Semiparametric Estimation Methods for Survival and Biomarker Data

by

Theresa P. Devasia

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in the University of Michigan
2021

Doctoral Committee:

Professor Alexander Tsodikov, Chair
Assistant Professor Walter Dempsey
Professor Susan Murray
Associate Professor Ted Skolarus

Theresa P. Devasia

tdevasia@umich.edu

ORCID ID: 0000-0002-0348-1407

For my family

ACKNOWLEDGEMENTS

This dissertation is the product of my five wonderful years at Michigan, which have been the most academically challenging yet rewarding of my life. I would like to thank all of the amazing professors and students I have met during my time here. You have all left an impression on me, and I will always remember my time as a graduate student fondly. I would also like to thank all of the mentors I had as an undergraduate student at the University of Georgia for encouraging me to pursue my Ph.D and the Emory Winship Cancer Institute's Summer Scholars Research Program for introducing me to biostatistics.

This work would not have been possible without the support and guidance of my advisor Alex, whose knowledge of statistical theory is unparalleled. He took what I knew about survival analysis and expanded it beyond measure. He has been the most kind and patient mentor, gently guiding me along the way while providing me the independence to explore my own ideas, however misguided they may be. He has taught me how to be a better statistician, writer, and researcher.

I would like to thank my other committee members as well. Susan made me fall in love with survival analysis and taught me everything I know about martingales. Walter provided invaluable insight on joint modeling and stochastic process models. Ted gave me the opportunity to learn about prostate cancer and work on impactful clinical research at the VA. They have all provided incredible feedback and questions throughout my dissertation.

I would also like to thank my GSRA advisor Matt. He has been a wonderful mentor

and has allowed me to gain experience working on cancer research that impacts patient care. He has taught me how to be a better applied statistician and collaborator. I thank the wonderful investigators I have worked with in the Departments of Radiology and Radiation Oncology at Michigan Medicine for allowing me to contribute to a diverse set of research projects.

Lastly, I would like to thank my family. Without their love and support, I would not have made it through graduate school. I would like to thank my parents for all the sacrifices they have made and all the love they have given me. I would like to thank my brother and sister for their constant encouragement. Together, they have all helped me tremendously over the past five years, and I cannot say "thank you" enough to each of them.

The bulk of this dissertation was completed during the COVID-19 pandemic, and I have been extremely fortunate to be able to continue pursuing my research during this difficult time. This work was supported by methodological grants from the National Cancer Institute's Division of Cancer Control and Populations Sciences' (DCCPS) portfolio in statistical and analytic methods, 1R01CA242559, and 1U01CA253915 (Cancer Interventions Surveillance Modeling Network, CISNET).

TABLE OF CONTENTS

DEDICATION ii

ACKNOWLEDGEMENTS iii

LIST OF FIGURES viii

LIST OF TABLES x

LIST OF APPENDICES xii

ABSTRACT xiii

CHAPTER

I. Introduction 1

II. Efficiency of the Breslow Estimator in Semiparametric Transformation Models 7

2.1 Introduction 7
2.2 Methods 9
2.2.1 Model 9
2.2.2 Likelihood 10
2.2.3 Nonparametric Maximum Likelihood Estimation . . 11
2.2.4 Estimating Equations 13
2.2.5 Estimation Algorithm for NPMLE 14
2.2.6 Relative Efficiency of the Breslow Estimator 15
2.3 Simulation Study 18
2.4 SEER Data Analysis 30
2.5 Discussion 32

III. Reconsidering the Role of Endogenous Covariates in Survival Models 34

3.1 Introduction . . . 34
3.2 Time-Varying Covariates . . . 36
3.2.1 Definitions . . . 36
3.2.2 Defining Survival Functions . . . 38
3.3 Exponential Form of Survival . . . 39
3.3.1 Definition . . . 39
3.3.2 Validity . . . 40
3.4 Valid Survival Functions for Endogenous Markers . . . 40
3.5 Bivariate Shared Frailty Model . . . 43
3.5.1 Conditional Survival Assuming X Fully Observed . 44
3.5.2 Conditional Survival Assuming X Partially Observed 44
3.6 Semicompeting Risks Type Model . . . 46
3.6.1 Conditional Survival Assuming X is Fully Observed 49
3.6.2 Conditional Survival Assuming X is Partially Observed 50
3.7 Threshold Regression Model . . . 51
3.7.1 Definition . . . 51
3.7.2 Fixed Threshold and Fully Observed Marker . . . 52
3.7.3 Fixed Threshold and Partially Observed Marker . . 53
3.7.4 Random Threshold and Fully Observed Marker . . . 54
3.7.5 Random Threshold and Partially Observed Marker . 55
3.8 Discussion . . . 56

IV. Joint Modeling of a Time-to-Event and Partially Observed Marker Process Using Lévy Processes . . . 60

4.1 Introduction . . . 60
4.2 Joint Model . . . 63
4.2.1 Survival Model . . . 63
4.2.2 Developing the Joint Model . . . 67
4.3 Likelihood Construction . . . 73
4.3.1 Completely Observed Marker . . . 73
4.3.2 Marker Observed at Informative Measurement Time 74
4.3.3 Marker Observed at Set of Uninformative Measurement Times . . . 77
4.3.4 Nonparametric Estimation of the Cumulative Baseline Hazard . . . 80
4.4 Simulation Study . . . 81
4.5 SEER Data Analysis . . . 84
4.6 Discussion . . . 89

V. Conclusion . . . 93

APPENDICES 98
A.1 Expectation of Weight in NPMLE 99
A.2 Derivation of Profile Likelihood Score Equations 100
A.3 Asymptotic Difference of $\widehat{dH_t}^B$ and $\widehat{dH_t}^W$ 102
A.4 Variances of Estimators 105
A.5 Asymptotic Distribution and Profile Likelihood Information . 107
A.5.1 Asymptotic Properties of NPMLE 107
A.5.2 Profile Likelihood Hessian and Information Matrix . 114
A.6 Asymptotic Variances of $\hat{\boldsymbol{\beta}}^B$ and $\hat{\boldsymbol{\beta}}^W$ 117
A.7 Details of Proportional Odds Model 119
A.8 Details of Simulations 120
A.9 Additional Simulation Results 122
B.1 Bivariate Shared Frailty Model 124
B.2 Semicompeting Risks Type Model Survival Functions 125
B.3 Validity of Univariate Survival Functions 125
B.4 Semicompeting Risks Type Model Assuming X Fully Observed 127
B.5 Threshold Regression Models 128
B.5.1 Fixed Threshold and Partially Observed Marker . . 128
B.5.2 Random Threshold and Partially Observed Marker . 128
C.1 Marker Observed at Informative Measurement Time 130
C.2 Marker Observed at Uninformative Measurement Times . . . 133
C.2.1 Single Measurement Time 133
C.2.2 Multiple Measurement Times 136
C.3 Details of Simulations 138
C.4 Derivation of Cross-Sectional Form of Likelihood 139
C.5 Additional SEER Analysis Results 140
C.5.1 Results for Year of Diagnosis 2011 140
C.5.2 Results for Year of Diagnosis 2012 142
C.5.3 Results for Year of Diagnosis 2013 144
C.5.4 Comparison of SEER Parameter Estimates 146

BIBLIOGRAPHY 148

LIST OF FIGURES

Figure

2.1	ARE for increasing hazards, $\tau = 5$.	21
2.2	ARE for decreasing hazards, $\tau = 5$.	22
2.3	ARE for increasing hazards, $\tau = 10$.	23
2.4	ARE for decreasing hazards, $\tau = 10$.	24
2.5	ARE for increasing hazards, $\tau = 15$.	25
2.6	ARE for decreasing hazards, $\tau = 15$.	26
2.7	Estimated ARE for simulation settings 1-6 in Table 2.2.	29
2.8	Observed and estimated survival curves by stage and estimation method. Observed survival summarized via Kaplan-Meier method. Estimated survival curves calculated using either Breslow (red) or Weighted Breslow estimator/NPMLE (blue). Disease stage indicated by line type: solid = localized/stage 1; dashed = regional/stage 2; dotted = distant/stage 3.	31
4.1	Observed age of prostate cancer diagnosis by race, 2010.	87
4.2	Distribution of PSA at age of prostate cancer diagnosis, 2010.	87
4.3	Observed (solid) and estimated (dashed) survival curves for time to prostate cancer diagnosis for men diagnosed in 2010. 95% Wald-based confidence bands (dotted) constructed using bootstrap standard errors.	88
A.1	Boxplots of bias for Breslow and Weighted Breslow estimators in each PO simulation setting. Weighted estimator = NPMLE.	123
C.1	Observed age of prostate cancer diagnosis by race, 2011.	141
C.2	Distribution of PSA at age of prostate cancer diagnosis, 2011.	141
C.3	Observed (solid) and estimated (dashed) survival curves for time to prostate cancer diagnosis for men diagnosed in 2011. 95% Wald-based confidence bands (dotted) constructed using bootstrap standard errors.	142
C.4	Observed age of prostate cancer diagnosis by race, 2012.	143
C.5	Distribution of PSA at age of prostate cancer diagnosis, 2012.	143
C.6	Observed (solid) and estimated (dashed) survival curves for time to prostate cancer diagnosis for men diagnosed in 2012. 95% Wald-based confidence bands (dotted) constructed using bootstrap standard errors.	144
C.7	Observed age of prostate cancer diagnosis by race, 2013.	145
C.8	Distribution of PSA at age of prostate cancer diagnosis, 2013.	145

C.9 Observed (solid) and estimated (dashed) survival curves for time to prostate cancer diagnosis for men diagnosed in 2013. 95% Wald-based confidence bands (dotted) constructed using bootstrap standard errors. 146
C.10 Comparison of estimated parameters and 95% bootstrapped confidence interval by year of diagnosis. 147

LIST OF TABLES

Table

2.1 Summary statistics for AREs across all times in each setting. Please refer to Table A.1 for a description of each simulation setting. . . . 20

2.2 Settings used in proportional odds (PO) simulation study. n = sample size. k and λ are parameters of the Weibull baseline hazard function. τ = administrative censoring time. $Z_1 \sim N(\mu, \sigma)$ and $Z_2 \sim Bernoulli(p)$. % censoring is approximate. 28

2.3 Simulation results for estimator $\hat{\boldsymbol{\beta}}$ for the PO model. ASE = average standard error. ESD = empirical standard deviation. CP (%) = coverage probability of 95% Wald-based confidence interval. Weighted Breslow estimator = NPMLE. 28

2.4 Parameter estimates of the PO model fit to prostate cancer specific survival data extracted from SEER. Weighted Breslow estimator = NPMLE. Stage refers to disease stage, which may be localized (stage 1–reference group), regional (stage 2), or distant (stage 3). OR = odds ratio. 95% Wald-based confidence interval is reported. 31

4.1 Settings used in simulation study. n represents the sample size. k and λ are parameters of the Weibull baseline hazard function. $Z \sim Bernoulli(p)$. ν_1 and (η_0, η_1) relate to mean and variance, respectively, of the Gamma process representing the marker. c = administrative censoring time. Exp Censoring = % censoring anticipated. Obs Censoring = % censoring observed. 82

4.2 Simulation study results with parametric Weibull H. True parameter values are $(\nu_1, \eta_0, \eta_1) = (0.3, -2.1, 0.5)$. Est = average estimate across all simulations. ASE = average standard error. ESD = empirical standard deviation. CP (%) = coverage probability for 95% Wald-based interval. 83

4.3 Simulation study results using Weibull H. True Weibull hazard parameter values are $(k, \lambda) = (2, 1)$. Est = average estimate across all simulations. ASE = average standard error. ESD = empirical standard deviation. CP (%) = coverage probability for 95% Wald-based interval. 83

4.4 Simulation study results with Breslow estimated H. True parameter values are $(\nu_1, \eta_0, \eta_1) = (0.3, -2.1, 0.5)$. Est = average estimate across all simulations. ASEB = average standard error estimated from bootstrap samples. ESD = empirical standard deviation. CPB (%) = coverage probability for 95% Wald-based interval constructed using bootstrap standard error. 84
4.5 Marker parameter estimates in 2010. 95% Wald-based interval constructed using bootstrap standard error. p-value estimated using Wald Z value $= \frac{\text{Est}}{\text{Bootstrap SE}}$. 87
A.1 Simulation settings for ARE of the Breslow estimator. 121
A.2 Average number of iterations to converge for each estimator and average percent censoring in PO model simulations. Weighted Breslow estimator = NPMLE. 122
A.3 Summary of differences in regression parameter variances for each estimator. $\hat{V}(\hat{\boldsymbol{\beta}}^B)$ = estimated variance of Breslow-derived $\boldsymbol{\beta}$ and $\hat{V}(\hat{\boldsymbol{\beta}}^W)$ = estimated variance of Weighted Breslow/NPMLE-derived $\boldsymbol{\beta}$. 123
C.1 Marker parameter estimates in 2011. 95% Wald interval constructed using bootstrap standard error. p-value estimated using Wald Z value $= \frac{\text{Est}}{\text{Bootstrap SE}}$. 141
C.2 Marker parameter estimates in 2012. 95% Wald interval constructed using bootstrap standard error. p-value estimated using Wald Z value $= \frac{\text{Est}}{\text{Bootstrap SE}}$. 143
C.3 Marker parameter estimates in 2013. 95% Wald interval constructed using bootstrap standard error. p-value estimated using Wald Z value $= \frac{\text{Est}}{\text{Bootstrap SE}}$. 145

LIST OF APPENDICES

Appendix

A. Efficiency of the Breslow Estimator in Semiparametric Transformation Models 99

B. Reconsidering the Role of Endogenous Covariates in Survival Models . 124

C. Joint Modeling of a Time-to-Event and Partially Observed Marker Process Using Lévy Processes 130

ABSTRACT

In many clinical studies, the outcome of interest is an event time. In addition, longitudinal data on biomarkers may be collected, and such information may provide insight into underlying disease risk or severity. The goal in this dissertation is to develop models and estimation procedures that can incorporate survival and longitudinal data to provide patients and clinicians with knowledge of a subject's disease risk, which may influence future treatment decisions.

Traditional survival analysis methods often place strong assumptions on the effect of covariates on a patient's predicted risk. More flexible models have been proposed, but estimation of the cumulative baseline hazard function, which is of infinite dimension, proves difficult. In Chapter II, we consider two estimators of the cumulative baseline hazard: the nonparametric maximum likelihood estimator (NPMLE) and a Breslow-type estimator derived from a martingale estimating equation. The Breslow estimator relies solely on current event information, while the NPMLE depends on future data for risk predictions. We derive the asymptotic relative efficiency of the Breslow estimator in comparison to the NPMLE and demonstrate that while theoretically the Breslow estimator might not be fully efficient, in practice it is virtually identical to the NPMLE. The practical implication of this result is that the Breslow estimator may be used with minimal loss of efficiency while being conceptually and computationally more straightforward.

In Chapter III, we consider the role of internal or endogenous time-varying covariates in joint models of survival and biomarker data. The current belief, based largely on intuition, is that the future history of the endogenous marker should not

be incorporated into the hazard function. To the contrary, we show that the hazard function conditional on an endogenous process is a construct incorporating missing data associated with the future unobserved trajectory of the process. In addition, in the presence of an endogenous covariate, the validity of the exponential relationship between the survival and the hazard function is questioned. In this chapter, we offer explicit theory and examples of such models and use it to derive a generalized hazard function that satisfies the exponential relationship.

In Chapter IV, we extend prior work on joint models of survival and biomarker data and utilize the framework developed in Chapter III. We consider a discretely observed marker process at measurement times that may be informative. We focus in particular on the case of marked survival, where the marker is measured at the event time for subjects who experience the event of interest. We assume that the marker process can be modeled as a Lévy process that is connected to a survival model through a time transformation. Considering the partially observed marker and an informative measurement time, we derive estimators of the marker parameters.

CHAPTER I

Introduction

Survival analysis entails the examination of time-to-event outcomes in order to understand the probability of occurrence of an outcome of interest at some future time. Most survival models are defined using two components: a fully specified function of subject-specific factors and an unspecified function of time shared across subjects. In addition to event data, longitudinal marker data may be available for individuals in clinical studies. The information supplied by the marker could provide additional insight into the underlying disease process, improve survival predictions, and inform future treatment decisions. In this dissertation, we focus on developing semiparametric estimation methods for complex survival models, particularly in the case where partially observed biomarker data are available.

In the first chapter, we focus on estimation of the cumulative baseline hazard function in the semiparametric transformation model. The most popular survival model is Cox's proportional hazards model (1972), wherein the risk of the event is specified as the product of a parametric regression function incorporating covariates and a nonparametric cumulative baseline hazard function. Due to the presence of both a parametric and nonparametric term, the proportional hazards model is "semiparametric." Estimation is done using the partial likelihood method (*Cox*, 1972), which "profiles out" the unspecified cumulative baseline hazard that is of infinite dimension.

The asymptotic properties of Cox's model have been well established both directly and using counting processes (*Tsiatis*, 1981; *Andersen and Gill*, 1982).

One consequence of Cox's original model is that it assumes proportionality of the covariate effects, meaning that the hazard ratio between two covariate values remains constant over time. Due to potential heterogeneity that may evolve over the course of a study, the proportional hazards assumption can be restrictive. More flexible survival models, known as semiparametric transformation models, have been proposed (*Chen et al.*, 2002; *Zeng and Lin*, 2007). However, these more general models still face difficulties in terms of estimation due to the unspecified cumulative baseline hazard. Previous work has established that the nonparametric likelihood estimator (NPMLE) relies on future event information for a given risk set, which can lead to conceptual and computational challenges (*Chen*, 2009). A Breslow-type estimator derived from a martingale estimating equation has been proposed as an alternative, and it has been shown to rely solely on current event information for a given risk set (*Chen et al.*, 2002).

In Chapter II, we investigate the Breslow-type estimator of the cumulative baseline hazard and derive its asymptotic relative efficiency (ARE) compared to the NPMLE. We rely on martingale process theory to determine the ARE and study its behavior under different model settings. Because the cumulative baseline hazard can be thought of as a nuisance parameter, we also consider the asymptotic properties of the regression parameters associated with both the NPMLE and the Breslow-type estimator. We apply both estimators to SEER registry data on prostate-cancer specific survival by disease stage and demonstrate that the regression estimates obtained using each estimator for the cumulative baseline hazard are nearly identical. The Breslow-type estimator can be used with minimal loss of efficiency in certain cases, which allows for potential interim analysis as it relies solely on current event information.

The simplest survival models include only baseline covariate information in the parametric regression function. However, in observational clinical studies, biomarker data that are dynamic may provide insight into a patient's future survival. In such instances, models that can incorporate time-varying covariate information are needed to calculate accurate risk predictions. There are two main types of time-varying covariates: (1) external or exogenous and (2) internal or endogenous. The difference between the two types is largely defined by intuition on how the covariate relates to the subject. Exogenous covariates are thought to exist outside of the subject, whereas endogenous covariates exist within the subject and rely on the subject's survival to be observed. While these intuitive definitions allow for simple classification of any time-varying covariate, they lack satisfactory theoretical development and reflect assumptions made before modeling.

However, mathematical definitions for both exogenous and endogenous covariates exist (*Kalbfleisch and Prentice*, 2002), and they relate to how the level of available marker information impacts the risk of failure. Equivalently, exogenous and endogenous covariates can be distinguished based on how occurrence of the event impacts the distribution of the covariate trajectory. Let W denote a stochastic covariate related to the underlying health status of a patient. Then by the definition put forth in Kalbfleisch and Prentice (2002), W will be exogenous if the hazard of failure at time s remains the same conditional on $\mathcal{W}(s)$ or $\mathcal{W}(t)$, where $\mathcal{W}(u) = \{W(x) : 0 \leq x \leq u\}$ represents the marker history up to time u and $s < t$. W will be endogenous if the hazard changes dependent on the covariate history. Thus, the difference between the two types of time-varying covariates is largely based on the how the current risk of the event is related to the future covariate information: for exogenous covariates, the future is irrelevant, but for endogenous covariates, the future impacts the hazard at the current point in time. Based on these definitions, survival functions at time t conditional on the covariate history of an endogenous marker up to time t will be

mathematically invalid.

In Chapter III, we focus on clarifying the differences between exogenous and endogenous covariates. We reformulate the difference as a missing data problem, where endogenous covariates are linked to the unobserved future, but exogenous covariates are not. Under this formulation, we reexamine the role endogenous covariates play in survival models. In particular, we develop mechanistic survival models conditional on an endogenous marker history by treating the unobserved future path as a latent variable and applying the definition of a conditional expectation. Related to survival functions defined to be conditional on a time-varying covariate is the validity of the exponential form, which provides a link between the survival and hazard functions for continuous event times. While it is commonly thought that exponential forms are invalid for endogenous covariates, the legitimacy is related to the continuity of the event time (*Aven and Jensen*, 1999; *Finkelstein*, 2004). As part of our exploration of endogenous covariates, we delineate and decouple time-varying covariates and exponential forms. To do so, we consider illustrative examples including bivariate shared frailty models (*Hougaard*, 2000), bivariate semicompeting risks-type models (*Fine et al.*, 2001; *Tran et al.*, 2018), and threshold regression models where the event time can be thought of as the first-hitting time of a marker reaching a boundary state (*Ting Lee and Whitmore*, 2006). Through these examples, we highlight the shortcomings associated with the current understanding of time-varying covariates, and we demonstrate the advantages of our proposed framework.

Stochastic markers can play an important role in risk prediction for patients in biomedical studies. Ideally, they would be continuously monitored to understand how their levels and any associated fluctuations impact the probability of experiencing a health-related event. In reality, however, such markers will only be intermittently observed, either at scheduled clinical visits or possibly at some time related to the event of interest, for example at the event time itself. In such cases, models that

can account for the impact the marker has on future survival and for the marker measurements themselves are needed in order to accurately predict the risk of failure. The most common way to examine both event and marker information is through joint modeling, wherein a model for survival conditional on the marker and a model for the marker are specified such that there are latent factors shared between the two submodels (*Tsiatis et al.*, 1995; *Jewell and Kalbfleisch*, 1996; *Henderson et al.*, 2000; *Tsiatis and Davidian*, 2004; *Rizopoulos*, 2008). Joint modeling makes explicit how the marker impacts survival, and it also assumes a functional form for the marker process. In reality, however, the marker might be sparsely observed, particularly if it is only collected at event-related times. Thus, it may be difficult to properly assess the assumed functional form.

In Chapter IV, we consider joint models where we treat the marker process as a stochastic process that is partially observed. We largely focus on the case where the marker is measured at the event time if the event is observed, which is known as marked survival. Because the marker may be observed at the event time, the measurement time of the marker is informative. The informative observation of the marker is built into our proposed joint model through the survival modeling framework proposed in Chapter III and the marginal distribution of the marker. The proposed joint model is flexible and can also be applied in settings where measurement times are uninformative. A popular choice for the marker process is the non-negative Lévy process, which has a tractable Laplace functional and preserves the non-negativity of the hazard function (*Hoyle*, 2010). Prior work has focused on applying Lévy frailties as a multiplicative effect (*Gjessing et al.*, 2003; *Suresh*, 2018), whereas we explore introducing a time transformation into the Lévy frailty itself for even greater flexibility. The specific Lévy process we use is the Gamma process, where the mean and variance of the process can be specified to be functions of baseline covariates (*Suresh*, 2018).

In our survival submodel, we treat the marker process as a time-transformed frailty process, where the time transformation allows us to consider time on the scale of the accumulated risk of failure. Because the marker is a partially observed endogenous covariate, naively specifying the hazard of failure at time t conditional on the marker history up to t will result in an invalid survival function. To overcome this problem, we utilize the framework proposed in Chapter III. Using both components of the joint model, we derive estimators of the marker parameters. We explore the use of both a parametric cumulative baseline hazard function and a Breslow-type estimator of the cumulative baseline hazard function. We demonstrate the validity of the model through various simulation settings, and we apply our joint model to SEER registry data on prostate cancer diagnosis, where the marker prostate-specific antigen is measured at the age of diagnosis.

This dissertation investigates semiparametric estimation methods for survival models and joint models of time-to-event and longitudinal data. The methods outlined in this dissertation can be applied to a variety of clinical research projects, but they may be particularly useful in the context where a biomarker that acts as a surrogate of the underlying disease process is partially observed at an event time. We hope that our work can lead to better knowledge of subject-specific risk and the development of further methods to explore the complex associations between various disease-related processes.

CHAPTER II

Efficiency of the Breslow Estimator in Semiparametric Transformation Models

2.1 Introduction

The proportional hazards (PH) model proposed by Cox (1972) is one of the most ubiquitous methods in modern survival analysis. The hazard of failure is modeled as the product of a nonparametric baseline hazard function and a parametric function of covariates, thus making Cox's model semiparametric. As interest lies mostly in the effect of the covariates on the hazard, the partial likelihood method proposed by Cox (1972) provides an elegant solution to estimate the regression parameters while eliminating the infinite-dimensional baseline hazard. However, the statistical properties of the partial likelihood estimators were not immediately known. The asymptotic properties of the partial likelihood regression estimator and the nonparametric maximum likelihood estimator (NPMLE) of the cumulative baseline hazard function proposed by Breslow (1972) were proved directly by Tsiatis (1981). Andersen and Gill (1982) later developed counting process models for failure times, and using properties of martingales, they established the asymptotic behavior of these estimators as well.

Recent history has seen the development of more general semiparametric transformation models (*Chen et al.*, 2002; *Zeng and Lin*, 2007). Nonparametric estimation

in semiparametric transformation models proceeds by maximizing the likelihood over the nonparametrically-specified baseline cumulative hazard, an infinite-dimensional parameter, jointly with the other finite-dimensional parameters of the model. Profile likelihood obtained by maximizing out the infinite-dimensional parameter is often used for estimation (*Murphy and van de Vaart*, 2000). General maximization methods and expectation-maximization (EM) algorithms were proposed to deal with the high dimensionality of the problem (*Tsodikov*, 2003; *Zeng and Lin*, 2007), with the asymptotic theory supplied by empirical processes providing consistency, asymptotic normality, and efficiency of the NPMLE (*Kosorok et al.*, 2004; *Zeng and Lin*, 2006, 2007). Chen (2009) elucidated the martingale structure of the NPMLE and proposed the so-called "Weighted Breslow" NPMLE. The NPMLE of the cumulative baseline hazard has been shown to rely on future event information for a given risk set in the form of weights built using a martingale transform (*Chen*, 2009). Despite such successful developments, the dependence on the future made estimation computationally and theoretically challenging.

Before the NPMLE theory and algorithms were established, a simpler non-MLE Breslow-type estimator was derived from a martingale estimating equation (MEE) setting observed and expected counts of failures under the model equal, conditional on the past history (*Chen et al.*, 2002). This estimator could be considered a special case of the Weighted Breslow when weights are set to one, and we will refer to it as the Breslow estimator to distinguish it from the Weighted Breslow NPMLE. Since the weights house the future path of the functional hazard parameter and the event data in the Weighted Breslow estimator, setting weights equal to one in the Breslow estimator makes it dependent on the past history only. Despite the perception of lacking full efficiency (*Chen*, 2009), the unweighted Breslow estimator continues to be used in practice due to its computational efficiency.

The purpose of this paper is to consider the relative efficiency of the Breslow es-

timator, which relies solely on event information at the current point in time, and explore how this estimator compares to the NPMLE in practice. First, we define our model and estimators in Sections 2.2.1-2.2.5. Then, we derive the form of the relative efficiency of the Breslow estimator compared to the NPMLE in Section 2.2.6. Next, we compare their performance in simulation studies in Section 2.3 and in a SEER prostate cancer data set used previously to illustrate the EM algorithm for the NPMLE in Section 2.4 (*Tsodikov*, 2003). Finally, we conclude in Section 2.5 with a general discussion and comparison of the two estimators of the cumulative baseline hazard.

2.2 Methods

2.2.1 Model

Consider the failure time T and covariate $\boldsymbol{Z}$, which may be time-dependent. Conditional on covariates $\boldsymbol{Z}$, we assume that T is absolutely continuous under any $\boldsymbol{Z}$. This is needed to ensure that the exponential relationship between the survival and hazard functions, represented by S and λ, respectively, holds. Consider

$$\begin{aligned} S(t|\boldsymbol{\beta},\boldsymbol{Z}) &= \mathcal{L}(H_t|\boldsymbol{\beta},\boldsymbol{Z}) \\ \Lambda(H_t|\boldsymbol{\beta},\boldsymbol{Z}) &= -\ln \mathcal{L}(H_t|\boldsymbol{\beta},\boldsymbol{Z}) \\ d\Lambda(H_t|\boldsymbol{\beta},\boldsymbol{Z}) &= -\frac{d\mathcal{L}(H_t|\boldsymbol{\beta},\boldsymbol{Z})}{\mathcal{L}(H_t|\boldsymbol{\beta},\boldsymbol{Z})} = \lambda(H_t|\boldsymbol{\beta},\boldsymbol{Z})dt \end{aligned} \tag{2.1}$$

where Λ is the cumulative hazard, H is the baseline cumulative hazard, and $\boldsymbol{\beta}$ is a finite-dimensional parameter vector, usually of regression coefficients. The differentials $d\Lambda$ and $d\mathcal{L}$ are taken with respect to H_t. In equation (2.2) below, $S, \Lambda, \lambda, \mathcal{L}$, and $\Theta^0(H_t|\boldsymbol{\beta},\boldsymbol{Z})$ are generally functionals of the past history of $H(t)$. To ensure the

Breslow-form of the NPMLE, we assume that

$$d\Lambda(H_t|\boldsymbol{\beta}, \boldsymbol{Z}) = \Theta^0(H_t|\boldsymbol{\beta}, \boldsymbol{Z})dH_t \tag{2.2}$$

Time-dependent covariates are assumed to be external or exogenous, in which case the observed hazard function $\lambda(H_t|\boldsymbol{\beta}, \boldsymbol{Z}) = \lambda(H_t|\boldsymbol{\beta}, \boldsymbol{Z}_t)$, which allows us to avoid discussion of the model for the covariate process that would otherwise need to be incorporated into the hazard. Dependent on the model, we will understand H_t and $\boldsymbol{Z}_t$ either as values of the function or process at t or as the respective histories up to t. We will use the same notation when the meaning is clear based on the context.

Examples of the framework in equation (2.2) include (1) the semiparametric transformation model considered by Zeng and Lin (2007)

$$\Lambda(H_t|\boldsymbol{\beta}, \boldsymbol{Z}) = G\Big\{ \int_0^t e^{\boldsymbol{\beta}\boldsymbol{Z}(s)} dH_s \Big\} \tag{2.3}$$

where G is a continuously differentiable, strictly increasing function, and (2) transformation models induced by univariate frailty (*Tsodikov*, 2003; *Kosorok et al.*, 2004), where $\mathcal{L}$ is the Laplace transform of a nonnegative frailty random variable U whose distributional parameters depend on covariates and regression coefficients, evaluated at $H(t)$ with $S = \mathcal{L}_U(H_t) = E_U\{e^{-UH_t}\}$.

2.2.2 Likelihood

Let C be the right censoring time with survival function G and assume that $T \perp C|\boldsymbol{Z}$. Let τ be the maximum follow-up time in the study. The observed data are $(X_i, \Delta_i, \boldsymbol{Z}_{X_i})_{i=1}^n$, where $X_i = \min(T_i, C_i)$ and $\Delta_i = I(T_i \leq C_i)$, with $I(\cdot)$ defining the indicator function. Subjects are assumed to be independent. Then, the observed-data

log-likelihood is given by

$$\begin{aligned} l(\boldsymbol{\beta}, H) &= \sum_{i=1}^{n} \Delta_i \ln \left\{ d\Lambda(H_{X_i}|\boldsymbol{\beta}, \boldsymbol{Z}_{X_i}) \right\} - \Lambda(H_{X_i}|\boldsymbol{\beta}, \boldsymbol{Z}_{X_i}) \\ &= \sum_{i=1}^{n} \int_0^{\tau} dN_i(u) \ln \left\{ d\Lambda(H_u|\boldsymbol{\beta}, \boldsymbol{Z}_{u,i}) \right\} - Y_i(u) d\Lambda(H_u|\boldsymbol{\beta}, \boldsymbol{Z}_{u,i}) \end{aligned} \tag{2.4}$$

where $dN_i(u) = I(X_i = u, \Delta_i = 1)$ and $Y_i(u) = I(X_i \geq u)$. Here, $dN_i(u)$ is the increment of the observed event process, which only jumps when an event occurs at u, and $Y_i(u)$ is the at-risk process. If we define our filtration as $\mathcal{F}_{t^-} = \sigma\{N_i(s), Y_i(s), \boldsymbol{Z}_s : s \in [0, t); i = 1, \ldots, n\}$, then under the true model, $E(dN_i(t)|\mathcal{F}_{t^-}) = Y_i(t) d\Lambda(H_t|\boldsymbol{\beta}, \boldsymbol{Z}_{t,i}) = Y_i(t)\Theta^0(H_t|\boldsymbol{\beta}, \boldsymbol{Z}_{t,i}) dH_t$ is the compensator of the increment of the counting process N. Note that given $\mathcal{F}_{t^-}$, Θ^0 is assumed to be predictable.

2.2.3 Nonparametric Maximum Likelihood Estimation

Recall that our model in (2.2) is semiparametric as H is a completely unspecified increasing function. For estimation purposes, we treat H as a step-function with a set of jumps $\{dH\}$ at each of the m observed event times $t_1 < t_2 < \cdots < t_m$. To jointly estimate the parameters $(\boldsymbol{\beta}, H)$, we make use of profile maximum likelihood estimation, wherein $l(\boldsymbol{\beta}, H)$ is first maximized over H for fixed $\boldsymbol{\beta}$ and then maximized over $\boldsymbol{\beta}$ with $\hat{H}(\boldsymbol{\beta})$ substituted in for H. Because H is an infinite-dimensional function, to maximize $l(\boldsymbol{\beta}, H)$ with respect to H, we take a functional derivative. For a functional $F(H)$, consider the argument function H_x perturbed in the direction of an indicator function $I(x - t)$ with a jump at t, i.e., $\hat{H}(x) = H(x) + aI(x - t)$. We can define a local directional functional derivative with respect to H as

$$\frac{\partial F(H(\cdot))}{\partial dH_t} = \left. \frac{dF(H(\cdot) + aI(\cdot - t))}{da} \right|_{a=0} \tag{2.5}$$

The definition in (2.5) corresponds to taking a derivative of $l(\boldsymbol{\beta}, H)$ with respect to

the set of jumps $\{dH\}$, and it is valid for both discrete (finite sample) and continuous (asymptotic) settings. Applying this definition, we have $\frac{\partial H_x}{\partial dH_t} = I(x \geq t)$ and $\frac{\partial dH_x}{\partial dH_t} = I(x = t)$. Define the increment of the martingale process as $dM_i(t) = dN_i(t) - Y_i(t)\Theta^0(H_t|\boldsymbol{\beta}, \boldsymbol{Z}_{t,i})dH_t$, $i = 1, \ldots, n$. Then, the log-likelihood can be rewritten as

$$l(\boldsymbol{\beta}, H) = \sum_{i=1}^{n} \int_0^{\tau} dN_i(u)\big[\ln(\Theta^0(H_u|\boldsymbol{\beta}, \boldsymbol{Z}_{u,i})) + \ln(dH_u)\big] - Y_i(u)\Theta^0(H_u|\boldsymbol{\beta}, \boldsymbol{Z}_{u,i})dH_u \quad (2.6)$$

The score equation $U_{dH_t} = \frac{\partial}{\partial dH_t} l(\boldsymbol{\beta}, H)$ and NPMLE of dH_t for fixed $\boldsymbol{\beta}$ are

$$\begin{aligned} U_{dH_t} &= \sum_{i=1}^{n} \frac{dN_i(t)}{dH_t} - Y_i(t)\Theta^0(H_t|\boldsymbol{\beta}, \boldsymbol{Z}_{t,i}) + \int_{t^+}^{\tau} [\Theta^0(H_u|\boldsymbol{\beta}, \boldsymbol{Z}_{u,i}) - \Theta^1(H_u|\boldsymbol{\beta}, \boldsymbol{Z}_{u,i})]dM_i(u) \\ \widehat{dH_t}^W(\boldsymbol{\beta}) &= \frac{\frac{1}{n}\sum_{i=1}^{n} dN_i(t)}{\frac{1}{n}\sum_{i=1}^{n} Y_i(t)\Theta^0(H_t|\boldsymbol{\beta}, \boldsymbol{Z}_{t,i}) - \int_{t^+}^{\tau} \big[\Theta^0(H_u|\boldsymbol{\beta}, \boldsymbol{Z}_{u,i}) - \Theta^1(H_u|\boldsymbol{\beta}, \boldsymbol{Z}_{u,i})\big]dM_i(u)} \end{aligned} \quad (2.7)$$

where $\Theta^1(H_t|\boldsymbol{\beta}, \boldsymbol{Z}_t) = \Theta^0(H_t|\boldsymbol{\beta}, \boldsymbol{Z}_t) - \frac{\partial}{\partial dH_t} \ln[\Theta^0(H_t|\boldsymbol{\beta}, \boldsymbol{Z}_t)]$ and the superscript W represents the NPMLE. The notation for the functions Θ^c, $c = 0, 1$, stems from frailty models where they have the meaning of the conditional expectation of the frailty variable U, given the observed data on the subject (*Tsodikov et al.*, 2020). In this context, $\Theta^0 = E(U|\text{censored})$ and $\Theta^1 = E(U|\text{failure})$. It can be shown that for a subject at risk at t, $\Theta^0(H_t|\cdot) = E\{\Theta^{\Delta}(H_X|\cdot)|X \geq t\}$, where Δ is the subject's random indicator of failure, and X is the subject's time at risk. This implies that a prediction of a subject's risk at t without knowledge of future data (Δ, X) is an average of the risk prediction with knowledge of prospective data over the subject's random future data.

The martingale transform over the future in (2.7) can be represented as

$$\int_{t^+}^{\tau} \big[\Theta^0(H_u|\boldsymbol{\beta}, \boldsymbol{Z}_{u,i}) - \Theta^1(H_u|\boldsymbol{\beta}, \boldsymbol{Z}_{u,i})\big]dM_i(u) = \Theta^0(H_t|\boldsymbol{\beta}, \boldsymbol{Z}_{t,i}) - \Theta^{\Delta_i}(H_{X_i}|\boldsymbol{\beta}, \boldsymbol{Z}_{t,i}) \quad (2.8)$$

Using the expression in (2.8), an equivalent representation of the NPMLE would be

$$\widehat{dH_t}^W = \frac{\frac{1}{n}\sum_{i=1}^{n} dN_i(t)}{\frac{1}{n}\sum_{i=1}^{n} Y_i(t)\Theta^{\Delta_i}(H_{X_i}|\boldsymbol{\beta}, \boldsymbol{Z}_{t,i})} \quad (2.9)$$

showing that future information (Δ_i, X_i) is used to predict the risk for subjects at risk at t in the denominator of the estimator. The NPMLE $\widehat{dH_t}$ can be expressed as a Weighted Breslow-type estimator, as suggested in Chen (2009), where

$$\begin{aligned} \widehat{dH_t}^W &= \frac{\frac{1}{n}\sum_{i=1}^n dN_i(t)}{\frac{1}{n}\sum_{i=1}^n Y_i(t)\Theta^0(H_t|\boldsymbol{\beta}, \boldsymbol{Z}_{t,i})w_i(H_t|\boldsymbol{\beta}, \boldsymbol{Z}_{t,i})} \\ w_i(H_t|\boldsymbol{\beta}, \boldsymbol{Z}_{t,i}) &= 1 - \frac{\int_{t^+}^{\tau}\left[\Theta^0(H_u|\boldsymbol{\beta}, \boldsymbol{Z}_{u,i}) - \Theta^1(H_u|\boldsymbol{\beta}, \boldsymbol{Z}_{u,i})\right]dM_i(u)}{\Theta^0(H_t|\boldsymbol{\beta}, \boldsymbol{Z}_{t,i})} \\ &= \frac{\Theta^{\Delta_i}(H_{X_i}|\boldsymbol{\beta}, \boldsymbol{Z}_{t,i})}{\Theta^0(H_t|\boldsymbol{\beta}, \boldsymbol{Z}_{t,i})} \end{aligned} \tag{2.10}$$

Based on the form in equation (2.10), we refer to the NPMLE as the Weighted estimator. Note that the weight at time t in (2.10) depends on future values of H. Given $\mathcal{F}_{t^-}$, the expected value of the weight is one (Appendix A.1). Using $\hat{H}$ obtained from integrating $\widehat{dH}$ in (2.10) with respect to time, which we will denote by $\hat{H}_t^W$, we can derive the estimator of $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}^W$, as the solution to the profile likelihood score

$$0 = U_{\boldsymbol{\beta}} = \sum_{i=1}^n \int_0^{\tau} \frac{\partial}{\partial \boldsymbol{\beta}} \ln(\Theta^0(\hat{H}_u(\boldsymbol{\beta})|\boldsymbol{\beta}, \boldsymbol{Z}_{u,i}))dM_i(u) \tag{2.11}$$

Derivations of the profile likelihood score equations for $\boldsymbol{\beta}$ and dH_t are given in Appendix A.2.

2.2.4 Estimating Equations

An alternative estimator for H can be obtained via a martingale estimating equation. The form of the estimating equation is based on the martingale property $E(dN_i(t)|\mathcal{F}_{t^-}) = Y_i(t)\Theta^0(H_t|\boldsymbol{\beta}, \boldsymbol{Z}_{t,i})dH_t$. This leads to

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n dN_i(t) &= \frac{1}{n}\sum_{i=1}^n Y_i(t)\Theta^0(H_t|\boldsymbol{\beta}, \boldsymbol{Z}_{t,i})dH_t \\ \implies \widehat{dH_t}^B &= \frac{\frac{1}{n}\sum_{i=1}^n dN_i(t)}{\frac{1}{n}\sum_{i=1}^n Y_i(t)\Theta^0(H_t|\boldsymbol{\beta}, \boldsymbol{Z}_{t,i})} \end{aligned} \tag{2.12}$$

where the superscript B denotes the estimating equations or Breslow estimator. The estimating equations estimator of H is referred to as the Breslow estimator. Previous work has shown this estimator to be consistent for H, yet it is not believed to be asymptotically efficient (*Chen*, 2009). Using $\hat{H}$ obtained by integrating $\widehat{dH}$ over time in (2.12), which we will denote by $\hat{H}_t^B$, we can obtain the estimator of $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}^B$, via equation (2.11). Note that the Breslow estimator is equivalent to the Weighted Breslow NPMLE with all weights w_i set to one.

2.2.5 Estimation Algorithm for NPMLE

The algorithm for the NPMLE $\hat{H}^W$ outlined below is adapted from the one proposed in Chen (2009). At the first step, assume $\boldsymbol{\beta}$ is known, all weights $w_i^{(0)}$ are set to one, and $\hat{H}^{W,(0)}$ can be computed from the Nelson-Aalen estimator given in equation (2.13).

$$\begin{aligned} \widehat{dH_t}^{NA} &= \frac{\frac{1}{n}\sum_{i=1}^n dN_i(t)}{\frac{1}{n}\sum_{i=1}^n Y_i(t)} \\ \hat{H}_t^{NA} &= \int_0^t \frac{\frac{1}{n}\sum_{i=1}^n dN_i(s)}{\frac{1}{n}\sum_{i=1}^n Y_i(s)} \end{aligned} \tag{2.13}$$

The following steps are repeated until convergence is achieved, meaning that the difference between consecutive estimators is less than some pre-specified tolerance ξ:

1. For fixed weights $w_i^{(k)}$, obtain $\hat{H}^{W,(k+1)}$ from equation (2.10) by estimating the set of jumps $\{dH\}$ at the observed failure times. This is done by solving an equation for each jump.

2. Update $w_i^{(k+1)}$ in equation (2.10) using $\hat{H}^{W,(k+1)}$.

Using this algorithm in combination with profile maximum likelihood estimation allows us to obtain estimates of $\boldsymbol{\beta}$ and H. For estimation of $\hat{H}^B$, this same algorithm

can be used, except now we only execute step 1. Note that because $\hat{H}^B$ relies only on current event information, a recurrent algorithm is used.

2.2.6 Relative Efficiency of the Breslow Estimator

Recall that our two estimators of H are

$$\begin{aligned}\hat{H}_t^B &= \int_0^t \widehat{dH_s}^B = \int_0^t \frac{\frac{1}{n}\sum_{i=1}^n dN_i(s)}{\frac{1}{n}\sum_{i=1}^n Y_i(s)\Theta^0(H_s|\boldsymbol{\beta}, \boldsymbol{Z}_{s,i})} \\ \hat{H}_t^W &= \int_0^t \widehat{dH_s}^W = \int_0^t \frac{\frac{1}{n}\sum_{i=1}^n dN_i(s)}{\frac{1}{n}\sum_{i=1}^n Y_i(s)\Theta^0(H_s|\boldsymbol{\beta}, \boldsymbol{Z}_{s,i})w_i(H_s|\boldsymbol{\beta}, \boldsymbol{Z}_{s,i})}\end{aligned} \tag{2.14}$$

where $\hat{H}_t^B$ refers to the Breslow estimator and $\hat{H}_t^W$ refers to the Weighted Breslow estimator or NPMLE. We explore the efficiency of the Breslow estimator using martingale properties. We assume that the standard regularity conditions used to show consistency and asymptotic normality of both estimators $\hat{H}_t^B$ and $\hat{H}_t^W$ hold (*Andersen and Gill*, 1982; *Zeng and Lin*, 2007; *Chen*, 2009). As a direct consequence of these conditions and previous arguments, we assume that martingale transforms are square integrable. Additionally, we assume that Uniform Laws of Large of Numbers and the Martingale Central Limit Theorem hold as $\widehat{dH_t}^B$ and $\widehat{dH_t}^W$ both $\xrightarrow{p} dH_t$ uniformly over t. Then,

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^n \varphi_i(t) \xrightarrow{p} E(\varphi(t)) \\ &\frac{1}{\sqrt{n}}\sum_{i=1}^n \psi_i(t) \xrightarrow{d} GP\big(0, Cov(\psi(x)\psi(y)) = E(\psi(x)\psi(y))\big), \\ &\qquad E(\psi(t)) = 0, \quad t \in [0, \infty)\end{aligned} \tag{2.15}$$

for a variety of subject-specific functions ψ, φ, and where GP denotes a Gaussian Process.

Since $\sqrt{n}(\widehat{dH_t}^B - dH_t)$ and $\sqrt{n}(\widehat{dH_t}^W - dH_t)$ have already been shown to be asymptotically Normal with finite variances (*Andersen and Gill*, 1982; *Zeng and Lin*,

2007; *Chen*, 2009), to compare asymptotic variances, it is enough to consider the normalized difference in the two estimators of the jump dH_t. For brevity $\Theta^c = \Theta^c(H_x|\boldsymbol{\beta}, \boldsymbol{Z}_x)$, with the understanding that x can be replaced with the appropriate time. First, we represent the difference between the two estimators as

$$\begin{aligned}\sqrt{n}(\widehat{dH_t}^B - \widehat{dH_t}^W) = &\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n dM_i(t)}{\frac{1}{n}\sum_{i=1}^n Y_i(t)\Theta_i^0} \cdot \frac{-\frac{1}{n}\sum_{i=1}^n \int_{t^+}^\tau [\Theta_i^0 - \Theta_i^1] dM_i(u)}{\frac{1}{n}\sum_{i=1}^n Y_i(t)\Theta_i^0 - \int_{t^+}^\tau [\Theta_i^0 - \Theta_i^1] dM_i(u)} \\ &+ \frac{-\frac{dH_t}{\sqrt{n}}\sum_{i=1}^n \int_{t^+}^\tau [\Theta_i^0 - \Theta_i^1] dM_i(u)}{\frac{1}{n}\sum_{i=1}^n Y_i(t)\Theta_i^0 - \int_{t^+}^\tau [\Theta_i^0 - \Theta_i^1] dM_i(u)}\end{aligned} \tag{2.16}$$

See Appendix A.3 for a detailed derivation of this expression. By the Law of Large Numbers, as noted above, $\frac{1}{n}\sum_{i=1}^n(\cdot)$ terms will converge to a non-random constant or function in probability. Specifically, using this fact and Lenglart's inequality,

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^n \int_{t^+}^\tau [\Theta_i^0 - \Theta_i^1] dM_i(u) \xrightarrow{p} 0 \\ &\frac{1}{n}\sum_{i=1}^n Y_i(t)\Theta_i^0 \xrightarrow{p} E_{\boldsymbol{Z}_t}\left\{S^*_{T|\boldsymbol{Z}_t}(t) G^*_{C|\boldsymbol{Z}_t}(t)\Theta^0\right\} > 0\end{aligned} \tag{2.17}$$

where $S^*_{T|\boldsymbol{Z}_t}$ and $G^*_{C|\boldsymbol{Z}_t}$ are the true survival functions of the event time and censoring time, respectively, conditional on the covariate trajectory $\boldsymbol{Z}_t$. The key term is $-\frac{dH_t}{\sqrt{n}}\sum_{i=1}^n \int_{t^+}^\tau [\Theta_i^0 - \Theta_i^1] dM_i(u)$, which is asymptotically Normal with a variance that is $O((dH_t)^2)$ (Appendix A.3). We can express this term as

$$-\frac{dH_t}{\sqrt{n}}\sum_{i=1}^n \int_{t^+}^\tau [\Theta_i^0 - \Theta_i^1] dM_i(u) = -\frac{dH_t}{\sqrt{n}}\sum_{i=1}^n X_{it} \tag{2.18}$$

While, $Var(-\frac{dH_t}{\sqrt{n}}\sum_{i=1}^n X_{it}) = O((dH(t))^2)$, we cannot say that $Var(\int_0^t -\frac{dH_x}{\sqrt{n}}\sum_{i=1}^n X_{ix}) = \int_0^t O((dH(x))^2)$ because the independent increments assumption no longer holds. This is because X_{ix} is not independent of X_{iy}, $x \neq y$, as they share $\int_{\max(x,y)}^\tau [\Theta_i^0 - \Theta_i^1] dM_i(u)$. Therefore, we will consider the relative efficiency of $\hat{H}_t^B$ in comparison to $\hat{H}_t^W$. To do so, let us derive the variances of $\sqrt{n}(\hat{H}_t^B - H_t)$ and $\sqrt{n}(\hat{H}_t^W - H_t)$.

First consider $\sqrt{n}(\hat{H}_t^B - H_t)$. For $\hat{H}_t^B$, independence of the martingale increments

will hold. Thus,

$$
\begin{aligned}
Var\big[\sqrt{n}(\hat{H}_t^B - H_t)\big] &= Var\bigg[\sqrt{n}\bigg(\int_0^t \frac{\frac{1}{n}\sum_{i=1}^n dN_i(x)}{\frac{1}{n}\sum_{i=1}^n Y_i(x)\Theta_i^0} - dH_x\bigg)\bigg] = Var\bigg(\int_0^t \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n dM_i(x)}{\frac{1}{n}\sum_{i=1}^n Y_i(x)\Theta_i^0}\bigg) \\
&= \int_0^t \frac{dH_x}{A_x},\ A_x = E_{\boldsymbol{Z}}(S^*_{T|\boldsymbol{Z}}(x)G^*_{C|\boldsymbol{Z}}(x)\Theta_x^0) \\
&= Var_B(t)
\end{aligned}
\tag{2.19}
$$

Details are provided in Appendix A.4. Now, consider $\sqrt{n}(\hat{H}_t^W - H_t)$. Then,

$$
\begin{aligned}
Var\big[\sqrt{n}(\hat{H}_t^W - H_t)\big] &= Var\bigg[\sqrt{n}\bigg(\int_0^t \frac{\frac{1}{n}\sum_{i=1}^n dN_i(x)}{\frac{1}{n}\sum_{i=1}^n Y_i(x)\Theta_i^0 - \int_{x^+}^\tau [\Theta_i^0 - \Theta_i^1] dM_i(u)} - dH_x\bigg)\bigg] \\
&= Var\bigg(\int_0^t \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n dM_i(x) + \frac{dH_x}{\sqrt{n}}\sum_{i=1}^n \int_{x^+}^\tau [\Theta_i^0 - \Theta_i^1] dM_i(u)}{\frac{1}{n}\sum_{i=1}^n Y_i(x)\Theta_i^0 - \int_{x^+}^\tau [\Theta_i^0 - \Theta_i^1] dM_i(u)}\bigg) \\
&= Var\bigg(\int_0^t \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n dM_i(x) - \frac{dH_x}{\sqrt{n}}\sum_{i=1}^n X_{ix}}{\frac{1}{n}\sum_{i=1}^n Y_i(x)\Theta_i^0}\bigg) + o_p(1)
\end{aligned}
\tag{2.20}
$$

The last line of (2.20) has an $o_p(1)$ by noting that the second term in the denominator $\xrightarrow{p} 0$. Let $dV_t = \frac{1}{\sqrt{n}}\sum_{i=1}^n dM_i(t) - X_{it}dH_t$, where $dM_i(t) \perp X_{it}$ by the independence of martingale increments. Then,

$$
\begin{aligned}
\sqrt{n}(\hat{H}_t^W - H_t) &= \int_0^t \frac{dV_x}{\frac{1}{n}\sum_{i=1}^n Y_i(x)\Theta_i^0} + o_p(1) \\
Var\big[\sqrt{n}(\hat{H}_t^W - H_t)\big] &= \int_0^t \int_0^t \frac{Cov(dV_x, dV_y)}{A_x A_y} + o_p(1)
\end{aligned}
\tag{2.21}
$$

Please see Appendix A.4. for further details. By the independence of subjects,

$$
\begin{aligned}
Cov(dV_x, dV_y) &= E(dV_x dV_y) = \frac{1}{n}\sum_{i=1}^n E\big\{(dM_i(x) - X_{ix}dH_x)(dM_i(y) - X_{iy}dH_y)\big\} \\
&= \frac{1}{n}\sum_{i=1}^n E(dM_i(x)dM_i(y)) - E(dM_i(x)X_{iy}dH_y) - E(dM_i(y)X_{ix}dH_x) + E(X_{ix}X_{iy}dH_x dH_y) \\
&= A_x dH_x I(x=y) - B_x dH_x dH_y I(x>y) - B_y dH_x dH_y I(x<y) + dH_x dH_y \int_{\max(x,y)}^t B_a dH_a, \\
B_x &= E_{\boldsymbol{Z}}([\Theta_x^0 - \Theta_x^1]^2 S^*_{T|\boldsymbol{Z}}(x)G^*_{C|\boldsymbol{Z}}(x)\Theta_x^0)
\end{aligned}
\tag{2.22}
$$

The expressions for each term in the second line of (2.22) are given in Appendix A.4. Then, applying the definition of variance in (2.21) and using symmetry of x and y,

$$Var\big[\sqrt{n}(\hat{H}_t^W - H_t)\big] = Var_B(t) - \int_0^t \left[B_x - \int_x^t B_a dH_a\right] dVar_B^2(x) + o_p(1) \qquad (2.23)$$

where $dVar_B^2(x) = 2\frac{dH_x}{A_x} Var_B(x) = 2\frac{dH_x}{A_x}\int_0^x \frac{dH_y}{A_y}$. Therefore, ignoring the $o_p(1)$ terms, the asymptotic relative efficiency (ARE) of the Breslow estimator will be

$$\begin{aligned} ARE_B(t) &= \frac{Var\big[\sqrt{n}(\hat{H}_t^B - H_t)\big]}{Var\big[\sqrt{n}(\hat{H}_t^W - H_t)\big]} = \frac{Var_B(t)}{Var_B(t) - \int_0^t \big[B_x - \int_x^t B_a dH_a\big] dVar_B^2(x)} \\ &= \frac{1}{1 - \frac{\int_0^t [B_x - \int_x^t B_a dH_a] dVar_B^2(x)}{Var_B(t)}} \end{aligned} \qquad (2.24)$$

The Breslow estimator is not expected to be efficient when the integral term in the denominator is positive. The asymptotic distribution and efficiency of the NPMLE of $\boldsymbol{\beta}$ and its profile-likelihood variance are established in Appendix A.5. Comparison of the asymptotic variances for the corresponding $\boldsymbol{\beta}$ estimators, $\hat{\boldsymbol{\beta}}^W$ and $\hat{\boldsymbol{\beta}}^B$, is given in Appendix A.6, using both a profile likelihood argument and an argument similar to Zucker (2005).

2.3 Simulation Study

We studied the behavior of the ARE of the Breslow estimator under various settings. We consider a single time-independent covariate Z, where $Z \sim Bernoulli(p)$, $p \in \{0.3, 0.5, 0.7\}$. The baseline cumulative hazard function is assumed to be Weibull, with either $H_t = \frac{t^2}{4}$ (increasing hazards) or $H_t = \sqrt{\frac{t}{2}}$ (decreasing hazards). The true survival function of the failure time T given covariate Z was assumed to be the proportional odds model in Tsodikov (2003), with frailty $U \sim Exp(e^{\beta Z})$ and $\beta \in \{-1, 0, 1\}$.

Details of the model are given in the Appendix A.7. The true survival function of censoring C was assumed to be the uniform survival function $1 - \frac{x}{\tau}$, for $0 < x < \tau$ and $\tau \in \{5, 10, 15\}$. A summary of all simulation settings is provided in Table A.1. Calculation of the ARE in the simulation settings is detailed in Appendix A.8.

Results of the simulations are given in Table 2.1 and Figures 2.1-2.6. Table 2.1 displays summary statistics of the AREs across time. The Breslow estimator performs worst on average in Settings 10-12, where $\beta = -1$, the baseline hazard is decreasing, and $\tau = 5$. The Breslow estimator performs best on average in Settings 40-45, where $\beta = 0$ or $\beta = 1$, the baseline hazard is increasing, and $\tau = 15$. From Figures 2.1-2.6, certain trends can be seen. The NPMLE is upwards of 150% more efficient than the Breslow estimator early on in settings where $\beta = -1$ and $p = 0.7$. When $\beta = 0$, the NPMLE is at most 30% more efficient than the Breslow estimator, and when $\beta = 1$, the NPMLE is at most 25% more efficient than the Breslow estimator (Table 2.1). For settings with increasing hazards, the Breslow estimator actually becomes slightly more efficient after $t = 5$ when $\tau = 10$ or $\tau = 15$. In settings with decreasing hazards, the Breslow estimator is never more efficient than the NPMLE. In each setting, the NPMLE is more efficient early on, but as time progresses, the Breslow estimator becomes more efficient. The differences in the denominators of the two estimators may be more prominent early on because there is more information contained in the future, but as the study progresses, the amount of future data decreases and becomes less important.

It may be useful to know how much more efficient the NPMLE is than the Breslow estimator at some interim analysis time. Assuming this time is 0.5τ, when $\tau = 5$, the NPMLE is between 13% and 55% more efficient than the Breslow estimator at $t = 2.5$. When $\tau = 10$, the NPMLE is between 1% and 37% more efficient than the Breslow estimator at $t = 5$. When $\tau = 15$ and the hazards are increasing, the NPMLE is actually 3% less efficient than the Breslow estimator at $t = 7.5$. When the

hazards are decreasing, the NPMLE is between 10% and 29% more efficient at $t = 7.5$.

Table 2.1: Summary statistics for AREs across all times in each setting. Please refer to Table A.1 for a description of each simulation setting.

Setting	**Min**	**Median**	**Mean**	**Max**	**ARE at** 0.5τ
1	1.01	1.21	1.28	1.69	1.30
2	1.01	1.24	1.37	2.04	1.35
3	1.02	1.27	1.47	2.44	1.40
4	1.00	1.13	1.14	1.28	1.22
5	1.00	1.13	1.14	1.28	1.22
6	1.00	1.13	1.14	1.28	1.22
7	1.00	1.11	1.12	1.23	1.18
8	1.00	1.10	1.10	1.19	1.15
9	1.00	1.08	1.08	1.14	1.12
10	1.12	1.38	1.41	1.73	1.38
11	1.13	1.46	1.53	2.12	1.46
12	1.15	1.55	1.66	2.58	1.55
13	1.10	1.24	1.22	1.28	1.24
14	1.10	1.24	1.22	1.28	1.24
15	1.10	1.24	1.22	1.28	1.24
16	1.08	1.20	1.19	1.24	1.20
17	1.07	1.16	1.16	1.20	1.17
18	1.06	1.13	1.12	1.15	1.13
19	0.98	1.03	1.14	1.72	1.03
20	0.98	1.03	1.19	2.10	1.03
21	0.98	1.03	1.25	2.54	1.04
22	0.98	1.02	1.07	1.29	1.03
23	0.98	1.02	1.07	1.29	1.03
24	0.98	1.02	1.07	1.29	1.03
25	0.98	1.02	1.06	1.24	1.02
26	0.98	1.01	1.05	1.20	1.02
27	0.98	1.01	1.03	1.15	1.01
28	1.08	1.28	1.33	1.74	1.28
29	1.08	1.33	1.41	2.13	1.33
30	1.09	1.37	1.50	2.60	1.37
31	1.07	1.21	1.20	1.29	1.21
32	1.07	1.21	1.20	1.29	1.21
33	1.07	1.21	1.20	1.29	1.21
34	1.06	1.17	1.17	1.24	1.17
35	1.05	1.15	1.14	1.20	1.15
36	1.04	1.12	1.11	1.15	1.12
37	0.95	0.98	1.08	1.73	0.97
38	0.95	0.98	1.11	2.12	0.97
39	0.95	0.98	1.15	2.57	0.97
40	0.95	0.98	1.03	1.29	0.97
41	0.95	0.98	1.03	1.29	0.97
42	0.95	0.98	1.03	1.29	0.97
43	0.95	0.98	1.02	1.24	0.97
44	0.95	0.98	1.01	1.20	0.97
45	0.95	0.97	1.01	1.15	0.97
46	1.06	1.23	1.28	1.74	1.23
47	1.06	1.26	1.35	2.14	1.26
48	1.07	1.29	1.41	2.60	1.29
49	1.05	1.18	1.19	1.29	1.18
50	1.05	1.18	1.19	1.29	1.18
51	1.05	1.18	1.19	1.29	1.18
52	1.05	1.15	1.15	1.24	1.15
53	1.04	1.13	1.13	1.20	1.13
54	1.04	1.10	1.10	1.15	1.10

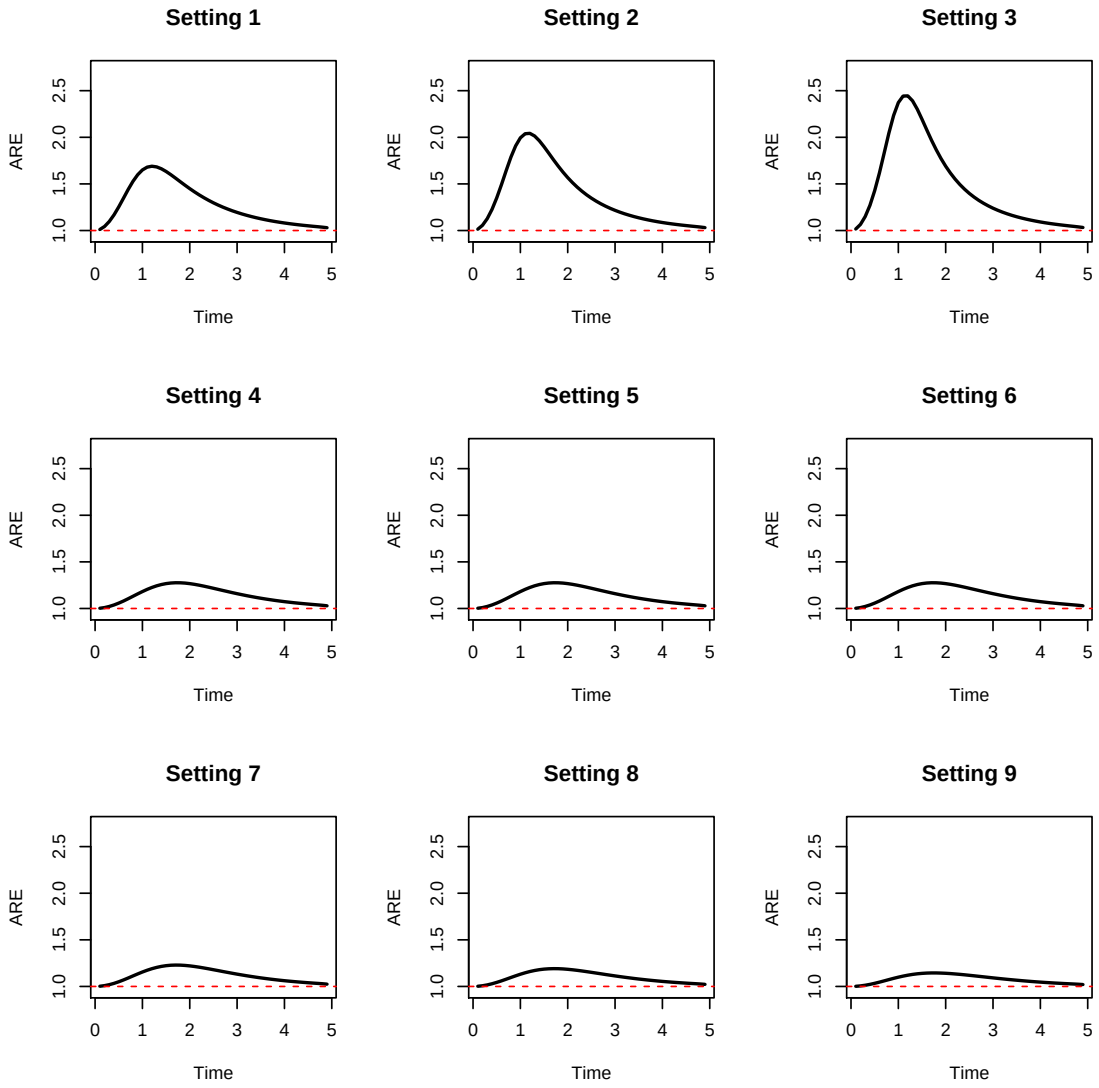


Figure 2.1: ARE for increasing hazards, $\tau = 5$.

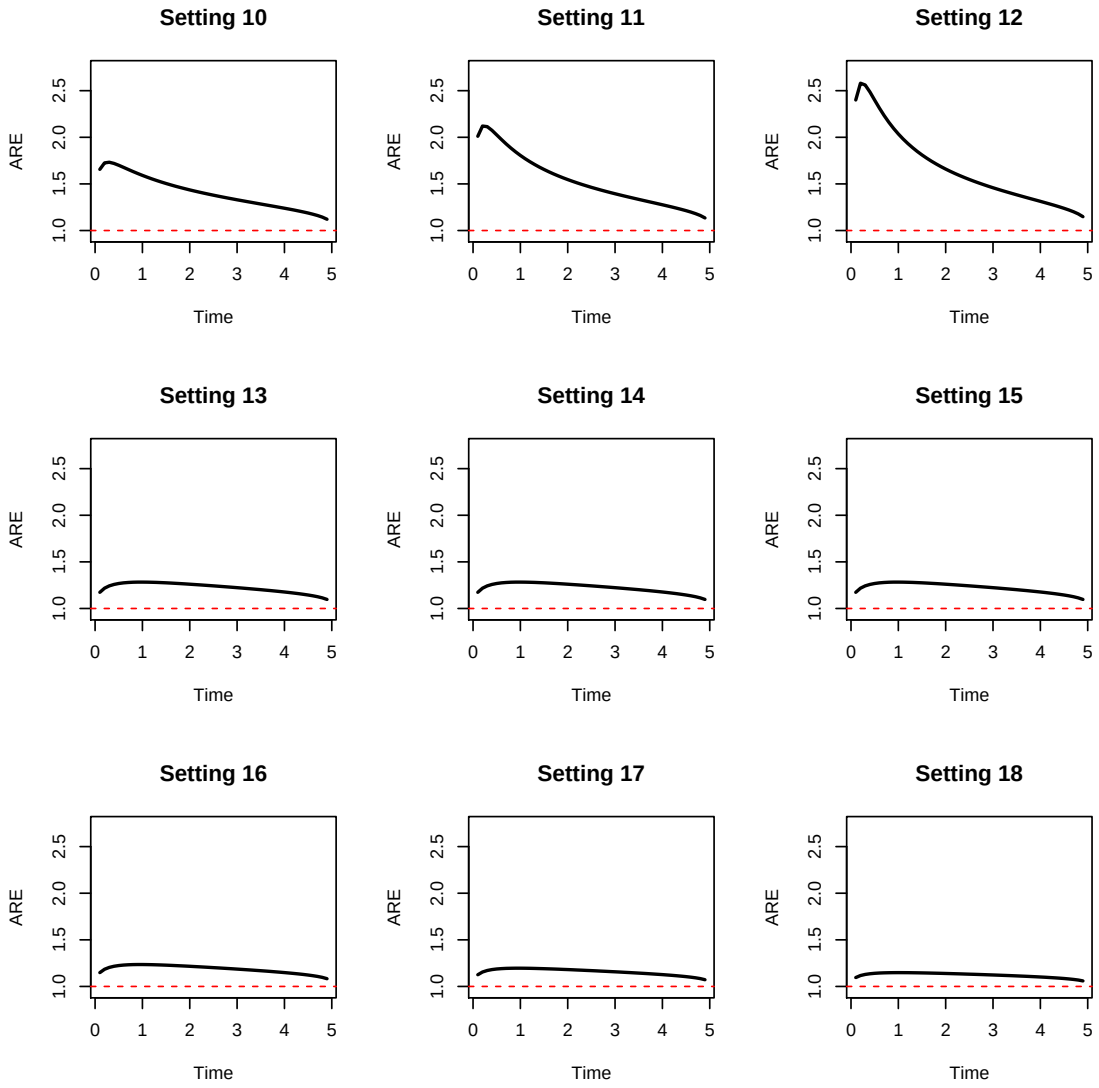


Figure 2.2: ARE for decreasing hazards, $\tau = 5$.

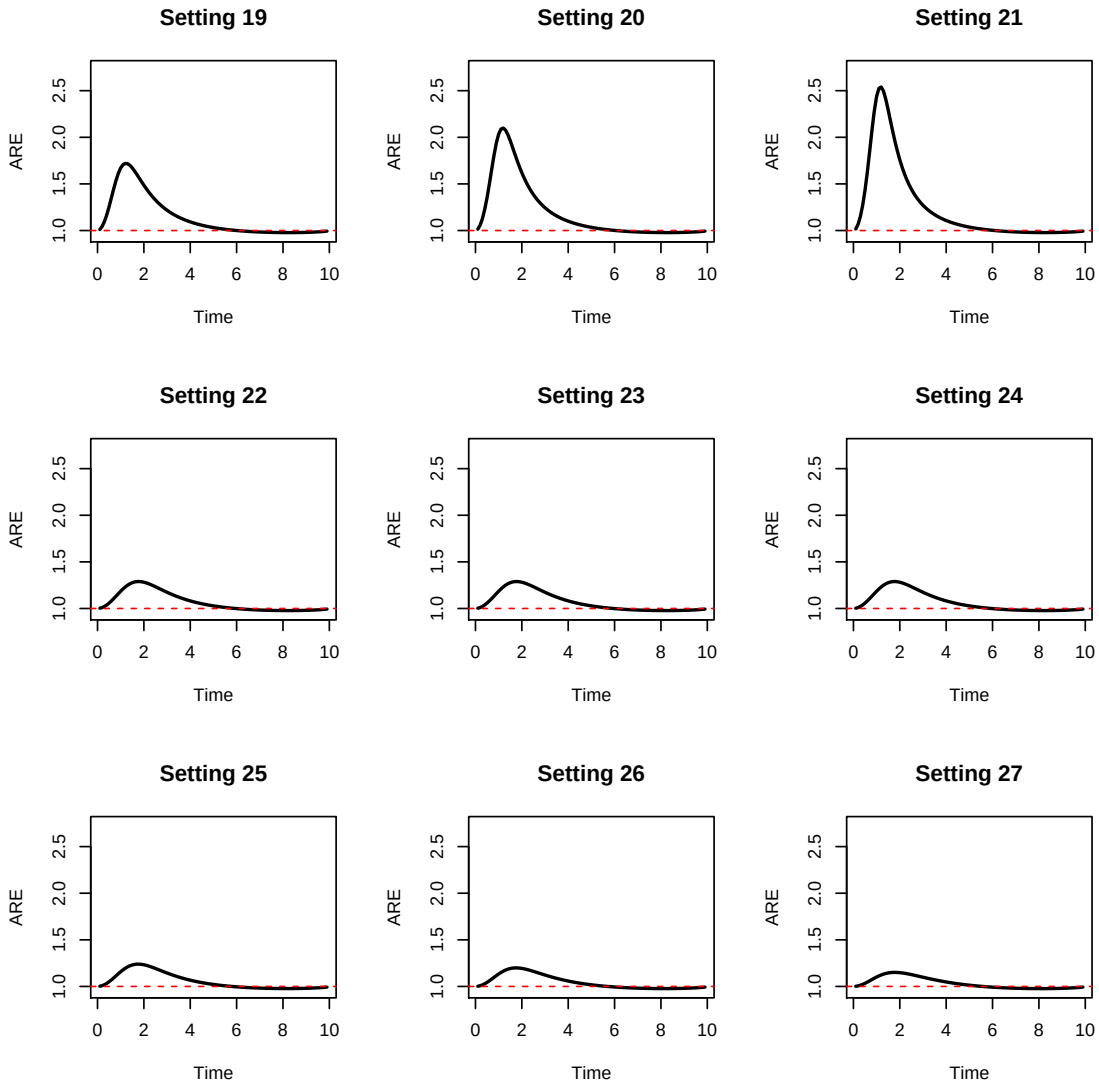


Figure 2.3: ARE for increasing hazards, $\tau = 10$.

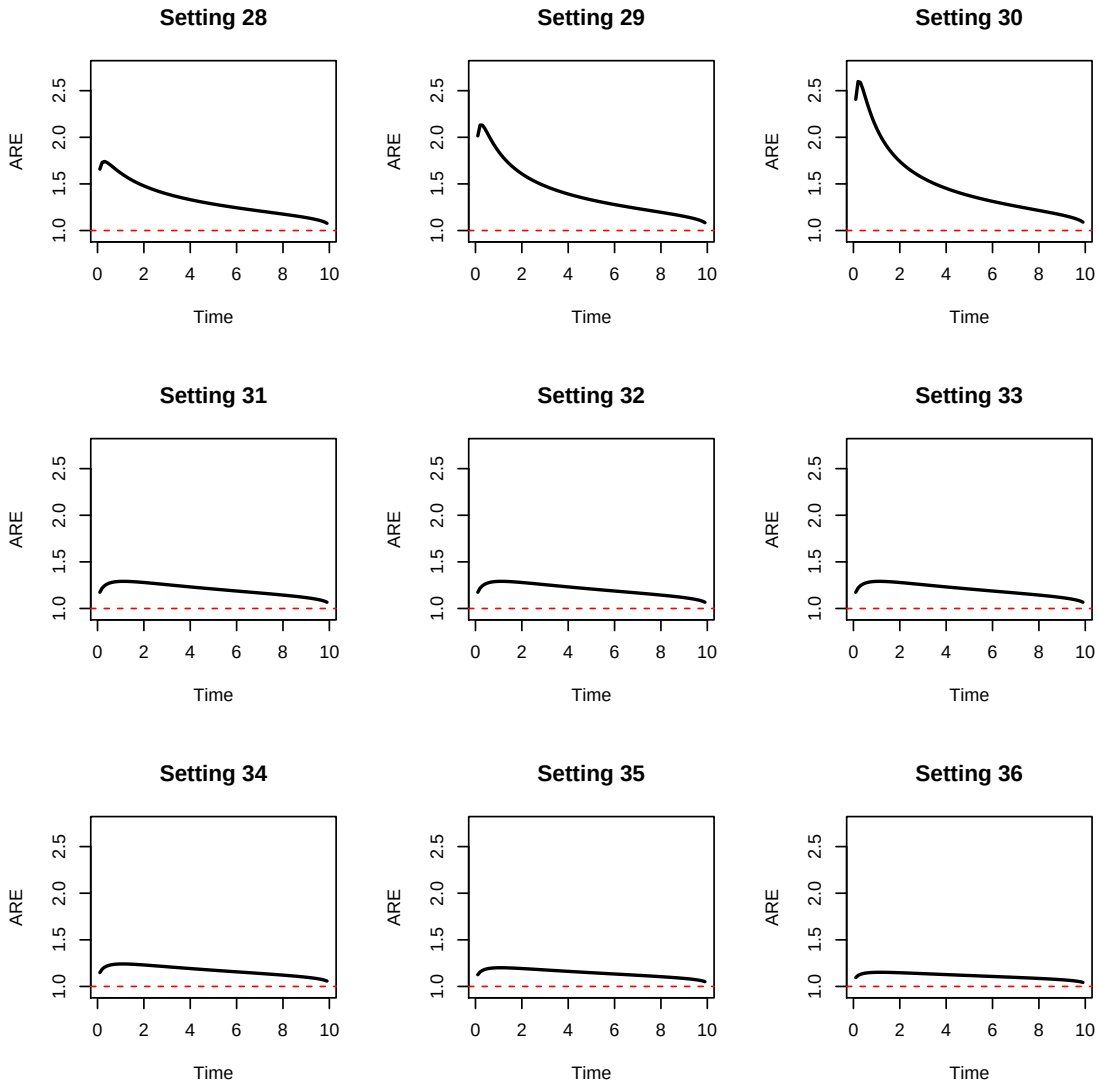


Figure 2.4: ARE for decreasing hazards, $\tau = 10$.

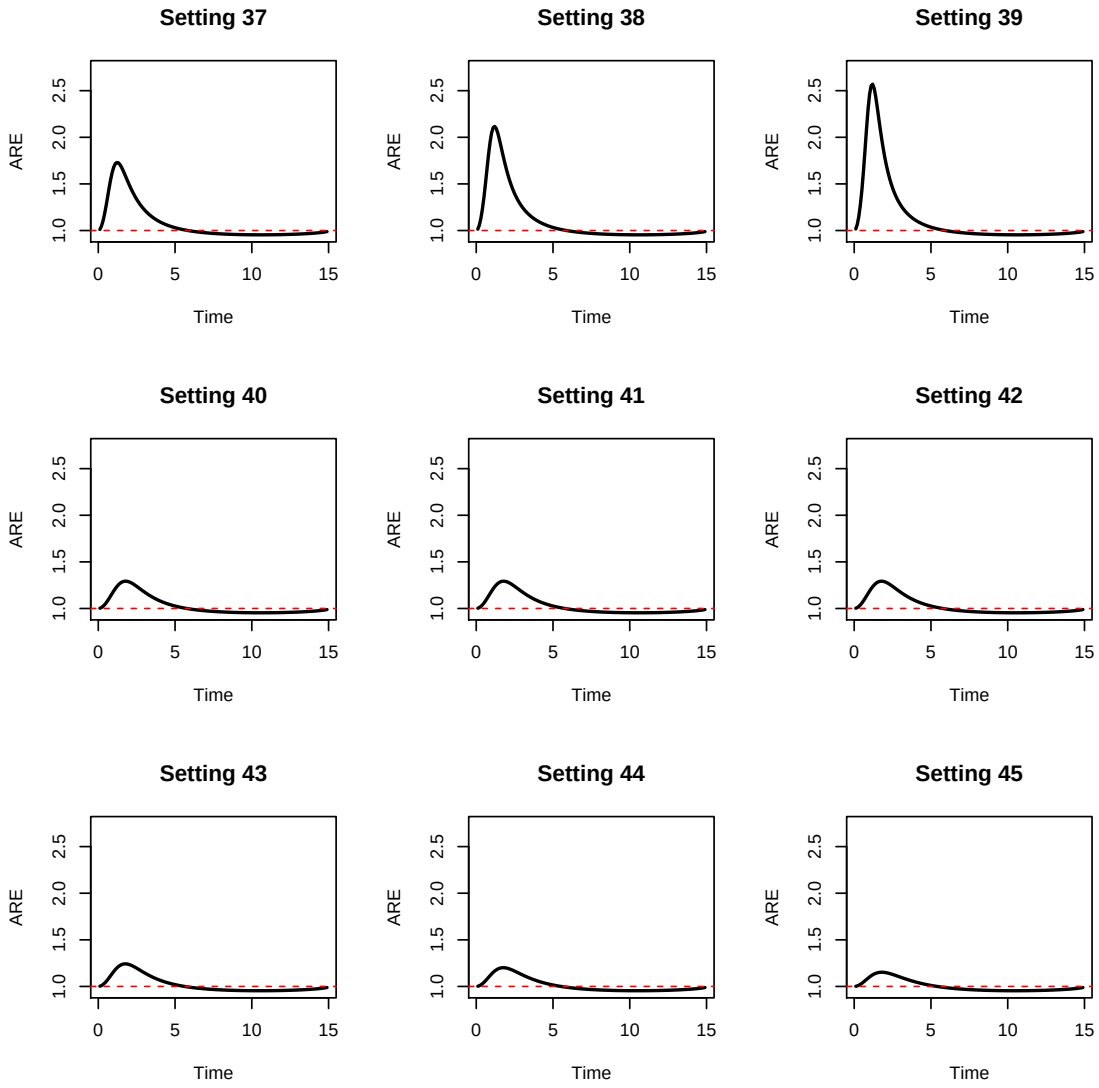


Figure 2.5: ARE for increasing hazards, $\tau = 15$.

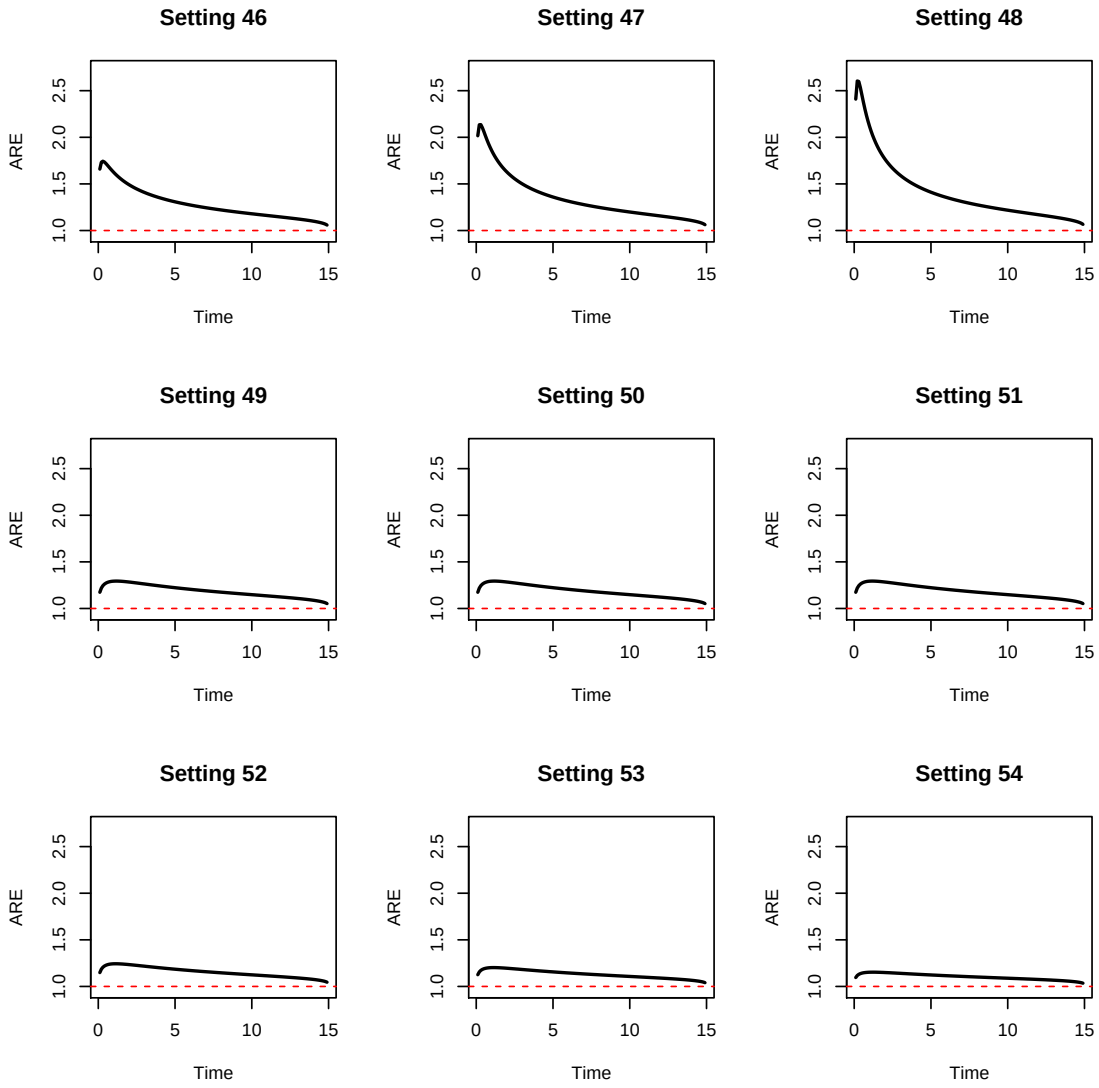


Figure 2.6: ARE for decreasing hazards, $\tau = 15$.

To assess the finite-sample properties of both estimators, we evaluated their performance in multiple simulation settings. The settings are modified from the ones used in Hu and Tsodikov (2014). We consider two time-independent covariates $\boldsymbol{Z} = (Z_1, Z_2)$, with $Z_1 \sim N(0,1)$ and $Z_2 \sim Bernoulli(0.5)$. The baseline cumulative hazard function is assumed to be Weibull, with either $H_t = \frac{t^2}{4}$ (increasing hazards) or $H_t = \sqrt{\frac{t}{2}}$ (decreasing hazards). The true failure time T was simulated from a proportional odds model as parameterized in Tsodikov (2003), where the frailty $U \sim Exp(e^{\boldsymbol{\beta Z}})$ with $\boldsymbol{\beta} = (\beta_1, \beta_2) = (-0.25, 0.75)$. Censoring time $C \sim Uniform(0, \tau = 15)$ was generated to yield approximately 25% censoring for settings with increasing baseline hazards and approximately 45% censoring for settings with decreasing baseline hazards. Sample sizes of 50, 250, and 500 were considered. For each simulation setting, 1,000 data sets were generated. A summary of all simulation settings is provided in Table 2.2.

Results of the simulations are given in Tables 2.3 and A.2 and Figures 2.7 and A.1. Figure 2.7 displays estimated and true AREs for the Breslow estimator in each of the settings outlined in Table 2.2. In the settings with increasing hazards, the Breslow estimator starts off as being less efficient but at some point it becomes more efficient than the NPMLE. Unlike the curves in Figures 2.1-2.6, the peaks of the estimated AREs tend to be at early event times, and the estimated curves do not converge to one at late event times. In the small sample settings, a few estimated ARE curves demonstrate very poor performance of the Breslow estimator, but this could be due to simulation error and small-sample bias. In the settings with decreasing hazards, the estimated curves are similar to those from the increasing hazards settings, but the Breslow estimator does not have as large of an efficiency gain at the late event times. For both increasing and decreasing hazards, the small-sample settings are reflective of the expected trends, though there is greater variation in the individual estimated curves. The estimated AREs are similar in shape to the true ARE curves in each

simulation setting.

From Table 2.3, both estimators generally result in unbiased estimates of the regression parameters, though there is small-sample bias present. In most simulation settings, the Breslow and Weighted Breslow estimators result in equal estimates of $\boldsymbol{\beta}$. For the small-sample settings, the Breslow estimator is slightly less biased than the Weighted Breslow estimator. In terms of variance, average standard errors (ASEs) and empirical standard deviations (ESDs) are in good agreement, and the two estimators are nearly identical in every scenario. Coverage probabilities for the two estimators are also close to the nominal 95% level in all simulation settings. Additional simulation results are presented in Appendix A.9.

Table 2.2: Settings used in proportional odds (PO) simulation study. n = sample size. k and λ are parameters of the Weibull baseline hazard function. τ = administrative censoring time. $Z_1 \sim N(\mu, \sigma)$ and $Z_2 \sim Bernoulli(p)$. % censoring is approximate.

Setting	n	k	λ	$\boldsymbol{\beta} = (\beta_1, \beta_2)$	τ	μ	σ	p	**% Censoring**
1	250	2	2	(-0.25, 0.75)	15	0	1	0.5	25
2	500	2	2	(-0.25, 0.75)	15	0	1	0.5	25
3	250	0.5	2	(-0.25, 0.75)	15	0	1	0.5	45
4	500	0.5	2	(-0.25, 0.75)	15	0	1	0.5	45
5	50	0.5	2	(-0.25, 0.75)	15	0	1	0.5	45
6	50	2	2	(-0.25, 0.75)	15	0	1	0.5	25

Table 2.3: Simulation results for estimator $\hat{\boldsymbol{\beta}}$ for the PO model. ASE = average standard error. ESD = empirical standard deviation. CP (%) = coverage probability of 95% Wald-based confidence interval. Weighted Breslow estimator = NPMLE.

Setting	**Estimator**	$\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2)$	$(\mathbf{ESD}_{\hat{\beta}_1}, \mathbf{ESD}_{\hat{\beta}_2})$	$(\mathbf{ASE}_{\hat{\beta}_1}, \mathbf{ASE}_{\hat{\beta}_2})$	$(\mathbf{CP}_{\beta_1}, \mathbf{CP}_{\beta_2})$
1	Breslow	(-0.25, 0.75)	(0.11, 0.25)	(0.12, 0.24)	(95.0, 93.3)
	Weighted	(-0.25, 0.75)	(0.11, 0.25)	(0.12, 0.24)	(95.1, 93.3)
2	Breslow	(-0.25, 0.77)	(0.09, 0.16)	(0.08, 0.17)	(94.2, 96.8)
	Weighted	(-0.25, 0.77)	(0.09, 0.16)	(0.08, 0.17)	(94.2, 96.7)
3	Breslow	(-0.25, 0.77)	(0.12, 0.24)	(0.12, 0.24)	(95.9, 95.3)
	Weighted	(-0.25, 0.77)	(0.12, 0.24)	(0.12, 0.24)	(95.8, 95.3)
4	Breslow	(-0.25, 0.75)	(0.08, 0.17)	(0.08, 0.17)	(95.9, 95.5)
	Weighted	(-0.25, 0.75)	(0.08, 0.17)	(0.08, 0.17)	(95.9, 95.6)
5	Breslow	(-0.27, 0.78)	(0.32, 0.56)	(0.29, 0.57)	(93.2, 96.0)
	Weighted	(-0.27, 0.79)	(0.32, 0.57)	(0.29, 0.57)	(93.1, 96.0)
6	Breslow	(-0.25, 0.80)	(0.29, 0.57)	(0.28, 0.55)	(94.8, 94.9)
	Weighted	(-0.26, 0.81)	(0.29, 0.57)	(0.28, 0.56)	(94.4, 94.9)

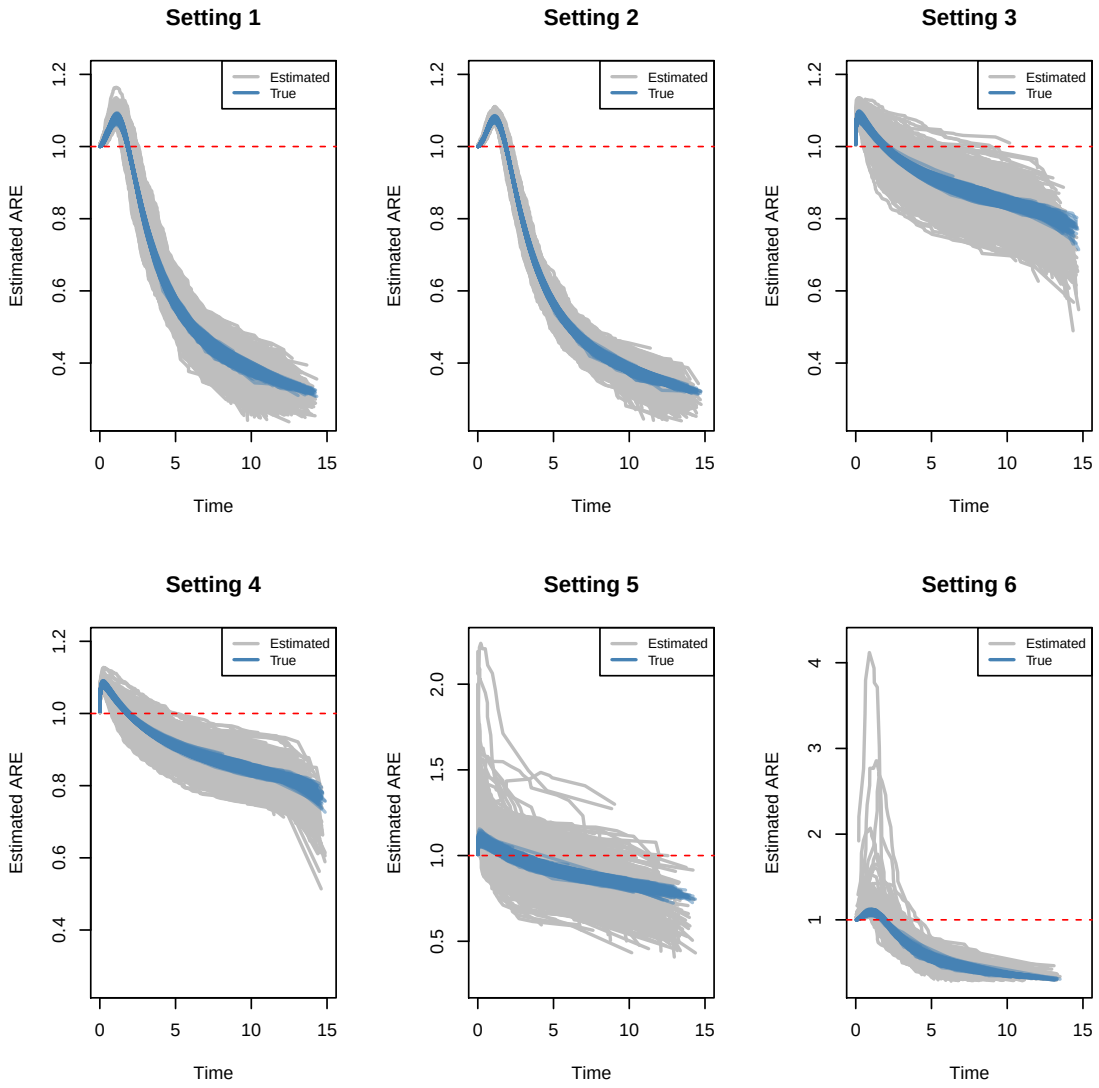


Figure 2.7: Estimated ARE for simulation settings 1-6 in Table 2.2.

2.4 SEER Data Analysis

We applied the Breslow estimator and the Weighted Breslow NPMLE to data from the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) program. Using the SEER database, a total of 19,819 cases of primary prostate cancer diagnosed in the state of Utah between 1988 and 1999 were identified (*Tsodikov*, 2003). The subset of 11,621 cases with positive survival time, valid stage of the disease, and age $\geq$ 18 years was used for analysis. Disease stage was defined as localized (stage 1), regional (stage 2), or distant (stage 3). For more details on the data and the definition of stages, please refer to SEER documentation (`http://seer.cancer.gov/`). Prostate cancer specific survival was analyzed using a proportional odds model with stage of disease as the sole covariate. We compared survival curves estimated using Breslow and Weighted Breslow estimators to observed survival summarized using the Kaplan-Meier method.

There were 9,052 (77.9%) men with localized prostate cancer, 1,713 (14.7%) men with regional prostate cancer, and 856 (7.4%) men with distant prostate cancer. Estimates of the effect of stage are given in Table 2.4. The Breslow and Weighted Breslow estimates are identical for regional and distant disease, and both methods converge in the same number of iterations. The 95% confidence intervals for both methods are identical as well, suggesting that there are negligible differences in the variances of the regression parameters. In fact, the differences between the Breslow and Weighted Breslow estimates of the effects of regional and distant stage are both < 0.00001 in magnitude. Predicted survival curves for both estimators are shown in Figure 2.8. For all stages, there are minimal differences in the estimated survival curves: curves for the Breslow and Weighted Breslow estimators appear superimposed. Furthermore, both estimated curves are close to the observed curves, suggesting that the proportional odds model is appropriate to describe observed trends in prostate cancer specific survival by disease stage: more advanced stage at diagnosis is associated

with increased risk of prostate-cancer related death.

Table 2.4: Parameter estimates of the PO model fit to prostate cancer specific survival data extracted from SEER. Weighted Breslow estimator = NPMLE. Stage refers to disease stage, which may be localized (stage 1–reference group), regional (stage 2), or distant (stage 3). OR = odds ratio. 95% Wald-based confidence interval is reported.

Method	**Stage**	ln(**OR**)	95% **CI**	$\hat{\boldsymbol{\beta}}$ **Iterations**	$\{\widehat{dH}\}$ **Iterations**
Breslow	Regional	-0.58	$(-0.76, -0.40)$	3	8
	Distant	-3.43	$(-3.61, -3.26)$		
Weighted	Regional	-0.58	$(-0.76, -0.40)$	3	8
	Distant	-3.43	$(-3.61, -3.26)$		

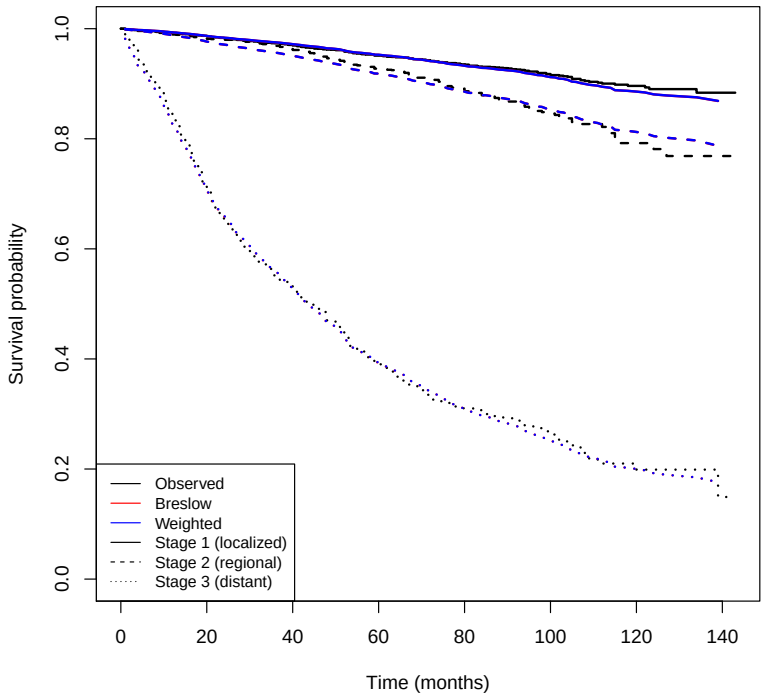


Figure 2.8: Observed and estimated survival curves by stage and estimation method. Observed survival summarized via Kaplan-Meier method. Estimated survival curves calculated using either Breslow (red) or Weighted Breslow estimator/NPMLE (blue). Disease stage indicated by line type: solid = localized/stage 1; dashed = regional/stage 2; dotted = distant/stage 3.

2.5 Discussion

We have derived the relative efficiency of the Breslow estimator of the cumulative baseline hazard for a general class of survival models. We have demonstrated via simulation that despite the theoretical inefficiency of the Breslow estimator, there are cases when it is nearly as efficient or even more efficient than the Weighted Breslow NPMLE. Through simulations and application to a large, clinically relevant data set, the Breslow estimator was shown to be virtually equivalent to the Weighted Breslow estimator in terms of the estimated regression parameters and survival curves. The practical implications of this finding are that investigators may utilize the Breslow estimator without large sacrifices in efficiency on average. This allows for faster estimation in semiparametric transformation models. While large differences were not observed in how quickly the estimators converged in simulation (Table A.2), the Breslow estimator of the regression parameter always took fewer iterations to converge on average. As the effect of the covariates on the hazard of failure is what is of greater interest, this finding is practically relevant. While the differences observed in Table A.2 are not large, the simulation scenarios considered had a maximum sample size of 500. In our analysis of prostate cancer specific survival by disease stage in SEER data, we observed virtually no differences in the Breslow and Weighted estimators: all parameter estimates were identical and the numbers of iterations to converge were equal.

The Breslow estimator relies solely on present information at any given event time, while the Weighted Breslow estimator incorporates information on future events which have not yet occurred. This makes the Weighted Breslow estimator conceptually and computationally more difficult. The dependence on the future precludes the development of recurrent algorithms for estimation of the set of jumps at event times. In contrast, the Breslow estimator can be estimated using a recurrent algorithm. Both the Breslow and Weighted Breslow estimators can be connected to EM algorithms for

semiparametric transformation models with frailty by recognizing that the denominators of both estimators involve the expectation of the latent frailty given the observed data.

While our simulations demonstrated that the Breslow estimator performed as well as the NPMLE and sometimes converged faster, we use a specific PO model for illustration. However, we expect the performance to be similar in other semiparametric transformation models, such as the proportional hazards frailty model with Gamma frailty. We also assumed that the covariates were either time-independent or external time-dependent covariates. There is a question of whether our findings will hold when the time-varying covariate is internal, considering the necessary modifications to the model and the likelihood associated with averaging hazards over the future path of the covariate process. We plan to investigate this further. We also demonstrated that for an interim analysis time equal to half the administrative censoring time, the Breslow estimator generally performed well. However, had an earlier interim analysis time been chosen, this may not have been the case. We caution investigators to carefully consider at what point it makes sense to perform such analyses. Because there is an abundance of future information available at the beginning of the study, the NPMLE may be preferred for early interim analysis.

Given that the Breslow estimator relies only on present event information, a natural extension to the current study would be to explore alternative estimators of the jump in the cumulative baseline hazard that make use of past and present event information. For example, an estimator similar in form to the Weighted Breslow estimator, except with the weight now incorporating an integral over the past rather than the future. We expect that such an estimator should also be practically equivalent to the NPMLE. This estimator is currently being explored.

CHAPTER III

Reconsidering the Role of Endogenous Covariates in Survival Models

3.1 Introduction

In many clinical studies, event information and longitudinal biomarker data are collected simultaneously. The marker of interest measures some aspect of a patient's health history over time. For example, in prostate cancer studies, time to various prostate cancer related events and prostate-specific antigen (PSA) levels at regular follow-up may be recorded for each patient. The level of the biomarker may be prognostic for disease status or severity, which motivates the use of models to capture the association between the event time and the biomarker as two manifestations of the underlying disease process.

To understand the impact of the prior marker history on the failure process, survival models incorporating the longitudinal biomarker data are needed. Because the marker is rarely observed continuously, the marker trajectory is only partially observed at discrete points in time. This implies that a model for the biomarker is also needed. The models for the event time conditional on the marker and the marker itself are often linked using shared latent factors, and this approach is known as joint modeling (*Henderson et al.*, 2000; *Tsiatis and Davidian*, 2004; *Rizopoulos*, 2008). In

the context of prostate cancer, frequentist and Bayesian joint models have been used to study the association between PSA and prostate cancer progression (*Pauler and Finkelstein*, 2002; *Law et al.*, 2002; *Yu et al.*, 2004, 2008; *Proust-Lima and Taylor*, 2009; *Taylor et al.*, 2013; *Serrat et al.*, 2015) and overall survival (*Desmée et al.*, 2015, 2017a,b). Joint latent class models, which assume similar marker trajectories within heterogeneous latent classes, have been used to model the association between PSA and prostate cancer onset (*Lin et al.*, 2002; *Proust-Lima et al.*, 2014).

Focusing on the survival submodel that incorporates the longitudinal biomarker as a covariate, the marker in this model constitutes a time-varying covariate, as its value may fluctuate over the course of the study. The magnitude of fluctuation may be informative for risk prediction. Two types of time-varying covariates exist: exogenous, also known as external, and endogenous, also known as internal. The covariates are intuitively defined by how they relate to the survival of the individual subject, with exogenous covariates assumed to exist outside the subject and endogenous covariates assumed to exist within the subject and directly depend on the subject's survival.

In many contexts, the biomarker is an endogenous covariate. To study the impact of the marker on the failure time, it can be included as a covariate in the model for the event time. However, naively specifying the survival at time t conditional on the marker history up to t results in an invalid survival function. Namely, while predicted survival probabilities at different points in time t_1 and t_2 can be obtained, the function used to obtain these probabilities is not continuous and cannot be used to "link" the two predictions. In contrast, if the marker were exogenous, the survival function would be valid and could be used to obtain predicted survival probability curves over a continuous time interval.

The purpose of this work is to clarify the role of endogenous markers in survival models. We focus on models where the marker is a covariate for the hazard of failure. In Section 3.2, we review current definitions for time-varying covariates, and we

formulate the difference between exogenous and endogenous covariates as a missing data problem. In Section 3.3, we review the validity of exponential forms for survival functions and discuss how it relates to time-varying covariates. We demonstrate the separation in the definition of a time-varying covariate from the validity of the exponential form of the survival function. In Section 3.4, we derive a mechanistic survival function that incorporates an endogenous covariate and is mathematically tractable. We illustrate the utility of our framework using the examples of bivariate shared frailty models in Section 3.5, semicompeting risks-type models in Section 3.6, and threshold regression models in Section 3.7. Our illustrative examples demonstrate the shortcomings of the current approach to specifying survival models that incorporate endogenous marker trajectories. We conclude with a discussion of our framework and future research directions in Section 3.8.

3.2 Time-Varying Covariates

3.2.1 Definitions

There are two types of time-varying covariates: (1) external or exogenous and (2) internal or endogenous (*Kalbfleisch and Prentice*, 2002). In practice, the difference between the two types of time-varying covariates is driven by intuition on how the covariate relates to the survival of the individual patient. Exogenous covariates are thought to exist outside the patient, while endogenous covariates exist within the patient and rely on the patient's survival, hence the alternative names of "external" and "internal." These definitions are largely based on one's philosophical view of any given covariate. Common examples of exogenous covariates include air pollution levels and daily temperature. Common examples of endogenous covariates include blood pressure, PSA in the context of prostate cancer, and CD4 cell counts in the context of HIV. While these intuitive definitions may be helpful to differentiate exogenous

and endogenous covariates conceptually, they offer little mathematically.

Mathematical definitions, however, do exist for both types of covariates (*Kalbfleisch and Prentice*, 2002). Let T represent the event time of interest. We assume that we observe a marker W whose value can change as a function of time t. Let $W(t)$ denote the value of the marker at time t and $\mathcal{W}(t) = \{W(u) : 0 \leq u \leq t\}$ represent the marker history up to time t. Then, for an exogenous W,

$$\lambda(u|\mathcal{W}(u)) = \lambda(u|\mathcal{W}(t)),\ 0 \leq u \leq t \tag{3.1}$$

where $\lambda(u|\mathcal{W}(u)) = \lim_{du \to 0} \frac{P(u<T\leq u+du|T\geq u,\mathcal{W}(u))}{du}$ represents the hazard of failure, conditional on the marker history up to u. The definition in (3.1) implies that the hazard of failure will remain the same whether we condition on the past or future trajectory of an exogenous W. Equivalently, the unobserved future trajectory at any point in time u does not impact the hazard at u and is therefore not needed. An alternative definition of an exogenous covariate can be formulated using the distribution of the trajectory of W:

$$P(\mathcal{W}(t)|\mathcal{W}(u), T \geq u) = P(\mathcal{W}(t)|\mathcal{W}(u), T = u),\ 0 < u \leq t \tag{3.2}$$

From the alternative definition in (3.2), the path of an exogenous covariate remains unchanged if the event occurs at $T = u$.

Endogenous covariates are largely defined by being "not exogenous," but mathematical definitions can be established by taking the converse of both (3.1) and (3.2):

$$\begin{aligned} \lambda(u|\mathcal{W}(u)) &\neq \lambda(u|\mathcal{W}(t)) \\ P(\mathcal{W}(t)|\mathcal{W}(u), T \geq u) &\neq P(\mathcal{W}(t)|\mathcal{W}(u), T = u),\ 0 < u \leq t \end{aligned} \tag{3.3}$$

Interpreting the first definition in (3.3), an endogenous covariate is one for which

the hazard of failure differs depending on the marker history or level of covariate information. Equivalently, the hazard depends on the unobserved future trajectory that exists beyond u. Using the second definition in (3.3), occurrence of the event at $T = u$ is informative of the future path of W. It is clear from the definitions in (3.1)-(3.3) that the main difference in the hazard functions for exogenous and endogenous W is due to the dependence on the unobserved future in the latter. Thus, the difference between the two types of covariates can be formulated as a missing data problem, where the unobserved future path of W constitutes the missing data.

3.2.2 Defining Survival Functions

The past trajectory can easily be incorporated into survival models for exogenous markers. In particular, applying the results of (3.1) and (3.2), the survival function evaluted at time t conditional on the marker history up to t for an exogenous W can be defined as

$$S(t|\mathcal{W}(t)) = e^{-\int_0^t \lambda(s|\mathcal{W}(s))ds} = e^{-\int_0^t \lambda(s|\mathcal{W}(t))ds}, \quad s \leq t \tag{3.4}$$

Defining survival functions conditional on the trajectory of an endogenous marker is more challenging as observation of the covariate depends on the survival of the subject. In fact, naively defining a survival function at t conditional on $\mathcal{W}(t)$ for an endogenous W results in an invalid survival function that is unrelated to its value at any other time s. Namely, $S(t|\mathcal{W}(t))$ and $S(s|\mathcal{W}(s))$, $s \neq t$, are two disjoint predicted probabilities that are not linked by a continuous function. Thus, $S(\cdot|\mathcal{W}(\cdot))$ is not a valid survival function for an endogenous W, but it can be used to generate dynamic predicted probabilities at various points in time that depend on the observed marker values. This approach is known as landmarking, and it allows for dynamic risk predictions based on a patient's past marker history (*Jewell and Nielsen*, 1993;

van Houwelingen and Putter, 2012; *Suresh et al.*, 2017).

Furthermore, for an endogenous marker, $S(t|\mathcal{W}(t))$ is not a mathematically tractable function. When conditioning on $\mathcal{W}(t)$, it is assumed to be fixed for the purposes of prediction. In reality, $\mathcal{W}(t)$ is a function of the argument t of S. This results in a contradiction as $\mathcal{W}(t)$ is actually dynamic with t. To resolve this inherent contradiction, we can consider conditioning on the marker history at some time x that is unrelated to t. Note that x is not restricted to be less than t. Thus if we consider $x > t$, we can specify the survival model conditional on future values of W that, by definition, are related to the trajectory at time t. Such ideas are explored further in Section 3.4.

3.3 Exponential Form of Survival

3.3.1 Definition

For an absolutely continuous event time T, the survival function is directly related to the hazard function through the exponential form:

$$S(t) = e^{-\int_0^t \lambda(u)du} \tag{3.5}$$

The relationship in (3.5) will hold in the presence of covariates $\boldsymbol{Z}$ assuming that $T|\boldsymbol{Z}$ remains absolutely continuous. For time-varying covariates, the exponential form of the survival function is thought to be satisfied for all exogenous markers. The observation of an endogenous marker relies on the survival of the subject, thus it is commonly thought that for such markers, the exponential relationship will be violated when we condition on the past history up to the time of survival prediction, but it will be satisfied if we condition on a fixed marker history up to and beyond the point of prediction.

3.3.2 Validity

The validity of the exponential form of the survival function conditional on a given marker history $\mathcal{W}$ is dependent on the continuity of $T|\mathcal{W}$. If T is a sudden failure time, then the form in (3.5) will be satisfied. A sudden failure time is one that cannot be predicted or fully explained given the full history $\mathcal{W}(\infty)$, which represents the trajectory of the marker W over all time. In most settings, it is reasonable to assume that T is sudden and W does not fully explain the failure process. T being a sudden failure time can be inferred if T is a totally inaccessible stopping time (*Aven and Jensen*, 1999; *Finkelstein*, 2004) and if no trajectory of W induces discreteness in T.

If the trajectory $\mathcal{W}(t)$ for an endogenous marker provides knowledge on the occurrence of T, the exponential form will be invalid due to the dicreteness that has been introduced on T. Thus, the validity of the exponential form of the survival function is a consequence of the continuity of T and not whether the marker is endogenous. While conditioning on the marker trajectory may induce discreteness in T, there is no one-to-one association between the type of time-varying covariate and the legitimacy of the exponential form in (3.5). In fact, it is possible to define tractable survival functions for endogenous markers that satisfy the exponential relationship in (3.5), as shown in Section 3.4.

3.4 Valid Survival Functions for Endogenous Markers

To define a valid survival function satisfying (3.5), let W be a marker process and W_t represent the value of W at time t. $\bar{W}_a^b$ will denote the trajectory of W between a and b, i.e., $\bar{W}_a^b = \{W_t\}|_{t=a}^{t=b}$. For brevity, we define $\bar{W}_t = \bar{W}_0^t$ and $\bar{W} = \bar{W}_0^\infty$. Let $\mathcal{F}_t^W$ and $\mathcal{F}^W$ denote the filtrations induced by $\bar{W}_t$ and $\bar{W}$, respectively, where the filtration $\mathcal{F}_s^W$ for some time s captures all past event, marker, and covariate information up to time s^-.

Regardless of whether W is endogenous or exogenous, mathematically, the exponential relationship

$$S(t|\mathcal{F}^W) = \exp\Big\{ -\int_0^t \lambda(x|\mathcal{F}^W)dx \Big\} \tag{3.6}$$

is satisfied as long as S and λ are conditional on the full information $\mathcal{F}^W$. Many survival models are specified conditional only on the past information $\mathcal{F}_t^W$. Under this level of conditioning, the exponential relationship in (3.6) will not be satisfied unless W is an exogenous marker as $S(t|\mathcal{F}_t^W) = S(t|\mathcal{F}^W)$. For endogenous W, the relationship will be satisfied conditional on fixed $\mathcal{F}_x^W$, where x can be $< t$ or $> t$ and including $x = \infty$.

Applying the Law of Total Expectation, a valid survival function conditional on fixed $\mathcal{F}_x^W$ for endogenous W can be derived as

$$S(t|\mathcal{F}_x^W) = S(t|\bar{W}_x) = E\big\{S(t|\mathcal{F}^W)\big|\bar{W}_x\big\} = E\Big\{ \exp\Big\{ -\int_0^t \lambda(u|\bar{W})du \Big\}\Big|\bar{W}_x\Big\} \tag{3.7}$$

In (3.7), the expectation is taken over the future trajectory $\bar{W}_{x^+}^{\infty}$ conditional on the past history $\bar{W}_x$. Similar expressions can be derived for the density and hazard functions:

$$\begin{aligned} f(t|\mathcal{F}_x^W) = f(t|\bar{W}_x) &= E\big\{f(t|\mathcal{F}^W)\big|\bar{W}_x\big\} = E\Big\{\lambda(t|\bar{W})\exp\Big\{ -\int_0^t \lambda(u|\bar{W})du \Big\}\Big|\bar{W}_x\Big\} \\ \lambda(t|\mathcal{F}_x^W) = \lambda(t|\bar{W}_x) &= \frac{f(t|\bar{W}_x)}{S(t|\bar{W}_x)} = \frac{E\Big\{\lambda(t|\bar{W})\exp\big\{ -\int_0^t \lambda(u|\bar{W})du\big\}\Big|\bar{W}_x\Big\}}{E\Big\{\exp\big\{ -\int_0^t \lambda(u|\bar{W})du\big\}\Big|\bar{W}_x\Big\}} \end{aligned} \tag{3.8}$$

Note that $f(t|\mathcal{F}_x^W) = -\frac{\partial}{\partial t}S(t|\mathcal{F}_x^W)$ and $\lambda(t|\mathcal{F}_x^W) = -\frac{\partial}{\partial t}\ln S(t|\mathcal{F}_x^W)$, but $f(t|\mathcal{F}_t^W) \neq -\frac{\partial}{\partial t}S(t|\mathcal{F}_t^W)$ and $\lambda(t|\mathcal{F}_t^W) \neq -\frac{\partial}{\partial t}\ln S(t|\mathcal{F}_t^W)$ unless W is assumed to be an exogenous

marker. For an endogenous W, $f(t|\mathcal{F}_t^W)$, $\lambda(t|\mathcal{F}_t^W)$, and $S(t|\mathcal{F}_t^W)$ are not legitimate distributions or functions for T. This implies that the distributional characteristics of T are bivariate functions of x and t when dynamic information is provided by the endogenous process as future behavior of the process at $x > t$ is informative for the process at t.

Using the quantities defined in (3.7) and (3.8), we can mechanistically define valid functions $f(t|\mathcal{F}_t^W)$, $\lambda(t|\mathcal{F}_t^W)$, and $S(t|\mathcal{F}_t^W)$ via conditional expectations and specifying functions conditional on a fixed marker history at time x:

$$
\begin{aligned}
S(t|\bar{W}_x)|_{x=t} &= E\big\{S(t|\mathcal{F}^W)\big|\bar{W}_x\big\}\big|_{x=t} = E\bigg\{\exp\bigg\{-\int_0^t \lambda(u|\bar{W})du\bigg\}\bigg|\bar{W}_x\bigg\}\bigg|_{x=t} \\
f(t|\bar{W}_x)|_{x=t} &= E\big\{f(t|\mathcal{F}^W)\big|\bar{W}_x\big\}\big|_{x=t} = E\bigg\{\lambda(t|\bar{W})\exp\bigg\{-\int_0^t \lambda(u|\bar{W})du\bigg\}\bigg|\bar{W}_x\bigg\}\bigg|_{x=t} \\
\lambda(t|\bar{W}_x)|_{x=t} &= \frac{f(t|\bar{W}_x)|_{x=t}}{S(t|\bar{W}_x)|_{x=t}} = \frac{E\bigg\{\lambda(t|\bar{W})\exp\big\{-\int_0^t \lambda(u|\bar{W})du\big\}\bigg|\bar{W}_x\bigg\}\bigg|_{x=t}}{E\bigg\{\exp\big\{-\int_0^t \lambda(u|\bar{W})du\big\}\bigg|\bar{W}_x\bigg\}\bigg|_{x=t}}
\end{aligned}
\tag{3.9}
$$

The mechanistic definitions in (3.9) are constructed by fully specifying each function conditional on $\bar{W}_x$ for a fixed x unrelated to t. Once each function has been fully specified, $x = t$ can be plugged in to represent the fact that the fixed trajectory of the marker is observed to occur at t. The definitions in (3.9) are demonstrated using illustrative examples in Sections 3.5-3.7. In addition, we use these examples to highlight the lack of direct correspondence between the type of marker process, which may be exogenous or endogenous, and the validity of the exponential form of the survival function.

3.5 Bivariate Shared Frailty Model

Consider two event times X and Y that are assumed to be conditionally independent given a random variable U, representing a shared frailty (*Hougaard*, 2000). We assume that there is no censoring on X and Y. We conceptualize X and Y as two correlated event times originating from a subject, for example time to prostate cancer progression and time to death from prostate cancer. The frailty U captures the dependence between X and Y. Under this assumption, the joint survival model conditional on U can be specified as

$$S_{X,Y|U}(x, y|U) = e^{-U[H_X(x)+H_Y(y)]} \tag{3.10}$$

where $H_X(x)$ and $H_Y(y)$ represent the cumulative hazard functions of X and Y at times x and y, respectively. To ensure identifiability of the model, the cumulative hazard functions are specified in the absence of any covariates $\boldsymbol{Z}$. As U is a latent frailty, it is unobserved in practice and must be averaged out in order to obtain the marginal joint survival function.

$$\begin{aligned} S_{X,Y}(x, y) &= E_U\big\{S_{X,Y|U}(x, y|U)\big\} = E_U\big\{e^{-U[H_X(x)+H_Y(y)]}\big\} \\ &= \mathcal{L}_U(H_X(x) + H_Y(y)) \end{aligned} \tag{3.11}$$

where $\mathcal{L}_U(v) = E\{e^{-Uv}\}$ denotes the Laplace transform of U. From the joint survival function $S_{X,Y}$, we can derive the joint density function:

$$f_{X,Y}(x, y) = \frac{\partial}{\partial x}\frac{\partial}{\partial y}S_{X,Y}(x, y) = \mathcal{L}''_U(H_X(x) + H_Y(y))h_X(x)h_Y(y) \tag{3.12}$$

where $h_X(x)$ and $h_Y(y)$ are the hazard functions of X and Y evaluated at times x and y, respectively.

3.5.1 Conditional Survival Assuming X Fully Observed

It is often of interest to derive the hazard and survival functions of Y conditional on X. Using the example stated earlier, these functions would model the risk of death from prostate cancer given prostate cancer recurrence. We assume X is fully observed and uncensored, meaning the exact event time $X = x$ is known for each subject. The conditional density and survival functions of $Y|X = x$ are

$$\begin{aligned} f_{Y|X=x}(y|X=x) &= \frac{f_{Y,X}(y,x)}{f_X(x)} = -\frac{\mathcal{L}_U''(H_X(x)+H_Y(y))}{\mathcal{L}_U'(H_X(x))}h_Y(y) \\ S_{Y|X=x}(y|X=x) &= \int_y^\infty f_{Y|X=x}(v|X=x)dv = \frac{\mathcal{L}_U'(H_X(x)+H_Y(y))}{\mathcal{L}_U'(H_X(x))} \end{aligned} \tag{3.13}$$

Detailed derivations of the functions in (3.13) are given in Appendix B.1. Because no discreteness has been induced on Y when we condition on X, the exponential form of survival will be valid, and the conditional hazard function can be written as

$$\begin{aligned} \lambda_{Y|X=x}(y|X=x) = \frac{f_{Y|X=x}(y|X=x)}{S_{Y|X=x}(y|X=x)} &= -\frac{\mathcal{L}_U''(H_X(x)+H_Y(y))}{\mathcal{L}_U'(H_X(x)+H_Y(y))}h_Y(y) \\ &= \Theta^1(H_X(x)+H_Y(y))h_Y(y) \end{aligned} \tag{3.14}$$

where Θ^1 is a functional of the cumulative hazard functions of X and Y and derivatives of the Laplace transform of the shared frailty U with respect to its argument. Θ^1 can be interpreted as the predicted risk of failure at y for a subject observed to fail at $X = x$.

3.5.2 Conditional Survival Assuming X Partially Observed

Rather than exactly observing $X = x$, we may only know that X occurred before Y. Using the example from the beginning of Section 3.5, the exact time of prostate cancer progression may be unknown, but knowledge of whether cancer progressed by a certain time t before death from prostate cancer may be available. In such a

setting, we can define a marker process $W_t = I(X \geq t)$. Then, it can easily be seen that $\bar{W} = \bar{W}_0^\infty$ is equivalent to full knowledge of X, i.e., $\bar{W} \Leftrightarrow \{X = x\}$. We now want to derive the hazard and survival functions for Y conditional on $\bar{W}_y$.

Naively applying the definition of the hazard function results in

$$\lambda_{Y|W}(y|\bar{W}_y) = \begin{cases} \lambda_{Y|X=x}(y|X=x), & x \leq y \\ \lim_{dy \to 0} \frac{P(Y \in (y, y+dy] | Y \geq y, X > y)}{dy}, & x > y \end{cases} \tag{3.15}$$

The hazard in (3.15) has discreteness introduced as its form depends on how x relates to y. In the first case when $x \leq y$, knowledge of when X occurs will be captured in $\bar{W}_y$, thus the conditional hazard is equal to the hazard function in (3.14) that assumes X is fully observed.

In the second case, where $x > y$, the conditional probability can be written as

$$\begin{aligned} P(Y \in (y, y+dy] | Y \geq y, X > y) &= \frac{P(Y \in (y, y+dy], X > y)}{P(Y \geq y, X > y)} = \frac{-\frac{\partial}{\partial y} S_{X,Y}(x,y)|_{x=y}}{S_{X,Y}(x,y)} \\ &= -\frac{\mathcal{L}'_U(H_X(x) + H_Y(y)) h_Y(y)|_{x=y}}{\mathcal{L}_U(H_X(y) + H_Y(y))} \\ &= -\frac{\mathcal{L}'_U(H_X(y) + H_Y(y))}{\mathcal{L}_U(H_X(y) + H_Y(y))} h_Y(y) \\ &= \Theta^0(H_X(x) + H_Y(y)) h_Y(y) \end{aligned} \tag{3.16}$$

where Θ^0 is a functional of the cumulative hazard functions of X and Y and the Laplace transform of U. Θ^0 represents the predicted risk of failure at y for a subject who has not yet experienced a failure event of type X by y. Therefore, the naive hazard function given $\bar{W}_y$ is

$$\lambda_{Y|W}(y|\bar{W}_y) = \begin{cases} \Theta^1(H_X(x) + H_Y(y)) h_Y(y), & x \leq y \\ \lim_{dy \to 0} \frac{\Theta^0(H_X(x) + H_Y(y)) h_Y(y)}{dy}, & x > y \end{cases} \tag{3.17}$$

Because of the discreteness introduced on Y, the survival function does not satisfy the exponential relationship and is not linked to the hazard in (3.17) through the exponential form.

To define mathematically valid survival and hazard functions that satisfy the exponential relationship, we can make use of the functions defined in (3.9). As $\bar{W} \Leftrightarrow \{X = x\}$, we can rewrite the functions in (3.9) as

$$
\begin{aligned}
S(y|\bar{W}_x)|_{x=y} &= E\big\{S(y|X=x)\big|\bar{W}_x\big\}\big|_{x=y} = E\left\{\frac{\mathcal{L}'_U(H_X(x)+H_Y(y))}{\mathcal{L}'_U(H_X(x))}\middle|\bar{W}_x\right\}\middle|_{x=y} \\
f(y|\bar{W}_x)|_{x=y} &= E\big\{f(y|X=x)\big|\bar{W}_x\big\}\big|_{x=y} = E\left\{-\frac{\mathcal{L}''_U(H_X(x)+H_Y(y))}{\mathcal{L}'_U(H_X(x))}h_Y(y)\middle|\bar{W}_x\right\}\middle|_{x=y} \\
\lambda(y|\bar{W}_x)|_{x=y} &= \frac{f(y|X=x)|_{x=y}}{S(y|X=x)|_{x=y}} = \frac{E\left\{-\frac{\mathcal{L}''_U(H_X(x)+H_Y(y))}{\mathcal{L}'_U(H_X(x))}h_Y(y)\middle|\bar{W}_x\right\}\middle|_{x=y}}{E\left\{\frac{\mathcal{L}'_U(H_X(x)+H_Y(y))}{\mathcal{L}'_U(H_X(x))}\middle|\bar{W}_x\right\}\middle|_{x=y}}
\end{aligned}
\tag{3.18}
$$

The expressions in (3.18) constitute mathematically valid functions for the failure time Y conditional on partial knowledge of X. In general, these expressions will not be equal to those in (3.13) and (3.14), which assume full observation of X. Thus, the marker W satisfies the definition of an endogenous time-varying covariate. This example illustrates that we can define a valid survival function for an endogenous covariate that can incorporate its observed history up to the point of survival prediction.

3.6 Semicompeting Risks Type Model

An alternative set of bivariate survival models that may be of interest can be conceptualized using two conditional models specifying how the event times X and Y are related: (1) Y does not depend on X until a failure of type X occurs and (2) X does not depend on Y until a failure of type Y occurs. Such a framework has been previously studied in the context of incidence, metastasis, and mortality for prostate

cancer (*Tran et al.*, 2018). This particular model setup can be thought of as a special case of semicompeting risks data (*Fine et al.*, 2001), where in model (1), Y can be likened to the "terminal event" and X can be likened to the "non-terminal" event and vice versa for model (2). For illustrative purposes, we focus on model (1) where Y does not depend on X until X occurs. An analogous derivation for model (2) can be obtained by switching the roles of X and Y in what follows.

Our setup is largely based on the models constructed in Tran et al (2018). Let $\boldsymbol{Z}_1$ be the set of baseline covariates, $\boldsymbol{Z}_2$ be the covariates related to Y, and $\boldsymbol{Z} = \boldsymbol{Z}_1 \cup \boldsymbol{Z}_2$ be the complete set of covariates. H_x and h_x represent the baseline cumulative hazard function and baseline hazard function, respectively, pertaining to X at time x. We assume that the cumulative baseline hazard for Y can be written as $H_y^{\delta_Y}$ at time y, where δ_Y is a parameter related to Y that may modify the cumulative baseline hazard function for X through some power transformation. Then, the baseline hazard function for Y is $\delta_Y H_y^{\delta_Y - 1} h_y$ at time y. Note that when $\delta_Y = 1$, we assume a common cumulative baseline hazard function for X and Y.

We can set up the hazard for X conditional on covariates $\boldsymbol{Z}$ as

$$\lambda_X(x|\boldsymbol{Z}) = h_x\big[\eta I(x \leq Y) + \tilde{\eta} I(x > Y)\big] \tag{3.19}$$

where $\eta = e^{\beta_\eta \boldsymbol{Z}}$ and $\tilde{\eta} = e^{\beta_{\eta_1} \boldsymbol{Z}_1}$. The hazard for Y conditional on observation of X and covariates $\boldsymbol{Z}$ can be specified as

$$\lambda_{Y|X}(y|X, \mathbf{Z}) = \delta_Y H_y^{\delta_Y - 1} h_y \theta \mu^{I(y \geq X)} \tag{3.20}$$

where $\theta = e^{\beta_\theta \boldsymbol{Z}_1}$ and $\mu = e^{\beta_\mu \boldsymbol{Z}_2}$. The expressions in (3.19) and (3.20) are similar to the ones expressed in Tran et al (2018). Based on (3.19) and (3.20), we can derive

the survival functions for both X and Y:

$$\begin{aligned} S_X(x|\boldsymbol{Z}) &= \exp\Big\{ -\int_0^x \lambda_X(u|\boldsymbol{Z})du \Big\} = \exp\big\{ -\eta H_y - \tilde{\eta}[H_x - H_y] \big\} \\ S_{Y|X}(y|X,\boldsymbol{Z}) &= \exp\Big\{ -\int_0^y \lambda_{Y|X}(u|X,\boldsymbol{Z})du \Big\} = \exp\big\{ -H_y^{\delta_Y}\theta\mu - H_x^{\delta_Y}\theta\bar{\mu} \big\} \end{aligned} \tag{3.21}$$

where $\bar{\mu} = 1 - \mu$. Detailed derivations for the survival functions in (3.21) can be found in Appendix B.2.

By the ideas of conditional modeling (*Arnold et al.*, 1999), the joint survival function from models (1) and (2) in the introduction of Section 3.6 should be equivalent, meaning

$$S_{Y|X}(y|X,\boldsymbol{Z}) \cdot S_X(x|\boldsymbol{Z}) = S_{X|Y}(x|Y,\boldsymbol{Z}) \cdot S_Y(y|\boldsymbol{Z}) = S_{X,Y}(x,y|\boldsymbol{Z}) \tag{3.22}$$

By Theorem 11.1 of Arnold et al (1999), we should have

$$\frac{S_{X|Y,\boldsymbol{Z}}}{S_{Y|X,\boldsymbol{Z}}} = \frac{a(x,y)}{b(x,y)} = \frac{u(x)}{v(y)} \tag{3.23}$$

where $u(x)$ and $v(y)$ are one-dimensional survival functions. Using the second survival function specified in (3.21),

$$\frac{S_{X|Y,\boldsymbol{Z}}}{S_{Y|X,\boldsymbol{Z}}} = \frac{\exp\big\{ -H_x^{\delta_X}\theta\mu - H_y^{\delta_X}\theta\bar{\mu} \big\}}{\exp\big\{ -H_y^{\delta_Y}\theta\mu - H_x^{\delta_Y}\theta\bar{\mu} \big\}} = \frac{\exp\big\{ -H_x^{\delta_X}\theta\mu + H_x^{\delta_Y}\theta\bar{\mu} \big\}}{\exp\big\{ -H_y^{\delta_Y}\theta\mu + H_y^{\delta_X}\theta\bar{\mu} \big\}} = \frac{u(x)}{v(y)} \tag{3.24}$$

It can be easily checked that $u(x)$ is a valid survival function as $u'(x) \leq 0$, $u(0) = 1$, $\lim_{x\to\infty} u(x) = 0$, and $u(x)$ is right continuous. Similarly, it can be shown that $v(y)$ is a valid survival function. Please see Appendix B.3 for further details.

3.6.1 Conditional Survival Assuming X is Fully Observed

As in the previous development and assuming there is no censoring, the survival models for X and $Y|X$ are given by

$$\begin{aligned} S_X(x|\boldsymbol{Z}) &= \exp\big\{ -\eta H_y - \tilde{\eta}[H_x - H_y]\big\} \\ S_{Y|X}(y|X, \boldsymbol{Z}) &= \exp\big\{ -H_y^{\delta_Y}\theta\mu - H_x^{\delta_Y}\theta\bar{\mu}\big\} \end{aligned} \tag{3.25}$$

Of interest is the survival function for Y conditional on observing $X = x$ which can be expressed as

$$\begin{aligned} S_{Y|X=x,\boldsymbol{Z}}(y|X = x, \boldsymbol{Z}) &= \frac{S_{Y,X=x|\boldsymbol{Z}}(y, X = x|\boldsymbol{Z})}{f_X(x|\boldsymbol{Z})} = \frac{-\frac{\partial}{\partial x}S_{Y,X|\boldsymbol{Z}}(y, x|\boldsymbol{Z})}{-\frac{\partial}{\partial x}S_{X|\boldsymbol{Z}}(x|\boldsymbol{Z})} \\ &= \left(1 + \frac{\delta_Y H_x^{\delta_Y - 1}\theta\bar{\mu}}{\tilde{\eta}}\right)\exp\big\{ -\theta[H_y^{\delta_Y}\mu + H_x^{\delta_Y}\bar{\mu}]\big\} \end{aligned} \tag{3.26}$$

where

$$\begin{aligned} -\frac{\partial}{\partial x}S_{Y,X|\boldsymbol{Z}}(y, x|\boldsymbol{Z}) &= h_x\big(\delta_Y H_x^{\delta_Y - 1}\theta\bar{\mu} + \tilde{\eta}\big)\exp\big\{ -H_y^{\delta_Y}\theta\mu - H_x^{\delta_Y}\theta\bar{\mu} - \eta H_y - \tilde{\eta}[H_x - H_y]\big\} \\ -\frac{\partial}{\partial x}S_{X|\boldsymbol{Z}}(x|\boldsymbol{Z}) &= h_x\tilde{\eta}\exp\big\{ -\eta H_y - \tilde{\eta}[H_x - H_y]\big\} \end{aligned} \tag{3.27}$$

Because X is assumed to be fully observed, the exponential form of the survival function is valid, and we can calculate the hazard function directly from the survival function.

$$\lambda_{Y|X=x,\boldsymbol{Z}}(y|X = x, \boldsymbol{Z}) = -\frac{\partial}{\partial y}\ln S_{Y|X=x,\boldsymbol{Z}}(y|X = x, \boldsymbol{Z}) = \theta\delta_Y H_y^{\delta_Y - 1}h_y\mu^{I(y \geq x)} \tag{3.28}$$

Please refer to Appendix B.4 for detailed derivations of the conditional survival and hazard functions in (3.27) and (3.28).

3.6.2 Conditional Survival Assuming X is Partially Observed

In certain contexts, we may only know whether X occurred before Y and not exactly when X occurs. In such cases, we can define a marker process $W_t = I(X \geq t)$. It can easily be seen that $\bar{W}$ is equivalent to full knowledge of X, i.e., $\bar{W} \Leftrightarrow \{X = x\}$. We now want to derive the hazard and survival functions for Y conditional on $\bar{W}_y$. Note that $\lambda_{Y|X=x,\boldsymbol{Z}}(y|X = x, \boldsymbol{Z}) = \lambda(y|\bar{W}, \boldsymbol{Z}) = \theta\delta_Y H_y^{\delta_Y - 1} h_y \mu^{I(y \geq x)}$. If only the partial trajectory of W_t is observed, the conditional hazard can be written as

$$\begin{aligned}
\lambda_{Y|\bar{W}_y,\boldsymbol{Z}}(y|\bar{W}_y, \boldsymbol{Z}) &= \begin{cases} \lambda_{Y|\boldsymbol{Z}}(y|\boldsymbol{Z}) = \theta\delta_Y H_y^{\delta_Y - 1} h_y, & x > y \\ \lambda_{Y|X=x,\boldsymbol{Z}}(y|X = x, \boldsymbol{Z}) = \theta\delta_Y H_y^{\delta_Y - 1} h_y \mu, & x \leq y \end{cases} \\
&= \theta\delta_Y H_y^{\delta_Y - 1} h_y \mu^{I(x \leq y)} \\
&= \lambda(y|X = x, \boldsymbol{Z}) \\
&= \lambda(y|\bar{W}, \boldsymbol{Z})
\end{aligned} \tag{3.29}$$

While the hazard of Y conditional on $\bar{W}_y$ and $\boldsymbol{Z}$ relies on how X relates to Y, it is equivalent in form to the hazard function conditional on the full marker trajectory $\bar{W}$. Thus, by the Kalbfleisch and Prentice definition (2002), W_t is an exogenous covariate.

In general, this bivariate survival model highlights that despite the marker W_t being defined in relation to a subject's event time X, it satisfies the mathematical definition of an exogenous covariate. Therefore, the intuitive definitions of endogenous and exogenous covariates as originating within and outside a subject, respectively, are not sufficient for this particular model. W_t clearly originates within the subject, but by the Kalbfleisch and Prentice (2002) definition, it is an exogenous covariate. By an analogous argument, the marker $V_t = I(Y \geq t)$ is also an exogenous covariate in the alternative model specification where X does not depend on Y until Y fails. This can easily be seen by switching the roles of X and Y above.

3.7 Threshold Regression Model

3.7.1 Definition

The event time T can often be interpreted as a first hitting time (FHT) of a boundary or threshold state by a stochastic process which may be latent or observed (*Ting Lee and Whitmore*, 2006). For example, the time to prostate cancer onset can be thought of as the first time PSA exceeds a latent threshold. In general, FHT models have two components: a parent stochastic process $X(t)$ and a boundary set $\mathcal{B}$. For the parent process $X(t)$, $t \in \mathcal{T}$ and $x \in \mathcal{X}$, where $\mathcal{T}$ is the time space and $\mathcal{X}$ is the state space of the process. Sample paths of the parent process can be observed or unobserved. The boundary set is assumed to be a subset of the state space, i.e., $\mathcal{B} \subset \mathcal{X}$. Then the first hitting time T of $\mathcal{B}$ is a random variable such that

$$T = \inf\{t : X(t) \in \mathcal{B}\} \tag{3.30}$$

The state $X(T) \in \mathcal{B}$ is the threshold state. As the boundary set $\mathcal{B}$ defines a stopping condition for the process, the FHT is usually a stopping time. If we define a set of filtrations $\{\mathcal{F}_t\}$ representing the past history up to t, then T is a stopping time if information in $\mathcal{F}_t$ indicates whether T has occurred by t. In the case where T is a stopping time, knowledge of the past history induces discreteness in T, and the exponential form of the survival function will be invalid because of this discreteness.

In the following sections, we consider cases where the threshold $\mathcal{B}$ is fixed or random and the stochastic marker process $X(t)$ is fully or partially observed. In each case, we determine whether the marker is endogenous or exogenous by considering the hazard functions conditional on the past and full trajectories of the marker. We also consider the validity of the exponential form in each case and illustrate the lack of a one-to-one correpondence between its legitimacy and whether a given marker is endogenous. In our examples, T represents the event time or FHT, B

represents the boundary which may be fixed or random, W_t represents a fully observed marker process, and V_t represents a partially observed marker process. We assume that W_t can be modeled via a process in the Lévy family, which is a large class of stochastic processes with stationary and independent increments (*Bertoin*, 1998; *Hoyle*, 2010). Common examples of Lévy processes include the Wiener process and the non-decreasing Gamma process. For the fully observed process W_t, we define $\bar{W}_t$ and $\bar{W}$ as in Section 3.4. For the partially observed process V_t, we can represent $\bar{V}_t$ as the collection of observed values V measured at a set of measurement times all less than t, i.e., $\bar{V}_t = \{V_{\tau_j} : \tau_j \leq t\}$. We assume that higher marker values are associated with increased risk of failure, and failure will occur once $W_t \geq B$ or $V_t \geq B$.

3.7.2 Fixed Threshold and Fully Observed Marker

When the threshold B is fixed and W_t is fully observed, we want to consider the hazard and survival functions when we observe $\bar{W}_t$ and $\bar{W}$. We assume that W_t is predictable given $\bar{W}_t$ and known given $\bar{W}$. Then,

$$\begin{aligned} S(t|\bar{W}_t) = P(T > t|\bar{W}_t) = P(W_t < B|\bar{W}_t) = \begin{cases} 1, & W_t < B \\ 0, & W_t \geq B \end{cases} \\ S(t|\bar{W}) = P(T > t|\bar{W}) = P(W_t < B|\bar{W}) = \begin{cases} 1, & W_t < B \\ 0, & W_t \geq B \end{cases} \end{aligned} \tag{3.31}$$

In (3.31), conditioning on the past or full history of the marker process induces discreteness in the survival function because at any point t, we can determine whether $W_t < B$ and whether failure has occurred. It can easily be seen that

$$\lambda(t|\bar{W}_t) = \lambda(t|\bar{W}) = \delta(W_t - B) = \begin{cases} \infty, & W_t = B \\ 0, & \text{otherwise} \end{cases} \tag{3.32}$$

where $\delta(\cdot)$ in (3.32) denotes Dirac's delta function. Therefore, W_t is an exogenous covariate by the definition provided in Kalbfleisch and Prentice (2002), but the exponential form of the survival model is not valid when B is fixed as T is a stopping time given the trajectory of W. This scenario illustrates that the exponential form can be illegitimate for an exogenous marker.

3.7.3 Fixed Threshold and Partially Observed Marker

We assume the partially observed marker V_t is observed at non-informative times $\tau_1, \ldots, \tau_k$, with the additional assumption that $V(\tau_j) = W(\tau_j)$, $j = 1, \ldots, k$, where W_t is now taken to be the latent true marker process. Some of the measurements may be taken after the event time if it is non-terminal. Let $\tau_{x^-} = \sup\{\tau_j : \tau_j \leq x\}$ represent the last measurement before x and $\tau_{x^+} = \inf\{\tau_j : \tau_j \geq x\}$ represent the first measurement after x. Then, we can define survival functions given $\bar{V}_x = \{V_{\tau_j} : \tau_j \leq x\}$ and $\bar{V} = \{V_{\tau_j}\}$ using the framework introduced in Section 3.4.

$$\begin{aligned} S(t|\bar{V}_x)|_{x=t} &= P(T > t|\bar{V}_x)|_{x=t} = 1 - E_{\bar{W}}\big\{S_W(B - V_{\tau_{x^-}})|\bar{V}_x\big\}|_{x=t} \\ S(t|\bar{V}) &= P(T > t|\bar{V}) = 1 - E_U\left\{S_U\left(\frac{B - V_{\tau_{t^-}}}{V_{\tau_{t^+}} - V_{\tau_{t^-}}}\right)\middle|\bar{V}\right\} \end{aligned} \tag{3.33}$$

Please see Appendix B.5 for more details on the survival models in (3.33). In $S(t|\bar{V})$ in (3.33), U represents a bridge stochastic process that describes the behavior of W between the two observed values $V_{\tau_{t^-}}$ and $V_{\tau_{t^+}}$. If we assume W_t can be modeled using a Lévy process, V_t will be a partially observed Lévy process and U will be a Lévy bridge process (*Hoyle*, 2010). Note that if $V_{\tau_{x^-}}|_{x=t} \geq B$ or $V_{\tau_{t^-}} \geq B$, the respective survival functions will be 0. If $V_{\tau_{t^+}} < B$, $S(t|\bar{V}) = 1$. The exponential form will be valid for both $S(t|\bar{V}_x)|_{x=t}$ and $S(t|\bar{V})$ by the arguments provided in Section 3.4. The

hazard functions are

$$\lambda(t|\bar{V}_x)|_{x=t} = \frac{\frac{\partial}{\partial t}E_{\bar{W}}\left\{S_W(B - V_{\tau_{x^-}})|\bar{V}_x\right\}}{1 - E_{\bar{W}}\left\{S_W(B - V_{\tau_{x^-}})|\bar{V}_x\right\}}\Bigg|_{x=t}$$
$$\lambda(t|\bar{V}) = \frac{\frac{\partial}{\partial t}E_U\left\{S\left(\frac{B-V_{\tau_{t^-}}}{V_{\tau_{t^+}}-V_{\tau_{t^-}}}\right)\middle|\bar{V}\right\}}{1 - E_U\left\{S\left(\frac{B-V_{\tau_{t^-}}}{V_{\tau_{t^+}}-V_{\tau_{t^-}}}\right)\middle|\bar{V}\right\}} \tag{3.34}$$

By the Kalbfleisch and Prentice (2002) definition, V_t is an endogenous covariate, but the exponential form of the survival function remains valid under the framework introduced in Section 3.4, which treats the unobserved marker trajectory as a latent variable. This example once more highlights the fact that the exponential form is not immediately violated in models that condition on the trajectory of an endogenous marker.

3.7.4 Random Threshold and Fully Observed Marker

We now assume that the threshold B is random, and we assume that its distribution is related to the trajectory of the marker W_t, i.e., $B|\bar{W}_t \sim f_B(b|\bar{W}_t)$. B is assumed to be absolutely continuous with a valid survival function that satisfies the exponential relationship. Then, the survival functions conditional on $\bar{W}_t$ and $\bar{W}$ can be expressed as

$$\begin{aligned} S(t|\bar{W}_t) &= P(T > t|\bar{W}_t) = P(W_t < B|\bar{W}_t) = S_{B|\bar{W}}(W_t) \\ S(t|\bar{W}) &= P(T > t|\bar{W}) = P(W_t < B|\bar{W}) = S_{B|\bar{W}}(W_t) \end{aligned} \tag{3.35}$$

The survival functions in (3.35) are defined in terms of the survival function of the random threshold B, and they will be identical. Thus, the hazard functions will be identical as well. The hazards can be derived through the exponential relationship

that is satisfied for the survival function of B.

$$\lambda(t|\bar{W}_t) = \lambda(t|\bar{W}) = -\frac{\partial}{\partial t} \ln S_{B|\bar{W}}(W_t) \tag{3.36}$$

Thus, by the Kalbfleisch and Prentice (2002) definition, W_t is an exogenous covariate, and the exponential relationship is satisfied. This example demonstrates that for a fully observed exogenous marker, there is a threshold regression model that satisfies the exponential relationship.

3.7.5 Random Threshold and Partially Observed Marker

We once more assume V_t is observed at non-informative times $\tau_1, \ldots, \tau_k$, with $V(\tau_j) = W(\tau_j)$, $j = 1, \ldots, k$. We let $\tau_{x^-} = \sup\{\tau_j : \tau_j \leq x\}$ represent the last measurement before x and $\tau_{x^+} = \inf\{\tau_j : \tau_j \geq x\}$ represent the first measurement after x. We also assume that $B|\bar{W}_t \sim f_B(b|\bar{W}_t)$ with a valid survival function that satisfies the exponential relationship. Then, the survival functions conditional on $\bar{V}_t$ and $\bar{V}$ can be expressed as

$$\begin{aligned} S(t|\bar{V}_x)|_{x=t} = P(T > t|\bar{V}_x)|_{x=t} &= E_{\bar{W}}\big\{S_{B|\bar{W}}(V_{\tau_{x^-}} + W_{t-\tau_{x^-}})|\bar{V}_x\big\}\big|_{x=t} \\ S(t|\bar{V}) = P(T > t|\bar{V}) &= E_U\big\{S_{B|U}(V_{\tau_{t^-}} + U_{[0,\tau_{t^+}-\tau_{t^-}]}(V_{\tau_{t^+}} - V_{\tau_{t^-}}))|\bar{V}\big\} \end{aligned} \tag{3.37}$$

Please refer to Appendix B.5 for more details. The survival functions in (3.37) are derived under the assumptions that $W_t|\bar{V}_x = V_{\tau_{x^-}} + W_{t-\tau_{x^-}}$ and $W_t|\bar{V} = V_{\tau_{t^-}} + U_{[0,\tau_{t^+}-\tau_{t^-}]}(V_{\tau_{t^+}} - V_{\tau_{t^-}})$, where U represents the bridge stochastic process related to W. Thus, the hazard functions can be derived through the exponential relationship

that is satisfied for the survival function of B:

$$\lambda(t|\bar{W}_t) = -\frac{\partial}{\partial t}\ln S(t|\bar{V}_x)|_{x=t} = -\frac{\partial}{\partial t}\ln E_{\bar{W}}\big\{S_{B|\bar{W}}(V_{\tau_{x^-}} + W_{t-\tau_{x^-}})|\bar{V}_x\big\}\big|_{x=t}$$
$$\lambda(t|\bar{W}) = -\frac{\partial}{\partial t}\ln S(t|\bar{V}) = -\frac{\partial}{\partial t}\ln E_U\big\{S_{B|U}(V_{\tau_{t^-}} + U_{[0,\tau_{t^+}-\tau_{t^-}]}(V_{\tau_{t^+}} - V_{\tau_{t^-}}))|\bar{V}\big\} \tag{3.38}$$

Thus, by the Kalbfleisch and Prentice (2002) definition, W_t is an endogenous covariate, and the exponential relationship is satisfied. This particular model demonstrates the validity of the exponential form for an endogenous covariate.

3.8 Discussion

We have carefully considered the definitions of time-varying covariates and illustrated through various examples how current beliefs regarding the definitions of such variables fall short. Rather than basing our understanding of these covariates off of intuition, we used the mathematical definitions outlined in Kalbfleisch and Prentice (2002) to reformulate the difference between exogenous and endogenous covariates as a missing data problem. By doing so, we were able to emphasize that the major difference between these covariates lies in how the future trajectory impacts the hazard of failure at any point in time. For exogenous covariates, the future is irrelevant, but for endogenous covariates, it is inextricably linked to the present. Thus, models that only condition on the past for an endogenous marker fail to account for the dependence on the future. Naively specifying survival at time t conditional on the trajectory of an endogenous marker up to t results in an invalid survival function due to the trajectory changing as a function of t but being held fixed for the purposes of prediction. Our newly developed mechanistic framework for specifying valid survival functions circumvents this issue by using the Law of Total Expectation to first specify survival conditional on the full trajectory of the marker and then average over the

unobserved future, given that a history up to some fixed time x is observed, where x can be less than or greater than t. In this sense, the unobserved future trajectory of the marker is treated as a latent variable, similar to a frailty or random effect term.

Another important point we sought to clarify was the lack of a one-to-one correspondence between the type of time-varying covariate and the validity of the exponential form of survival. Once more, whether a covariate is endogenous or exogenous is related to how the future trajectory of the marker affects the hazard of failure. The validity of the exponential form, however, is influenced by the continuity of the event time T. This continuity may depend on the trajectory of the marker process, namely through any possible discreteness that may be induced on T as a result of the path of the marker. However, as seen through our illustrative examples, there are cases where the exponential form is valid for an endogenous covariate and invalid for an exogenous covariate. One specific example we and others (*Ting Lee and Whitmore*, 2006) have explored is the threshold regression model where the marker is fully observed and the boundary is fixed. In this model, the marker is exogenous due to being fully observed, but knowledge of the marker at any point in time allows one to fully determine whether the event has occurred at time t. More specifically, T is now a stopping time for which the exponential form, by definition, is no longer valid (*Aven and Jensen*, 1999; *Finkelstein*, 2004). Decoupling the legitimacy of the exponential form and the definitions of time-varying covariates allows us to more accurately define survival functions in different settings where complex marker data may impact the risk of failure.

Our framework sits opposite the intuitive understanding about time-varying covariates. Current practice is to treat time-varying covariates originating outside a subject as being exogenous and those being measured on a subject as being endogenous. From our theoretical construction and illustrative examples, we highlight where the current belief falls short. Our approach is to formulate the difference as a missing

data problem, which allows for quantitative examination of whether a covariate is endogenous or exogenous. This is philosophically different from the current approach, where covariates are deemed internal or external before analysis is done in order to determine how to define our models. Under our framework, legitimate survival models can be defined regardless of the type of time-varying marker, and this allows us to incorporate both exogenous and endogenous markers into our models. Therefore, not only is our framework quantitatively different from what it done in practice, it is also a philosophical shift in how we think about endogenous markers. Recently, there has been work by others to try to extend the current understanding of how markers play a role in survival models (*Dempsey and McCullagh*, 2018). Our framework stands in contrast to the one explored in Dempsey and McCullagh (2018), which formulates a model for the marker conditional on the event time backwards in time. In fact, our approach can be seen as complementary: we condition on the latent marker trajectory looking forward in time. Regardless of the difference, both the approach outlined in Dempsey and McCullagh (2018) and the approach we develop herein seek to elucidate the role that endogenous covariates can play in survival models.

An important aspect of time-varying covariates that needs to be carefully considered is their role in the context of clinical trials. As interest lies in understanding the effect of a treatment or intervention on a health-related outcome, models for data originating from a clinical trial should include the treatment and baseline factors only. Incorporating time-varying marker measures after the initiation of treatment can obfuscate any treatment effect that may have occurred. Thus, time-varying markers should not be incorporated into survival models in the clinical trial setting. We do not dispute that this should be the case. Our framework was largely motivated by observational data, particularly screening studies for prostate cancer where men will have serially measured PSA values. In this context, there is no intervention or treatment being studied in a carefully selected sample of patients. Therefore, our methods

are appropriate for observational data, which are common in clinical studies of longitudinal biomarkers and survival.

While our framework was set up in a general way and should be applicable to a range of survival model specifications, we only studied its application in specific examples. These examples were chosen based on common settings where a marker process would be appropriate to define the time-varying covariate. Exploration of other common survival models with marker processes is needed. In the context of joint modeling, our framework largely focuses on the survival model. There is a question about how we can also define a marker process model that can be linked to the survival model in order to define the joint model. Our framework should still be valid in such models, but the difficulty may lay in how to define the latent factors needed to link the two submodels. Such work is ongoing. This framework should also apply in the competing risks and recurrent event settings, and we plan to explore these models in future work.

CHAPTER IV

Joint Modeling of a Time-to-Event and Partially Observed Marker Process Using Lévy Processes

4.1 Introduction

Often in clinical research, event information and longitudinal biomarker data will be collected for each subject. In prostate cancer studies, time to diagnosis or recurrence and prostate-specific antigen (PSA) levels may be recorded at each clinical visit. The various events that may be observed for a given patient can be conceptualized as resulting from an underlying disease process that accumulates over time until failure occurs. One specific instance of this setup is the first-hitting time model, wherein failure is defined as the first time at which a stochastic process hits a threshold or boundary state (*Ting Lee and Whitmore*, 2006). In general, the latent disease process can be conceptualized as a stochastic process (*Jewell and Kalbfleisch*, 1996), and it may be captured by some known biomarker that is partially observed over time through longitudinal measurements. The level of the marker may be prognostic for disease status or severity, which motivates the use of joint models to capture the association between failure and the marker as two manifestations of the latent disease process.

Joint models are specified through submodels for the failure process and longi-

tudinal marker process and terms that link the two individual models (*Henderson et al.*, 2000; *Tsiatis and Davidian*, 2004; *Rizopoulos*, 2008). The most common joint model specification is a shared random effects model where the longitudinal submodel is specified as a linear mixed effects model and the failure model is given as a Cox proportional hazards model incorporating the marker as a covariate. This model specification assumes a linear form for the marker. Sparse observation of the marker via longitudinal measurements may make it difficult to accurately assess the assumption of linearity. Thus, the most common joint model specification does not allow for nonlinear biological variability in the marker process. Others have explored the use of nonlinear mixed models for the longitudinal marker data (*Wu et al.*, 2008; *Murawska et al.*, 2012; *Desmée et al.*, 2015, 2017a; *Köhler et al.*, 2018), with a particular focus on Bayesian algorithms for estimation and individualized prediction.

An alternative approach to modeling the longitudinal marker data uses a stochastic process (*Henderson et al.*, 2000; *Wang and Taylor*, 2001; *Hashemi et al.*, 2003; *Arbeev et al.*, 2014). Due to the partially observed nature of the marker process, the model for observed data is averaged over the latent trajectory that is unobserved between longitudinal measurement times. This average is taken conditional on the observed value; thus, the latent process being averaged over is a bridge process. The difficulty with stochastic process marker models stems from mathematical complexity, as averaging over the latent trajectory may involve an infinite dimensional integral. To overcome this difficulty, most joint models involving stochastic processes make use of approximations or place strong assumptions on the stochastic process itself. While motivated by tractability, these approximations and assumptions can be hard to justify if there are few measurements for each subject.

However, there are stochastic process models that are precisely specified. One example considered by Yashin and Manton (1997) proposed a model using a squared Gaussian process to link a hazard function with a discretely observed stochastic pro-

cess. However, in their model, a U- or J-shaped hazard is possible, which may not be reasonable in cases where the risk of a disease-related event correlates with higher marker values. Maximum likelihood is used for parameter estimation, and it involves solving stochastic differential equations which are conceptually and technically challenging. In contrast to the approach proposed by Yashin and Manton (1997), Dempsey and McCullagh (2018) model the stochastic process in reverse time conditional on the failure time and define "revival models." Their proposed model avoids any averaging over the latent trajectory, but it assumes that the survival time will be finite with probability one and fully explained by the marker process. In practice, however, the failure process may be related to but not fully explained by the biomarker process of interest. Dempsey and McCullagh's (2018) approach is complementary to joint modeling and works in the case of extreme dependence between the marker and the failure time when the marker cannot exist, theoretically or counterfactually, after failure.

When longitudinal measurements on an individual subject are sparse, making strong assumptions on the marker process, either through a mixed effects model or a stochastic process model with approximations, can potentially lead to an inflexible joint model. Ideally, we want a marker process model that can properly explain the assumed latent and actually observed trajectory. In this work, we propose a joint model specification that makes use of a specific type of Lévy subordinator process and a time transformation to model the accumulated risk of failure. Our model has a tractable form and avoids the computational cost associated with other joint models. We outline the longitudinal marker and failure process submodels of the joint model in Section 4.2. In Section 4.3, we describe our model formulation and estimation under different observation scenarios for the marker process. In Section 4.4, we examine the performance of our model in simulation studies. We apply our model to SEER prostate cancer incidence data, where PSA is measured at the age of diagnosis for

men diagnosed with prostate cancer, in Section 4.5. We conclude with a discussion of the strengths and limitations of our model in Section 4.6.

4.2 Joint Model

4.2.1 Survival Model

Let T be the failure time for the event of interest, $\boldsymbol{Z}$ be a vector of baseline covariates, and W_t be a time-dependent covariate. W_t may be observed continuously or intermittently at discrete time points. The most common survival model is the Cox proportional hazards model (1972), which specifies the survival function conditional on baseline covariates as

$$S(t|\boldsymbol{Z}) = e^{-\int_0^t h_0(u)\exp\{\boldsymbol{\beta Z}\}du} = e^{-H_0(t)\exp\{\boldsymbol{\beta Z}\}} = e^{-H(t|\boldsymbol{Z})} \tag{4.1}$$

In (4.1), $h_0(t)$ and $H_0(t)$ represent the baseline hazard and baseline cumulative hazard, respectively. $H(t|\boldsymbol{Z})$ represents the cumulative hazard that is conditional on baseline covariates, and $\boldsymbol{\beta}$ is a vector of regression parameters corresponding to $\boldsymbol{Z}$. While the Cox model in (4.1) is the most common approach to modeling survival, it comes with a strong assumption of proportionality. In addition, it assumes the same level of risk for each subject. To account for the heterogeneity that exists among individuals, extensions to the Cox model involving frailty terms (*Vaupel et al.*, 1979), which are essentially random effect terms, have been developed. It is assumed that at baseline, each subject has a non-negative frailty random variable U_i that acts multiplicatively on the baseline hazard $h(t|\boldsymbol{Z})$. As the frailty represents a latent effect that is never observed in practice, it must be integrated over to obtain the marginal survival function.

$$S(t|\boldsymbol{Z}) = E_U\big\{e^{-U\exp\{\boldsymbol{\beta Z}\}H_0(t)}|\boldsymbol{Z}\big\} = E_U\{e^{-UH(t|\boldsymbol{Z})}|\boldsymbol{Z}\} \tag{4.2}$$

The expectation in (4.2) has the form of a Laplace transform of U with $H(t|\boldsymbol{Z})$ as the argument. Recall that the Laplace transform is defined as $\mathcal{L}_U(s) = E_U\{e^{-sU}\}$. The frailty random variable U is typically drawn from a parametric family, most often the Gamma or compound Poisson distributions (*Vaupel et al.*, 1979; *Aalen*, 1992).

While standard frailty models account for heterogeneity across subjects, they fail to account for potential heterogeneity across time within a given subject. Assuming that a patient's risk can change dynamically over time, we can consider a generalization of proportional hazards frailty models where the frailty is now allowed to vary with time. In other words, the frailty random variable for subject i, U_i, now becomes a stochastic process $U_i(t)$. Such process frailty models have been previously considered, particularly in the context of multivariate survival data (*Aalen*, 1994; *McGilchrist and Yau*, 1996; *Huibin and Chan*, 1997; *Perperoglou et al.*, 2006; *Aalen et al.*, 2008). In these models, the marginal survival function is now an average over all possible trajectories of U up to any time t, represented by $\bar{U}_{[0,t]}$.

$$S(t|\boldsymbol{Z}) = E_{\bar{U}_{[0,t]}}\left\{e^{-\int_0^t U_s h_0(s)\exp\{\boldsymbol{\beta Z}\}ds}\Big|\boldsymbol{Z}\right\} = E_{\bar{U}_{[0,t]}}\left\{e^{-\int_0^t U_s h(s|\boldsymbol{Z})ds}\Big|\boldsymbol{Z}\right\} \tag{4.3}$$

The expectation in (4.3) now represents a Laplace function for the process $U(t)$ with respect to the hazard $h(s|\boldsymbol{Z})$, i.e., $E_{\bar{U}_{[0,t]}}\{e^{-\int_0^t U_s f(s)ds}\}$, where $f(s) = h(s|\boldsymbol{Z})$.

As $U(t)$ acts multiplicatively on the hazard, the process must be non-negative in order to preserve the non-negative property of the hazard function. Prior work by Gjessing et al (2003) modeled the stochastic process using a non-negative Lévy process. Lévy processes are a class of stochastic processes with independent and stationary increments (*Bertoin*, 1998; *Hoyle*, 2010). The Lévy process family includes Wiener processes, Gaussian processes, and compound Poisson processes. A particular sub-family of non-negative Lévy processes, known as Lévy subordinators, includes the compound Poisson process and its limits, which include the Gamma process, and ex-

cludes Gaussian processes. Using a Lévy subordinator in (4.3) will lead to a tractable form for the survival and hazard function. However, this model specification assumes that all subjects will have proportional hazards and that the random effect acting on the hazard is increasing over time.

Work by Putter and Van Houwelingen (2015) focused on dynamic frailty models that used Lévy processes to flexibly model serial correlation of the frailty process. Under their framework, the time-dependent frailty $U(t)$ was constructed from independent frailty components $X(u, v)$, representing a compound birth-death stochastic process. $X(u, v)$ contributes to the hazard only if $u \leq t$ and $v \geq t$, meaning that each component will only contribute in the time period after they are "born" and before they "die." Therefore, the frailty components were specified as a Lévy process in two dimensions. Parameters were estimated using the expectation-maximization (EM) algorithm for a multivariate survival model. Estimation via the EM algorithm can be computationally demanding and slow in this context.

In both the Gjessing et al (2003) and Putter and Van Houwelingen (2015) frameworks, the frailty process is fully unobserved and the marginal survival model is used in maximum likelihood estimation. However, if we think of the frailty $U(t)$ as a stochastic marker process, such as PSA in the context of prostate cancer, it may be observed intermittently at discrete time points. Let us assume that $U(t)$ is observed at time points $\boldsymbol{\tau} = \{0, \tau_1, \tau_2, \ldots, \tau_k\}$ and remains a latent process in the intervals between measurement times $(0, \tau_1), (\tau_1, \tau_2), \ldots, (\tau_{k-1}, \tau_k)$. We can assume that $U(t)$ is itself the marker process, but it is also possible to model $U(t)$ given the observed measurements as a function of the marker process. As the process is observed only at discrete measurement times, the Lévy bridge process can be used to link the discrete measurements in order to model the full trajectory of $U(t)$ over a given interval (*Hoyle*, 2010). The Lévy bridge process will be scaled by the observed values of the marker process at each point in time. In general, a bridge process is a stochastic

process conditioned on a known value at a fixed future time point (*Mansuy and Yor*, 2008). A Lévy bridge is a Lévy process defined over a finite interval, and it is assumed that the initial and final values over the process needed to construct the bridge are known at baseline (*Hoyle*, 2010).

In the dynamic frailty framework, the non-negative Lévy process is applied multiplicatively to the hazard. An alternative model specification explored in Suresh (2018) uses the Lévy frailty process as a multiplicative effect in the cumulative hazard function in order to account for the non-decreasing assumption. Specifically, survival was defined as $S(t|U) = e^{-U(t)H(t)}$, where $H(t)$ represents the cumulative baseline hazard function. The cumulative baseline hazard is a transformation of time and can be thought of as an underlying risk of disease common to all subjects. Then, the stochastic frailty $U(t)$, which can be thought of as a subject-specific transformation of the risk, could be recast as a function of $H(t)$, namely $U(H(t))$. Essentially, we can introduce a time transformation into the frailty process $U(t)$, where the transformation is the cumulative baseline hazard. Under this model, each subject's risk of disease is the combination of an underlying level of risk $H(t)$ and an individual level of risk which may be modified by subject-specific covariate values through the parameters of $U(t)$. Thus, an alternative model specification we focus on assumes that $U(H(t)) = U_{H_t}$ is itself the cumulative hazard function, and the survival function can be written as $S(t|U) = e^{-U_{H_t}}$. This model specification has greater tractability and convenience due to being able to use the Laplace transform of U_{H_t} rather than a Laplace functional to obtain the marginal survival function.

Survival models involving a time-transformed Lévy frailty process have been previously considered by Gjessing et al (2003). Others have examined alternative times scales in survival and threshold regression models based on an underlying stochastic process (*Ting Lee and Whitmore*, 1993; *Oakes*, 1995; *Whitmore and Schenkelberg*, 1997; *Duchesne and Lawless*, 2000; *Ting Lee and Whitmore*, 2006). Our pro-

posed model introducing a time transformation on the latent process $U(t)$ allows for greater flexibility relative to other models that assume a multiplicative effect (*Peng and Huang*, 2007; *Suresh*, 2018).

4.2.2 Developing the Joint Model

4.2.2.1 General Model Notation, Assumptions, and Likelihood

Much of what follows is based on the notation and models for valid survival functions that incorporate endogenous marker histories outlined in Chapter III, Section 3.4. Let M be a marker process and let M_t represent the value of M at time t. $\bar{M}_a^b$ will denote the trajectory of M between a and b, i.e., $\bar{M}_a^b = \{M_t\}|_{t=a}^{t=b}$. For brevity, we define $\bar{M}_t = \bar{M}_0^t$ and $\bar{M} = \bar{M}_0^\infty$. Let $\mathcal{F}_t^M$ and $\mathcal{F}^M$ denote the filtrations induced by $\bar{M}_t$ and $\bar{M}$, respectively, where the filtration $\mathcal{F}_s^M$ for some time s captures all past event, marker, and covariate information up to s^-.

Let the random variables T be the survival time and M be the generally endogenous, or internal, stochastic marker process that provides dynamic information on T. M is potentially observed as it may be informatively censored by T and by the non-informative right censoring time C that may censor both T and M. We assume that T and M satisfy the following conditions:

1. The process M is left continuous and therefore predictable. This assumption is needed for the integrals of the process over time to be well defined.

2. The conditional random variable $T|\bar{M}$ is absolutely continuous. This assumption implies that T is a sudden failure, which is one that cannot be predicted or fully explained if the full history $\mathcal{F}^M$ were available. In most situations when the marker does not fully explain when failure occurs, this assumption makes biological sense. This condition can also be implied by a weaker set of assumptions by postulating that T is a totally inaccessible stopping time (*Aven and Jensen*,

1999) and that no behavior of M can induce discreteness in T. For example, when T is defined as the first time that M passes a threshold, the assumption is violated if the threshold is non-random as T is non-random or discrete for fixed M. However, the assumption will be satisfied when the threshold is random, unobserved, and absolutely continuous (*Finkelstein*, 2004).

The above conditions ensure that the well-known exponential relationship between the survival function S and the hazard function λ of $T|\mathcal{F}$ holds true.

$$S(t|\mathcal{F}^M) = \exp\left\{ -\int_0^t \lambda(x|\mathcal{F}^M)dx \right\} \tag{4.4}$$

For the relationship in (4.4) to be satisfied, both S and λ should be conditional on the full information $\mathcal{F}^M$, including values in the future of time t or x.

Many survival models are specified conditional only on the past information $\mathcal{F}_t^M$. Under this level of conditioning, the exponential relationship in (4.4) will not be satisfied unless M is an exogenous marker as $S(t|\mathcal{F}_t^M) = S(t|\mathcal{F}^M)$ (*Kalbfleisch and Prentice*, 2002). For endogenous M, the relationship will be satisfied conditional on fixed $\mathcal{F}_x^M$, where x can be $< t$ or $> t$ and including $x = \infty$. Applying the Law of Total Expectation, a valid survival function conditional on fixed $\mathcal{F}_x^M$ for endogenous M can be derived as

$$S(t|\mathcal{F}_x^M) = S(t|\bar{M}_x) = E\big\{S(t|\mathcal{F}^M)\big|\bar{M}_x\big\} = E\left\{ \exp\left\{ -\int_0^t \lambda(u|\bar{M})du \right\}\middle|\bar{M}_x\right\} \tag{4.5}$$

In (4.5), the expectation is taken over the future trajectory $\bar{M}_{x^+}^{\infty}$ conditional on the past history $\bar{M}_x$. Similar expressions can be derived for the density function f and

the hazard function λ:

$$f(t|\mathcal{F}_x^M) = f(t|\bar{M}_x) = E\big\{f(t|\mathcal{F}^M)\big|\bar{M}_x\big\} = E\bigg\{\lambda(t|\bar{M})\exp\bigg\{-\int_0^t \lambda(u|\bar{M})du\bigg\}\bigg|\bar{M}_x\bigg\}$$

$$\lambda(t|\mathcal{F}_x^M) = \lambda(t|\bar{M}_x) = \frac{f(t|\bar{M}_x)}{S(t|\bar{M}_x)} = \frac{E\bigg\{\lambda(t|\bar{M})\exp\big\{-\int_0^t \lambda(u|\bar{M})du\big\}\bigg|\bar{M}_x\bigg\}}{E\bigg\{\exp\big\{-\int_0^t \lambda(u|\bar{M})du\big\}\bigg|\bar{M}_x\bigg\}} \tag{4.6}$$

Using the quantities defined in (4.5) and (4.6), we can mechanistically define valid functions $f(t|\mathcal{F}_t^M)$, $\lambda(t|\mathcal{F}_t^M)$, and $S(t|\mathcal{F}_t^M)$ via equation (3.9) in Chapter III, Section 3.4:

$$S(t|\bar{M}_x)|_{x=t} = E\big\{S(t|\mathcal{F}^M)\big|\bar{M}_x\big\}\big|_{x=t} = E\bigg\{\exp\bigg\{-\int_0^t \lambda(u|\bar{M})du\bigg\}\bigg|\bar{M}_x\bigg\}\bigg|_{x=t}$$

$$f(t|\bar{M}_x)|_{x=t} = E\big\{f(t|\mathcal{F}^M)\big|\bar{M}_x\big\}\big|_{x=t} = E\bigg\{\lambda(t|\bar{M})\exp\bigg\{-\int_0^t \lambda(u|\bar{M})du\bigg\}\bigg|\bar{M}_x\bigg\}\bigg|_{x=t}$$

$$\lambda(t|\bar{M}_x)|_{x=t} = \frac{f(t|\bar{M}_x)|_{x=t}}{S(t|\bar{M}_x)|_{x=t}} = \frac{E\bigg\{\lambda(t|\bar{M})\exp\big\{-\int_0^t \lambda(u|\bar{M})du\big\}\bigg|\bar{M}_x\bigg\}\bigg|_{x=t}}{E\bigg\{\exp\big\{-\int_0^t \lambda(u|\bar{M})du\big\}\bigg|\bar{M}_x\bigg\}\bigg|_{x=t}} \tag{4.7}$$

The expressions in (4.7) can be conceptualized by first defining the functions S, λ, and f conditional on a fixed marker history up to time x, and then equating x to t.

In this study, we are interested in the process V representing a partially observed biomarker process W, where W represents the process if it were continuously observed. We assume that $\mathcal{F}^V \subset \mathcal{F}^W$. By the Law of Total Expectation, the survival, hazard, and density functions conditional on $\bar{V}_x$ are those given in (4.7) for general $x > 0$, including $x = \infty$. When the marker W is external, we can write $\bar{W}_u$ in the expressions in (4.7) under the integral and $\bar{W}_t$ in λ outside the integral. The philosophy

underpinning this model is similar to modeling dependence using a frailty term. In such models, conditional independence given the frailty is assumed. In our model, a simpler form of dependence is assumed given the latent, exogenous W that explains the dependence between the failure time T and the endogenous partially observed biomarker process V.

Given a set of observation times $\boldsymbol{\tau} = \{\tau_1, \tau_2, \ldots, \tau_k\}$, we assume that $W(\tau_j) = V(\tau_j)$, $j = 1, \ldots, k$. In other words, the observed marker values V at each measurement time are assumed to be observations of the true underlying marker process W_t. The true W is otherwise unobserved between measurement times $\boldsymbol{\tau}$ and $\bar{W}|\bar{V}$ can be represented by a bridge process. The observed V are measured at discrete time points, thus to define V in continuous time, we assume that it is constant between measurement times. Such a construction induces an endogenous structure on V, and future observations of V provide information on the latent W between measurement times.

Based on the general formulation of the joint model, we can write the joint log-likelihood for a sample of size n as

$$\begin{aligned} l &= \sum_{i=1}^{n} (1 - \Delta_i) \ln S(x_i|\bar{V}_{x_i}) + \Delta_i \ln f(x_i|\bar{V}_{x_i}) + \ln g(\bar{V}_{x_i}) \\ &= \sum_{i=1}^{n} \Delta_i \ln d\Lambda(x_i|\bar{V}_{x_i}) - \Lambda(x_i|\bar{V}_{x_i}) + \ln g(\bar{V}_{x_i}) \end{aligned} \tag{4.8}$$

where Δ_i is an indicator of failure, $x_i = \min(t_i, c_i)$ is the observed failure time for each subject (t_i =failure time; c_i = censoring time), g is the joint pdf of the subject-specific observed biomarker measurements $\bar{V}_{x_i}$, and

$$\begin{aligned} d\Lambda(x|\bar{V}_x) &= \lambda(x|\bar{V}_y)|_{y=x} dx \\ \Lambda(x|\bar{V}_x) &= \int_0^x \lambda(\xi|\bar{V}_y) d\xi \Big|_{y=x} \end{aligned}$$

4.2.2.2 Marker Process Model

We assume that the risk of failure is associated with an underlying stochastic process $\{W_t\}_{0\leq t\leq\infty}$ that is 0 at baseline and increasing over time. W_t is rarely continuously observed, and a set of discrete measurements $\{V_{\boldsymbol{\tau}}\}$ collected at observation times $\boldsymbol{\tau} = \{\tau_1, \ldots, \tau_k\}$ may be observed instead. We thus make the assumption that W_t is partially observed, and the observed $V_{\boldsymbol{\tau}}$ at each τ_j, $j = 1, \ldots, k$, represent the true values of the underlying process W. Between the measurement times $\boldsymbol{\tau}$, the latent process W_t takes the form

$$W_t = B\{t\}_{[0,\tau_j-\tau_{j-1}]}(V_{\tau_j} - V_{\tau_{j-1}}), \quad j = 1, \ldots, k \tag{4.9}$$

where $[B\{t\}]_{0\leq t\leq\tau_j-\tau_{j-1}}$ is a Lévy bridge process over the interval $[0, \tau_j - \tau_{j-1}]$ that takes on values between 0 to 1. We scale $B\{t\}$ by the difference in observed values at τ_j and τ_{j-1} in order to have the latent and observed marker values on the same scale. Thus, at baseline, the bridge process is assumed to be 0, and it increases to $V_{\tau_j} - V_{\tau_{j-1}}$. Under our modeling framework outlined in Section 4.2.1, the process W_t will be subordinated by the cumulative baseline hazard H_t, which represents the accumulated risk inherent to each subject. As we introduce a time transformation $H(t)$ into the latent marker process W, any processes related to W will also be specified on the transformed time scale. By the assumption of no measurement error on the observed marker values, we have that $W_{H_{\boldsymbol{\tau}}} = V_{H_{\boldsymbol{\tau}}}$. Additionally, the bridge process can be written on the transformed time scale as $B\{H(t)\}$.

The use of a Lévy process to model the true marker process is largely done for mathematical tractability. In the model where the marker is partially observed, the latent trajectory over intervals where the process is unobserved must be averaged over. This generally involves high dimensional integration and can be computationally difficult, but the assumption of a process in the Lévy family leads to closed form

solutions. There are many choices for the marker model, but we will focus on the Gamma process model. The Gamma process is often used to model marker processes due largely to its convenient distributional properties (*Gjessing et al.*, 2003; *Lawless and Crowder*, 2004; *Putter and van Houwelingen*, 2015). In financial mathematics, the Gamma bridge process has had applications in aggregate claims data, and prior work has derived useful properties of the bridge process (*Brody et al.*, 2008; *Hoyle*, 2010; *Suresh*, 2018). The Gamma process is suitable to describe the accumulation of disease risk associated with a stochastic marker process, and we apply the Gamma process in a joint model for survival and longitudinal data.

Under this assumption, the underlying marker W_t can be modeled using the Gamma process. For our specific model, where we consider the subordinated process W_{H_t}, we assume that W_{H_t} can also be modeled via the Gamma process. By assumption, the process has independent increments, initial value $W_{H_0} = 0$, and W_{H_t} follows a Gamma distribution with mean μH_t and variance $\sigma^2 H_t$. We define a subordinated bridge process $B\{H_t\}$ that is independent of the final observed V_{H_τ}. Because the underlying W_{H_t} follows a Gamma distribution, it can be shown that $B\{H_t\}$ conditional on the observed values follows a Beta distribution (*Brody et al.*, 2008). Therefore, to average over the unobserved Gamma bridge process when constructing the observed data likelihood, we must make use of the Laplace transform of the bridge process, which corresponds to the Laplace transform of a Beta random variable.

The survival model outlined in Section 4.2.1 and the marker model developed above fully specify a joint model for the marker process and event time, where we choose a model for the marker process that is mathematically tractable. This joint model allows us to use the complete set of observed marker values for a given subject, including those that may be observed after the event time in the case of a non-terminal event. Baseline covariates $\boldsymbol{Z}$ can be incorporated into the Gamma process

parameters of the underlying marker W or into the time transformation $H(t)$. By construction, there is positive dependence between the marker process and the event time that stems from the variance and covariance of W.

4.3 Likelihood Construction

Our model can be formulated under different observation mechanisms for the marker. In particular, we focus on three scenarios: (1) the marker process is completely observed; (2) the marker process is observed at an informative measurement time; (3) the marker process is partially observed at a set of uninformative measurement times. In each scenario, estimation is done by maximizing the likelihood with respect to the marker parameters. If H is assumed to be a parametric function, standard errors for the parameters can be obtained from the Hessian matrix using numerical differentiation. If H is estimated nonparametrically, standard errors of the estimators must be obtained in an alternative way, either by deriving the appropriate variance expressions or by using an empirical procedure such as the bootstrap.

4.3.1 Completely Observed Marker

Let T_i and C_i denote the true event and censoring times, respectively. Let $X_i = \min(T_i, C_i)$ be the observed event time and $\Delta_i = I(T_i \leq C_i)$ be the event indicator. Let $\boldsymbol{Z}$ represent a vector of baseline covariates that can include the initial value of the marker process. Let $\{W_{H_t}\}_{0 \leq t \leq X}$ be the Lévy process representing the completely observed marker process on the H-transformed time scale. The observed data for each subject can be represented as $\mathcal{O}_i = \{X_i, \Delta_i, \boldsymbol{Z}_i, \bar{W}_{X_i}\}_{i=1}^n$, where $\bar{W}_{X_i}$ is the history of the process W up to the observed event time X_i. The conditional survival function can be expressed as

$$S(t|\bar{W}_t, \boldsymbol{Z}) = e^{\Lambda(t|\bar{W}_t)} = e^{-W(H(t))} \tag{4.10}$$

where $\Lambda(t)$ is the cumulative hazard and $H(t)$ is the cumulative baseline hazard. Recall that the baseline covariates $\boldsymbol{Z}$ can be incorporated into the parameters governing W_{H_t} or into $H(t)$ itself. Also recall that because W_{H_t} is assumed to be completely observed or exogenous, the model specification in (4.10) is mathematically valid. Going forward, we will exclude $\boldsymbol{Z}$ from our model specifications for brevity. Then, the hazard function corresponding to the survival model represented by (4.10) is

$$d\Lambda(t|\bar{W}_t) = dW(H(t)) \tag{4.11}$$

Estimation is done by maximizing the log-likelihood

$$\begin{aligned} l &= \sum_{i=1}^{n} \Delta_i \ln\left[d\Lambda(X_i|\bar{W}_{X_i})S(X_i|\bar{W}_{X_i})\right] - (1-\Delta_i)\Lambda(X_i|\bar{W}_{X_i}) \\ &= \sum_{i=1}^{n} \Delta_i \ln\left[d\Lambda(X_i|\bar{W}_{X_i})\right] - \Lambda(X_i|\bar{W}_{X_i}) \end{aligned} \tag{4.12}$$

4.3.2 Marker Observed at Informative Measurement Time

It may be the case that the marker is measured at some informative observation time related to the event process. The most straightforward example is when the marker is observed at the event time, in which case we have a marked survival endpoint. For simplicity, we will assume that the marker is observed at a single informative measurement time that is the event time. To set up a mathematically valid and tractable model, we will make use of the mechanistic model specifications in Section 4.2.2.1.

Let $\{W_t\}_{0\leq t\leq \tau}$ be the underlying Lévy process defined on the horizon $[0, \tau]$. As our model is specified on the H-transformed time scale, we can write W_{H_t}. We assume a parameterization similar to Suresh (2018), where W_{H_t} is a scaled gamma process with mean μH_t and variance $\sigma^2 H_t$. Thus, W_{H_t} will follow a Gamma distribution with shape αH_t and scale β, where $\alpha = \frac{\mu^2}{\sigma^2}$ and $\beta = \frac{\sigma^2}{\mu}$. We assume subjects will have a

marker value V measured at their observed event time t only if the event is observed to occur, i.e., we will have a bivariate outcome (t, V_t) for subjects who experience the event and a univariate outcome t for those that do not. We assume that (1) $V_t = W_t$ for an underlying Lévy process W and (2) both V and W can be expressed on the H-transformed time scale.

In the case of a marked survival endpoint, the joint distribution of the event time T and the marker value observed at T can be written as

$$f(t, V_t) = f(t|V_x)|_{x=t} f(V_t) = f(V_t|T = t)f(t) \tag{4.13}$$

Considering the first expression in (4.13),

$$\begin{aligned}
f(t|V_x)|_{x=t} &= \Big[-\frac{d}{dt} S(t|V_x) \Big]_{x=t} = \Big[-\frac{d}{dt} E\big\{ S(t|\bar{W}) \big| V_x \big\} \Big]_{x=t} \\
&= \Big[-\frac{d}{dt} E\big\{ e^{-W(H_t)} \big| V_x \big\} \Big]_{x=t} \\
&= \Big[-\frac{d}{dt} E\big\{ e^{-[V(H_x)+W(H_t-H_x)]} \big| V_x \big\} \Big]_{x=t} \\
&= \Big[-\frac{d}{dt} E\big\{ e^{-B\{H_t\}V(H_x)} \big| V_x \big\} \Big]_{x=t} \\
f(V_t) &= \int_0^\infty f(V_y|T = y)f(y)dy \\
\implies f(t|V_x)|_{x=t} f(V_t) &= \Big[-\frac{d}{dt} E\big\{ e^{-B\{H_t\}V(H_x)} \big| V_x \big\} \Big]_{x=t} \times \int_0^\infty f(V_y|T = y)f(y)dy
\end{aligned} \tag{4.14}$$

Use of the first expression in (4.13) involves a complicated integral for the marginal distribution of the marker in order to account for informative observation at the event time. The explicit form of this integral is derived in Appendix C.1. In addition, the final two lines in the expression for $f(t|V_x)|_{x=t}$ in (4.14) reflect the fact that the same density should result whether we assume $x < t$ or $x > t$.

Use of the second expression in (4.13), however, results in a much simpler and more straightforward joint model specification:

$$\begin{aligned} f(V_t|T=t) &= \frac{v^{\alpha H_t - 1} e^{-v/\beta}}{\Gamma(\alpha H_t)\beta^{\alpha H_t}} \\ f(t) &= -\frac{d}{dt} E_W\{e^{-W(H_t)}\} = -\frac{d}{dt}\mathcal{L}_W(1) = -\frac{d}{dt}(1+\beta)^{-\alpha H_t} \\ &= \alpha h_t \ln(1+\beta)(1+\beta)^{-\alpha H_t} \\ \implies f(V_t|T=t)f(t) &= \frac{v^{\alpha H_t - 1} e^{-v/\beta}}{\Gamma(\alpha H_t)\beta^{\alpha H_t}} \alpha h_t \ln(1+\beta)(1+\beta)^{-\alpha H_t} \end{aligned} \tag{4.15}$$

where the marginal distribution of T in (4.15) is derived from the completely observed model in (4.10) by taking the expectation with respect to the latent W. Note that this expectation corresponds to the Laplace transform of a Gamma random variable.

If a subject does not have a marker value measured at their event time t, the marginal survival and hazard functions can be derived from the model specified using the completely observed marker W.

$$\begin{aligned} S(t) &= E_W\{S(t|\bar{W})\} = E_W\{e^{-W(H_t)}\} = \mathcal{L}_W(1) = (1+\beta)^{-\alpha H_t} \\ d\Lambda(t) &= -d\ln S(t) = \alpha \ln(1+\beta) dH_t \end{aligned} \tag{4.16}$$

The marginal functions in (4.16) are derived from the Laplace transform of a Gamma random variable. Details for (4.13)-(4.16) are provided in Appendix C.1.

To set up the log-likelihood for estimation, we assume that subjects will have a marker value observed only if they experience the event of interest. In this case, the marker value $V_{H_{X_i}}$ will be observed at X_i for subjects with $\Delta_i = 1$. If $\Delta_i = 0$ and the subject does not experience the event of interest, no marker measure will be observed. Under this setup, the observed data for each subject is $\mathcal{O}_i = \{X_i, \Delta_i, \mathbf{Z}_i, V_{H_\tau}|_{\tau=X_i}\}_{i=1}^n$.

Estimation is done by maximizing the log-likelihood

$$l = \sum_{i=1}^{n} \Delta_i \ln \left[d\Lambda(X_i|V_{H_\tau})|_{\tau=X_i} \cdot S(X_i|V_{H_\tau})|_{\tau=X_i} \cdot g_{V_{H_\tau}}(V_i)|_{\tau=X_i} \right] + (1-\Delta_i) \ln \left[S(X_i) \right] \tag{4.17}$$

The likelihood in (4.17) uses the first joint model specification from (4.13). The computationally simpler log-likelihood based on the equivalent model specification is

$$l = \sum_{i=1}^{n} \Delta_i \ln \left[f(V_{H_{X_i}}|X_i) \cdot f(X_i) \right] + (1-\Delta_i) \ln \left[S(X_i) \right] \tag{4.18}$$

Thus, estimation can be done by maximizing the log-likelihood in (4.18) instead.

4.3.3 Marker Observed at Set of Uninformative Measurement Times

We may also assume that each subject has a set of observed marker values measured at noninformative measurement times. In the simplest case, each subject may have a single measurement at some time τ. Once more, let $\{W_t\}_{0 \le t \le \tau}$ be the underlying Lévy process defined on the horizon $[0, \tau]$. Let $B\{t\}_{[0,\tau]}$ be the Lévy bridge process starting at 0 and ending at 1 at time τ. Then, we can write the process W_t as the bridge process $B\{t\}_{[0,\tau]}$ scaled by the observed value at time τ. As our model is specified on the H-transformed time scale, we can write $W_{H_t} = B\{H_t\}_{[0,H_\tau]} V_{H_\tau}$. We assume the same parameterization as in Section 4.3.2, where W_{H_t} will follow a Gamma distribution with shape αH_t and scale β, where $\alpha = \frac{\mu^2}{\sigma^2}$ and $\beta = \frac{\sigma^2}{\mu}$. Then, $B\{H_t\}_{[0,H_\tau]}$ will follow a Beta distribution with parameters $a = \alpha H_t$ and $b = \alpha(H_\tau - H_t)$, $0 \le t \le \tau$.

Thus, the survival function conditional on a single observed marker value will be

$$S(t|V_{H_\tau}) = \begin{cases} E_{\bar{W}}\big\{S(t|\bar{W}_{H_\tau})\big|V_{H_\tau}\big\} = E_B\big\{e^{-B\{H_t\}V_{H_\tau}}\big|V_{H_\tau}\big\}, & 0 \le t < \tau \\ E_{\bar{W}}\big\{S(t|\bar{W}_{H_\tau})\big|V_{H_\tau}\big\} = E_{\bar{W}}\big\{e^{-(V_{H_\tau}+W_{H_t-H_\tau})}\big|V_{H_\tau}\big\}, & t \ge \tau \end{cases}$$
$$= \begin{cases} \mathcal{L}_B(V_{H_\tau}) = M(\alpha H_t, \alpha H_\tau, -V_{H_\tau}), & 0 \le t < \tau \\ e^{-V_{H_\tau}}(1+\beta)^{-\alpha(H_t-H_\tau)}, & t \ge \tau \end{cases} \tag{4.19}$$

where M represents Kummer's confluent hypergeometric function of the first kind, which is the Laplace transform of a Beta-distributed random variable (*Hoyle*, 2010; *Suresh*, 2018).

In many cases, subjects may have multiple marker measurements $\mathbf{V}$ collected at a set of noninformative times $\boldsymbol{\tau} = \{\tau_1, \ldots, \tau_k\}$. At each τ_j, we assume $V_{H_{\tau_j}} = W_{H_{\tau_j}}$. Then the survival function conditional on $\mathbf{V}$ will be

$$S(t|\mathbf{V}) = \begin{cases} e^{-V_{H_{\tau_{k-1}}}} M\big(\alpha(H_t - H_{\tau_{k-1}}), \alpha(H_{\tau_k} - H_{\tau_{k-1}}), -(V_{H_\tau} - V_{H_{\tau_{k-1}}})\big), & \tau_{k-1} \le t < \tau_k \\ e^{-V_{H_{\tau_k}}}(1+\beta)^{-\alpha(H_t - H_{\tau_k})}, & t \ge \tau_k \end{cases} \tag{4.20}$$

The survival function in (4.20) depends only on the marker values immediately before or after t. Without loss of generality, we assume that t falls between the last two consecutive measurement times in the first case in (4.20). Note that in this particular case, we assume that occurrence of the event does not terminate observation of the marker process, i.e., the event is non-terminal. From the survival function, the conditional hazard can be calculated as

$$d\Lambda(t|\mathbf{V}) = \begin{cases} -\frac{d}{dt}\ln M\big(\alpha(H_t - H_{\tau_{k-1}}), \alpha(H_{\tau_k} - H_{\tau_{k-1}}), -(V_{H_{\tau_k}} - V_{H_{\tau_{k-1}}})\big)dt, & \tau_{k-1} \le t < \tau_k \\ \alpha \ln(1+\beta) dH_t, & t \ge \tau_k \end{cases} \tag{4.21}$$

For subjects with no observed marker measurements, the marginal survival and

hazard functions are given by

$$\begin{aligned} S(t) &= (1+\beta)^{-\alpha H_t} \\ d\Lambda(t) &= \alpha \ln(1+\beta) dH_t \end{aligned} \tag{4.22}$$

The survival function in (4.22) will be equivalent to the one in (4.20) if $\tau_k = 0$ and $V_{H_{\tau_k}} = 0$, for $t \geq \tau_k$. Note that the marginal functions in (4.22) also match those in (4.16) from Section 4.3.2, as expected. Assuming $\tau_0 = 0$ and $V_{H_{\tau_0}} = 0$ and using the independent increments assumption, the joint distribution of all marker measurements is given by

$$\begin{aligned} g(\mathbf{V}) &= g(V_{H_{\tau_1}}, V_{H_{\tau_2}}, \ldots, V_{H_{\tau_k}}) \\ &= g(V_{H_{\tau_1}}) g(V_{H_{\tau_2}} | V_{H_{\tau_1}}) \cdot \cdots \cdot g(V_{H_{\tau_k}} | V_{H_{\tau_{k-1}}}) \\ &= \prod_{j=1}^{k} \frac{(V_{H_{\tau_j}} - V_{H_{\tau_{j-1}}})^{\alpha(H_{\tau_j} - H_{\tau_{j-1}})-1} e^{-(V_{H_{\tau_j}} - V_{H_{\tau_{j-1}}})/\beta}}{\Gamma\big(\alpha(H_{\tau_j} - H_{\tau_{j-1}})\big) \beta^{\alpha(H_{\tau_j} - H_{\tau_{j-1}})}} \end{aligned} \tag{4.23}$$

Details are provided in Appendix C.2.

To set up the log-likelihood, we let δ_i be an indicator of whether subject i has at least one observed marker value. The observed data for each subject is $\mathcal{O}_i = \{X_i, \Delta_i, \boldsymbol{Z}_i, \delta_i, \mathbf{V}_i, \boldsymbol{\tau}_i\}_{i=1}^n$, where $\boldsymbol{\tau}_i = \{\tau_{ij} : j = 1, \ldots, k_i\}$ is the set of k_i measurement times at which we observe W for subject i, and $\mathbf{V}_i = \{V_{H_{\tau_{ij}}} : j = 1, \ldots, k_i\}$ is the set of corresponding observed marker values. Estimation is done by maximizing the log-likelihood:

$$\begin{aligned} l = \sum_{i=1}^{n} \Delta_i \ln \Big\{ \big[d\Lambda(X_i|\mathbf{V}_i) S(X_i|\mathbf{V}_i) g(\mathbf{V}_i) \big]^{\delta_i} \big[d\Lambda(X_i) S(X_i) \big]^{1-\delta_i} \Big\} \\ + (1-\Delta_i) \ln \Big\{ \big[S(X_i|\mathbf{V}_i) g(\mathbf{V}_i) \big]^{\delta_i} S(X_i)^{1-\delta_i} \Big\} \end{aligned} \tag{4.24}$$

In the present study, we focus largely on the case of a single marker observed at an informative event time for methodological development. This case can be extended further by incorporating (1) multiple uninformative observation times and a single

informative observation time; (2) multiple informative observation times; (3) multiple uninformative and informative observation times. Using the derivations in Sections 4.3.2 and 4.3.3, extensions to these additional cases are straightforward.

4.3.4 Nonparametric Estimation of the Cumulative Baseline Hazard

In the preceding sections, the likelihood was constructed using a general function $H(t)$ to represent the cumulative baseline hazard time transformation. The simplest estimation method would assume a parametric function for H, as in Suresh (2018). However, the assumption of a particular parametric form may not accurately describe the true underlying hazard process. An alternative approach uses a nonparametric estimator to approximate $H(t)$.

Using ideas from Chapter II, we may use a Breslow-type estimator for the set of jumps $\{dH_t\}$ observed at each event time to approximate the cumulative baseline hazard function. We construct a martingale estimating equation based on the marginal survival model $S(t) = (1+\beta)^{-\alpha H_t}$.

$$\begin{aligned}
\sum_{i=1}^{n} dN_i(t) &= \sum_{i=1}^{n} Y_i(t) d\Lambda(t|\boldsymbol{Z}_i) = \sum_{i=1}^{n} Y_i(t)\alpha_i \ln(1+\beta_i) dH_t \\
\implies \widehat{dH_t} &= \frac{\sum_{i=1}^{n} dN_i(t)}{\sum_{i=1}^{n} Y_i(t)\alpha_i \ln(1+\beta_i)} \\
&= \frac{\sum_{i=1}^{n} I(X_i = t)\Delta_i}{\sum_{i=1}^{n} I(X_i \geq t)\alpha_i \ln(1+\beta_i)}
\end{aligned} \tag{4.25}$$

where $dN_i = I(X_i = t, \Delta_i = 1)$ is the increment of the observed counting process N and $Y_i(t) = I(X_i \geq t)$ is the at-risk process. The subscripts on α and β in (4.25) indicate that the parameters are subject-specific. Note that the expression for $\widehat{dH_t}$ in (4.25) is obtained by applying the definition of a functional derivative. Use of the marginal survival model allows for both subjects with marker values and subjects without marker values to contribute to estimation of the cumulative baseline hazard function.

When H is estimated using the Breslow estimator in (4.25), the marker parameters can be estimated using an iterative quasi-profile maximum likelihood algorithm. At the first step, we assume α and β are known and $\hat{H}^{(0)}$ can be computed from the Nelson-Aalen estimator given in equation (4.26).

$$\begin{aligned} \widehat{dH_t}^{NA} &= \frac{\frac{1}{n}\sum_{i=1}^{n} dN_i(t)}{\frac{1}{n}\sum_{i=1}^{n} Y_i(t)} \\ \widehat{H_t}^{NA} &= \int_0^t \frac{\frac{1}{n}\sum_{i=1}^{n} dN_i(s)}{\frac{1}{n}\sum_{i=1}^{n} Y_i(s)} \end{aligned} \tag{4.26}$$

The following steps are repeated until convergence is achieved, meaning that the difference between consecutive estimators is less than a pre-specified tolerance ξ:

1. For fixed $\alpha^{(k)}$, $\beta^{(k)}$, obtain $\hat{H}^{(k+1)}$ from equation (4.25) by estimating the set of jumps $\{dH\}$ at the observed failure times. This is done by recurrently solving an equation for each jump.

2. Update $\alpha^{(k+1)}$, $\beta^{(k+1)}$ by maximizing the log-likelihood using $\hat{H}^{(k+1)}$.

Asymptotic results for the marker parameters estimated using the Breslow-estimated H are expected to be close to the maximum likelihood estimators based on the results of Chapter II. Due to the structure of the proposed joint model, the variances may have complex expressions, but they can be approximated empirically via the bootstrap procedure.

4.4 Simulation Study

To assess the finite-sample properties of the proposed estimators, we evaluated their performance in multiple simulation settings for a marked survival endpoint. The settings are modified from the ones used in Suresh (2018). We consider a single time-independent baseline covariate $Z \sim Bernoulli(0.3)$. The baseline cumulative

hazard function is assumed to be Weibull with $H_t = t^2$, implying increasing baseline hazards. The true failure time T for each subject was simulated from the marginal model given in (4.16), with the appropriate marker parameters substituted. A Gamma marker value was simulated at the event time T assuming $\mu = \exp\{\nu_0 + \nu_1 Z\}$ and $\sigma^2 = \exp\{\eta_0 + \eta_1 Z\}$. In all simulations, ν_0 was restricted to be 0 for identifiability. The true parameter values were $(\nu_1, \eta_0, \eta_1) = (0.3, -2.1, 0.5)$, as in Suresh (2018). Uniform censoring over the interval $(0, c)$ was assumed with $c = 4$ and $c = 3$ corresponding to approximately 20% and 30% censoring, respectively. Sample sizes of 250, 500, and 1,000 were considered. For each simulation setting, 1,000 data sets were generated. A summary of all simulation settings is provided in Table 4.1.

In each simulation, estimates resulting from use of a parametric Weibull cumulative baseline hazard function and a nonparametric Breslow-type estimator of H were compared. For simulations based on the Breslow estimator of H, standard errors were estimated using parametric bootstrap samples. One-hundred bootstrap samples were taken per simulation, and the standard errors of each parameter were estimated as the standard deviation of the respective bootstrap sample estimates. Further details on the simulations are provided in Appendix C.3.

Table 4.1: Settings used in simulation study. n represents the sample size. k and λ are parameters of the Weibull baseline hazard function. $Z \sim Bernoulli(p)$. ν_1 and (η_0, η_1) relate to mean and variance, respectively, of the Gamma process representing the marker. c = administrative censoring time. Exp Censoring = % censoring anticipated. Obs Censoring = % censoring observed.

Setting	n	k	λ	p	ν_1	(η_0, η_1)	c	**Exp Censoring**	**Obs Censoring**
1	250	2	1	0.3	0.3	(-2.1, 0.5)	4	20%	21.8%
2	250	2	1	0.3	0.3	(-2.1, 0.5)	3	30%	29.2%
3	500	2	1	0.3	0.3	(-2.1, 0.5)	4	20%	21.9%
4	500	2	1	0.3	0.3	(-2.1, 0.5)	3	30%	29.1%
5	1000	2	1	0.3	0.3	(-2.1, 0.5)	4	20%	21.9%
6	1000	2	1	0.3	0.3	(-2.1, 0.5)	3	30%	29.2%

Results of the simulations based on a parametric H are given in Tables 4.2 and 4.3. Overall, the estimates exhibit little bias and have coverage near the expected level of 95%. Average standard errors (ASEs) and empirical standard deviations (ESDs) are in good agreement for each parameter. In general, the ASEs and ESDs for the settings with $c = 4$, corresponding to approximately 20% censoring, are slightly lower. From Table 4.3, estimates of k and λ, the parameters governing the underlying Weibull cumulative baseline hazard function, are unbiased with ASEs and ESDs that are equal in each setting. Coverage probabilities for k and λ are close to the nominal 95% level as well.

Table 4.2: Simulation study results with parametric Weibull H. True parameter values are $(\nu_1, \eta_0, \eta_1) = (0.3, -2.1, 0.5)$. Est = average estimate across all simulations. ASE = average standard error. ESD = empirical standard deviation. CP (%) = coverage probability for 95% Wald-based interval.

Setting	ν_1				η_0				η_1			
	Est	**ASE**	**ESD**	**CP**	**Est**	**ASE**	**ESD**	**CP**	**Est**	**ASE**	**ESD**	**CP**
1	0.30	0.06	0.06	95.4	-2.12	0.12	0.12	94.6	0.48	0.24	0.24	94.3
2	0.30	0.06	0.06	95.3	-2.13	0.13	0.13	94.8	0.50	0.25	0.26	94.7
3	0.30	0.04	0.04	95.0	-2.11	0.09	0.09	96.0	0.48	0.17	0.17	95.2
4	0.30	0.04	0.04	93.8	-2.11	0.09	0.09	95.0	0.49	0.18	0.18	94.8
5	0.30	0.03	0.03	94.5	-2.10	0.06	0.06	95.2	0.49	0.12	0.12	95.5
6	0.30	0.03	0.03	94.4	-2.10	0.06	0.06	94.7	0.49	0.12	0.13	93.8

Table 4.3: Simulation study results using Weibull H. True Weibull hazard parameter values are $(k, \lambda) = (2, 1)$. Est = average estimate across all simulations. ASE = average standard error. ESD = empirical standard deviation. CP (%) = coverage probability for 95% Wald-based interval.

Setting	k				λ			
	Est	**ASE**	**ESD**	**CP**	**Est**	**ASE**	**ESD**	**CP**
1	2.00	0.05	0.05	95.4	1.00	0.02	0.02	94.7
2	2.00	0.06	0.06	94.0	1.00	0.02	0.02	94.7
3	2.00	0.04	0.04	95.5	1.00	0.01	0.01	95.4
4	2.00	0.04	0.04	95.2	1.00	0.01	0.01	95.5
5	2.00	0.03	0.03	93.5	1.00	0.01	0.01	95.3
6	2.00	0.03	0.03	94.9	1.00	0.01	0.01	94.5

Simulation results based on the Breslow-estimated H are presented in Table 4.4. Parameter estimates exhibit slight bias for small sample sizes, particularly for the variance parameters η_0 and η_1. However, as the sample size increases, bias decreases and average parameter estimates approach the true values. There is generally good agreement between the bootstrap ASEs and ESDs. For, η_1, however, there are noticeable differences for small samples, but these decrease as the sample size increases. Coverage probabilities for ν_1 are close to the nominal 95% level, but for η_0 and η_1, they are inflated, suggesting that the intervals are conservative. These parameters model the variance of the marker process, which may be difficult to ascertain based on a single observed value. With multiple marker measures at informative or uninformative times, coverage would likely be closer to 95%. Nonetheless, with large sample sizes, coverage should approach the nominal level. Based on these trends, we expect that asymptotically, the estimators will be unbiased with 95% confidence intervals that achieve proper coverage.

Table 4.4: Simulation study results with Breslow estimated H. True parameter values are $(\nu_1, \eta_0, \eta_1) = (0.3, -2.1, 0.5)$. Est = average estimate across all simulations. ASEB = average standard error estimated from bootstrap samples. ESD = empirical standard deviation. CPB (%) = coverage probability for 95% Wald-based interval constructed using bootstrap standard error.

Setting	ν_1				η_0				η_1			
	Est	**ASEB**	**ESD**	**CPB**	**Est**	**ASEB**	**ESD**	**CPB**	**Est**	**ASEB**	**ESD**	**CPB**
1	0.28	0.14	0.12	96.9	-1.94	0.24	0.21	97.3	0.36	0.46	0.35	98.8
2	0.29	0.14	0.13	96.6	-1.95	0.25	0.22	96.2	0.39	0.48	0.37	99.3
3	0.29	0.09	0.09	95.7	-2.01	0.15	0.14	97.2	0.42	0.30	0.24	98.5
4	0.29	0.10	0.09	96.0	-2.01	0.16	0.14	96.6	0.43	0.32	0.25	98.7
5	0.30	0.06	0.06	94.9	-2.05	0.10	0.09	96.9	0.46	0.21	0.17	98.2
6	0.30	0.07	0.06	95.7	-2.06	0.11	0.10	97.0	0.46	0.22	0.17	98.2

4.5 SEER Data Analysis

We apply the proposed joint model to prostate cancer incidence data collected by the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER)

program. In SEER incidence data, PSA is collected at the age of diagnosis for men diagnosed with prostate cancer in a given year, thus this is an example of a marked survival endpoint. Incidence data for 1973-2013 were obtained from the SEER registry. This database includes PSA values in the range of 0.0 ng/mL to 97.9 ng/mL. Any values $\geq$ 98.0 ng/mL were discarded due to these representing missing values. Prior work has investigated the error rate of PSA values reported in SEER and found that the error was less than 10% (*Adamo et al.*, 2016; *Schultheiss et al.*, 2016), thus it is reasonable to use the PSA values reported in SEER for the purposes of methodological illustration. More details about the SEER registry and how PSA information is collected and coded can be found at `https://seer.cancer.gov/`.

In the incidence data, PSA was only recorded for patients diagnosed between 2010 and 2013, so analysis was restricted to patients diagnosed during this interval. In addition, analysis was restricted to men aged 40-84. The primary covariate we consider is race, which is coded as White, Black, or Other. Less than 5% of men diagnosed with prostate cancer between 2010 and 2013 were of Other races, thus analysis focused on the subset of White and Black men.

SEER incidence data only records men diagnosed with prostate cancer. In order to calculate the survival curve for time to diagnosis with prostate cancer, information on the risk set at each age of diagnosis in a given calendar year is needed. Population counts for the years 1969-2017 were downloaded from the SEER registry. Risk set information for men of each race category considered for analysis (White or Black) for the years 2010-2013 was used.

Because the risk set is presented in terms of group counts, to perform estimation, we must rewrite the log-likelihood in (4.18) to be compatible with cross-sectional risk set information. This can be done by expressing the log-likelihood in counting process

notation:

$$l = \int_0^{\tau} \sum_{i=1}^{n} dN_i(u) \big[\ln \big(f(V_{H_u} | u, \boldsymbol{Z}_i) \big) + \ln \big(d\Lambda(u | \boldsymbol{Z}_i) \big) \big] - \sum_{\boldsymbol{z_m} \in \mathcal{Z}} Y_m(u) d\Lambda(u | \boldsymbol{z_m}) \quad (4.27)$$

where $dN_i(u) = I(X_i = u, \Delta_i = 1)$ is the increment of the counting process N, $\boldsymbol{z_m}$ is a covariate value in the set of all unique covariate values $\mathcal{Z}$, and $Y_m(u) = \sum_{i=1}^{n} I(X_i \geq u, \boldsymbol{Z}_i = \boldsymbol{z_m})$ is the at-risk indicator for all subjects at risk at time u with covariate value $\boldsymbol{z_m}$. A detailed derivation of (4.27) is given in Appendix C.4. Note that the first sum in (4.27) is taken over individual subjects who are diagnosed with prostate cancer and have PSA available, whereas the second sum is over unique covariate values and is for the entire set at risk at time u.

PSA is modeled as a Gamma process with mean $\mu H(t)$ and variance $\sigma^2 H(t)$, where μ is specified as $\exp\{\nu_0 + \nu_1 \text{I(Race=Black)}\}$ and σ^2 is specified as $\exp\{\eta_0 + \eta_1 \text{I(Race=Black)}\}$. A common $H(t)$ is estimated using the Breslow-type estimator defined in (4.25). Standard errors of the marker parameters were estimated using 100 bootstrap samples of the set of men diagnosed with prostate cancer. Marker parameters were estimated separately for years 2010, 2011, 2012, and 2013. For illustrative purposes, we present only the results of the year 2010.

In total, 157,857 men had PSA available at the age of diagnosis for the years 2010 to 2013. Of this group of men, 43,018 (27.3%) were diagnosed in 2010, 43,406 (27.5%) were diagnosed in 2011, 36,538 (23.1%) were diagnosed in 2012, and 34,895 (22.1%) were diagnosed in 2013. Observed survival curves summarized using the Kaplan-Meier method for time to prostate cancer diagnosis stratified by race and the distribution of PSA values for the year 2010 are displayed in Figures 4.1 and 4.2, respectively. From Figure 4.1, Black men tend to be diagnosed earlier compared to White men. The marginal distribution of PSA values in Figure 4.2 is right-skewed with mean and median values of 9.3 ng/mL and 6.1 ng/mL, respectively.

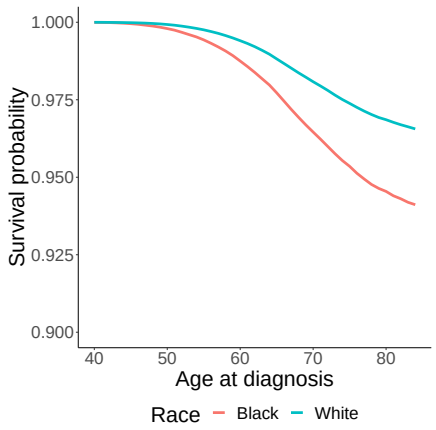


Figure 4.1: Observed age of prostate cancer diagnosis by race, 2010.

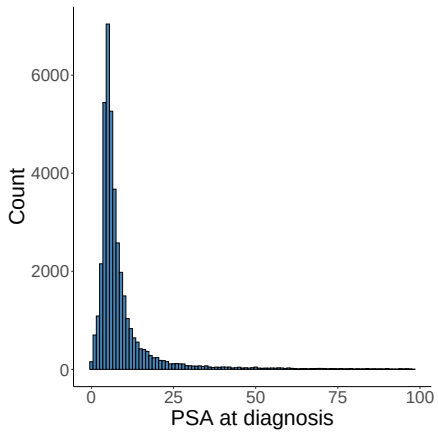


Figure 4.2: Distribution of PSA at age of prostate cancer diagnosis, 2010.

A summary of our marker parameter estimates for 2010 is given in Table 4.5. Using a Wald-based test with the bootstrap standard error, each estimate is statistically significant at the $\alpha = 0.05$ level. In particular, the significance of ν_1 and η_1 indicate significant differences in the mean and variance of the underlying PSA process observed at age of diagnosis by race.

Table 4.5: Marker parameter estimates in 2010. 95% Wald-based interval constructed using bootstrap standard error. p-value estimated using Wald Z value $= \frac{\text{Est}}{\text{Bootstrap SE}}$.

Parameter	**Estimate**	**Bootstrap SE**	**95% CI**	p
ν_0	2.811	0.006	$(2.798, 2.824)$	< 0.001
ν_1	0.775	0.022	$(0.733, 0.818)$	< 0.001
η_0	6.255	0.014	$(6.228, 6.283)$	< 0.001
η_1	1.036	0.040	$(0.958, 1.114)$	< 0.001

Estimated and observed curves for time to diagnosis are displayed in Figure 4.3 with 95% Wald-based confidence bands obtained using the bootstrapped standard errors. Use of the Breslow estimator for H results in a close fit between the estimated and observed curves, particularly for White men. For Black men, there is separation

between the observed and estimated curves at the tail, but overall, the fitted survival curve accurately describes the shape of the observed curve. We conduct a likelihood ratio test for the null hypothesis that there is no difference between μ and σ^2 by race, which is equivalent to testing $H_0 : \nu_1 = 0, \eta_1 = 0$. We observe a significant difference ($p < 0.001$), confirming the results in Table 4.5 and providing further evidence of a difference in the underlying distribution of the PSA process related to prostate cancer diagnosis by race.

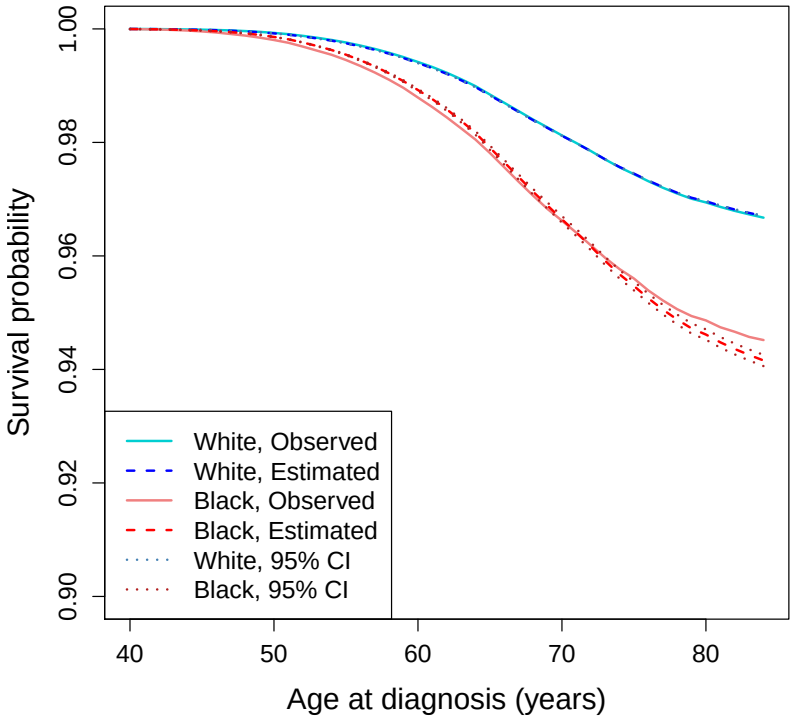


Figure 4.3: Observed (solid) and estimated (dashed) survival curves for time to prostate cancer diagnosis for men diagnosed in 2010. 95% Wald-based confidence bands (dotted) constructed using bootstrap standard errors.

The tables and figures analogous to Table 4.5 and Figures 4.1-4.3 for years 2011, 2012, and 2013 are included in Appendix C.5.

4.6 Discussion

We have developed a novel joint model where the partially observed biomarker is modeled using a stochastic process from the Lévy family. Use of a stochastic model for the biomarker avoids restriction of patterns to a given functional form. In addition, it accounts for biological variability by incorporating individual fluctuations. While we make use of the Gamma process due to its tractability and attractive mathematical properties, our model can be generalized by considering alternative Lévy processes as long as they satisfy the conditions for the hazard and cumulative hazard function. In our model, the marker is subordinated by the cumulative baseline hazard function $H(t)$, essentially representing the fact that time is on the scale of accumulated risk. The benefit of using this approach is that is allows for greater flexibility in modeling the marker process.

We consider both a parametric function for H and estimating H using the Breslow-type estimator derived from a martingale estimating equation based on the marginal survival model. Our results demonstrate the potential to use a Breslow-approximated H in large samples. The poorer performance of the Breslow estimator relative to a parametric form for H in small samples is in line with the general loss in power associated with nonparametric methods. Based on the results in Chapter II, we expect estimates based on the Breslow estimator to be close to those derived using the nonparametric maximum likelihood estimator (NPMLE) of H, which can be obtained by maximizing the log-likelihood with respect to the jump at any event time. While we approximate the standard errors of our marker parameter estimators using the parametric bootstrap procedure, closed forms may be derived using a partial likelihood argument. We plan to investigate this in future work.

For model development and illustration, we focused in particular on the case of marked survival wherein the continuous marker is observed at the event time for subjects who experience the event of interest. While our work largely focused on the

marked survival model, the methods we develop can also be applied to settings with longitudinally collected biomarkers. Previous work has investigated marked survival endpoints with a categorical mark (*Hu and Tsodikov*, 2014). In Hu and Tsodikov (2014), the mark of interest was stage of disease for investigating time to recurrence for breast cancer. Using the framework developed in Chapter III, we construct mathematically valid joint models in the case of marked survival with a continuous mark.

In our study, we consider application of our method to the continuous biomarker PSA and time to diagnosis with prostate cancer. In SEER data where PSA is known at the age of diagnosis, our method performed well and provided good fit to the observed survival curves stratified by race. We observed a significant difference in the mean and variance of the underlying PSA process observed at the age of prostate cancer diagnosis by race. This difference may be due to socioeconomic differences between the two groups, differences in access to healthcare, or differences in PSA screening. Our findings are hypothesis generating and may be used to drive changes in screening practices in order to eliminate the observed disparity between White and Black men.

Part of what sets our method apart from other joint modeling approaches is the survival model we propose conditional on the observed history of the marker process. We make use of the ideas explored in Chapter III regarding the incorporation of endogenous time-varying covariates into survival models. Namely, we first specify the survival at time t conditional on the fixed marker history observed up to time x and then substitute $x = t$ once the model is fully specified. This allows us to construct a model that is mathematically well-defined, satisfies the exponential relationship between the survival and hazard functions, and is consistent with intuition. However, the complexity of the marginal distribution in this specification of the joint model makes maximum likelihood estimation difficult. Because we consider a marked survival endpoint, we may reexpress the joint distribution in a way that makes es-

timation more straightforward. Our simulation results demonstrate that use of this equivalent form of the joint model results in proper inference.

One potential weakness of our model is the need for a large sample size in order to achieve identifiability of the parameters and appropriate performance of the Breslow estimator. Estimation of parameters governing the variance of the marker process may be especially difficult to estimate when only a single marker value is observed, and this may be even more difficult in small samples. Thus, a large sample size and adequate variability among subjects in marker values may be needed to reliably estimate parameters. This may be why the results based on SEER data, which is collected at the population level, exhibit much better fit relative to the simulation results, which were based on samples consisting of no more than 1,000 subjects, with only a subset of subjects having marker data available. In small samples, it may be better to assume a parametric form for H in order to obtain unbiased estimates of the marker parameters. In our SEER analysis, we use PSA as the marker of interest. While there has been controversy surrounding the use of PSA in the diagnosis and surveillance of prostate cancer, we focus on its use as a marker for the purposes of model development. In this work, we do not seek to comment on its performance as a diagnostic marker for prostate cancer. In addition, we consider race as the sole covariate, but there is the potential to adjust for other factors such as socioeconomic status or healthcare access that may explain part of the difference we observed between White and Black men.

There are many potential extensions to this work. First, we can explore incorporating potential measurement error in the marker process as PSA may be observed with some uncertainty. In addition, we may want to extend our method to incorporate multiple markers. One relatively straightforward way is to assume an additive effect of the markers on survival. Assuming we have markers M_1 and M_2 both > 0, we can write the conditional survival function as $S(t|\bar{M}_1, \bar{M}_2) = e^{-[M_1(H_t)+M_2(H_t)]}$.

Alternative ways to incorporate multiple continuous marks can be explored further. Another natural extension would be to consider combinations of uninformative and informative observation times and specifying joint models in the stochastic marker framework for a multivariate survival outcome. We can also consider relaxing the proportional hazards structure inherent to the marginal survival model we developed and utilizing alternative estimators of the cumulative baseline hazard function $H(t)$.

CHAPTER V

Conclusion

In this dissertation, we aim to extend semiparametric survival models to settings where longitudinal biomarker data may provide additional information regarding a patient's risk of experiencing disease-related events. In particular, we focus on the case where the marker of interest is endogenous or "internal." The methods explored in this dissertation can be widely applied in clinical research. With the advent of screening for a variety of diseases and increased interest in personalized treatment decisions, the methods developed in this dissertation can hopefully extend the utility of longitudinally collected markers that may act as surrogates for the underlying disease process.

In Chapter II, we focused on comparing two estimators of the cumulative baseline hazard function in the class of semiparametric transformation models for survival data. The two estimators of interest were the nonparametric maximum likelihood estimator (NPMLE) and a Breslow-type estimator derived from a martingale estimating equation. The principal difference between the two estimators stems from dependence on the future: the NPMLE relies on future data for predictions of estimated risk, whereas the Breslow-type estimator relies solely on event information at the current point in time. We derive the asymptotic relative efficiency of the Breslow-type estimator and demonstrate via simulation that there are settings in which the

Breslow-type estimator is slightly more efficient than the NPMLE. The asymptotic relative efficiency was shown to be reliant on the censoring horizon, the covariate distribution, and true regression parameter values.

As interest in semiparametric survival models lies in the regression parameters, we compared the estimates of such parameters resulting from applying either the NPMLE or Breslow-type estimator in finite samples. We demonstrated that despite the perceived inefficiency of the Breslow-type estimator, it had nearly identical performance to the NPMLE in terms of variance, and average estimates of regression parameters were close as well. Thus, the Breslow-type estimator may be utilized with minimal loss of efficiency in finite samples, and it may be preferred for certain interim analyses. However, caution is needed before applying the Breslow-type estimator in place of the NPMLE in all survival models, and further investigation of the relative efficiency under more complex censoring mechanisms and in the presence of endogenous time-varying covariates is needed. Nonetheless, the Breslow-type estimator represents a conceptually and computationally simpler alternative to the NPMLE that may be used in a variety of survival models with minimal loss of efficiency.

In Chapter III, we reexamine the role of internal or endogenous time-varying covariates in survival models. We explore the differences between endogenous covariates and external or exogenous covariates. The current approach, based largely on intuition, ignores any potentially useful information these covariates may provide beyond the event time. Ignoring future covariate information may result in less accurate risk predictions. In fact, we demonstrate via the mathematical definition of an endogenous covariate that the hazard of failure at any time u conditional on the marker history up to u is related to the future path of the marker up to time t, where $t > u$. Through this inherent dependence on the future, we formulate the difference between endogenous and exogenous covariates as a missing data problem, where the future trajectory of the marker constitutes the missing data.

Through this formulation, we mechanistically define a survival function conditional on the history of an endogenous marker. Our survival function utilizes the future marker trajectory as a latent variable that is averaged out given the observed marker trajectory. As a consequence, our survival function satisfies all the necessary properties and also has a legitimate exponential form that links the hazard and survival functions. As part of our discussion on exogenous and endogenous covariates, we clarify that the legitimacy of the exponential form relates to continuity of the event time T, which may be influenced by the marker. However, there is no one-to-one correspondence between the type of time-varying covariate and whether the exponential form will be valid.

We illustrate our novel survival function through bivariate shared frailty models, bivariate semicompeting risks-type models, and threshold regression models. Through each example, we highlight the shortcomings of the current understanding of time-varying covariates and the utility of our modeling approach. The examples we consider represent a subset of settings where endogenous marker information may be available, and further study is needed to grasp how our survival function may be applied to other models. Regardless, the benefit of our framework is that it is general, and we anticipate it being applied to a variety of settings with endogenous markers.

In Chapter IV, we apply the model developed in Chapter III to the setting of a joint model of survival and biomarker data. In traditional joint models, a model for survival conditional on the marker and a model for the marker are separately specified but linked through shared latent factors. In the specific context we consider, the biomarker is measured longitudinally at a set of measurement times, where these times may be uninformative (e.g. scheduled clinic visits) or related to the event time of interest. The simplest informative observation time we consider is the event time T itself, known as "marked survival." We specify the survival submodel using the general mechanistic model developed in Chapter III, where we take the expectation

over the latent marker trajectory conditional on the observed marker values. We specify the marker as a time-transformed Lévy stochastic process, which allows for greater flexibility and maintains necessary properties of the hazard and cumulative hazard functions. The specific transformation we consider is the cumulative baseline hazard function itself. We demonstrate that our model results in unbiased estimates when a parametric time transformation is used, and there is potential to use a non-parametrically estimated transformation.

One drawback associated with our proposed joint model is that it is heavily dependent on the specific properties of the marker process model. In our setup, we considered the Gamma process due mathematical tractability and its inherent monotonicity. However, further extensions of our model could make use of alternative processes from the Lévy family. Additionally, our method utilizes either a parametric form or a Breslow-type estimator for the cumulative baseline hazard. Use of the Breslow estimator required a large sample size for unbiased estimates. Based on results from Chapter II, we expect minimal loss of efficiency relative to the NPMLE due to the use of this estimator. While our simulation results demonstrate the promise of using the Breslow-type estimator in estimation of the marker parameters, the asymptotic properties of our estimators need to be investigated further to ensure proper inference.

An extension to the marker process model we developed would be to incorporate measurement error as the observed value may not be exactly equal to the true value. To ensure desirable mathematical properties, we may be limited to a subset of possible models. Two examples that may be explored, as in Suresh (2018), are the compound Gamma model and the Normal-Gamma model. Introducing an additional layer of modeling could also allow marker values to exhibit non-monotonic behavior over time, which is likely the case for longitudinally collected biomarkers.

The methods explored in this dissertation are widely applicable to clinical research

studies. They are particularly suitable to settings in which survival and biomarker data are collected with the aim of using both types of data to predict subject-specific risk of experiencing a failure of interest. Our methods can be applied when the biomarker is collected at uninformative or informative observation times, most notably when the marker is measured at the event time. We make use of previous ideas in survival analysis and stochastic process theory to develop our methods. We hope that this work extends the current understanding of incorporating biomarkers into survival models, and we anticipate further avenues of research related to the ideas explored in this dissertation.

APPENDICES

APPENDIX A

Efficiency of the Breslow Estimator in Semiparametric Transformation Models

A.1 Expectation of Weight in NPMLE

Using the Law of Total Expectation, martingale properties, and assuming $u \geq t$

$$
\begin{aligned}
E(w_i(H_t|\boldsymbol{\beta}, \boldsymbol{Z}_{t,i})) &= E\bigg(1 - \frac{\int_{t^+}^{\tau} \big[\Theta^0(H_u|\boldsymbol{\beta}, \boldsymbol{Z}_{u,i}) - \Theta^1(H_u|\boldsymbol{\beta}, \boldsymbol{Z}_{u,i})\big] dM_i(u)}{\Theta^0(H_t|\boldsymbol{\beta}, \boldsymbol{Z}_{t,i})}\bigg) \\
&= 1 - E\bigg(\frac{\int_{t^+}^{\tau} \big[\Theta^0(H_u|\boldsymbol{\beta}, \boldsymbol{Z}_{u,i}) - \Theta^1(H_u|\boldsymbol{\beta}, \boldsymbol{Z}_{u,i})\big] dM_i(u)}{\Theta^0(H_t|\boldsymbol{\beta}, \boldsymbol{Z}_{t,i})}\bigg) \\
&= 1 - E\bigg\{E\bigg(\frac{\int_{t^+}^{\tau} \big[\Theta^0(H_u|\boldsymbol{\beta}, \boldsymbol{Z}_{u,i}) - \Theta^1(H_u|\boldsymbol{\beta}, \boldsymbol{Z}_{u,i})\big] dM_i(u)}{\Theta^0(H_t|\boldsymbol{\beta}, \boldsymbol{Z}_{t,i})}\bigg|\mathcal{F}_{t^-}\bigg)\bigg\} \\
&= 1 - E\bigg\{\frac{\int_{t^+}^{\tau} E\big(E\big(\big[\Theta^0(H_u|\boldsymbol{\beta}, \boldsymbol{Z}_{u,i}) - \Theta^1(H_u|\boldsymbol{\beta}, \boldsymbol{Z}_{u,i})\big] dM_i(u)|\mathcal{F}_{u^-}\big)|\mathcal{F}_{t^-}\big)}{\Theta^0(H_t|\boldsymbol{\beta}, \boldsymbol{Z}_{t,i})}\bigg\} \\
&= 1 - E\bigg\{\frac{\int_{t^+}^{\tau} E\big(\big[\Theta^0(H_u|\boldsymbol{\beta}, \boldsymbol{Z}_{u,i}) - \Theta^1(H_u|\boldsymbol{\beta}, \boldsymbol{Z}_{u,i})\big] E\big(dM_i(u)|\mathcal{F}_{u^-}\big)|\mathcal{F}_{t^-}\big)}{\Theta^0(H_t|\boldsymbol{\beta}, \boldsymbol{Z}_{t,i})}\bigg\} \\
&= 1 - 0 \\
&= 1
\end{aligned}
\tag{A.1}
$$

A.2 Derivation of Profile Likelihood Score Equations

Let $l = l(\boldsymbol{\beta}, H)$ denote the log-likelihood. The score equation for the jump dH_t given $\boldsymbol{\beta}$ is

$$
\begin{aligned}
U_{dH_t}(\boldsymbol{\beta}) &= \frac{\partial}{\partial dH_t}\Big\{\sum_{i=1}^{n}\int_0^\tau dN_i(u)\big[\ln(\Theta^0(H_u|\boldsymbol{\beta},\boldsymbol{Z}_{u,i})) + \ln(dH_u)\big] - Y_i(u)\Theta^0(H_u|\boldsymbol{\beta},\boldsymbol{Z}_{u,i})dH_u\Big\} \\
&= \sum_{i=1}^{n}\int_0^\tau dN_i(u)\Big[\frac{\partial}{\partial dH_t}\ln(\Theta^0(H_u|\boldsymbol{\beta},\boldsymbol{Z}_{u,i})) + \frac{\partial}{\partial dH_t}\ln(dH_u)\Big] - Y_i(u)\frac{\partial}{\partial dH_t}\Theta^0(H_u|\boldsymbol{\beta},\boldsymbol{Z}_{u,i})dH_u \\
&= \sum_{i=1}^{n}\int_0^\tau dN_i(u)\Big[\frac{\frac{\partial}{\partial dH_t}\Theta^0(H_u|\boldsymbol{\beta},\boldsymbol{Z}_{u,i})}{\Theta^0(H_u|\boldsymbol{\beta},\boldsymbol{Z}_{u,i})} + \frac{\frac{\partial}{\partial dH_t}dH_u}{dH_u}\Big] - Y_i(u)\Big[\frac{\partial\Theta^0(H_u|\boldsymbol{\beta},\boldsymbol{Z}_{u,i})}{\partial dH_t}dH_u + \Theta^0(H_u|\boldsymbol{\beta},\boldsymbol{Z}_{u,i})\frac{\partial dH_u}{\partial dH_t}\Big] \\
&= \sum_{i=1}^{n}\int_0^\tau dN_i(u)\Big[\frac{\dot\Theta^0(H_u|\boldsymbol{\beta},\boldsymbol{Z}_{u,i})I(u\geq t)}{\Theta^0(H_u|\boldsymbol{\beta},\boldsymbol{Z}_{u,i})} + \frac{I(u=t)}{dH_u}\Big] - Y_i(u)\Big[\dot\Theta^0(H_u|\boldsymbol{\beta},\boldsymbol{Z}_{u,i})I(u\geq t)dH_u + \Theta^0(H_u|\boldsymbol{\beta},\boldsymbol{Z}_{u,i})I(u=t)\Big] \\
&= \sum_{i=1}^{n}\frac{dN_i(t)}{dH_t} - Y_i(t)\Theta^0(H_t|\boldsymbol{\beta},\boldsymbol{Z}_{t,i}) + \int_{t^+}^\tau dN_i(u)\frac{\dot\Theta^0(H_u|\boldsymbol{\beta},\boldsymbol{Z}_{u,i})}{\Theta^0(H_u|\boldsymbol{\beta},\boldsymbol{Z}_{u,i})} - Y_i(u)\dot\Theta^0(H_u|\boldsymbol{\beta},\boldsymbol{Z}_{u,i})dH_u \\
&= \sum_{i=1}^{n}\frac{dN_i(t)}{dH_t} - Y_i(t)\Theta^0(H_t|\boldsymbol{\beta},\boldsymbol{Z}_{t,i}) + \int_{t^+}^\tau \frac{\partial}{\partial H_u}\ln(\Theta^0(H_u|\boldsymbol{\beta},\boldsymbol{Z}_{u,i}))\big[dN_i(u) - Y_i(u)\Theta^0(H_u|\boldsymbol{\beta},\boldsymbol{Z}_{u,i})dH_u\big] \\
&= \sum_{i=1}^{n}\frac{dN_i(t)}{dH_t} - Y_i(t)\Theta^0(H_t|\boldsymbol{\beta},\boldsymbol{Z}_{t,i}) + \int_{t^+}^\tau \{\Theta^0(H_u|\boldsymbol{\beta},\boldsymbol{Z}_{u,i}) - \Theta^1(H_u|\boldsymbol{\beta},\boldsymbol{Z}_{u,i})\}\big[dN_i(u) - Y_i(u)\Theta^0(H_u|\boldsymbol{\beta},\boldsymbol{Z}_{u,i})dH_u\big] \\
&= \sum_{i=1}^{n}\frac{dN_i(t)}{dH_t} - Y_i(t)\Theta^0(H_t|\boldsymbol{\beta},\boldsymbol{Z}_{t,i}) + \int_{t^+}^\tau [\Theta^0(H_u|\boldsymbol{\beta},\boldsymbol{Z}_{u,i}) - \Theta^1(H_u|\boldsymbol{\beta},\boldsymbol{Z}_{u,i})]dM_i(u)
\end{aligned}
\tag{A.2}
$$

where $\dot\Theta^c$ denotes the first partial derivative of Θ^c with respect to the argument H. The profile likelihood score equation for $\boldsymbol{\beta}$ given the estimator $\hat H(\boldsymbol{\beta})$ is

$$
\begin{aligned}
U_{\boldsymbol{\beta}} &= \frac{\partial l(\boldsymbol{\beta},\hat H(\boldsymbol{\beta}))}{\partial\boldsymbol{\beta}} \\
&= \frac{\partial}{\partial\boldsymbol{\beta}}\Big\{\sum_{i=1}^{n}\int_0^\tau dN_i(u)\big[\ln(\Theta^0(\hat H_u(\boldsymbol{\beta})|\boldsymbol{\beta},\boldsymbol{Z}_{u,i})) + \ln(\widehat{dH_u}(\boldsymbol{\beta}))\big] - Y_i(u)\Theta^0(\hat H_u(\boldsymbol{\beta})|\boldsymbol{\beta},\boldsymbol{Z}_{u,i})\widehat{dH_u}(\boldsymbol{\beta})\Big\} \\
&= \sum_{i=1}^{n}\int_0^\tau dN_i(u)\frac{\frac{\partial}{\partial\boldsymbol{\beta}}\Theta^0(\hat H_u(\boldsymbol{\beta})|\boldsymbol{\beta},\boldsymbol{Z}_{u,i})}{\Theta^0(\hat H_u(\boldsymbol{\beta})|\boldsymbol{\beta},\boldsymbol{Z}_{u,i})} - Y_i(u)\frac{\partial}{\partial\boldsymbol{\beta}}\Theta^0(\hat H_u(\boldsymbol{\beta})|\boldsymbol{\beta},\boldsymbol{Z}_{u,i})\widehat{dH_u}(\boldsymbol{\beta}) \\
&= \sum_{i=1}^{n}\int_0^\tau \frac{\partial}{\partial\boldsymbol{\beta}}\ln(\Theta^0(\hat H_u(\boldsymbol{\beta})|\boldsymbol{\beta},\boldsymbol{Z}_{u,i}))\big[dN_i(u) - Y_i(u)\Theta^0(\hat H_u(\boldsymbol{\beta})|\boldsymbol{\beta},\boldsymbol{Z}_{u,i})\widehat{dH_u}(\boldsymbol{\beta})\big] \\
&= \sum_{i=1}^{n}\int_0^\tau \frac{\partial}{\partial\boldsymbol{\beta}}\ln(\Theta^0(\hat H_u(\boldsymbol{\beta})|\boldsymbol{\beta},\boldsymbol{Z}_{u,i}))\widehat{dM}_i(u)
\end{aligned}
\tag{A.3}
$$

where $\widehat{dM}$ is understood to be the increment of the martingale process with $\widehat{dH}(\boldsymbol{\beta})$ substituted for dH.

We want to express both of these scores as martingale transforms. For brevity, we use Θ^c to represent $\Theta^c(H_t|\boldsymbol{\beta}, \boldsymbol{Z}_t)$ and suppress the profile likelihood notation. The normalized scores based on a sample of size n can be expressed as

$$U_n(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \frac{\partial}{\partial\boldsymbol{\beta}}\ln(\Theta_i^0)dM_i(u)$$

$$U_n(H_t) = \frac{1}{n}\sum_{i=1}^{n}\frac{dM_i(t)}{dH_t} + \int_{t^+}^\tau [\Theta_i^0 - \Theta_i^1]dM_i(u)$$

Note that the score for $\boldsymbol{\beta}$ is already in the form of a martingale transform. Let us consider the score for H.

$$\begin{aligned}
\int_0^t U_n(H_x)dH_x &= \int_0^t \left\{\frac{1}{n}\sum_{i=1}^{n}\frac{dM_i(x)}{dH_x} + \int_{x^+}^\tau [\Theta_i^0 - \Theta_i^1]dM_i(u)\right\}dH_x \\
&= \frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{dM_i(x)}{dH_x}dH_x + \int_0^t\int_{x^+}^\tau [\Theta_i^0(u) - \Theta_i^1(u)]dM_i(u)dH_x \\
&= \frac{1}{n}\sum_{i=1}^{n}\int_0^t \frac{dM_i(x)}{dH_x}\int_{x^+}^\tau I(u=x)dH_u + \int_0^t\int_{x^+}^\tau [\Theta_i^1(u) - \Theta_i^1(u)]dM_i(u)dH_x \\
&= \frac{1}{n}\sum_{i=1}^{n}\int_0^t\int_{x^+}^\tau \frac{dM_i(u)}{dH_u}I(u=x)dH_u + [\Theta_i^0(u) - \Theta_i^1(u)]dM_i(u)dH_x \\
&= \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \int_0^{\min(u,t)} \frac{dM_i(u)}{dH_u}I(u=x)dH_u + [\Theta_i^0(u) - \Theta_i^1(u)]dH_x dM_i(u) \\
&\text{(switch order of integration)} \\
&= \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau dM_i(u)\int_0^{\min(u,t)} I(u=x) + \int_0^\tau [\Theta_i^0(u) - \Theta_i^1(u)]dM_i(u)\int_0^{\min(u,t)} dH_x \\
&= \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau dM_i(u)I(u<t) + \int_0^\tau [\Theta_i^0(u) - \Theta_i^1(u)]dM_i(u)H(\min(u,t)) \\
&= \frac{1}{n}\sum_{i=1}^{n}\int_0^\tau \big\{I(u<t) + [\Theta_i^0(u) - \Theta_i^1(u)]H(\min(u,t))\big\}dM_i(u)
\end{aligned}$$

Since this expression does not depend on t for $u < t$, $U_n(H_t)$ can be expressed as a martingale transform.

A.3 Asymptotic Difference of $\widehat{dH_t}^B$ and $\widehat{dH_t}^W$

The asymptotic difference between the two estimators of the jump dH_t is

$$
\begin{aligned}
&\sqrt{n}(\widehat{dH_t}^B - dH_t) - \sqrt{n}(\widehat{dH_t}^W - dH_t) = \sqrt{n}(\widehat{dH_t}^B - \widehat{dH_t}^W) \\
&= \sqrt{n}\Bigg(\frac{\frac{1}{n}\sum_{i=1}^n dN_i(t)}{\frac{1}{n}\sum_{i=1}^n Y_i(t)\Theta_i^0} - \frac{\frac{1}{n}\sum_{i=1}^n dN_i(t)}{\frac{1}{n}\sum_{i=1}^n Y_i(t)\Theta_i^0 - \int_{t^+}^{\tau}[\Theta_i^0 - \Theta_i^1]dM_i(u)} \Bigg) \\
&= \sqrt{n}\Bigg(\frac{\frac{1}{n}\sum_{i=1}^n dN_i(t)\big\{\frac{1}{n}\sum_{i=1}^n Y_i(t)\Theta_i^0 - \int_{t^+}^{\tau}[\Theta_i^0 - \Theta_i^1]dM_i(u)\big\} - \frac{1}{n}\sum_{i=1}^n dN_i(t)\big\{\frac{1}{n}\sum_{i=1}^n Y_i(t)\Theta_i^0\big\}}{\frac{1}{n}\sum_{i=1}^n Y_i(t)\Theta_i^0\big\{\frac{1}{n}\sum_{i=1}^n Y_i(t)\Theta_i^0 - \int_{t^+}^{\tau}[\Theta_i^0 - \Theta_i^1]dM_i(u)\big\}} \Bigg) \\
&= \sqrt{n}\Bigg(- \frac{\big\{\frac{1}{n}\sum_{i=1}^n dN_i(t)\big\}\big\{\frac{1}{n}\sum_{i=1}^n \int_{t^+}^{\tau}[\Theta_i^0 - \Theta_i^1]dM_i(u)\big\}}{\big\{\frac{1}{n}\sum_{i=1}^n Y_i(t)\Theta_i^0\big\}\big\{\frac{1}{n}\sum_{i=1}^n Y_i(t)\Theta_i^0 - \int_{t^+}^{\tau}[\Theta_i^0 - \Theta_i^1]dM_i(u)\big\}} \Bigg) \\
&= \sqrt{n}\Bigg(\frac{\frac{1}{n}\sum_{i=1}^n dN_i(t)}{\frac{1}{n}\sum_{i=1}^n Y_i(t)\Theta_i^0} \cdot \frac{-\frac{1}{n}\sum_{i=1}^n \int_{t^+}^{\tau}[\Theta_i^0 - \Theta_i^1]dM_i(u)}{\frac{1}{n}\sum_{i=1}^n Y_i(t)\Theta_i^0 - \int_{t^+}^{\tau}[\Theta_i^0 - \Theta_i^1]dM_i(u)} \Bigg) \\
&= \sqrt{n}\Bigg(\frac{\frac{1}{n}\sum_{i=1}^n dN_i(t) - Y_i(t)\Theta_i^0 dH_t + Y_i(t)\Theta_i^0 dH_t}{\frac{1}{n}\sum_{i=1}^n Y_i(t)\Theta_i^0} \cdot \frac{-\frac{1}{n}\sum_{i=1}^n \int_{t^+}^{\tau}[\Theta_i^0 - \Theta_i^1]dM_i(u)}{\frac{1}{n}\sum_{i=1}^n Y_i(t)\Theta_i^0 - \int_{t^+}^{\tau}[\Theta_i^0 - \Theta_i^1]dM_i(u)} \Bigg) \\
&= \sqrt{n}\Bigg(\Bigg[\frac{\frac{1}{n}\sum_{i=1}^n dM_i(t)}{\frac{1}{n}\sum_{i=1}^n Y_i(t)\Theta_i^0} + \frac{\frac{1}{n}\sum_{i=1}^n Y_i(t)\Theta_i^0 dH_t}{\frac{1}{n}\sum_{i=1}^n Y_i(t)\Theta_i^0} \Bigg] \cdot \frac{-\frac{1}{n}\sum_{i=1}^n \int_{t^+}^{\tau}[\Theta_i^0 - \Theta_i^1]dM_i(u)}{\frac{1}{n}\sum_{i=1}^n Y_i(t)\Theta_i^0 - \int_{t^+}^{\tau}[\Theta_i^0 - \Theta_i^1]dM_i(u)} \Bigg) \\
&= \sqrt{n}\Bigg(\Bigg[\frac{\frac{1}{n}\sum_{i=1}^n dM_i(t)}{\frac{1}{n}\sum_{i=1}^n Y_i(t)\Theta_i^0} + dH_t \Bigg] \cdot \frac{-\frac{1}{n}\sum_{i=1}^n \int_{t^+}^{\tau}[\Theta_i^0 - \Theta_i^1]dM_i(u)}{\frac{1}{n}\sum_{i=1}^n Y_i(t)\Theta_i^0 - \int_{t^+}^{\tau}[\Theta_i^0 - \Theta_i^1]dM_i(u)} \Bigg) \\
&= \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n dM_i(t)}{\frac{1}{n}\sum_{i=1}^n Y_i(t)\Theta_i^0} \cdot \frac{-\frac{1}{n}\sum_{i=1}^n \int_{t^+}^{\tau}[\Theta_i^0 - \Theta_i^1]dM_i(u)}{\frac{1}{n}\sum_{i=1}^n Y_i(t)\Theta_i^0 - \int_{t^+}^{\tau}[\Theta_i^0 - \Theta_i^1]dM_i(u)} + \frac{-\frac{dH_t}{\sqrt{n}}\sum_{i=1}^n \int_{t^+}^{\tau}[\Theta_i^0 - \Theta_i^1]dM_i(u)}{\frac{1}{n}\sum_{i=1}^n Y_i(t)\Theta_i^0 - \int_{t^+}^{\tau}[\Theta_i^0 - \Theta_i^1]dM_i(u)} \\
&= A + B
\end{aligned}
\tag{A.4}
$$

First consider A. By Lenglart's inequality, with $u' = u - t$ as a transformation of the time scale,

$$
\begin{aligned}
P\Bigg(\sup_{u' \le \tau'} \Bigg\{ -\frac{1}{n}\sum_{i=1}^n \int_0^{t'} [\Theta_i^0 - \Theta_i^1]dM_i(u') \Bigg\} \ge \epsilon \Bigg) &\le \frac{\eta}{\epsilon} + P\Bigg(\int_0^{\tau'} \frac{1}{n^2}\sum_{i=1}^n [\Theta_i^0 - \Theta_i^1]^2 Y_i(u')\Theta_i^0 dH_{u'} \ge \eta \Bigg) \\
&\le \frac{\eta}{\epsilon} + P\Bigg(\int_0^{\tau'} \frac{1}{n^2} \cdot nc \cdot dH_{u'} \ge \eta \Bigg) \text{ assuming } \boldsymbol{Z} \text{ is finite} \\
&= \frac{\eta}{\epsilon} + P\Bigg(\frac{c}{n}\int_0^{\tau'} dH_{u'} \ge \eta \Bigg) \\
&= \frac{\eta}{\epsilon} + P\Bigg(\frac{c}{n} H_{\tau'} \ge \eta \Bigg)
\end{aligned}
\tag{A.5}
$$

For any $\epsilon > 0$, we can find η to make (A.5) arbitrarily close to 0 as $n \to \infty$. Thus,

$$\frac{1}{n}\sum_{i=1}^{n}\int_{0}^{\tau'}[\Theta_i^0 - \Theta_i^1]dM_i(u') = \frac{1}{n}\sum_{i=1}^{n}\int_{t^+}^{\tau}[\Theta_i^0 - \Theta_i^1]dM_i(u) \xrightarrow{p} 0 \tag{A.6}$$

By the Law of Large Numbers,

$$\frac{1}{n}\sum_{i=1}^{n} Y_i(t)\Theta_i^0 \xrightarrow{p} E_{\boldsymbol{Z}_t}\{S^*_{T|\boldsymbol{Z}_t}(t)G^*_{C|\boldsymbol{Z}_t}(t)\Theta^0\} > 0 \tag{A.7}$$

Applying the Continuous Mapping Theorem, the Law of Large Numbers, Lenglart's inequality, and Slutsky's Theorem with $\xrightarrow{p}$ gives

$$\frac{\frac{1}{n}\sum_{i=1}^{n}\int_{t^+}^{\tau}[\Theta_i^0 - \Theta_i^1]dM_i(u)}{\frac{1}{n}\sum_{i=1}^{n} Y_i(t)\Theta_i^0 - \frac{1}{n}\sum_{i=1}^{n}\int_{t^+}^{\tau}[\Theta_i^0 - \Theta_i^1]dM_i(u)} \xrightarrow{p} 0 \cdot \{S^*_{T|\boldsymbol{Z}_t}(t)G^*_{C|\boldsymbol{Z}_t}(t)\Theta^0 - 0\}^{-1} = 0 \tag{A.8}$$

Now,

$$\begin{aligned}
E\left(\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} dM_i(t)}{\frac{1}{n}\sum_{i=1}^{n} Y_i(t)\Theta_i^0}\right) &= E\left(E\left(\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} dM_i(t)}{\frac{1}{n}\sum_{i=1}^{n} Y_i(t)\Theta_i^0}\middle|\mathcal{F}_{t^-}\right)\right) = E\left(\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} E(dM_i(t)|\mathcal{F}_{t^-})}{\frac{1}{n}\sum_{i=1}^{n} Y_i(t)\Theta_i^0}\right) = 0 \\
Var\left(\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} dM_i(t)}{\frac{1}{n}\sum_{i=1}^{n} Y_i(t)\Theta_i^0}\right) &= E\left(Var\left(\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} dM_i(t)}{\frac{1}{n}\sum_{i=1}^{n} Y_i(t)\Theta_i^0}\middle|\mathcal{F}_{t^-}\right)\right) = E\left(\frac{\frac{1}{n}\sum_{i=1}^{n} Var(dM_i(t)|\mathcal{F}_{t^-})}{\{\frac{1}{n}\sum_{i=1}^{n} Y_i(t)\Theta_i^0\}^2}\right) \\
&= E\left(\frac{\frac{1}{n}\sum_{i=1}^{n} Y_i(t)\Theta_i^0 dH_t}{\{\frac{1}{n}\sum_{i=1}^{n} Y_i(t)\Theta_i^0\}^2}\right) = E\left(dH_t\frac{\frac{1}{n}\sum_{i=1}^{n} Y_i(t)\Theta_i^0}{\{\frac{1}{n}\sum_{i=1}^{n} Y_i(t)\Theta_i^0\}^2}\right) \\
&= dH_t E\left(\frac{1}{\frac{1}{n}\sum_{i=1}^{n} Y_i(t)\Theta_i^0}\right) \\
&= \frac{dH_t}{E_{\boldsymbol{Z}_t}(S^*_{T|\boldsymbol{Z}_t}(t)G^*_{C|\boldsymbol{Z}_t}(t)\Theta^0)} + o_p(1)
\end{aligned} \tag{A.9}$$

By the Central Limit Theorem, $\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} dM_i(t)}{\frac{1}{n}\sum_{i=1}^{n} Y_i(t)\Theta_i^0} \xrightarrow{d} N\big(0, \frac{dH_t}{E_{\boldsymbol{Z}_t}(S^*_{T|\boldsymbol{Z}_t}(t)G^*_{C|\boldsymbol{Z}_t}(t)\Theta^0)}\big)$. Therefore, by Slutsky's Theorem,

$$\begin{aligned}
A &= \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} dM_i(t)}{\frac{1}{n}\sum_{i=1}^{n} Y_i(t)\Theta_i^0} \cdot \frac{-\frac{1}{n}\sum_{i=1}^{n}\int_{t^+}^{\tau}[\Theta_i^0 - \Theta_i^1]dM_i(u)}{\frac{1}{n}\sum_{i=1}^{n} Y_i(t)\Theta_i^0 - \int_{t^+}^{\tau}[\Theta_i^0 - \Theta_i^1]dM_i(u)\}} \\
&\xrightarrow{d} N\left(0, \frac{dH_t}{E_{\boldsymbol{Z}_t}(S^*_{T|\boldsymbol{Z}_t}(t)G^*_{C|\boldsymbol{Z}_t}(t)\Theta^0)}\right) \times 0 \\
&\xrightarrow{d} 0
\end{aligned} \tag{A.10}$$

Now, let us consider B. We know from equations (A.6) and (A.7) that

$$\left(\frac{1}{n}\sum_{i=1}^{n} Y_i(t)\Theta_i^0 - \int_{t^+}^{\tau}[\Theta_i^0 - \Theta_i^1]dM_i(u)\right)^{-1} \xrightarrow{p} E_{\boldsymbol{Z}_t}\{S^*_{T|\boldsymbol{Z}_t}(t)G^*_{C|\boldsymbol{Z}_t}(t)\Theta^0\}^{-1} \tag{A.11}$$

Then, by the independence of subjects and by the independence among increments of a martingale,

$$
\begin{aligned}
E\Big(\frac{-dH_t}{\sqrt{n}}\sum_{i=1}^{n}\int_{t^+}^{\tau}[\Theta_i^0-\Theta_i^1]dM_i(u)\Big) &= \frac{-dH_t}{\sqrt{n}}\sum_{i=1}^{n}\int_{t^+}^{\tau}E\big([\Theta_i^0-\Theta_i^1]dM_i(u)\big)\\
&= \frac{-dH_t}{\sqrt{n}}\sum_{i=1}^{n}\int_{t^+}^{\tau}E\big(E([\Theta_i^0-\Theta_i^1]dM_i(u)|\mathcal{F}_{u^-})\big)\\
&= \frac{-dH_t}{\sqrt{n}}\sum_{i=1}^{n}\int_{t^+}^{\tau}E\big([\Theta_i^0-\Theta_i^1]E(dM_i(u)|\mathcal{F}_{u^-})\big)=0\\
Var\Big(\frac{-dH_t}{\sqrt{n}}\sum_{i=1}^{n}\int_{t^+}^{\tau}[\Theta_i^0-\Theta_i^1]dM_i(u)\Big) &= \frac{(dH_t)^2}{n}\sum_{i=1}^{n}\int_{t^+}^{\tau}Var\big([\Theta_i^0-\Theta_i^1]dM_i(u)\big)\\
&= \frac{(dH_t)^2}{n}\sum_{i=1}^{n}\int_{t^+}^{\tau}Var\big(E([\Theta_i^0-\Theta_i^1]dM_i(u)|\mathcal{F}_{u^-})\big)+E\big(Var([\Theta_i^0-\Theta_i^1]dM_i(u)|\mathcal{F}_{u^-})\big)\\
&= \frac{(dH_t)^2}{n}\sum_{i=1}^{n}\int_{t^+}^{\tau}E\big([\Theta_i^0-\Theta_i^1]^2Var(dM_i(u)|\mathcal{F}_{u^-})\big)\\
&= (dH_t)^2E\Big(\frac{1}{n}\sum_{i=1}^{n}\int_{t^+}^{\tau}E\big([\Theta_i^0-\Theta_i^1]^2Y_i(u)\Theta_i^0dH_u\big)\Big)\\
&= O((dH(t))^2)
\end{aligned}
\tag{A.12}
$$

Let GP represent a Gaussian Process. Then, by the Martingale Central Limit Theorem (for which all conditions are satisfied),

$$
\begin{aligned}
\frac{-dH_t}{\sqrt{n}}\sum_{i=1}^{n}\int_{t^+}^{\tau}[\Theta_i^0-\Theta_i^1]dM_i(u) &\xrightarrow{d} GP\big(0,O((dH(t))^2)\big)\\
\implies B=\frac{-\frac{dH_t}{\sqrt{n}}\sum_{i=1}^{n}\int_{t^+}^{\tau}[\Theta_i^0-\Theta_i^1]dM_i(u)}{\frac{1}{n}\sum_{i=1}^{n}Y_i(t)\Theta_i^0-\int_{t^+}^{\tau}[\Theta_i^0-\Theta_i^1]dM_i(u)} &\xrightarrow{d} GP\Big(0,\frac{O((dH(t))^2)}{\{E_{\boldsymbol{Z}_t}(S^*_{T|\boldsymbol{Z}_t}(t)G^*_{C|\boldsymbol{Z}_t}(t)\Theta^0)\}^2}\Big)\\
&= GP\big(0,O((dH(t))^2)\big)
\end{aligned}
\tag{A.13}
$$

Therefore, by Slutsky's Theorem,

$$
\sqrt{n}(\widehat{dH_t}^{B}-\widehat{dH_t}^{W})=A+B\xrightarrow{d}0+GP\big(0,O((dH(t))^2)\big)=GP\big(0,O((dH(t))^2)\big) \tag{A.14}
$$

This result implies that asymptotically, the difference in the Breslow and Weighted estimators of the jump at time t is a Gaussian Process with mean 0 and a variance that is negligible.

A.4 Variances of Estimators

The variance of $\hat{H}_t^B$ is

$$
\begin{aligned}
Var\big[\sqrt{n}(\hat{H}_t^B - H_t)\big] &= Var\bigg[\sqrt{n}\bigg(\int_0^t \frac{\frac{1}{n}\sum_{i=1}^n dN_i(x)}{\frac{1}{n}\sum_{i=1}^n Y_i(x)\Theta_i^0} - dH_x\bigg)\bigg] = Var\bigg(\int_0^t \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n dM_i(x)}{\frac{1}{n}\sum_{i=1}^n Y_i(x)\Theta_i^0}\bigg) \\
&= \int_0^t E\bigg(\frac{\frac{1}{n}\sum_{i=1}^n Var(dM_i(x)|\mathcal{F}_{x^-})}{\{\frac{1}{n}\sum_{i=1}^n Y_i(x)\Theta_i^0\}^2}\bigg) \\
&= \int_0^t E\bigg(\frac{\frac{1}{n}\sum_{i=1}^n Y_i(x)\Theta_i^0 dH_x}{\{\frac{1}{n}\sum_{i=1}^n Y_i(x)\Theta_i^0\}^2}\bigg) \\
&= \int_0^t E\bigg(\frac{dH_x}{\frac{1}{n}\sum_{i=1}^n Y_i(x)\Theta_i^0}\bigg) \\
&= \int_0^t \frac{dH_x}{E_{\boldsymbol{Z}_x}(S^*_{T|\boldsymbol{Z}_x}(x)G^*_{C|\boldsymbol{Z}_x}(x)\Theta^0)} + o_p(1) \\
&= \int_0^t \frac{dH_x}{A_x},\ A_x = E_{\boldsymbol{Z}_x}(S^*_{T|\boldsymbol{Z}_x}(x)G^*_{C|\boldsymbol{Z}_x}(x)\Theta^0) \\
&= Var_B(t)
\end{aligned}
\tag{A.15}
$$

The variance of $\hat{H}_t^W$ can be expressed as

$$
\begin{aligned}
\sqrt{n}(\hat{H}_t^W - H_t) &= \int_0^t \frac{dV_x}{\frac{1}{n}\sum_{i=1}^n Y_i(x)\Theta_{ix}^0} + o_p(1) \\
Var\big[\sqrt{n}(\hat{H}_t^W - H_t)\big] &= \int_0^t\int_0^t E(dV_x dV_y)\frac{1}{\{\frac{1}{n}\sum_{i=1}^n Y_i(x)\Theta_{ix}^0\}}\frac{1}{\{\frac{1}{n}\sum_{i=1}^n Y_i(y)\Theta_{iy}^0\}} + o_p(1) \\
&= \int_0^t\int_0^t Cov(dV_x, dV_y)E_{\boldsymbol{Z}_x}\bigg(\frac{1}{\frac{1}{n}\sum_{i=1}^n Y_i(x)\Theta_{ix}^0}\bigg)E_{\boldsymbol{Z}_y}\bigg(\frac{1}{\frac{1}{n}\sum_{i=1}^n Y_i(y)\Theta_{iy}^0}\bigg) + o_p(1) \\
&= \int_0^t\int_0^t \frac{Cov(dV_x, dV_y)}{E_{\boldsymbol{Z}_x}\{S^*_{T|\boldsymbol{Z}_x}(x)G^*_{C|\boldsymbol{Z}_x}(x)\Theta_x^0\}E_{\boldsymbol{Z}_y}\{S^*_{T|\boldsymbol{Z}_y}(y)G^*_{C|\boldsymbol{Z}_y}(y)\Theta_y^0\}} + o_p(1) \\
&= \int_0^t\int_0^t \frac{Cov(dV_x, dV_y)}{A_x A_y} + o_p(1)
\end{aligned}
\tag{A.16}
$$

The covariance can be expressed as

$$
\begin{aligned}
&Cov(dV_x, dV_y) = E(dV_x dV_y) \\
&= \frac{1}{n}\sum_{i=1}^{n} E\big\{(dM_i(x) - X_{ix}dH_x)(dM_i(y) - X_{iy}dH_y)\big\} \\
&= \frac{1}{n}\sum_{i=1}^{n} E(dM_i(x)dM_i(y)) - E(dM_i(x)X_{iy}dH_y) - E(dM_i(y)X_{ix}dH_x) + E(X_{ix}X_{iy}dH_x dH_y)
\end{aligned}
\tag{A.17}
$$

Each term in (A.17) can be simplified as follows:

$$
\begin{aligned}
E(dM_i(x)dM_i(y)) &= E_{\boldsymbol{Z}_x}(S^*_{T|\boldsymbol{Z}_x}(x)G^*_{C|\boldsymbol{Z}_x}(x)\Theta^0_x)dH_x I(x=y) \\
&= A_x dH_x I(x=y) \\
E(dM_i(x)X_{iy}dH_y) &= E_{\boldsymbol{Z}_x}([\Theta^0_x - \Theta^1_x]^2 S^*_{T|\boldsymbol{Z}_x}(x)G^*_{C|\boldsymbol{Z}_x}(x)\Theta^0_x)dH_x dH_y I(x>y) \\
&= B_x dH_x dH_y I(x>y), \\
B_x &= E_{\boldsymbol{Z}_x}([\Theta^0_x - \Theta^1_x]^2 S^*_{T|\boldsymbol{Z}_x}(x)G^*_{C|\boldsymbol{Z}_x}(x)\Theta^0_x) \\
E(dM_i(y)X_{ix}dH_x) &= B_y dH_x dH_y I(x<y) \\
E(X_{ix}X_{iy})dH_x dH_y &= dH_x dH_y \int\limits_{\max(x,y)}^{t} B_a dH_a
\end{aligned}
$$

Then,

$$
\begin{aligned}
Cov(dV_x, dV_y) =& A_x dH_x I(x=y) - B_x dH_x dH_y I(x>y) - B_y dH_x dH_y I(x<y) \\
&+ dH_x dH_y \int\limits_{\max(x,y)}^{t} B_a dH_a
\end{aligned}
\tag{A.18}
$$

Based on (A.16), the variance of $\hat{H}_t^W$ is

$$
\begin{aligned}
&Var\big[\sqrt{n}(\hat{H}_t^W - H_t)\big] \\
&= \int_0^t \int_0^t \frac{A_x dH_x I(x=y)}{A_x A_y} - \int_0^t \int_0^t \frac{B_x dH_x dH_y I(x>y)}{A_x A_y} - \int_0^t \int_0^t \frac{B_y dH_x dH_y I(x<y)}{A_x A_y} \\
&+ \int_0^t \int_0^t \frac{dH_x dH_y \int_{\max(x,y)}^t B_a dH_a}{A_x A_y} + o_p(1) \\
&= \int_0^t \frac{A_x dH_x}{A_x^2} - \int_0^t \frac{B_x dH_x}{A_x} \int_0^x \frac{dH_y}{A_y} - \int_0^t \frac{B_y dH_y}{A_y} \int_0^y \frac{dH_x}{A_x} + 2\int_0^t \int_0^y \frac{\int_y^t B_a dH_a}{A_x A_y} dH_x dH_y + o_p(1) \\
&= \int_0^t \frac{dH_x}{A_x} - \int_0^t \frac{B_x dH_x}{A_x} Var_B(x) - \int_0^t \frac{B_y dH_y}{A_y} Var_B(y) + 2\int_0^t \frac{\int_y^t B_a dH_a}{A_y} \int_0^y \frac{dH_x}{A_x} dH_y + o_p(1) \\
&= Var_B(t) - \int_0^t B_x \frac{dVar_B^2(x)}{2} - \int_0^t B_y \frac{dVar_B^2(y)}{2} + 2\int_0^t \int_y^t B_a dH_a \frac{dH_y}{A_y} Var_B(y) + o_p(1) \\
&= Var_B(t) - \int_0^t B_x dVar_B^2(x) + \int_0^t \int_y^t B_a dH_a dVar_B^2(y) + o_p(1) \\
&= Var_B(t) - \int_0^t \Big[B_x - \int_x^t B_a dH_a\Big] dVar_B^2(x) + o_p(1)
\end{aligned}
\tag{A.19}
$$

A.5 Asymptotic Distribution and Profile Likelihood Information

A.5.1 Asymptotic Properties of NPMLE

We now prove the asymptotic consistency and convergence of the Weighted Breslow estimator/NPMLE. Much of what follows mirrors the arguments in Hu and Tsodikov (2014) and Rice and Tsodikov (2017). Let

$\|\cdot\|_\infty$ denote the supremum norm in $[0, \tau]$

$\|w\|_{TV}$ denote the total variation of $w(t)$ in $[0, \tau]$

$Q = \{w(t) : \|w\|_{TV} \leq 1\}$

$\hat{H}_t$ be regarded as a bounded linear functional in $L^\infty[Q]$

$\{\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*, \hat{H} - H^*\}$ be a random element in the metric space $\mathbb{R}^p \times L^\infty[Q]$

$\mathcal{H}$ denote a compact convex set in the metric space $\mathbb{R}^p \times L^\infty[Q]$ in which $\Omega^* = (\boldsymbol{\beta}^*, H^*)$ is contained

We assume the following **regularity conditions** (*Fleming and Harrington*, 1991):

(1) The true cumulative baseline hazard H^* is strictly increasing and differentiable, and Ω^* is in the interior of the compact convex set $\mathcal{H}$.

(2) With probability 1, the covariate process $\boldsymbol{Z}(t)$ is left continuous with bounded total variation on $[0, \tau]$. $\boldsymbol{Z}(t)$ is linearly independent, i.e., if there exist $a(t)$ and c such that $a(t) + c^T \boldsymbol{Z}(t) = 0$ with probability 1, then $a(t) = 0$ and $c = 0$.

(3) With probability 1, $E(Y(\tau)|\boldsymbol{Z}) > 0$, i.e., the risk set will not shrink to zero at time τ

(4) Hessian matrix I_n evaluated at $\boldsymbol{\beta}^*, H^*$ is positive definite and converges in probability to I^0, a deterministic and invertible operator.

(5) **Identifiability condition:** The model is identifiable such that $H^* = H$ uniformly over $\Omega \Longrightarrow \Omega = \Omega^*$. This ensures that for any sequence $\Omega_n \in \mathcal{H}$,

$$\liminf_{n\to\infty} l(\Omega_n) \geq l(\Omega^*) \Longrightarrow \|\Omega_n - \Omega^*\| \xrightarrow{p} 0$$

(6) **Uniform convergence condition:** For any sequence $\Omega \in \mathcal{H}$, we have uniform convergence, i.e.,

$$\sup_{\Omega\in\mathcal{H}} |l_n(\Omega) - l(\Omega)| \xrightarrow{p} 0$$

Theorem 1. *Assuming regularity conditions (1)-(6) hold, then with probability 1,* $\hat{\boldsymbol{\beta}}$ *converges to* $\boldsymbol{\beta}^*$ *and* $\hat{H}$ *converges to* H^* *uniformly on the interval* $[0, \tau]$.

Proof. By regularity conditions (5) and (6) and Theorem 2.12 of Kosorok (2008), we have $l_n(\hat{\Omega}) = \sup_{\Omega \in \mathcal{H}} l_n(\Omega) + o_p(1)$, which implies that $\left\| \hat{\Omega} - \Omega^* \right\| \xrightarrow{p} 0$. To demonstrate consistency, we will show 1) convexity and unique maximum of the likelihood; 2) identifiability of the parameters; 3) uniform convergence of the parameters.

1) *Convexity and unique maximum of the likelihood:* The model may be characterized as $d\Lambda_t = dH_t \Theta^0(\Omega)$, which is a functional depending on the processes $H(\cdot)$ and $\boldsymbol{Z}(\cdot)$ on $[0, t]$. Let F_t be the cumulative incidence function and R_t be the survival function in the presence of censoring. The true likelihood can be written as

$$l(\Omega, \Omega^*) = E_{\boldsymbol{Z}} \left\{ \int_0^\tau \ln(d\Lambda_t) dF_t^* - R_t^* d\Lambda_t \right\} \tag{A.20}$$

where the expectation is taken with respect to the covariate process. Consider the true negative Kullback-Leibler distance.

$$\begin{aligned} D &= l(\Omega, \Omega^*) - l(\Omega^*, \Omega^*) \\ &= E_{\boldsymbol{Z}} \left\{ \int_0^\tau \ln(d\Lambda_t) dF_t^* - R_t^* d\Lambda_t \right\} - E_{\boldsymbol{Z}} \left\{ \int_0^\tau \ln(d\Lambda_t^*) dF_t^* - R_t^* d\Lambda_t^* \right\} \\ &= E_{\boldsymbol{Z}} \left\{ \int_0^\tau \ln \left(\frac{d\Lambda_t}{d\Lambda_t^*} \right) dF_t^* - (d\Lambda_t - d\Lambda_t^*) R_t^* \right\} \\ &= E_{\boldsymbol{Z}} \left\{ \int_0^\tau \left[\ln \left(\frac{d\Lambda_t}{d\Lambda_t^*} \right) - \left(\frac{d\Lambda_t}{d\Lambda_t^*} - 1 \right) \right] dF_t^* \right\} \\ &= E_{\boldsymbol{Z}} \left\{ \int_0^\tau \nu \left(\frac{d\Lambda_t}{d\Lambda_t^*} \right) dF_t^* \right\} \end{aligned} \tag{A.21}$$

where $\nu(x) = \ln(x) - x + 1$ is a convex, non-positive function with $\nu'(x) = \frac{1}{x} - 1 = 0 \implies x = 1$ and $\nu''(x) = -\frac{1}{x^2} < 0$ for all $x \implies x = 1$ is the unique maximizer. Therefore, D has a unique maximum when $d\Lambda_t = d\Lambda_t^*$ uniformly. Under an identifiable model, this implies that the unique maximum of D occurs

at Ω^*.

2) *Identifiability of parameters:* Λ^* is assumed to be a continuous, differentiable functional of H, implying that the likelihood $l(\Omega)$ is also a continuous, differentiable functional of H. From (A.21), $\Omega^* = \text{ argmax}_{\Omega\in\mathcal{H}} l(\Omega)$ is unique, i.e., the model $l(\Omega, \Omega^*)$ is identifiable. Let us assume that the model is identifiable in the sense that $\Lambda = \Lambda^*$ uniformly over Ω implies that $\Omega = \Omega^*$ uniformly. Therefore, by Lemma 14.3 of Kosorok (2008), $\liminf_{n\to\infty} l(\Omega_n) \geq l(\Omega^*)$, meaning that the parameters are identifiable.

3) *Uniform convergence of parameters:* By regularity condition (1), Ω is in the class of functions of bounded variation with integrable envelope, implying that H_t is bounded. Therefore $\mathcal{H}$ is a Glivenko-Cantelli class whose ϵ-entropy with bracketing number is bounded by $\frac{A}{\epsilon}$, where A is some constant. By the assumption of continuity of functionals Λ and $l(\Omega)$ and the integrability of the envelope of Ω, the integrand in $l(\Omega)$ is also Glivenko-Cantelli. Therefore, we can apply the Uniform Law of Large Numbers to the empirical process counterparts of D and l

$$\begin{aligned} D_n &= l_n(\Omega, \Omega^*) - l_n(\Omega^*, \Omega^*) \\ l_n(\Omega, \Omega^*) &= \frac{1}{n}\sum_{i=1}^{n} \int_0^\tau \left[\ln(dH_t) + \ln(\Theta_i^0(\boldsymbol{\beta}, H_t))\right] dN_i(t) - Y_i(t)\Theta_i^0(\boldsymbol{\beta}, H_t) dH_t \end{aligned} \tag{A.22}$$

such that

$$\begin{aligned} \sup_{\Omega\in\mathcal{H}} |D_n(\Omega) - D(\Omega)| &\xrightarrow{p} 0 \\ \sup_{\Omega\in\mathcal{H}} |l_n(\Omega) - l(\Omega)| &\xrightarrow{p} 0 \end{aligned} \tag{A.23}$$

■

Consider the linear functional

$$\sqrt{n}\left[a^T(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) + \int_0^\tau b(t)d\big(\hat{H}_t - H_t^*\big)\right] \tag{A.24}$$

where a^T is a real-valued vector and $b(t)$ is a function with bounded total variation. Let $\mathbf{B}$ denote the vector of values of $b(t)$ evaluated at the observed failure times. Let $\mathcal{C}^T = (a^T, \mathbf{B}^T)$.

Theorem 2: *Assuming all regularity conditions hold,* $\sqrt{n}\left\{\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*\right)^T, \hat{H}_t - H_t^*\right\}^T$ *converges weakly to a zero-mean Gaussian process. Additionally,* $n\mathcal{C}^T\mathcal{I}_n^{-1}\mathcal{C}$ *converges in probability to the asymptotic covariance function of the linear functional (*A.*24), where* $\mathcal{I}_n$ *is the negative Hessian matrix of the log-likelihood.*

Proof. Consider the score $U(\Omega) = (U_{\boldsymbol{\beta}}, U_H)^T$, which is a multivariate martingale transform under Ω^*. Then, by the Martingale Central Limit Theorem, $\frac{1}{\sqrt{n}}U(\Omega^*)$ converges weakly to $\mathcal{U}(t) = (\mathcal{U}_{\boldsymbol{\beta}}, \mathcal{U}_{H_t})^T$, where $\mathcal{U}_{\boldsymbol{\beta}}$ is a mean-zero multivariate Normal random variable and $\mathcal{U}_{H_t}$ is a mean-zero Gaussian process. The variance-covariance function of $(\mathcal{U}_{\boldsymbol{\beta}}, \mathcal{U}_{H_t})^T$ is characterized by the limiting values of the scaled predictable variation processes $V(\boldsymbol{\beta}), V(H), V(\boldsymbol{\beta}, H)$. Note that the scaled score equations can be written as the following martingale transforms

$$\begin{aligned}
\sqrt{n}U_n(\boldsymbol{\beta}) =& \frac{1}{\sqrt{n}}\sum_{i=1}^n \int_0^\tau \frac{\partial}{\partial \boldsymbol{\beta}} \ln(\Theta_i^0) dM_i(u) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \int_0^\tau \frac{\frac{\partial}{\partial \boldsymbol{\beta}}\Theta_i^0}{\Theta_i^0} dM_i(u) \\
\sqrt{n}U_n(H) =& \frac{1}{\sqrt{n}}\sum_{i=1}^n \int_0^\tau \epsilon_i(u,t) dM_i(u), \\
\epsilon_i(u,t) =& I(u<t) + [\Theta_i^0 - \Theta_i^1]H(\min(u,t))
\end{aligned} \tag{A.25}$$

Then, using the fact that $Var(dM_i(u)|\mathcal{F}_{u^-}) = Y_i(u)\Theta_i^0 dH_u$,

$$
\begin{aligned}
V(\boldsymbol{\beta}) &= \sum_{i=1}^{n} \int_0^\tau \frac{1}{n} \frac{[\frac{\partial}{\partial \boldsymbol{\beta}}\Theta_i^0]^2}{[\Theta_i^0]^2} Y_i(u)\Theta_i^0 dH_u = \sum_{i=1}^{n} \int_0^\tau \frac{1}{n} \frac{[\frac{\partial}{\partial \boldsymbol{\beta}}\Theta_i^0]^2}{\Theta_i^0} Y_i(u) dH_u \\
&\xrightarrow{p} \int_0^\tau \frac{[\frac{\partial}{\partial \boldsymbol{\beta}}\Theta^0]^2}{\Theta^0} P(X \geq u) dH_u = \int_0^\tau \frac{[\frac{\partial}{\partial \boldsymbol{\beta}}\Theta^0]^2}{\Theta^0} S_T^*(u) G_C^*(u) dH_u = \sigma_{\boldsymbol{\beta}}^2 \\
V(H) &= \sum_{i=1}^{n} \int_0^\tau \frac{1}{n} \epsilon_i^2(u,t) Y_i(u)\Theta_i^0 dH_u \\
&\xrightarrow{p} \int_0^\tau \epsilon(u,t)\epsilon(u,s) P(X \geq u)\Theta^0 dH_u = \int_0^\tau \epsilon(u,t)\epsilon(u,s) S_T^*(u) G_C^*(u)\Theta^0 dH_u = \sigma_H^2 \\
V(\boldsymbol{\beta}, H) &= \sum_{i=1}^{n} \int_0^\tau \frac{1}{n} \frac{\frac{\partial}{\partial \boldsymbol{\beta}}\Theta_i^0}{\Theta_i^0} \epsilon_i(u,t) Y_i(u)\Theta_i^0 dH_u = \frac{1}{n}\sum_{i=1}^{n} \int_0^\tau \frac{\partial}{\partial \boldsymbol{\beta}} \Theta_i^0 \epsilon_i(u,t) Y_i(u)\Theta_i^0 dH_u \\
&\xrightarrow{p} \int_0^\tau \epsilon(u,t) \frac{\partial}{\partial \boldsymbol{\beta}} \Theta^0 P(X \geq u) dH_u = \int_0^\tau \epsilon(u,t) \frac{\partial}{\partial \boldsymbol{\beta}} \Theta^0 S_T^*(u) G_C^*(u) dH_u = \sigma_{\boldsymbol{\beta},H}^2
\end{aligned}
\tag{A.26}
$$

Define a linear transformation operator as

$$
\mathcal{I}^*(t,s) = \frac{\partial U^*}{\partial \Omega} = -\begin{pmatrix} \frac{\partial^2 l_\infty}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} & \frac{\partial^2 l_\infty}{\partial \boldsymbol{\beta} \partial dH_s} \\ \frac{\partial^2 l_\infty}{\partial \boldsymbol{\beta}^T \partial dH_t} & \frac{\partial^2 l_\infty}{\partial dH_t \partial dH_s} \end{pmatrix}_{\Omega=\Omega^*} \tag{A.27}
$$

where l_∞ is the limit in probability of $\frac{1}{n} l(\boldsymbol{\beta}, H)$ and $U^* = (\frac{\partial l_\infty}{\partial \boldsymbol{\beta}}, \frac{\partial l_\infty}{\partial dH_t})^T$. The operator $\mathcal{I}^*$ will act on an arbitrary vector element $\Omega_s = (\boldsymbol{\beta}^T, dH_s)^T$ as

$$
\mathcal{I}^*(t,s)\Omega_s = -\begin{pmatrix} \frac{\partial^2 l_\infty}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T}\boldsymbol{\beta} + \int_0^\tau \frac{\partial^2 l_\infty}{\partial \boldsymbol{\beta} \partial dH_s} dH_s \\ \frac{\partial^2 l_\infty}{\partial \boldsymbol{\beta}^T \partial dH_t}\boldsymbol{\beta} + \int_0^\tau \frac{\partial^2 l_\infty}{\partial dH_t \partial dH_s} dH_s \end{pmatrix} \tag{A.28}
$$

Expanding the score $U(\hat{\Omega})$ about the true Ω^* gives

$$
\mathcal{I}^*(t,s)\sqrt{n}(\hat{\Omega}_s - \Omega_s^*) = \mathcal{U}(t) + o_p(1) \tag{A.29}
$$

Assume that the Fredholm operator expressed by the kernel $\mathcal{I}^*$ of the Fredholm

integral equations of the first kind in (A.29) is square integrable and that $\mathcal{I}^*\Omega = 0$ only has the trivial solution $\Omega = 0$. Then equation (A.29) has a unique solution, and there exists an inverse information operator $\mathcal{I}^{*-1}(s,t)$ such that, by Theorem 3.3.1 of van der Vaart and Wellner (1996)

$$\sqrt{n}(\hat{\Omega}_s - \Omega_s^*) = \mathcal{I}^{*-1}(s,t)\mathcal{U}(t) + o_p(1) \tag{A.30}$$

Differentiating the equation $E\{U(\Omega^*)\} = 0$ with respect to Ω at the true Ω^* gives the usual equivalence between $\mathcal{I}^*$ represented by second derivatives and

$$\mathcal{I}^*(s,t) = \begin{pmatrix} \frac{\partial l_\infty}{\partial \boldsymbol{\beta}} \frac{\partial l_\infty}{\partial \boldsymbol{\beta}^T} & \frac{\partial l_\infty}{\partial \boldsymbol{\beta}} \frac{\partial l_\infty}{\partial dH_s} \\ \frac{\partial l_\infty}{\partial \boldsymbol{\beta}^T} \frac{\partial l_\infty}{\partial dH_t} & \frac{\partial l_\infty}{\partial dH_t} \frac{\partial l_\infty}{\partial dH_s} \end{pmatrix}_{\Omega=\Omega^*} \tag{A.31}$$

which represents the variance of the normalized score Gaussian process $\mathcal{U}(t)$. By the functional delta method from Section 2.2.4 of Kosorok (2008), for a differentiable functional $F(\Omega)$, $\sqrt{n}\{F(\hat{\Omega}) - F(\Omega)\}$ converges weakly to a mean zero Gaussian process with variance-covariance function given by $\dot{F}(\Omega)^T\mathcal{I}^{*-1}\dot{F}(\Omega)$, where $\dot{F}(\Omega) = \frac{\partial F}{\partial \Omega}$. Operator products are defined as in equation (A.28). Applying this to the linear functional in (A.24) and replacing operator products by matrix multiplication and $\mathcal{I}^*$ by its consistent matrix estimator $n^{-1}\hat{\mathcal{I}}_n$ gives the desired result. ■

The elements of the observed information matrix for $\hat{\Omega} = (\hat{\boldsymbol{\beta}}, \{\widehat{dH}\})$ are explicitly given in (A.32). Note that asymptotically, the integrals over the future, $\int_{t^+}^{\tau}$, will converge to zero in probability due to the presence of the martingale increment. Thus, the elements of the expected information matrices for the NPMLE and the

Breslow estimator are expected to be close for large samples.

$$
\begin{aligned}
J_{\boldsymbol{\beta}\boldsymbol{\beta}} &= -\frac{\partial}{\partial \boldsymbol{\beta}} U_n(\boldsymbol{\beta}) = -\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} \frac{\partial}{\partial \boldsymbol{\beta}}\frac{\partial}{\partial \boldsymbol{\beta}^T}\ln(\Theta_i^0) dM_i(u) \\
J_{dH_t\boldsymbol{\beta}} &= -\frac{\partial}{\partial dH_t} U_n(\boldsymbol{\beta}) = -\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} \frac{\partial}{\partial dH_t}\frac{\partial}{\partial \boldsymbol{\beta}}\ln(\Theta_i^0) dM_i(u) \\
J_{\boldsymbol{\beta} dH_t} &= -\frac{\partial}{\partial \boldsymbol{\beta}} U_n(dH_t) = -\frac{1}{n}\sum_{i=1}^{n} -Y_i(t)\frac{\partial}{\partial \boldsymbol{\beta}}\Theta_i^0 + \int_{t^+}^{\tau}\frac{\partial}{\partial \boldsymbol{\beta}}[\Theta_i^0 - \Theta_i^1] dM_i(u) \\
J_{dH_s dH_t} &= -\frac{\partial}{\partial dH_s} U_n(H) = -\frac{1}{n}\sum_{i=1}^{n} -Y_i(t)\frac{\partial}{\partial dH(s)}\Theta_i^0 - \frac{dN_i(t)I(t=s)}{dH_t dH_s} + \int_{t^+}^{\tau}\frac{\partial}{\partial dH(s)}[\Theta_i^0 - \Theta_i^1] dM_i(u) \\
J_{dH_t dH_t} &= -\frac{\partial}{\partial dH_t} U_n(H) = -\frac{1}{n}\sum_{i=1}^{n} -Y_i(t)\frac{\partial}{\partial dH(t)}\Theta_i^0 - \frac{dN_i(t)}{dH_t^2} + \int_{t^+}^{\tau}\frac{\partial}{\partial dH(t)}[\Theta_i^0 - \Theta_i^1] dM_i(u)
\end{aligned}
\tag{A.32}
$$

A.5.2 Profile Likelihood Hessian and Information Matrix

We now show, using the arguments of Rice and Tsodikov (2017), that the variance of $\hat{\boldsymbol{\beta}}$ derived via profile likelihood is a consistent estimator of the covariance matrix of $\hat{\boldsymbol{\beta}}$, which will be the $\boldsymbol{\beta}\boldsymbol{\beta}$ submatrix of $\mathcal{I}^{*-1}$, the asymptotic covariance "matrix" operator of the score, where

$$
\mathcal{I}^* = \begin{pmatrix} \mathcal{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} & \mathcal{I}_{\boldsymbol{\beta}H} \\ \mathcal{I}_{H\boldsymbol{\beta}} & \mathcal{I}_{HH} \end{pmatrix} = \begin{pmatrix} -\frac{\partial^2 l_\infty}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^T} & -\frac{\partial^2 l_\infty}{\partial \boldsymbol{\beta}\partial dH_s} \\ -\frac{\partial^2 l_\infty}{\partial dH_t \partial \boldsymbol{\beta}} & -\frac{\partial^2 l_\infty}{\partial dH_t \partial dH_s} \end{pmatrix} \tag{A.33}
$$

For fixed $\boldsymbol{\beta}$, let $\hat{H}_{\boldsymbol{\beta}}$ be the solution to $U_H = 0$. Then, the profile likelihood can be written as $\hat{l} = l(\hat{H}_{\boldsymbol{\beta}}(\cdot), \boldsymbol{\beta})$. Note that for any block matrix or "operator" I_∞, the inverse has the form

$$
\begin{aligned}
I_\infty^{-1} &= \begin{pmatrix} Q^{-1} & -Q^{-1} I_{\boldsymbol{\beta}H} I_{HH}^{-1} \\ -I_{HH}^{-1} I_{H\boldsymbol{\beta}} Q^{-1} & I_{HH}^{-1} + I_{HH}^{-1} I_{H\boldsymbol{\beta}} Q^{-1} I_{\boldsymbol{\beta}H} I_{HH}^{-1} \end{pmatrix} \\
Q &= I_{\boldsymbol{\beta}\boldsymbol{\beta}} - I_{\boldsymbol{\beta}H} I_{HH}^{-1} I_{H\boldsymbol{\beta}}
\end{aligned}
\tag{A.34}
$$

Thus, we want to show that asymptotically, $Var\big[\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)\big] = Q^{-1} = I_{pr}^{-1}$ up to $o_p(1)$ terms in the profile matrix. Let $J_{H\boldsymbol{\beta}} = \frac{\partial \widehat{dH}_{\boldsymbol{\beta}}(s)}{\partial \boldsymbol{\beta}}$ be the Jacobian. The full derivative of the profile likelihood $\hat{l}$ with respect to $\boldsymbol{\beta}$ (which includes a derivative over $\boldsymbol{\beta}$ in $\hat{H}_{\boldsymbol{\beta}}$) is given by

$$\frac{d\hat{l}}{d\boldsymbol{\beta}} = \int \frac{\partial \hat{l}}{\partial dH(s)} \cdot \frac{\partial \widehat{dH}_{\boldsymbol{\beta}}(s)}{\partial \boldsymbol{\beta}} + \frac{\partial \hat{l}}{\partial \boldsymbol{\beta}} \tag{A.35}$$

Then, since $U_{\hat{H}_{\boldsymbol{\beta}}} = 0$, the profile score is given by

$$U_{pr} = U_{\hat{H}_{\boldsymbol{\beta}}} J_{H\boldsymbol{\beta}} + U_{\boldsymbol{\beta}}|_{H=\hat{H}_{\boldsymbol{\beta}}} = U_{\boldsymbol{\beta}}|_{H=\hat{H}_{\boldsymbol{\beta}}} \tag{A.36}$$

The profile likelihood Hessian is

$$\begin{aligned}
-I_{pr} &= \frac{d^2\hat{l}}{d\boldsymbol{\beta} d\boldsymbol{\beta}^T} \\
&= \int_y \int_s \frac{\partial}{\partial dH(y)} \left[\frac{\partial \hat{l}}{\partial dH(s)} \cdot \frac{\partial \widehat{dH}_{\boldsymbol{\beta}}(s)}{\partial \boldsymbol{\beta}} \right] \frac{\partial \widehat{dH}_{\boldsymbol{\beta}}(y)}{\partial \boldsymbol{\beta}^T} + \int_y \frac{\partial \hat{l}}{\partial \boldsymbol{\beta} \partial dH(y)} \cdot \frac{\partial \widehat{dH}_{\boldsymbol{\beta}}(y)}{\partial \boldsymbol{\beta}^T} \\
&+ \int_s \frac{\partial}{\partial \boldsymbol{\beta}^T} \left[\frac{\partial \hat{l}}{\partial dH(s)} \cdot \frac{\partial \widehat{dH}_{\boldsymbol{\beta}}(s)}{\partial \boldsymbol{\beta}} \right] + \frac{\partial^2 \hat{l}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \\
&= \int_y \int_s \left[\frac{\partial^2 \hat{l}}{\partial dH(s) \partial dH(y)} \cdot \frac{\partial \widehat{dH}_{\boldsymbol{\beta}}(s)}{\partial \boldsymbol{\beta}} \cdot \frac{\partial \widehat{dH}_{\boldsymbol{\beta}}(y)}{\partial \boldsymbol{\beta}^T} + \frac{\partial \hat{l}}{\partial dH(s)} \cdot \frac{\partial^2 \widehat{dH}_{\boldsymbol{\beta}}(s)}{\partial \boldsymbol{\beta} \partial dH(y)} \cdot \frac{\partial \widehat{dH}_{\boldsymbol{\beta}}(y)}{\partial \boldsymbol{\beta}^T} \right] \\
&+ \int_s \left[\frac{\partial^2 \hat{l}}{\partial \boldsymbol{\beta} \partial dH(s)} \cdot \frac{\partial \widehat{dH}_{\boldsymbol{\beta}}(s)}{\partial \boldsymbol{\beta}^T} + \frac{\partial^2 \hat{l}}{\partial dH(s) \partial \boldsymbol{\beta}^T} \cdot \frac{\partial \widehat{dH}_{\boldsymbol{\beta}}(s)}{\partial \boldsymbol{\beta}} + \frac{\partial \hat{l}}{\partial dH(s)} \cdot \frac{\partial^2 \widehat{dH}_{\boldsymbol{\beta}}(s)}{\partial \beta \partial \boldsymbol{\beta}^T} \right] + \frac{\partial^2 \hat{l}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \\
&= -J_{\boldsymbol{\beta}H} \hat{I}_{HH} J_{H\boldsymbol{\beta}} + \hat{U}_{\hat{H}_{\boldsymbol{\beta}}} J_{H\boldsymbol{\beta}H} J_{H\boldsymbol{\beta}} - \hat{I}_{\boldsymbol{\beta}H} J_{H\boldsymbol{\beta}} - J_{\boldsymbol{\beta}H} \hat{I}_{H\boldsymbol{\beta}} + \hat{U}_{\hat{H}_{\boldsymbol{\beta}}} J_{H\boldsymbol{\beta}\boldsymbol{\beta}} - \hat{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} \\
&= -\hat{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} - J_{\boldsymbol{\beta}H} \hat{I}_{HH} J_{H\boldsymbol{\beta}} - \hat{I}_{\boldsymbol{\beta}H} J_{H\boldsymbol{\beta}} - J_{\boldsymbol{\beta}H} \hat{I}_{H\boldsymbol{\beta}} \\
\implies I_{pr} &= \hat{I}_{\boldsymbol{\beta}\boldsymbol{\beta}} + J_{\boldsymbol{\beta}H} \hat{I}_{HH} J_{H\boldsymbol{\beta}} + \hat{I}_{\boldsymbol{\beta}H} J_{H\boldsymbol{\beta}} + J_{\boldsymbol{\beta}H} \hat{I}_{H\boldsymbol{\beta}}
\end{aligned} \tag{A.37}$$

Note that

$$\begin{aligned}
0 = \frac{d}{d\boldsymbol{\beta}} \left(\frac{\partial l}{\partial dH(s)} \Big|_{H=\hat{H}_{\boldsymbol{\beta}}} \right) &= \frac{\partial^2 \hat{l}}{\partial dH(s) \partial \boldsymbol{\beta}} + \int \frac{\partial^2 \hat{l}}{\partial dH(s) \partial dH(y)} \cdot \frac{\partial \widehat{dH}_{\boldsymbol{\beta}}(y)}{\partial \boldsymbol{\beta}} = -\hat{I}_{H\boldsymbol{\beta}} - \hat{I}_{HH} J_{H\boldsymbol{\beta}} \\
\implies \hat{I}_{HH} J_{H\boldsymbol{\beta}} &= -\hat{I}_{H\boldsymbol{\beta}} \\
J_{H\boldsymbol{\beta}} &= -\hat{I}_{HH}^{-1} \hat{I}_{H\boldsymbol{\beta}}
\end{aligned} \tag{A.38}$$

Then,

$$\begin{aligned} I_{pr} &= \hat{I}_{\boldsymbol{\beta\beta}} + \hat{I}_{\boldsymbol{\beta}H}\hat{I}_{HH}^{-1}\hat{I}_{HH}\hat{I}_{HH}^{-1}\hat{I}_{H\boldsymbol{\beta}} - \hat{I}_{\boldsymbol{\beta}H}\hat{I}_{HH}^{-1}\hat{I}_{H\boldsymbol{\beta}} - \hat{I}_{\boldsymbol{\beta}H}\hat{I}_{HH}^{-1}\hat{I}_{H\boldsymbol{\beta}} \\ &= \hat{I}_{\boldsymbol{\beta\beta}} + \hat{I}_{\boldsymbol{\beta}H}\hat{I}_{HH}^{-1}\hat{I}_{H\boldsymbol{\beta}} - 2\hat{I}_{\boldsymbol{\beta}H}\hat{I}_{HH}^{-1}\hat{I}_{H\boldsymbol{\beta}} \\ &= \hat{I}_{\boldsymbol{\beta\beta}} - \hat{I}_{\boldsymbol{\beta}H}\hat{I}_{HH}^{-1}\hat{I}_{H\boldsymbol{\beta}} \\ &= Q \end{aligned} \tag{A.39}$$

Then, I^{*-1} has $\boldsymbol{\beta\beta}$ submatrix $Q^{-1} = I_{pr}^{-1}$. Direct algebraic manipulations lead to an expression for the normalized observed information matrix, given as

$$I_{n,\Omega\Omega^T} = \frac{1}{n}\sum_{i=1}^{n}\int_0^\infty \frac{\partial \log d\Lambda_i(x)}{d\Omega}\frac{\partial \log d\Lambda_i(x)}{\partial\Omega^T}Y_i(x)d\Lambda_i(x) + \frac{\partial^2 \log d\Lambda_i(x)}{\partial\Omega\partial\Omega^T}dM_i(x) \tag{A.40}$$

where $\Omega = (\boldsymbol{\beta}, H)$, $\Lambda_i =$ subject-specific cumulative hazard, $Y_i =$ at-risk process, $N_i =$ subject's failure counting process, and $dM_i = dN_i - Y_i d\Lambda_i =$ martingale increment under the true model. Note that $\frac{\partial\Lambda(x)}{\partial dH(s)} = 0, s > x$, meaning that terms corresponding to the $dH(s)$ functional component of Ω under the integral are 0 until $x \geq s$. As the martingale term turns into an $o_p(1)$ term and the first term is a consistent estimator of the covariance of the normalized score, we have

$$\begin{aligned} Cov\big(\sqrt{n}U^*\big) &= I_\infty + o_p(1) \\ \sqrt{n}(\hat{\Omega} - \Omega^*) &= I_\infty^{-1}\sqrt{n}U^* + o_p(1) \\ \implies Var\big[\sqrt{n}(\hat{\Omega} - \Omega^*)\big] &= I_\infty^{-1}Cov\big(\sqrt{n}U^*\big)I_\infty^{-1} = I_\infty^{-1}I_\infty I_\infty^{-1} = I_\infty^{-1} \\ \implies Var\big[\sqrt{n}(\hat{\beta} - \beta^*)\big] &= Q^{-1} = I_{pr}^{-1}, \text{ up to } o_p(1) \end{aligned} \tag{A.41}$$

Thus, the profile likelihood estimator $\hat{\boldsymbol{\beta}}$ is asymptotically efficient for $\boldsymbol{\beta}$ when we use the NPMLE $\hat{H}(\boldsymbol{\beta})$.

A.6 Asymptotic Variances of $\hat{\boldsymbol{\beta}}^B$ and $\hat{\boldsymbol{\beta}}^W$

From Rice and Tsodikov (2017), the variance of $\hat{\boldsymbol{\beta}}^W$ derived via profile likelihood is a consistent estimator of the covariance matrix of $\hat{\boldsymbol{\beta}}$, which will be the $\boldsymbol{\beta}\boldsymbol{\beta}$ submatrix of $\mathcal{I}^{*-1}$, given in (A.33). From (A.41), $Var\left[\sqrt{n}(\hat{\boldsymbol{\beta}}^W - \boldsymbol{\beta}^*)\right] = Q^{-1} = I_{pr}^{-1}$, up to $o_p(1)$, where $\boldsymbol{\beta}^*$ is the true parameter value. We establish the asymptotic variance of $\hat{\boldsymbol{\beta}}^B$. In Rice and Tsodikov (2017), it was assumed that all scores were normalized (i.e. multiplied by $\frac{1}{n}$). We make the same assumption here. Consider the estimating equation $U^B(\boldsymbol{\beta}, \hat{H}^B(\boldsymbol{\beta}))$ that $\hat{\boldsymbol{\beta}}^B$ solves, i.e., $U^B(\hat{\boldsymbol{\beta}}^B, \hat{H}^B(\hat{\boldsymbol{\beta}}^B)) = 0$. This estimating equation is identical in form to the profile likelihood score for $\boldsymbol{\beta}$ in Chapter II, Section 2.2.3, equation (2.11), which can be represented as $U(\boldsymbol{\beta}, \hat{H}^W(\boldsymbol{\beta}))$ when using the NPMLE of H. Then, we can express the normalized estimating equation for $\boldsymbol{\beta}$ as

$$0 = \sqrt{n}U^B(\hat{\boldsymbol{\beta}}^B, \hat{H}^B(\hat{\boldsymbol{\beta}}^B)) = \sqrt{n}U(\hat{\boldsymbol{\beta}}^B, \hat{H}^B(\hat{\boldsymbol{\beta}}^B)) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i(\hat{\boldsymbol{\beta}}^B, \hat{H}^B(\hat{\boldsymbol{\beta}}^B)) \tag{A.42}$$

Now, we take a Taylor expansion of the profile likelihood "score" for $\boldsymbol{\beta}$ about the true $\boldsymbol{\beta}^*$. This leads to

$$\begin{aligned} 0 =& \sqrt{n}U^B(\hat{\boldsymbol{\beta}}^B, \hat{H}^B(\hat{\boldsymbol{\beta}}^B)) \\ =& \frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i(\boldsymbol{\beta}^*, \hat{H}^B(\boldsymbol{\beta}^*)) + \left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \dot{U}_{i,\boldsymbol{\beta}}(\boldsymbol{\beta}^*, \hat{H}^B(\boldsymbol{\beta}^*)) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \dot{U}_{i,H}(\boldsymbol{\beta}^*, \hat{H}^B(\boldsymbol{\beta}^*))\hat{J}^B_{H\boldsymbol{\beta}}\right](\hat{\boldsymbol{\beta}}^B - \boldsymbol{\beta}^*) \\ &+ \frac{1}{\sqrt{n}}o(||\hat{\boldsymbol{\beta}}^B - \boldsymbol{\beta}^*||^2) \end{aligned} \tag{A.43}$$

where $\dot{U}_{i,\boldsymbol{\beta}} = \frac{\partial}{\partial \boldsymbol{\beta}}U_i(\boldsymbol{\beta})$, $\dot{U}_{i,H} = \frac{\partial}{\partial H}U_i(\boldsymbol{\beta})$, and $\hat{J}^B_{H\boldsymbol{\beta}} = \frac{\partial}{\partial \boldsymbol{\beta}}\hat{H}^B(\boldsymbol{\beta}) =$ the Jacobian of $\hat{H}^B$ with respect to $\boldsymbol{\beta}$. For brevity, we drop the $(\boldsymbol{\beta}^*, \hat{H}^B(\boldsymbol{\beta}^*))$ argument from U_i^B and U_i. Note that $\frac{1}{\sqrt{n}}o(||\hat{\boldsymbol{\beta}}^B - \boldsymbol{\beta}^*||^2)$ and all higher order terms will be asymptotically negligible. Thus,

$$0 = \sqrt{n}U^B(\hat{\boldsymbol{\beta}}^B, \hat{H}^B(\hat{\boldsymbol{\beta}}^B)) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} U_i + \left[\frac{1}{n}\sum_{i=1}^{n} \dot{U}_{i,\boldsymbol{\beta}} + \frac{1}{n}\sum_{i=1}^{n} \dot{U}_{i,H}\hat{J}^B_{H\boldsymbol{\beta}}\right]\sqrt{n}(\hat{\boldsymbol{\beta}}^B - \boldsymbol{\beta}^*) + o_p(1) \tag{A.44}$$

Moving the term involving $\sqrt{n}(\hat{\boldsymbol{\beta}}^B - \boldsymbol{\beta}^*)$ to the left hand side results in

$$
\begin{aligned}
-\left[\frac{1}{n}\sum_{i=1}^{n}\dot{U}_{i,\boldsymbol{\beta}} + \frac{1}{n}\sum_{i=1}^{n}\dot{U}_{i,H}\hat{J}^B_{H\boldsymbol{\beta}}\right]\sqrt{n}(\hat{\boldsymbol{\beta}}^B - \boldsymbol{\beta}^*) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i + o_p(1) \\
\implies \sqrt{n}(\hat{\boldsymbol{\beta}}^B - \boldsymbol{\beta}^*) &= -\left[\frac{1}{n}\sum_{i=1}^{n}\dot{U}_{i,\boldsymbol{\beta}} + \frac{1}{n}\sum_{i=1}^{n}\dot{U}_{i,H}\hat{J}^B_{H\boldsymbol{\beta}}\right]^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i + o_p(1)
\end{aligned}
\tag{A.45}
$$

By results from Rice and Tsodikov (2017) and consistency of $\hat{H}^B$,

$$
\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i &\xrightarrow{d} N(0, E\{UU^T\}) \\
\frac{1}{n}\sum_{i=1}^{n}\dot{U}_{i,\boldsymbol{\beta}} &\xrightarrow{p} -I_{\boldsymbol{\beta}\boldsymbol{\beta}} \\
\frac{1}{n}\sum_{i=1}^{n}\dot{U}_{i,H}\hat{J}^B_{H\boldsymbol{\beta}} &\xrightarrow{p} -I_{\boldsymbol{\beta}H}J_{H\boldsymbol{\beta}} = -I_{\boldsymbol{\beta}\boldsymbol{\beta}}(-I^{-1}_{HH}I_{H\boldsymbol{\beta}}) = I_{\boldsymbol{\beta}\boldsymbol{\beta}}I^{-1}_{HH}I_{H\boldsymbol{\beta}}
\end{aligned}
\tag{A.46}
$$

Therefore,

$$
-\left[\frac{1}{n}\sum_{i=1}^{n}\dot{U}_{i,\boldsymbol{\beta}} + \frac{1}{n}\sum_{i=1}^{n}\dot{U}_{i,H}\hat{J}^B_{H\boldsymbol{\beta}}\right]^{-1} \xrightarrow{p} -\left\{-I_{\boldsymbol{\beta}\boldsymbol{\beta}} + I_{\boldsymbol{\beta}H}I^{-1}_{HH}I_{H\boldsymbol{\beta}}\right\}^{-1} = \left\{I_{\boldsymbol{\beta}\boldsymbol{\beta}} - I_{\boldsymbol{\beta}H}I^{-1}_{HH}I_{H\boldsymbol{\beta}}\right\}^{-1} = Q^{-1}
\tag{A.47}
$$

Thus,

$$
\begin{aligned}
\sqrt{n}(\hat{\boldsymbol{\beta}}^B - \boldsymbol{\beta}^*) &= -\left[\frac{1}{n}\sum_{i=1}^{n}\dot{U}_{i,\boldsymbol{\beta}} + \frac{1}{n}\sum_{i=1}^{n}\dot{U}_{i,H}\hat{J}^B_{H\boldsymbol{\beta}}\right]^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i + o_p(1) \\
&\xrightarrow{d} Q^{-1}N(0, E\{UU^T\}) = N(0, Q^{-1}E\{UU^T\}Q^{-1})
\end{aligned}
\tag{A.48}
$$

From (A.46), the estimating equation for $\boldsymbol{\beta}$ when using $\hat{H}^B$ does not have the same property as the score where $E\{UU^T\} = E\{-\frac{\partial}{\partial\boldsymbol{\beta}}U\}$. Therefore, the asymptotic variance of $\hat{\boldsymbol{\beta}}^B$ contains additional terms that do not cancel. Nonetheless, the difference in the variances of $\hat{\boldsymbol{\beta}}^W$ and $\hat{\boldsymbol{\beta}}^B$ is expected to be small, especially for large n.

Alternatively, following an argument similar to Zucker (2005), we can write

$$
\begin{aligned}
0 = \sqrt{n}U(\hat{\boldsymbol{\beta}}^W, \hat{H}^W(\hat{\boldsymbol{\beta}}^W)) &= \sqrt{n}U(\boldsymbol{\beta}^*, H^*(\boldsymbol{\beta}^*)) + \sqrt{n}\big[U(\boldsymbol{\beta}^*, \hat{H}^W(\boldsymbol{\beta}^*)) - U(\boldsymbol{\beta}^*, H^*(\boldsymbol{\beta}^*))\big] \\
&\quad + \sqrt{n}\big[U(\hat{\boldsymbol{\beta}}^W, \hat{H}^W(\hat{\boldsymbol{\beta}}^W)) - U(\boldsymbol{\beta}^*, \hat{H}^W(\boldsymbol{\beta}^*))\big] \\
0 = \sqrt{n}U^B(\hat{\boldsymbol{\beta}}^B, \hat{H}^B(\hat{\boldsymbol{\beta}}^B)) &= \sqrt{n}U(\hat{\boldsymbol{\beta}}^B, \hat{H}^B(\hat{\boldsymbol{\beta}}^B)) \\
&= \sqrt{n}U(\boldsymbol{\beta}^*, H^*(\boldsymbol{\beta}^*)) + \sqrt{n}\big[U(\boldsymbol{\beta}^*, \hat{H}^B(\boldsymbol{\beta}^*)) - U(\boldsymbol{\beta}^*, H^*(\boldsymbol{\beta}^*))\big] \\
&\quad + \sqrt{n}\big[U(\hat{\boldsymbol{\beta}}^B, \hat{H}^B(\hat{\boldsymbol{\beta}}^B)) - U(\boldsymbol{\beta}^*, \hat{H}^B(\boldsymbol{\beta}^*))\big]
\end{aligned}
\tag{A.49}
$$

Once more, we assume all scores and estimating equations are normalized, i.e., multiplied by $\frac{1}{n}$. It can be shown that when we take the appropriate Taylor expansions about the true $\boldsymbol{\beta}^*$ and H^*, the asymptotic distributions of $\hat{\boldsymbol{\beta}}^B$ and $\hat{\boldsymbol{\beta}}^W$ are

$$\begin{aligned}\sqrt{n}(\hat{\boldsymbol{\beta}}^B - \boldsymbol{\beta}^*) &\xrightarrow{d} N(0, Q^{-1}\{I_{\boldsymbol{\beta}\boldsymbol{\beta}} + I_{\boldsymbol{\beta}H} Var_B I_{H\boldsymbol{\beta}}\}Q^{-1}) \\ \sqrt{n}(\hat{\boldsymbol{\beta}}^W - \boldsymbol{\beta}^*) &\xrightarrow{d} N(0, Q^{-1}\{I_{\boldsymbol{\beta}\boldsymbol{\beta}} + I_{\boldsymbol{\beta}H}(Var_B - R) I_{H\boldsymbol{\beta}}\}Q^{-1})\end{aligned} \tag{A.50}$$

where Var_B represents the Breslow estimator variance in Chapter II, Section 2.2.6, equation (2.19) and R is the integral term in the NPMLE variance in Chapter II, Section 2.2.6, equation (2.23). Because the asymptotic variance of $\hat{\boldsymbol{\beta}}^W$ has an additional term being subtracted, its variance will be smaller than that of $\hat{\boldsymbol{\beta}}^B$. However, this difference is expected to be small, particularly for large n.

A.7 Details of Proportional Odds Model

The proportional odds (PO) model as parameterized in Tsodikov (2003) is given by

$$S(H_t|\boldsymbol{\beta}, \boldsymbol{Z}) = \frac{\exp\{\boldsymbol{\beta}\boldsymbol{Z}\}}{H_t + \exp\{\boldsymbol{\beta}\boldsymbol{Z}\}} \tag{A.51}$$

where $\boldsymbol{Z}$ is a vector of baseline and external time-dependent covariates. Assume we have underlying time-independent frailty U. Then,

$$\begin{aligned} S(H_t|\boldsymbol{\beta}, \boldsymbol{Z}) = \frac{\exp\{\boldsymbol{\beta}\boldsymbol{Z}\}}{H_t + \exp\{\boldsymbol{\beta}\boldsymbol{Z}\}} &= \mathcal{L}_U(H_t) \\ U &\sim \text{Exp}(\exp\{\boldsymbol{\beta}\boldsymbol{Z}\}) \\ d\Lambda(H_t|\boldsymbol{\beta}, \boldsymbol{Z}) = d\{-\ln S(H_t|\boldsymbol{\beta}, \boldsymbol{Z})\} &= \frac{1}{H_t + \exp\{\boldsymbol{\beta}\boldsymbol{Z}\}} \cdot dH_t = \Theta^0 dH_t \end{aligned} \tag{A.52}$$

In our simulation study, we assume that the true baseline hazard function for the event time T is specified by the Weibull cumulative baseline hazard function $H_t = \left(\frac{t}{\lambda}\right)^k$ and baseline hazard function $h_t = \frac{kt^{k-1}}{\lambda^k}$. In the hazard function, k controls whether the hazards are increasing ($k > 1$) or decreasing ($k < 1$).

A.8 Details of Simulations

The ARE of the Breslow estimator is

$$ARE_B(t) = \frac{1}{1 - \frac{\int_0^t [B_x - \int_x^t B_a dH_a] dVar_B^2(x)}{Var_B(t)}} = \frac{1}{1 - \frac{\int_0^t \int_0^x \frac{dH_y}{A_y} 2B_x \frac{dH_x}{A_x} - \int_0^t \int_0^x \int_x^t B_a dH_a 2 \frac{dH_y}{A_y} \frac{dH_x}{A_x}}{\int_0^t \frac{dH_x}{A_x}}} \tag{A.53}$$

where

$$\begin{aligned}
Var_B(t) &= \int_0^t \frac{dH_x}{A_x} \\
dVar_B^2(x) &= 2\frac{dH_x}{A_x} Var_B(x) = 2\frac{dH_x}{A_x} \int_0^x \frac{dH_y}{A_y} \\
A_x &= E_{\boldsymbol{Z}_x}\{S_x^* G_x^* \Theta_x^0\} \\
B_x &= E_{\boldsymbol{Z}_x}\{[\Theta_x^0 - \Theta_x^1]^2 S_x^* G_x^* \Theta_x^0\} \\
S_x^* &= P(T > x | \boldsymbol{Z}) \\
G_x^* &= P(C > x | \boldsymbol{Z})
\end{aligned} \tag{A.54}$$

We assume

$$\begin{aligned}
P(T > t|\boldsymbol{Z}) = S_T^*(t|\boldsymbol{\beta}, \boldsymbol{Z}) &= \frac{e^{\boldsymbol{\beta Z}}}{H_t + e^{\boldsymbol{\beta Z}}} \Longrightarrow d\Lambda_T^*(t|\boldsymbol{\beta}, \boldsymbol{Z}) = \frac{1}{H_t + e^{\boldsymbol{\beta Z}}} \cdot dH_t \\
\Theta^0 &= \frac{1}{H_t + e^{\boldsymbol{\beta Z}}} \\
\Theta^1 &= \frac{2}{H_t + e^{\boldsymbol{\beta Z}}} \\
[\Theta^0 - \Theta^1]^2 &= \frac{1}{(H_t + e^{\boldsymbol{\beta Z}})^2} \\
P(C > t|\boldsymbol{Z}) &= 1 - \frac{t}{\tau},\ C \sim Uniform(0, \tau) \\
H_t &= \left(\frac{t}{\lambda}\right)^k \Longrightarrow h_t = \frac{kt^{k-1}}{\lambda^k} \text{ (Weibull hazard)} \\
Z &\sim Bernoulli(p) \\
A_x &= E_Z\{S_x^* G_x^* \Theta_x^0\} = \left(1 - \frac{x}{\tau}\right)\left\{\frac{1-p}{(H_x+1)^2} + \frac{pe^\beta}{(H_x + e^\beta)^2}\right\} \\
B_x &= E_Z\{[\Theta_x^0 - \Theta_x^1]^2 S_x^* G_x^* \Theta_x^0\} = \left(1 - \frac{x}{\tau}\right)\left\{\frac{1-p}{(H_x+1)^4} + \frac{pe^\beta}{(H_x + e^\beta)^4}\right\}
\end{aligned} \tag{A.55}$$

We can plug the expressions for A_x and B_x into (A.53) over a range of $t \in (0, \tau)$ to study the behavior of the ARE over time.

Table A.1: Simulation settings for ARE of the Breslow estimator.

Setting	p	β	λ	k	τ
1	0.3	-1	2	2	5
2	0.5	-1	2	2	5
3	0.7	-1	2	2	5
4	0.3	0	2	2	5
5	0.5	0	2	2	5
6	0.7	0	2	2	5
7	0.3	1	2	2	5
8	0.5	1	2	2	5
9	0.7	1	2	2	5
10	0.3	-1	2	0.5	5
11	0.5	-1	2	0.5	5
12	0.7	-1	2	0.5	5
13	0.3	0	2	0.5	5
14	0.5	0	2	0.5	5
15	0.7	0	2	0.5	5
16	0.3	1	2	0.5	5
17	0.5	1	2	0.5	5
18	0.7	1	2	0.5	5
19	0.3	-1	2	2	10
20	0.5	-1	2	2	10
21	0.7	-1	2	2	10
22	0.3	0	2	2	10
23	0.5	0	2	2	10
24	0.7	0	2	2	10
25	0.3	1	2	2	10
26	0.5	1	2	2	10
27	0.7	1	2	2	10
28	0.3	-1	2	0.5	10
29	0.5	-1	2	0.5	10
30	0.7	-1	2	0.5	10
31	0.3	0	2	0.5	10
32	0.5	0	2	0.5	10
33	0.7	0	2	0.5	10
34	0.3	1	2	0.5	10
35	0.5	1	2	0.5	10
36	0.7	1	2	0.5	10
37	0.3	-1	2	2	15
38	0.5	-1	2	2	15
39	0.7	-1	2	2	15
40	0.3	0	2	2	15
41	0.5	0	2	2	15
42	0.7	0	2	2	15
43	0.3	1	2	2	15
44	0.5	1	2	2	15
45	0.7	1	2	2	15
46	0.3	-1	2	0.5	15
47	0.5	-1	2	0.5	15
48	0.7	-1	2	0.5	15
49	0.3	0	2	0.5	15
50	0.5	0	2	0.5	15
51	0.7	0	2	0.5	15
52	0.3	1	2	0.5	15
53	0.5	1	2	0.5	15
54	0.7	1	2	0.5	15

A.9 Additional Simulation Results

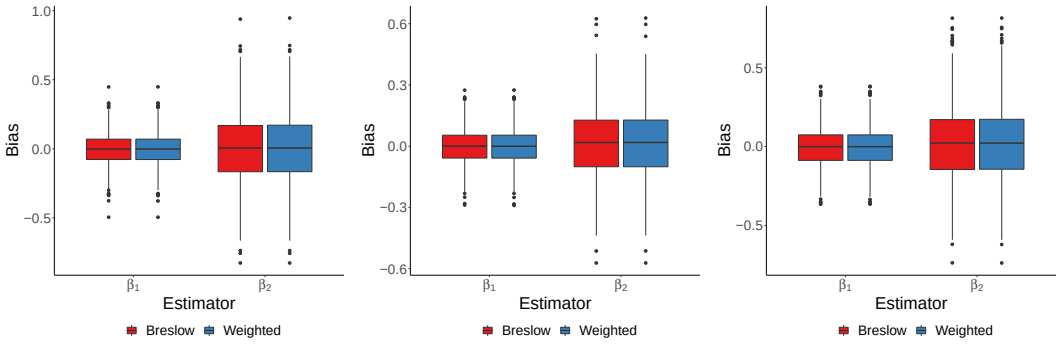
From Table A.2, for estimation of $\boldsymbol{\beta}$, the Breslow estimator takes fewer iterations to converge on average than the Weighted Breslow NPMLE. For estimation of $\{dH\}$, the Breslow and Weighted Breslow estimators have the same average number of iterations in the large sample settings, while the Breslow estimator has slightly fewer iterations in the small sample settings. Table A.3 summarizes differences in the estimated variances of $\hat{\boldsymbol{\beta}}$ between the Breslow and Weighted Breslow estimators. In each setting, including the small-sample cases, the differences in the estimated variances are small. Figure A.1 displays boxplots of bias in $\boldsymbol{\beta}$ for both estimators in each simulation setting. We observe that the bias in estimation for both estimators is nearly identical in all settings. Furthermore, the range in bias for both estimators is comparable, suggesting that they are similar in terms of consistency and asymptotic variance.

Table A.2: Average number of iterations to converge for each estimator and average percent censoring in PO model simulations. Weighted Breslow estimator = NPMLE.

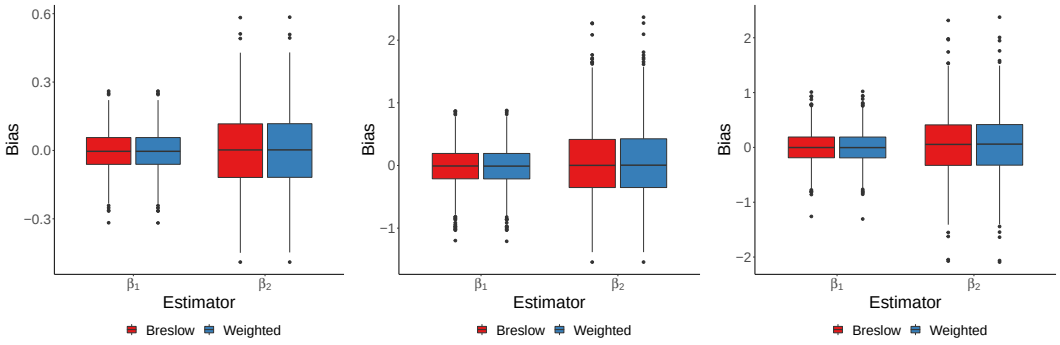
Setting	**Estimator**	$\hat{\boldsymbol{\beta}}$ **Iterations**	$\{\widehat{dH}\}$ **Iterations**	**% Censoring**
1	Breslow	3.6	22.0	23.2
	Weighted	4.3	22.0	
2	Breslow	3.6	21.0	23.1
	Weighted	4.0	21.0	
3	Breslow	3.6	10.8	46.6
	Weighted	4.2	10.8	
4	Breslow	3.7	10.4	46.6
	Weighted	4.0	10.4	
5	Breslow	3.6	12.4	47.0
	Weighted	5.1	12.5	
6	Breslow	3.6	23.4	23.8
	Weighted	5.5	23.6	

Table A.3: Summary of differences in regression parameter variances for each estimator. $\hat{V}(\hat{\boldsymbol{\beta}}^B)$ = estimated variance of Breslow-derived $\boldsymbol{\beta}$ and $\hat{V}(\hat{\boldsymbol{\beta}}^W)$ = estimated variance of Weighted Breslow/NPMLE-derived $\boldsymbol{\beta}$.

Setting	$\hat{V}(\hat{\beta_1}^B) - \hat{V}(\hat{\beta_1}^W)$ **Mean [Min, Max]**	$\hat{V}(\hat{\beta_2}^B) - \hat{V}(\hat{\beta_2}^W)$ **Mean [Min, Max]**
1	-2.5×10^{-5} $[-2.1 \times 10^{-4}, 1.9 \times 10^{-4}]$	-5.0×10^{-5} $[-1.8 \times 10^{-3}, 9.9 \times 10^{-4}]$
2	-5.2×10^{-7} $[-4.9 \times 10^{-5}, 7.9 \times 10^{-5}]$	-1.2×10^{-5} $[-3.4 \times 10^{-4}, 5.0 \times 10^{-4}]$
3	-7.7×10^{-7} $[-1.7 \times 10^{-4}, 2.7 \times 10^{-4}]$	-2.7×10^{-5} $[-8.5 \times 10^{-4}, 1.1 \times 10^{-3}]$
4	-1.4×10^{-6} $[-4.5 \times 10^{-5}, 5.0 \times 10^{-5}]$	-1.3×10^{-5} $[-3.0 \times 10^{-4}, 2.6 \times 10^{-4}]$
5	-5.4×10^{-5} $[-6.6 \times 10^{-3}, 2.8 \times 10^{-3}]$	-1.0×10^{-3} $[-7.8 \times 10^{-2}, 2.2 \times 10^{-2}]$
6	-6.3×10^{-5} $[-8.9 \times 10^{-3}, 4.8 \times 10^{-3}]$	-1.4×10^{-3} $[-4.1 \times 10^{-2}, 2.8 \times 10^{-2}]$



(a) Setting 1 (b) Setting 2 (c) Setting 3



(d) Setting 4 (e) Setting 5 (f) Setting 6

Figure A.1: Boxplots of bias for Breslow and Weighted Breslow estimators in each PO simulation setting. Weighted estimator = NPMLE.

APPENDIX B

Reconsidering the Role of Endogenous Covariates in Survival Models

B.1 Bivariate Shared Frailty Model

Assuming X is fully observed, the conditional density and survival functions of $Y|X = x$ in Chapter III, Section 3.5.1 can be derived as

$$\begin{aligned}
f_{Y|X=x}(y|X=x) = \frac{f_{Y,X}(y,x)}{f_X(x)} &= \frac{\mathcal{L}_U''(H_X(x)+H_Y(y))h_X(x)h_Y(y)}{-\frac{\partial}{\partial x}S_X(x)} \\
&= \frac{\mathcal{L}_U''(H_X(x)+H_Y(y))h_X(x)h_Y(y)}{-\frac{\partial}{\partial x}\mathcal{L}_U(H_X(x))} \\
&= -\frac{\mathcal{L}_U''(H_X(x)+H_Y(y))}{\mathcal{L}_U'(H_X(x))}h_Y(y) \\
S_{Y|X=x}(y|X=x) = \int_y^\infty f_{Y|X=x}(v|X=x)dv &= \int_y^\infty -\frac{\mathcal{L}_U''(H_X(x)+H_Y(v))}{\mathcal{L}_U'(H_X(x))}h_Y(v)dv \\
&= \int_{H_Y(y)}^\infty -\frac{\mathcal{L}_U''(H_X(x)+v)}{\mathcal{L}_U'(H_X(x))}dv \\
&= \frac{\mathcal{L}_U'(H_X(x)+H_Y(y))}{\mathcal{L}_U'(H_X(x))}
\end{aligned} \tag{B.1}$$

B.2 Semicompeting Risks Type Model Survival Functions

The survival functions for X and Y given in Chapter III, Section 3.6, equation (3.21) can be derived as

$$\begin{aligned}
S_X(x|\boldsymbol{Z}) &= \exp\Big\{ -\int_0^x \lambda_X(u|\boldsymbol{Z})du \Big\} = \exp\Big\{ -\int_0^x h_u[\eta I(u \le Y) + \tilde{\eta} I(u > Y)]du \Big\} \\
&= \exp\Big\{ -\int_0^x \eta h_u I(u \le Y)du - \int_0^x \tilde{\eta} h_u I(u > Y)du \Big\} \\
&= \exp\Big\{ -\int_0^y \eta h_u du - \int_y^x \tilde{\eta} h_u du \Big\} \\
&= \exp\big\{ -\eta H_y - \tilde{\eta}[H_x - H_y] \big\} \\
S_{Y|X}(y|X, \boldsymbol{Z}) &= \exp\Big\{ -\int_0^y \lambda_{Y|X}(u|X, \boldsymbol{Z})du \Big\} = \exp\Big\{ -\int_0^y \delta_Y H_u^{\delta_Y - 1} h_u \theta \mu^{I(u \ge X)} du \Big\} \\
&= \exp\Big\{ -\int_0^x \delta_Y H_u^{\delta_Y - 1} h_u \theta du - \int_x^y \delta_Y H_u^{\delta_Y - 1} h_u \theta \mu du \Big\} \\
&= \exp\big\{ -H_x^{\delta_Y}\theta - [H_y^{\delta_Y}\theta\mu - H_x^{\delta_Y}\theta\mu] \big\} \\
&= \exp\big\{ -H_y^{\delta_Y}\theta\mu - H_x^{\delta_Y}\theta[1-\mu] \big\} \\
&= \exp\big\{ -H_y^{\delta_Y}\theta\mu - H_x^{\delta_Y}\theta\bar{\mu} \big\}
\end{aligned} \tag{B.2}$$

B.3 Validity of Univariate Survival Functions

We demonstrate that the univariate survival function $u(x) = \exp\{-H_x^{\delta_X}\theta\mu + H_x^{\delta_Y}\theta\bar{\mu}\}$ in (3.24) is a legitimate survival function by showing that it is nonincreasing and right continuous with $u(0) = 1$ and $\lim_{x\to\infty} u(x) = 0$.

(1) $\boldsymbol{u(x)}$ **nonincreasing:**

$$\begin{aligned} u'(x) &= e^{-H_x^{\delta_X}\theta\mu + H_x^{\delta_Y}\theta\bar{\mu}} \cdot \left\{ -\delta_X H_x^{\delta_X - 1} h_x \theta\mu + \delta_Y H_x^{\delta_Y - 1} h_x \theta\bar{\mu} \right\} \\ &= -\theta h_x e^{-H_x^{\delta_X}\theta\mu + H_x^{\delta_Y}\theta\bar{\mu}} \left\{ \delta_X H_x^{\delta_X - 1}\mu - \delta_Y H_x^{\delta_Y - 1}\bar{\mu} \right\} \\ &= -\theta h_x e^{-H_x^{\delta_X}\theta\mu + H_x^{\delta_Y}\theta\bar{\mu}} \left\{ \delta_X H_x^{\delta_X - 1}\mu - \delta_Y H_x^{\delta_Y - 1}(1-\mu) \right\} \\ &\leq 0 \end{aligned}$$

(2) $\boldsymbol{u(0) = 1}$: Assuming $H_0^{\delta_X} = H_0^{\delta_Y} = 0$,

$$u(0) = e^{-H_0^{\delta_X}\theta\mu + H_0^{\delta_Y}\theta\bar{\mu}} = e^0 = 1$$

(3) $\boldsymbol{\lim_{x\to\infty} u(x) = 0}$:

$$\begin{aligned} \lim_{x\to\infty} u(x) &= \lim_{x\to\infty} e^{-H_x^{\delta_X}\theta\mu + H_x^{\delta_Y}\theta\bar{\mu}} \\ &= \lim_{x\to\infty} e^{-H_x^{\delta_X}\theta\mu + H_x^{\delta_Y}\theta(1-\mu)} \\ &= 0 \end{aligned}$$

This assumption in combination with (1) and (2) ensures that the total variation of $u(x)$ is one.

(4) $\boldsymbol{u(x)}$ **right continuous:** Assuming the true H is right continuous, H^{δ_X} and H^{δ_Y} will also be continuous and therefore right continuous. Because $u(x)$ is a transformation of a right continuous function, it will also be right continuous

An analogous derivation for $v(y)$ in Chapter III, Section 3.6, equation (3.24) can be done using the same justification as above.

B.4 Semicompeting Risks Type Model Assuming X Fully Observed

Assuming X is fully observed, the survival function for Y conditional on $X = x$ is

$$\begin{aligned}
-\frac{\partial}{\partial x}S_{Y,X|\boldsymbol{Z}}(y,x|\boldsymbol{Z}) &= -\frac{\partial}{\partial x}\Big\{\exp\big\{-H_y^{\delta_Y}\theta\mu - H_x^{\delta_Y}\theta\bar{\mu} - \eta H_y - \tilde{\eta}[H_x - H_y]\big\}\Big\} \\
&= \big(\delta_Y H_x^{\delta_Y-1}h_x\theta\bar{\mu} + \tilde{\eta}h_x\big)\exp\big\{-H_y^{\delta_Y}\theta\mu - H_x^{\delta_Y}\theta\bar{\mu} - \eta H_y - \tilde{\eta}[H_x - H_y]\big\} \\
&= h_x\big(\delta_Y H_x^{\delta_Y-1}\theta\bar{\mu} + \tilde{\eta}\big)\exp\big\{-H_y^{\delta_Y}\theta\mu - H_x^{\delta_Y}\theta\bar{\mu} - \eta H_y - \tilde{\eta}[H_x - H_y]\big\} \\
-\frac{\partial}{\partial x}S_{X|\boldsymbol{Z}}(x|\boldsymbol{Z}) &= -\frac{\partial}{\partial x}\Big\{\exp\big\{-\eta H_y - \tilde{\eta}[H_x - H_y]\big\}\Big\} \\
&= h_x\tilde{\eta}\exp\big\{-\eta H_y - \tilde{\eta}[H_x - H_y]\big\} \\
S_{Y|X=x,\boldsymbol{Z}}(y|X=x,\boldsymbol{Z}) &= \frac{-\frac{\partial}{\partial x}S_{Y,X|\boldsymbol{Z}}(y,x|\boldsymbol{Z})}{-\frac{\partial}{\partial x}S_{X|\boldsymbol{Z}}(x|\boldsymbol{Z})} \\
&= \frac{h_x\big(\delta_Y H_x^{\delta_Y-1}\theta\bar{\mu} + \tilde{\eta}\big)\exp\big\{-H_y^{\delta_Y}\theta\mu - H_x^{\delta_Y}\theta\bar{\mu} - \eta H_y - \tilde{\eta}[H_x - H_y]\big\}}{h_x\tilde{\eta}\exp\big\{-\eta H_y - \tilde{\eta}[H_x - H_y]\big\}} \\
&= \left(1 + \frac{\delta_Y H_x^{\delta_Y-1}\theta\bar{\mu}}{\tilde{\eta}}\right)\exp\big\{-\theta[H_y^{\delta_Y}\mu + H_x^{\delta_Y}\bar{\mu}]\big\}
\end{aligned} \tag{B.3}$$

Because X is fully observed, the survival function in (B.3) has a valid exponential form, and we can directly calculate the conditional hazard function.

$$\begin{aligned}
\lambda_{Y|X=x,\boldsymbol{Z}}(y|X=x,\boldsymbol{Z}) &= -\frac{\partial}{\partial y}\ln S_{Y|X=x,\boldsymbol{Z}}(y|X=x,\boldsymbol{Z}) \\
&= -\frac{\partial}{\partial y}\left\{\ln\left(1 + \frac{\delta_Y H_x^{\delta_Y-1}\theta\bar{\mu}}{\tilde{\eta}}\right) - \theta[H_y^{\delta_Y}\mu + H_x^{\delta_Y}\bar{\mu}]\right\} \\
&= -\{-\theta\delta_Y H_y^{\delta_Y-1}h_y\mu\},\ x \le y \\
&= \theta\delta_Y H_y^{\delta_Y-1}h_y\mu^{I(y\ge x)}
\end{aligned} \tag{B.4}$$

B.5 Threshold Regression Models

B.5.1 Fixed Threshold and Partially Observed Marker

Consider the survival functions conditional on $\bar{V}_x$ and $\bar{V}$. First, we recognize that

$$\begin{aligned} W_t|\bar{V}_x &= V_{\tau_{x^-}} + W_{t-\tau_{x^-}},\ x \le t \\ W_t|\bar{V} &= V_{\tau_{t^-}} + U\{t\}_{[0,\tau_{t^+}-\tau_{t^-}]}(V_{\tau_{t^+}} - V_{\tau_{t^-}}) \end{aligned} \tag{B.5}$$

where U represents a bridge stochastic process. Then, the survival functions can be written as

$$\begin{aligned} S(t|\bar{V}_x)|_{x=t} &= P(T > t|\bar{V}_x)|_{x=t} = E_{\bar{W}}\big\{P(W_t < B)|\bar{V}_x\big\}|_{x=t} \\ &= E_{\bar{W}}\big\{P(V_{\tau_{x^-}} + W_{t-\tau_{x^-}} < B)|\bar{V}_x\big\}|_{x=t} \\ &= E_{\bar{W}}\big\{P(W_{t-\tau_{x^-}} < B - V_{\tau_{x^-}})|\bar{V}_x\big\}|_{x=t} \\ &= 1 - E_{\bar{W}}\big\{S_W(B - V_{\tau_{x^-}})|\bar{V}_x\big\}|_{x=t} \\ &= 1 - E_{\bar{W}}\big\{S_W(B - V_{\tau_{t^-}})|\bar{V}_t\big\} \\ S(t|\bar{V}) &= P(T > t|\bar{V}) = E_{\bar{W}}\big\{P(W_t < B)|\bar{V}\big\} \\ &= E_{\bar{W}}\big\{P(V_{\tau_{t^-}} + U(V_{\tau_{t^+}} - V_{\tau_{t^-}}) < B)|\bar{V}\big\} \\ &= E_U\left\{P\left(U < \frac{B - V_{\tau_{t^-}}}{V_{\tau_{t^+}} - V_{\tau_{t^-}}}\right)\middle|\bar{V}\right\} \\ &= 1 - E_U\left\{S_U\left(\frac{B - V_{\tau_{t^-}}}{V_{\tau_{t^+}} - V_{\tau_{t^-}}}\right)\middle|\bar{V}\right\} \end{aligned} \tag{B.6}$$

B.5.2 Random Threshold and Partially Observed Marker

Consider the survival functions conditional on $\bar{V}_x$ and $\bar{V}$. Once more, we have $W_t|\bar{V}_x = V_{\tau_{x^-}} + W_{t-\tau_{x^-}}$, $x \le t$, and $W_t|\bar{V} = V_{\tau_{t^-}} + U\{t\}_{[0,\tau_{t^+}-\tau_{t^-}]}(V_{\tau_{t^+}} - V_{\tau_{t^-}})$. Then,

the survival functions are

$$
\begin{aligned}
S(t|\bar{V}_x)|_{x=t} &= P(T > t|\bar{V}_x)|_{x=t} = E_{\bar{W}}\big\{P(W_t < B|\bar{W})\big|\bar{V}_x\big\}\big|_{x=t} \\
&= E_{\bar{W}}\big\{S_{B|\bar{W}}(W_t)\big|\bar{V}_x\big\}\big|_{x=t} \\
&= E_{\bar{W}}\big\{S_{B|\bar{W}}(V_{\tau_{x^-}} + W_{t-\tau_{x^-}})\big|\bar{V}_x\big\}\big|_{x=t} \\
&= E_{\bar{W}}\big\{S_{B|\bar{W}}(V_{\tau_{t^-}} + W_{t-\tau_{t^-}})|\bar{V}_t\big\} \\
S(t|\bar{V}) &= P(T > t|\bar{V}) = E_{\bar{W}}\big\{P(W_t < B|\bar{W})\big|\bar{V}\big\} = E_{\bar{W}}\big\{S_{B|\bar{W}}(W_t)\big|\bar{V}\big\} \\
&= E_{\bar{W}}\big\{S_{B|U}(V_{\tau_{t^-}} + U_{[0,\tau_{t^+}-\tau_{t^-}]}(V_{\tau_{t^+}} - V_{\tau_{t^-}}))|\bar{V}\big\}
\end{aligned}
\tag{B.7}
$$

APPENDIX C

Joint Modeling of a Time-to-Event and Partially Observed Marker Process Using Lévy Processes

Many of the derivations detailed below are based on those outlined in Appendix B of Suresh (2018), p. 169-179.

C.1 Marker Observed at Informative Measurement Time

The joint model of interest can be written as

$$f(t, V_t) = f(t|V_x)|_{x=t} f(V_t) = f(V_t|T = t)f(t) \tag{C.1}$$

The first specification is based on the general survival function proposed in Chapter III. Namely, we condition the event time distribution on the marker observed at the the fixed time x and then substitute $x = t$. Note that the proper marginal distribution of the marker V in this joint model is derived by averaging across all possible event times T. This is done to account for the informative observation of the marker at T. For brevity, we have dropped the conditioning on baseline covariates $\boldsymbol{Z}$ from our expressions, but it is implicit that the distributions of the event time and marker will

depend on these covariates as the shape and scale parameters α and β are functions of $\boldsymbol{Z}$. Then, the marginal distribution of V (on the H-transformed time scale) is given by

$$\begin{aligned} f_V(v) &= \int_0^\infty f_{V_T|T=x}(v|T=x) f_T(x) dx = \int_0^\infty f_{V_T|T=x}(v|T=x) \cdot \Big\{ -\frac{d}{dx} S_T(x) \Big\} dx \\ &= \int_0^\infty \frac{v^{\alpha H_x - 1} e^{-v/\beta}}{\Gamma(\alpha H_x)\beta^{\alpha H_x}} \cdot \alpha h_x \ln(1+\beta)(1+\beta)^{-\alpha H_x} dx \\ &= \alpha \ln(1+\beta) e^{-v/\beta} \int_0^\infty \frac{v^{\alpha H_x - 1}}{\Gamma(\alpha H_x)\beta^{\alpha H_x}} h_x (1+\beta)^{-\alpha H_x} dx \end{aligned} \tag{C.2}$$

Using substitution with $u = H_x$ and $du = h_x dx = dH_x$, we have

$$\begin{aligned} f_V(v) &= \alpha \ln(1+\beta) e^{-v/\beta} \int_0^\infty \frac{v^{\alpha u - 1}}{\Gamma(\alpha u)\beta^{\alpha u}} (1+\beta)^{-\alpha u} du \\ &= \alpha \ln(1+\beta) e^{-v/\beta} \int_0^\infty \frac{v^{\alpha u - 1}}{\Gamma(\alpha u)[\beta(1+\beta)]^{\alpha u}} du \\ &= \frac{\alpha \ln(1+\beta) e^{-v/\beta}}{\beta(1+\beta)} \int_0^\infty \frac{(v/\beta(1+\beta))^{\alpha u - 1}}{\Gamma(\alpha u)} du \end{aligned} \tag{C.3}$$

Focusing on the integral term, another substitution with $g = \alpha u$, $dg = \alpha du \Longrightarrow du = \frac{dg}{\alpha}$, and $s = \frac{v}{\beta(1+\beta)}$ results in

$$\begin{aligned} \int_0^\infty \frac{s^{g-1}}{\Gamma(g)} \frac{dg}{\alpha} &= \frac{1}{\alpha} \int_0^\infty \frac{s^{g-1}}{\Gamma(g)} dg = \frac{1}{\alpha} \int_0^\infty \frac{g}{g} \cdot \frac{s^{g-1}}{\Gamma(g)} dg = \frac{1}{\alpha} \int_0^\infty \frac{g s^{g-1}}{g\Gamma(g)} dg \\ &= \frac{1}{\alpha s} \int_0^\infty \frac{g s^g}{\Gamma(2)\Gamma(g+1)} dg \\ &= \frac{1}{\alpha s} \int_0^\infty \frac{g^1 s^g}{\Gamma(1+1)\Gamma(g+1)} dg \end{aligned} \tag{C.4}$$

The final integral expression in (C.4) can be recognized as the integral function $\mu(s,1)$ from Gradshteyn and Ryzhik (2014). Thus, we can write the marginal distribution of the marker in the simplified form

$$f_V(v) = \frac{\alpha \ln(1+\beta) e^{-v/\beta}}{\beta(1+\beta)} \cdot \frac{\mu\big(\frac{v}{\beta(1+\beta)}, 1\big)}{\alpha \frac{v}{\beta(1+\beta)}} = \frac{e^{-v/\beta}\ln(1+\beta)}{v} \mu\bigg(\frac{v}{\beta(1+\beta)}, 1\bigg), \quad v > 0 \tag{C.5}$$

For the distribution of the event time conditional on the marker value, we can write it in terms of the bridge process $B\{H_t\}$ assuming $x > t$. Note that an equivalent expression could be derived assuming $x < t$. This results in

$$f(t|V_x)|_{x=t} = \bigg[-\frac{d}{dt} E\big\{ e^{-B\{H_t\}V(H_x)} \big| V_x \big\} \bigg]_{x=t} \tag{C.6}$$

This expression will have a complicated form as it involves the derivative of a bridge stochastic process with a time transformation. However, we can derive the explicit form of this conditional distribution by applying Bayes rule:

$$f(t|V_x)|_{x=t} = \frac{f(t, V_t)}{f(V_t)} = \frac{f(V_t|T=t)f(t)}{f(V_t)} \tag{C.7}$$

The numerator in (C.7) will be the product of a Gamma density, by the assumption that the marker process at any fixed meaurement time is Gamma distributed, and the derivative of the Laplace transform of a Gamma random variable with respect to t. The denominator is the marginal distribution of the marker given in (C.5). Specifically,

$$f(t|V_x)|_{x=t} = \frac{\frac{v^{\alpha H_t - 1} e^{-v/\beta}}{\Gamma(\alpha H_t)\beta^{\alpha H_t}} \alpha h_t \ln(1+\beta)(1+\beta)^{-\alpha H_t}}{\frac{e^{-v/\beta}\ln(1+\beta)}{v} \mu\big(\frac{v}{\beta(1+\beta)}, 1\big)} = \frac{\frac{v^{\alpha H_t - 1}}{\Gamma(\alpha H_t)\beta^{\alpha H_t}} \alpha h_t (1+\beta)^{-\alpha H_t}}{\frac{1}{v}\mu\big(\frac{v}{\beta(1+\beta)}, 1\big)} \tag{C.8}$$

Therefore, the joint model of interest is specified as

$$f(t, V_t) = \frac{v^{\alpha H_t - 1} e^{-v/\beta}}{\Gamma(\alpha H_t)\beta^{\alpha H_t}} \alpha h_t \ln(1+\beta)(1+\beta)^{-\alpha H_t} \tag{C.9}$$

The terms that contribute to the log-likelihood in Chapter IV, Section 4.3.2, equation (4.18) are

$$\begin{aligned} f(V_{H_t}|T=t) &= \frac{v^{\alpha H_t - 1} e^{-v/\beta}}{\Gamma(\alpha H_t)\beta^{\alpha H_t}} \\ f(t) &= \alpha h_t \ln(1+\beta)(1+\beta)^{-\alpha H_t} \\ S(t) &= (1+\beta)^{-\alpha H_t} \end{aligned}$$

C.2 Marker Observed at Uninformative Measurement Times

C.2.1 Single Measurement Time

To calculate the survival function $S(t)$ conditional on an observed marker value V_{H_τ}, we must consider how τ relates to t when applying the definition in Chapter IV, Section 4.2.2.1, equation (4.7). There are two possibilities: $\tau \leq t$ and $\tau > t > 0$. Let us consider the first case. If $\tau \leq t$, then we will have

$$\begin{aligned} S(t|V_{H_\tau}) &= E_{\bar{W}}\left\{S(t|\bar{W})\middle|V_{H_\tau}\right\} = E_{\bar{W}}\left\{e^{-(V_{H_\tau}+W_{H_t-H_\tau})}\middle|V_{H_\tau}\right\} \\ &= e^{-V_{H_\tau}}\mathcal{L}_W(1) \\ &= e^{-V_{H_\tau}}(1+\beta)^{-\alpha(H_t-H_\tau)} \end{aligned} \tag{C.10}$$

as $\bar{W}|V_{H_\tau} = V_{H_\tau} + W_{H_t-H_\tau}$. The hazard conditional on the marker can be found by

applying the definition

$$\begin{aligned} d\Lambda(t|V_{H_\tau}) &= -\frac{d}{dt}\ln S(t|V_{H_\tau})dt \\ &= -\frac{d}{dt}\ln\left[e^{-V_{H_\tau}}(1+\beta)^{-\alpha(H_t-H_\tau)}\right]dt \\ &= \frac{d}{dt}\left[V_{H_\tau}+\alpha(H_t-H_\tau)\ln(1+\beta)\right]dt \\ &= dH_t\alpha\ln(1+\beta) \end{aligned} \tag{C.11}$$

In the case where $0 < t < \tau$, we have $\bar{W}|V_{H_\tau} = B\{H_t\}V_{H_\tau}$, where B represents the scaled Gamma bridge process. Then the conditional survival function will be

$$\begin{aligned} S(t|V_{H_\tau}) &= E_{\bar{W}}\left\{S(t|\bar{W})\middle|V_{H_\tau}\right\} = E_B\left\{e^{-B\{H_t\}V_{H_\tau}}\middle|V_{H_\tau}\right\} = \mathcal{L}_B(V_{H_\tau}) \\ &= M(mH_t, mH_\tau, -V_{H_\tau}) \end{aligned} \tag{C.12}$$

Once again, when $0 \leq t < \tau$, the hazard conditional on the marker can be found by applying the definition

$$\begin{aligned} d\Lambda(t|V_{H_\tau}) &= -\frac{d}{dt}\ln S(t|V_{H_\tau})dt \\ &= -\frac{d}{dt}\ln M\big(mH_t, mH_\tau, -V_{H_\tau}\big)dt \\ &= -\frac{1}{M\big(mH_t, mH_\tau, -V_{H_\tau}\big)}\cdot\frac{d}{dt}M\big(mH_t, mH_\tau, -V_{H_\tau}\big)dt \end{aligned} \tag{C.13}$$

Focusing on the derivative, we have

$$\begin{aligned} \frac{d}{dt}M\big(mH_t, mH_\tau, -V_{H_\tau}\big) &= \frac{d}{dt}\left\{1+\sum_{k=1}^{\infty}\left[\prod_{r=1}^{k}\frac{mH_t+r-1}{mH_\tau+r-1}\right]\frac{(-1)^kV_{H_\tau}^k}{k!}\right\} \\ &= \frac{d}{dt}\sum_{k=1}^{\infty}\frac{(mH_t)_{(k)}}{(mH_\tau)_{(k)}}\frac{(-1)^kV_{H_\tau}^k}{k!} \end{aligned} \tag{C.14}$$

where $(x)_{(n)} = \frac{\Gamma(x+n)}{\Gamma(x)}$ is the Pochhammer symbol. Then,

$$\begin{aligned} \frac{d}{dt}M\big(mH_t, mH_\tau, -V_{H_\tau}\big) &= \frac{d}{dt}\sum_{k=1}^{\infty}\frac{\Gamma(mH_t+k)}{\Gamma(mH_t)}\frac{\Gamma(mH_\tau)}{\Gamma(mH_\tau+k)}\frac{(-1)^kV_{H_\tau}^k}{k!} \\ &= \frac{d}{dt}\sum_{k=1}^{\infty}\frac{1}{B(mH_t,k)}\frac{\Gamma(mH_\tau)}{\Gamma(mH_\tau+k)}\frac{(-1)^kV_{H_\tau}^k}{k} \end{aligned} \tag{C.15}$$

where $B(mH_t, k) = \frac{\Gamma(mH_t)\Gamma(k)}{\Gamma(mH_t+k)}$ and $\Gamma(k+1) = k! = k\Gamma(k)$ when k is an integer. Therefore, applying the quotient and chain rules for differentiation results in

$$
\begin{aligned}
&\frac{d}{dt}\sum_{k=1}^{\infty}\frac{1}{B(mH_t,k)}\frac{\Gamma(mH_\tau)}{\Gamma(mH_\tau+k)}\frac{(-1)^k V_{H_\tau}^k}{k} \\
&=\sum_{k=1}^{\infty}\frac{d}{dt}\left[\frac{1}{B(mH_t,k)}\right]\frac{\Gamma(mH_\tau)}{\Gamma(mH_\tau+k)}\frac{(-1)^k V_{H_\tau}^k}{k} \\
&\quad+\frac{1}{B(mH_t,k)}\frac{d}{dt}\left[\frac{\Gamma(mH_\tau)}{\Gamma(mH_\tau+k)}\right]\frac{(-1)^k V_{H_\tau}^k}{k} \\
&\quad+\frac{1}{B(mH_t,k)}\frac{\Gamma(mH_\tau)}{\Gamma(mH_\tau+k)}\frac{d}{dt}\left[\frac{(-1)^k V_{H_\tau}^k}{k}\right] \\
&=\sum_{k=1}^{\infty}\frac{mh_t}{B(mH_t,k)}\big\{\psi(mH_t+k)-\psi(mH_t)\big\}\frac{\Gamma(mH_\tau)}{\Gamma(mH_\tau+k)}\frac{(-1)^k V_{H_\tau}^k}{k} \\
&\quad+\frac{1}{B(mH_t,k)}\frac{\Gamma(mH_\tau)}{\Gamma(mH_\tau+k)}mh_t\big\{\psi(mH_\tau)-\psi(mH_\tau+k)\big\}\frac{(-1)^k V_{H_\tau}^k}{k} \\
&\quad+\frac{1}{B(mH_t,k)}\frac{\Gamma(mH_\tau)}{\Gamma(mH_\tau+k)}(-1)^k V_{H_\tau}^{k-1}V'_{H_\tau}h_t \\
&=\sum_{k=1}^{\infty}\frac{1}{B(mH_t,k)}\frac{\Gamma(mH_\tau)}{\Gamma(mH_\tau+k)}\frac{(-1)^k V_{H_\tau}^k}{k} \\
&\quad\times\left[mh_t\big\{\psi(mH_t+k)-\psi(mH_t)+\psi(mH_\tau)-\psi(mH_\tau+k)\big\}+\frac{d}{dt}\ln(V_{H_\tau})k\right]
\end{aligned}
\tag{C.16}
$$

where $h_t = \frac{d}{dt}H_t$ and $\psi(x) = \frac{d}{dx}\ln\Gamma(x)$ is the digamma function. Note that because $\tau > t$ and we consider our marker on the H-transformed time scale, there are additional terms capturing the fact that H_τ contains information relevant to time t. Thus, the hazard conditional on the observed marker value is

$$
\begin{aligned}
d\Lambda(t|V_{H_\tau}) &= -\frac{1}{M\big(mH_t, mH_\tau, -V_{H_\tau}\big)}\frac{d}{dt}M\big(mH_t, mH_\tau, -V_{H_\tau}\big)dt \\
&= -\frac{\sum_{k=1}^{\infty}\frac{1}{B(mH_t,k)}\frac{\Gamma(mH_\tau)}{\Gamma(mH_\tau+k)}\frac{(-1)^k V_{H_\tau}^k}{k}\left[mh_t\big\{\psi(mH_t+k)-\psi(mH_t)+\psi(mH_\tau)-\psi(mH_\tau+k)\big\}+\frac{d}{dt}\ln(V_{H_\tau})k\right]}{M\big(mH_t, mH_\tau, -V_{H_\tau}\big)}dt
\end{aligned}
\tag{C.17}
$$

Therefore, the functions needed to maximize the log-likelihood in Chapter IV, Section

4.3.3, equation (4.24) when only a single marker measure is available are

$$
\begin{aligned}
S(t|V_{H_\tau}) &= \begin{cases} M(mH_t, mH_\tau, -V_{H_\tau}), & 0 \le t < \tau \\ e^{-V_{H_\tau}}(1+\beta)^{-m(H_t - H_\tau)}, & t \ge \tau \end{cases} \\
d\Lambda(t|V_{H_\tau}) &= \begin{cases} -\frac{\frac{d}{dt}M\left(mH_t, mH_\tau, -V_{H_\tau}\right)dt}{M\left(mH_t, mH_\tau, -V_{H_\tau}\right)}, & 0 \le t < \tau \\ dH_t m \ln(1+\beta), & t \ge \tau \end{cases} \\
g(V_{H_\tau})(v) &= \frac{v^{mH_t - 1}e^{-v/\beta}}{\Gamma(mH_t)\beta^{mH_t}} \\
S(t) &= (1+\beta)^{-mH_t} \\
d\Lambda(t) &= dH_t m \ln(1+\beta)
\end{aligned}
$$

C.2.2 Multiple Measurement Times

When the marker is observed at multiple measurement times, we can use the same ideas as in (C.10)-(C.12). Let $\tau = \{\tau_1, \ldots, \tau_k\}$ be the set of measurement times, which are assumed to be uninformative. The corresponding set of observed marker values can be expressed as $\mathbf{V} = \{V_{H_{\tau_1}}, \ldots, V_{H_{\tau_k}}\}$. The goal is to derive the survival function $S(t)$ conditional on $\mathbf{V}$. Once more, there are two cases to consider: $t \ge \tau_k$ and $0 \le \tau_1 \le \cdots \le \tau_{k-1} \le t < \tau_k$. In the latter, we assume, without loss of generality, that the time of survival prediction t is between the last two measurement times. This assumption can be relaxed further by having t between any two consecutive measurement times.

In the first case where $t \ge \tau_k$, we have

$$
\begin{aligned}
S(t|\mathbf{V}) &= E_{\bar{W}}\left\{S(t|\bar{W})\middle|\mathbf{V}\right\} = E_{\bar{W}}\left\{e^{-(V_{H_{\tau_k}} + W_{H_t - H_{\tau_k}})}\middle|\mathbf{V}\right\} = e^{-V_{H_{\tau_k}}}\mathcal{L}_W(1) \\
&= e^{-V_{H_{\tau_k}}}(1+\beta)^{-\alpha(H_t - H_{\tau_k})}
\end{aligned}
\tag{C.18}
$$

as $\bar{W}|\mathbf{V} = V_{H_{\tau_k}} + W_{H_t - H_{\tau_k}}$, similar to (C.10). When $\tau_{k-1} \le t < \tau_k$, we assume that $\bar{W}|\mathbf{V} = V_{H_{\tau_{k-1}}} + B\{H_t\}(V_{H_{\tau_k}} - V_{H_{\tau_{k-1}}})$. Then, the Laplace transform of the Lévy process can once more be used to obtain the survival function conditional on observed

data.

$$
\begin{aligned}
S(t|\mathbf{V}) &= E_{\bar{W}}\big\{S(t|\bar{W})\big|\mathbf{V}\big\} \\
&= E_B\big\{e^{-(V_{H_{\tau_{k-1}}}+B(V_{H_{\tau_k}}-V_{H_{\tau_{k-1}}}))}\big|\mathbf{V}\big\} \\
&= e^{-V_{H_{\tau_{k-1}}}}\mathcal{L}_B(V_{H_{\tau_k}}-V_{H_{\tau_{k-1}}}) \\
&= e^{-V_{H_{\tau_{k-1}}}}M\big(m(H_t-H_{\tau_{k-1}}),m(H_{\tau_k}-H_{\tau_{k-1}}),-(V_{H_{\tau_k}}-V_{H_{\tau_{k-1}}})\big)
\end{aligned}
\tag{C.19}
$$

The hazard conditional on the observed marker can be generalized from (C.11) and (C.17):

$$
d\Lambda(t|\mathbf{V}) = \begin{cases} dH_t m\ln(1+\beta), & t\geq\tau_k \\ \frac{-\frac{d}{dt}M\left(m(H_t-H_{\tau_{k-1}}),m(H_{\tau_k}-H_{\tau_{k-1}}),-(V_{H_{\tau_k}}-V_{H_{\tau_{k-1}}})\right)}{M\left(m(H_t-H_{\tau_{k-1}}),m(H_{\tau_k}-H_{\tau_{k-1}}),-(V_{H_{\tau_k}}-V_{H_{\tau_{k-1}}})\right)}dt, & \tau_{k-1}\leq t<\tau_k \end{cases}
\tag{C.20}
$$

where

$$
\begin{aligned}
&\frac{d}{dt}M\big(m(H_t-H_{\tau_{k-1}}),m(H_{\tau_k}-H_{\tau_{k-1}}),-(V_{H_{\tau_k}}-V_{H_{\tau_{k-1}}})\big) \\
=&\sum_{k=1}^{\infty}\frac{1}{B(m(H_t-H_{\tau_{k-1}}),k)}\frac{\Gamma(m(H_{\tau_k}-H_{\tau_{k-1}}))}{\Gamma(m(H_{\tau_k}-H_{\tau_{k-1}})+k)}\frac{(-1)^k(V_{H_{\tau_k}}-V_{H_{\tau_{k-1}}})^k}{k} \\
&\times\Big[mh_t\big\{\psi(m(H_t-H_{\tau_{k-1}})+k)-\psi(m(H_t-H_{\tau_{k-1}}))+\psi(m(H_{\tau_k}-H_{\tau_{k-1}}))-\psi(m(H_{\tau_k}-H_{\tau_{k-1}})+k)\big\} \\
&+\frac{d}{dt}\ln(V_{H_{\tau_k}}-V_{H_{\tau_{k-1}}})k\Big]
\end{aligned}
$$

following an argument similar to the one in (C.16).

With multiple measurements, the contributions to the log-likelihood in Chapter IV, Section 4.3.3, equation (4.24) that involve $\mathbf{V}$ are

$$
S(t|\mathbf{V}) = \begin{cases} e^{-V_{H_{\tau_{k-1}}}}M\big(m(H_t-H_{\tau_{k-1}}),m(H_{\tau_k}-H_{\tau_{k-1}}),-(V_{H_{\tau_k}}-V_{H_{\tau_{k-1}}})\big), & \tau_{k-1}\leq t<\tau_k \\ e^{-V_{H_{\tau_k}}}(1+\beta)^{-m(H_t-H_{\tau_k})}, & t\geq\tau_k \end{cases}
$$

$$
d\Lambda(t|\mathbf{V}) = \begin{cases} \frac{-\frac{d}{dt}M\left(m(H_t-H_{\tau_{k-1}}),m(H_{\tau_k}-H_{\tau_{k-1}}),-(V_{H_{\tau_k}}-V_{H_{\tau_{k-1}}})\right)}{M\left(m(H_t-H_{\tau_{k-1}}),m(H_{\tau_k}-H_{\tau_{k-1}}),-(V_{H_{\tau_k}}-V_{H_{\tau_{k-1}}})\right)}dt, & \tau_{k-1}\leq t<\tau_k \\ dH_t m\ln(1+\beta), & t\geq\tau_k \end{cases}
$$

$$
g(\mathbf{V}) = \prod_{j=1}^{k}\frac{(V_{H_{\tau_j}}-V_{H_{\tau_{j-1}}})^{m(H_{\tau_j}-H_{\tau_{j-1}})-1}e^{-(V_{H_{\tau_j}}-V_{H_{\tau_{j-1}}})/\beta}}{\Gamma\big(m(H_{\tau_j}-H_{\tau_{j-1}})\big)\beta^{m(H_{\tau_j}-H_{\tau_{j-1}})}}
$$

C.3 Details of Simulations

The setup of our simulations in the case of marked survival follows what is reported in Suresh (2018) with some modification:

1. Define the mean and variance governing the true marker process $W_{H(t)}$ on the H-transformed time scale as $\mu H(t)$ and $\sigma^2 H(t)$, respectively, where $\mu = e^{\boldsymbol{\nu} Z}$ and $\sigma^2 = e^{\boldsymbol{\eta} Z}$. We specify $\boldsymbol{\nu} = \nu_0 + \nu_1$ and $\boldsymbol{\eta} = \eta_0 + \eta_1$. Z represents a single binary covariate that is Bernoulli distributed with probability 0.3. The true parameter values are $\boldsymbol{\nu} = (0, 0.3)$ and $\boldsymbol{\eta} = (-2.1, 0.5)$, where ν_0 is restricted to 0 for identifiability. $H(t)$ represents the cumulative baseline hazard, which will be assumed to be Weibull with $\lambda = 1$ and $k = 2$, i.e., $H(t) = t^2$. Then, $W_{H(t)}$ will be assumed to be follow a Gamma distribution with shape $\alpha H(t)$ and scale β, where $\alpha = \frac{\mu^2}{\sigma^2}$ and $\beta = \frac{\sigma^2}{\mu}$.

2. First generate the event time T using the marginal survival function via the probability integral transform by generating a random variable $u \sim$ Uniform(0,1) and solving $S(t) = (1 + \beta)^{-\alpha H(t)} = u$ for t.

3. Generate the observed marker value V measured at T by simulating it from a Gamma distribution with shape $\alpha H(T)$ and scale β.

4. Generate administrative censoring time $C \sim$ Uniform$(0, c)$, and calculate observed event time $X = \min(T, C)$ and event indicator $\Delta = I(T \leq C)$

5. Calculate the log-likelihood as in Chapter IV, Section 4.3.2, equation (4.18). When the parametric Weibull function is used, $H_t = \left(\frac{t}{\lambda}\right)^k$, where the maximum likelihood estimators of λ and k are plugged in. The parameters $\boldsymbol{\nu}$, $\boldsymbol{\eta}$, λ, and k will be estimated using standard maximum likelihood estimation with a parametric $H(t)$. Alternatively, we can use the Breslow estimator in Chapter IV, Section 4.3.4, equation (4.25) to approximate the cumulative baseline hazard

function $H(t)$. When H is estimated using the Breslow estimator, $\boldsymbol{\nu}$ and $\boldsymbol{\eta}$ will be estimated using an iterative quasi-profile maximum likelihood algorithm.

C.4 Derivation of Cross-Sectional Form of Likelihood

Recall that the log-likelihood for a standard survival model assuming longitudinal follow-up for each of the n subjects is given by

$$l = \sum_{i=1}^{n} \Delta_i \ln\left(d\Lambda_T(X_i|\boldsymbol{Z}_i)\right) - \Lambda_T(X_i|\boldsymbol{Z}_i) \tag{C.21}$$

Rewriting the log-likelihood in (C.21) in counting process notation gives

$$l = \int_0^{\xi} \sum_{i=1}^{n} dN_i(u) \ln\left(d\Lambda_T(u|\boldsymbol{Z}_i)\right) - Y_i(u) d\Lambda_T(u|\boldsymbol{Z}_i) \tag{C.22}$$

where $dN_i(u) = I(X_i = u, \Delta_i = 1)$ and $Y_i = I(X_i \geq u)$. The log-likelihoods in both (C.21) and (C.22) assume that longitudinal information is available for each subject in the risk set at time u. In reality, this may not be the case. Instead, life-table risk set information may only be available, wherein the number of subjects with covariate value $\boldsymbol{z}$ at risk at time u is known. In such cases, the log-likelihood in (C.22) can be written as

$$l = \int_0^{\xi} \sum_{i=1}^{n} dN_i(u) \ln\left(d\Lambda_T(u|\boldsymbol{Z}_i)\right) - \sum_{\boldsymbol{z}\in\mathcal{Z}} Y_{\boldsymbol{z}}(u) d\Lambda_T(u|\boldsymbol{z}) \tag{C.23}$$

The first sum in (C.23) is taken over the subjects with event information, and the second sum is over the unique covariate values. $Y_{\boldsymbol{z}}(u)$ is a count of the number of subjects with covariate value $\boldsymbol{z}$ in the set of unique covariate values $\mathcal{Z}$ who are at risk at time u.

Using the form of (C.23), it is straightforward to see how the log-likelihood in (4.18) can be written to conform with cross-sectional risk information.

C.5 Additional SEER Analysis Results

C.5.1 Results for Year of Diagnosis 2011

Observed survival curves summarized using the Kaplan-Meier method for time to prostate cancer diagnosis stratified by race and the distribution of PSA values for the year 2011 are displayed in Figures C.1 and C.2, respectively. Black men tend to be diagnosed earlier than to White men. The marginal distribution of PSA values in Figure C.2 is right-skewed with mean and median values of 9.0 ng/mL and 6.0 ng/mL, respectively.

Estimated marker parameters obtained from the proposed joint model are given in Table C.1. Estimated and observed survival curves are displayed in Figure C.3. The fitted curve for White men is nearly identical to the observed survival curve. For Black men, there is some slight deviation at the tail, but overall, the estimated curve accurately describes the shape of the observed curve. Based on a likelihood ratio test of the hypothesis $H_0 : \nu_1 = 0, \eta_1 = 0$, we reject the null hypothesis ($p < 0.001$) and conclude that there is enough evidence to suggest a significant difference in the parameters governing the underlying PSA process between Black and White men.

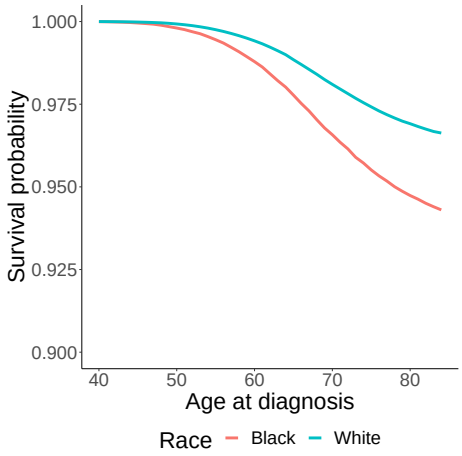


Figure C.1: Observed age of prostate cancer diagnosis by race, 2011.

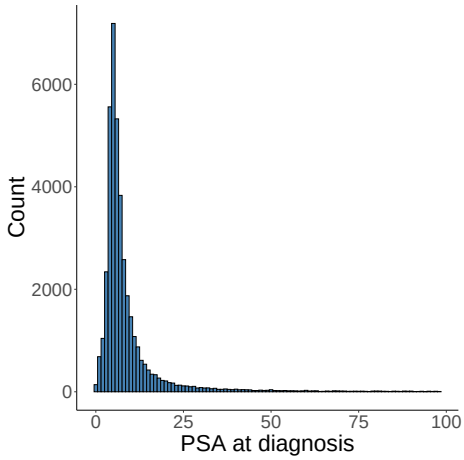


Figure C.2: Distribution of PSA at age of prostate cancer diagnosis, 2011.

Table C.1: Marker parameter estimates in 2011. 95% Wald interval constructed using bootstrap standard error. p-value estimated using Wald Z value $= \frac{\text{Est}}{\text{Bootstrap SE}}$.

Parameter	**Estimate**	**Bootstrap SE**	**95% CI**	p
ν_0	2.799	0.007	$(2.784, 2.813)$	< 0.001
ν_1	0.715	0.022	$(0.672, 0.758)$	< 0.001
η_0	6.222	0.017	$(6.189, 6.255)$	< 0.001
η_1	0.922	0.042	$(0.840, 1.004)$	< 0.001

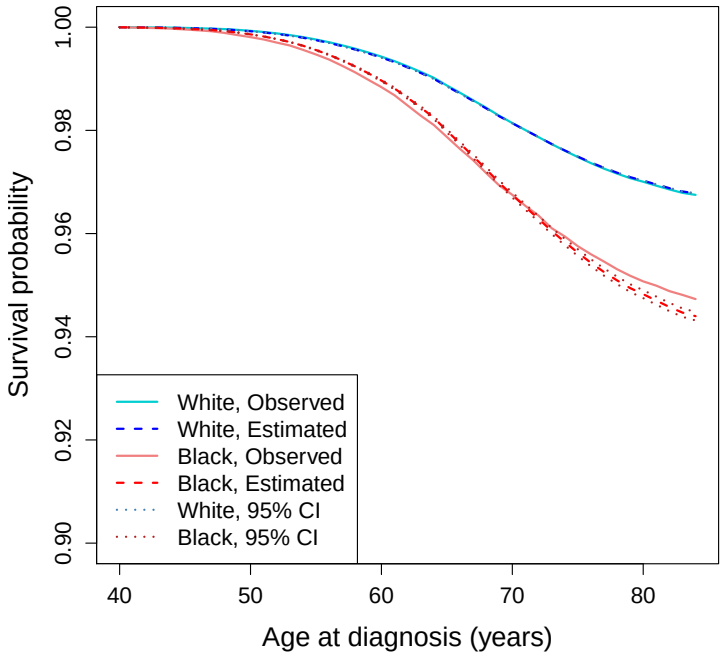


Figure C.3: Observed (solid) and estimated (dashed) survival curves for time to prostate cancer diagnosis for men diagnosed in 2011. 95% Wald-based confidence bands (dotted) constructed using bootstrap standard errors.

C.5.2 Results for Year of Diagnosis 2012

Observed survival curves summarized using the Kaplan-Meier method for time to prostate cancer diagnosis stratified by race and the distribution of PSA values for the year 2012 are displayed in Figures C.4 and C.5, respectively. Black men are diagnosed earlier compared to White men. The marginal distribution of PSA values in Figure C.5 is right-skewed with mean and median values of 9.7 ng/mL and 6.3 ng/mL, respectively.

Estimated marker parameters obtained from the proposed joint model are given in Table C.2. Estimated and observed survival curves are displayed in Figure C.6. The fitted curve for White men is nearly identical to the observed survival curve. For Black men, there is noticeable deviation at the tail, but overall, the estimated curve

accurately describes the shape of the empirical curve. Based on a likelihood ratio test of the hypothesis $H_0 : \nu_1 = 0, \eta_1 = 0$, we reject the null hypothesis ($p < 0.001$) and conclude that there is enough evidence to suggest a significant difference in the parameters governing the underlying PSA process between Black and White men.

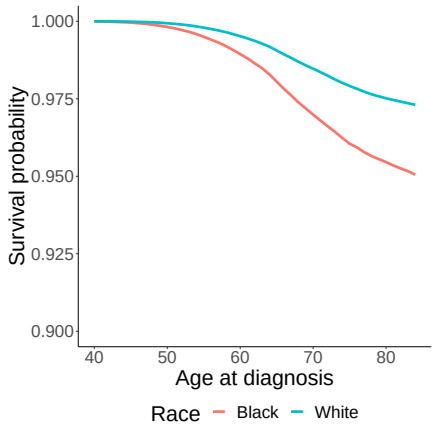


Figure C.4: Observed age of prostate cancer diagnosis by race, 2012.

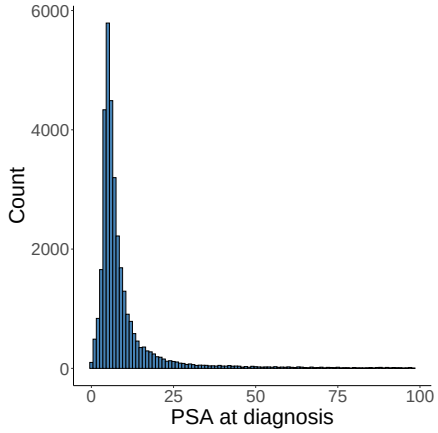


Figure C.5: Distribution of PSA at age of prostate cancer diagnosis, 2012.

Table C.2: Marker parameter estimates in 2012. 95% Wald interval constructed using bootstrap standard error. p-value estimated using Wald Z value $= \frac{\text{Est}}{\text{Bootstrap SE}}$.

Parameter	**Estimate**	**Bootstrap SE**	**95% CI**	p
ν_0	2.853	0.010	$(2.834, 2.872)$	< 0.001
ν_1	0.802	0.036	$(0.732, 0.872)$	< 0.001
η_0	6.370	0.021	$(6.329, 6.411)$	< 0.001
η_1	1.023	0.074	$(0.877, 1.169)$	< 0.001

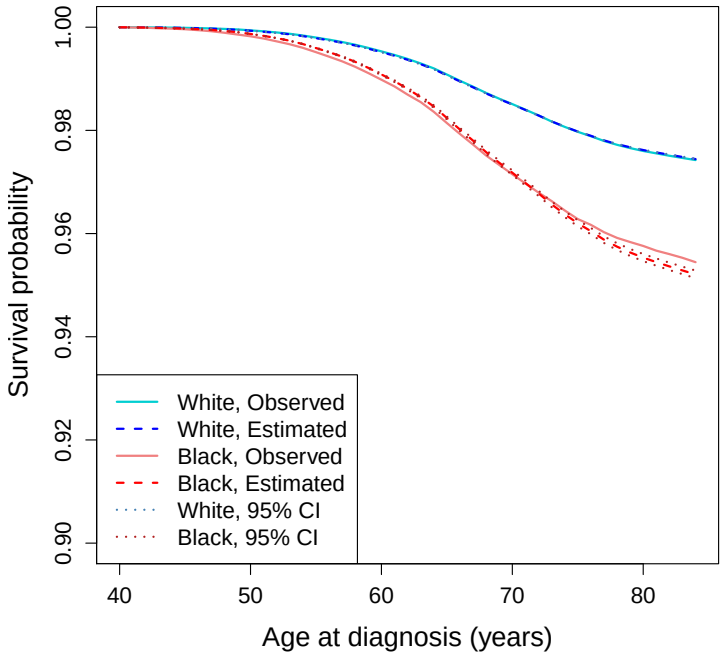


Figure C.6: Observed (solid) and estimated (dashed) survival curves for time to prostate cancer diagnosis for men diagnosed in 2012. 95% Wald-based confidence bands (dotted) constructed using bootstrap standard errors.

C.5.3 Results for Year of Diagnosis 2013

Observed survival curves summarized using the Kaplan-Meier method for time to prostate cancer diagnosis stratified by race and the distribution of PSA values for the year 2013 are displayed in Figures C.7 and C.8, respectively. Black men are diagnosed earlier than White men as evidenced by the observed survival curves. The marginal distribution of PSA values in Figure C.8 is right-skewed with mean and median values of 9.9 ng/mL and 6.4 ng/mL, respectively.

Estimated marker parameters obtained from the proposed joint model are given in Table C.3. Estimated and observed survival curves are displayed in Figure C.9. The fitted curve for White men is nearly identical to the observed survival curve. For Black men, there is slight deviation at the tail of the survival curve, but overall, the

estimated curve accurately describes the shape of the empirical curve. Based on a likelihood ratio test of the hypothesis $H_0 : \nu_1 = 0, \eta_1 = 0$, we reject the null hypothesis ($p < 0.001$) and conclude that there is enough evidence to suggest a significant difference in the parameters governing the underlying PSA process between Black and White men.

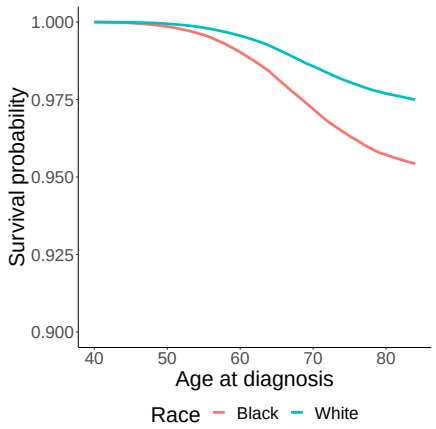


Figure C.7: Observed age of prostate cancer diagnosis by race, 2013.

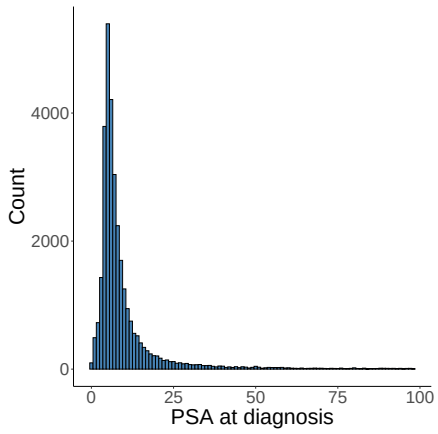


Figure C.8: Distribution of PSA at age of prostate cancer diagnosis, 2013.

Table C.3: Marker parameter estimates in 2013. 95% Wald interval constructed using bootstrap standard error. p-value estimated using Wald Z value $= \frac{\text{Est}}{\text{Bootstrap SE}}$.

Parameter	**Estimate**	**Bootstrap SE**	**95% CI**	p
ν_0	2.880	0.008	$(2.864, 2.896)$	< 0.001
ν_1	0.786	0.021	$(0.745, 0.828)$	< 0.001
η_0	6.436	0.018	$(6.401, 6.470)$	< 0.001
η_1	0.988	0.040	$(0.910, 1.066)$	< 0.001

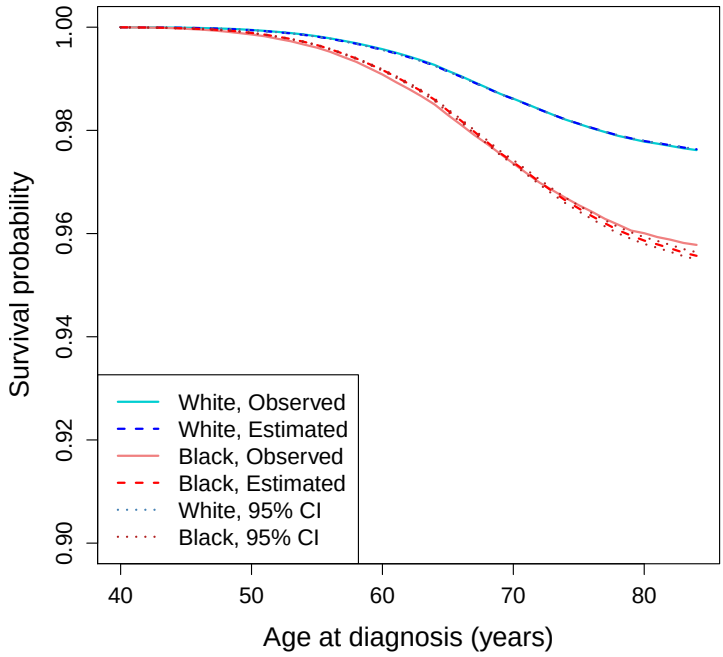


Figure C.9: Observed (solid) and estimated (dashed) survival curves for time to prostate cancer diagnosis for men diagnosed in 2013. 95% Wald-based confidence bands (dotted) constructed using bootstrap standard errors.

C.5.4 Comparison of SEER Parameter Estimates

Plots of the marker parameter estimates by year of diagnosis are displayed in Figure C.10. Estimates of all parameters except for η_1 tend to increase with year, though estimates for the year 2011 are noticeably lower than those of other years. Estimates for the intercept term η_0 related to the variance of the marker process were the most variable from year to year. Overall, the point estimates and bootstrap standard errors are similar across years.

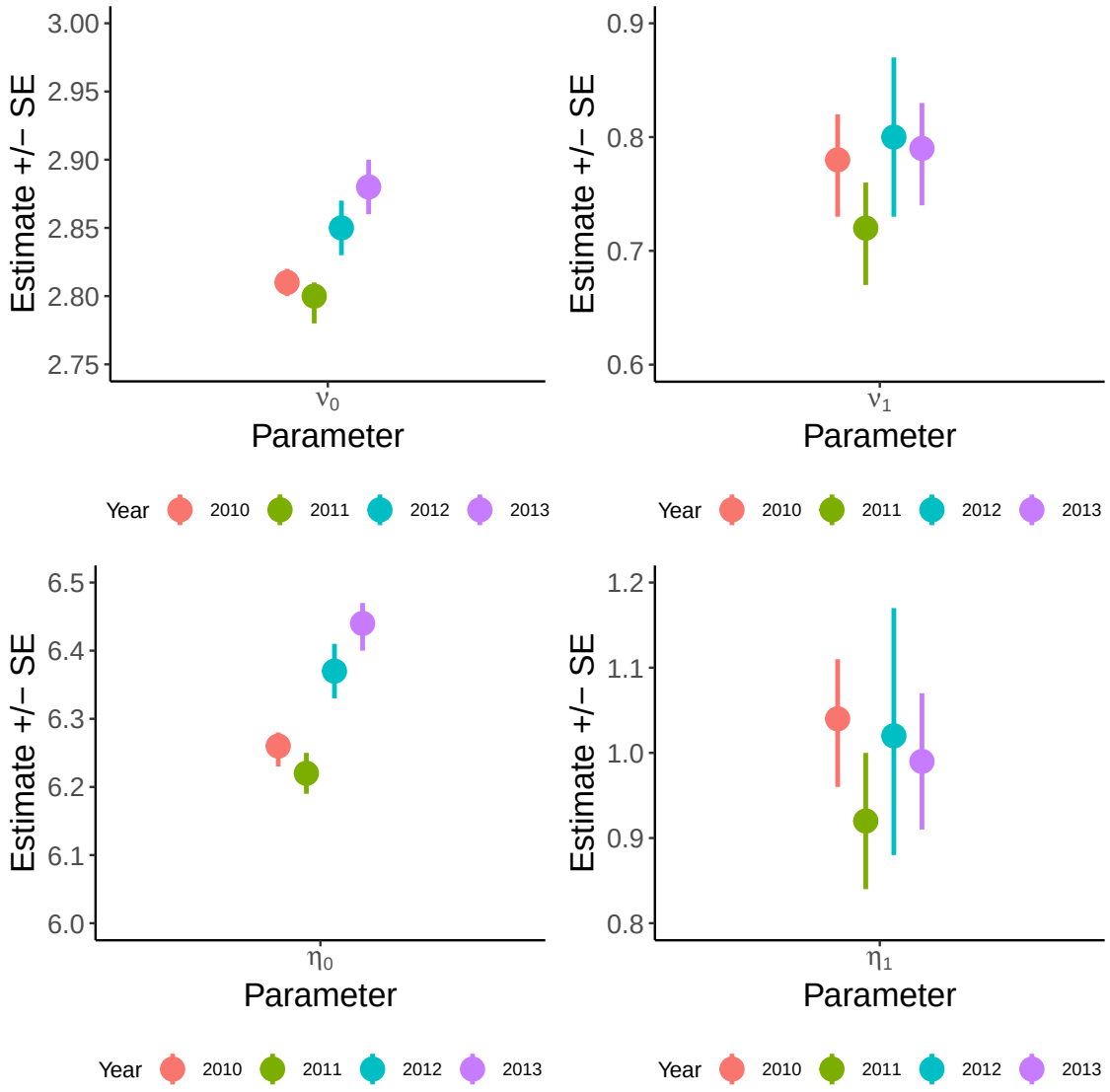


Figure C.10: Comparison of estimated parameters and 95% bootstrapped confidence interval by year of diagnosis.

BIBLIOGRAPHY

BIBLIOGRAPHY

Aalen, O. (1992), Modelling heterogeneity in survival analysis by the compound poisson distribution, *The Annals of Applied Probability*, pp. 951–972.

Aalen, O. (1994), Effects of frailty in survival analysis, *Statistical Methods in Medical Research*, pp. 227–243.

Aalen, O., Ø. Borgan, and H. Gjessing (2008), *Survival and Event History Analysis: A Process Point of View*, Springer, New York.

Adamo, M., J. Boten, L. Coyle, K. Cronin, C. Lam, S. Negoita, L. Penberthy, J. Stevens, and K. Ward (2016), Validation of prostate-specific antigen laboratory values recorded in surveillance, epidemiology, and end results registries, *Cancer*, *123*, 697–703.

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

Arbeev, K., I. Akushevich, A. Kulminski, S. Ukraintseva, and A. Yashin (2014), Joint analyses of longitudinal and time-to-event data in research on aging: implications for predicting health and survival, *Frontiers in Public Health*, *2*, 1–12.

Arnold, B., E. Castillo, and J. Sarabia (1999), *Conditional Specification of Statistical Models*, Springer, New York.

Aven, T., and U. Jensen (1999), *Stochastic Models in Reliability*, Springer, New York.

Bertoin, J. (1998), *Lévy Processes*, *Cambridge Tracts in Mathematics*, vol. 121, Cambridge University Press, Cambridge.

Breslow, N. (1972), Discussion on *Regression models and life tables*, *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*, 216–217.

Brody, D., L. Hughston, and A. Macrina (2008), Dam rain and cumulative gain, *Proceedings of the Royal Society of London A: Mathematical, Physical, and Engineering Sciences*, *464*, 1801–1822.

Chen, K., Z. Jin, and Z. Yin (2002), Semiparametric analysis of transformation models with censored data, *Biometrika*, *89*, 659–668.

Chen, Y. (2009), Weighted breslow-type and maximum likelihood estimation in semiparametric transformation models, *Biometrika*, *96*, 591–600.

Cox, D. (1972), Regression models and life tables, *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*, 187–200.

Dempsey, W., and P. McCullagh (2018), Survival models and health sequences, *Lifetime Data Analysis*, *24*, 550–584.

Desmée, S., F. Mentré, C. Veyrat-Follet, and J. Guedj (2015), Nonlinear mixed-effect models for prostate-specific antigen kinetics and link with survival in the context of metastatic prostate cancer: A comparison by simulation of two-stage and joint approaches, *The AAPS Journal*, *17*, 691–699.

Desmée, S., F. Mentré, C. Veyrat-Follet, B. Sébastian, and J. Guedj (2017a), Nonlinear joint models for individual dynamic prediction of risk of death using hamiltonian monte carlo: application to metastatic prostate cancer, *BMC Medical Research Methodology*, *17*.

Desmée, S., F. Mentré, C. Veyrat-Follet, B. Sébastian, and J. Guedj (2017b), Using the saem algorithm for mechanistic joint models characterizing the relationship between nonlinear psa kinetics and survival in prostate cancer patients, *Biometrics*, *73*, 305–312.

Duchesne, T., and J. Lawless (2000), Alternative time scales and failure time models, *Lifetime Data Analysis*, *6*, 157–179.

Fine, J., H. Jiang, and R. Chappell (2001), On semi-competing risks data, *Biometrika*, *88*, 907–919.

Finkelstein, M. (2004), On the exponential form for reliability, *IEEE Transactions on Reliability*, *53*, 265–268.

Fleming, T., and D. Harrington (1991), *Counting Processes and Survival Analysis*, Wiley, Hoboken.

Gjessing, H., O. Aalen, and N. Hjort (2003), Frailty models based on lévy processes, *Advances in Applied Probability*, *35*, 532–550.

Gradshteyn, I. S., and I. M. Ryzhik (2014), *Table of Integrals, Series, and Products*, eighth ed., Academic Press, Amsterdam.

Hashemi, R., H. Jacqmin-Gadda, and D. Commenges (2003), A latent process model for joint modeling of events and marker, *Lifetime Data Analysis*, *9*, 331–343.

Henderson, R., P. Diggle, and A. Dobson (2000), Joint modelling of longitudinal measurements and event time data, *Biostatistics*, *1*, 465–480.

Hougaard, P. (2000), *Analysis of Multivariate Survival Data*, Springer, New York.

Hoyle, A. (2010), Information-based models of finance and insurance, Ph.D. thesis, Imperial College London, Department of Mathematics, Imperial College London, London SW7 2AZ, United Kingdom.

Hu, C., and A. Tsodikov (2014), Semiparametric regression analysis for time-to-event marked endpoints in cancer studies, *Biostatistics*, *15*, 513–525.

Huibin, Y., and K. Chan (1997), A dynamic frailty model for multivariate survival data, *Biometrics*, *53*, 785–793.

Jewell, N., and J. Kalbfleisch (1996), Marker processes in survival analysis, *Lifetime Data Analysis*, *2*, 15–29.

Jewell, N., and J. Nielsen (1993), A framework for consistent prediction rules based on markers, *Biometrika*, *80*, 153–164.

Kalbfleisch, J., and R. Prentice (2002), *The Statistical Analysis of Failure Time Data, Second Edition*, Wiley, Hoboken.

Köhler, M., N. Umlauf, and S. Greven (2018), Nonlinear association structures in flexible bayesian additive joint models, *Statistics in Medicine*, *37*, 4771–4788.

Kosorok, M. (2008), *Introduction to Empirical Processes and Semiparametric Inference*, Springer, New York.

Kosorok, M., B. Lee, and J. Fine (2004), Robust inference for univariate proportional hazards frailty regression models, *The Annals of Statistics*, *32*, 1448–1491.

Law, N., J. Taylor, and H. Sandler (2002), The joint modeling of a longitudinal disease progression marker and the failure time process in the presence of cure, *Biostatistics*, *3*, 547–563.

Lawless, J., and M. Crowder (2004), Covariates and random effects in a gamma process model with application to degradation and failure, *Lifetime Data Analysis*, *10*, 213–227.

Lin, H., B. Turnbull, C. McCulloch, and E. Slate (2002), Latent class models for joint analysis of longitudinal biomarker and event process data, *Journal of the American Statistical Association*, *97*, 53–65.

Mansuy, R., and M. Yor (2008), *Aspects of Brownian Motion*, Springer-Verlag, Berlin.

McGilchrist, C., and K. Yau (1996), Survival analysis with time dependent frailty using a longitudinal model, *Australian Journal of Statistics*, *38*, 53–60.

Murawska, M., D. Rizopoulos, and E. Lesaffre (2012), A two-stage joint model for nonlinear longitudinal response and a time-to-event with application in transplantation studies, *Journal of Probability and Statistics*, *2012*.

Murphy, S., and A. van de Vaart (2000), On profile likelihood, *Journal of the American Statistical Association*, *95*, 449–465.

Oakes, D. (1995), Multiple time scales in survival analysis, *Lifetime Data Analysis*, *1*, 7–18.

Pauler, D., and D. Finkelstein (2002), Predicting time to prostate cancer recurrence based on joint models for non-linear longitudinal biomarkers and event time outcomes, *Statistics in Medicine*, *21*, 3897–3911.

Peng, L., and Y. Huang (2007), Survival analysis with temporal covariate effects, *Biometrika*, *94*, 719–733.

Perperoglou, A., H. van Houwelingen, and R. Henderson (2006), A relaxation of the gamma frailty (burr) model, *Statistics in Medicine*, *25*, 4253–4266.

Proust-Lima, C., and J. Taylor (2009), Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of post-treatment psa: a joint modeling approach, *Biostatistics*, *10*, 535–549.

Proust-Lima, C., S. Mbéry, J. Taylor, and H. Jacqmin-Gadda (2014), Joint latent class models for longitudinal and time-to-event data: a review, *Statistical Methods in Medical Research*, *23*, 74–90.

Putter, H., and H. van Houwelingen (2015), Dynamic frailty models based on compound birth-death processes, *Biostatistics*, *16*, 550–564.

Rice, J., and A. Tsodikov (2017), Semiparametric time-to-event modeling in the presence of a latent progression event, *Biometrics*, *73*, 463–472.

Rizopoulos, D. (2008), *Joint Models for Longitudinal and Time-to-Event Data with Applications in R*, CRC Press, Florida.

Schultheiss, T., M. Hall, G. Farino, and J. Wong (2016), Analysis of surveillance, epidemiology, and end results program (seer) psa coding error using primary data, *Journal of Clinical Oncology*, *34, no.2 supplement*, 21–21.

Serrat, C., M. Rué, C. Armero, X. Piulachs, H. Perpiñán, A. Forte, Á. Páez, and G. Gómez (2015), Frequentist and bayesian approaches for a joint model for prostate cancer risk and longitudinal prostate-specific antigen data, *Journal of Applied Statistics*, *42*, 1223–1239.

Suresh, K. (2018), Dynamic methods for the prediction of survival outcomes using longitudinal biomarkers, Ph.D. thesis, University of Michigan, Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA.

Suresh, K., J. Taylor, D. Spratt, S. Daignault, and A. Tsodikov (2017), Comparison of joint modeling and landmarking for dynamic prediction under an illness-death model, *Biometrical Journal*, *59*, 1277–1300.

Taylor, J., Y. Park, D. Ankerst, C. Proust-Lima, S. Williams, L. Kestin, K. Bae, T. Pickles, and H. Sandler (2013), Real-time individual predictions of prostate cancer recurrence using joint models, *Biometrics*, *69*, 206–213.

Ting Lee, M., and G. Whitmore (1993), Processes directed by randomized time, *Journal of Applied Probability*, *30*, 302–314.

Ting Lee, M., and G. Whitmore (2006), Threshold regression for survival analysis: Modeling event times by a stochastic process reaching a boundary, *Statistical Science*, *21*, 501–513.

Tran, Q., K. Kidwell, and A. Tsodikov (2018), A joint model of cancer incidence, metastasis, and mortality, *Lifetime Data Analysis*, *24*, 385–406.

Tsiatis, A. (1981), A large sample study of cox's regression model, *The Annals of Statistics*, *9*, 93–108.

Tsiatis, A., and M. Davidian (2004), Joint modeling of longitudinal and time-to-event data: an overview, *Statistica Sinica*, *14*, 809–834.

Tsiatis, A., V. DeGruttola, and M. Wulfsohn (1995), Modeling the relationship of survival to longitudinal data measured with error: applications to survival and cd4 counts in patients with aids, *Journal of the American Statistical Association*, *90*, 27–37.

Tsodikov, A. (2003), Semiparametric models: a generalized self-consistency approach, *Journal of the Royal Statistical Society: Series B (Methodological)*, *65*, 759–774.

Tsodikov, A., L. Liu, and C. Tseng (2020), Likelihood transformations and artificial mixtures, in *Statistical Modeling for Biological Systems*, edited by A. Almudevar, D. Oakes, and J. Hall, pp. 191–209, Springer.

van der Vaart, A., and J. Wellner (1996), *Weak Convergence and Empirical Processes: With Applications to Statistics*, Springer, New York.

van Houwelingen, H., and H. Putter (2012), *Dynamic Prediction in Clinical Survival Analysis*, Taylor & Francis, Boca Raton.

Vaupel, J., K. Manton, and E. Stallard (1979), The impact of heterogeneity in individual frailty on the dynamics of mortality, *Demography*, *16*, 439–454.

Wang, Y., and J. Taylor (2001), Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome, *Journal of the American Statistical Association*, *96*, 895–905.

Whitmore, G., and F. Schenkelberg (1997), Modelling accelerated degradation data using wiener diffusion with a time scale transformation, *Lifetime Data Analysis*, *3*, 27–45.

Wu, L., X. Hu, and H. Wu (2008), Joint inference for nonlinear mixed-effects models and time to event at the presence of missing data, *Biostatistics*, *9*, 308–320.

Yashin, A., and K. Manton (1997), Effects of unobserved and partially observed covariate processes on system failure: a review of models and estimation strategies, *Statistical Science*, pp. 20–34.

Yu, M., N. Law, J. Taylor, and H. Sandler (2004), Joint longitudinal-survival-cure models and their applications to prostate cancer, *Statistica Sinica*, *14*, 835–862.

Yu, M., J. Taylor, and H. Sandler (2008), Individual prediction in prostate cancer studies using a joint longitudinal survival–cure model, *Journal of the American Statistical Association*, *103*, 178–187.

Zeng, D., and D. Lin (2006), Efficient estimation of semiparametric transformation models for counting processes, *Biometrika*, *93*, 627–640.

Zeng, D., and D. Lin (2007), Maximum likelihood estimation in semiparametric regression models with censored data (with discussion), *Journal of the Royal Statistical Society: Series B (Methodological)*, *69*, 507–564.

Zucker, D. (2005), A pseudo-partial likelihood method for semiparametric survival regression with covariate errors, *Journal of the American Statistical Association*, *100*, 1264–1277.