]\lambda_i(t)\right)\Omega(t).$$

To simplify calculations, let

$$A(t) = \Omega(t)R'[R\Omega(t)R']^{-1}. \tag{B.7}$$

and let

$$\Phi^*(\tau) \;=\; \int_0^\tau [I - A(t)R]\Omega(t)\mathrm{E}[W_1^2(t)[X_1(t)X_1'(t)]X_1(t)\beta(t)]\Omega(t)[I - R'A'(t)]dt$$

$$= \int_0^\tau \alpha(t)m(t)\alpha'(t)dt,$$

where

$$\alpha(t) = I - A(t)R$$

and

$$m(t) = \Omega(t)\mathrm{E}[W_1^2(t)[X_1(t)X_1'(t)]X_1(t)\beta(t)]\Omega.$$

Combining Lemma B.1.4 and Proposition B.1.8, we note that $\alpha_n(t)$ converges in probability to $\alpha(t)$ and $m_n(t)$ converges in probability to $m(t)$, both uniformly on $[0, \tau]$,

$$\text{i.e.} \quad \sup_{0\le t\le\tau} |\alpha_n(t) - \alpha(t)| \xrightarrow[n\to\infty]{P} 0 \quad \text{and} \quad \sup_{0\le t\le\tau} |m_n(t) - m(t)| \xrightarrow[n\to\infty]{P} 0.$$

Then

$$\sup_{0\le t\le\tau} |\alpha_n(t)m_n(t)\alpha_n'(t) - \alpha(t)m(t)\alpha'(t)| \xrightarrow[n\to\infty]{P} 0.$$

Further, one can verify that

$$\int_0^\tau |\alpha(t)m(t)\alpha'(t)|\, dt < \infty.$$

Therefore, using Lemma B.1.2, we get

$$\sup_{0\le t\le\tau} |\langle P_{1,n}\rangle(t) - \Phi^*(t)| = \sup_{0\le t\le\tau} \left| \int_0^\tau \alpha_n(t)m_n(t)\alpha_n'(t) - \alpha(t)m(t)\alpha'(t) \right| dt \xrightarrow[n\to\infty]{P} 0.$$

Hence,

$$\langle P_{1,n}\rangle(t) \xrightarrow[n\to\infty]{P} \Phi^*(t), \tag{B.8}$$

uniformly on $[0,\tau]$.

Now consider the process containing all jumps with absolute value greater than ε, for $\varepsilon > 0$, of the j^{th} component of P_1:

$$P_{1j\varepsilon,n}(t) = \sum_{i=1}^{n}\int_0^t n^{-\frac{1}{2}}\sum_{l=1}^{k}\tilde{Q}_{ji}(s)\ I_{\left\{|n^{-\frac{1}{2}}\tilde{Q}_{ji}(s)|>\varepsilon\right\}}dM_i(s)$$

where

$$\tilde{Q}_{ji}(s) = \{[I - A_n(s)R]\Omega(s)\}_{jl}\, W_i(s)X_{il}(s).$$

The predictable variation of this process is

$$\langle P_{1j\varepsilon,n}\rangle(t) = \sum_{i=1}^{n}\int_0^t \left[n^{-\frac{1}{2}}\tilde{Q}_{ji}(s)\right]^2\ I_{\left\{|n^{-\frac{1}{2}}\tilde{Q}_{ji}(s)|>\varepsilon\right\}}\lambda_i(s)ds,$$

and then, using relations (B.3) and (B.6), we get

$$\langle P_{1j\varepsilon,n}\rangle(t) \le \sum_{i=1}^{n}\int_0^t \left[\tilde{H}_n\right]^2\ I_{\left\{|\tilde{H}_n|>\varepsilon\right\}}\lambda_i(s)ds,$$

where $\tilde{H}_n$ is defined as in (B.6) .

So,

$$\begin{aligned}
\langle P_{1j\varepsilon,n}\rangle(t) &\le \sum_{i=1}^{n}\tilde{H}_n^2 I_{\left\{\tilde{H}_n>\varepsilon\right\}}\int_0^t \lambda_i(s)ds \\
&\le n\tilde{H}_n^2 I_{\left\{\tilde{H}_n>\varepsilon\right\}}\left[\frac{1}{n}\int_0^t\sum_{i=1}^{n}X_i(s)\beta(s)ds\right] \\
&\le n\tilde{H}_n^2 I_{\left\{\tilde{H}_n>\varepsilon\right\}}\sup_{0\le t\le\tau}\left(\int_0^t \frac{1}{n}\sum_{i=1}^{n}X_i(s)\beta(s)ds\right).
\end{aligned} \tag{B.9}$$

Then,

$$\langle P_{1j\varepsilon,n}\rangle \;\leq\; n\tilde{H}_n^2 I_{\left\{\tilde{H}_n>\varepsilon\right\}}\tau \sup_{0\leq t\leq\tau}\left|\frac{1}{n}\sum_{i=1}^{n} X_i(t)\right| \sup_{0\leq t\leq\tau}|\beta(t)|.$$

Now, from relation (B.5), $\tilde{H}_n \xrightarrow[n\to\infty]{P} 0$. Then, by Lemma B.1.6

$$n\tilde{H}_n^2 I_{\left\{\tilde{H}_n>\varepsilon\right\}} \xrightarrow[n\to\infty]{P} 0. \tag{B.10}$$

Further, by the uniform strong law of large numbers,

$$\sup_{0\leq t\leq\tau}\left|\frac{1}{n}\sum_{i=1}^{n} X_i(t) - \mathrm{E}[X_1(t)]\right| \xrightarrow[n\to\infty]{a.s.} 0. \tag{B.11}$$

Therefore, combining (B.9) , (B.10) and (B.11), we get

$$\langle P_{1j\varepsilon,n}\rangle(t) \xrightarrow[n\to\infty]{a.s.} 0. \tag{B.12}$$

Finally, using relations (B.8) and (B.12) along with the Martingale Central Limit Theorem, we conclude that $P_{1,n}(t)$ converges in distribution to a Gaussian martingale, on $\mathcal{D}\left([0,\tau]\right)$, with covariance function $\Phi^*(t)$.

□

Proof of Proposition 2.0.4 . $P_{2,n}$ is given by

$$P_{2,n}(t) = \int_0^t [I - A_n(s)R]n^{-\frac{1}{2}}V(s)dM(s),$$

and as in the proof of Proposition B.1.10, $P_{2,n}$ is a locally square integrable $\mathcal{F}_t$-martingale whose predictable variation process is

$$\langle P_{2,n}\rangle(t) = n^{-1}\int_0^t [I - A_n(s)R]V(s)\; diag(\lambda_i(s))V'(s)[I - R'A_n'(s)]ds.$$

To simplify the proof, let us deal with $P_{2,n}$ component-wise. The j^{th} component of $\langle P_{2,n}\rangle(t)$ is

$$\langle P_{2j,n}\rangle(\tau) = n^{-1}\sum_{i=1}^{n}\int_0^{\tau} Q_{ji}^2(t)\lambda_i(t)dt,$$

where

$$\begin{aligned} Q_{ji}(t) &= \sum_{l=1}^{k}\left\{[I - A_n(t)R][(n^{-1}\Gamma(t))^{-1} - \Omega(t)\right\}_{jl} X_{il}(t)W_i(t) \\ &= \sum_{l=1}^{k}\left[\sum_{h=1}^{k}\{I - A_n(t)R\}_{jh}\left\{(n^{-1}\Gamma(t))^{-1} - \Omega(t)\right\}_{hl}\right] X_{il}(t)W_i(t). \end{aligned}$$

Then,

$$\langle P_{2j,n}\rangle(\tau) = n^{-1}\sum_{i=1}^{n}\int_0^{\tau} Q_{ji}^2(t)\lambda_i(t)dt$$

$$= \int_0^{\tau} n^{-1}\sum_{i=1}^{n}\left[\sum_{l=1}^{k}\sum_{h=1}^{k}\{I - A_n(t)R\}_{jh}\left\{(n^{-1}\Gamma(t))^{-1} - \Omega(t)\right\}_{hl} X_{il}(t)W_i(t)\right]^2 \lambda_i(t)dt.$$

Using the Cauchy-Schwarts Inequality, we get

$$\begin{aligned} \langle P_{2j,n}\rangle(\tau) \le \sum_{i=1}^{n}\int_0^{\tau} n^{-1}\sum_{l=1}^{k}\left[\sum_{h=1}^{k}\{I - A_n(t)R\}_{jh}\left\{(n^{-1}\Gamma(t))^{-1} - \Omega(t)\right\}_{hl}\right]^2 \\ \times \sum_{l=1}^{k}(X_{il}(t)W_i(t))^2\lambda_i(t)dt. \end{aligned}$$

Applying Cauchy-Schwartz again, we get that

$$\langle P_{2j,n}\rangle(\tau) \le \sum_{i=1}^{n}\int_0^{\tau} n^{-1}\sum_{l=1}^{k}\left[\sum_{h=1}^{k}\{I - A_n(t)R\}_{jh}^2\right]\left[\sum_{h=1}^{k}\left\{(n^{-1}\Gamma(t))^{-1} - \Omega(t)\right\}_{hl}^2\right]$$

$$\times \sum_{l=1}^{k}(X_{il}(t)W_i(t))^2\lambda_i(t)dt.$$

This gives,

$$\langle P_{2j,n}\rangle(\tau) \leq \sum_{i=1}^{n}\int_0^\tau n^{-1}\left[\sum_{h=1}^{k}\{I-A_n(t)R\}_{jh}^2\right]\left(\sum_{l=1}^{k}\sum_{h=1}^{k}\left\{(n^{-1}\Gamma(t))^{-1}-\Omega(t)\right\}_{hl}^2\right) \times \left(\sum_{l=1}^{k}\left(X_{il}(t)W_i(t)\right)^2\right)\lambda_i(t)dt.$$

Hence,

$$\langle P_{2j,n}\rangle(\tau) \leq \int_0^\tau n^{-1}\left[\sum_{h=1}^{k}\{I-A_n(t)R\}_{jh}^2\right] \times \left[\sum_{l=1}^{k}\sum_{h=1}^{k}\sum_{u=1}^{k}\left\{(n^{-1}\Gamma(t))^{-1}-\Omega(t)\right\}_{hu}^2\left(\sum_{i=1}^{n}(X_{il}(t)W_i(t))^2\lambda_i(t)\right)\right]dt,$$

and then,

$$\langle P_{2j,n}\rangle(\tau) \leq \sup_{0\leq t\leq\tau}\left(\sum_{h=1}^{k}\{I-A_n(t)R\}_{jh}^2\right)\left[\int_0^\tau\sum_{h=1}^{k}G_h(t)dt\right]$$

where

$$G_h(t) = \sum_{l=1}^{k}\sum_{u=1}^{k}\left\{(n^{-1}\Gamma(t))^{-1}-\Omega(t)\right\}_{hu}^2\left[n^{-1}\sum_{i=1}^{n}(X_{il}(t)W_i(t))^2\right]\lambda_i(t).$$

Thus

$$\langle P_{2j,n}\rangle(\tau) \leq \left[\sum_{h=1}^{k}\left(\sup_{0\leq t\leq\tau}|\{I-A_n(t)R\}_{jh}|\right)^2\right]\left[\int_0^\tau\sum_{h=1}^{k}G_h(t)dt\right]. \tag{B.13}$$

Note that this supremum is

$$\sup_{0\le t\le\tau}|I-A_n(t)R| = \sup_{0\le t\le\tau}|I-[X(t)X'(t)]^{-1}R'[R[X(t)X'(t)]^{-1}R']^{-1}R|$$

so the limit of this supremum is

$$1+\lim_{n\to\infty}\left[\sup_{0\le t\le\tau}|[X(t)X'(t)]^{-1}R'[R[X(t)X'(t)]^{-1}R']^{-1}R|\right]$$
$$\le\ 1+\sup_{0\le t\le\tau}|\Omega(t)R'[R\Omega(t)R']^{-1}R|<\infty.$$

This means that

$$\lim_{n\to\infty}\left(\sup_{0\le t\le\tau}|\left\{I-A_n(t)R\right\}_{jh}|\right)^2\ <\infty.$$

Therefore, by using Proposition B.1.8,

$$\lim_{n\to\infty}\sum_{h=1}^{k}\left[\left(\sup_{0\le t\le\tau}|\left\{I-A_n(t)R\right\}_{jh}|\right)^2\right]\ <\infty \tag{B.14}$$

since this is simply a finite sum of finite values.

In addition, using Proposition B.1.5, we conclude that

$$\int_0^\tau\sum_{h=1}^{k}G_h(t)dt\overset{P}{\to}0. \tag{B.15}$$

Hence, combining (B.13), (B.14) and (B.15), we get

$$\langle P_{2j,n}\rangle(\tau)\ \xrightarrow[n\to\infty]{P}\ 0.$$

Hence, using Lenglart's inequality (Proposition B.1.3), we have

$$\sup_{0 \leq t \leq \tau} |P_{2j}(t)| \xrightarrow[n \to \infty]{P} 0.$$

This means that $P_{2j}(t)$ converges in probability to 0, uniformly over $[0, \tau]$, for each $j = 1, ..., k$.

Therefore $P_2(t)$ converges in probability to 0, uniformly over $[0, \tau]$.

$\square$

Proof of Proposition 2.0.5 . Recall that $A_n(t)$ converges in probability to $A(t)$, uniformly on $[0, \tau]$ (by Proposition B.1.8). Therefore

$$\sup_{0 \leq t \leq \tau} |A_n(t)\delta_1(t) - A(t)\delta_1(t)| \xrightarrow[n \to \infty]{P} 0.$$

Further,

$$\int_0^\tau |A(t)\delta_1(t)| \, dt < \infty.$$

Then, by Lemma B.1.2,

$$\sup_{0 \leq t \leq \tau} \left| \int_0^t A_n(s)\delta_1(s)ds - \int_0^t \Omega(s)R'[R\Omega(s)R']^{-1}\delta_1(s)ds \right| \xrightarrow[n \to \infty]{P} 0,$$

uniformly on $[0, \tau]$.

Thus,

$$P_{3,n}(t) \xrightarrow[n \to \infty]{P} \int_0^t \Omega(s)R'[R\Omega(s)R']^{-1}\delta_1(s)ds,$$

uniformly on $[0, \tau]$.

$\square$

Proof of Proposition 2.0.6 . Notice that we can rewrite $(\xi_n'(t), \eta_n'(t))'$ in the following way:

$$(\xi_n'(t), \eta_n'(t))' = \sqrt{n}\int_0^t \begin{pmatrix} -A_n(s)R \\ I_k - A_n(s)R \end{pmatrix} \Omega(s)X(s)W(s)dM(s)$$
$$+ \sqrt{n}\int_0^t \begin{pmatrix} -A_n(s)R \\ I_k - A_n(s)R \end{pmatrix} \left\{(n^{-1}\Gamma(s))^{-1} - \Omega(s)\right\} X(s)W(s)dM(s)$$
$$+ \int_0^t \begin{pmatrix} I_k \\ I_k \end{pmatrix} A_n(s)\delta_1(s)ds.$$

Then by similar arguments to those used to show the asymptotic normality of $\sqrt{n}(\widetilde{B}_R(t) - B(t))$, (i.e. replacing $[I - A_n(t)R]$ with our new matrix) we get

$$(\xi_n'(t), \eta_n'(t))' \quad \xrightarrow[n\to\infty]{D} \quad (\xi'(t), \eta'(t))' \qquad \text{on } \mathcal{D}\left([0, \tau]\right)$$

where

$$(\xi'(t), \eta'(t))' \sim N_{2k}\left[\int_0^t \begin{pmatrix} I_k \\ I_k \end{pmatrix} \Omega(s)R'[R\Omega(s)R']^{-1}\delta_1(s)ds \ , \ \Phi^{**}(t)\right]$$

where

$$\Phi^{**}(t) = \int_0^t \begin{pmatrix} -A(s)R \\ I_k - A(s)R \end{pmatrix} \Omega(s)\mathrm{E}[W_1^2(s)[X_1(s)X_1'(s)]X_1(s)\beta(s)]\Omega(s) \begin{pmatrix} -R'A'(s) & I_k - R'A'(s) \end{pmatrix} ds$$

and $A(s) = \Omega(s)R'[R\Omega(s)R']^{-1}$ (by Proposition B.1.8) . This can be written as

$$\Phi^{**}(t) = \int_0^t \begin{pmatrix} A_{11}(s) & A_{12}(s) \\ A_{21}(s) & A_{22}(s) \end{pmatrix} ds.$$

□

Proof of Corollary 2.1.

Recall from Proposition 2.0.6 that the general form of Φ^{**} is

$$\Phi^{**}(t) = \int_0^t \begin{pmatrix} A_{11}(s) & A_{12}(s) \\ A_{21}(s) & A_{22}(s) \end{pmatrix} ds,$$

where

$$\begin{aligned} A_{11}(s) &= A(s)RC(s)R'A'(s), \\ A_{12}(s) &= -A(s)RC(s)[I_k - R'A'(s)], \\ A_{21}(s) &= [I_k - A(s)R]C(s)R'A'(s), \\ A_{22}(s) &= [I_k - A(s)R]C(s)[I_k - R'A'(s)] \end{aligned}$$

and

$$C(s) = \Omega(s)\mathrm{E}[W_1^2(s)[X_1(s)X_1'(s)]X_1(s)\beta(s)]\Omega(s).$$

Now, if we let

$$W_1(s) = \frac{1}{X_1(s)\beta(s)},$$

then

$$\begin{aligned} C(s) &= \Omega(s)\mathrm{E}\left[\frac{X_1(s)X_1'(s)}{X_1(s)\beta(s)}\right]\Omega(s) \\ &= \Omega(s) \end{aligned}$$

because $\Omega^{-1}(s)$ is simply $\mathrm{E}\left[\frac{X_1(s)X_1'(s)}{X_1(s)\beta(s)}\right]$.

Then, the covariance becomes

$$\int_0^t \begin{pmatrix} \kappa(s) & \kappa(s) - A(s)R\Omega(s) \\ [\kappa(s) - A(s)R\Omega(s)]' & \kappa(s) - A(s)R\Omega(s) - [A(s)R\Omega(s)]' + \Omega(s) \end{pmatrix}$$

where

$$\kappa(s) = A(s)R\Omega(s)R'A'(s).$$

Then,

$$\begin{aligned} A_{12}(s) &= \kappa(s) - A(s)R\Omega(s) \\ &= A(s)R\Omega(s)R'A'(s) - A(s)R\Omega(s) \\ &= \Omega(s)R'[R\Omega(s)R']^{-1}R\Omega(s)R'[R\Omega(s)R']^{-1}R\Omega(s) - \Omega(s)R'[R\Omega(s)R']^{-1}R\Omega(s) \\ &= \Omega(s)R'[R\Omega(s)R']^{-1}R\Omega(s) - \Omega(s)R'[R\Omega(s)R']^{-1}R\Omega(s) \\ &= 0. \end{aligned}$$

The other 3 values simplify in a similar way, and therefore the covariance becomes:

$$\Phi^{***} = \int_0^t \begin{pmatrix} \Omega(s)R'[R\Omega(s)R']^{-1}R\Omega(s) & 0 \\ 0 & \Omega(s) - \Omega(s)R'[R\Omega(s)R']^{-1}R\Omega'(s) \end{pmatrix} ds,$$

and this completes the proof.

□

In order to prove Proposition 2.2.1, and later Proposition 2.3.3, we will use a theorem from Nkurunziza (2012b). The following proposition shows that the conditions of that theorem are satisfied.

Proposition B.1.11. *Let $J(t)$ and $\Sigma_{11}(t)$ be matrices defined as in Section 2.2. Then*

1.) $\Sigma_{11}(t)J(t)$ *is an idempotent matrix,*

and

2.) $J(t)\Sigma_{11}(t)J(t) = J(t)$.

Proof. First, note that

$$\begin{aligned}
\Sigma_{11}(t)J(t) &= \left(\int_0^t \Omega(s)R'[R\Omega(s)R']^{-1}R\Omega(s)ds\right)R'\left(\int_0^t R\Omega(s)R'ds\right)^{-1}R \\
&= \left(\int_0^t \Omega(s)R'[R\Omega(s)R']^{-1}[R\Omega(s)R']ds\right)\left(\int_0^t R\Omega(s)R'ds\right)^{-1}R \\
&= \left(\int_0^t \Omega(s)R'ds\right)\left(\int_0^t R\Omega(s)R'ds\right)^{-1}R. \qquad \text{(B.16)}
\end{aligned}$$

Then

$$\begin{aligned}
&\Sigma_{11}(t)J(t)\Sigma_{11}(t)J(t) \\
&= \left(\int_0^t \Omega(s)R'ds\right)\left(\int_0^t R\Omega(s)R'ds\right)^{-1}R\left(\int_0^t \Omega(s)R'ds\right)\left(\int_0^t R\Omega(s)R'ds\right)^{-1}R \\
&= \left(\int_0^t \Omega(s)R'ds\right)\left(\int_0^t R\Omega(s)R'ds\right)^{-1}\left(\int_0^t R\Omega(s)R'ds\right)\left(\int_0^t R\Omega(s)R'ds\right)^{-1}R
\end{aligned}$$

$$= \left(\int_0^t \Omega(s)R'ds\right)\left(\int_0^t R\Omega(s)R'ds\right)^{-1} R.$$

Therefore, by relation B.16,

$$\Sigma_{11}(t)J(t)\Sigma_{11}(t)J(t) = \Sigma_{11}(t)J(t).$$

In addition,

$$J(t)\Sigma_{11}(t)J(t) = J(t)\left[\left(\int_0^t \Omega(s)R'ds\right)\left(\int_0^t R\Omega(s)R'ds\right)^{-1} R\right],$$

by relation (B.16). Then,

$$\begin{aligned}
& J(t)\Sigma_{11}(t)J(t) \\
= \; & R'\left(\int_0^t R\Omega(s)R'ds\right)^{-1} R\left(\int_0^t \Omega(s)R'ds\right)\left(\int_0^t R\Omega(s)R'ds\right)^{-1} R \\
= \; & R'\left(\int_0^t R\Omega(s)R'ds\right)^{-1}\left(\int_0^t R\Omega(s)R'ds\right)\left(\int_0^t R\Omega(s)R'ds\right)^{-1} R \\
= \; & R'\left(\int_0^t R\Omega(s)R'ds\right)^{-1} R \\
= \; & J(t),
\end{aligned}$$

and this ends the proof. □

The proposition which is proven below is important for determining the ADR of each of the proposed estimators.

Proof of Proposition 2.2.1. By Slutsky's Theorem

$$\xi_n'(t)J(t)\xi_n(t) \xrightarrow[n\to\infty]{D} \xi'(t)J(t)\xi(t).$$

It suffices to prove that

$$\xi'(t)J(t)\xi(t) \sim \chi_q^2(\Delta(t)).$$

We will do this using Theorem 4 of Styan (1970), so we must show that the four conditions of that theorem are satisfied.

By Proposition B.1.11, $\Sigma_{11}(t)J(t)$ is idempotent. Therefore,

$$\Sigma_{11}(t)J(t)\Sigma_{11}(t)J(t) = \Sigma_{11}(t)J(t),$$

and so

$$\Sigma_{11}(t)J(t)\Sigma_{11}(t)J(t)\Sigma_{11}(t) = \Sigma_{11}(t)J(t)\Sigma_{11}(t).$$

Thus, the first condition is satisfied.

Now we need to show that $\text{trace}(J(t)\Sigma_{11}(t)) = q = \text{rank}(\Sigma_{11}(t)J(t)\Sigma_{11}(t))$. First,

$$\begin{aligned}
\text{trace}\left(J(t)\Sigma_{11}(t)\right) &= \text{trace}\left(\Sigma_{11}(t)J(t)\right) \\
&= \text{trace}\left(\left(\int_0^t \Omega(s)R'ds\right)\left(\int_0^t R\Omega(s)R'ds\right)^{-1} R\right) \\
&= \text{trace}\left(\left(\int_0^t R\Omega(s)R'ds\right)\left(\int_0^t R\Omega(s)R'ds\right)^{-1}\right) \\
&= \text{trace}\left(I_q\right)
\end{aligned}$$

$$= \quad q.$$

We can also write

$$\begin{aligned}
\operatorname{rank}\left(\Sigma_{11}(t)J(t)\Sigma_{11}(t)\right) &= \operatorname{rank}\left[\left(\int_0^t \Omega(s)R'ds\right)\left(\int_0^t R\Omega(s)R'ds\right)^{-1}\left(\int_0^t R\Omega(s)ds\right)\right] \\
&= \operatorname{rank}\left(\int_0^t \Omega(s)R'ds\right),
\end{aligned}$$

and since $\int_0^t \Omega(s)ds$ is invertible, we have

$$\operatorname{rank}\left(\int_0^t \Omega(s)R'ds\right) = \operatorname{rank}(R) = q.$$

Therefore the second condition is satisfied. The third condition is satisfied since $J(t)\Sigma_{11}(t)$ is idempotent. Indeed,

$$\begin{aligned}
J(t)\Sigma_{11}(t) &= R'\left(\int_0^t R\Omega(s)R'ds\right)^{-1} R\left(\int_0^t \Omega(s)R'[R\Omega(s)R']^{-1}R\Omega(s)ds\right) \\
&= R'\left(\int_0^t R\Omega(s)R'ds\right)^{-1}\left(\int_0^t R\Omega(s)ds\right),
\end{aligned}$$

and thus

$$\begin{aligned}
&(J(t)\Sigma_{11}(t))(J(t)\Sigma_{11}(t)) \\
&= R'\left(\int_0^t R\Omega(s)R'ds\right)^{-1}\left(\int_0^t R\Omega(s)ds\right)R'\left(\int_0^t R\Omega(s)R'ds\right)^{-1}\left(\int_0^t R\Omega(s)ds\right) \\
&= R'\left(\int_0^t R\Omega(s)R'ds\right)^{-1}\left(\int_0^t R\Omega(s)ds\right) \\
&= J(t)\Sigma_{11}(t).
\end{aligned}$$

Therefore the third condition is satisfied. By Proposition B.1.11, $J(t)\Sigma_{11}(t)J(t) = J(t)$, so

$$\left(\int_0^t \delta^{*\prime}(s)ds\right) J(t)\Sigma_{11}(t)J(t)\left(\int_0^t \delta^*(s)ds\right)$$

$$
\begin{aligned}
&= \left(\int_0^t \delta^{*\prime}(s)ds\right) J(t) \left(\int_0^t \delta^*(s)ds\right) \\
&= \left(\int_0^t \delta^{*\prime}(s)ds\right) \times R' \left(\int_0^t R\Omega(s)R'ds\right)^{-1} R \times \left(\int_0^t \delta^*(s)ds\right) \\
&= \Delta(t),
\end{aligned}
$$

which satisfies the last condition.

Therefore all of the conditions of Theorem 4 of Styan (1970) are satisfied and, from this theorem, we conclude that

$$\xi'(t)J(t)\xi(t) \sim \chi_q^2(\Delta(t)).$$

□

In the next three proofs, the ADRs of the unrestricted, restricted and shrinkage estimators are calculated.

In order to calculate the ADR of any of the estimators for $B(t)$, we must first recall from Proposition 2.1 that, if Conditions $\mathcal{A}$, $\mathcal{B}$ and (2.9) hold, then

$$(\xi_n'(t), \eta_n'(t))' \xrightarrow[n\to\infty]{D} (\xi'(t), \eta'(t))'$$

on $\mathcal{D}([0,\tau])$, where $\{(\xi'(t), \eta'(t))', t \geq 0\}$ is the Gaussian martingale with

$$(\xi'(t), \eta'(t))' \sim N_{2k}\left[\int_0^t \begin{pmatrix} \delta^*(t) \\ \delta^*(t) \end{pmatrix}, \begin{pmatrix} \Sigma_{11}(t) & 0 \\ 0 & \Omega_{11}(t) - \Sigma_{11}(t) \end{pmatrix}\right],$$

for $0 \leq t \leq \tau$, where

$$\Omega_{11}(t) = \int_0^t \Omega(s)ds,$$

and

$$\Sigma_{11}(t) = \int_0^t \Sigma(s) ds.$$

We will begin with the ADR of the unrestricted estimator.

Proof of Proposition 2.3.1.

Using the loss function at (2.17), the ADR of the unrestricted estimator, $\widehat{B}_{LS}$ of B is given by

$$ADR(\widehat{B}_{LS}, B; W^*) = \int_0^\tau \mathrm{E}[\rho'(t) W^*(t) \rho(t)] dt,$$

where

$$\rho(t) \;\sim\; N_k\left(0, \int_0^t \Omega(s) ds\right).$$

Therefore, since this is a single value,

$$\begin{aligned} ADR(\widehat{B}_{LS}, B; W^*) &= \int_0^\tau tr\left(W^*(t) \int_0^t \Omega(s) ds\right) dt \\ &= tr\left(\int_0^\tau \int_0^t W^*(t) \Omega(s) ds \; dt\right) \\ &= tr\left(\int_0^\tau \int_s^\tau W^*(t) \Omega(s) dt \; ds\right) \\ &= tr\left(\int_0^\tau \Omega(s) \left(\int_s^\tau W^*(t) dt\right) ds\right). \end{aligned}$$

Then, by (2.18),

$$\begin{aligned} ADR(\widehat{B}_{LS}, B; W^*) &= tr\left(\int_0^\tau \Omega(s) \overline{W}^*(s) ds\right) \\ &= \int_0^\tau tr\left(\Omega(s) \overline{W}^*(s)\right) ds, \end{aligned}$$

and this completes the proof.

□

Proof of Proposition 2.3.2.

Using the loss function at (2.17), the ADR of the restricted estimator, $\widetilde{B}_R$ of B is given by

$$ADR(\widetilde{B}_R, B; W^*) = \int_0^\tau \mathrm{E}[\eta'(t)W^*(t)\eta(t)]dt.$$

Similar to the proof of Proposition 2.3.1, since this is a single value,

$$ADR(\widetilde{B}_R, B; W^*) = \int_0^\tau tr\left(\Omega_{11}(s)W^*(s) - \Sigma_{11}(s)W^*(s)\right)ds + \int_0^\tau \delta^{*\prime}(s)W^*(s)\delta^*(s)ds. \tag{B.17}$$

Notice that

$$\begin{aligned}
& \int_0^\tau tr\left(\Omega_{11}(t)W^*(t) - \Sigma_{11}(t)W^*(t)\right)dt \\
= \ & tr\left(\int_0^\tau \left(\int_0^t \Omega(s)ds\right) W^*(t)dt - \int_0^\tau \left(\int_0^t \Sigma(s)ds\right) W^*(t)dt\right).
\end{aligned}$$

Then, similar to the proof of 2.3.1, we have

$$\begin{aligned}
& \int_0^\tau tr\left(\Omega_{11}(t)W^*(t) - \Sigma_{11}(t)W^*(t)\right)dt \\
= \ & tr\left(\int_0^\tau \int_s^\tau \Omega(s)W^*(t)dtds - \int_0^\tau \int_s^\tau \Sigma(s)W^*(t)dtds\right) \\
= \ & tr\left(\int_0^\tau \Omega(s)\overline{W}^*(s)ds - \int_0^\tau \Sigma(s)\overline{W}^*(s)ds\right) \\
= \ & \int_0^\tau tr\left(\Omega(s)\overline{W}^*(s)\right)ds - \int_0^\tau tr\left(\Sigma(s)\overline{W}^*(s)\right)ds.
\end{aligned}$$

So, combining this with (B.17), we have that

$$\begin{aligned}
ADR(\widetilde{B}_R, B; W^*) = & \int_0^\tau tr\left(\Omega(s)\overline{W}^*(s)\right)ds - \int_0^\tau tr\left(\Sigma(s)\overline{W}^*(s)\right)ds \\
& + \int_0^\tau \delta^{*\prime}(s)W^*(s)\delta^*(s)ds.
\end{aligned}$$

□

Proof of Proposition 2.3.3.

Using the loss function at (2.17), the ADR of the shrinkage estimator, $\widehat{B}^S$ of B is given by

$$
\begin{aligned}
& ADR(\widehat{B}^S, B; W^*) \\
= & \int_0^\tau E\left[\left(\eta'(t) - \left(1 - \frac{q-2}{\varphi(t)}\right)\xi(t)\right)' W^*(t)\left(\eta'(t) - \left(1 - \frac{q-2}{\varphi(t)}\right)\xi(t)\right)\right] dt \\
= & \quad \int_0^\tau \mathrm{E}[\eta'(t)W^*(t)\eta(t)]dt \qquad \text{(B.18)} \\
& -2\int_0^\tau E\left[\eta'(t)W^*(t)\left(1 - \frac{q-2}{\varphi(t)}\right)\xi(t)\right] dt \qquad \text{(B.19)} \\
& +\int_0^\tau E\left[\xi'(t)W^*(t)\xi(t)\left(1 - \frac{q-2}{\varphi(t)}\right)^2\right] dt. \qquad \text{(B.20)}
\end{aligned}
$$

We will consider (B.18), (B.19) and (B.20) separately. Beginning with (B.18), we have

$$
\begin{aligned}
\int_0^\tau \mathrm{E}[\eta'(t)W^*(t)\eta(t)]dt \quad = & \quad \int_0^\tau tr\left(\Omega(s)\overline{W}^*(s)\right) ds \; - \; \int_0^\tau tr\left(\Sigma(s)\overline{W}^*(s)\right) ds \\
+ & \quad \int_0^\tau \delta^{*\prime}(s)W^*(s)\delta^*(s)ds \qquad \text{(B.21)} \\
= & \quad ADR(\widehat{B}_{LS}, B; W^*) - \int_0^\tau tr\left(\Sigma(s)\overline{W}^*(s)\right) ds \qquad \text{(B.22)} \\
& \quad + \int_0^\tau \delta^{*\prime}(s)W^*(s)\delta^*(s)ds. \qquad \text{(B.23)}
\end{aligned}
$$

by Proposition 2.3.2 .

From (B.19) and the definition of $\varphi(t)$, we have,

$$
\begin{aligned}
& -2\int_0^\tau E\left[\eta'(t)W^*(t)\left(1 - \frac{q-2}{\varphi(t)}\right)\xi(t)\right] dt \\
= & \; -2\int_0^\tau E\left[\eta'(t)W^*(t)\left(1 - \frac{q-2}{\mathrm{trace}(\xi'(t)\widehat{J}(t)\xi(t))}\right)\xi(t)\right] dt \\
= & \; -2\int_0^\tau E\left[h\left(\; \mathrm{trace}(\xi'(t)\widehat{J}(t)\xi(t))\right)\eta'(t)W^*(t)\xi(t)\right] dt,
\end{aligned}
$$

where

$$h\left(z(t)\right)=1-\frac{q-2}{z(t)}$$

for $z(t)\neq 0$.

Then, by Theorem 2.1 of Nkurunziza (2012b),

$$\begin{aligned}
& -2\int_0^\tau E\left[\eta'(t)W^*(t)\left(1-\frac{q-2}{\varphi(t)}\right)\xi(t)\right]dt \\
= & -2\int_0^\tau E\left[1-\frac{q-2}{D_1(t)}\right]\delta^{*\prime}(t)W^*(t)\delta^*(t)dt, \qquad \text{(B.24)}
\end{aligned}$$

where $D_1(t)$ is defined as in (2.19). Note that the conditions of that theorem are satisfied by Proposition B.1.11.

From (B.20) and the definition of $\varphi(t)$, we have,

$$\begin{aligned}
& \int_0^\tau E\left[\xi'(t)W^*(t)\xi(t)\left(1-\frac{q-2}{\varphi(t)}\right)^2\right]dt \\
= & \int_0^\tau E\left[h\Big(\operatorname{trace}(\xi'(t)\widehat{J}(t)\xi(t))\Big)\operatorname{trace}(\xi'(t)W^*(t)\xi(t))\right]dt,
\end{aligned}$$

where

$$h\left(z(t)\right)=\left(1-\frac{q-2}{z(t)}\right)^2$$

for $z(t)\neq 0$.

Then, by Theorem 2.2 of Nkurunziza (2012b),

$$\begin{aligned}
& \int_0^\tau E\left[\xi'(t)W^*(t)\xi(t)\left(1-\frac{q-2}{\varphi(t)}\right)^2\right]dt \\
= & \int_0^\tau E\left[\left(1-\frac{q-2}{D_1(t)}\right)^2\right]\operatorname{trace}(W^*(t)\Sigma_{11}(t))dt \\
& +\int_0^\tau E\left[\left(1-\frac{q-2}{D_2(t)}\right)^2\right]\operatorname{trace}(\delta^{*\prime}(t)W^*(t)\delta^*(t))dt. \qquad \text{(B.25)}
\end{aligned}$$

Once again, note that the conditions of this theorem are satistfied by Proposition B.1.11. Therefore, combining (B.21), (B.24) and (B.25), we have

$$ADR(\widehat{B}^S,B;W^*) = ADR(\widehat{B}_{LS},B;W^*)-\int_0^\tau tr\left(\Sigma(t)\overline{W}^*(t)\right)dt$$

$$
\begin{aligned}
&+\int_0^\tau \delta^{*\prime}(t)W^*(t)\delta^*(t)dt \\
&-2\int_0^\tau E\left[1-\frac{q-2}{D_1(t)}\right]\delta^{*\prime}(t)W^*(t)\delta^*(t)dt \\
&+\int_0^\tau E\left[\left(1-\frac{q-2}{D_1(t)}\right)^2\right]\ \text{trace}(W^*(t)\Sigma_{11}(t))dt \\
&+\int_0^\tau E\left[\left(1-\frac{q-2}{D_2(t)}\right)^2\right]\ \text{trace}(\delta^{*\prime}(t)W^*(t)\delta^*(t))dt.
\end{aligned}
$$

□

In order to prove Proposition 2.3.4, we need Courant's Theorem. We provide this theorem, as given in Saleh (2006), below.

Proposition B.1.12. *Let $\lambda_1, \lambda_2, ..., \lambda_n$ be the characteristic roots of an $n \times n$ matrix A such that $min\lambda_i = \lambda_1$, $max\lambda_i = \lambda_n$, and let $v_1, ..., v_n$ be the characteristic vectors. Then $A = \lambda_1 v_1 v_1' + ... + \lambda_n v_n v_n'$, $I = v_1 v_1' + ... + v_n v_n'$: $sup\left(\frac{x'Ax}{xx'}\right) = \lambda_n$ and $inf\left(\frac{x'Ax}{xx'}\right) = \lambda_1$. Hence,*

$$ch_{min}(A) \leq \frac{x'Ax}{xx'} \leq ch_{max}(A),$$

where

$$min_i\,(\lambda_i) = ch_{min}(A)$$

and

$$max_i\,(\lambda_i) = ch_{max}(A).$$

Proof of Proposition 2.3.4 .

The ADR of the shrinkage estimator is given by

$$ADR\left(\widehat{B}^S, B; W^*\right)$$

$$
\begin{aligned}
= \; & ADR\left(\widehat{B}_{LS}, B; W^*\right) \\
& - \int_0^\tau \operatorname{trace}\left(\Sigma(t)\overline{W}^*(t)\right) dt + \int_0^\tau \delta^{*\prime}(t) W^*(t) \delta^*(t) dt \\
& -2 \int_0^\tau \mathrm{E}\left[1 - \frac{q-2}{D_1(t)}\right] \delta^{*\prime}(t) W^*(t) \delta^*(t) dt \\
& + \int_0^\tau \mathrm{E}\left[\left(1 - \frac{q-2}{D_1(t)}\right)^2\right] \operatorname{trace}(W^*(t)\Sigma_{11}(t)) dt \\
& + \int_0^\tau \mathrm{E}\left[\left(1 - \frac{q-2}{D_2(t)}\right)^2\right] \delta^*(t)' W^*(t) \delta^*(t) dt
\end{aligned}
$$

$$
\begin{aligned}
= \; & ADR\left(\widehat{B}_{LS}, B; W^*\right) \\
& +2(q-2) \int_0^\tau E[D_1^-(t)] \operatorname{trace}(\delta^{*\prime}(t) W^*(t) \delta^*(t) dt) - \int_0^\tau \delta^{*\prime}(t) W^*(t) \delta^*(t) dt \\
& -2(q-2) \int_0^\tau \mathrm{E}[D_1^{-1}(t)] \operatorname{trace}(W^*(t)\Sigma_{11}(t)) dt \\
& +(q-2)^2 \int_0^\tau \mathrm{E}[D_1^{-2}(t)] \operatorname{trace}(W^*(t)\Sigma_{11}(t)) dt + \int_0^\tau \delta^{*\prime}(t) W^*(t) \delta^*(t) dt \\
& -2(q-2) \int_0^\tau \mathrm{E}[D_2^{-1}(t)] \operatorname{trace}(\delta^{*\prime}(t) W^*(t) \delta^*(t)) dt \\
& +(q-2)^2 \int_0^\tau \mathrm{E}[D_2^{-2}(t)] \operatorname{trace}(\delta^{*\prime}(t) W^*(t) \delta^*(t)) dt.
\end{aligned}
$$

This can be further simplified, leaving us with

$$
ADR\left(\widehat{B}^S, B; W^*\right)
$$

$$
\begin{aligned}
= \; & ADR\left(\widehat{B}_{LS}, B; W^*\right) \\
& +(q-2) \int_0^\tau \operatorname{trace}(\delta^{*\prime}(t) W^*(t) \delta^*(t)) \mathrm{E}[2D_1^{-1}(t) - 2D_2^{-1}(t) + (q-2) D_2^{-2}(t)] dt
\end{aligned}
$$

$$-(q-2)\int_0^\tau \text{trace}(W^*(t)\Sigma_{11}(t))\left\{2\text{E}[D_1^{-1}(t)]-(q-2)\text{E}[D_1^{-2}(t)]\right\}dt.$$

Using the definitions of D_1 and D_2, along with the identity

$$\text{E}[\chi_{d+2}^{-2}(\Delta(t))]-\text{E}[\chi_{d+4}^{-2}(\Delta(t))]=2\text{E}[\chi_{d+4}^{-4}(\Delta(t))],$$

we have

$$\begin{aligned} & ADR\left(\widehat{B}^S, B; W^*\right) \\ = \; & ADR\left(\widehat{B}_{LS}, B; W^*\right) \\ & +(q^2-4)\int_0^\tau c_2(t)\text{E}[D_2^{-2}(t)]dt \\ & -(q-2)\int_0^\tau c_1(t)\left\{2\text{E}[D_1^{-1}(t)]-(q-2)\text{E}[D_1^{-2}(t)]\right\}dt, \end{aligned}$$

where $c_1(t) = \text{trace}(W^*(t)\Sigma_{11}(t))$ and $c_2(t) = \text{trace}(\delta^{*\prime}(t)W^*(t)\delta^*(t))$. We also have the identity

$$(q-2)\text{E}[\chi_{d+2}^{-4}(\Delta(t))]=\text{E}[\chi_{d+2}^{-2}(\Delta(t))]-2\Delta(t)\text{E}[\chi_{d+4}^{-4}(\Delta(t))],$$

which allows us to write

$$\begin{aligned} & ADR\left(\widehat{B}^S, B; W^*\right) \\ = \; & ADR\left(\widehat{B}_{LS}, B; W^*\right) \\ & -(q-2)\int_0^\tau \left\{c_1(t)\text{E}[D_1^{-1}(t)]+2c_1(t)\Delta(t)\text{E}[D_2^{-2}(t)]-(q+2)c_2(t)\text{E}[D_2^{-2}(t)]\right\}dt. \end{aligned}$$

Thus, it suffices to prove that $2\Delta(t)c_1(t)-(q+2)c_2(t)\geq 0$, for all t, since $q-2>0$.

If $c_2(t)=0$, we are done, so let $c_2(t)>0$. Then we need to prove that

$\Delta(t)\dfrac{c_1(t)}{c_2(t)} \geq \dfrac{(q+2)}{2}$, for all t. Following the proof of Corollary 2.1 in Nkurunziza (2010), we can use Courant's Theorem to get

$$\frac{c_1(t)}{ch_{max}(W^*(t)\Sigma_{11}(t))} \leq \Delta(t)\frac{c_1(t)}{c_2(t)} \leq \frac{c_1(t)}{ch_{min}(W^*(t)\Sigma_{11}(t))},$$

where $ch_{min}(A)$ and $ch_{max}(A)$ are the smallest and largest eigenvalues of the matrix A, respectively. Then $\Delta(t)\dfrac{c_1(t)}{c_2(t)} \geq \dfrac{(q+2)}{2}$, for all t, and we are done.

□

Proof of Proposition 2.3.6 .

Proposition 2.3.5 states that

$$\begin{aligned}
ADR\left(\widehat{B}^{S+}, B; W^*\right) &= \int_0^\tau tr\left(\left[\Omega_{11}(t) - \Sigma_{11}(t)\right] W^*(t)\right) dt + \int_0^\tau \delta^{*\prime}(t)W^*(t)\delta^*(t)dt \\
&\quad -2\int_0^\tau E\left[\left(1 - \frac{q-2}{D_1(t)}\right)^+\right] \delta^{*\prime}(t)W^*(t)\delta^*(t)dt \\
&\quad +\int_0^\tau E\left[\left(\left[1 - \frac{q-2}{D_1(t)}\right]^+\right)^2\right] \text{trace}(W^*(t)\Sigma_{11}(t))dt \\
&\quad +\int_0^\tau E\left[\left(\left[1 - \frac{q-2}{D_2(t)}\right]^+\right)^2\right] \delta^*(t)'W^*(t)\delta^*(t)dt,
\end{aligned}$$

Notice that we can write

$$\begin{aligned}
E\left[\left(1 - \frac{q-2}{D_1(t)}\right)^+\right] &= E\left[\mathbf{I}_{(0,\infty)}\left\{1 - \frac{q-2}{D_1(t)}\right\}\left(1 - \frac{q-2}{D_1(t)}\right)\right] \\
&= E\left[\left(1 - \mathbf{I}_{(-\infty,0)}\left\{1 - \frac{q-2}{D_1(t)}\right\}\right)\left(1 - \frac{q-2}{D_1(t)}\right)\right].
\end{aligned}$$

Then

$$E\left[\left(1 - \frac{q-2}{D_1(t)}\right)^+\right] = E\left[1 - \frac{q-2}{D_1(t)}\right] - E\left[\mathbf{I}_{(-\infty,0)}\left\{1 - \frac{q-2}{D_1(t)}\right\}\left(1 - \frac{q-2}{D_1(t)}\right)\right].$$

Similarly,

$$
\begin{aligned}
E\left[\left(\left[1-\frac{q-2}{D_1(t)}\right]^+\right)^2\right] &= E\left[\left(1-\frac{q-2}{D_1(t)}\right)^2\right] \\
&\quad -E\left[\mathbf{I}_{(-\infty,0)}\left\{1-\frac{q-2}{D_1(t)}\right\}\left(1-\frac{q-2}{D_1(t)}\right)^2\right].
\end{aligned}
$$

Therefore we can write $ADR\left(\widehat{B}^{S+}, B; W^*\right)$ as a function of $ADR\left(\widehat{B}^{S}, B; W^*\right)$ in the following way:

$$
\begin{aligned}
& ADR\left(\widehat{B}^{S+}, B; W^*\right) \\
= \; & ADR\left(\widehat{B}^{S}, B; W^*\right) \\
& +2\int_0^\tau E\left[\mathbf{I}_{(-\infty,0)}\left\{1-\frac{q-2}{D_1(t)}\right\}\left(1-\frac{q-2}{D_1(t)}\right)\right]\delta^{*\prime}(t)W^*(t)\delta^*(t)dt \\
& +\int_0^\tau E\left[\mathbf{I}_{(-\infty,0)}\left\{1-\frac{q-2}{D_1(t)}\right\}\left(1-\frac{q-2}{D_1(t)}\right)^2\right] \operatorname{trace}(W^*(t)\Sigma_{11}(t))dt \\
& +\int_0^\tau E\left[\mathbf{I}_{(-\infty,0)}\left\{1-\frac{q-2}{D_1(t)}\right\}\left(1-\frac{q-2}{D_1(t)}\right)^2\right] \delta^*(t)'W^*(t)\delta^*(t)dt.
\end{aligned}
$$

This concludes the proof.

□

Proof of Proposition 2.3.10 .

The ADB of $\widehat{B}^S$ is given by

$$
ADB(\widehat{B}^S, B) = \int_0^\tau E\left[\xi(t)+\left(1-\frac{q-2}{\varphi(t)}\right)\eta'(t)\right]dt
$$

$$= \int_0^\tau E\left[\xi(t)\right] + \int_0^\tau E\left[h\left(trace\left(\eta'(t)J(t)\eta(t)\right)\right)\eta'(t)\right]dt,$$

where $h(z) = 1 - \dfrac{q-2}{z}$.

Then, by Theorem 2.2 of Nkurunziza (2012a),

$$ADB(\widehat{B}^S, B) = \int_0^\tau \delta^*(t)dt + \int_0^\tau E\left[1 - \frac{q-2}{D_1(t)}\right]\delta^*(t)dt.$$

This concludes the proof.

$\square$

Bibliography

[1] Ahmed S. E., Nicol C. J. (1999) *Shrinkage Estimators for the Nonlinear Regression Model*. University of Regina, Saskatchewan.

[2] Andersen P. K. , Gill R. D. (1982) *Cox's Regression Model for Counting Processes: A Large Sample Study* . The Annals of Statistics Vol. 10, No. 4, p 1100-1120.

[3] Andersen P. K., Borgan 0., Gill R. D., Keiding N. (1993) *Statistical Models Based on Counting Processes*. Springer, New York.

[4] Athreya K., Lahiri S. (2006) *Measure Theory and Probability Theory*. Springer, New York.

[5] Billingsley P. (1995) *Probability and Measure (Third Edition)* . Wiley, New York.

[6] Borovskikh Y. V., Semenovi V. (1997) *Martingale Approximation*. VSP, The Netherlands.

[7] Burrill C. W. (1972) *Measure Integration and Probability*. McGraw-Hill, New York.

[8] Doob J. L. (1953) *Stochastic Processes*. Wiley, New York.

[9] Fleming T. R., Harrington D. P. (1991) *Counting Processes and Survival Analysis*. Wiley, New York.

[10] Gill R. D. (1980) *Censoring and Stochastic Intergrals.* Mathematical Center Tract 124, Amsterdam: Mathematische Centrum.

[11] Kalbfleisch J. D., Prentice R. L. (1980) *The Statistical Analysis of Failure Time Data.* Wiley, New Jersey.

[12] Klebaner F. C. (2005) *Introduction to Stochastic Calculus With Applications (Second Edition).* Imperial College Press, London.

[13] Lenglart E. (1977) *Transformation de martingales locales par changement absolue continu de probabilits.* Zeitschrift fr Wahrscheinlichkeit 39, p 65-70.

[14] Lipster R. S., Shiryayev A. N., (1989) *Theory of Martingales.* Kluwer, Dordrecht.

[15] Martinussen T., Scheike T. H. (2006) *Dynamic Regression Models for Survival Data.* Springer, New York.

[16] Nkurunziza S. (2010) *Shrinkage Strategies In Some Multiple Multi-factor Dynamical Systems.* ESAIM: PS, doi: 10.1051/ps/2010015.

[17] Nkurunziza S. (2012a) *The risk of pretest and shrinkage estimators.* Statistics: A Journal of Theoretical and Applied Statistics, Vol. 46, 3, 305-312.

[18] Nkurunziza S. (2012b) *A Simple Formula for Asymptotic Distributional Risk of Some Estimators.* Brazilian Journal of Probability and Statistics Vol. 26, No. 2, p 113-122.

[19] Rao K. M. (1969) *Quasi-Martingales.* Math. Scand., Sweden, p 79-92 .

[20] Saleh A. K. (2006) *Theory of Preliminary Test and Stein-Type Estimation with Applications.* Wiley, New Jersey.

[21] Stein C. (1965) *Inadmissibility of the usual Estimator for the Mean of a Multivariate Distribution*. Proc. Third Berkeley Symposium on Mathematical Statistics and Probability, 1, p 197-206 .

[22] Styan G. (1970) *Notes on the Distribution of Quadratic Forms in Singular Normal Variables*. Biometrika, 57, 3, 567-572.

[23] Taylor S. J. (1973) *Introduction to Measure and Integration*. CUP archive.

[24] Therneau T., Grambsch P. (2000) *Modeling Survival Data: Extending the Cox Model*. Springer, New York.

[25] Whittle P. (2000) *Probability Via Expectation (Fourth Edition)* . Springer-Verlag, New York.

[26] Williams D. (1991) *Probability With Martingales* . Cambridge University Press, UK.

Vita Auctoris

NAME: Katrina Tomanelli

PLACE OF BIRTH: Leamington, Ontario

YEAR OF BIRTH: 1988

EDUCATION:

Cardinal Carter Catholic Secondary School, Leamington, Ontario

2002-2006

University of Windsor, Windsor, Ontario

Bachelor of Mathematics *(Honours Mathematics and Statistics)*

2006-2010

University of Windsor, Windsor, Ontario

Master of Science *(in Statistics)*

2010-2012