Sparse Selection in Cox Models with Functional Predictors

Yulei Zhang

Submitted in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2012

ABSTRACT

Sparse Selection in Cox Models with Functional Predictors

Yulei Zhang

This thesis investigates sparse selection in the Cox regression models with functional predictors. Interest in sparse selection with functional predictors (Lindquist and McKeague [24]; McKeague and Sen [29]) can arise in biomedical studies. A functional predictor is a predictor with a trajectory which is usually indexed by time, location or other factors. When the trajectory of a covariate is observed for each subject, and we need to identify a common "sensitive" point of these trajectories which drives outcome, the problem can be formulated as sparse selection with functional predictors. For example, we may locate a gene that is associated to cancer risk along a chromosome.

The functional linear regression [37] method is widely used for the analysis of functional covariates. However, it could lack interpretability. The method we develop in this thesis has straightforward interpretation since it relates the hazard to some sensitive components of functional covariates.

The Cox regression model has been extensively studied in the analysis of time-to-event data. In this thesis, we extend it to allow for sparse selection with functional predictors. Using the partial likelihood as the criterion function, and following the 3-step procedure for M-estimators established in van der Vaart and Wellner [54], the consistency, rate of convergence and asymptotic distribution are obtained for M-estimators of the sensitive point and the regression coefficients. In this thesis, to study these large sample properties of the estimators, the fractional Brownian motion

assumption is posed for the trajectories for mathematical tractability.

Simulations are conducted to evaluate the finite sample performance of the methods, and a way to construct the confidence interval for the location parameter, i.e., the sensitive point, is proposed.

The proposed method is applied to an adult brain cancer study and a breast cancer study to find the sensitive point, here the locus of a chromosome, which is closely related to cancer mortality. Since the breast cancer data set has missing values, we investigate the impact of varying proportions of missingness in the data on the accuracy of our estimator as well.

Table of Contents

List of Figures

List of Tables

Acknowledgments

No words can convey my gratitude to Dr. Ian W. McKeague. I can not meet a better advisor. The strong desire to learn more about skills of Empirical Processes theory drove me to choose him as my advisor. Thanks to his patience and encouragement, I never lost my confidence when facing so many hurdles in my research. I spent so much time reading van der Vaart and Wellner's book without much progress for my dissertation in the beginning year and he encouraged me instead of blaming me. Thanks to his considerate deed, when I met family financial crisis in 2010, allowing me to go to Chicago working for half year to pass the hard time.

Special thanks go to Dr. Wei-Yann Tsai who was an active academic advisor and now chair of my dissertation committee. The training I received from his survival analysis course is very useful for my dissertation. He also tried to help me find an RA position. Dr. ZheZhen Jin, Dr. Todd Ogden, Dr. Susser have been in my PhD research proposal committee and continue to be in my defense committee. I feel so grateful for their careful reading of my draft and constructive advise proposed. Dr. Jin and Dr. Ogden's courses in inference and statistical computing played important role in my research as well.

I also want to say thanks to Dr. Bruce Levin, Dr. Xinhua Liu, Dr. Markatou, Dr. Myunghee Cho Paik, Dr. John L. P. Thompson, Dr. Shuang Wang, Dr. Ying Wei who has helped my job hunting, either RA position or fulltime ones, especially Dr. Levin and Dr. Paik. I enjoyed working as Teaching Assistant for Dr. Ken Cheung very much. I appreciate Dr. Tai-Tsang Chen and other colleagues In Bristol-Myers Squibb who provide me a chance to learn more about pharmaceutical statistics. Also the experience working with Ms. Lei Zhang, Ms. Yea-Lin Lin, Ms. Karen Sachs, Ms. Tracy Miller and other colleagues in Abbott Molecular Diagnostics was happy and memorable. I am deeply indebted to Lizeng Zhang, Alireza Javaheri, Lei Fang, Lin Zhu and other colleagues during my internship at J.P. Morgan.

Thanks also go to Dr. Edward Nunes, Dr. Huiping Jiang, Dr. Naihua Duan, Ms. Carrie Davies and other colleagues in NYSPI. I enjoyed working with and learning from them and wish I could have worked harder and contributed more if time came back again.

During my stay in Columbia Biostatistics, lots of fellow students in Columbia enriched my life: Xianhuang Zhou, Yi-Hsuan Tu, Hong Tian, Xiaodong Luo, Chung Chang, Philip Reiss, Yimeng Lu, Wanling Tai, Hui Zhang, Tao Duan, Yihong Zhao, Lin Huang, Jia Cao, Rui Liu, Shing Lee, Rositsa Dimova, Chenbo Zhu, Jimmy Duong, Nanshi Sha, Zhigang Li, John Spivack, Huaihou Chen, Xin Li, Wei Xiong, Wenfei Zhang, Yi Wang, Bingzhi Zhang from Biostatistics department, Yixin Fang, Man Jin, Jinfeng Xu, Qinghua Li from Statistics department and Shean Xiang Zhou, Qiangfeng Zhang, Jung-Wei Fan, Lixia Yao, Peng Liu, Xuelian Huang, Yunfei Cai, Xianhua Peng from other departments. Their names made me remember the life in Washington Heights from time to time even after I moved away from Manhattan. I am highly indebted to Jimmy Duong who spent lots of time helping me polish the language of my thesis.

Also thanks go to Ms. Leslie McHale, Ms. Michelle McClave, Mr. Zeffrey Rodrigues and Ms. Carolyn Kaufman, Ms. Justine Herrera, Ms. Nancy Zai, Mr. Zeffrey Rodrigues, Mr. Anthony Guerrero, Mr. Carlos, who have helped every student here.

Last but most important, I thank my parents and my wife. My parents, Xiang'e He and Qingjun Zhang, have been supporting and encouraging me in various ways throughout my study ever since I was a child; my wife, Xuesong Yu, accompanied me, sacrificing her promising career in China. Their love will continue to encourage me to proceed.

dedicated to my dream and struggle

Chapter 1

Introduction

This thesis proposes a novel variant of the Cox model, to allow for sparse selection with functional predictors. The Cox model ([10]) is a widely used statistical model in surival analysis. It models the hazard of experiencing events in a semiparametric way, the product of a baseline hazard function which is the unspecified nonparametric part and a parametric part which involves the covariates of interest. The covariates of interest can be either functional covariate or non-functional covariate. The classical Cox models with non-functional covariates have been extensively studied and widely used in medical research. A functional covariate is a covariate with a continuously observed trajectory which can be indexed by time, location among others. A non-functional covariate is a random variable which takes scalar values.

When functional covariates are correlated to a scalar outcome, the functional linear regression method (see [37]) can be used to study the varying effect of the functional covariate along the trajectory. The resulting model can have good fitting but is often hard to interpret. To overcome this drawback, some alternative methods are proposed, e.g., the functional linear regression that is interpretable [17] and the Point Impact model [29], which assume the response variable is correlated to some "sensitive" regions/points of the trajectories. In this thesis, we will extend the Point Impact model in the linear regression setting to the Cox model setting.

We are interested in finding the sensitive location of the functional covariate and its impact strength, allowing for other non-functional covariates to appear in the Cox model as well. To investigate the parameters of interest, the maximum partial likelihood principle is used to get the estimators. The large sample properties of the estimators including the consistency, rate of convergence, and asymptotic distribution are proved under assumptions about the trajectories. We also propose methods to construct confidence intervals for the maximum partial likelihood estimators.

The proposed methods can be used in a wide range of fields including genetics, environmental science, network traffic and finance, as long as the study interest is to locate the most sensitive point of the trajectories for a covariate. Here the "most sensitive" means that among all the points, this point provides the best prediction of the time-to-event outcome.

A brief outline of the thesis is as follows. In the first chapter, we introduce the background and motivation of this thesis. Then our model is proposed and its theoretical properties are studied in Chapter 2. To make the theoretical development in Chapter 2 more accessible, a simplified model is explored first before an extended model is fully investigated. The proofs are presented in Chapter 2 except for some more technical proof steps (which are collected in Appendices). To make the construction of Wald-type confidence intervals feasible, Chapter 3 is devoted to the Monte Carlo calibration of quantiles. A survey of the proposed procedure is given at the end of Chapter 3. We perform extensive simulations in Chapter 4 to study the finite sample performance of the proposed methods. Our methods are applied to an adult brain cancer study and a breast cancer study in Chapter 5. Chapter 6 summarizes the thesis and discusses possible further research. The thesis ends with Appendices, which collect proofs of lemmas used in Chapter 2, and some more technical proof steps omitted in Chapter 2.

In this chapter, we start with the biomedical background and motivation of our thesis. Then we present the survival analysis background, propose our model, and review some related literature. We conclude this chapter with an introduction of theorems in empirical process theory that will be used in Chapter 2 and the Appendices.

1.1 Biomedical background

With the development of modern medicine, targeted therapies are becoming popular in cancer treatment. Targeted cancer therapies aim to "block the growth and spread of cancer by interfering with specific molecules involved in tumor growth and progression", and "may be more effective than current treatments and less harmful to normal cells" (National Cancer Institute [30]).

To make targeted cancer therapies feasible, the specific biomarkers which are most closely related to cancer risk have to be identified. Biologists seek sensitive biomarkers by testing and comparing healthy tissues and tumor tissues. We statisticians develop statistical methods to identify predictive biomarkers for cancer risk.

A biomarker is a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention ([6]). Traditional biomarkers include body temperature, blood pressure, and blood test. New technologies in genomics and proteomics help scientists find molecular biomarkers for diseases including cancers. These new technologies have received increasing attention in recent years.

In 2008, a team in Washington University at St. Louis discovered ten genes with acquired mutations by sequencing a typical acute myeloid leukemia genome, and its matched normal counterpart from the same patient ([22]). Researchers in Canada sequenced a lobular breast cancer genome and found 5 prevalent somatic mutations in DNA from the primary tumour ([43]). In 2009, a research consortium led by the Wellcome Trust Sanger Institute found more than 23000 mutations caused by smoking in the DNA of a lung cancer ([35]) and 33000 mutations in the DNA of a skin cancer ([34]). Now the consortium is focusing on finding the key genetic mutations that fueled these cancers. All these indicate the importance of extracting key elements out of high dimensional (functional) information. Our method is proposed to meet this challenge from a statistical perspective.

Biomarker discovery provides promising prospects for health care. With biomarkers determined, subjects can be screened or surveilled for diseases using established diagnostic biomarkers, and/or targeted therapy can be applied to therapeutic biomarkers. In breast cancer surveillance, serum tumor markers such as CA 15-3, carcinoembyonic antigen (CEA), and CA 27-29 are widely used today ([9]). Experiments are conducted to explore targeted therapy based on genetic biomarkers. E.g., in 2010, a targeted therapy study was conducted on rodents with spinal cord injuries at the University of Maryland School of Medicine in Baltimore ([44]). A specific single strand of DNA was given to the rodents to block Abcc8 gene activity. The Abcc8 gene activates SUR1 protein, which allows sodium into cells, increases the risk for cells to inflate, explode, and die in severe injury. Injured rodents given the new gene-targeted therapy had lesions that were one-fourth to one-third the size of lesions in those not treated. They also recovered much better. This experiment demonstrates the potential usefulness of targeted therapy at the DNA level, and inspires the exploration of disease-sensitive biomarkers in genomics.

1.2 Motivation

This thesis is motivated by the interest to locate genes related to cancers. We now give two biomedical examples to be analyzed later in this thesis. One of the examples concerns an adult brain cancer study ([48]), and the other a breast cancer study ([45]). In the adult brain cancer study, the complete gene expression profile is available for each subject. In the breast cancer study, on the other hand, the gene expression profiles for some subjects are not completely observed. In each study, we are interested in locating the locus on a chromosome that is related to the risk of dying of cancer.

Glioblastoma is the most common primary brain tumor for adults ([14]). The median survival of newly diagnosed patients is only about 1 year. If sensitive genes for glioblastoma can be identified through genome-wide profiling studies, subjects can be stratified by the sensitive genes into subgroups. Then different treatment regimes can be applied to these subgroups, with the goal of improving treatment outcome. This is the idea of "personalized medicine" and "stratified medicine". We do not discuss it further here. For more information, see [51].

In the glioblastoma study [48], the complete gene expression profile for each subject is collected, and the survival outcome is also obtained. We will develop statistical methods to locate the sensitive locus on a chromosome that predicts the patients' risk of dying of glioblastoma.

Breast cancer is the second leading cause of death from cancer for American women, and one of the major causes of death from cancer worldwide. Finding the key locus on a chromosome that is sensitive to the risk of dying of breast cancer is urgent. If the key locus can be determined, scientists can target it to find ways to lessen the risk of dying of breast cancer.

This breast cancer study has 78 carcinomas and 3 fibroadenomas breast tissues collected. Each tissue corresponds to a subject except for 2 carcinomas breast tissues from one patient diagnosed at different times (one of the two will be excluded in the data analysis). Each subject's gene expression profile and clinical outcome are available. An important scientific question is which locus on a chromosome has the most significant influence on the subjects' risk of dying of breast cancer.

The biological problem is what location of the trajectory best predicts the risk of dying of cancer. For each subject, the gene expression levels of all the loci along a chromosome can be viewed as a trajectory of a stochastic process. In other words, the gene expression process is indexed by the location along the chromosome. Then the question becomes locating this point along the trajectory that is associated with risk of death from cancer. In this way, statistical strategies can be posed to answer the scientific question.

This thesis proposes statistical methods to select the sensitive location. We will study their large sample properties by theoretical derivation and finite sample performance by simulation studies.

1.3 Survival Analysis background

In this section, we will give background on the Cox model, the maximum partial likelihood estimator, and the counting process approach to the Cox model. They are frequently used in survival analysis.

1.3.1 Cox model

The Cox regression model ([10]), also known as the proportional hazards model, has been widely used in the analysis of time-to-event data. We review the Cox model and some notation which will be used frequently in this thesis. More results on the Cox model can be found in classical books on survival analysis, e.g., [2] and [19].

Denote the failure time as T^0 and the censoring time as C. Instead of observing T^0 and C, we can only observe the follow-up time $T = \min(T^0, C)$ and the non-censoring indicator $\delta = 1_{T^0 \leq C}$, where 1_A is an indicator function of event A.

Assume that T^0 and C are conditionally independent given a p-vector of predictable covariates $\mathbf{X}(t)$. For each subject i, we observe $T_i, \delta_i, \{\mathbf{X}_i(t) : t \in [0, T_i]\}$. Instead of modeling the time-to-event T^0 directly, the Cox model sets up the relationship between the hazard function of T^0, $\lambda(t|\mathbf{X}) = \lim_{\Delta \to 0+} \frac{P(t \leq T^0 < t+\Delta | T^0 \geq t; \mathbf{X})}{\Delta}$ and the covariate $\mathbf{X}(t)$ in a semiparametric form,

$$\lambda(t|\mathbf{X}) = \lambda_0(t) \exp\{\beta^T \mathbf{X}(t)\}.$$

The hazard function $\lambda(t)$, the failure intensity of failure times, is modeled as the product of the baseline hazard function $\lambda_0(t)$ and an exponential regression function with regression coefficient vector β and covariate vector $\mathbf{X}(t)$, and β^T means the transpose of β. The baseline hazard function $\lambda_0(t)$ is unspecified except that it is non-negative. The cumulative baseline hazard function $\Lambda_0(t) = \int_0^t \lambda_0(u)du$.

1.3.2 Partial likelihood

Statistical inference for the Cox model is usually based on maximizing the partial likelihood function. The partial likelihood, originally unnamed, was used in Cox [10] where D. R. Cox proposed the Cox proportional hazards model. Kalbfleisch and Prentice [18] examined the Cox model with covariates not depending on time, and without censoring. They found that if there were no ties in event times, the formula that Cox [10] used was the marginal likelihood of the ranks of the event times, not dependent upon the specific time values themselves. Then D. R. Cox [11] justified the use of his formula and named it the partial likelihood. The same formula applies to time-dependent covariates.

The partial likelihood function for the Cox model is

$$PL(\beta) = \prod_{j=1}^{n} \left(\frac{\exp\{\beta^T \mathbf{X}_j(t_j)\}}{\sum_{i=1}^{n} 1_{T_i \geq T_j} \exp\{\beta^T \mathbf{X}_i(t_j)\}} \right)^{\delta_j}.$$

In this formula, the nonparametric element, i.e. the baseline hazard function $\lambda_0(t)$ has been eliminated, and hence the partial likelihood can be used for inference of the parametric element of the Cox regression model.

Efron [13] and Oakes [32] considered the efficiency of the partial likelihood estimator $\hat{\beta}_{PL}$ versus the maximum likelihood estimator $\hat{\beta}_{MLE}$ of parametric submodels. In parametric submodels, $\lambda_0(\cdot)$ is specified up to certain unknown parameters. They argued that the asymptotic variance of $\hat{\beta}_{PL}$ will be close to that of $\hat{\beta}_{MLE}$ given that the parametric family is reasonably rich. Efron [13] concluded that the asymptotic variance of $\hat{\beta}_{MLE}$ approaches that of $\hat{\beta}_{PL}$ if the number of independent parameters in the parametric setting goes to infinity. This implies that the efficiency of the estimator of β can not be improved from the partial likelihood estimator $\hat{\beta}_{PL}$ without constraints on $\lambda_0(\cdot)$.

Thus the maximum partial likelihood estimator $\hat{\beta}_{PL}$ (which will be abbreviated as $\hat{\beta}$ hereafter) is a good choice to estimate the regression parameters β. To estimate the baseline hazard function $\lambda_0(t)$, Breslow ([7] and [8]) gave an estimator of the cumulative baseline hazard function $\Lambda_0(t)$ which is called the Nelson–Aalen estimator or Breslow estimator.

$$\hat{\Lambda}_0(t) = \sum_{T_j \le t} \frac{\delta_j}{\sum_{i=1}^{n} 1_{T_i \ge T_j} \exp\{\hat{\beta}^T \mathbf{Z}_i(T_j)\}}.$$

1.3.3 Counting process approach

The Cox regression model was extended by Anderson and Gill [1] with a counting process approach. They used a predictable at-risk process $Y(t) = 1_{T \ge t}$ and a counting process $N(t) = \delta 1_{T \le t}$ as an alternative to record T and δ, and obtained

$$M(t) = N(t) - \int_0^t Y(u)\lambda_0(u) e^{\beta^T \mathbf{X}} du$$

as a local square integrable martingale. Using a martingale central limit theorem, they proved the asymptotics of the maximum partial likelihood estimator $\hat{\beta}$ under mild conditions. See Anderson and Gill [1] for more details. We will adopt the counting process approach to develop our theoretical results throughout this thesis.

1.4 Proposed model

For a time-to-event outcome, to find the most influential point of the trajectories for a functional covariate, we set up the following Cox regression model,

$$\lambda(t|\mathbf{Z};\mathbf{X}) = \lambda_0(t)\exp\{\beta Z(\theta) + \gamma^T \mathbf{X}\}.$$

The baseline hazard function, $\lambda_0(t)$, is unspecified except that it is non-negative. Here θ is our key interest, the most important point of the trajectories that drives the subjects' event risk. The covariate $Z(\theta)$ is the sensitive point with location θ, where θ is unknown. The regression coefficient β captures the impact strength of the sensitive point and the regression coefficient vector γ captures the impact strength of the other non-functional covariate vector $\mathbf{X}$.

Even though we observe the realized trajectory of the stochastic process $\mathbf{Z} \equiv \{Z(\tilde{\theta}) : \tilde{\theta} \in [0, \theta_M]\}$ for each subject, the proposed model considers only one point on the trajectory, $Z(\theta)$ where θ is shared by every subject, to be correlated to the time-to-event outcome.

Since θ is unknown when we set up the model and needs to be estimated from data, the parameter estimation process performs sparse selection of the functional predictor $\{Z(\tilde{\theta}) : \tilde{\theta} \in [0, \theta_M]\}$. After the selection, only $Z(\theta)$ is left and the values of all other locations $\{Z(\tilde{\theta}) : 0 \leq \tilde{\theta} < \theta \quad \text{or} \quad \theta < \tilde{\theta} \leq \theta_M\}$ are eliminated from the model. This explains the title of this thesis.

Except for the sparsely selected functional predictor $Z(\theta)$, the proposed model allows for a non-functional covariate vector $\mathbf{X}$ as well. These non-functional covariates are random variables, independent of the stochastic process $\mathbf{Z}$. There is no intercept term in the exponential part of the Cox model since it is already absorbed into the $\lambda_0(\cdot)$.

In later development in Chapter 2, we will see that in deriving the large sample properties, a key assumption we make is that $\mathbf{Z}$ follows a 1-dimensional fractional Brownian motion (abbreviated as fBm hereafter) with Hurst parameter H starting from θ (i.e. $Z(\cdot + \theta) - Z(\theta)$ follows a standard 1-dimensional 2-sided fBm with Hurst parameter H), where $Z(\theta)$ is a random variable independent of the fBm. We observe the trajectory of $Z(\cdot)$ from location 0 to θ_M for each subject.

Now we explain why the fBm assumption is made. Fractional Brownian motion is a model of fractal phenomena (Mandelbrot [28], [27]) and has been successfully used in environmental science, finance and network traffic studies. Since gene expression data along a chromosome also displays a fractal pattern (see Figure 1.1), which is consistent with results by Lieberman et al. [23], we make this assumption in our

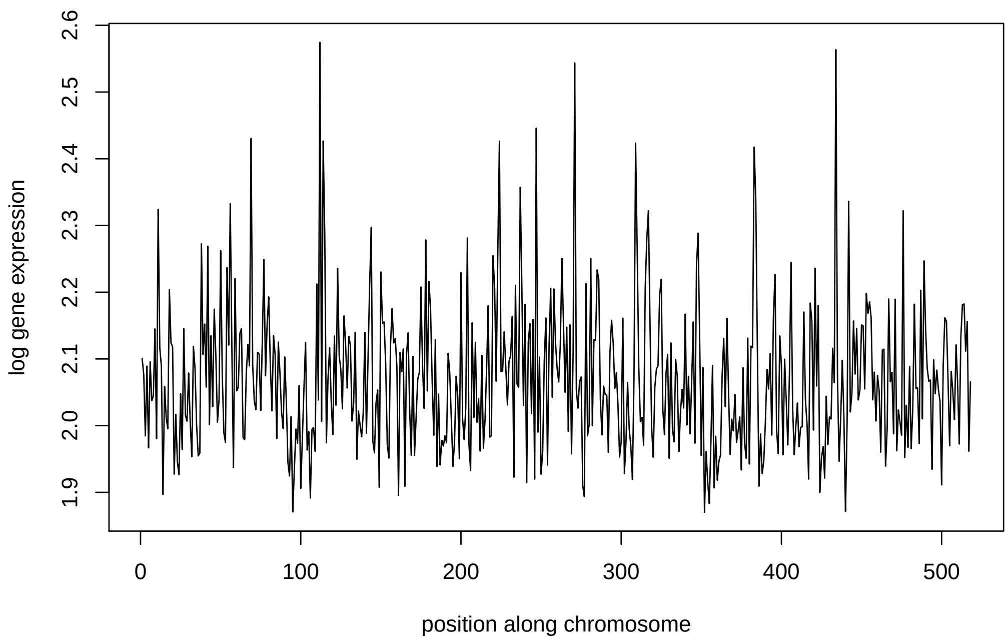


Figure 1.1: Log(gene expression level) at 518 loci along Chromosome 17 from one breast cancer tissue

model setting. Another reason is for mathematical tractability in the investigation of the proposed model.

We emphasize here that the fBm assumption is used only to derive the large sample properties, but not a prerequisite for the statistical method to work. When the fBm assumption does not hold, we can still use the maximum partial likelihood principle to obtain the estimates of the parameters (θ, β, γ). However, the statistical inference, e.g. confidence interval construction and hypothesis testing, can't be based on the large sample properties derived in Chapter 2. In such cases, alternative methods may be posed to address the statistical inferences, e.g., a bayesian method is briefly discussed in Chapter 4.

Since we need to refer to the properties of the fBm frequently in the theoretical development of our model, we now introduce the definition and the basic properties of fBm.

A standard fBm with Hurst parameter $H \in (0,1)$ is a Gaussian process $G_H = \{G_H(t) : t \in \mathbb{R}\}$ with continuous sample paths, having zero mean and covariance function

$$\mathrm{Cov}\left(G_H(s), G_H(t)\right) = \frac{1}{2}\left(|s|^{2H} + |t|^{2H} - |s-t|^{2H}\right).$$

The fBm has a self similarity property: for any $c > 0$, $G_H(ct) =_d c^H G_H(t)$ as processes, where $_d$ mean equal in distribution. Brownian motion is a special case of the fBm ($H = \frac{1}{2}$). H is an index indicating the trajectory roughness of the fBm, where higher H corresponds to smoother sample paths.

Assume that we observe n i.i.d. copies of $\{T, \delta, \mathbf{Z}, \mathbf{X}\}$. Our inference will be based on these data.

1.5 Related literature

We review some literature of point impact models. Point impact models were introduced by McKeague and Sen [29] in the linear regression setting, and extended to the generalized linear model setting by Lindquist and McKeague [24]. The model proposed in this thesis extends these two papers to the Cox model setting. Becoming familiar with these related literature can help us understanding this thesis. We also need to know these results so that we can compare them to our results in subsequent chapters.

McKeague and Sen [29] considers a scalar outcome for a point impact model,

$$Y = \alpha + \beta X(\theta) + \epsilon,$$

where θ is the parameter of interest. To estimate all the unknown parameters (α, β, θ), the least squares method is used,

$$\left(\hat{\alpha}_n, \hat{\beta}_n, \hat{\theta}_n\right) = \arg\min_{\alpha,\beta,\theta} \sum_{i=1}^{n} \left[Y_i - \alpha - \beta X_i(\theta)\right]^2.$$

Suppose that $(X_i, Y_i), i = 1, ..., n,$ are independent and identically distributed (i.i.d.) satisfying the model, with unknown parameters $(\alpha, \beta, \theta) \in \mathbb{R}^2 \times [0, 1]$, where the true values are $(\alpha_0, \beta_0, \theta_0)$.

The following conditions are used to obtain the asymptotic distribution of the least square estimators:

(1) X follows a fractional Brownian motion with Hurst exponent $H \in (0, 1)$,

(2) $0 < \theta_0 < 1$ and $\beta_0 \neq 0$,

(3) $E|\epsilon|^{2+\delta} < \infty$ for some $\delta > 0$.

Under conditions (1) and (2), the estimator $\left(\hat{\beta}_n, \hat{\gamma}_n, \hat{\theta}_n\right)$ is consistent. With the additional condition (3),

$$\left(\sqrt{n}(\hat{\alpha}_n - \alpha_0), \sqrt{n}(\hat{\beta}_n - \beta_0), n^{1/(2H)}(\hat{\theta}_n - \theta_0)\right)$$
$$\to d\left(\sigma Z_1, |\theta_0|^{-H}\sigma Z_2, \arg\min_{t\in\mathbb{R}}\{2\frac{\sigma}{|\beta_0|}B_H(t) + |t|^{2H}\}\right),$$

where Z_1 and Z_2 are i.i.d. $N(0, 1)$ and independent of the fBm B_H.

Since the asymptotic distribution involves the nuisance parameter H, we need to estimate H from data in order to apply the asymptotic distribution for inference. A residual-based bootstrap was proposed to avoid estimating H, which made the application of the method easier and more appealing.

Lindquist and McKeague [24] extends this to the generalized linear model setting, where the conditional density of a scalar response Y given X is modeled by a canonical exponential family

$$p(y|X) = \exp\left([X(\theta)y - b(X(\theta))]/a(\phi) + r(y, \phi)\right)$$

for some known functions $a(\cdot), b(\cdot)$, and $r(\cdot, \cdot)$.

They obtained a similar asymptotic distribution to that of McKeague and Sen [29], and applied the method to two real data sets. One is a functional magnetic resonance imaging (fMRI) study to locate an anxiety-provoking period for subjects, and the other is a gene expression study to estimate the most sensitive locus along a chromosome to classify breast cancer patients and normal subjects.

1.6 Theorems in Empirical Process Theory

In this section, we review the 3-step procedure in van der Vaart and Wellner (hereafter abbreviated as VW) [54] to establish the asymptotics of the M-estimators using the empirical process theory. The first part discusses M-estimation theorems to be used in each of the three steps. The second part presents three fundamental theorems from the empirical process theory which are powerful tools to justify the conditions in these M-estimation theorems.

To understand the following theorems from the empirical process theory, it is necessary to familiarize ourselves with some definitions and notation in Chapter 2 of VW [54], including *empirical measure*, *empirical process*, *outer probability*, *bracketing number*, *entropy with bracketing*, the *Glivenko−Cantelli class* and the *Donsker class.*

1.6.1 M-estimation theorems

An M-estimator $\hat{\theta}_n$ is the value of θ that maximizes a random criterion function $\mathbb{M}_n(\theta)$, which is usually an empirical version of a criterion function, i.e. $\mathbb{M}_n(\theta) = \mathbb{P}_n m(\theta)$. To derive the asymptotic distribution of an M-estimator, there is an elegant 3-step procedure: prove its consistency, then obtain its rate of convergence, and finally derive its limiting distribution.

Consistency

To prove consistency, two common practices are the (generalized) Wald's method and the method based on establishing the uniform convergence property of the empirical criterion function. We will adopt the latter one to establish consistency of the M-estimator by Corollary 3.2.3 in VW [54].

In the case of i.i.d data and the empirical criterion function $\mathbb{M}_n(\theta) = \mathbb{P}_n m_\theta$, the uniform convergence condition in Corollary 3.2.3 in VW [54] is equivalent to the condition that $\{m_\theta : \theta \in \Theta\}$ is a Glivenko−Cantelli class.

Rate of convergence

To obtain the rate of convergence for the maximizer $\hat{\theta}_n$ of the random criterion function $\mathbb{M}_n(\theta)$, we assume that θ_0 maximizes the non-random criterion function $\mathbb{M}(\theta)$. To make θ_0 the maximizer, we expect $\mathbb{M}(\theta)$ to have a local quadratic property near θ_0,

$$\mathbb{M}(\theta) - \mathbb{M}(\theta_0) \lesssim -d^2(\theta, \theta_0),$$

where $\lesssim$ means less than or equal to up to a universal constant. Given this, we can get an upper bound for the convergence rate of $\hat{\theta}_n$ based on the modulus of continuity of $\mathbb{M}_n - \mathbb{M}$ at θ_0. Theorem 3.2.5 in VW [54] provides a powerful tool for this purpose.

In Chapter 3.4 of VW [54], this theorem is generalized to apply to sieved M-estimators, where sieves $\Theta_n (n \geq 1)$ is a sequence of subsets of the parameter space. The sieved M-estimator $\hat{\theta}_n$ is defined as the maximizer of $\mathbb{M}_n(\theta)$ over the sieve Θ_n. Generally, for the sieved M-estimator $\hat{\theta}_n$ to be consistent, the sieves $\Theta_n (n \geq 1)$ must be constructed to grow dense in Θ as $n \to \infty$. The simplest sieves series with this property is the whole space, i.e., $\Theta_n = \Theta$ for every $n \geq 1$.

We will use another slightly generalized version of this theorem. Define an event sequence Ω_n as a sequence of event sets on the sample space such that $P^*(\Omega_n) \to 1$, where P^* is the outer probability. If the modulus of continuity condition in the theorem holds on Ω_n, the conclusion of the theorem still holds. This generalization helps in the calculation of the modulus of continuity when applying this theorem, and has been used in Banerjee and McKeague [4] as well.

Asymptotic distribution

Once we obtain the convergence rate $\tilde{r}_n$ of the M-estimator $\hat{\theta}_n$ (i.e. $\tilde{r}_n(\hat{\theta}_n - \theta_0) = O_P^*(1)$), the next step is to establish its limiting distribution.

First we show that a suitably rescaled version of the empirical criterion function, $\mathbb{Q}_n(h) \equiv s_n \left(\mathbb{M}_n(\theta_0 + \frac{h}{\tilde{r}_n}) - \mathbb{M}_n(\theta_0) \right)$ converges in distribution to a process $\mathbb{Q}$ in the space $l^\infty(h : \|h\| \leq K)$ for every K, where θ_0 is the true parameter.

Then if the sample paths $h \mapsto \mathbb{Q}(h)$ of the limit process are upper semicontinuous and possess a unique maximizer $\hat{h}$, we conclude that the sequence $\tilde{r}_n(\hat{\theta}_n - \theta_0)$ converges in distribution to $\hat{h}$ by Theorem 3.2.2 in VW [54].

1.6.2 Fundamental theorems

The following fundamental theorems from empirical process theory are expected to play key roles in verifying the conditions of the M-estimation theorems mentioned in Chapter 1.6.1.

Glivenko−Cantelli Theorem

The Glivenko−Cantelli Theorem is the uniform version of the Strong Law of Large Numbers over a class of functions. It is often used to verify the uniform convergence condition for the *consistency* theorem in Chapter 1.6.1. We will use Theorem 2.4.1 in VW [54], the version based on entropy with bracketing, in this thesis.

Maximal Inequality for empirical process

Various maximal inequalities are presented in Chapter 2.14 of VW [54]. They are often used to verify the modulus of continuity condition for the *rate of convergence* theorem in Chapter 1.6.1. Theorem 2.14.2 and Theorem 2.14.5 in VW [54] will be used in our study. By these theorems, we can bound the $L_1(P)$ and $L_2(P)$ norms of the supremum of the empirical processes $\mathbb{G}_n$ over a class of functions $\mathcal{F}$ that possesses a finite bracketing entropy integral.

Lindeberg–Feller Theorem for stochastic processes

Theorem 2.11.9 in VW [54] is often used to verify the convergence in $l^\infty(\mathcal{F})$ condition of the rescaled criterion functions of Theorem 3.2.2 in VW [54] to obtain the *asymptotic distribution* of the M-estimators. We call it the Lindeberg–Feller Theorem for stochastic processes here since it parallels the Lindeberg–Feller Theorem for random variables. The stochastic processes need to be independent but not necessarily identically distributed to apply this theorem.

Chapter 2

Large sample properties

In this chapter, we will study the proposed model in two stages. In the first stage, we explore a special case of our general model. In this case, we set the coefficient of $Z(\theta)$ to be 1 and consider no other non-functional covariates. In this way, we focus on the essential element of interest, to select a sensitive point on the trajectories of a stochastic process that predicts time-to-event outcomes. This stage helps us both capture the essential feature of the proposed model and keep the proof from being formidable. In the second stage, the special-case model is extended to the full model, adding both the coefficient for $Z(\theta)$ and other non-functional covariates.

We focus on the large sample properties of the maximum partial likelihood estimator for the proposed model in both stages. The structures of the two stages are the same. In each stage, following the 3-step procedure for M-estimators in Chapter 3 of VW [54], the consistency, rate of convergence and asymptotic distribution of the maximum partial likelihood estimator are obtained. Each section ends with proofs of these large sample properties.

To learn about the finite sample performance of the maximum partial likelihood estimator of this model, Monte Carlo simulations are conducted and the corresponding results are collected in Chapter 3.

2.1 Model setting

Suppose for each subject, we observe his/her gene expression level profile $Z(\cdot)$ over $[0, \theta_M]$ on a chromosome with length θ_M, survival time (possibly right censored) T, non-censoring indicator variable δ ($\delta = 0$ means being right censored). The followup time for all the patients is set to be τ, a prespecified fixed time.

Neither the survival time T^0 nor the censoring time C is always observed. Instead we observe $T = T^0 \wedge C$, $\delta = 1_{T^0 \le C}$, where $T^0 \wedge C \equiv \min\{T^0, C\}$. Using the counting process approach in Chapter 1.3.3 to record them, we observe the at risk process $Y(t) = 1_{T \ge t}$ and the counting process $N(t) = 1_{(\delta=1, T \le t)}$. Note the relationship $Y(t) = 1_{(T^0 \ge t, C \ge t)}$ and $N(t) = 1_{(T^0 \le C, T^0 \le t)}$ hold. Notice that in practice, the values of $Z(\cdot)$ over $[0, \theta_M]$ are commonly observed on a grid fine enough instead of continuously on $[0, \theta_M]$.

Even though we observe the whole gene expression profile along a chromosome for each subject, we assume that subjects' event risk depends on only one unknown locus on the chromosome, θ ($0 \le \theta \le \theta_M$). The hazard function

$$\lambda(t|\mathbf{Z}; \mathbf{X}) = \lambda_0(t) \exp\{\beta Z(\theta) + \gamma^T \mathbf{X}\}.$$

By estimating the model parameter θ, we can pick out the specific location from the gene expression profile. This idea of extracting a single point from the functional data (here the gene expression levels at continuous loci along a chromosome formed a functional data),is named the *point impact model*, and its relationship to the widely used *functional linear regression model* has been discussed by McKeague and Sen [29].

The point impact model could be extended to include multiple impact points, which is a compromise between point impact model that emphasizes model interpretability and functional linear model that emphasizes fitting accuracy.

A well-known method for sparse selection and shrinkage estimation is the LASSO ([49], [50]). The LASSO selects important variables out of a set of covariates which could be either dependent or independent. For the model of this thesis, the variables to be selected from are correlated to each other, with higher correlation for closer distance.

2.2 Simplified model

In this section, we will study a special case of the proposed model. The exploration in the simplified model can reveal the key property of the proposed model by including the key feature, while keep the proof from being formidable. Once we understand the simplified model, we have been familiar with the study tools and got some prospect for the results of the general model (i.e., full model) as well, both of which will help our investigation of the general model.

To keep things simple but retain the essential element of our interest, we assume $\beta = 1$ and omit other non-functional covariates $\mathbf{X}$ in this section. Then we have a simplified model with hazard function

$$\lambda(t|\mathbf{Z}) = \lambda_0(t)e^{Z(\theta)}, \quad t \geq 0. \tag{2.1}$$

To estimate the unknown parameter θ, we will use the partial likelihood principle and adopt the M-estimation framework. For this model, the log partial likelihood function is used as the empirical criterion function,

$$\begin{aligned}\mathbb{M}_n(\theta) &= \mathbb{P}_n\left[\int_0^\tau Z(\theta)dN(s) - \int_0^\tau \log[\mathbb{P}_n(Y(u)e^{Z(\theta)})]dN(u)\right] \\ &= \mathbb{P}_n\left[Z(\theta)N(\tau) - \int_0^\tau \log[\mathbb{P}_n(Y(u)e^{Z(\theta)})]dN(u)\right].\end{aligned}$$

The maximum partial likelihood estimator of $\hat{\theta}_n = \mathrm{argmax}_\theta \mathbb{M}_\mathrm{n}(\theta)$.

Consider

$$\mathbb{M}(\theta) = P\left[Z(\theta)N(\tau) - \int_0^\tau \log[P(Y(u)e^{Z(\theta)})]dN(u)\right]$$

and suppose the true value of θ is θ_0 (i.e., the underlying probability measure that generates data $\{\delta_i, T_i\}_{i=1,\dots,n}$ corresponds to parameter θ_0), then by Lemma 2.2.4,

$\theta_0 = \text{argmax}_\theta \mathbb{M}(\theta)$. By this way, the MPLE of θ is put into the framework of the M-estimators.

In order to study the large sample properties of the MPLE for this model, we need to make some assumptions.

Assumptions 2.2.1.

1. *$Z(\cdot)$ is a 2-sided Brownian motion starting from θ (i.e. $Z(\cdot + \theta) - Z(\theta)$ is a 2-sided Brownian motion starting from 0) scaled by σ, i.e. $W(\cdot) \equiv \frac{Z(\cdot+\theta)-Z(\theta)}{\sigma}$ follows 2-sided standard Brownian motion (abbreviated as S.B.M. in the sequel) starting from 0. The trajectory of $\{Z(\tilde{\theta}) : 0 \le \tilde{\theta} \le \theta_M\}$ is observed.*

2. *$Z(\theta)$ is independent of the process $Z(\cdot + \theta) - Z(\theta)$ and satisfies $PZ^2(\theta) < \infty$, $Pe^{2Z(\theta)} < \infty$.*

3. *Both the distributions of T^0 and C depend on $Z(\theta)$ only; T^0 and C are conditionally independent given $Z(\theta)$.*

4. *$P(C > \tau | Z(\theta)) > 0$.*

5. *$\int_0^\tau \lambda_0(u) du < \infty$.*

The first assumption, i.e., the fractional Brownian motion assumption, does capture the fractal feature of many functional data observed in practice, e.g., gene expression data, financial data, and so on. However, the assumption is too strong to be fully satisfied for data collected in practice. We have two reasons to make this assumption. One is that it is convenient for mathematical and statistical handling. Another is that (fractional) Brownian motion is the most fundamental one in stochastic process which is the counterpart to the normal distribution in probability theory. So the fBm assumption is at least a good starting point to study the proposed model.

The second assumption is a technical one which is used to facilitate the proof. The first statement in the third assumption is the essential idea of the proposed model, the hazard of experiencing events depends on only one component of a functional covariate. The second statement, i.e., the conditional independence of event time and censoring time, is a widely used assumption in survival analysis. Both the forth and fifth assumptions are commonly used assumptions in survival analysis.

Notation

We will use the following notation in Chapter 2.2.

$$S(\theta, u) \equiv \mathbb{P}_n \left[Y(u) \exp(Z(\theta)) \right] = \frac{1}{n} \sum_{i=1}^{n} Y_i(u) \exp(Z_i(\theta)),$$

$$s(\theta, u) \equiv P \left[Y(u) \exp(Z(\theta)) \right] = P \left[Y(u) \exp(Z(\theta)) \right].$$

2.2.1 Main Results

In this subsection, we will summarize the large sample properties for the MPLE of θ and make some comments.

Relying on the empirical processes theory, the consistency, rate of convergence and asymptotic distribution are obtained for the MPLE of θ in this model.

Theorem 2.2.2. *Under Assumptions 2.2.1,* $\hat{\theta}_n \to_{P^*} \theta_0$, $n(\hat{\theta}_n - \theta_0) = O_P^*(1)$,

$$n(\hat{\theta}_n - \theta_0) \overset{\mathcal{W}}{\longrightarrow} \arg\max_h \left(W(h) - \frac{|h|}{2} \sigma \sqrt{PN(\tau)} \right),$$

where $W(\cdot)$ is a 2-sided standard Brownian motion starting from zero with unit variance scale (i.e., $W(1) =_d W(-1) \sim N(0, 1)$).

Comparing Theorem 2.2.2 above to Theorem 2.1 in McKeague and Sen [29], we find our Theorem 2.2.2 is similar to their result except for two aspects. One is there is a coefficient $PN(\tau)$ in front of the drift term in our Theorem 2.2.2; another is the σ instead of $1/\sigma$ appeared in the coefficient of the drift term in our Theorem 2.2.2. The reason for the second difference is simple: the same notation σ means different things in these two works.

Investigating the probability distribution function of $\text{argmax}_h \big(W(h) - \frac{1}{2} c|h| \big)$ for $c > 0$ finds that smaller c results in wider spread distribution for positive c.

Remark 2.2.3.

1. *Note that $PN(\tau)$ corresponds to the expected proportion of uncensored events among all subjects under the proposed model. The closer to 1 it is, the more information the data carries about the relationship between $Z(\theta)$ and event risk. Larger $PN(\tau)$ corresponds to larger c, and hence the asymptotic distribution of $\hat{\theta}_n$ is less spread out, which implies more information about $\hat{\theta}_n$. In the extreme case that $PN(\tau)$ approaches 1, which means all subjects experience events without being censored before the follow up endpoint τ, the asymptotic distribution of the MPLE (Maximum Partial Likelihood Estimator) $\hat{\theta}_n$ approaches its counterpart in the linear regression setting (if we do not consider the effect of σ for the moment). However, this limit scenario is not covered by our model since by our model assumption, $PN(\tau) = P(T^0 \leq \tau, T^0 \leq C) \leq P(T^0 \leq \tau) = 1 - P(T^0 > \tau) = 1 - P\left[\exp\left(-e^{Z(\theta_0)} \int_0^\tau \lambda_0(s)ds\right)\right] < 1$ which means there is always a non-ignorable proportion of subjects at risk by time τ.*

2. *Larger σ corresponds to larger c and hence less spread out asymptotic distribution for $\hat{\theta}_n$. Since σ^2 is the Brownian motion's infinitesimal variance, so larger σ means bigger difference between the value of $Z(\theta)$ and values of $Z(\cdot)$ (other than $Z(\theta)$), hence easier to extract θ out of $[0, \theta_M]$, which implies less spread out asymptotic distribution of $\hat{\theta}_n$.*

Denote $c = \sigma\sqrt{PN(\tau)}$, then the analytic formula of distribution function F of $\mathrm{argmax_h}\ \mathbb{Q}(h)$ can be obtained from Bhattacharya and Brockwell [5] and Henrik [46]. Its P.D.F. (probability distribution function) is symmetric about zero and can be written as

$$f(x) = \frac{3}{2}c^2 e^{c^2 x}\Phi(\frac{-3c\sqrt{x}}{2}) - \frac{1}{2}c^2\Phi(\frac{-c\sqrt{x}}{2})$$

for $x \geq 0$; its C.D.F. (cumulative distribution function) can be expressed as

$$F(x) = 1 + c\sqrt{\frac{x}{2\pi}}e^{\frac{-c^2 x}{8}} + \frac{3}{2}e^{c^2 x}\Phi(\frac{-3c\sqrt{x}}{2}) - \frac{c^2 x + 5}{2}\Phi(\frac{-c\sqrt{x}}{2})$$

for $x \geq 0$, where Φ is the C.D.F. of the standard normal distribution.

2.2.1.1 Non-identifiability of (θ_M, σ)

By the self-similarity of Brownian motion, the distributional property of Brownian motion $Z(\theta)$ on $\theta \in [0, \theta_M]$ with infinitesimal variance σ^2, is not distinguishable from that of Brownian motion $Z(\tilde{\theta})$ on $\tilde{\theta} \in [0, \tilde{\theta}_M]$ with infinitesimal variance $\tilde{\sigma}^2$, if

$$\theta_M \sigma^2 = \tilde{\theta}_M \tilde{\sigma}^2.$$

Thus to make model (2.1) identifiable, we can always set $\theta_M = 1$ during the parameter estimation process.

2.2.1.2 Wald-type confidence interval

By the asymptotic distribution obtained in Theorem 2.2.2, we can construct Wald-type confidence intervals for θ. Thanks to the analytical form of the asymptotic distribution's C.D.F. $F(x)$, its quantiles can be determined easily. With consistent estimates of σ and $PN(\tau)$, the Wald-type confidence interval of θ can be constructed.

2.2.2 Proofs

Without loss of generality, we assume $\sigma = 1$ and hence $W(\cdot) \equiv Z(\cdot + \theta) - Z(\theta)$ follows 2-sided S.B.M. starting from 0. The case for general $\sigma > 0$ can be deduced in exactly the same way.

2.2.2.1 Local quadratic property

To apply Corollary 3.2.3 (i) in [54] to prove consistency, we need to show θ_0 is the unique maximizer of $\mathbb{M}$ over $[0, \theta_M]$.

Lemma 2.2.4 (local quadratic property)**.** *Under Assumptions 2.2.1, there exists a metric* $d(\theta_1, \theta_2) = \sqrt{|\theta_1 - \theta_2|}$*, such that for any* $\theta \in [0, \theta_M]$*,* $\mathbb{M}(\theta) - \mathbb{M}(\theta_0) \lesssim -d^2(\theta, \theta_0)$*.*

Proof.

$$\mathbb{M}(\theta) - \mathbb{M}(\theta_0) = P\left[(Z(\theta) - Z(\theta_0))N(\tau)\right] - \int_0^\tau \log\left[\frac{P(Y(u)e^{Z(\theta)})}{P(Y(u)e^{Z(\theta_0)})}\right] P\left(dN(u)\right). \tag{2.2}$$

Note that $$N(u) = 1_{(T^0 \le C, T^0 \wedge C \le u)} = 1_{(T^0 \le C, T^0 \le u)} \quad \text{for} \quad 0 \le u \le \tau,$$

we have $$P\left[(Z(\theta) - Z(\theta_0))N(\tau)\right] = P\left(P\left[(Z(\theta) - Z(\theta_0))N(\tau)|\mathbb{Z}\right]\right)$$
$$= P\left[(Z(\theta) - Z(\theta_0))P\left(N(\tau)|\mathbb{Z}\right)\right] = P\left[(Z(\theta) - Z(\theta_0))P\left(1_{(T^0 \le C, T^0 \le \tau)}|\mathbb{Z}\right)\right].$$

By Assumptions 2.2.1, T^0 and C's marginal distributions only depend on $Z(\theta_0)$, and T^0 and C are conditionally independent given $Z(\theta_0)$, we have

$$P\left(1_{(T^0 \le C, T^0 \le \tau)}|\mathbb{Z}\right) = P\left(1_{(T^0 \le C, T^0 \le \tau)}|Z(\theta_0)\right).$$

Then $$P\left[(Z(\theta) - Z(\theta_0))N(\tau)\right] = P\left[(Z(\theta) - Z(\theta_0))P\left(1_{(T^0 \le C, T^0 \le \tau)}|Z(\theta_0)\right)\right]$$
$$= P\left[Z(\theta) - Z(\theta_0)\right] P\left[P\left(1_{(T^0 \le C, T^0 \le \tau)}|Z(\theta_0)\right)\right] = 0, \tag{2.3}$$

where we used the independence of $Z(\theta) - Z(\theta_0)$ and $Z(\theta_0)$ from Assumptions 2.2.1 in the second to last equality, and mean zero property of Brownian motion in the last equality.

On the other hand,

$$P\left[Y(u)e^{Z(\theta)}\right] = P\left[P(Y(u)e^{Z(\theta)}|\mathbb{Z})\right] = P\left[e^{Z(\theta)}P(Y(u)|\mathbb{Z})\right]$$
$$= P\left[e^{Z(\theta)}P(1_{\{T^0 \wedge C \ge u\}}|\mathbb{Z})\right] = P\left[e^{Z(\theta)}P(1_{\{T^0 \wedge C \ge u\}}|Z(\theta_0))\right]$$

by the assumptions that T^0 and C are conditionally independent given $Z(\theta_0)$, and that the marginal distributions of T^0 and C depend on only $Z(\theta_0)$ (out of the whole process $\mathbb{Z} = \{Z(\theta) : 0 \le \theta \le \theta_M\}$).

Decompose $e^{Z(\theta)}P(1_{\{T^0\wedge C\geq u\}}|Z(\theta_0))$ into the product of $e^{Z(\theta_0)}P(1_{\{T^0\wedge C\geq u\}}|Z(\theta_0))$ (depending on $Z(\theta_0)$) and $e^{Z(\theta)-Z(\theta_0)}$ (depending on $Z(\theta)-Z(\theta_0)$); due to the independence of $Z(\theta)-Z(\theta_0)$ and $Z(\theta_0)$,

$$\begin{aligned}P\left[Y(u)e^{Z(\theta)}\right] =&P\left[e^{Z(\theta_0)}P(1_{\{T^0\wedge C\geq u\}}|Z(\theta_0))\right]P\left[e^{Z(\theta)-Z(\theta_0)}\right]\\ =&P\left[Y(u)e^{Z(\theta_0)}\right]P\left[e^{Z(\theta)-Z(\theta_0)}\right]=P\left[Y(u)e^{Z(\theta_0)}\right]\cdot e^{\frac{|\theta-\theta_0|}{2}}.\end{aligned}$$

Here we used the property of Brownian motion $Z(\cdot+\theta_0)-Z(\theta_0)$.

Thus we obtain a relationship to be used frequently in the sequel,

$$P\left[Y(u)e^{Z(\theta)}\right]=e^{\frac{|\theta-\theta_0|}{2}}\cdot P\left[Y(u)e^{Z(\theta_0)}\right],\quad \text{for}\quad 0\leq\theta\leq\theta_M; \tag{2.4}$$

or equivalently,

$$s^{(0)}(\theta,u)=e^{\frac{|\theta-\theta_0|}{2}}\cdot s^{(0)}(\theta_0,u),\quad \text{for}\quad 0\leq\theta\leq\theta_M.$$

Suppose the underlying counting process for T^0 is $N^0(t)$, which is not always completely observed, in contrast to the counting process $N(t)$ for T. Then by the second paragraph on P.151 of Kalbfleisch and Prentice [19],

$$N(t)=\int_0^t Y(u)dN^0(u).$$

It follows that $dN(u)=Y(u)dN^0(u)$ and

$$\begin{aligned}P\left(dN(u)\right)&=P\left(Y(u)dN^0(u)\right)=P\left(1_{(T^0\wedge C\geq u)}d1_{(T^0\leq u)}\right)\\ &=P\left(1_{(C\geq u)}1_{(T^0\geq u)}d1_{(T^0\leq u)}\right)=P\left(1_{(C\geq u)}d1_{(T^0\leq u)}\right).\end{aligned}$$

The last equality holds since $d1_{(T^0\leq u)}=1$ if and only if $T^0=u$ which implies $T^0\geq u$

and otherwise $d1_{(T^0 \leq u)}$ is 0. It can further be written as

$$
\begin{aligned}
P\left(dN(u)\right) &= P\left(1_{(C \geq u)} d1_{(T^0 \leq u)}\right) \\
&= P\left[P\left(1_{(C \geq u)} d1_{(T^0 \leq u)} | Z(\theta_0)\right)\right] \\
&= P\left[P\left(C \geq u | Z(\theta_0)\right) dP\left(T^0 \leq u | Z(\theta_0)\right)\right] \\
&= P\left[P(C \geq u | Z(\theta_0)) \lambda_0(u) e^{Z(\theta_0)} e^{-\int_0^u \lambda_0(s) e^{Z(\theta_0)} ds} du\right] \\
&= \lambda_0(u) \cdot P\left[P(C \geq u | Z(\theta_0)) e^{Z(\theta_0)} e^{-\int_0^u \lambda_0(s) e^{Z(\theta_0)} ds} du\right], \qquad (2.5)
\end{aligned}
$$

where the third equality holds by the conditional independence of T^0 and C given $Z(\theta_0)$ and the forth equality holds by formula (2.11) of Kalbfleisch and Prentice [19].

Consider that

$$
\begin{aligned}
s^{(0)}(\theta_0, u) &= P\left[Y(u) e^{Z(\theta_0)}\right] = P\left[1_{(T^0 \wedge C \geq u)} e^{Z(\theta_0)}\right] \\
&= P\left(P\left[1_{(T^0 \geq u)} 1_{(C \geq u)} e^{Z(\theta_0)} | \mathbb{Z}\right]\right) \\
&= P\left(P\left[1_{(T^0 \geq u)} 1_{(C \geq u)} e^{Z(\theta_0)} | Z(\theta_0)\right]\right) \\
&= P\left[e^{Z(\theta_0)} \cdot P\left(T^0 \geq u | Z(\theta_0)\right) P\left(C \geq u | Z(\theta_0)\right)\right] \\
&= P\left[e^{Z(\theta_0)} e^{-\int_0^u \lambda_0(s) e^{Z(\theta_0)} ds} \cdot P\left(C \geq u | Z(\theta_0)\right)\right], \qquad (2.6)
\end{aligned}
$$

where we used the the conditional independence of T^0 and C given $Z(\theta_0)$ in the second to last equality and the property of hazard function in the last equality.

Compare (2.5) with (2.6), we obtain

$$
P\left(dN(u)\right) = \lambda_0(u) s^{(0)}(\theta_0, u) du. \qquad (2.7)
$$

To summarize (2.3), (2.4) and (2.7) and considering (2.2), we have

$$
\mathbb{M}(\theta) - \mathbb{M}(\theta_0) = -\frac{1}{2}|\theta - \theta_0| \int_0^\tau s^{(0)}(\theta_0, u) \lambda_0(u) du.
$$

Since $\int_0^\tau s^{(0)}(\theta_0, u)\lambda_0(u)du > 0$, check with the local quadratic condition in the rate of convergence theorem, i.e., Theorem 3.2.5 in VW [54], we can choose the metric $d(\theta, \theta_0) = \sqrt{|\theta - \theta_0|}$.

□

Remark 2.2.5. *: An alternative way to prove this lemma is to argue by the fact that* $N(t) - \int_0^t Y(u)\lambda_0(u)e^{Z(\theta_0)}du$ *is a local martingale whose distributional property depends only on* $Z(\theta_0)$ *and independent of* $Z(\theta) - Z(\theta_0)$.

Now by Lemma 2.2.4, it follows θ_0 is the unique maximizer of $\mathbb{M}$ over $[0, \theta_M]$, and $\mathbb{M}(\theta_0) > \sup_{\theta \notin G} \mathbb{M}(\theta)$, for every open set G that contains θ_0. If we can obtain the uniform convergence of $\mathbb{M}_n - \mathbb{M}$ to 0 in outer probability P^*, then by Corollary 3.2.3 of VW [54], the consistency of $\hat{\theta}_n$ is proved.

2.2.2.2 Consistency

Since $\mathbb{M}_n(\theta) - \mathbb{M}(\theta)$ can be decomposed into

$$\begin{aligned}\mathbb{M}_n(\theta) - \mathbb{M}(\theta) =& (\mathbb{P}_n - P)(Z(\theta)N(\tau)) + \int_0^\tau \log\left[\frac{\mathbb{P}_n Y(u)e^{Z(\theta)}}{PY(u)e^{Z(\theta)}}\right]\mathbb{P}_n dN(u) \\ &+ \int_0^\tau \log s^{(0)}(\theta, u)(\mathbb{P}_n - P)dN(u),\end{aligned}$$

it suffices to prove the uniform convergence to 0 in P^*-probability or in $L_1(P^*)$ of the three terms.

For the first term,

$$P^* \sup_{\theta \in [0, \theta_M]} |(\mathbb{P}_n - P)(Z(\theta)N(\tau))| = \frac{1}{\sqrt{n}} P^* \sup_{\theta \in [0, \theta_M]} |\mathbb{G}_n(Z(\theta)N(\tau))|,$$

where $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P)$.

Since $\{Z(\theta) : \theta \in \mathbb{R}\}$ is a 2-sided Brownian motion starting from θ_0, we can bound $P^* \sup_{\theta \in [0, \theta_M]} |\mathbb{G}_n(Z(\theta)N(\tau))|$ by

$$\begin{aligned}&P\left|\mathbb{G}_n[Z(\theta_0)N(\tau)]\right| + P^* \sup_{\theta \in [0, \theta_M]} \left|\mathbb{G}_n[(Z(\theta) - Z(\theta_0))N(\tau)]\right| \\ &\leq \left(P\left(\mathbb{G}_n[Z(\theta_0)N(\tau)]\right)^2\right)^{\frac{1}{2}} + P^* \sup_{t \in [-\theta_0, \theta_M - \theta_0]} \left|\mathbb{G}_n[W(t)N(\tau)]\right|,\end{aligned}$$

where $\{W(t) : W(t) \equiv Z(t + \theta_0) - Z(\theta_0), t \in \mathbb{R}\}$ is a 2-sided S.B.M. starting from 0.

It is easy to show

$$P\left(\mathbb{G}_n[Z(\theta_0)N(\tau)]\right)^2 = P[Z(\theta_0)N(\tau) - P(Z(\theta_0)N(\tau))]^2$$
$$\leq P[Z(\theta_0)N(\tau)]^2 \leq P[Z^2(\theta_0)],$$

and $P^* \sup_{t\in[-\theta_0,\theta_M-\theta_0]} |\mathbb{G}_n[W(t)N(\tau)]|$ can be controlled by

$$\begin{aligned} &P^* \sup_{t\in[-\theta_0,0]} |\mathbb{G}_n[W(t)N(\tau)]| + P^* \sup_{t\in[0,\theta_M-\theta_0]} |\mathbb{G}_n[W(t)N(\tau)]| \\ &= P^* \sup_{t\in[0,\theta_0]} |\mathbb{G}_n[W(t)N(\tau)]| + P^* \sup_{t\in[0,\theta_M-\theta_0]} |\mathbb{G}_n[W(t)N(\tau)]| \\ &\leq 2P^* \sup_{t\in[0,\theta_M]} |\mathbb{G}_n[W(t)N(\tau)]|, \end{aligned}$$

where the equality holds by the symmetry of 2-sided S.B.M. $W(\cdot)$.

Since $\{W(t) : 0 \leq t \leq \theta_M\}$ has bounded bracketing entropy by example 3.2.12 of [54] and the Lipschitz property of $W(t)$ (see the proof of Lemma 8.1 in [29]), the class formed by multiplying it to a function $N(\tau)$ which is bounded by 1, $\mathcal{M}_{WN,\theta_M} \equiv \{W(t)N(\tau) : 0 \leq t \leq \theta_M\}$, still has bounded bracketing entropy (by changing brackets from $[l_i, u_i]$ to $[l_i \cdot N(\tau), u_i \cdot N(\tau)]$, it follows that $J_{[]}(1, \mathcal{M}_{WN,\theta_M}, L_2(P)) < \infty$). Then by Theorem 2.14.2 of [54],

$$\begin{aligned} &P^* \sup_{t\in[0,\theta_M]} |\mathbb{G}_n[W(t)N(\tau)]| \\ &\leq J_{[]}(1, \mathcal{M}_{WN,\theta_M}, L_2(P)) \sqrt{P^* \sup_{t\in[0,\theta_M]} W^2(t)N^2(1)} \lesssim \sqrt{P^* \sup_{t\in[0,\theta_M]} W^2(\theta)}, \end{aligned}$$

with $\lesssim$ meaning $\leq$ up to a universal constant.

Since $\{W(t) : t \geq 0\}$ is a martingale, by Doob's maximal inequality (Theorem 2.1.7 in [38]),

$$P^* \sup_{t\in[0,\theta_M]} W^2(t) \leq 4 \sup_{t\in[0,\theta_M]} P[W^2(t)] = 4\theta_M.$$

So altogether, for the supremum of the first term in the decomposition of $\mathbb{M}_n(\theta) - \mathbb{M}(\theta)$, we have

$$P^* \sup_{\theta\in[0,\theta_M]} |(\mathbb{P}_n - P)(Z(\theta)N(\tau))| \lesssim \frac{1}{\sqrt{n}} \left(\sqrt{P\left(Z^2(\theta_0)\right)} + \sqrt{\theta_M} \right).$$

Obviously the first term in the decomposition of $\mathbb{M}_n(\theta) - \mathbb{M}(\theta)$ converges uniformly to 0 in $L_1(P^*)$ and hence in P^*-probability.

Now consider the second term,

$$\int_0^\tau \log \left[\frac{\mathbb{P}_n Y(u) e^{Z(\theta)}}{PY(u)e^{Z(\theta)}} \right] \mathbb{P}_n dN(u).$$

First we aim to prove $\sup_{\theta,u} \frac{|(\mathbb{P}_n - P)Y(u)e^{Z(\theta)}|}{PY(u)e^{Z(\theta)}}$ converge to 0 P^*-a.s.. Since

$$\begin{aligned} PY(u)e^{Z(\theta)} &\geq PY(\tau)e^{Z(\theta)} = e^{\frac{|\theta-\theta_0|}{2}} PY(\tau)e^{Z(\theta_0)} \geq PY(\tau)e^{Z(\theta_0)} \\ &= P\left[e^{Z(\theta_0)} e^{-\int_0^\tau \lambda_0(s) e^{Z(\theta_0)} ds} P(C > \tau | Z(\theta_0)) \right] \equiv B(\theta_0), \end{aligned} \tag{2.8}$$

where the first equality holds by (2.4) and the second equality holds by (2.6).

Considering the condition $P(C > \tau | Z(\theta_0)) > 0$ in Assumptions 2.2.1, $PY(u)e^{Z(\theta)}$ is bounded away from 0 by (2.8). Hence if $\{Y(u)e^{Z(\theta)} : u \in [0,\tau], \theta \in [0, \theta_M]\}$ is a P-Glivenko$-$Cantelli class, then

$$\lim_{n\to\infty} \sup_{\theta,u} \frac{|(\mathbb{P}_n - P)Y(u)e^{Z(\theta)}|}{PY(u)e^{Z(\theta)}} \leq \lim_{n\to\infty} \sup_{\theta,u} \frac{|(\mathbb{P}_n - P)Y(u)e^{Z(\theta)}|}{B(\theta_0)} = 0 \quad P^*\text{-a.s..} \tag{2.9}$$

Furthermore,

$$\begin{aligned} &\lim_{n\to\infty} \sup_{\theta\in[0,\theta_M]} \left| \int_0^\tau \log \left[\frac{\mathbb{P}_n Y(u)e^{Z(\theta)}}{PY(u)e^{Z(\theta)}} \right] \mathbb{P}_n dN(u) \right| \\ &\leq \lim_{n\to\infty} \sup_{\theta,u} \left| \log \frac{\mathbb{P}_n Y(u)e^{Z(\theta)}}{PY(u)e^{Z(\theta)}} \right| \mathbb{P}_n N(\tau) \leq \lim_{n\to\infty} \sup_{\theta,u} \left| \log \frac{\mathbb{P}_n Y(u)e^{Z(\theta)}}{PY(u)e^{Z(\theta)}} \right|. \end{aligned}$$

Since the supremum value is equal to either $\log\left(1 + \sup_{u,\theta} \frac{|(\mathbb{P}_n - P)Y(u)e^{Z(\theta)}|}{PY(u)e^{Z(\theta)}}\right)$ or $-\log\left(1 - \sup_{u,\theta} \frac{|(\mathbb{P}_n - P)Y(u)e^{Z(\theta)}|}{PY(u)e^{Z(\theta)}}\right)$, whose values both go to 0 P^*-a.s. as $n \to \infty$ (by continuous mapping theorem and (2.9)), so the P^*-a.s. uniform convergence of the second item in the decomposition of $\mathbb{M}_n(\theta) - \mathbb{M}(\theta)$ is obtained, then its uniform convergence in P^*-probability is proved.

Now it suffices to prove $\{Y(u)e^{Z(\theta)} : u \in [0,\tau], \theta \in [0, \theta_M]\}$ is a P-Glivenko$-$Cantelli class. According to page 82 of VW [54], every Donsker class is a Glivenko$-$Cantelli class almost surely. The P-Glivenko−Cantelli property of the class $\{Y(u)e^{Z(\theta)} : u \in [0,\tau], \theta \in [0,\theta_M]\}$ follows from Lemma A.0.1 in Appendix A and the first Donsker theorem on page 85 of VW [54].

For the third term $\int_0^\tau \log s^{(0)}(\theta,u)(\mathbb{P}_n - P)dN(u)$,

$$P^* \sup_{\theta\in[0,\theta_M]} \left| \int_0^\tau \log s^{(0)}(\theta,u)(\mathbb{P}_n - P)dN(u) \right| = \frac{1}{\sqrt{n}} P^* \sup_{\theta\in[0,\theta_M]} |\mathbb{G}_n f_\theta| ,$$

where $\quad f_\theta = 1_{(T\le C)} 1_{(0<T\le\tau)} \log s^{(0)}(\theta,T) - e^{Z(\theta_0)} \int_0^\tau 1_{(T\ge u)} \lambda_0(u) \log s^{(0)}(\theta,u)du.$

Since $s^{(0)}(\theta,u) = e^{\frac{|\theta-\theta_0|}{2}} s^{(0)}(\theta_0,u)$ by (2.4),

$$\left| \log \frac{s^{(0)}(\theta_1,u)}{s^{(0)}(\theta_2,u)} \right| = 1/2 \, ||\theta_1 - \theta_0| - |\theta_2 - \theta_0|| \le 1/2|\theta_1 - \theta_2|,$$

then $\quad |f_{\theta_1} - f_{\theta_2}|$

$$\le \left| 1_{(T\le C)} 1_{(0<T\le\tau)} \log \frac{s^{(0)}(\theta_1,T)}{s^{(0)}(\theta_2,T)} \right| + \left| e^{Z(\theta_0)} \int_0^\tau 1_{(T\ge u)} \lambda_0(u) \log \frac{s^{(0)}(\theta_1,u)}{s^{(0)}(\theta_2,u)} du \right|$$

$$\le 1/2|\theta_1 - \theta_2| 1_{(T\le C)} 1_{(0<T\le\tau)} + 1/2|\theta_1 - \theta_2| e^{Z(\theta_0)} \int_0^\tau 1_{(T\ge u)} \lambda_0(u)du$$

$$\le 1/2 \left(1 + e^{Z(\theta_0)} \Lambda_0(\tau) \right) |\theta_1 - \theta_2| \equiv L_f \cdot |\theta_1 - \theta_2|,$$

where $\Lambda_0(\tau) = \int_0^\tau \lambda_0(u)du$ and $L_f \equiv 1/2 \left(1 + e^{Z(\theta_0)}\Lambda_0(\tau)\right)$.

Since $P(L_f^2) \le \left(1 + \Lambda_0^2(\tau)P[e^{2Z(\theta_0)}]\right) < \infty$ by the conditions $\int_0^\tau \lambda_0(u)du < \infty$ and $P[e^{2Z(\theta_0)}] < \infty$ in Assumptions 2.2.1, then f_θ is "Lipschitz in parameter" and hence $J_{[]}(1, \mathcal{M}_f, L_2(P)) < \infty$ (p.294 of [54]), where $\mathcal{M}_f = \{f_\theta, \theta \in [0,\theta_M]\}$. Now we prove its envelope function, $\sup_{\theta\in[0,\theta_M]} f_\theta^2$, has finite second moment. Due to the Lipschitz property, we have

$$\left| \sup_{\theta\in[0,\theta_M]} f_\theta \right| \le |f_{\theta_0}| + L_f \cdot \theta_M,$$

$$P \sup_{\theta\in[0,\theta_M]} f_\theta^2 \lesssim \theta_M^2 \cdot P \sup_{\theta\in[0,\theta_M]} (L_f^2) + P f_{\theta_0}^2. \tag{2.10}$$

On one hand,

$$P \sup_{\theta\in[0,\theta_M]} (L_f^2) \leq \theta_M^2 \left(1 + \Lambda_0^2(\tau) P e^{2Z(\theta_0)}\right) < \infty;$$

on the other hand,

$$\begin{aligned} P f_{\theta_0}^2 &\leq \sup_u \left|\log s^{(0)}(\theta_0, u)\right|^2 \cdot P\left(1 + \Lambda_0(\tau) e^{Z(\theta_0)}\right)^2 \\ &\leq 2 \sup_u \left|\log P\left(Y(u) e^{Z(\theta_0)}\right)\right|^2 \cdot P\left(1 + \Lambda_0^2(\tau) e^{2Z(\theta_0)}\right) \\ &\leq 2\left[\left|\log P e^{Z(\theta_0)}\right|^2 + \left|\log P Y(\tau) e^{Z(\theta_0)}\right|^2\right] \cdot P\left(1 + \Lambda_0^2(\tau) e^{2Z(\theta_0)}\right) < \infty, \end{aligned}$$

where we used $Y(\tau) \leq Y(u) \leq 1$ in the third inequality and $0 < P(e^{Z(\theta_0)}) < \infty$, $0 < P(Y(\tau)e^{Z(\theta_0)}) < \infty$ which follow from Assumptions 2.2.1 in the last inequality.

Since both $P\sup_{\theta\in[0,\theta_M]}(L_f^2)$ and $Pf_{\theta_0}^2$ are finite, summarizing them and considering (2.10), we obtained that

$$P \sup_{\theta\in[0,\theta_M]} f_\theta^2 < \infty.$$

Then by Theorem 2.14.2 of [54],

$$\frac{1}{\sqrt{n}} P^* \sup_{\theta\in[0,\theta_M]} |\mathbb{G}_n f_\theta| \leq \frac{1}{\sqrt{n}} J_{[]}(1, \mathcal{M}_f, L^2(P)) \sqrt{P \sup_{\theta\in[0,\theta_M]} f_\theta^2} \lesssim \frac{1}{\sqrt{n}}.$$

It follows that the third term in the decomposition of $\mathbb{M}_n(\theta) - \mathbb{M}(\theta)$ converges uniformly to zero in $L_1(P^*)$ and hence in P^*-probability as $n \to \infty$.

So to summarize the results in this section, as $n \to \infty$, the summation of the three terms converges to 0 in P^*-probability uniformly over $\theta \in [0, \theta_M]$, i.e.,

$$\sup_{\theta\in[0,\theta_M]} |\mathbb{M}_n(\theta) - \mathbb{M}(\theta)| \to 0 \quad \text{in} \quad P^*\text{-probability}.$$

We had proved that θ_0 is the unique maximizer of $\mathbb{M}$ over $[0, \theta_M]$, and $\mathbb{M}(\theta_0) > \sup_{\theta\notin G} \mathbb{M}(\theta)$, for every open set G that contains θ_0 in Chapter 2.2.2.1. Now we have proved the uniform convergence of $\mathbb{M}_n - \mathbb{M}$ to 0 in outer probability P^* in this section, so the consistency of $\hat{\theta}_n$ follows by Corollary 3.2.3 of VW [54].

2.2.2.3 Rate of Convergence

In Chapter 2.2.2.1, we already found a metric d, such that

$$\mathbb{M}(\theta) - \mathbb{M}(\theta_0) \lesssim -d^2(\theta, \theta_0).$$

The next step would be to find out a suitable function $\phi_n(\delta)$ which can satisfy the modulus of continuity condition in the rate of convergence theorem. Then by this theorem, an upper bound for the convergence rate $\tilde{r}_n$ of $\hat{\theta}_n$ is established.

The proof of this part is too techinical and lengthy, so we put it into Appendix A. In Appendix A, we finally get an upper bound for the rate of converge of $\hat{\theta}_n$, $\tilde{r}_n = n$.

2.2.2.4 Asymptotics of rescaled criterion function

Consider $\tilde{r}_n(\hat{\theta}_n - \theta_0)$, and rewrite it as $\tilde{r}_n(\hat{\theta}_n - \theta_0) = \hat{h}_n = \mathrm{argmax}_{\mathrm{h}\in\mathbb{R}}\ \mathbb{Q}_n(h)$.

Write $\mathbb{Q}_n(h)$ as the form of

$$\mathbb{Q}_n(h) = s_n\left(\mathbb{M}_n\left(\theta_0 + \frac{h}{\tilde{r}_n}\right) - \mathbb{M}_n(\theta_0)\right).$$

If the uniform weak convergence of $\mathbb{Q}_n(h)$ can be established, i.e. $\mathbb{Q}_n \xrightarrow{\mathcal{W}} \mathbb{Q}$, then apply the Argmax Continuous Mapping Theorem, the asymptotic distribution of $\hat{h}_n$ can be established, $\hat{h}_n \xrightarrow{\mathcal{W}} \mathrm{argmax}_{\mathrm{h}}\ \mathbb{Q}(h)$. So $\tilde{r}_n(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{W}} \mathrm{argmax}_{\mathrm{h}}\ \mathbb{Q}(h)$. The asymptotic distribution of $\hat{\theta}_n$ is established.

Now let's consider $\mathbb{Q}_n(h) = s_n(\mathbb{M}_n(\theta_0 + h/\tilde{r}_n) - \mathbb{M}_n(\theta_0)), \forall h \in [-K, K], \forall K > 0$, where we use $s_n = n$ and $\tilde{r}_n = n$.

$$\begin{aligned}
\mathbb{Q}_n(h) &= n(\mathbb{M}_n(\theta_0 + h/n) - \mathbb{M}_n(\theta_0)) \\
&= n\left(\mathbb{P}_n\left[\int_0^\tau Z(\theta_0 + \frac{h}{n})dN(s) - \int_0^\tau \log[\mathbb{P}_n Y(u)e^{Z(\theta_0+\frac{h}{n})}]dN(u)\right]\right. \\
&\qquad\qquad \left. -\mathbb{P}_n\left[\int_0^\tau Z(\theta_0)dN(s) - \int_0^\tau \log[\mathbb{P}_n Y(u)e^{Z(\theta_0)}]dN(u)\right]\right) \\
&= n\mathbb{P}_n\int_0^\tau \left[Z(\theta_0 + h/n) - Z(\theta_0)\right]dN(s) - n\int_0^\tau \log\left[\frac{\mathbb{P}_n Y(u)e^{Z(\theta_0+\frac{h}{n})}}{\mathbb{P}_n Y(u)e^{Z(\theta_0)}}\right]\mathbb{P}_n dN(u).
\end{aligned}$$

If we rewrite it using the S.B.M. $W(t) \equiv Z(t+\theta_0) - Z(\theta_0)$, then

$$\mathbb{Q}_n(h) = n\mathbb{P}_n W(h/n)N(\tau) - n\int_0^\tau \log\left[1 + \frac{\mathbb{P}_n Y(u)e^{Z(\theta_0)}\left(e^{W(\frac{h}{n})}-1\right)}{\mathbb{P}_n Y(u)e^{Z(\theta_0)}}\right]\mathbb{P}_n dN(u).$$

By similar arguments to that on $\mathcal{M}_{Z,\delta}$ in Appendix A,

$$\lim_{n\to\infty}\sup_{u\in[0,\tau]}|(\mathbb{P}_n - P)Y(u)e^{Z(\theta_0)}| = 0, \quad P^*\text{-a.s.},$$

$$\lim_{n\to\infty}\sup_{|h|\le K, u\in[0,\tau]}\left|(\mathbb{P}_n - P)Y(u)e^{Z(\theta_0)}\left(e^{W(\frac{h}{n})}-1\right)\right| = 0, \quad P^*\text{-a.s.}.$$

Then

$$\mathbb{P}_n Y(u)e^{Z(\theta_0)}\left(e^{W(\frac{h}{n})}-1\right) = PY(u)e^{Z(\theta_0)}\left(e^{W(\frac{h}{n})}-1\right) + o_{uP}(1) = o_{uP}(1),$$
$$\mathbb{P}_n Y(u)e^{Z(\theta_0)} = PY(u)e^{Z(\theta_0)} + o_{uP}(1).$$

Notice that $PY(u)e^{Z(\theta_0)}$ is bounded away from zero by (2.8), it follows that

$$\frac{\mathbb{P}_n Y(u)e^{Z(\theta_0)}\left(e^{W(\frac{h}{n})}-1\right)}{\mathbb{P}_n Y(u)e^{Z(\theta_0)}} = o_{uP}(1),$$

where $A_{h,n} = o_{uP}(1)$ means $A_{h,n} = o_P(1)$ uniformly over $(h,u) \in [-K,K]\times[0,\tau]$.

Since by Taylor expansion, $\log(1+x) = x + o(x) = x(1+o(1))$ as $x \to 0$,

$$\log\left[1 + \frac{\mathbb{P}_n Y(u)e^{Z(\theta_0)}\left(e^{W(\frac{h}{n})}-1\right)}{\mathbb{P}_n Y(u)e^{Z(\theta_0)}}\right] = \frac{\mathbb{P}_n Y(u)e^{Z(\theta_0)}\left(e^{W(\frac{h}{n})}-1\right)}{\mathbb{P}_n Y(u)e^{Z(\theta_0)}}[1+o_{uP}(1)]$$
$$= \frac{\mathbb{P}_n Y(u)e^{Z(\theta_0)}\left(e^{W(\frac{h}{n})}-1\right)}{PY(u)e^{Z(\theta_0)} + o_{uP}(1)}[1+o_{uP}(1)] = \frac{\mathbb{P}_n Y(u)e^{Z(\theta_0)}\left(e^{W(\frac{h}{n})}-1\right)}{PY(u)e^{Z(\theta_0)}}[1+o_{uP}(1)].$$

Since in this section, we are only interested in the asymptotics, we can omit those

$o_{uP}(1)$ terms. Then $\mathbb{Q}_n(h)$ can be decomposed as

$$\text{So} \quad \mathbb{Q}_n(h) = n\mathbb{P}_n W(h/n)N(\tau) - n\int_0^\tau \frac{\mathbb{P}_n Y(u)e^{Z(\theta_0)}\left(e^{W(h/n)}-1\right)}{PY(u)e^{Z(\theta_0)}}\mathbb{P}_n dN(u)$$
$$= n\mathbb{P}_n W(\frac{h}{n})N(\tau) - n\int_0^\tau \frac{\mathbb{P}_n Y(u)e^{Z(\theta_0)}W(\frac{h}{n})}{PY(u)e^{Z(\theta_0)}}\mathbb{P}_n dN(u)$$
$$- n\int_0^\tau \frac{\mathbb{P}_n Y(u)e^{Z(\theta_0)}\left(e^{W(\frac{h}{n})}-1-W(\frac{h}{n})\right)}{PY(u)e^{Z(\theta_0)}}\mathbb{P}_n dN(u)$$

We have transformed the nonlinear log function into a linear term plus a remainder term, which make it easier to utilize the empirical process tools to prove the asymptotic properties.

On one hand,

$$n\mathbb{P}_n W(h/n)N(\tau) - n\int_0^\tau \frac{\mathbb{P}_n Y(u)e^{Z(\theta_0)}W(\frac{h}{n})}{PY(u)e^{Z(\theta_0)}}\mathbb{P}_n dN(u)$$
$$=_d \sqrt{n}\mathbb{P}_n W(h)N(\tau) - \sqrt{n}\int_0^\tau \frac{\mathbb{P}_n Y(u)e^{Z(\theta_0)}W(h)}{PY(u)e^{Z(\theta_0)}}\mathbb{P}_n dN(u)$$
$$= \sqrt{n}\mathbb{P}_n W(h)N(\tau) - \sqrt{n}\int_0^\tau \frac{PY(u)e^{Z(\theta_0)}W(h) + o_{uP}(1)}{PY(u)e^{Z(\theta_0)}}\mathbb{P}_n dN(u)$$
$$= \sqrt{n}\mathbb{P}_n W(h)N(\tau) - \sqrt{n}\int_0^\tau \frac{0 + o_{uP}(1)}{PY(u)e^{Z(\theta_0)}}\mathbb{P}_n dN(u)$$
$$= \sqrt{n}\mathbb{P}_n W(h)N(\tau) - o_{uP}(1)\sqrt{n}\int_0^\tau \frac{\mathbb{P}_n dN(u)}{PY(u)e^{Z(\theta_0)}}$$
$$= \sqrt{n}\mathbb{P}_n W(h)N(\tau) - o_{uP}(1)\cdot O_{uP}(1) = \sqrt{n}\mathbb{P}_n W(h)N(\tau) - o_{uP}(1),$$

where we used the self-similarity property of Brownian motion in the first equality (in distribution), the independence between $(Y(u), Z(\theta_0))$ and $W(h) \equiv Z(\theta_0+h) - Z(\theta_0)$, and the mean-zero property of Brownian motion $W(h)$ in the third equality.

On the other hand,

$$n\int_0^\tau \frac{\mathbb{P}_n Y(u)e^{Z(\theta_0)}\left(e^{W(\frac{h}{n})}-1-W(\frac{h}{n})\right)}{PY(u)e^{Z(\theta_0)}}\mathbb{P}_n dN(u)$$
$$= \sqrt{n}\int_0^\tau \frac{\mathbb{P}_n Y(u)e^{Z(\theta_0)}\sqrt{n}\left(e^{W(\frac{h}{n})}-1-W(\frac{h}{n})\right)}{PY(u)e^{Z(\theta_0)}}\mathbb{P}_n dN(u)$$

$$
\begin{aligned}
=&\sqrt{n}\int_0^\tau \frac{PY(u)e^{Z(\theta_0)}\sqrt{n}\left(e^{W(\frac{h}{n})}-1-W(\frac{h}{n})\right)+o_{uP}(1)}{PY(u)e^{Z(\theta_0)}}\mathbb{P}_n dN(u)\\
=&\sqrt{n}\int_0^\tau \frac{PY(u)e^{Z(\theta_0)}\sqrt{n}P\left(e^{W(\frac{h}{n})}-1-W(\frac{h}{n})\right)+o_{uP}(1)}{PY(u)e^{Z(\theta_0)}}\mathbb{P}_n dN(u)\\
=&\sqrt{n}\int_0^\tau \frac{PY(u)e^{Z(\theta_0)}\sqrt{n}\left(e^{\frac{|h|}{2n}}-1\right)+o_{uP}(1)}{PY(u)e^{Z(\theta_0)}}\mathbb{P}_n dN(u)\\
=&\int_0^\tau \left[\sqrt{n}\left(e^{\frac{|h|}{2n}}-1\right)+o_{uP}(1)\sqrt{n}\right]\mathbb{P}_n dN(u)\\
=&n\left(e^{\frac{|h|}{2n}}-1\right)\mathbb{P}_n N(\tau)+o_{uP}(1)\sqrt{n}\mathbb{P}_n N(\tau)\\
=&n\left(e^{\frac{|h|}{2n}}-1\right)\mathbb{P}_n N(\tau)+o_{uP}(1)\cdot O_{uP}(1)=n\left(e^{\frac{|h|}{2n}}-1\right)\mathbb{P}_n N(\tau)+o_{uP}(1)\\
=&n\left(e^{\frac{|h|}{2n}}-1\right)(PN(\tau)+o_{uP}(1))+o_{uP}(1)=n\left(e^{\frac{|h|}{2n}}-1\right)PN(\tau)+o_{uP}(1),
\end{aligned}
$$

where we used the Glivenko−Cantelli property of the class of functions

$$\{Y(u)e^{Z(\theta_0)}\sqrt{n}\left(e^{W(\frac{h}{n})}-1-W(h/n)\right): u\in[0,\tau], h\in[-\theta_0,\theta_M-\theta_0]\}$$

in the second equality, the independence between $(Y(u),Z(\theta_0))$ and $W(h)\equiv Z(\theta_0+h)-Z(\theta_0)$ in the third equality, and the property of Brownian motion $W(h)$ in the forth equality. The Glivenko−Cantelli property of the class of functions

$$\{Y(u)e^{Z(\theta_0)}\sqrt{n}\left(e^{W(\frac{h}{n})}-1-W(h/n)\right): u\in[0,\tau], h\in[-\theta_0,\theta_M-\theta_0]\}$$

can be proved similarly to that of $\mathcal{M}$ in Lemma A.0.1 in Appendix A. We omit it here.

Then putting them together, we have

$$\mathbb{Q}_n(h)=_d\sqrt{n}\mathbb{P}_n W(h)N(\tau)-n\left(e^{\frac{|h|}{2n}}-1\right)PN(\tau)+o_{uP}(1).$$

For the first term,

$$
\begin{aligned}
\sqrt{n}\mathbb{P}_n W(h)N(\tau)&=\sqrt{\frac{\sum_{i=1}^n N_i(\tau)}{n}}\frac{1}{\sqrt{\sum_{i=1}^n N_i(\tau)}}\sum_{i=1}^n W_i(h)N_i(\tau)\\
&=\sqrt{\mathbb{P}_n N(\tau)}\sqrt{\sum_{i=1}^n N_i(\tau)}\mathbb{P}_{\sum_{i=1}^n N_i(\tau)}W(h).
\end{aligned}
$$

By the independence of $(W_i(\cdot), N_i(\tau))$ $(i = 1, .., n)$ and properties of Brownian motion,

$$\sqrt{\sum_{i=1}^{n} N_i(\tau)}\mathbb{P}_{\sum_{i=1}^{n} N_i(\tau)} W(h) =_d W(h).$$

Since $\lim_{n\to\infty} \mathbb{P}_n N(\tau) = PN(\tau)$, then

$$\lim_{n\to\infty} \sqrt{n}\mathbb{P}_n W(h)N(\tau) =_d \lim_{n\to\infty} \sqrt{\mathbb{P}_n N(\tau)} W(h) = \sqrt{PN(\tau)} \cdot W(h).$$

For the second term, we have

$$\lim_{n\to\infty} \sup_{h\in[-K,K]} \left| n\left(e^{\frac{|h|}{2n}} - 1\right) PN(\tau) - \frac{|h|}{2} PN(\tau) \right| = 0,$$

Summing them up, we have $\mathbb{Q}_n(h)$ converges uniformly to the process $\sqrt{PN(\tau)} \cdot W(h) - \frac{|h|}{2} PN(\tau)$.

2.2.2.5 Asymptotic distribution of estimator $\hat{\theta}_n$

By the *Argmax Continuous Mapping Theorem* (i.e., Theorem 3.2.2 of VW [54]), $\hat{h}_n \xrightarrow{\mathcal{W}} \mathrm{argmax_h}\ \mathbb{Q}(h)$. So

$$n(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{W}} \mathrm{argmax}_h \mathbb{Q}(h).$$

Because

$$\begin{aligned} \mathrm{argmax}_h \mathbb{Q}(h) &= \mathrm{argmax}_h \left[\sqrt{PN(\tau)} \cdot W(h) - \frac{|h|}{2} PN(\tau) \right] \\ &= \mathrm{argmax}_h \left(W(h) - \frac{|h|}{2} \sqrt{PN(\tau)} \right), \end{aligned}$$

the asymptotic distribution of $\hat{\theta}_n$ can be established.

For the case of general $\sigma > 0$, following the lines throughout Chapter 2.2.2, we will obtain $\mathbb{Q}_n(h)$ converges uniformly to the process $\mathbb{Q}(h) = \sqrt{PN(\tau)} \cdot \sigma W(h) - \frac{|h|}{2}\sigma^2 PN(\tau)$, and

$$\begin{aligned} \mathrm{argmax}_h \mathbb{Q}(h) &= \mathrm{argmax}_h \left[\sqrt{PN(\tau)} \cdot \sigma W(h) - \frac{|h|}{2} \sigma^2 PN(\tau) \right] \\ &= \mathrm{argmax}_h \left(W(h) - \frac{|h|}{2} \sigma \sqrt{PN(\tau)} \right), \end{aligned}$$

where $W(\cdot)$ is a 2-sided standard Brownian motion starting from zero with unit variance scale (i.e.$W(1) =_d W(-1) \sim N(0,1)$).

2.3 Extended model

In this section, we extend the simplified model to the general model, i.e., the full model. The extensions are in two aspects. First, the trajectories of the functional covariate follow fractional Brownian motion with Hurst parameter H instead of Brownian motion. Such an extension from B.M. to fBm allows the functional covariate to have varied roughness, i.e., the extended model covers a wider range of functional covariate types. Second, the extended model allows for other non-functional covariates besides the functional covariate. Using the extended model, we can study the the effects of both the functional covariate and other non-functional covariates together in the Cox model.

These two extensions make the proposed model more applicable in data analysis. The extended model has more flexibility, however, deriving the large sample properties of its estimators is more challenging.

We will present the setting of the model, then state the model assumptions and study the large sample properties of the proposed estimators based on these assumptions. Most proofs are relegated to the last subsection, and some others are put into Appendix B.

Simulations to evaluate the finite sample performance of the estimators for the extended model can be found in Chapter 3.2.

2.3.1 Extended model setting

Consider a more complicated working model:

$$\lambda(t|\mathbf{Z};\mathbf{X}) = \lambda_0(t)\exp\{\beta Z(\theta) + \gamma^T\mathbf{X}\}.$$

Even though we observe a functional covariate, i.e., the realized trajectory of a stochastic process $\{Z(\tilde{\theta}) : \tilde{\theta} \in [0, \theta_M]\}$ for each subject, only one common location θ (shared by every subject) on the trajectories predicts the subjects' risk of experiencing the event of interest.

Note that in practice, the realized trajectories are not always observed continuously. They are commonly observed on grids that are fine enough instead. Besides the functional predictor $\{Z(\tilde{\theta}) : \tilde{\theta} \in [0, \theta_M]\}$, other non-functional covariates $\mathbf{X}$ are also included in the model.

By estimating the parameters in this model, we make a sparse selection of the functional predictor $\{Z(\tilde{\theta}) : \tilde{\theta} \in [0, \theta_M]\}$. After the sparse selection, only one element of the functional predictor, $Z(\theta)$ $(0 \leq \theta \leq \theta_M)$, is retained in the model. To better understand this model, refer to Chapter 1.4 for more explanations.

Parameter estimation for this model is based on the partial likelihood principle. For every grid $\tilde{\theta}$ on the trajectory $\{Z(\tilde{\theta}) : \tilde{\theta} \in [0, \theta_M]\}$, we can treat $Z(\tilde{\theta})$ as a non-functional covariate and fit the Cox proportional hazards model using $\Big(Z(\tilde{\theta}), \mathbf{X}\Big)$. As a byproduct of the model fitting process, we obtain the partial likelihood value for the fitted Cox model. We do this for every grid on the trajectory. Compare their partial likelihood values to find the grid that has the maximum partial likelihood value. This grid is the maximum partial likelihood estimator of the sensitive location on the trajectory. The corresponding fitted Cox model based on this chosen grid is the final Cox model estimated from data. The estimates of (β, γ) can be obtained within this fitted model.

This idea of parameter estimation is simple and can be applied to any stochastic process $\mathbf{Z}$ and other non-functional covariates $\mathbf{X}$. However, to derive the large sample properties of the maximum partial likelihood estimators, we have to make some assumptions.

Before we state the model assumptions, to make notation simple in this chapter, we include only one non-functional covariate, i.e., X, in the model. We also give a subscript H to $Z(\theta)$, which will be referred to in the following Assumptions.

$$\lambda(t|\mathbf{Z}_H, \mathbf{X}) = \lambda_0(t) \exp(\beta Z_H(\theta) + \gamma X).$$

All the theoretical results based on this model with one non-functional covariate can be extended to the model with multiple non-functional covariates.

All the model assumptions are almost the same as those of Chapter 2.2 besides those adjustments to adapt to the model complexity.

Assumptions 2.3.1.

1. *$Z_H(\cdot)$ is a 2-sided fractional Brownian motion (abbreviated as fBm in the sequel) with Hurst parameter H starting from θ scaled by σ, i.e. $W_H(\cdot) \equiv \frac{Z_H(\cdot+\theta)-Z_H(\theta)}{\sigma}$ follows 2-sided standard fBm with Hurst parameter H starting from 0. The trajectory of $\{Z_H(\tilde{\theta}) : 0 \leq \tilde{\theta} \leq \theta_M\}$ is observed.*

2. *$(Z_H(\theta), X)$ is independent of the process $Z_H(\cdot+\theta)-Z_H(\theta)$, $Z_H(\theta)$ is independent of X, and they satisfy $Pe^{2\beta Z_H(\theta)} < \infty$, $Pe^{2\gamma X} < \infty$.*

3. *Both the distributions of T^0 and C depend on $(Z_H(\theta), X)$ only; T^0 and C are conditionally independent given $(Z_H(\theta), X)$.*

4. *$P[Z_H^2(\tilde{\theta})e^{\tilde{\beta}Z_H(\tilde{\theta})+\tilde{\gamma}X}] < \infty$, $P[X^2 e^{\tilde{\beta}Z_H(\tilde{\theta})+\tilde{\gamma}X}] < \infty$ for all $(\tilde{\beta}, \tilde{\gamma}, \tilde{\theta}) \in [-\beta_M, \beta_M] \times [-\gamma_M, \gamma_M] \times [0, \theta_M]$.*

5. *$P(C > \tau | Z_H(\theta), X) > 0$.*

6. *$\int_0^\tau \lambda_0(u)du < \infty$.*

7. *$0 < |\beta| \leq \beta_M, |\gamma| \leq \gamma_M$.*

All these assumptions are similar to their counterparts in the simplified model except for some necessary extensions. Notice that in the second assumption, we need the independence of $Z(\theta)$ and X. There are two reasons to have this assumption. The first is it will make the model simpler and the investigation of the theoretical properties easier. The second is that if the correlation between $Z(\theta)$ and X is high, then the variable $Z(\theta)$ already includes part of the information about X; hence adding X into the model will not bring much more information. So to assume the independence of $Z(\theta)$ and X is reasonable.

Note that by the seventh assumption, γ can be zero in this model but β can not be zero. If $\beta = 0$, not only the parameter θ in this model can not be identified, but also the model will lose its key feature proposed by this thesis, becoming the classical Cox model.

Notation

We will assume the true value of (θ, β, γ) is $(\theta_0, \beta_0, \gamma_0)$ and use the following notation in this chapter. Denote $\pi \equiv (\beta, \gamma, \theta), N(u) = 1_{\delta=1, T^0 \wedge C \leq u}, Y(u) = 1_{T^0 \wedge C \geq u}$. For any $\pi \in [-\beta_M, \beta_M] \times [-\gamma_M, \gamma_M] \times [0, \theta_M]$, we set

$$m_\pi(T^0, C, \delta) = (\beta Z_H(\theta) + \gamma X)N(\tau) - \int_0^\tau \log\left[PY(u)\exp(\beta Z_H(\theta) + \gamma X)\right] dN(u),$$

$$\mathbb{M}(\pi) = P\left[(\beta Z_H(\theta) + \gamma X)N(\tau) - \int_0^\tau \log\left[PY(u)\exp(\beta Z_H(\theta) + \gamma X)\right] dN(u)\right],$$

$$\mathbb{M}_n(\pi) = \mathbb{P}_n\left[(\beta Z_H(\theta) + \gamma X)N(\tau) - \int_0^\tau \log\left[\mathbb{P}_n Y(u)\exp(\beta Z_H(\theta) + \gamma X)\right] dN(u)\right],$$

$$U(\beta, \gamma, \theta, u) = Y(u)\exp(\beta Z_H(\theta) + \gamma X),$$

$$F(\beta, \gamma, u) = \exp\left(\beta Z_H(\theta_0) + \gamma X - e^{\beta_0 Z_H(\theta_0) + \gamma_0 X}\int_0^u \lambda_0(s)ds\right) \cdot P\left(C \geq u | Z_H(\theta_0), X\right),$$

$$f(\beta, \gamma, u) = \log F(\beta, \gamma, u),$$

$$S(\pi, u) \equiv \mathbb{P}_n\left[Y(u)\exp(\beta Z_H(\theta) + \gamma X)\right] = \frac{1}{n}\sum_{i=1}^n Y_i(u)\exp(\beta Z_{H,i}(\theta) + \gamma X_i),$$

$$s(\pi, u) \equiv P\left[Y(u)\exp(\beta Z_H(\theta) + \gamma X)\right] = P\left[Y(u)\exp(\beta Z_H(\theta) + \gamma X)\right].$$

2.3.2 Main Results

Theorem 2.3.2. *Under Assumptions 2.3.1, for $H \in [1/2, 1)$,*

$$\hat{\beta}_n \to_{P^*} \beta_0, \quad \hat{\gamma}_n \to_{P^*} \gamma_0, \quad \hat{\theta}_n \to_{P^*} \theta_0.$$

With additional moment conditions B.2.1 in Appendix B.2 satisfied, for $H \in [1/2, 1)$,

$$\sqrt{n}(\hat{\beta}_n - \beta_0) = O_P^*(1), \quad \sqrt{n}(\hat{\gamma}_n - \gamma_0) = O_P^*(1), \quad n^{1/(2H)}(\hat{\theta}_n - \theta_0) = O_P^*(1);$$

and the asymptotic distributions of $(\sqrt{n}(\hat{\beta}_n - \beta_0), \sqrt{n}(\hat{\gamma}_n - \gamma_0))$ *and* $n^{1/(2H)}(\hat{\theta}_n - \theta_0)$ *are independent, with*

$$\begin{pmatrix} \sqrt{n}(\hat{\beta}_n - \beta_0) \\ \sqrt{n}(\hat{\gamma}_n - \gamma_0) \end{pmatrix} \to_w N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} s_1^2 & \rho s_1 s_2 \\ \rho s_1 s_2 & s_2^2 \end{pmatrix}\right),$$

$$n^{1/(2H)}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{W}} \textit{argmax}_h\left(W_H(h) - \frac{|h|^{2H}}{2}|\beta_0|\sigma\sqrt{PN(\tau)}\right),$$

$$\text{where} \quad s_1^2 = \frac{P(X^2N(\tau))}{P(Z_H^2(\theta_0)N(\tau))P(X^2N(\tau)) - P^2(Z_H(\theta_0)XN(\tau))},$$

$$s_2^2 = \frac{P(Z_H^2(\theta_0)N(\tau))}{P(Z_H^2(\theta_0)N(\tau))P(X^2N(\tau)) - P^2(Z_H(\theta_0)XN(\tau))},$$

$$\rho = \frac{-P(Z_H(\theta_0)XN(\tau))}{\sqrt{P(X^2N(\tau))P(Z_H^2(\theta_0)N(\tau))}},$$

while $W_H(\cdot)$ is a standard 2-sided fractional Brownian motion with Hurst parameter H starting from zero with unit variance scale (i.e., $W_H(1) =_d W_H(-1) \sim N(0,1)$).

Comparing to $PN(\tau)$ in Remark 2.2.1, here

$$PN(\tau) = 1 - P\left[\exp\left(-e^{\beta_0 Z_H(\theta_0) + \gamma_0 X}\int_0^\tau \lambda_0(s)ds\right)\right].$$

For general H other than $H = 0.5$, we do not have closed form C.D.F. for the asymptotic distribution $\text{argmax}_h\left(W_H(h) - \frac{|h|^{2H}}{2}|\beta_0|\sigma\sqrt{PN(\tau)}\right)$. However, we can use simulations (see Chapter 3.2) to learn about its properties.

In Chapter 2.2, we have made some remarks on the theoretical results of the simplified model. Now, we add some additional remarks based on Theorem 2.3.2.

Remark 2.3.3. *1. The rates of convergence for the regression coefficients estimators* $\hat{\beta}_n, \hat{\gamma}_n$ *are both* $\sqrt{n}$*, while the rate of convergence of location estimator* $\hat{\theta}_n$ *is* $n^{1/(2H)}$ $(0 < H < 1)$*. It means as* $n \to \infty$*,* $\hat{\theta}_n$ *converges to* θ_0 *in a faster rate compared to* $\hat{\beta}_n$ *and* $\hat{\gamma}_n$*. Due to the roughness of fBm's paths, a small shift of location on the trajectory of fBm can lead to a big change of the value of the trajectory. So it is easy to capture the location of interest. The smaller* H *is,* *the rougher the trajectory, and hence it is easier to estimate the location. This explains why the convergence rate of $\hat{\theta}_n$ is regulated by H, the Hurst parameter which describes the roughness of fBm's trajectories.*

2. *The estimation of (β, γ) is based on the estimation of θ. Especially, β is the regression coefficient of $Z_H(\theta)$ where θ is unknown as well. So it is not a surprise if the convergence rate of $\hat{\beta}_n$ is slower than those of $\hat{\gamma}_n$. However, thanks to the fast convergence of $\hat{\theta}_n$, the convergence rate of $\hat{\beta}_n$ is not impacted (by the uncertainty of θ) compared to that of $\hat{\gamma}_n$.*

2.3.3 The case H strictly less than 0.5

The theoretical results presented above require $H \in [1/2, 1)$. Such a restriction comes from the unavailability of a maximal inequality for the exponential function of fractional Brownian motion in the case $H \in (0, 1/2)$. Such a maximal inequality is used in the proof of Lemma C.2.3 in Appendix C.

However, our simulation results imply these theoretical results probably still apply for the case $H \in (0, 1/2)$, even though it is not mathematically justified by our proof due to the absence of the maximal inequality.

2.3.3.1 Non-identifiability of (θ_M, σ)

Similarly to the counterpart in the simplified model, to resolve the problem caused by the self-similarity of fractional Brownian motion and to make model (2.1) identifiable, we always set $\theta_M = 1$ when we estimate parameters for this model.

2.3.3.2 Wald-type confidence interval for θ

Comparing to its counterpart for the simplified model, the Wald-type confidence interval construction for the extended model is more involved. Since for general H, there is no analytical C.D.F. for the asymptotic distribution of $n^{1/(2H)}(\hat{\theta}_n - \theta_0)$, to get its quantile is a challenge. Even simulation methods won't work here since they require generating a fractional Brownian motion with drift on an infinite interval. For an infinite interval, it is not executable in simulations.

To get Wald-type confidence interval for θ in practice, we will modify the asymptotic distribution with the actual sample size information incorporated. The details are put into Chapter 3.2.

2.3.4 Proofs

We will use the procedure in Chapter 3 of VW [54] to establish the asymptotic properties of our M-estimators. Same as for the simplified model, WLOG we will assume $\sigma = 1$.

2.3.4.1 Local quadratic property

In this section, we will prove three results. The first is the strict concavity of the function $\mathbb{M}(\pi)$. The second is π_0 is the unique global maximum point. The third is the local quadratic property of $\mathbb{M}(\pi)$ at π_0. The proofs of the first and second results are done in the subsection **Strict Concavity and Unique Global Maximum Point** and the proof of the third will appear in the subsection **Local Quadratic Property at** π_0.

$$\mathbb{M}(\pi) = P\left[(\beta Z_H(\theta) + \gamma X)N(\tau) - \int_0^\tau \log\left[PY(u)\exp(\beta Z_H(\theta) + \gamma X)\right] dN(u)\right].$$

In this section, we need to take derivatives of expectations. We give an example in Appendix D to show how to justify the exchange of differentiation and expectation.

Take $d(\pi, \pi_0)$ to be a function of $|\beta - \beta_0|, |\gamma - \gamma_0|, |\theta - \theta_0|$, in order to check the local quadratic property of $\mathbb{M}(\pi)$ in the neighborhood of π_0, we investigate the difference of $\mathbb{M}(\pi)$ and $\mathbb{M}(\pi_0)$.

$$
\begin{aligned}
&\mathbb{M}(\pi) - \mathbb{M}(\pi_0) \\
=&P\left[(\beta Z_H(\theta) + \gamma X)N(\tau) - \int_0^\tau \log\left[PY(u)\exp(\beta Z_H(\theta) + \gamma X)\right] dN(u)\right] \\
&- P\left[(\beta_0 Z_H(\theta_0) + \gamma_0 X)N(\tau) - \int_0^\tau \log\left[PY(u)\exp(\beta_0 Z_H(\theta_0) + \gamma_0 X)\right] dN(u)\right] \\
=&P\left[(\beta Z_H(\theta) + \gamma X - (\beta_0 Z_H(\theta_0) + \gamma_0 X))\, N(\tau)\right] \\
&- P\left[\int_0^\tau \log\left[\frac{PY(u)\exp(\beta Z_H(\theta) + \gamma X)}{PY(u)\exp(\beta_0 Z_H(\theta_0) + \gamma_0 X)}\right] dN(u)\right]
\end{aligned}
$$

To deal with the last term of this decomposition,

$$
\begin{aligned}
&PY(u)\exp(\beta Z_H(\theta) + \gamma X) \\
&= P\left(C \geq u | Z_H(\theta_0), X\right) \exp\left(\beta Z_H(\theta) + \gamma X - e^{\beta_0 Z_H(\theta_0) + \gamma_0 X} \int_0^u \lambda_0(s) ds\right) \\
&= P\left(C \geq u | Z_H(\theta_0), X\right) \cdot P\exp(\beta(Z_H(\theta) - Z_H(\theta_0))) \\
&\qquad \cdot P\exp\left(\beta Z_H(\theta_0) + \gamma X - e^{\beta_0 Z_H(\theta_0) + \gamma_0 X} \int_0^u \lambda_0(s) ds\right) \\
&= \exp(-1/2\beta^2|\theta - \theta_0|^{2H}) PY(u)\exp(\beta Z_H(\theta_0) + \gamma X),
\end{aligned}
\tag{2.11}
$$

where we used the distributional property of fBm $Z_H(\theta) - Z_H(\theta_0)$ in the last equality.

So $\mathbb{M}(\pi) - \mathbb{M}(\pi_0)$ can be further written as

$$
\begin{aligned}
=&(\beta - \beta_0)P\left[Z_H(\theta_0)N(\tau)\right] + (\gamma - \gamma_0)P\left[XN(\tau)\right] - 1/2\beta^2|\theta - \theta_0|^{2H} PN(\tau) \\
&- P\left[\int_0^\tau \log\left[\frac{PY(u)\exp(\beta Z_H(\theta_0) + \gamma X)}{PY(u)\exp(\beta_0 Z_H(\theta_0) + \gamma_0 X)}\right] dN(u)\right],
\end{aligned}
$$

where $P[\beta(Z_H(\theta) - Z_H(\theta_0))N(\tau)]$ disappeared since

$$
\begin{aligned}
P[\beta(Z_H(\theta) - Z_H(\theta_0))N(\tau)] &= P[\beta(Z_H(\theta) - Z_H(\theta_0))1_{T^0 \leq C, T^0 \leq \tau}] \\
&= P[\beta(Z_H(\theta) - Z_H(\theta_0))] P[1_{T^0 \leq C, T^0 \leq \tau}] = 0,
\end{aligned}
$$

where the second equality holds because the vector (T^0, C) depends on $(Z_H(\theta_0), X)$ only and hence is independent of $Z_H(\theta) - Z_H(\theta_0)$ by model assumption, the last equality holds by the zero-mean property of fractional Brownian motion $Z_H(\theta) - Z_H(\theta_0)$.

Strict concavity and unique global maximum point

In this subsection, we will prove that the function $\mathbb{M}(\pi)$ is strictly concave, and that π_0 is the unique global maximum point of $\mathbb{M}(\pi)$.

Using the notation $F(\beta,\gamma,u)$ set earlier in Chapter 2.3.1, we have

$$PY(u)\exp(\beta Z_H(\theta_0)+\gamma X) = PF(\beta,\gamma,u).$$

The first and second order derivatives of $\mathbb{M}(\pi)$ w.r.t. β are (in all the following derivatives calculation, we need to justify the exchange of differentiation and expectation as that in Appendix D, which we omit here not to obscure our focus)

$$\begin{aligned}
\frac{\partial \mathbb{M}(\pi)}{\partial \beta} &= PZ_H(\theta_0)N(\tau) - \int_0^\tau \frac{PZ_H(\theta_0)F(\beta,\gamma,u)}{PF(\beta,\gamma,u)} dPN(u), \\
\frac{\partial^2 \mathbb{M}(\pi)}{\partial^2 \beta} &= \int_0^\tau \frac{\left(PZ_H(\theta_0)F(\beta,\gamma,u)\right)^2 - P\left(Z_H^2(\theta_0)F(\beta,\gamma,u)\right)PF(\beta,\gamma,u)}{\left(PF(\beta,\gamma,u)\right)^2} dPN(u),
\end{aligned}$$

and $\frac{\partial \mathbb{M}(\pi)}{\partial \gamma}, \frac{\partial^2 \mathbb{M}(\pi)}{\partial^2 \gamma}$ can be obtained with similar formulas. Notice that in deriving $\frac{\partial^2 \mathbb{M}(\pi)}{\partial^2 \beta}$, we need $P\left(Z_H^2(\theta_0)F(\beta,\gamma,u)\right) < \infty$ which is guaranteed by Assumptions 2.3.1.

The term $\frac{\partial^2 \mathbb{M}(\pi)}{\partial^2 \beta}$ can be proved to be strictly negative deterministic functions of β by using Cauchy–Schwartz Inequality for the integrand; be equal to zero-valued function only if model are degenerated: $Z_H(\theta_0)$ is a degenerated random variable. Similar results holds for $\frac{\partial^2 \mathbb{M}(\pi)}{\partial^2 \gamma}$.

$$\frac{\partial^2 \mathbb{M}(\pi)}{\partial \beta \partial \gamma} = \int_0^\tau \frac{P\left(Z_H(\theta_0)F_{\beta,\gamma,u}\right)P(X) - P\left(Z_H(\theta_0)XF_{\beta,\gamma,u}\right)PF_{\beta,\gamma,u}}{\left(PF_{\beta,\gamma,u}\right)^2} dPN(u),$$

where we denote $F(\beta,\gamma,u)$ as $F_{\beta,\gamma,u}$ due to space limit.

We want to prove

$$\frac{\partial^2 \mathbb{M}(\pi)}{\partial^2 \gamma} \cdot \frac{\partial^2 \mathbb{M}(\pi)}{\partial^2 \beta} - \left(\frac{\partial^2 \mathbb{M}(\pi)}{\partial \beta \partial \gamma}\right)^2 > 0.$$

To make notation simpler, denote $Z \equiv Z_H(\theta_0), F \equiv F(\beta, \gamma, u)$, it suffices to prove

$$\left[\int_0^\tau \frac{P(Z^2F)P(F) - P^2(ZF)}{P^2F} dPN(u)\right] \left[\int_0^\tau \frac{P(X^2F)P(F) - P^2(XF)}{P^2F} dPN(u)\right] - \left[\int_0^\tau \frac{P(ZXF)P(F) - P(ZF)P(XF)}{P^2F} dPN(u)\right]^2 > 0. \tag{2.12}$$

The proof can be done as follows.

For any random variable U, we have $P(U^2F)P(F) \geq P^2(UF)$ by Cauchy–Schwartz Inequality with equality holds only if $U = c$ for a constant c, P -a.s..

Let $U = X + aZ$, where a is any real number, we will have $P((X+aZ)^2F)P(F) \geq P^2[(X + aZ)F]$, with equality holds only if $X + aZ = c$ for a constant c, P -a.s..

Since the condition for equality, $X + aZ = c$ for a constant c, P -a.s., does not hold by the independence of $Z_H(\theta_0)$ and X, it follows that $P((X + aZ)^2F)P(F) > P^2[(X + aZ)F]$. By the monotonically increasing property of $PN(u)$, we have

$$\int_0^\tau \frac{P[(X + aZ)^2F]P(F)}{P^2F} dPN(u) > \int_0^\tau \frac{P^2[(X + aZ)F)}{P^2F} dPN(u).$$

Expand and reorganize the inequality,

$$\left[\int_0^\tau \frac{P(Z_H^2F)P(F) - P^2(ZF)}{P^2F} dPN(u)\right] a^2 + \left[\int_0^\tau \frac{P(X^2F)P(F) - P^2(XF)}{P^2F} dPN(u)\right] + 2\left[\int_0^\tau \frac{P(ZXF)P(F) - P(ZF)P(XF)}{P^2F} dPN(u)\right] a > 0, \qquad \forall a \in \mathbb{R}.$$

By discriminant of a quadratic, (2.12) holds; and hence

$$\frac{\partial^2\mathbb{M}(\pi)}{\partial^2\gamma} \cdot \frac{\partial^2\mathbb{M}(\pi)}{\partial^2\beta} - \left(\frac{\partial^2\mathbb{M}(\pi)}{\partial\beta\partial\gamma}\right)^2 > 0.$$

Besides, it is obvious that $\frac{\partial^2\mathbb{M}(\pi)}{\partial^2\beta} < 0$ and $\frac{\partial^2\mathbb{M}(\pi)}{\partial^2\gamma} < 0$. All the three conditions hold throughout $(\beta, \gamma) \in [-\beta_M, \beta_M] \times [-\gamma_M, \gamma_M]$, so $\mathbb{M}(\pi)$ is a strictly concave function of (β, γ).

By the derived expression of $\mathbb{M}(\pi) - \mathbb{M}(\pi_0)$ which has $-1/2\beta^2|\theta - \theta_0|^{2H}PN(\tau)$ as the only term that includes θ, it follows that for any given value of (β, γ), $\mathbb{M}(\pi)$ takes a unique maximum at $(\beta, \gamma, \theta_0)$ for all θ.

It follows that to look for the global maximum of $\mathbb{M}(\pi)$, we just fix θ_0 and look for the (β, γ) that maximize $\mathbb{M}(\beta, \gamma, \theta_0)$.

The first derivative of $\mathbb{M}(\pi)$ w.r.t. β at π_0

$$\frac{\partial \mathbb{M}(\pi)}{\partial \beta}|_{\pi=\pi_0} = PZ_H(\theta_0)N(\tau) - P\left[\int_0^\tau \frac{PZ_H(\theta_0)F(\beta_0,\gamma_0,u)}{PF(\beta_0,\gamma_0,u)}dN(u)\right] = 0$$

by direct calculation, and the same equality holds for $\frac{\partial \mathbb{M}(\pi)}{\partial \gamma}|_{\pi=\pi_0}$.

By $\frac{\partial \mathbb{M}(\pi)}{\partial \beta}|_{\pi=\pi_0} = \frac{\partial \mathbb{M}(\pi)}{\partial \gamma}|_{\pi=\pi_0} = 0$, $\mathbb{M}(\beta,\gamma,\theta) - \mathbb{M}(\beta,\gamma,\theta_0) = 1/2\beta^2|\theta-\theta_0|^{2H}PN(\tau)$ and considering its strict concavity w.r.t. (β,γ), it follows that $\pi_0 = (\beta_0,\gamma_0,\theta_0)$ is the unique global maximum point of $\mathbb{M}(\pi)$, and $\sup_{\pi:d(\pi,\pi_0)\geq\epsilon}\mathbb{M}(\pi) < \mathbb{M}(\pi_0)$.

Local quadratic property at π_0

To obtain the local quadratic property of $\mathbb{M}(\pi)$ at π_0, we want to find a metric (or more general, a semi-metric) $d(\pi,\pi_0)$, such that for π near π_0, $\mathbb{M}(\pi) - \mathbb{M}(\pi_0) \lesssim -d^2(\pi,\pi_0)$. Looking at the difference of $\mathbb{M}(\pi)$ and $\mathbb{M}(\pi_0)$, and also considering Taylor expansion about β,γ near π_0, it can be decomposed as

$$\begin{aligned}
&\frac{\partial \mathbb{M}(\pi)}{\partial \beta}|_{\pi=\pi_0}(\beta-\beta_0) + \frac{\partial \mathbb{M}(\pi)}{\partial \gamma}|_{\pi=\pi_0}(\gamma-\gamma_0) + \frac{\partial^2 \mathbb{M}(\pi)}{\partial\beta\partial\gamma}|_{\pi=\pi_0}(\beta-\beta_0)(\gamma-\gamma_0)\\
&+\frac{1}{2}\frac{\partial^2 \mathbb{M}(\pi)}{\partial^2\beta}|_{\pi=\pi_0}(\beta-\beta_0)^2 + \frac{1}{2}\frac{\partial^2 \mathbb{M}(\pi)}{\partial^2\gamma}|_{\pi=\pi_0}(\gamma-\gamma_0)^2 - \frac{1}{2}|\beta(\theta-\theta_0)|^{2H}PN(\tau)\\
&+o\left((\beta-\beta_0)^2\right) + o\left((\gamma-\gamma_0)^2\right)\\
=&\frac{\partial^2 \mathbb{M}(\pi)}{\partial\beta\partial\gamma}|_{\pi=\pi_0}(\beta-\beta_0)(\gamma-\gamma_0) + \frac{1}{2}\frac{\partial^2 \mathbb{M}(\pi)}{\partial^2\beta}|_{\pi=\pi_0}(\beta-\beta_0)^2 + \frac{1}{2}\frac{\partial^2 \mathbb{M}(\pi)}{\partial^2\gamma}|_{\pi=\pi_0}(\gamma-\gamma_0)^2\\
&-\frac{1}{2}|\beta(\theta-\theta_0)|^{2H}PN(\tau) + o\left((\beta-\beta_0)^2\right) + o\left((\gamma-\gamma_0)^2\right).
\end{aligned}$$

Reorganize $\frac{\partial^2 \mathbb{M}(\pi)}{\partial\beta\partial\gamma}|_{\pi=\pi_0}(\beta-\beta_0)(\gamma-\gamma_0)+\frac{1}{2}\frac{\partial^2 \mathbb{M}(\pi)}{\partial^2\beta}|_{\pi=\pi_0}(\beta-\beta_0)^2+\frac{1}{2}\frac{\partial^2 \mathbb{M}(\pi)}{\partial^2\gamma}|_{\pi=\pi_0}(\gamma-\gamma_0)^2$ as the form of

$$-\frac{1}{2}c_1(\beta-\beta_0)^2 - \frac{1}{2}c_1(\gamma-\gamma_0)^2 - \frac{1}{2}(c_2(\beta-\beta_0) - \frac{\frac{\partial^2 \mathbb{M}(\pi)}{\partial\beta\partial\gamma}|_{\pi=\pi_0}}{c_2}(\gamma-\gamma_0))^2,$$

let $c_1 + c_2^2 = -\frac{\partial^2 \mathbb{M}(\pi)}{\partial^2 \beta}|_{\pi=\pi_0}$ and $c_1 + \left(\frac{\frac{\partial^2 \mathbb{M}(\pi)}{\partial \beta \partial \gamma}|_{\pi=\pi_0}}{c_2} \right)^2 = -\frac{\partial^2 \mathbb{M}(\pi)}{\partial^2 \gamma}|_{\pi=\pi_0}$, and solve the equations, we obtain the solution as real numbers:

$$c_1 = \frac{1}{2} \left(a + b - \sqrt{(a-b)^2 + 4c^2} \right),$$
$$c_2^2 = \frac{a - b + \sqrt{(a-b)^2 + 4c^2}}{2},$$

where $a = -\frac{\partial^2 \mathbb{M}(\pi)}{\partial^2 \beta}|_{\pi=\pi_0}$, $b = -\frac{\partial^2 \mathbb{M}(\pi)}{\partial^2 \gamma}|_{\pi=\pi_0}$, $c = \frac{\partial^2 \mathbb{M}(\pi)}{\partial \beta \partial \gamma}|_{\pi=\pi_0}$.

Note: The preceding inequality $ab - c^2 > 0$ proved earlier guarantees that $c_1 > 0$.

$$\begin{aligned}
&\frac{\partial^2 \mathbb{M}(\pi)}{\partial \beta \partial \gamma}|_{\pi=\pi_0}(\beta - \beta_0)(\gamma - \gamma_0) + \frac{1}{2}\frac{\partial^2 \mathbb{M}(\pi)}{\partial^2 \beta}|_{\pi=\pi_0}(\beta - \beta_0)^2 + \frac{1}{2}\frac{\partial^2 \mathbb{M}(\pi)}{\partial^2 \gamma}|_{\pi=\pi_0}(\gamma - \gamma_0)^2 \\
&- \frac{1}{2}|\beta(\theta - \theta_0)|^{2H} PN(\tau) \\
=& -\frac{1}{2}c_1 \left[(\beta - \beta_0)^2 + (\gamma - \gamma_0)^2 \right] - \frac{1}{2}(c_2(\beta - \beta_0) - \frac{\frac{\partial^2 \mathbb{M}(\pi)}{\partial \beta \partial \gamma}|_{\pi=\pi_0}}{c_2}(\gamma - \gamma_0))^2 \\
&- \frac{1}{2}|\beta(\theta - \theta_0)|^{2H} PN(\tau) \\
\leq& -\frac{1}{2}c_1 \left[(\beta - \beta_0)^2 + (\gamma - \gamma_0)^2 \right] - \frac{1}{2}|\beta(\theta - \theta_0)|^{2H} PN(\tau)
\end{aligned}$$

Since $\beta_0 \neq 0$ by Assumptions 2.3.1, when the neighborhood is small enough, we can always make all the β in the neighborhood to be bounded away from zero ($|\beta| \geq \beta_m > 0$), e.g., $\beta_m = \frac{|\beta_0|}{2}$ is one possible choice.

Take constant $\tilde{c} \equiv \frac{c_1 \wedge \beta_m^{2H} PN(\tau)}{2}$, then

$$\mathbb{M}(\theta) - \mathbb{M}(\theta_0) \leq -\tilde{c} d^2(\pi, \pi_0) \lesssim -d^2(\pi, \pi_0),$$

where $d^2(\pi, \pi_0) = (\beta - \beta_0)^2 + (\gamma - \gamma_0)^2 + |\theta - \theta_0|^{2H}$, and $\tilde{c}$ does not depend on β, γ, θ.

2.3.4.2 Consistency

The next step is to prove $\hat{\pi}_n \to_{P^*} \pi_0$, the consistency of $\hat{\pi}_n$. We already proved the local quadratic property of $\mathbb{M}(\pi)$ at π_0, π_0 as its unique global maximizing point of $\mathbb{M}(\pi)$, and the strict concavity of $\mathbb{M}(\pi)$ over the whole domain of $(\beta, \gamma) \in [-\beta_M, \beta_M] \times$

$[-\gamma_M, \gamma_M]$, and $\sup_{\pi:d(\pi,\pi_0)\geq\epsilon} \mathbb{M}(\pi) < \mathbb{M}(\pi_0)$. By Theorem 5.7 of van der Vaart [53], to prove $\sup_{\pi\in\Phi} |\mathbb{M}_n(\pi) - \mathbb{M}(\pi)| \to 0$ in P^* suffices, where $\Phi \equiv [-\beta_M, \beta_M] \times [-\gamma_M, \gamma_M] \times [0, \theta_M]$.

$$
\begin{aligned}
\forall \epsilon > 0, \qquad & P^*\left(\sup_{\pi\in\Phi} |\mathbb{M}_n(\pi) - \mathbb{M}(\pi)| \geq \epsilon\right) \\
= P^* \Bigg(& \sup_{\pi\in\Phi} \Big| (\mathbb{P}_n - P)\left((\beta Z_H(\theta) + \gamma X) N(\tau)\right) \\
& - \left[\int_0^\tau \log\left[\mathbb{P}_n Y(u) e^{\beta Z_H(\theta)+\gamma X}\right] \mathbb{P}_n dN(u) - \int_0^\tau \log s^{(0)}(\pi, u) P dN(u)\right]\Big| \geq \epsilon \Bigg) \\
\leq P^* \Bigg(& \sup_{\pi\in\Phi} \left| (\mathbb{P}_n - P)\left((\beta Z_H(\theta) + \gamma X) N(\tau)\right)\right| \\
& + \sup_{\pi\in\Phi} \left| \int_0^\tau \log\left[\frac{\mathbb{P}_n Y(u) e^{\beta Z_H(\theta)+\gamma X}}{P Y(u) e^{\beta Z_H(\theta)+\gamma X}}\right] \mathbb{P}_n dN(u)\right| \\
& + \sup_{\pi\in\Phi} \left| \int_0^\tau \log s^{(0)}(\pi, u) (\mathbb{P}_n - P) dN(u)\right| \geq \epsilon \Bigg) \\
\leq P^* \Bigg(& \sup_{\pi\in\Phi} \left| (\mathbb{P}_n - P)\left((\beta Z_H(\theta) + \gamma X) N(\tau)\right)\right| > \frac{\epsilon}{3} \Bigg) \\
& + P^*\left(\sup_{\pi\in\Phi} \left| \int_0^\tau \log\left[\frac{\mathbb{P}_n Y(u) e^{\beta Z_H(\theta)+\gamma X}}{P Y(u) e^{\beta Z_H(\theta)+\gamma X}}\right] \mathbb{P}_n dN(u)\right| > \frac{\epsilon}{3}\right) \\
& \quad + P^*\left(\sup_{\pi\in\Phi} \left| \int_0^\tau \log s^{(0)}(\pi, u) (\mathbb{P}_n - P) dN(u)\right| \geq \frac{\epsilon}{3}\right) \\
\leq \frac{3}{\epsilon} P^* & \sup_{\pi\in\Phi} \left| (\mathbb{P}_n - P)\left((\beta Z_H(\theta) + \gamma X) N(\tau)\right)\right| \\
& + P^*\left(\sup_{\pi\in\Phi} \left| \int_0^\tau \log\left[\frac{\mathbb{P}_n Y(u) e^{\beta Z_H(\theta)+\gamma X}}{P Y(u) e^{\beta Z_H(\theta)+\gamma X}}\right] \mathbb{P}_n dN(u)\right| \geq \epsilon\right) \\
& + \frac{3}{\epsilon} P^* \sup_{\pi\in\Phi} \left| \int_0^\tau \log s^{(0)}(\pi, u) (\mathbb{P}_n - P) dN(u)\right| \\
\equiv I_1 & + I_2 + I_3.
\end{aligned}
$$

$$
I_1 = \frac{3}{\sqrt{n}\epsilon} P^* \sup_{\pi\in\Phi} \left| \mathbb{G}_n\left((\beta Z_H(\theta) + \gamma X) N(\tau)\right)\right|.
$$

Since $\mathcal{Q} \equiv \{\beta Z_H(\theta) + \gamma X : |\beta| \leq \beta_M, \theta \in [0, \theta_M], |\gamma| \leq \gamma_M\}$ has finite integral of $L_2(P)$ entropy with bracketing (see Lemma C.3.1 in Appendix C) , the class formed by multiplying it to a function $N(\tau)$ which is bounded by 1, $\mathcal{Q}_N = \{(\beta Z_H(\theta)+\gamma X)N(\tau) : \pi \in \Phi\}$, still has bounded bracketing entropy and $J_{[]}(1, \mathcal{Q}_N, L_2(P)) < \infty$ as well (multiplying the brackets of $\mathcal{Q}$ by $N(\tau)$ provides brackets for $\mathcal{Q}_N$), then by Theorem 2.14.2 of VW [54],

$$\begin{aligned}
&P^* \sup_{\pi\in\Phi} |\mathbb{G}_n\left((\beta Z_H(\theta) + \gamma X)N(\tau)\right)| \\
&\leq J_{[]}(1, \mathcal{Q}_N, L_2(P))P^* \sqrt{\sup_{\pi\in\Phi}(\beta Z_H(\theta) + \gamma X)^2 N^2(1)} \lesssim P^* \sqrt{\sup_{\pi\in\Phi}(\beta Z_H(\theta) + \gamma X)^2} \\
&\lesssim \sqrt{2\left(4\beta_M^2 C_{2,H}\theta_M^{2H} + \gamma_M^2 PX^2\right)} \lesssim \sqrt{\beta_M^2\theta_M^{2H} + \gamma_M^2 PX^2},
\end{aligned}$$

where we used the maximal inequality for fractional Brownian motion from Novikov and Valkeila [31] in the second to last inequality. It follows that $I_1 \to 0$ as $n \to \infty$.

$$I_2 = P^*\left(\sup_{\pi\in\Phi}\left|\int_0^\tau \log\left[\frac{\mathbb{P}_n Y(u)e^{\beta Z_H(\theta)+\gamma X}}{PY(u)e^{\beta Z_H(\theta)+\gamma X}}\right]\mathbb{P}_n dN(u)\right| \geq \epsilon\right).$$

To prove $\lim_{n\to\infty} I_2 = 0$, it suffices to prove the supremum term in the preceding display converges to 0 P^*-a.s. as $n \to \infty$.

Following the same lines as that of the simplified model, we only need to prove the P^*-Glivenko−Cantelli a.s. of $\mathcal{N} \equiv \{Y(u)e^{\beta Z_H(\theta)+\gamma X} : u \in [0,\tau], |\beta| \leq \beta_M, \theta \in [0,\theta_M], |\gamma| \leq \gamma_M\}$ and find out the lower bound of $PY(u)e^{\beta Z_H(\theta)+\gamma X}$ over $(u,\beta,\gamma,\theta) \in [0,\tau] \times [-\beta_M,\beta_M] \times [-\gamma_M,\gamma_M] \times [0,\theta_M]$.

By Lemma C.2.1, $\mathcal{N}$ has finite integral of $L_2(P)$ entropy with bracketing, hence $\mathcal{N}$ is P^*-Glivenko−Cantelli a.s..

On the other hand, $PY(u)e^{\beta Z_H(\theta)+\gamma X} = e^{-1/2\beta^2|\theta-\theta_0|^{2H}} PF(\beta,\gamma,u)$. We can obtain its continuity over $(\beta,\gamma,\theta) \in [-\beta_M,\beta_M] \times [-\gamma_M,\gamma_M] \times [0,\theta_M]$. Its monotonicity w.r.t. u is also obvious. So its minimum value, denoted as C_m, is attained in the bounded and closed region $(u,\beta,\gamma,\theta) \in [0,\tau] \times [-\beta_M,\beta_M] \times [-\gamma_M,\gamma_M] \times [0,\theta_M]$. Since

$$\begin{aligned}
PY(u)e^{\beta Z_H(\theta)+\gamma X} &= P\left[e^{\beta Z_H(\theta)+\gamma X} e^{-\int_0^\tau \lambda_0(s)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}ds} P(C \geq u|Z_H(\theta_0), X)\right] \\
&\geq P\left[e^{\beta Z_H(\theta)+\gamma X} e^{-\int_0^\tau \lambda_0(s)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}ds} P(C \geq \tau|Z_H(\theta_0), X)\right],
\end{aligned}$$

where the right hand side of the inequality is strictly positive throughout the region by the condition $P(C \geq \tau | Z_H(\theta_0), X) > 0$ from Assumptions 2.3.1, the attained infimum $C_m > 0$ is bounded away from 0. (Note that C_m does not depend on $(u, \theta, \beta, \gamma)$.)

$$I_3 = \frac{3}{\epsilon} P^* \sup_{\pi \in \Phi} \left| \int_0^\tau \log s^{(0)}(\pi, u)(\mathbb{P}_n - P) dN(u) \right| \leq \frac{3}{\sqrt{n}\epsilon} P^* \sup_{\pi \in \Phi} |\mathbb{G}_n g_\pi|,$$

where $g_\pi = 1_{(T \leq C)} 1_{(0 < T \leq \tau)} \log s^{(0)}(\pi, T) - e^{\beta_0 Z_H(\theta_0) + \gamma_0 X} \int_0^\tau 1_{(T \geq u)} \lambda_0(u) \log s^{(0)}(\pi, u) du$. It also converges to zero as n goes to infinity. The detailed proof is put in Appendix B.1.

So altogether, as $n \to \infty$, $I_1 + I_2 + I_3 \to 0$, and hence

$$\forall \epsilon > 0, \qquad P^* \left(\sup_{\pi \in \Phi} |\mathbb{M}_n(\pi) - \mathbb{M}(\pi)| \geq \epsilon \right) \to 0. \tag{2.13}$$

In Chapter 2.3.4.1, it has been proved that $\mathbb{M}(\pi)$ is strictly concave on its whole domain of $(\beta, \gamma) \in [-\beta_M, \beta_M] \times [-\gamma_M, \gamma_M]$, , has unique global maximum point $\pi_0 = (\beta_0, \gamma_0, \theta_0)$ and local quadratic property holds for $\mathbb{M}(\pi)$ at π_0, and $\mathbb{M}(\pi_0) > \sup_{d(\pi, \pi_0) > \epsilon} \mathbb{M}(\pi)$. By Theorem 5.7 of van der Vaart [53] and (2.13), the consistency of $\hat{\pi}_n = (\hat{\beta}_n, \hat{\gamma}_n, \hat{\theta}_n)$ is proved.

2.3.4.3 Rate of convergence

By Theorem 3.2.5 in VW [54], the upper bounds for the convergence rates of $\hat{\beta}_n, \hat{\gamma}_n, \hat{\theta}_n$ are $\sqrt{n}, \sqrt{n}, n^{1/(2H)}$ respectively. Details of the proof are too lengthy and put into Appendix B.2.

2.3.4.4 Asymptotics of rescaled criterion function

We have obtained the rates of convergence in the previous section. Following the three-step procedure for M-estimators, the next step is to establish the uniform convergence of a rescaled localized criterion function.

Let $\beta_n = \beta_0 + \frac{h_\beta}{\sqrt{n}}$, $\gamma_n = \gamma_0 + \frac{h_\gamma}{\sqrt{n}}$, $\theta_n = \theta_0 + h_\theta/n^{1/(2H)}$. Denote $h_\pi = (h_\theta, h_\beta, h_\gamma)$, $h_{\pi_{r_n}} = \left(h_\theta/n^{1/(2H)}, \frac{h_\beta}{\sqrt{n}}, \frac{h_\gamma}{\sqrt{n}}\right)$. Now consider $\mathbb{Q}_n(h_{\pi_{r_n}}) = s_n(\mathbb{M}_n(\pi_0 + h_{\pi_{r_n}}) - \mathbb{M}_n(\pi_0))$, where $h_\pi \in [-K, K]^3, \forall K > 0$ and $s_n = n$.

$$
\begin{aligned}
& n(\mathbb{M}_n(\pi_0 + h_{\pi_{r_n}}) - \mathbb{M}_n(\pi_0)) \\
=& n\left(\mathbb{P}_n\left[\int_0^\tau (\beta_n Z_H(\theta_n) + \gamma_n X)\, dN(s) - \int_0^\tau \log\left[\mathbb{P}_n Y(u) e^{\beta_n Z_H(\theta_n) + \gamma_n X}\right] dN(u)\right]\right. \\
&\left. -\mathbb{P}_n\left[\int_0^\tau [\beta_0 Z_H(\theta_0) + \gamma_0 X]\, dN(s) - \int_0^\tau \log[\mathbb{P}_n Y(u) e^{\beta_0 Z_H(\theta_0) + \gamma_0 X}] dN(u)\right]\right) \\
=& \, n\mathbb{P}_n\left[\left(\beta_n\left(Z_H(\theta_n) - Z_H(\theta_0)\right) + \frac{h_\beta}{\sqrt{n}} Z_H(\theta_0) + \frac{h_\gamma}{\sqrt{n}} X\right) N(\tau)\right] \\
& - n\int_0^\tau \log\left[\frac{\mathbb{P}_n Y(u) \exp\left(\beta_n Z_H(\theta_n) + \gamma_n X\right)}{\mathbb{P}_n Y(u) e^{\beta_0 Z_H(\theta_0) + \gamma_0 X}}\right] \mathbb{P}_n dN(u) \\
\equiv& \, I_8 - I_9,
\end{aligned}
$$

where

$$
\begin{aligned}
I_8 &= n\mathbb{P}_n\left(\left[\beta_n\left(Z_H(\theta_n) - Z_H(\theta_0)\right) + \frac{h_\beta}{\sqrt{n}} Z_H(\theta_0) + \frac{h_\gamma}{\sqrt{n}} X\right] N(\tau)\right) \\
&= n\mathbb{P}_n\left(\left[\beta_n W\left(h_\theta/n^{1/(2H)}\right) + \frac{h_\beta}{\sqrt{n}} Z_H(\theta_0) + \frac{h_\gamma}{\sqrt{n}} X\right] N(\tau)\right), \\
I_9 &= n\int_0^\tau \log\left(\frac{\mathbb{P}_n Y(u) \exp\left[\beta_n Z_H(\theta_n) + \gamma_n X\right]}{\mathbb{P}_n Y(u) e^{\beta_0 Z_H(\theta_0) + \gamma_0 X}}\right) \mathbb{P}_n dN(u).
\end{aligned}
$$

The numerator of the integrand can be decomposed as

$$
\begin{aligned}
& \mathbb{P}_n Y(u) \exp\left[\left(\beta_0 + \frac{h_\beta}{\sqrt{n}}\right) Z_H\left(\theta_0 + h_\theta/n^{1/(2H)}\right) + \left(\gamma_0 + \frac{h_\gamma}{\sqrt{n}}\right) X\right] \\
=& \mathbb{P}_n Y(u) e^{\beta_0 Z_H(\theta_0) + \gamma_0 X}\left(\exp\left[\beta_n\left(Z_H(\theta_n) - Z_H(\theta_0)\right) + \frac{h_\beta}{\sqrt{n}} Z_H(\theta_0) + \frac{h_\gamma}{\sqrt{n}} X\right]\right) \\
=& \mathbb{P}_n Y(u) e^{\beta_0 Z_H(\theta_0) + \gamma_0 X}\left(\exp\left[\beta_n W_H\left(h_\theta/n^{1/(2H)}\right) + \frac{h_\beta}{\sqrt{n}} Z_H(\theta_0) + \frac{h_\gamma}{\sqrt{n}} X\right]\right),
\end{aligned}
$$

where $W_H\left(h_\theta/n^{1/(2H)}\right) \equiv Z_H(\theta_n) - Z_H(\theta_0)$.

Denote $\Delta_n \equiv \beta_n W_H\left(h_\theta/n^{1/(2H)}\right) + \frac{h_\beta}{\sqrt{n}} Z_H(\theta_0) + \frac{h_\gamma}{\sqrt{n}} X$, then

$$
\begin{aligned}
I_9 =& n\int_0^\tau \log\left[\frac{\mathbb{P}_n Y(u) e^{\beta_0 Z_H(\theta_0) + \gamma_0 X} \exp\left(\Delta_n\right)}{\mathbb{P}_n Y(u) e^{\beta_0 Z_H(\theta_0) + \gamma_0 X}}\right] \mathbb{P}_n dN(u) \\
=& \int_0^\tau \log\left[1 + \frac{\mathbb{P}_n Y(u) e^{\beta_0 Z_H(\theta_0) + \gamma_0 X}\left(\exp\left(\Delta_n\right) - 1\right)}{\mathbb{P}_n Y(u) e^{\beta_0 Z_H(\theta_0) + \gamma_0 X}}\right] n\mathbb{P}_n dN(u).
\end{aligned}
$$

Considering $\lim_{n\to\infty}\left|(\mathbb{P}_n - P)\,Y(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\right| = 0$ and

$$\lim_{n\to\infty}\sup_{\left(h_\beta,h_\gamma,h_\theta\right)\in[-K,K]^3}\left|(\mathbb{P}_n - P)\,Y(u)\exp\left[\beta_n Z_H\left(\theta_n\right)+\gamma_n X\right]\right| = 0,$$

which follow by similar argument as that on $\mathcal{N}_{\delta,-}$ in Appendix B.2, then

$$\begin{aligned}&\mathbb{P}_n Y(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\left(\exp\left(\Delta_n\right)-1\right)\\&\qquad = PY(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\left(\exp\left(\Delta_n\right)-1\right)+o_{uP}(1) = o_{uP}(1),\\&\mathbb{P}_n Y(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X} = PY(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}+o_{uP}(1).\end{aligned}$$

Notice that $PY(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}(\geq C_m)$ is bounded away from zero, it follows that

$$\frac{\mathbb{P}_n Y(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\left(\exp\left(\Delta_n\right)-1\right)}{\mathbb{P}_n Y(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}} = o_{uP}(1),$$

where $A_{h_\pi,n} = o_{uP}(1)$ means $A_{h_\pi,n} = o_P(1)$ uniformly over $h_\pi \in [-K,K]^3$.

Taylor expansion

Since $\log\left(1+x\right) = x + O(x^2) = x(1+O(x)) = x(1+o(1))$ as $x\to 0$,

$$\begin{aligned}&\log\left[1+\frac{\mathbb{P}_n Y(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\left(\exp\left(\Delta_n\right)-1\right)}{\mathbb{P}_n Y(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}}\right]\\&\qquad = \frac{\mathbb{P}_n Y(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\left(\exp\left(\Delta_n\right)-1\right)}{\mathbb{P}_n Y(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}}\left[1+o_{uP}(1)\right].\end{aligned}$$

For the denominator, $\mathbb{P}_n Y(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X} = PY(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}+o_{uP}(1)$.

For the numerator, since $e^x - 1 = x + \frac{1}{2}x^2 + O(x^3) = x + \frac{1}{2}x^2(1+O(x)) = x + \frac{1}{2}x^2(1+o(1))$ as $x\to 0$,

$$\begin{aligned}&\mathbb{P}_n Y(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\left(\exp\left(\Delta_n\right)-1\right)\\=&\mathbb{P}_n\left[Y(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\Delta_n\right]+1/2\mathbb{P}_n\left[Y(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\Delta_n^2\right]\left[1+o_{uP}(1)\right].\end{aligned}$$

Since in this section, we are interested in asymptotics only, all those $o_{uP}(1)$ terms are uniformly negligible and can be omitted. Then

$$I_9 = \int_0^\tau \frac{\mathbb{P}_n\left[Y(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\Delta_n\right] + \frac{1}{2}\mathbb{P}_n\left[Y(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\Delta_n^2\right]}{PY(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}} n\mathbb{P}_n dN(u).$$

Combine I_8 and I_9 to get

$$\begin{aligned}
&n(\mathbb{M}_n(\pi_0 + h_{\pi_{r_n}}) - \mathbb{M}_n(\pi_0)) = I_8 - I_9 \\
&= n\mathbb{P}_n\left[\Delta_n N(\tau)\right] - \int_0^\tau \frac{\mathbb{P}_n\left[Y(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\Delta_n\right]}{PY(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}} n\mathbb{P}_n dN(u) \\
&\qquad - \int_0^\tau \frac{\frac{1}{2}\mathbb{P}_n\left[Y(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\Delta_n^2\right]}{PY(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}} n\mathbb{P}_n dN(u) \\
&=_d \sqrt{n}\mathbb{P}_n\left[\left(\left(\beta_0 + \frac{h_\beta}{\sqrt{n}}\right) W_H(h_\theta) + h_\beta Z_H\left(\theta_0\right) + h_\gamma X\right) N(\tau)\right] \\
&- \int_0^\tau \frac{\mathbb{P}_n Y(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\left[\left(\beta_0 + \frac{h_\beta}{\sqrt{n}}\right) W_H(h_\theta) - h_\beta Z_H(\theta_0) + h_\gamma X\right]}{PY(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}} \sqrt{n}\mathbb{P}_n dN(u) \\
&- \int_0^\tau \frac{\frac{1}{2}\mathbb{P}_n Y(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\left[\left(\beta_0 + \frac{h_\beta}{\sqrt{n}}\right) W_H(h_\theta) - h_\beta Z_H(\theta_0) + h_\gamma X\right]^2}{PY(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}} \mathbb{P}_n dN(u),
\end{aligned}$$

where the second equality (in distribution) holds by the self-similarity property of fractional Brownian motion.

The right hand side of the last equality (in distribution) can be further written as

$$\begin{aligned}
&\sqrt{n}\mathbb{P}_n\left[\left(\beta_0 W_H(h_\theta) + h_\beta Z_H\left(\theta_0\right) + h_\gamma X\right) N(\tau)\right] \\
&- \int_0^\tau \frac{\mathbb{P}_n Y(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\left[\beta_0 W_H(h_\theta) + h_\beta Z_H(\theta_0) + h_\gamma X\right]}{PY(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}} \sqrt{n}\mathbb{P}_n dN(u) \\
&- \int_0^\tau \frac{\frac{1}{2}\mathbb{P}_n Y(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\left[\beta_0 W_H(h_\theta) + h_\beta Z_H(\theta_0) + h_\gamma X\right]^2}{PY(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}} \mathbb{P}_n dN(u),
\end{aligned} \tag{2.14}$$

where we omitted all the $\frac{h_\beta}{\sqrt{n}}$ terms. Since we are interested in asymptotics, all the $\frac{h_\beta}{\sqrt{n}}$ terms are uniformly negligible. We also used $PY(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}W_H(h_\theta) = 0$, $PW_H(h_\theta)N(\tau) = 0$ and Glivenko$-$Cantelli Theorem here.

Denote $\Delta \equiv \beta_0 W_H(h_\theta) + h_\beta Z_H(\theta_0) + h_\gamma X$, then (2.14) can be rewritten as

$$\begin{aligned}\sqrt{n}\mathbb{P}_n\left[N(\tau)\Delta\right] - \int_0^\tau \frac{\mathbb{P}_n\left[Y(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\Delta\right]}{PY(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}}\sqrt{n}\mathbb{P}_n dN(u) \\ - \int_0^\tau \frac{\frac{1}{2}\mathbb{P}_n\left[Y(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\Delta^2\right]}{PY(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}}\mathbb{P}_n dN(u).\end{aligned} \tag{2.15}$$

Further simplification by approximation

The last term of (2.15) converges uniformly (by Glivenko−Cantelli Theorem) to a deterministic function

$$\int_0^\tau \frac{\frac{1}{2}P\left[Y(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\Delta^2\right]}{P\left[Y(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\right]}PdN(u) = \frac{1}{2}\int_0^\tau \lambda_0(u)P\left[Y(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\Delta^2\right]du.$$

The first and second terms of (2.15) can be written in the following form and further decomposed into empirical process part and expectation part.

$$\begin{aligned}&\sqrt{n}\mathbb{P}_n\left(\Delta\cdot\left[N(\tau) - \int_0^\tau \frac{Y(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}}{PY(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}}\mathbb{P}_n dN(u)\right]\right)\\ =&\sqrt{n}(\mathbb{P}_n - P)\left(\Delta\cdot\left[N(\tau) - \int_0^\tau \frac{Y(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}}{PY(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}}\mathbb{P}_n dN(u)\right]\right)\\ &+\sqrt{n}P\left(\Delta\cdot\left[N(\tau) - \int_0^\tau \frac{Y(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}}{PY(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}}\mathbb{P}_n dN(u)\right]\right).\end{aligned}$$

The expectation part

$$\begin{aligned}&\sqrt{n}P\left(\Delta\cdot\left[N(\tau) - \int_0^\tau \frac{Y(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}}{PY(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}}\mathbb{P}_n dN(u)\right]\right)\\ =&\sqrt{n}P\left(\Delta\cdot\left[N(\tau) - \int_0^\tau \frac{Y(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}}{PY(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}}dN(u)\right]\right)\\ =&\sqrt{n}P\left(\Delta\cdot\left[N(\tau) - \int_0^\tau Y(u)\lambda_0(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}du\right]\right) = 0,\end{aligned}$$

where conditioning argument is used to get through the second equality.

The empirical process part (the leading term) can be approximated by

$$\begin{aligned}&\sqrt{n}(\mathbb{P}_n - P)\left(\Delta \cdot \left[N(\tau) - \int_0^\tau \frac{Y(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}}{PY(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}} PdN(u)\right]\right)\\ &= \sqrt{n}(\mathbb{P}_n - P)\left(\Delta \cdot \left[N(\tau) - \int_0^\tau Y(u)\lambda_0(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}du\right]\right),\end{aligned}$$

and the approximation error

$$\sqrt{n}(\mathbb{P}_n - P)\left(\Delta \cdot \left[\int_0^\tau \frac{Y(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}}{PY(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}}(\mathbb{P}_n - P)dN(u)\right]\right)$$

can be shown to converge uniformly to 0 by a slight generalization of Theorem 2.1 in van der Vaart and Wellner [55]. To apply Theorem 2.1, let $H_0 = [0,1]$ so that $\mathbb{P}_n N(u) \in H_0$ for any $u \in [0,\tau]$, we just need to verify two conditions. One is the class of functions

$$\begin{aligned}&\left\{\left(\beta_0 W_H(h_\theta) + h_\beta Z_H(\theta_0) + h_\gamma X\right)\left[N(\tau) - \int_0^\tau Y(u)\lambda_0(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}du\right]:\right.\\ &\quad \left.|h_\theta| \le K, |h_\beta| \le K, |h_\gamma| \le K\right\}\end{aligned}$$

is P-Donsker, which will be proved later in next section (Empirical process part). Another is

$$\sup_{|h_\theta|\le K,|h_\beta|\le K,|h_\gamma|\le K} P\left(\Delta \cdot \int_0^\tau \frac{Y(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}}{PY(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}}(\mathbb{P}_n - P)dN(u)\right)^2 \to 0.$$

We verify the second condition as follows.

$$\begin{aligned}&\sup_{|h_\theta|\le K,|h_\beta|\le K,|h_\gamma|\le K} P\left(\Delta \cdot \int_0^\tau \frac{Y(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}}{PY(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}}(\mathbb{P}_n - P)dN(u)\right)^2\\ &= \sup_{|h_\theta|\le K,|h_\beta|\le K,|h_\gamma|\le K} P\left(\Delta^2 \cdot P\left(\int_0^\tau \left(\frac{Y(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}}{PY(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}}\right)^2 \frac{dN(u)}{n} \mid Z_H(\theta_0), X\right)\right)\\ &= \sup_{|h_\theta|\le K,|h_\beta|\le K,|h_\gamma|\le K} P\left(\frac{\Delta^2}{n} \cdot \int_0^\tau \frac{P(Y(u)dN(u)|Z_H(\theta_0), X)}{(PY(u)|Z_H(\theta_0), X)^2}\right)\end{aligned}$$

$$= \sup_{|h_\theta|\le K, |h_\beta|\le K, |h_\gamma|\le K} P\left(\frac{\Delta^2}{n}\int_0^\tau \frac{P(dN(u)|Z_H(\theta_0),X)}{\left[P\left(Y(u)|Z_H(\theta_0),X\right)\right]^2}\right)$$

$$= \sup_{|h_\theta|\le K, |h_\beta|\le K, |h_\gamma|\le K} P\left(\frac{\Delta^2}{n}e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\int_0^\tau \frac{\lambda_0(u)du}{P\left(Y(u)|Z_H(\theta_0),X\right)}\right)$$

$\sim n^{-1}$, which goes to 0 as $n\to\infty$ uniformly over $H_{\pi_{r_n}}$.

Here

$$P\left(Y(u)|Z_H(\theta_0),X\right) = P(T\ge u, C\ge u|Z_H(\theta_0),X)$$
$$= \exp(-e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\int_0^u \lambda_0(s)ds)P(C\ge u|Z_H(\theta_0),X).$$

Now the rescaled localized criterion function is uniformly approximated by

$$\sqrt{n}(\mathbb{P}_n - P)\left(\Delta\cdot M(\tau)\right) - \frac{1}{2}P\left(\Delta^2\cdot\int_0^\tau Y(u)\lambda_0(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}du\right),$$

where $M(\tau)\equiv\left[N(\tau)-\int_0^\tau Y(u)\lambda_0(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}du\right]$.

The deterministic part of this approximation

$$-\frac{1}{2}P\left(\Delta^2\cdot\int_0^\tau Y(u)\lambda_0(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}du\right)$$
$$= -1/2\beta_0^2|h_\theta|^{2H}PN(\tau) - 1/2h_\beta^2 P[Z_H^2(\theta_0)N(\tau)] - 1/2h_\gamma^2 P[X^2N(\tau)]$$
$$- h_\beta h_\gamma P[Z_H(\theta_0)XN(\tau)].$$

The empirical process part of this approximation will be handled in the next section.

Empirical process part

We will prove the empirical process part converges to a mean-zero Gaussian process by the Donsker property of the collection of functions

$$\left\{\left(\beta_0 W_H(h_\theta) + h_\beta Z_H\left(\theta_0\right) + h_\gamma X\right)\left[N(\tau) - \int_0^\tau Y(u)\lambda_0(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}du\right]:\right.$$
$$\left.|h_\theta|\le K, |h_\beta|\le K, |h_\gamma|\le K\right\}.$$

The finite entropy integral property with $L_2(P)$ bracketing for the collection of functions

$$\left\{\beta_0 W_H(h_\theta) + h_\beta Z_H\left(\theta_0\right) + h_\gamma X: \quad |h_\theta|\le K, |h_\beta|\le K, |h_\gamma|\le K\right\}$$

can be obtained similarly as Lemma C.3.1 in Appendix C, so by adjusting its brackets (timing a function $\left[N(\tau) - \int_0^\tau Y(u)\lambda_0(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}du\right]$), it follows that the collection of functions in the preceding paragraph also has finite entropy integral with $L_2(P)$ bracketing. The $L_2(P)$ norm of its envelope is also bounded (which is easy to show), hence the uniform convergence to a mean-zero Gaussian process is justified.

To determine the asymptotic distribution of the empirical process part, we can just show

$$\begin{gathered}\mathbb{G}_n\left[(W_H(h_\theta)M(\tau), Z_H(\theta_0)M(\tau), XM(\tau))\right] \\ \to_w \left(\sqrt{PN(\tau)}W_H(h_\theta), \sqrt{P(Z_H^2(\theta_0)M^2(\tau))}Z_1, \sqrt{P(X^2M^2(\tau))}X_1\right),\end{gathered}$$

where (Z_1, X_1) follows 2-dimensional normal distribution with mean $(0,0)$, variance $(1,1)$ and covariance

$$\frac{P\left[Z_H(\theta_0)XM^2(\tau)\right]}{\sqrt{P\left[Z_H^2(\theta_0)M^2(\tau)\right]P\left[X^2M^2(\tau)\right]}}$$

and independent of 2-sided standard fBm $W_H(h_\theta)$ starting from zero with unit variance.

Since $W_H(h_\theta)$ is independent of $M(\tau)$, by the property of fBm,

$$\mathbb{G}_n\left[W_H(h_\theta)M(\tau)\right] =_d W_H(h_\theta)\sqrt{\mathbb{P}_n M^2(\tau)} \to \sqrt{PN(\tau)}W_H(h_\theta),$$

where we used the result $PM^2(\tau) = PN(\tau)$ from counting process theory.

The covariance structure of the joint asymptotic distribution is justified by the covariance terms between the three terms

$$\begin{aligned}\mathrm{Cov}(W_H(h_\theta)M(\tau), Z_H(\theta_0)M(\tau)) &= P\left[Z_H(\theta_0)M^2(\tau)\right]P\left[W_H(h_\theta)\right] = 0, &(2.16)\\ \mathrm{Cov}(W_H(h_\theta)M(\tau), XM(\tau)) &= P\left[XM^2(\tau)\right]P\left[W_H(h_\theta)\right] = 0, &(2.17)\\ \mathrm{Cov}(Z_H(\theta_0)M(\tau), XM(\tau)) &= P\left[Z_H(\theta_0)XM^2(\tau)\right].\end{aligned}$$

2.3.4.5 Asymptotics of $(\hat{\theta}_n, \hat{\beta}_n, \hat{\gamma}_n)$

We find $\hat{\theta}_n$ is asymptotically independent of $(\hat{\beta}_n, \hat{\gamma}_n)$, becasue the corresponding covariance terms in (2.16) and (2.17) vanish.

For the sum of the asymptotic distribution of the empirical process part and the deterministic part, we extract the terms which are relevant to h_θ. They are

$$\beta_0\sqrt{PN(\tau)}W_H(h_\theta) - 1/2\beta_0^2|h_\theta|^{2H}PN(\tau).$$

By the Argmax Continuous Mapping theorem and the symmetry of $W_H(\cdot)$ about zero, we obtain the estimator $\hat{\theta}_n$ has asymptotic distribution

$$n^{1/(2H)}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{W}} \text{argmax}_h\left(W_H(h) - \frac{|\beta_0||h|^{2H}}{2}\sqrt{PN(\tau)}\right).$$

For the case of general $\sigma > 0$, following the lines starting from Chapter 2.3.4.1, we will obtain

$$n^{1/(2H)}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{W}} \text{argmax}_h\left(W_H(h) - \frac{|\beta_0||h|^{2H}}{2}\sigma\sqrt{PN(\tau)}\right),$$

where $W_H(\cdot)$ is a standard 2-sided fractional Brownian motion starting from zero with unit variance scale (i.e. $W_H(1) =_d W_H(-1) \sim N(0,1)$).

Besides the terms relevant to h_θ, all other parts are relevant to h_β and h_γ and converges uniformly to a process which is equivalent (in distribution) to

$$\begin{aligned}&\left(h_\beta\sqrt{P(Z_H^2(\theta_0)M^2(\tau))}Z_1 + h_\gamma\sqrt{P(X^2M^2(\tau))}X_1\right)\\ &-1/2h_\beta^2P[Z_H^2(\theta_0)N(\tau)] - 1/2h_\gamma^2P[X^2N(\tau)] - h_\beta h_\gamma P[Z_H(\theta_0)XN(\tau)],\end{aligned} \tag{2.18}$$

where (Z_1, X_1) follows 2-dimensional normal distribution with mean $(0,0)$, variance $(1,1)$ and covariance

$$\frac{P\left[Z_H(\theta_0)XM^2(\tau)\right]}{\sqrt{P\left[Z_H^2(\theta_0)M^2(\tau)\right]P\left[X^2M^2(\tau)\right]}}.$$

It is easy to show the strict concavity of (2.18) w.r.t. (h_β, h_γ), so there is a unique maximizer of the process for each realized sample path of the limit process. By setting the first derivatives to be zero,

$$\begin{cases} h_\beta P[Z_H^2(\theta_0)N(\tau)] + h_\gamma P[Z_H(\theta_0)XN(\tau)] &= \sqrt{P(Z_H^2(\theta_0)M^2(\tau))}Z_1; \\ h_\gamma P[X^2N(\tau)] + h_\beta P[Z_H(\theta_0)XN(\tau)] &= \sqrt{P(X^2M^2(\tau))}X_1. \end{cases}$$

Solving the equations jointly, we obtain

$$\begin{cases} h_\beta = \frac{\sqrt{-P[X^2M^2(\tau)]}P[Z_H(\theta_0)XN]X_1 + \sqrt{P[Z_H^2(\theta_0)M^2(\tau)]}P[X^2N]Z_1}{P\left[Z_H^2(\theta_0)N(\tau)\right]P[X^2N(\tau)] - P^2[Z_H(\theta_0)XN(\tau)]}; \\ h_\gamma = \frac{\sqrt{P[X^2M^2(\tau)]}P[Z_H^2(\theta_0)N]X_1 - \sqrt{P[Z_H^2(\theta_0)M^2(\tau)]}P[Z_H(\theta_0)XN]Z_1}{P\left[Z_H^2(\theta_0)N(\tau)\right]P[X^2N(\tau)] - P^2[Z_H(\theta_0)XN(\tau)]}. \end{cases}$$

By the Argmax Continuous Mapping Theorem, the asymptotic distribution of the vector $(\sqrt{n}(\hat\beta_n - \beta), \sqrt{n}(\hat\gamma_n - \gamma))$ converges to that of

$$\left(\frac{\sqrt{-P[X^2M^2(\tau)]}P[Z_H(\theta_0)XN]X_1 + \sqrt{P[Z_H^2(\theta_0)M^2(\tau)]}P[X^2N]Z_1}{P\left[Z_H^2(\theta_0)N(\tau)\right]P\left[X^2N(\tau)\right] - P^2\left[Z_H(\theta_0)XN(\tau)\right]}, \right.$$
$$\left. \frac{\sqrt{P[X^2M^2(\tau)]}P[Z_H^2(\theta_0)N]X_1 - \sqrt{P[Z_H^2(\theta_0)M^2(\tau)]}P[Z_H(\theta_0)XN]Z_1}{P\left[Z_H^2(\theta_0)N(\tau)\right]P\left[X^2N(\tau)\right] - P^2\left[Z_H(\theta_0)XN(\tau)\right]} \right).$$

Since (Z_1, X_1) follows 2-dimensional normal distribution with mean $(0,0)$, variance $(1,1)$ and covariance

$$\frac{P\left[Z_H(\theta_0)XM^2(\tau)\right]}{\sqrt{P\left[Z_H^2(\theta_0)M^2(\tau)\right]P\left[X^2M^2(\tau)\right]}},$$

let (Y_1, Y_2) follow the asymptotic distribution of $\left(\sqrt{n}(\hat\beta_n - \beta), \sqrt{n}(\hat\gamma_n - \gamma)\right)$, then it is 2-dimensional normal distribution with mean $(0,0)$ and variance-covariance components

$$\begin{cases} \mathrm{Var}_1 = \frac{P(X^2M^2)P^2(ZXN) + P(Z^2M^2)P^2(X^2N) - 2P(X^2N)P(ZXN)P(ZXM^2)}{(P(Z^2N)P(X^2N) - P^2(ZXN))^2}, \\ \mathrm{Var}_2 = \frac{P(X^2M^2)P^2(Z^2N) + P(Z^2M^2)P^2(ZXN) - 2P(Z^2N)P(ZXN)P(ZXM^2)}{(P(Z^2N)P(X^2N) - P^2(ZXN))^2}, \\ \mathrm{Cov} = \frac{P(ZXM^2)[P^2(ZXN) + P(X^2N)P(Z^2N)] - P(ZXN)[P(X^2M^2)P(Z^2N) - P(Z^2M^2)P(X^2N)]}{(P(Z^2N)P(X^2N) - P^2(ZXN))^2}, \end{cases}$$

where we abbreviated $(Z_H(\theta_0), M(\tau))$ as (Z, M), and $(\mathrm{Var}(Y_1), \mathrm{Var}(Y_2), \mathrm{Cov}(Y_1, Y_2))$ as $(\mathrm{Var}_1, \mathrm{Var}_2, \mathrm{Cov})$ respectively.

Using the property of counting process and conditioning argument, we have

$$P[f(Z_H(\theta_0), X)(N(\tau) - M^2(\tau))] = 0,$$

for $f(x_1, x_2) = x_1$ or $f(x_1, x_2) = x_2$ or $f(x_1, x_2) = x_1^2$ or $f(x_1, x_2) = x_2^2$ or $f(x_1, x_2) = x_1 x_2$. Plugging in the formulas for $\mathrm{Var}(Y_1), \mathrm{Var}(Y_2), \mathrm{Cov}(Y_1, Y_2)$, we have

$$\left\{ \begin{array}{l} \mathrm{Var}(Y_1) = \frac{P(X^2 N(\tau))}{P(Z^2 N)P(X^2 N) - P^2(ZXN)}, \\ \mathrm{Var}(Y_2) = \frac{P(Z^2 N(\tau))}{P(Z^2 N)P(X^2 N) - P^2(ZXN)}, \\ \mathrm{Cov}(Y_1, Y_2) = \frac{-P(ZXN(\tau))}{P(Z^2 N)P(X^2 N) - P^2(ZXN)}. \end{array} \right.$$

Denote
$$s_1^2 = \frac{P(X^2 N(\tau))}{P(Z_H^2(\theta_0)N(\tau))P(X^2 N(\tau)) - P^2(Z_H(\theta_0)XN(\tau))},$$
$$s_2^2 = \frac{P(Z_H^2(\theta_0)N(\tau))}{P(Z_H^2(\theta_0)N(\tau))P(X^2 N(\tau)) - P^2(Z_H(\theta_0)XN(\tau))},$$
$$\rho = \frac{-P(Z_H(\theta_0)XN(\tau))}{\sqrt{P(X^2 N(\tau))P(Z_H^2(\theta_0)N(\tau))}},$$

then
$$\begin{pmatrix} \sqrt{n}(\hat{\beta}_n - \beta) \\ \sqrt{n}(\hat{\gamma}_n - \gamma) \end{pmatrix} \rightarrow_w N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} s_1^2 & \rho s_1 s_2 \\ \rho s_1 s_2 & s_2^2 \end{pmatrix} \right).$$

Chapter 3

CI calibrated by Monte Carlo

To construct the Wald-type confidence intervals for θ, we need to determine the quantiles of the asymptotic distribution of $\hat{\theta}_n$ in Theorems 2.2.2 and 2.3.2.

For the asymptotic distribution in Theorem 2.2.2, there is a closed form C.D.F. which can be used to obtain the quantiles and hence the confidence intervals, easily and accurately. For the asymptotic distribution in Theorem 2.3.2, however, there is no such closed form C.D.F. in general to the author's best knowledge. One possible way is to get the quantiles and hence the confidence intervals, by Monte Carlo calibration.

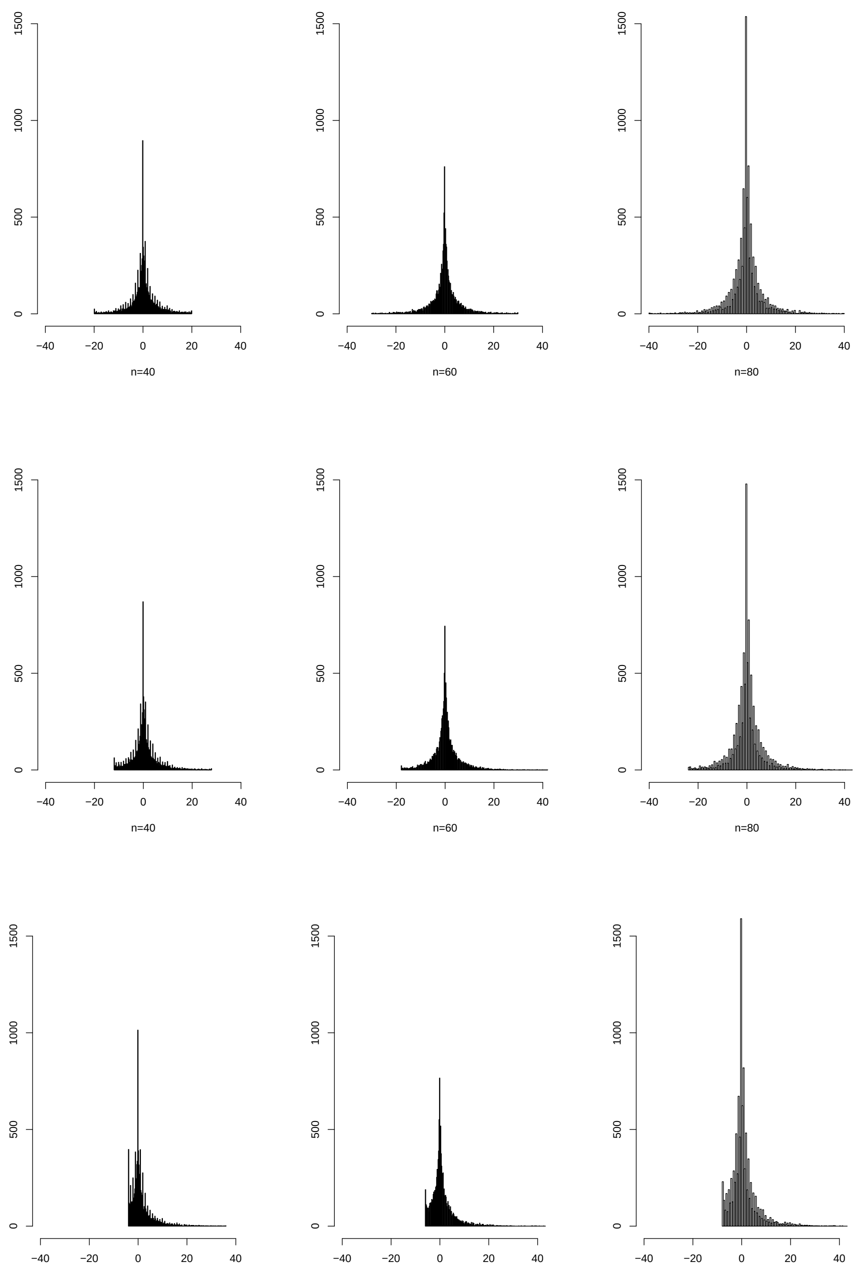
For reasons to be discussed in Chapter 4.1.2, instead of calibrating the quantiles of the asymptotic distribution, we use Monte Carlo method to calibrate the quantiles of the "Domain-Restricted Asymptotic Distribution" (abbreviated as "DRAD" hereafter). The motivation and definition of the DRAD can be found in Chapter 4.1.2.

At the end of this chapter, we give a survey of the proposed procedure.

3.1 Quantiles for the simple case of the Cox model

Even though the analytical distribution function is available for the asymptotic distribution of $\hat{\theta}_n$ in the simplified model, the analytical distribution function for the DRAD

Figure 3.1: Histograms of $n(\hat{\theta}_n - \theta_0)$ for simulated random variable $\hat{\theta}_n$ that follows the domain-restricted asymptotic C.D.F. for finite sample size n, $(\theta_0, \sigma) = (0.5, 1)$ for the upper row, $(0.1, 1)$ for the middle row, $(0.1, 3)$ for the lower row



is unavailable to the author's best knowledge. However, we can run simulations to learn about the DRAD.

For each of $\theta = 0.5, 0.3$ and 0.1, each of $n = 40, 60$ and 80, we generate 10,000 replicates of the trajectory $\{W(h) - \frac{|h|}{2}\sigma\sqrt{PN(\tau)} : h \in [-n\theta_0, n(1-\theta_0)]\}$ on a fine grid with $J = 240$ evenly spaced points. For each replicate of the trajectory, compare the 240 grid points to find the grid point that has the maximum value of the trajectory. Then we obtain 10,000 simulated random variables that follow the DRAD.

Looking at their histograms in Figure 3.1, we find their distributions capture the features of the empirical distributions of $\hat{\theta}_n$ (i.e., asymmetry and boundary-cluster phenomena) that we obtained for the estimates of the simple Cox model in Chapter 4.1.2. Numerical results in Table 3.1, i.e., the empirical tail probabilities of quantiles based on the asymptotic distribution and the DRAD show that DRAD is more preferable for the purpose of confidence interval construction.

Table 3.1 shows that for $\theta = 0.5$, which is in the middle of $[0, 1]$, the two quantiles' empirical tail coverage probabilities are comparable. For θ which deviates from 0.5, for the tail which is further away from θ, the DRADs' quantiles and the asymptotic distribution's quantiles have comparable empirical tail coverage proabilities. However, on the tail which is on the same side as of θ_0, the DRADs' quantiles have more reasonable empirical tail coverage probabilities. Such an advantage is more obvious for smaller n.

3.2 Quantiles for the extended case of the Cox model

In Chapter 4.1.2, we observe that the empirical distributions (the histograms of the estimates obtained from simulated Cox model) are asymmetric, which is contrary to the symmetric distribution of the asymptotic distribution. In the extended Cox model, similar situation appears for $n^{1/(2H)}(\hat{\theta}_n - \theta_0)$ as well. So we need to consider the DRAD.

Table 3.1: (H=0.5) Empirical tail coverage probabilities of (simulated) DRADs' quantiles (the upper part) and (analytical) asymptotic distribution's quantiles (the lower part) for finite sample sizes

θ_0	σ	n	$q_{.975}$	$q_{.95}$	$q_{.05}$	$q_{.025}$
0.5	1	40	.046	.073	.080	.047
		60	.041	.075	.075	.046
		80	.033	.061	.075	.041
0.3	1	40	.054	.093	.085	.040
		60	.048	.077	.060	.029
		80	.042	.075	.072	.040
0.1	1	40	.063	.095	.056	.029
		60	.052	.079	.055	.022
		80	.054	.080	.068	.028
0.5	1	40	.040	.067	.072	.038
		60	.041	.075	.073	.041
		80	.033	.066	.073	.041
0.3	1	40	.048	.092	.038	.000
		60	.050	.080	.051	.016
		80	.041	.077	.068	.035
0.1	1	40	.062	.095	.000	.000
		60	.050	.079	.000	.000
		80	.051	.081	.000	.000

There is yet another incentive to investigate the DRAD for the extended Cox model. If we choose the asymptotic distribution as the benchmark distribution, for inference purposes, we have to obtain the quantiles of the asymptotic distribution

$$\text{argmax}_h\bigg(W_H(h) - \frac{|h|^{2H}}{2}|\beta|\sigma\sqrt{PN(\tau)}\bigg).$$

For general H (instead of the special case $H = 1/2$), however, we do not have closed-form C.D.F. for this distribution. Hence we can't solve an analytic equation to get the quantiles. An alternative way is to get the quantiles through Monte Carlo. To run Monte Carlo, it is technically not feasible to maximize the trajectory of the $W_H(h) - \frac{|h|^{2H}}{2}|\beta|\sigma\sqrt{PN(\tau)}$ over an infinite interval. So we have to restrict the domain to be of finite length. Hence the DRAD is more appealing compared to the asymptotic distribution for the purpose of determining quantiles as well.

We study the quantiles of the DRAD

$$\text{argmax}_{h\in[-n^{1/(2H)}\theta_0, n^{1/(2H)}(1-\theta_0)]}\bigg(W_H(h) - \frac{|h|^{2H}}{2}|\beta|\sigma\sqrt{PN(\tau)}\bigg),$$

by Monte Carlo.

We simulate 10,000 replicates for each of $H = 0.5, 0.7$, $n = 120, 180, 240$, $\theta_0 = 0.5, 0.3, 0.1$ and $\sigma = 1, 2, 3$. To make the calibrated quantiles useful for the simulation setting described in Chapter 4.2.1, we take $(\beta_0, \gamma_0) = (1, 0)$.

To simulate this distribution, we need the value of $PN(\tau)$. $PN(\tau)$ can be calculated based on the given parameters.

To match the setting of steps 3 and 5 of the simulation procedure in Chapter 4.2.1, $Z(\theta_0) \sim U[-0.75, 0.75]$, $X \sim U[-1, 1]$, $T^0 \sim \text{Exp}\left(\exp(\beta Z(\theta_0) + \gamma X)\right)$ with $(\beta, \gamma) = (1, 0)$. $C \sim \text{Exp}\left(|Z(\theta_0)|\right)$, $\tau = 50$ which are the same as of the simulation procedure of the simple model in Chapter 4.1.1.

Then

$$\begin{aligned}P[N(\tau)] &= P[1_{T^0\le\tau,T^0\le C}] = P\left[P\left(T^0 \le \tau, T^0 \le C|Z(\theta_0), X\right)\right]\\ &= P\left[\int_0^\tau e^{Z(\theta_0)}e^{-e^{Z(\theta_0)}u}\cdot e^{-u|Z(\theta_0)|}du\right] = P\left[\cdot e^{Z(\theta_0)}\frac{1-e^{-\tau(e^{Z(\theta_0)}+|Z(\theta_0)|)}}{e^{Z(\theta_0)}+|Z(\theta_0)|}\right]\\ &= P\left[e^{Z(\theta_0)}\frac{1-e^{-\tau(e^{Z(\theta_0)}+|Z(\theta_0)|)}}{e^{Z(\theta_0)}+|Z(\theta_0)|}\right] = P\left[e^{Z(\theta_0)}\frac{1-e^{-50(e^{Z(\theta_0)}+|Z(\theta_0)|)}}{e^{Z(\theta_0)}+|Z(\theta_0)|}\right],\end{aligned}$$

where we used the properties of the distribution of T^0, C in the third equality.

Here $Z(\theta_0) \sim U[-0.75, 0.75]$, there is no analytical answer to $P[N(\tau)]$. Using 10,000,000 Monte-Carlo replicates we obtain the estimate of $P[N(u)] = 0.7404991$.

To obtain precise quantiles of the DRADs requires a dense grid on its domain, and hence the number of grid points J on the interval $[-n^{1/(2H)}\theta_0, n^{1/(2H)}(1-\theta_0)]$ needs to be large. Since the interval length grow with the decrease of H, it requires larger J for smaller H. To obtain acceptable precision under the computing ability constraint, the number of grid points are taken to be $J = 720$ for $H = 0.5$ and $J = 240$ for $H = 0.7$.

Their quantiles are listed in Tables 3.2 (for $H = 0.5$) and 3.3 (for $H = 0.7$). For $H = 0.3$, to obtain quantiles with acceptable precision poses formidable computing challenge, so we do not calculate their quantiles here.

3.3 Summary of the proposed procedure

1. For each component $Z(\theta_j)(j = 1, \ldots, J$, where J is the number of grid points observed) of the functional covariate Z, we choose it as the predictor and may add other non-functional covariates (either scalar or vector) X as another predictor for time-to-event risk. Then we can fit the classical Cox model using $\left(Z(\theta_j), X\right)$ as predictors by the "coxph" function in R, using the package "survival". As a result, we obtain a log-partial-likelihood value $\log \mathrm{PL}(\theta_j)$, and the estimated coefficients $\left(\hat\beta_j, \hat\gamma_j\right)$.

Table 3.2: (H=0.5) Quantiles of the DRADs for finite sample sizes by Monte Carlo

θ_0	σ	n	$q_{.95}$	$q_{.975}$	$q_{.05}$	$q_{.025}$
0.5	1	120	10.33	15.00	-10.17	-14.33
		180	10.75	15.26	-10.50	-15.25
		240	10.67	15.67	-10.67	-15.00
	2	120	2.50	3.50	-2.50	-3.83
		180	2.50	3.75	-2.75	-3.75
		240	2.67	3.67	-2.33	-3.67
	3	120	1.17	1.67	-1.17	-1.67
		180	1.25	1.75	-1.25	-1.75
		240	1.00	1.67	-1.00	-1.67
0.3	1	120	10.01	14.17	-10.17	-14.34
		180	10.25	14.75	-11.00	-15.25
		240	10.33	15.00	-10.33	-14.33
	2	120	2.50	3.67	-2.83	-3.83
		180	2.50	3.75	-2.50	-3.50
		240	2.33	3.33	-2.67	-3.67
	3	120	1.17	1.67	-1.17	-1.67
		180	1.25	1.75	-1.00	-1.75
		240	1.00	1.67	-1.00	-1.67
0.1	1	120	9.83	14.50	-8.17	-10.17
		180	10.25	15.00	-9.25	-12.25
		240	10.67	15.67	-10.00	-13.67
	2	120	2.67	3.83	-2.67	-3.67
		180	2.50	3.50	-2.50	-3.75
		240	2.67	3.33	-2.67	-3.67
	3	120	1.17	1.67	-1.17	-1.67
		180	1.00	1.75	-1.25	-1.75
		240	1.33	1.67	-1.00	-1.67

Table 3.3: (H=0.7) Quantiles of the DRADs for finite sample sizes by Monte Carlo

θ_0	σ	n	$q_{.95}$	$q_{.975}$	$q_{.05}$	$q_{.025}$
0.5	1	120	3.82	5.35	-3.95	-5.22
		180	3.91	5.10	-3.91	-5.27
		240	3.97	5.22	-3.97	-5.01
	2	120	1.53	1.91	-1.53	-1.91
		180	1.53	1.87	-1.53	-2.04
		240	1.46	1.88	-1.46	-1.88
	3	120	0.76	1.15	-0.89	-1.15
		180	0.85	1.19	-0.85	-1.02
		240	0.84	1.04	-0.84	-1.04
0.3	1	120	3.95	5.35	-4.20	-5.35
		180	3.91	5.27	-3.91	-5.27
		240	3.97	5.22	-3.97	-5.22
	2	120	1.53	2.04	-1.53	-1.91
		180	1.53	2.04	-1.53	-2.04
		240	1.46	2.09	-1.46	-2.09
	3	120	0.76	1.15	-0.89	-1.15
		180	0.85	1.02	-0.85	-1.02
		240	0.84	1.04	-0.84	-1.04
0.1	1	120	4.07	5.48	-2.67	-3.06
		180	3.91	5.27	-3.23	-3.91
		240	3.97	5.22	-3.55	-4.38
	2	120	1.53	2.04	-1.40	-1.91
		180	1.53	2.04	-1.53	-2.04
		240	1.46	1.88	-1.46	-1.88
	3	120	0.89	1.15	-0.76	-1.15
		180	0.85	1.02	-0.85	-1.02
		240	0.84	1.04	-0.84	-1.04

2. Compare $\log \mathrm{PL}(\theta_j)(j = 1, \ldots, J)$, locate the maximizer θ_{j^*} among them, and denote it as $\hat{\theta}_n$ (here n is the sample size, i.e., the number of subjects in the data set). The corresponding $\left(\hat{\beta}_{j^*}, \hat{\gamma}_{j^*}\right)$ can be denoted as $\left(\hat{\beta}_n, \hat{\gamma}_n\right)$.

3. The maximum partial likelihood estimator of (θ, β, γ) is obtained as $\left(\hat{\theta}_n, \hat{\beta}_n, \hat{\gamma}_n\right)$.

4. To prepare for statistical inference, we estimate the nuisance parameters (H, σ^2) in the following way.

5. The estimate of the Hurst exponent H can be obtained using the function "pengFit()" in the package "fArma" of R. For each subject's trajectory, we get an estimate of H. Take the mean of these n estimated Hs, we can get an estimate of H, $\hat{H}_n$.

6. The parameter σ^2 can be estimated by quadratic variation method:
$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{J-1} \left(Z_i(\theta_{j+1}) - Z_i(\theta_j)\right)^2.$$

7. The value of $PN(\tau)$ can be estimated by $\mathbb{P}_n N(\tau)$, the proportion of subjects who experienced events in the data set.

8. By Monte Carlo calibration, we get the .025 and .975 quantiles of the distribution
$$\operatorname{argmax}_{h \in [-n^{1/(2\hat{H}_n)}\hat{\theta}_n, n^{1/(2\hat{H}_n)}(1-\hat{\theta}_n)]} \left(W^{\hat{H}_n}(h) - \frac{|h|^{2\hat{H}_n}}{2} |\hat{\beta}_n| \hat{\sigma}_n \sqrt{\mathbb{P}_n N(\tau)} \right).$$

9. Denote the quantiles obtained in the previous step as $q_{.025}$ and $q_{.975}$, the 95% confidence interval of θ is $[\hat{\theta}_n + q_{.025} \cdot n^{-1/(2\hat{H})}, \hat{\theta}_n + q_{.975} \cdot n^{-1/(2\hat{H})}]$.

10. The confidence intervals of β and γ can be obtained easily using the quantiles of normal distributions by checking the quantile table of standard normal distribution, if we can estimate the covariance matrix in Theorem 2.3.2. The covariance matrix can be estimated by replacing all the $P[\cdot]$ with $\mathbb{P}_n[\cdot]$. For example, $P(X^2 N(\tau))$, the numerator of s_1^2, can be estimated by $\mathbb{P}_n(X^2 N(\tau))$.

Chapter 4

Simulations

The large sample properties of the maximum partial likelihood estimators for the proposed model have been explored in Chapter 2. In this chapter, we will evaluate the finite sample performance of these estimators by simulation studies.

For finite sample sizes, the asymptotic distribution of the location estimator $\hat{\theta}_n$ is symmetric, while the empirical distribution of $\hat{\theta}_n$ is asymmetric. To seek a more reasonable, i.e. asymmetric, approximation for finite sample size empirical distribution, we restrict the domain of the asymptotic distribution to define the (asymmetric) *"Domain-Restricted Asymptotic Distribution"* (abbreviated as "DRAD"). The .025, .05, .95 and .975 quantiles of the DRADs can be obtained through Monte Carlo calibration as shown in Chapter 3, and we call them *empirical critical values* or *empirical confidence limits*.

For both the simple model and the extended model, we simulate data sets from specified model parameters and get the estimates of (θ, β, γ) by the maximum partial likelihood method.

The empirical distributions of $n^{1/(2H)}(\hat{\theta}_n - \theta)$ are compared to the DRADs. Simulation results show that the DRADs provide a reasonable approximation for the empirical distributions when the sample size is relatively large.

The empirical distributions of $\sqrt{n}(\hat{\beta}_n - \beta)$ and $\sqrt{n}(\hat{\gamma}_n - \gamma)$ are compared to their asymptotic distributions (obtained in Theorem 3.2.3) respectively. Simulation results for $\sqrt{n}(\hat{\beta}_n - \beta)$ show that the approximations are still poor even for relatively large sample sizes (although the trend of better approximations with larger sample sizes is obvious). In contrast, simulation results for $\sqrt{n}(\hat{\gamma}_n - \gamma)$ show that the approximations are reasonable for all the sample sizes no less than 120.

All the simulations in this chapter and throughout this thesis are conducted using the statistical software R (version 2.13.0).

4.1 Simple case of the Cox model

For the simplified model proposed in Chapter 2.2, we describe the simulation procedure in Chapter 4.1.1. In Chapter 4.1.2, the histograms of $\hat{\theta}_n$ are displayed which are estimated from the data simulated from the Cox model. These empirical distributions (i.e., histograms) show asymmetry, in contrast to the symmetric property of asymptotic distributions obtained in Chapter 2.2. To resolve this issue, for finite n, we propose the "Domain-Restricted Asymptotic Distribution" to replace the asymptotic distribution.

4.1.1 Simulation procedure

In this section, we will describe how to simulate data sets from the simplified Cox model and obtain the estimates of θ.

For each of n subjects, we generate random variables $Z(\theta_0)$ and X, the trajectory of stochastic process $Z(\cdot)$ on a fine grid. then according to the simplified Cox model in Chapter 2.2, we generate its censoring time and time-to-event outcome. Then by the maximum partial likelihood principle, we can get an estimate of θ for the simplified Cox model based on these n subjects' data.

The procedure of simulations:

1. Without loss of generality, set the length of the interval $[0, \theta_M]$ to be $\theta_M = 1$,

so that we observe the value of covariate trajectory from 0 to 1, even though only to the fineness of grid size equal to 1/240 (i.e., the number of grid points on which we can observe the value of $\mathbb{Z}$ is $J = 240$).

2. Set the relative location of θ_0 on the whole interval $[0, \theta_M]$, i.e. $\theta_0 : \theta_M = 0.5$ (or 0.3, 0.1).

3. Generate $n = 40$ (or 60, 80) I.I.D. random variables $Z_1(\theta_0), \cdots, Z_n(\theta_0)$ which follows uniform distribution on $[-0.75, 0.75]$.

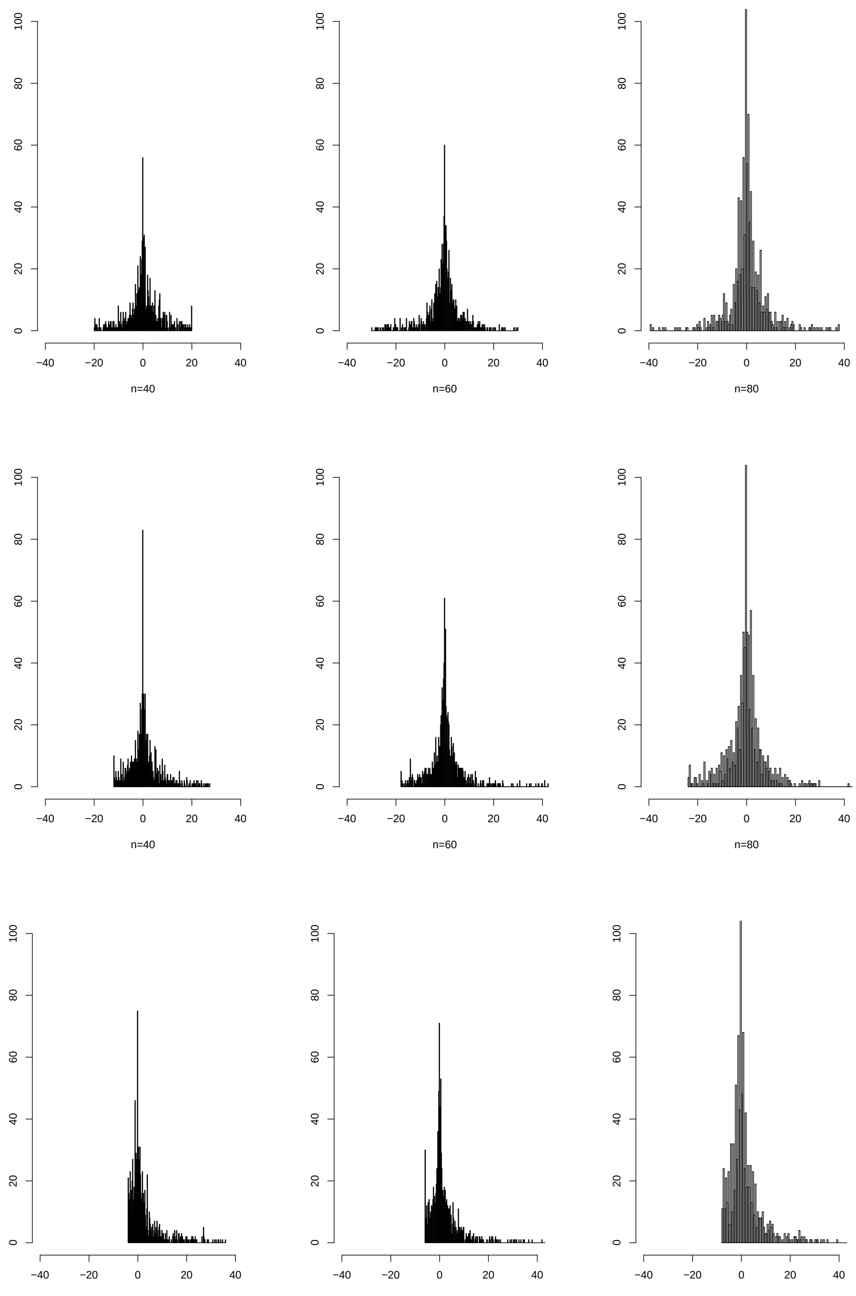
4. For the ith of the n subjects, generate 2 independent 1-sided S.B.M.s $W_{l,i}, W_{r,i}$ starting from 0 (also independent of $Z_i(\theta_0)$), and by transformation $Z_i(\theta) = Z_i(\theta_0) + W_{r,i}(\theta - \theta_0)$ for $\theta > \theta_0$ and $Z_i(\theta) = Z_i(\theta_0) + W_{l,i}(\theta_0 - \theta)$ for $\theta < \theta_0$, get a 2-sided S.B.M. $Z(\cdot)$ starting from θ_0 with variance $\sigma^2 = 1$ (or $2^2, 3^2$) (i.e., $W_l(1) =_d W_r(1) \sim N(0, \sigma^2)$). Do this step for $i = 1, \cdots, n$ independently. The simulation of S.B.M.s in this step is conducted by the function "fbmSim()" in package "fArma" in R.

5. For the ith of the n subjects, generate the censoring time C_i which follows exponential distribution with parameter $|Z_i(\theta_0)|$ and event time T_i^0 which follows exponential distribution with parameter $\exp(Z_i(\theta_0))$. So the censoring time T_i^0 and event time C_i are conditionally independent given $Z_i(\theta_0)$. Do this step for $i = 1, \cdots, n$ independently. The followup time $\tau = 50$. We observe only $(T_i^0 \wedge C_i \wedge \tau, 1_{T_i^0 \leq C_i, T_i \leq \tau})$.

6. For each grid point $j = 1, \cdots, J$, we calculate the corresponding value of the partial likelihood function, and compare them to get the j^* which maximize the partial likelihood.

7. Repeat the steps 3-6 for Rep=1000 times to obtain 1000 replicates of $\hat{\theta}_n$.

Figure 4.1: Histograms of $n(\hat{\theta}_n - \theta_0)$ for $\hat{\theta}_n$ that specified by the simple Cox model, $(\theta_0, \sigma) = (0.5, 1)$ for the upper row, $(0.3, 1)$ for the middle row, $(0.1, 1)$ for the lower row



4.1.2 Simulation results

For different parameter settings, the histograms of estimated $\hat{\theta}_n$ are shown in Figure 3.1.

The histograms show that as θ_0 (i.e., the true value of θ) deviates from 0.5 (i.e., the middle of the interval), the empirical distributions of the rescaled estimates $n(\hat{\theta}_n - \theta_0)$ become more asymmetric. We can observe the "boundary cluster" phenomena on the left boundary if the true θ are close to the left end of the interval $[0, 1]$. Similar phenomena will be observed on the right boundary if θ_0 gets close to the right end of the interval $[0, 1]$.

This phenomena can be explained as follows. The estimates $\hat{\theta}_n$ tend to lie around the true θ. If the interval length is infinite, the empirical distribution of $\hat{\theta}_n$ will be symmetric about θ. However, since the interval $[0, 1]$ has finite length, due to the constraint that $\hat{\theta}_n$ has to lie within $[0, 1]$, the half probability (which should have been assigned to the left of the true θ on the histogram of $\hat{\theta}_n$ if no constraint) that $\hat{\theta}_n < \theta_0$ has to be distributed within a short interval $[0, \theta]$ (for θ close to 0). This results in the boundary cluster phenomena on the left boundary of $[0, 1]$ (i.e., those values which should have been beyond the left end of the interval if no constraint exists are forced to cluster around the left end of the interval). Looking at the histograms of $n(\hat{\theta}_n - \theta)$, we can see the distributions are truncated at $-n\theta_0$.

The asymmetric empirical distributions suggest the asymptotic distribution derived in Chapter 2.2 does not provide a reasonable approximation for $n(\hat{\theta}_n - \theta)$ with finite sample sizes.

To resolve this issue, we refer to the proof of the asymptotic distribution in Chapter 2.2.2.4-2.2.2.5. For asymptotics, we considered $\mathbb{Q}_n(h) = s_n(\mathbb{M}_n(\theta_0 + h/n) - \mathbb{M}_n(\theta_0)), \forall h \in [-K, K], \forall K > 0$. For finite sample size, if we incorporate the constraint that $\theta_0 + h/n \in [0, 1]$, then $h \in [-n\theta_0, n(1 - \theta_0)]$.

Then in the finite sample scenario, we can replace the asymptotic distribution

$$\text{argmax}_{h\in(-\infty,\infty)}\left(W(h)-\frac{|h|}{2}\sigma\sqrt{PN(\tau)}\right),$$

by its domain-restricted version

$$\text{argmax}_{h\in[-n\theta_0,n(1-\theta_0)]}\left(W(h)-\frac{|h|}{2}\sigma\sqrt{PN(\tau)}\right),$$

and expect the DRAD to provide a more reasonable, i.e., asymmetric, approximation to empirical distributions of simulated estimates.

Figure 3.1 shows the shape of the approximate distributions of $n(\hat{\theta}_n - \theta_0)$ for different n and also implies that of the analytical asymptotic distribution as the limit case (i.e., n goes to infinity). Examining further reveals the ranges of the random variables are wider for larger n, which are predetermined by the definition of DRAD.

4.2 Extended case of the Cox model

For the extended model in Chapter 2.3, we follow the similar simulation procedure to the simplified model. Some adjustments to adapt the complexity of the extended model are described in Chapter 4.2.1.

The asymptotic distributions of $\hat{\beta}_n$ and $\hat{\gamma}_n$ for a specific parameter setting in Chapter 4.2.1 are further studied and compared to their corresponding empirical distributions of $\hat{\beta}_n$ and $\hat{\gamma}_n$ in Chapter 4.2.3.

4.2.1 Simulation procedure

We follow the same procedure as that of Chapter 4.1.1 except for the following changes.

1. The fineness of grid is set to be grid size equal to 1/100. So the number of grid points on which we can observe the value of $\mathbb{Z}$ is $J = 100$. We change it from 240 as of Chapter 4.1.1 to 100 due to the computing ability constraint.

2. Sample sizes $n = 120, 180, 240$ replaced $n = 40, 60, 80$ as of Chapter 4.1. Since in the extended Cox model, with more parameters to estimate, larger sample sizes are necessary for the asymptotic distribution to provide a reasonable approximation to the empirical distributions.

3. Besides $Z_1(\theta_0), \ldots, Z_n(\theta_0)$ which follow uniform distribution on $[-0.75, 0.75]$, we also generate n replicates of another covariate $X_1, \ldots, X_n$ (independent of $Z_1(\theta_0), \ldots, Z_n(\theta_0)$) which follow uniform distribution on $[-1, 1]$.

4. S.B.M. is changed to be fBm with Hurst parameter $H = 0.3, 0.5, 0.7$ to allow more flexible depiction of trajectories' roughness.

5. Simulate event times $T_1^0, \ldots, T_n^0$ which follow exponential distribution with parameters $\exp(\beta Z_1(\theta_0) + \gamma X_1), \ldots, \exp(\beta Z_n(\theta_0) + \gamma X_n)$ respectively. In this simulation we set $(\beta, \gamma) = (1, 0)$ to make our results more comparable to those obtained in Chapter 4.1. Generate the censoring times $C_1, \ldots, C_n$ which follow exponential distribution with parameter $|Z_1(\theta_0)|, \ldots, |Z_n(\theta_0)|$. So the censoring time T_i^0 and event time C_i are conditionally independent given $Z_i(\theta_0)$. The followup time $\tau = 50$. For the ith subject, we observe only $(T_i^0 \wedge C_i \wedge \tau, 1_{T_i^0 \le C_i, T_i \le \tau})$.

6. For each grid point $j = 1, \ldots, J$, we obtain $(\hat{\beta}_n^j, \hat{\gamma}_n^j)$ by maximizing partial likelihood function using $(Z(\theta_j), X)$ as covariates, and corresponding partial likelihood value PL_j.

7. By picking the maximum PL_j out of $j = 1, \ldots, J$, the maximizer index j^* and corresponding $(\hat{\beta}_n^{j^*}, \hat{\gamma}_n^{j^*})$ is obtained. The estimator of (θ, β, γ) is $(j^*/J, \hat{\beta}_n^{j^*}, \hat{\gamma}_n^{j^*})$.

An elaboration on the algorithm of looking for the maximizer of the partial likelihood function $\mathrm{PL}(\theta, \beta, \gamma)$ is as follows.

In the step 6 above, for each fixed grid point j, we have covariates $(Z(\theta_j), X)$ and survival outcomes $(T^0 \wedge C \wedge \tau, 1_{T^0 \le C, T \le \tau})$ observed for every subject. The problem of estimating parameters (β, γ) is achieved by maximizing the partial likelihood function

$\mathrm{PL}(\theta_j;\beta,\gamma)$. This can be easily solved by the "coxph" function in the R package "survival", and the corresponding partial likelihood value PL_j is obtained from the "coxph" function.

After we obtain PL_j for every $j = 1,\ldots,J$, we can follow step 7 and get the maximizer j^* and the corresponding $(\hat{\beta}_n^{j^*},\hat{\gamma}_n^{j^*})$.

By step 6 and step 7, we transformed the problem of maximizing the partial likelihood function $\mathrm{PL}(\theta,\beta,\gamma)$ in the three dimensional space $[0,1]\times(-\infty,+\infty)\times(-\infty,+\infty)$ into a problem easily solved by 2 steps. Since we only observe finite grid points for $\theta\in[0,1]$, the first element θ^* of the three element maximizer $(\theta^*,\beta^*,\gamma^*)$ must be among $\{\theta_1,\ldots,\theta_J\}$. So the maximizer of $\mathrm{PL}(\theta,\beta,\gamma)$ is no larger than the supremum of $\{\mathrm{PL}(\theta_j,\beta,\gamma) : j = 1,\ldots,J,\beta\in(-\infty,+\infty),\gamma\in(-\infty,+\infty)\}$. If we divide the set into J subset (without overlap) $\{\mathrm{PL}(\theta_j,\beta,\gamma) : \beta\in(-\infty,+\infty),\gamma\in(-\infty,+\infty)\}$, (j=1,...,J), the supremum of the original set is the maximum of the J subsets' supremums. The PL_j obtained in step 6 is exactly the jth subset's supremum. So the maximum obtained in step 7 is exactly the maximum value of $\mathrm{PL}(\theta,\beta,\gamma)$.

4.2.2 Confidence intervals of $\hat{\theta}_n$

To evaluate the finite sample performance of $\hat{\theta}_n$, we need a benchmark distribution, which is usually the asymptotic distribution, and see how close the empirical distributions of simulated results are to the benchmark distribution. The closer they are, statistical inferences based on the benchmark distribution perform better for data sets with finite sample sizes.

As the symmetric asymptotic distribution of $n^{1/(2H)}(\hat{\theta}_n-\theta_0)$ does not capture the features of empirical distributions (see Figures 4.2 and 4.3) well, we choose "Domain-Restricted Asymptotic Distribution" as the benchmark distribution. Statistical inferences can be based on the DRAD.

The investigation of the benchmark distribution, i.e., DRAD, is done in Chapter 3.2.

4.2.2.1 Simulation study of $n^{1/(2H)}(\hat{\theta}_n - \theta_0)$

The simulation results for $n^{1/(2H)}(\hat{\theta}_n - \theta_0)$ following the procedure in Chapter 4.2.1 are presented in Table 4.1, Figure 4.2 (for H=0.5), Table 4.2 and Figure 4.3 (for H=0.7), and Figure 4.4 (as comparison of H=0.3, 0.5, 0.7).

From Tables 4.1 and 4.2, as sample size n increases, the empirical tail probabilities of the nominal quantiles based on the simulated DRADs (and hence the empirical coverage probabilities of the nominal confidence intervals) approach their nominal levels.

Overall speaking, the approximation of the asymptotic distribution to the empirical distributions is not ideal, however, it is reasonable for sample sizes as big as 240. Note that some neighboring values in Table 4.1 are the same (for example, empirical tail probabilities for $(\theta_0, \sigma, n) = (0.5, 3, 240)$ are the same for $> q_{.05}$ and $> q_{.025}$) is due to the fact the grid with grid size 1/100 is not fine enough. Due to the computing ability constraint, we can not use a finer grid.

The "boundary-cluster" phenomena for $n^{1/(2H)}(\hat{\theta}_n - \theta_0)$ appears as expected (see the lower row of Figures 4.2, 4.3) however, it weakens gradually with the decrease of H and/or $|\theta_0 - 0.5|$, and/or increase of n.

Comparison of $\hat{\theta}_n$ for different H

Examining Figure 4.4 reveals that convergence rate $n^{1/(2H)}$ becomes slower as H increases. It also implies that the main results in Chapter 2.3 hold not only for $H \in [1/2, 1)$, but also are expected to hold for $(0, 1/2)$.

Remark 4.2.1.

1. *Lots of simulations are conducted for $H = 0.3$ as well, even though they are not reported here. It is hard to get the quantiles (actually, the empirical critical values) with acceptable precision of the benchmark distribution (i.e., the DRAD) in $H = 0.3$ case because to get quantiles with a given precision by simulation,*

Figure 4.2: (H=0.5): Histograms of $n(\hat{\theta}_n - \theta_0)$ for $\hat{\theta}_n$ estimated for Cox model,$(\theta_0, \sigma) = (0.5, 1)$ for the upper row, $(\theta_0, \sigma) = (0.3, 1)$ for the middle row, $(0.1, 1)$ for the lower row

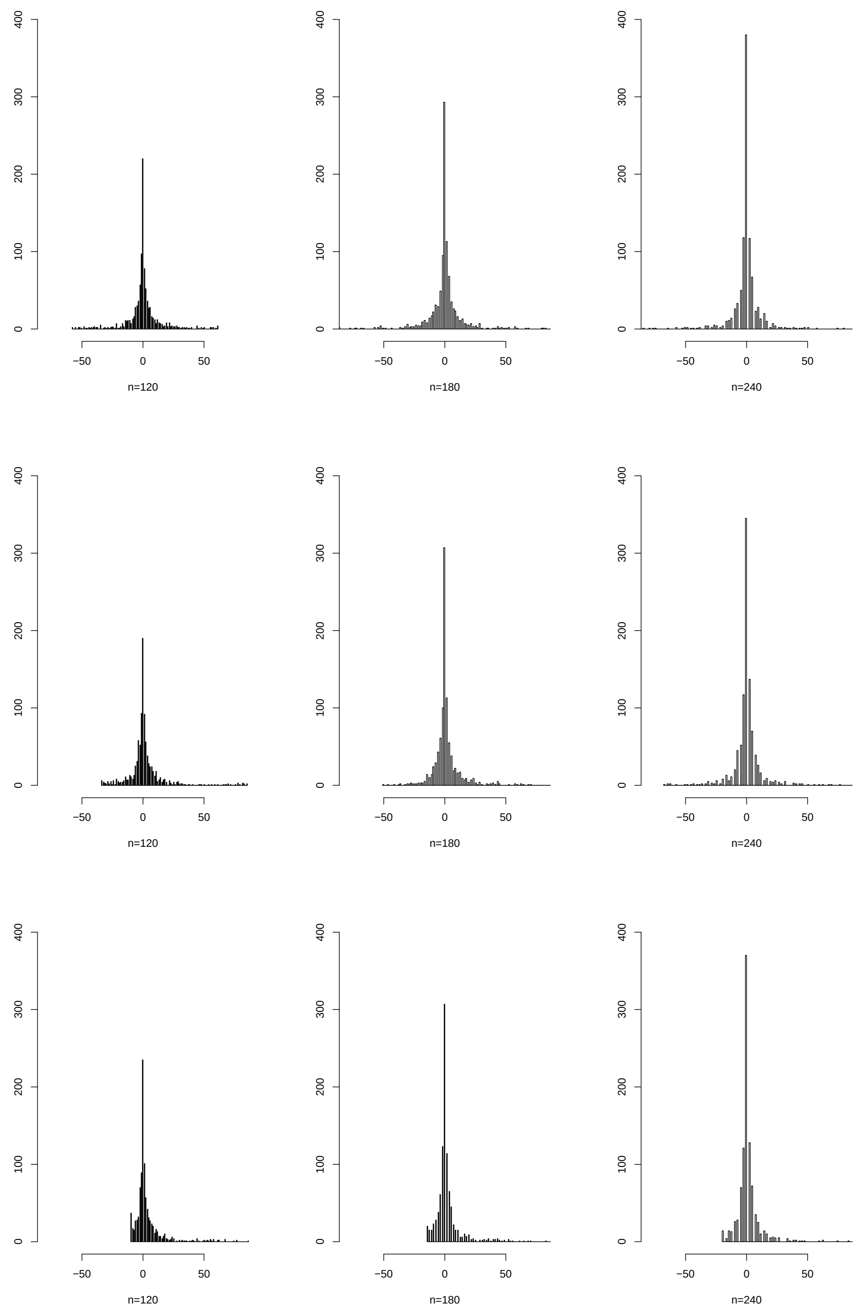


Table 4.1: (H=0.5): Empirical tail probabilities of nominal .025, .05, .95, .975 quantiles for $\hat{\theta}_n$; data simulated from extended Cox model

θ_0	σ	n	$> q_{.95}$	$> q_{.975}$	$< q_{.05}$	$< q_{.025}$
0.5	1	120	.120	.086	.113	.081
		180	.106	.066	.110	.071
		240	.077	.044	.081	.056
	2	120	.127	.101	.122	.106
		180	.095	.068	.093	.061
		240	.059	.059	.062	.062
	3	120	.200	.106	.180	.103
		180	.093	.093	.111	.111
		240	.057	.057	.068	.068
0.3	1	120	.124	.084	.103	.076
		180	.119	.077	.070	.032
		240	.076	.054	.073	.036
	2	120	.120	.098	.098	.083
		180	.089	.055	.096	.063
		240	.063	.063	.054	.054
	3	120	.193	.117	.190	.110
		180	.098	.098	.104	.104
		240	.064	.064	.071	.071
0.1	1	120	.125	.088	.080	.068
		180	.111	.072	.074	.061
		240	.076	.045	.071	.051
	2	120	.122	.097	.094	.071
		180	.091	.075	.089	.048
		240	.062	.034	.058	.045
	3	120	.164	.105	.148	.073
		180	.106	.095	.091	.091
		240	.061	.061	.041	.041

Figure 4.3: (H=0.7): Histograms of $n^{1/(2H)}(\hat{\theta}_n - \theta_0)$ for $\hat{\theta}_n$ estimated for Cox model, $(\theta_0, \sigma) = (0.5, 1)$ for the upper row, $(\theta_0, \sigma) = (0.3, 1)$ for the middle row, $(0.1, 1)$ for the lower row

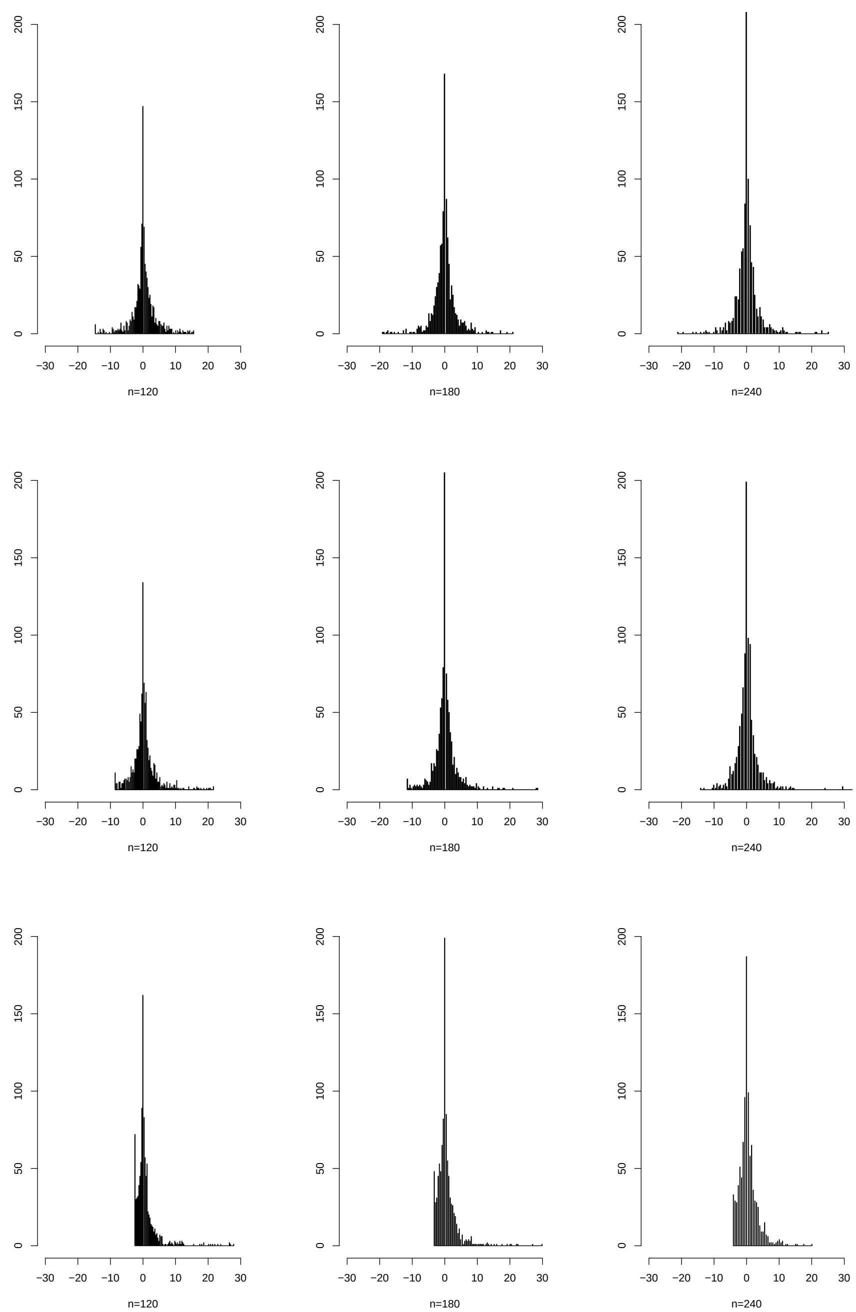


Table 4.2: (H=0.7): Empirical tail probabilities of nominal .025, .05, .95, .975 quantiles for $\hat{\theta}_n$; data simulated from extended Cox model

θ_0	σ	n	$> q_{.95}$	$> q_{.975}$	$< q_{.05}$	$< q_{.025}$
0.5	1	120	.109	.072	.088	.057
		180	.085	.062	.083	.049
		240	.087	.050	.069	.051
	2	120	.081	.066	.086	.071
		180	.080	.061	.088	.033
		240	.091	.056	.102	.062
	3	120	.114	.078	.125	.093
		180	.065	.065	.064	.064
		240	.080	.034	.101	.046
0.3	1	120	.100	.064	.077	.051
		180	.085	.058	.076	.051
		240	.089	.056	.070	.033
	2	120	.094	.078	.079	.063
		180	.086	.044	.100	.048
		240	.102	.046	.096	.041
	3	120	.111	.088	.112	.068
		180	.064	.064	.062	.062
		240	.080	.039	.057	.023
0.1	1	120	.106	.080	.073	.049
		180	.078	.055	.072	.045
		240	.079	.049	.056	.037
	2	120	.076	.060	.089	.053
		180	.085	.067	.081	.057
		240	.072	.057	.072	.053
	3	120	.093	.063	.093	.063
		180	.061	.061	.069	.069
		240	.077	.038	.072	.041

it requires finer grid for small H (number of grid points $\sim n^{1/(2H)}$). Hence it is hard to report any quantitative results for simulations conducted for $H = 0.3$. However, their shapes (not shown in this thesis in general, except for $\theta_0 = 0.5$ case in Figure 4.4) has exactly the same trend and features as those of $H = 0.5$ and $H = 0.7$, which implies all the large sample properties presented in Theorem 2.3.2 are expected to hold for $H \in (0, 1/2)$.

2. *Considering the asymmetry of left and right confidence limits for θ, we adopt the so-called Domain-Restricted Asymptotic Distribution as the benchmark distribution, which has better performance than the asymptotic distribution. Professor Tsai suggests another idea. That is to make a tranformation $f(\theta)$ for θ ($\theta \in (0, 1)$), such that the asymptotic distribution of transformed estimated parameter $f(\hat{\theta}_n)$ has some symmetric distribution. Then we can construct confidence interval based on this symmetric distribution, and then transform back to obtain the confidence interval. This idea is very promising considering it is convenient and do not involve extra Monte Carlo calibration for the quantiles of the DRADs. A natural candidate for the transformation is the logit function, $f(\theta) = \log(\theta/(1-\theta))$. We test its use with "delta method" in both $H = 0.5$ and $H = 0.7$ cases. For $H = 0.5$ case, it works as accurately as the DRAD method and is more desirable. However, in the $H = 0.7$ case, it performs worse compared to the DRAD method. Further studies are needed to evaluate the potential of the transformation method.*

3. *Examining the simulation results in Tables 4.1 and 4.2, the sample sizes 120, 180, 240 do not warrant accurate approximation of the coverage probability to the nominal levels. It is desirable to show the coverage probability converges to the nominal level when the sample size increases further. Due to the computing facility constraint, we simulate only one scenario: $\theta_0 = 0.5$ and $\sigma = 1$. The sample size $n = 720$ and the number of grid points $J = 1000$ on the interval*

$[0,1]$ *(if the grid size keeps fixed at 100 and merely increase sample size, we will get cruder rescaled empirical distribution of* $n^{1/(2H)}(\hat{\theta}_n - \theta_0)$*, which does not help much improving the approximation accuracy). The empirical tail probabilities of nominal .95,.975,.05,.025 quantiles are .061,.036,.068,.037 respectively. These results are much closer to the nominal levels (.05,.025,.05,.025) compared to the* $n = 240$ *and number of grid points* $J = 100$ *case, where these values are .077, .044,.081,.056. To achieve better approximation, we expect to have larger sample size which may be conducted on faster computers in the future.*

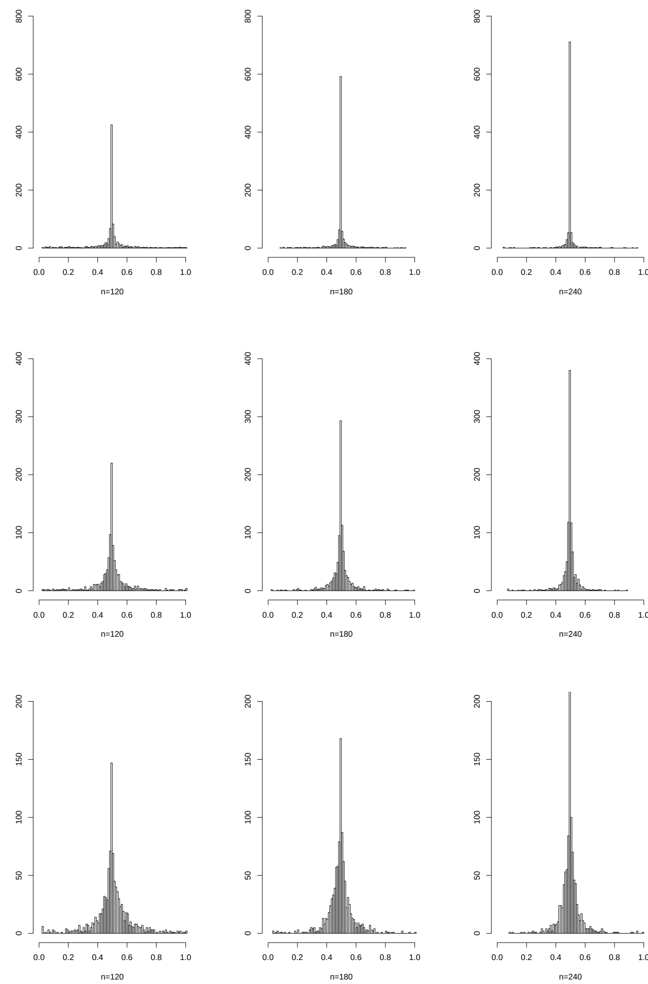
4. *The evaluation of the empirical tail probabilities in Tables 4.1 and 4.2 are based on the calibrated quantiles of the DRAD with restricted domain* $[-n\theta_0, n(1-\theta_0)]$*. In practice, the true value* θ_0 *is unknown and* $\hat{\theta}_n$ *is used instead. It is computationally too expensive to calibrate the quantiles of the DRAD with each value of* $\hat{\theta}_n$*. To roughly check the impact of this simplified handling, we choose the scenario* $(\theta_0, \sigma, n) = (0.5, 1, 120)$*, use among* $\{0.1, 0.3, 0.5, 0.7, 0.9\}$ *the closest value to* $\hat{\theta}_n$ *to get the calibrated quantiles, and then calculate the empirical tail probabilities. The results show it improved the approximation accuracy comparing to the table. It implies that the simplified handling (i.e., replacing* $[-n\hat{\theta}_n, n(1-\hat{\theta}_n)]$ *by* $[-n\theta_0, n(1-\theta_0)]$*) gives a conservative evaluation of the proposed method.*

4.2.3 Confidence intervals of $\hat{\beta}_n$ and $\hat{\gamma}_n$

To evaluate the finite sample performance of $\hat{\beta}_n$ and $\hat{\gamma}_n$, we can just choose their asymptotic distributions as the benchmark distributions, and see how the empirical distributions of the simulated results approach their asymptotic distributions. The closer they are, statistical inferences based on their asymptotic distributions performs better for data sets with finite sample sizes.

We begin with the investigation of the asymptotic distributions of $\hat{\beta}_n$ and $\hat{\gamma}_n$.

Figure 4.4: (H=0.3,0.5,0.7): Histograms of non-rescaled $\hat{\theta}_n$ for $\hat{\theta}_n$ estimated for Cox model, $(\theta_0, \sigma) = (0.5, 1)$, H=(0.3,0.5,0.7) for the upper, middle, lower row respectively



4.2.3.1 Wald-type confidence intervals of $\sqrt{n}(\hat{\beta}_n - \beta_0)$ and $\sqrt{n}(\hat{\gamma}_n - \gamma_0)$ based on the asymptotic distributions

In this subsection, we will calculate the asymptotic distributions of $\sqrt{n}(\hat{\beta}_n - \beta_0)$ and $\sqrt{n}(\hat{\gamma}_n - \gamma_0)$ based on the parameters set by the simulation procedure in Chapter 4.2.1.

By Theorem 2.3.2, the asymptotic distribution of $\left(\sqrt{n}(\hat{\beta}_n - \beta_0), \sqrt{n}(\hat{\gamma}_n - \gamma_0)\right)$ is a 2-dimensional normal distribution. To determine this distribution, we only need the values of s_1^2, s_2^2 and ρ.

By steps 3 and 5 of the simulation procedure in Chapter 4.2.1, $X \sim U[-1,1]$, $Z(\theta_0) \sim U[-0.75, 0.75]$, $T^0 \sim \text{Exp}\left(\exp(\beta Z(\theta_0) + \gamma X)\right)$ with $(\beta, \gamma) = (1, 0)$. $C \sim \text{Exp}\left(|Z(\theta_0)|\right)$, $\tau = 50$ which are the same as of the simulation procedure of the simple model in Chapter 4.1.1.

Then

$$
\begin{aligned}
P[X^2 N(\tau)] &= P[X^2 1_{T^0 \le \tau, T^0 \le C}] = P\left[X^2 P\left(T^0 \le \tau, T^0 \le C | Z(\theta_0), X\right)\right] \\
&= P\left[X^2 \int_0^\tau e^{Z(\theta_0)} e^{-e^{Z(\theta_0)} u} \cdot e^{-u|Z(\theta_0)|} du\right] \\
&= P\left[X^2 \cdot e^{Z(\theta_0)} \frac{1 - e^{-\tau(e^{Z(\theta_0)} + |Z(\theta_0)|)}}{e^{Z(\theta_0)} + |Z(\theta_0)|}\right] \\
&= P[X^2] \cdot P\left[e^{Z(\theta_0)} \frac{1 - e^{-\tau(e^{Z(\theta_0)} + |Z(\theta_0)|)}}{e^{Z(\theta_0)} + |Z(\theta_0)|}\right] \\
&= \frac{1}{3} P\left[e^{Z(\theta_0)} \frac{1 - e^{-50(e^{Z(\theta_0)} + |Z(\theta_0)|)}}{e^{Z(\theta_0)} + |Z(\theta_0)|}\right],
\end{aligned}
$$

where we used the properties of the distribution of T^0, C in the third equality and the property of the distribution of X in the last equality.

Here $Z(\theta_0) \sim U[-0.75, 0.75]$, there is no analytical solution to $P[X^2 N(\tau)]$. Using 10,000,000 Monte-Carlo simulations we obtain the estimate of $P[X^2 N(u)] = 0.246833033$.

$$
\begin{aligned}
P[Z^2(\theta_0)N(\tau)] &= P[Z^2(\theta_0)1_{T^0\leq\tau,T^0\leq C}] = P\left[Z^2(\theta_0)P\left(T^0\leq\tau,T^0\leq C|Z(\theta_0),X\right)\right]\\
&= P\left[Z^2(\theta_0)\int_0^\tau e^{Z(\theta_0)}e^{-e^{Z(\theta_0)}u}\cdot e^{-u|Z(\theta_0)|}du\right]\\
&= P\left[Z^2(\theta_0)\cdot e^{Z(\theta_0)}\frac{1-e^{-50(e^{Z(\theta_0)}+|Z(\theta_0)|)}}{e^{Z(\theta_0)}+|Z(\theta_0)|}\right]
\end{aligned}
$$

There is no analytical solution to $P[Z^2(\theta_0)N(\tau)]$ as well. Using 10,000,000 Monte-Carlo simulations we obtain the estimate of $P[Z^2(\theta_0)N(u)] = 0.1195684$.

$$
\begin{aligned}
P[Z(\theta_0)XN(\tau)] &= P[Z(\theta_0)X1_{T^0\leq\tau,T^0\leq C}] = P\left[Z^2(\theta_0)P\left(T^0\leq\tau,T^0\leq C|Z(\theta_0),X\right)\right]\\
&= P\left[Z(\theta_0)X\int_0^\tau e^{Z(\theta_0)}e^{-e^{Z(\theta_0)}u}\cdot e^{-u|Z(\theta_0)|}du\right]\\
&= P\left[XZ(\theta_0)\cdot e^{Z(\theta_0)}\frac{1-e^{-50(e^{Z(\theta_0)}+|Z(\theta_0)|)}}{e^{Z(\theta_0)}+|Z(\theta_0)|}\right]\\
&= P[X]P\left[Z(\theta_0)\cdot e^{Z(\theta_0)}\frac{1-e^{-50(e^{Z(\theta_0)}+|Z(\theta_0)|)}}{e^{Z(\theta_0)}+|Z(\theta_0)|}\right] = 0,
\end{aligned}
$$

where the last equality comes from $PX = 0$ (since $X \sim U[-1, 1]$) and the second-to-last equality comes from the independence of $Z(\theta_0)$ and X. Notice that $P[Z(\theta_0)XN(\tau)]$ is equal to 0 here because we set $\gamma = 0$ for simulations conducted in Chapter 4.2.

Putting the values of $P[Z^2(\theta_0)N(\tau)], P[X^2N(\tau)]$ and $P[Z(\theta_0)XN(\tau)]$ back to the formulas for s_1^2, s_2^2 and ρ ,

$$
s_1^2 = \frac{1}{P[Z^2(\theta_0)N(\tau)]} = 8.36341, \quad s_2^2 = \frac{1}{P[X^2N(\tau)]} = 4.05132, \quad \rho = 0.
$$

The 0.95 and 0.975 quantiles of the standard normal distribution are 1.645 and 1.96 respectively,

$$
1.96\cdot\sqrt{8.36341} = 5.67, \quad 1.645\cdot\sqrt{8.36341} = 4.76;
$$

$$
1.96\cdot\sqrt{4.05132} = 3.94, \quad 1.645\cdot\sqrt{4.05132} = 3.31.
$$

Then the 0.95 and 0.975 quantiles of the asymptotic distribution of $\sqrt{n}(\hat{\beta}_n - \beta_0)$ are 4.76 and 5.67; the 0.95 and 0.975 quantiles of the asymptotic distribution of $\sqrt{n}(\hat{\gamma}_n - \gamma_0)$ are 3.31 and 3.94.

4.2.3.2 Simulation study of $\sqrt{n}(\hat{\beta}_n - \beta_0)$ and $\sqrt{n}(\hat{\gamma}_n - \gamma_0)$

For $(\theta_0, \sigma, \beta, \gamma) = (0.5, 1, 1, 0)$ and $H = 0.3, 0.5, 0.7$, the histograms of $\sqrt{n}(\hat{\beta}_n - \beta_0)$ and $\sqrt{n}(\hat{\gamma}_n - \gamma_0)$ (with data simulated from the extended Cox model with these specified parameters) are shown in Figure 4.5 (for $\sqrt{n}(\hat{\beta}_n - \beta_0)$) and Figure 4.6 (for $\sqrt{n}(\hat{\gamma}_n - \gamma_0)$).

We did not present their histograms in the case of $(\theta_0, \sigma) = (0.3, 1)$ or $(\theta_0, \sigma) = (0.1, 1)$ since they did not change much compared to those in the case $(\theta_0, \sigma) = (0.5, 1)$.

Tables 4.3 and 4.4 show the empirical coverage probabilities of the nominal 90% and 95% confidence intervals of β and γ respectively.

By Table 4.3, there is obvious under coverage for the confidence intervals of β, especially for small H ($H = 0.3$ or 0.5). With the increase of H, there are significant gains of the coverage probabilities. With the increase of sample sizes, there are gradual gains of the coverage probabilities as well. Overall speaking, the confidence intervals based on the asymptotic normal distribution of β do not perform satisfactorily. It requires quite large sample sizes (n much larger than 240) to make the coverage probabilities of the confidence intervals approach their nominal levels.

In contrast, the confidence intervals of γ (see Table 4.4) show quite accurate coverage probabilities compared to their nominal levels in all the cases listed in the table.

The difference in the finite sample performance of β and γ could be due to the fact that β is the coefficient of $Z(\theta)$ where θ is unknown, while β is the coefficient of X which is observed. Such a difference could lead to a better higher order accuracy of $\hat{\gamma}_n$ compared to $\hat{\beta}_n$ even though they have the same $\sqrt{n}$-order accuracy.

Remark 4.2.2. *In this section, we used the nominal confidence intervals calculated from the true parameters' values. We could also consider to use the nominal confidence intervals constructed from estimated parameters' values. That looks at the problem from a slightly different perspective. However, we expect to see similar trends regarding the approximation accuracy.*

Table 4.3: Coverage probabilities of nominal 90% and 95% confidence intervals for β; data simulated from the extended Cox model, $\theta_0 = 0.5$

σ	n	$H = 0.3$		$H = 0.5$		$H = 0.7$	
		90%	95%	90%	95%	90%	95%
1	120	.786	.870	.816	.902	.845	.912
	180	.789	.854	.824	.894	.872	.918
	240	.791	.858	.846	.910	.871	.921
2	120	.682	.758	.824	.893	.853	.904
	180	.773	.837	.822	.884	.874	.925
	240	.840	.896	.848	.903	.867	.923
3	120	.720	.767	.777	.844	.839	.891
	180	.847	.893	.818	.871	.847	.911
	240	.876	.926	.834	.889	.847	.908

Figure 4.5: Histograms of $\sqrt{n}(\hat{\beta}_n - \beta_0)$ for $\hat{\beta}_n$; data simulated from the extended Cox model, $(\theta_0, \sigma) = (0.5, 1)$, H=(0.3,0.5,0.7) for the upper, middle, lower row respectively

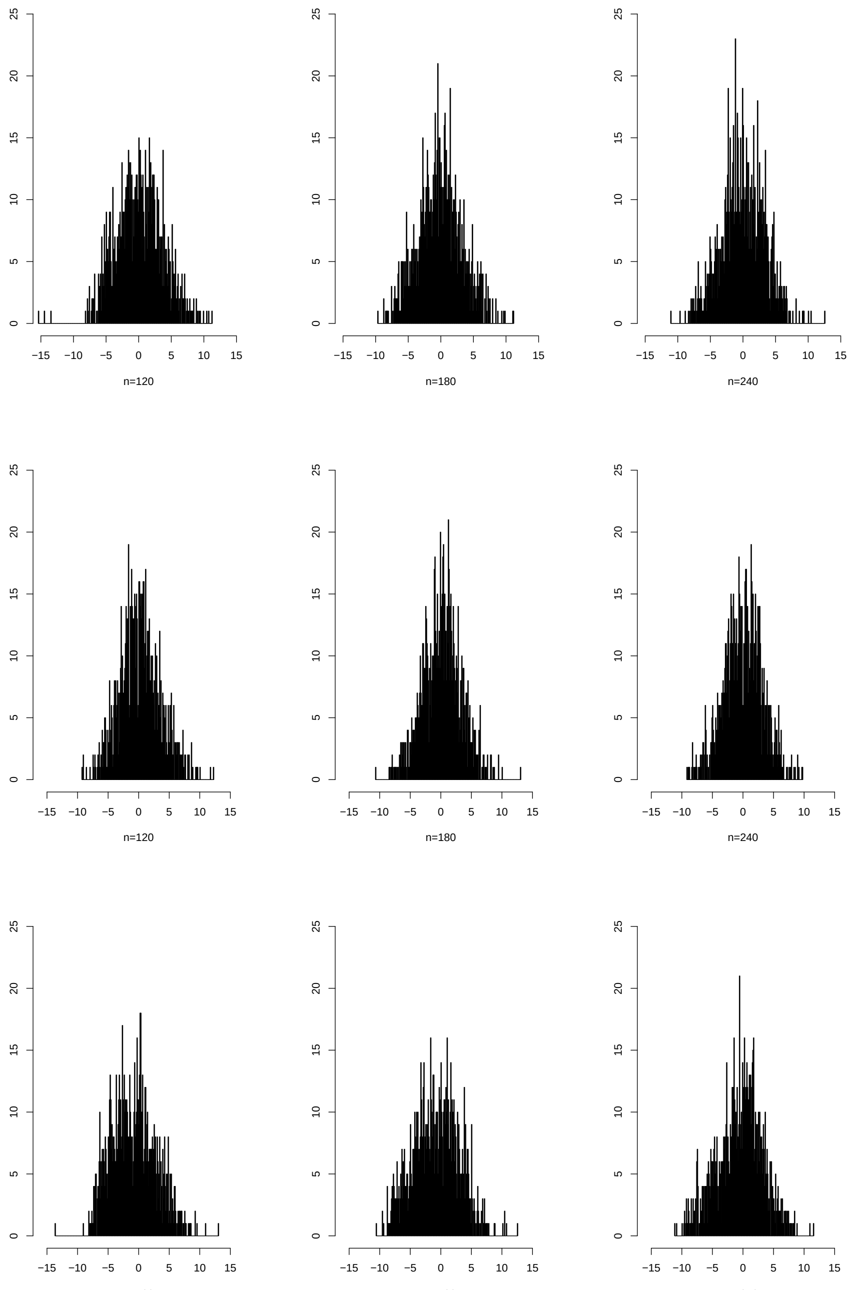


Figure 4.6: Histograms of $\sqrt{n}(\hat{\gamma}_n - \gamma_0)$ for $\hat{\gamma}_n$; data simulated from the extended Cox model, $(\theta_0, \sigma) = (0.5, 1)$, H=(0.3,0.5,0.7) for the upper, middle, lower row respectively

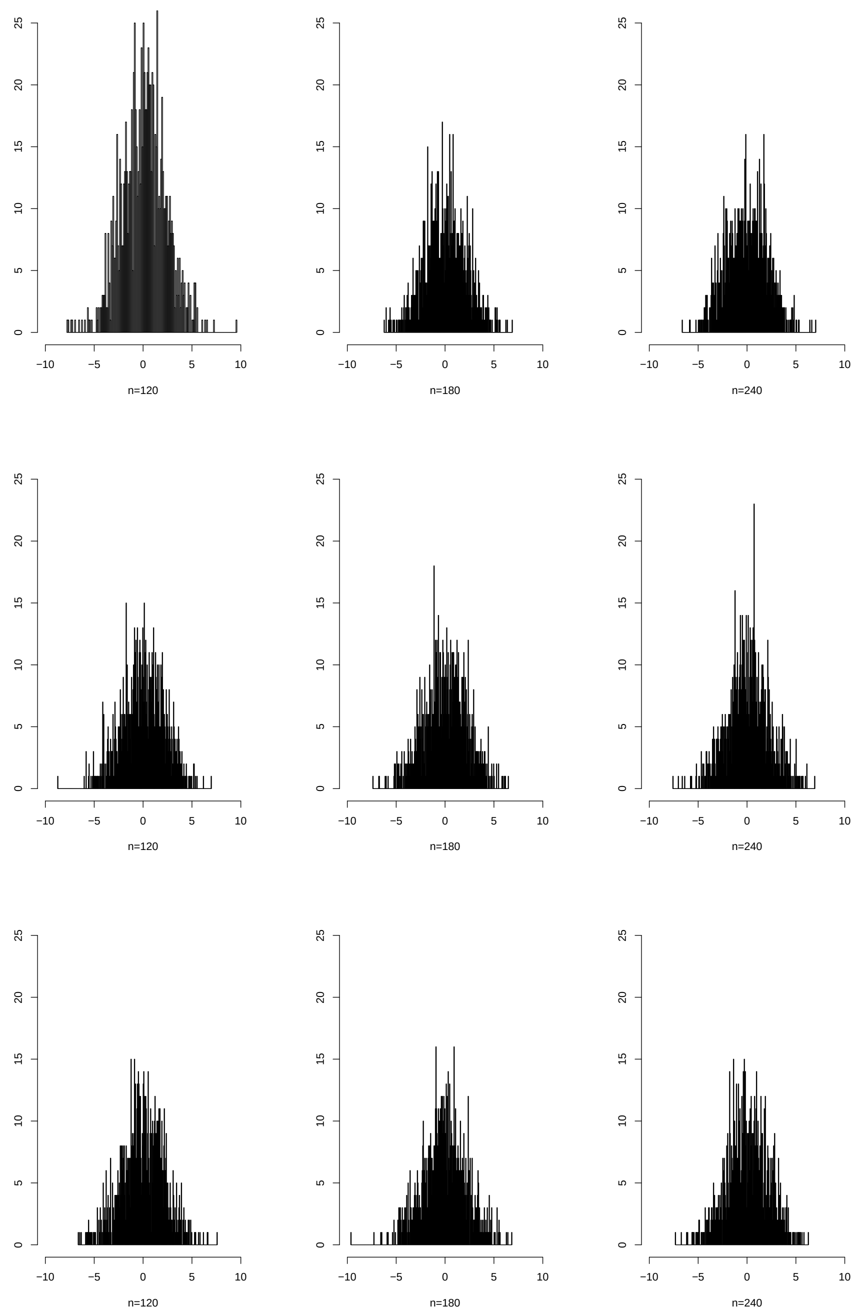


Table 4.4: Coverage probabilities of nominal 90% and 95% confidence intervals for γ; data simulated from the extended Cox model, $\theta_0 = 0.5$

σ	n	$H = 0.3$		$H = 0.5$		$H = 0.7$	
		90%	95%	90%	95%	90%	95%
1	120	.870	.929	.876	.928	.872	.929
	180	.857	.913	.875	.933	.866	.923
	240	.871	.932	.892	.945	.859	.921
2	120	.882	.927	.866	.924	.868	.923
	180	.876	.934	.874	.932	.866	.929
	240	.877	.929	.899	.942	.901	.948
3	120	.875	.930	.862	.918	.866	.924
	180	.884	.933	.887	.941	.866	.923
	240	.888	.943	.877	.934	.880	.942

Chapter 5

Analysis of genomic data

In this chapter, we will apply the methods developed in this thesis to two real data sets, one from an adult brain cancer study and another from a breast cancer study. For the adult brain cancer study, complete data is observed; for the breast cancer study, however, partial data is missing. They are studied in the first and second sections respectively. In the last section, we will summarize the basic algorithm of applying the proposed method for real data analysis.

In the first section, relying on the MPLE method and the DRAD developed in earlier chapters, we get the point estimates and confidence intervals of the sensitive locus on Chromosome 1 and its impact strength on hazard of dying of adult brain cancer. Simulations are conducted to evaluate the empirical coverage probability of the Wald-type confidence intervals based on the DRAD.

In the second section, to handle missing values in the functional covariate, inverse probability weighting method is used. We get the point estimates and confidence intervals for the sensitive locus on Chromosome X, the impact strength of this locus, and the impact strength of a non-functional covariate: tumor category.

Furthermore, the inverse probability weighting method is studied using simulated data.

5.1 Adult brain cancer study

Glioblastoma, the most common brain cancer in adults, is the first cancer studied by The Cancer Genome Atlas (TCGA). There are 174 subjects in this study, with 156 subjects experiencing death, 17 subjects survival times censored and 1 subject survival information missing. Their complete gene expression profiles are observed on 1599 loci of Chromosomes 1-22. Some of the chromosomes were found to be closely related to the risk of glioblastoma occurrence ([16], [20], [14] and [26]). As an example, we will use our proposed model to select the most sensitive locus to the risk of dying of glioblastoma on Chromosome 1.

Chromosome 1 has $J = 181$ locus in this data set. We can use the survival and gene expression profile data from the $n = 173$ subjects (we exclude the subject whose survival information is missing) to estimate the most sensitive locus. By implementing the MPL estimating procedure for our model, we get the estimate of (θ, β), with $\hat{\theta}_n = 40/181$ and $\hat{\beta}_n = 0.287$. The name of the 40th locus on Chromosome 1 is "DIRAS3".

5.1.1 Wald-type confidence intervals

In order to obtain the Wald-type confidence intervals for θ and β, we have to estimate the nuisance parameters in our model, H and σ^2. To estimate the Hurst parameter H, we adopt the function "pengFit()" in the R-package "fArma". This function estimates the Hurst parameter by Peng's variance of residuals method [33]. It divides the time series into blocks of size m. The cumulated sums within each block are computed up to time t, then least-square method is used to fit the cumulated sums by $a+bt$. The sample variance of these residuals is proportional to m^{2H}. The "mean" or "median" are computed over these blocks. The slope $2H$ from the least square provides an estimate for the Hurst parameter H.

By Peng's method, for each subject's gene expression profile, an estimate of H is obtained. For the 173 estimates, the mean is 0.535 and 80% percent of them falls within $[0.41, 0.66]$. So we can use 0.535 as the H estimate.

To estimate σ^2, we can adopt the quadratic variation method. For each subject, we can get an estimate of σ^2. Take an average of these estimates, we get an estimate

$$\hat{\sigma^2} = \frac{1}{173}\sum_{i=1}^{173}\sum_{j=1}^{180}\left(Z_i\left(\frac{j+1}{181}\right) - Z_i\left(\frac{j}{181}\right)\right)^2 \approx 236.14, \quad \text{hence} \quad \hat{\sigma} \approx 15.37.$$

To construct the Wald-type confidence intervals for θ and β, by Theorem 2.3.2, it remains to get the quantiles of

$$\text{argmax}_{h\in[-n^{1/(2H)}\theta_0, n^{1/(2H)}(1-\theta_0)]}\left(W^H(h) - \frac{|h|^{2H}}{2}|\beta|\sigma\sqrt{PN(\tau)}\right)$$

and the value of $1/P(Z^2(\theta_0)N(\tau))$.

Since $\theta_0, \beta, H, PN(\tau), P(Z^2(\theta_0)N(\tau))$ are unknown, we replace them by their estimates. The MPLEs of θ_0 and β are $\hat{\theta}_n = 40/181, \hat{\beta}_n = 0.287$. The estimate of H is 0.535. The empirical estimates of $PN(\tau), P(Z^2(\theta_0)N(\tau))$ are

$$\mathbb{P}_n N(\tau) = 156/173, \quad \mathbb{P}_n(Z^2(\theta_0)N(\tau)) = 0.0598.$$

To get the quantiles of

$$\text{argmax}_{h\in[-n^{1/(2H)}\theta_0, n^{1/(2H)}(1-\theta_0)]}\left(W^H(h) - \frac{|h|^{2H}}{2}|\beta|\sigma\sqrt{PN(\tau)}\right),$$

we resort to the Monte Carlo simulation. Through 10,000 sample paths generation and maximum index search for each path, we get the desired 0.975 and 0.025 quantiles: 0.74 and -0.74. Then we can easily get the 95% confidence interval of θ as

$$\left[40/181 - 0.74/173^{1/(2\cdot 0.535)}, 40/181 + 0.74/173^{1/(2\cdot 0.535)}\right] = [0.215, 0.227].$$

It means the 95% confidence interval of the locus is the 39th, 40th and 41st loci (with slight undercoverage since the grids are not fine enough).

By Theorem 2.3.2, $\sqrt{n}(\hat{\beta}_n - \beta_0)$ follows a normal distribution with variance given by $1/P(Z^2(\theta_0)N(\tau))$. Then the 95% Wald-type confidence interval for β is

$$\left[\hat{\beta}_n - 1.96/\sqrt{173\cdot 0.0598}, \hat{\beta}_n + 1.96/\sqrt{173\cdot 0.0598}\right],$$

which turns out to be $[-0.322, 0.896]$.

5.1.2 Assess the accuracy of the theoretical Wald-type confidence interval by simulation

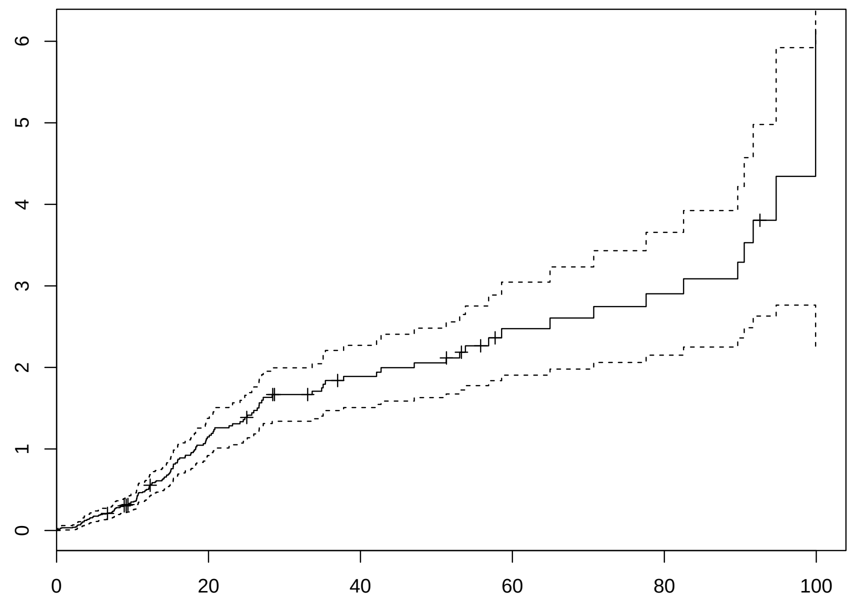
Since the Wald-type confidence intervals are based on theoretical results in Chapter 2 which depends on assumptions, part of which could be unrealistic. So we can't take it for granted that the actual coverage would be close to its nominal level. To evaluate the 95% confidence interval for θ assuming that the true θ corresponds to the 40th locus, one possible method is to simulate survival outcomes of subjects based on their gene expression levels at 40th locus and the corresponding hazard ratio is set as 0.287. In other words, we are pretending that the fitted Cox model is the actual data generating process. The purpose is to assess the actual coverage probability of the constructed confidence interval without thinking of the possibility of model misspecification. We will elaborate on the simulation method in the following paragraphs.

For the ith subject, we simulate its time-to-event outcome with hazard function $\hat{\lambda}_0(t)\exp(\hat{\beta}_n Z_i(\hat{\theta}_n))$. Since $\hat{\beta}_n = 0.287$, $Z_i(\hat{\theta}_n)$ is directly observed from the original gene expression level data set, we only need to set $\hat{\lambda}_0(t)$.

When estimating the Cox model with covariate as the 40th locus gene expression level, besides $\hat{\beta}_n$, the estimate of the cumulative baseline hazard function is also obtained, which is displayed in Figure 5.1.

The figure shows the estimated cumulative baseline hazard function $\hat{\Lambda}_0(t) = \int_0^t \hat{\lambda}_0(u)du$ and its pointwise 95% confidence interval. Since the baseline hazard function is the derivative of the cumulative baseline hazard function, we have to make some assumption about their smoothness. We can assume the baseline hazard function is piecewise constant and hence the cumulative baseline hazard function is a piecewise linear function. From Figure 5.1, it seems reasonable to smooth the cumulative baseline hazard function to be 3-phase piecewise linear. The baseline hazard function $\hat{\lambda}_0(t)$ is piecewise linear with 3 different values h_1, h_2 and h_3 for $t \in [0, 28.5], t \in [28.5, 89.6]$, and $t \in [89.6, 100]$ respectively. The estimated $h_1 = 0.059, h_2 = 0.023, h_3 = 0.185$.

Figure 5.1: The estimated cumulative baseline hazard function and its pointwise 95% confidence interval of the brain cancer study



Now we have the ith subject's hazard function $\hat{\lambda}_0(t)\exp(\hat{\beta}_n Z_i(\hat{\theta}_n))$. Its survival function $S_i(t) = \exp(-\int_0^t \hat{\lambda}_0(u)\exp(\hat{\beta}_n Z_i(\hat{\theta}_n))du)$ and the C.D.F. for the time-to-event variable is $F_i(t) = 1 - S_i(t)$. By the monotonicity of F_i, it is easy to solve for its inverse function F_i^{-1}.

By a well known result in simulation, if we generate a random variable U with uniform distribution on $[0,1]$, then $F_i^{-1}(U)$ follows the distribution with C.D.F. $F_i(\cdot)$. So now it is easy to simulate the time-to-event outcome for the ith subject with hazard function $\hat{\lambda}_0(t)\exp(\hat{\beta}_n Z_i(\hat{\theta}_n))$.

We can make the censoring proportion to be comparable to the actual one by tuning the parameter of the censoring variable's distribution (for simplicity, we use the exponential distribution for the simulation of the censoring variable).

We made Rep=1000 simulations and get the estimates of (θ, β) for each replicate. Their histograms are shown in Figure 5.2.

In this simulation, there are 630 out of 1000 estimated θ taking value 39,40 or 41, much less than the nominal coverage of around 95%. This is because in deriving the asymptotic distribution, we used the fBm assumption, which is not realistic. The simulation study can help us get a more realistic evaluation about the confidence interval of θ.

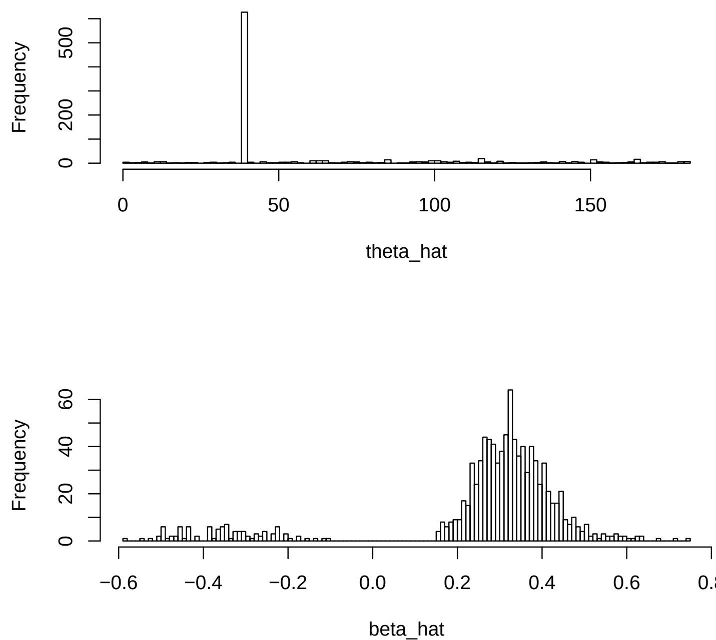
As for β, its distribution is approximately normal except for some negative values. It implies that not all loci on Chromosome 1 are positively correlated to subjects' hazard of dying of brain cancer. When the estimation of θ fall on those loci which are negatively correlated to the hazard, the estimation of β would be negative. Founding these loci might be of interest for physicians as well. Even though not an excellent fit to normal distribution, the nominal 95% confidence interval $[-0.322, 0.896]$, derived in last section based on the asymptotic distribution, has reasonable coverage probability 93.9%.

The strong contrast of the actual coverage probabilities of θ and β's nominal 95% confidence intervals implies that the asymptotic distribution of θ is highly dependent on the unrealistic fBm assumption, while the the asymptotic distribution of β may not. To keep us from abusing the asymptotic distribution of $\hat{\theta}_n$, which is highly dependent on the fBm assumption, simulation is a way to evaluate the actual coverage of θ's nominal confidence intervals.

5.2 Breast cancer study

In this section, we will study a breast cancer data set by methods developed in this thesis. We aim to find the most sensitive locus to breast cancer death disk on Chromosome X. After briefly introducing the data set, we will focus on dealing with missing values in the gene expression profiles. Besides getting the estimates and

Figure 5.2: Histograms of $\hat{\theta}_n$ and $\hat{\beta}_n$ from simulation



confidence intervals for parameters in the proposed Cox model, the impact of different proportions of missingness in the data is also explored by simulation.

5.2.1 Description of the data set

The breast cancer study has 84 subjects. To evaluate the influence of the genes on survival outcome, we need both clinical data (i.e. survival length, status, tumor category) and complete gene expression data on the chromosome we are interested in.

The data set can be found in the supporting information on the website of P-NAS, http://www.pnas.org/cgi/doi/10.1073/pnas.162471999. The clinical data and gene expression data are listed in different tables. To match them by subject ID and leave out those subjects without survival outcome data, we have 36 subjects left. In the clinical data table, the variables collected include age, patient survival, survival months, event for relapse free survival, relapse free survival months, p53 status, clinical ER, tumor category, node status, met, grade, histology and special comments. In the gene expression table, the gene expression levels of different loci on each chromosome (chromosome 1-22 and X) are collected for each subject, even though some values are missing.

5.2.2 Data analysis

For the clinical data table, we first pick up one variable as the non-functional covariate X in the extended Cox model, tumor category. It takes values 1, 2, 3, 4 and is expected to influence the risk of dying of breast cancer. The endpoint of interest is survival months and the event status. Event status comes from the column "patient survival" which takes 4 values, 0=no evidence of disease, 1=alive with disease, 2=dead of disease, 3=dead of other causes. Event status is set to be 1 for "patient survival" taking values 2 and to be 0 otherwise.

In the gene expression table, we choose chromosome X as our interest since a recent paper suggests its relationship to breast cancer ([39]). For the gene expression levels along chromosome X, missing values are observed on some loci.

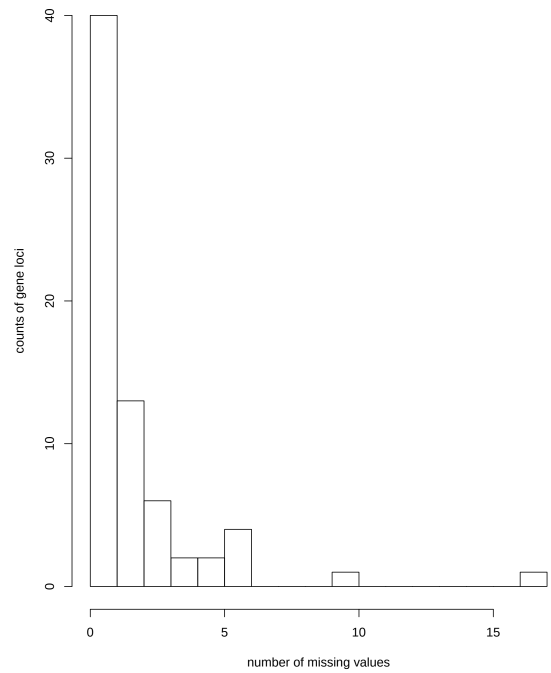
5.2.2.1 Missing values handling

Missing values occurs on 69 out of all 206 loci on chromosome X for the gene expression data. The count of missing values among these 69 loci varies from 1 to 17. 40 loci have 1 subject's expression level missing, 13 loci have 2 missings, the distribution of count of missingness is shown in Table 5.1 and Figure 5.3.

Table 5.1: Distribution of count of missingness among the 69 loci

count of missingness	1	2	3	4	5	6	10	17
counts of loci	40	13	6	2	2	4	1	1

Figure 5.3: Distribution of count of missingness among the 69 loci



We can see that the missingness are not severe except for two loci, the 148th locus (with 17 subjects' gene expression level values missing) and the 155th locus (with 10 subjects' missing). Note that we have altogether 36 subjects.

Assume the missings are completely at random (MCAR, see [41]), i.e. the probability of missing is equal among different loci for each subject, among different subjects, not depending on the survival outcome, other covariate of the subject, and the (unobserved) gene expression level.

Suppose we have J different loci $\{\theta_1, \ldots, \theta_J\}$ and n subjects. For each locus θ_j, if there is no missing values at all, the log-partial-likelihood function would be

$$\log \mathrm{PL}(\beta, \gamma, \theta_j) = \sum_{i=1}^{n} \log \mathrm{PL}_i(\beta, \gamma, \theta_j).$$

For each θ_j, we treat $Z(\theta_j)$ as a non-functional covariate, and fit a Cox model with covariates $Z(\theta_j)$ and X, get the MPLE (Maximum Partial Likelihood Estimator) (β_j^*, γ_j^*) of the regression coefficients vector (β, γ). For this fitted Cox model, we also obtain the log-partial-likelihood value

$$\log \mathrm{PL}(\beta_j^*, \gamma_j^*, \theta_j) = \sum_{i=1}^{n} \log \mathrm{PL}_i(\beta_j^*, \gamma_j^*, \theta_j).$$

According to the maximum likelihood principle, we choose the θ_{j^*} which maximizes $\log \mathrm{PL}(\beta_j^*, \gamma_j^*, \theta_j)$ out of $j = 1, 2, \ldots, J$. The large sample properties of this estimator have been studied in Chapter 3 and its finite sample performance has been evaluated by simulations in Chapter 4.

However, now the log-partial-likelihood value is not available; we only have

$$\log \mathrm{PL}_M(\beta, \gamma, \theta_j) = \sum_{i=1}^{n} \delta_{ij} \log \mathrm{PL}_i(\beta, \gamma, \theta_j).$$

Here $\delta_{ij} = 1$ if gene expression level $Z_i(\theta_j)$ of the locus θ_j is observed for subject i, and 0 if $Z_i(\theta_j)$ is missing.

Even though we do not have $\log \mathrm{PL}(\beta, \gamma, \theta_j)$, if the missing is completely random and not severe (as that of our data, except for 148th and 155th loci), we can expect a reasonably good and unbiased estimate of $\log \mathrm{PL}(\beta, \gamma, \theta_j)$ by

$$\hat{p}_\cdot^{-1} \log \mathrm{PL}_M(\beta, \gamma, \theta_j), \quad \text{where} \quad \hat{p}_\cdot = \frac{1}{n} \sum_{i=1}^{n} \delta_{ij}.$$

Comparing such values for different $\theta_j (j = 1, \ldots, J)$, the maximizing $\theta_{j^{**}}$ would be a reasonable estimate. (We need to be cautious about the estimate if it is equal to 148th or 155th locus.)

The procedure described above is the widely used "inverse probability weighting" (abbreviated as IPW hereafter) method. The IPW method was originally proposed by Horvitz and Thompson [15], and introduced to the Cox model setting by Pugh et al. [36] and further developed by Robins et al. [40] and Xu et al. [56], to name a few.

We call the estimator produced by the "IPW" procedure described above the IPW MPL estimator as opposed to the original MPL estimator without missing data.

Actually besides the IPW method, there are many more missing data methods available for Cox models. A survey of these methods is available in [52]. As long as these methods can be used for the classical Cox models, it can also be used for the method developed in this thesis. The key is that once we fix an individual grid point of the trajectories, our model can be estimated as a classical Cox model.

5.2.2.2 Application of the IPW method

The IPW MPL method is used to estimate the most sensitive locus for the risk of dying of breast cancer. The most sensitive locus is found to be the 199th, out of 206 loci. The name of the locus is "chrX|nt160764988|Xq28|R87497|2.19|2.19 gene".

$$\hat{\theta}_n = 199/206, \quad \hat{\beta}_n = -4.97, \quad \hat{\gamma}_n = 0.69.$$

For this given $\hat{\theta}_n = 199/206$, the standard errors of $\hat{\beta}_n$ and $\hat{\gamma}_n$ obtained from the Cox model estimation are 2.268 and 0.602 respectively. So for the 199th locus, its impact on patients' risk of dying of breast cancer is significant. Higher gene expression levels of the 199th locus are associated with lower risk of dying of breast cancer.

To estimate the Hurst parameter H, the function "pengFit()" in package "fArma" in the statistical software R is used for each subject's chromosome X gene expression level sequence. For each subject, we get an estimate of H. Altogether we have 36 estimates with mean 0.532, 80% falling in $(0.42, 0.64)$. So we take $H = 0.532$ as the estimate of the Hurst parameter.

To estimate σ^2, we still adopt the quadratic variation method.

$$\hat{\sigma^2} = \frac{1}{36}\sum_{i=1}^{36}\sum_{j=1}^{205} SD^2(Z_i((j+1)/206) - Z_i(j/206)) \approx 29.2, \quad \text{hence} \quad \hat{\sigma} \approx 5.4.$$

The estimate of $PN(\tau)$ is $11/36$, the proportion of uncensored events among all these subjects.

To construct the Wald-type confidence interval for θ, β and γ, by Theorem 2.3.2, it remains to get the quantiles of

$$\text{argmax}_{h\in[-n^{1/(2H)}\theta_0, n^{1/(2H)}(1-\theta_0)]}\left(W^H(h) - \frac{|h|^{2H}}{2}|\beta|\sigma\sqrt{PN(\tau)}\right)$$

and the values of s_1^2, s_2^2 and ρ in Theorem 2.3.2.

Since $\theta_0, \beta, PN(\tau), P(X^2N(\tau)), P(Z_H(\theta_0)^2N(\tau)), P(Z_H(\theta_0)XN(\tau))$ are unknown, we replace them by their estimates. The MPLEs of θ_0 and β are $\hat{\theta}_n = 199/206, \hat{\beta}_n = -4.97$. The estimate of H is 0.532. The empirical estimates of the other terms are

$$\mathbb{P}_nN(\tau) = 11/36, \qquad \mathbb{P}_n(Z^2(\theta_0)N(\tau)) = 1.1565/34 = 0.034,$$

$$\mathbb{P}_n(X^2N(\tau)) = 134/36 = 3.722, \quad \mathbb{P}_n(XZ(\theta_0)N(\tau)) = -10/34 = -0.294.$$

Then using the Monte Carlo simulation, by 10,000 replicates, we get the 0.975 and 0.025 quantiles of $n^{1/(2H)}(\hat{\theta}_n - \theta)$'s asymptotic distribution as -0.07 and 0.06. Then we can get the 95% confidence interval for θ as

$$[199/206 - 0.07/36^{1/(2\cdot 0.532)}, 199/206 + 0.06/36^{1/(2\cdot 0.532)}] = [0.9636, 0.9681].$$

Then 95% confidence interval for the loci is [198.5, 199.4]. Due to the coarse grids, we can choose 199th loci as the 95% confidence interval of θ.

The Wald-type confidence intervals for β and γ can also be obtained by their asymptotic normal distributions. Their 95% confidence intervals are $[-8.12, -1.82]$ and $[0.39, 0.99]$ respectively.

5.2.2.3 Empirical study of proposed IPW method

To evaluate whether the new estimator proposed for the missing data has the similar small sample performance to the original one (without missing data), we conduct some simulations under the scenario which mimics the real data set. The statistical software R (version 2.13.0) is used for this simulation.

The simulation conducted here will be different from that in Chapter 5.1.2. In Chapter 5.1.2, we used the true gene expression profiles to generate survival outcomes.

Here, this would not be feasible due to the missing values. If we use the true gene expression profiles to generate survival outcomes, the MCAR (i.e., Missing Completely At Random) assumption does not hold, hence the IPW MPLE method is not valid to estimate θ anymore. In fact, if we still use the IPW MPLE method, all the estimated $\hat{\theta}_n$ from such simulations would be those loci with most severe missings. So in this section, we will simulate the gene expression profiles from randomly generated trajectories of fBm.

The simulation procedure is as follows:

1. Generate a data set which has comparable characteristics to the real data. We have the fitted model parameters $(\hat{\theta}_n, \hat{\beta}_n, \hat{\gamma}_n)$ from the previous section. According to the corresponding estimated cumulative baseline hazard function $\hat{\Lambda}_0(t)$ (see Figure 5.4), we can estimate a 2-phase piecewise constant baseline hazard function $\hat{\lambda}_0(t)$. The empirical values of H, σ, $PN(\tau)$ is known as well, we can mimic the distribution of $Z(\theta_0)$ and X by fitted normal distributions $N(-0.12, 0.22^2)$ and $N(3, 0.75^2)$. The generated data set also have 206 fine grid points on each covariate trajectory for 36 subjects. Given the simulated trajectories, the mechanism to generate time-to-event outcome and censoring variable is the same as that of Chapter 5.1.2.

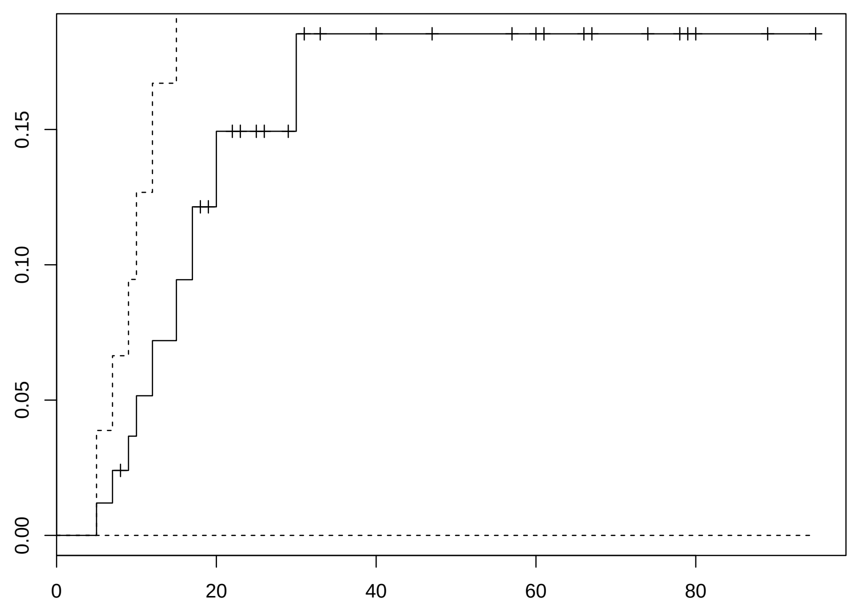
2. Generate a permutation of count of missingness from the true distribution of the count of missingness (137 zeros, 40 ones, 13 twos, 6 threes, 2 fours, 2 fives, 4 sixes, 1 ten and 1 seventeen), and assign them to loci 1,2,3 ,...,206 respectively.

3. For each locus, assign the assigned count of missingness to the 36 subjects with equal probability.

4. Adopt the IPW MPL method to obtain the estimate of (θ, β, γ).

5. Repeat steps 1-4 for 1000 times.

Figure 5.4: The estimated cumulative baseline hazard function and its pointwise 95% confidence interval of the breast cancer study



To obtain the small sample property of the original MPL estimator without missing, we just do step 1 and calculate the estimate of (θ, β, γ), and repeat for Rep=1000 times.

The histograms of 1,000 estimates of $(n^{1/(2H)}(\hat{\theta}_n - \theta_0), \sqrt{n}(\hat{\beta}_n - \beta), \sqrt{n}(\hat{\gamma}_n - \gamma_0))$ for the MPLE without missing and IPW MPLE with missing are presented in the left column and middle column in Figure 5.5. Significant change is observed between the histograms of $\hat{\theta}_n$ in these two scenarios.

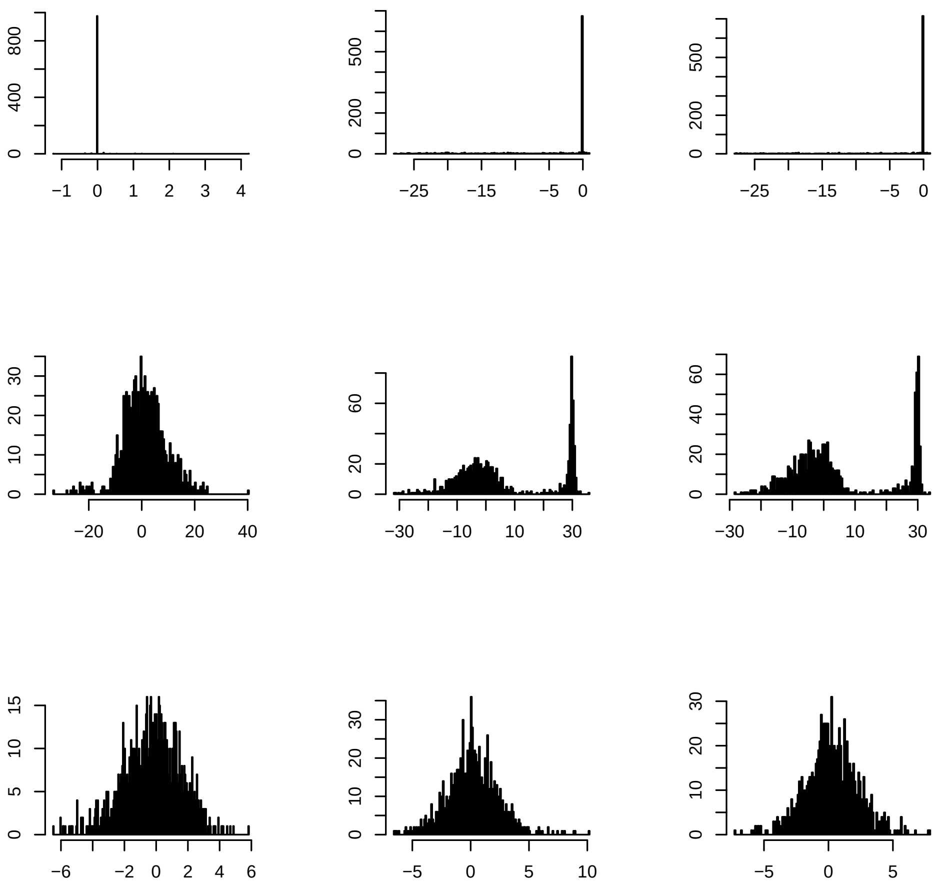
A question is to what extent such a big change is due to the two severe missing grid points, 10 missings (on 155th locus) and 17 missings (on 148th locus). We changed both of them to 6 missings (the maximum missing besides), so we permutate a missing distribution with 137 zeros, 40 ones, 13 twos, 6 threes, 2 fours, 2 fives, 6(=4+2) sixes for Rep=1,000 times, and the histograms of $(n^{1/(2H)}(\hat{\theta}_n - \theta_0), \sqrt{n}(\hat{\beta}_n - \beta), \sqrt{n}(\hat{\gamma}_n - \gamma_0))$ for the IPW MPLE with such a modified missing is shown in the third column.

The shape shows the histogram of $n^{1/(2H)}(\hat{\theta}_n - \theta_0)$ does not really change much compared to its counterpart in the middle column (the IPW MPLE with missing), which implies that the much flatter distribution of $n^{1/(2H)}(\hat{\theta}_n - \theta_0)$ under the IPW MPL is more due to the overall missing effect instead of the two severe missings (i.e. 10 and 17 missings).

An interesting feature of the histograms of $\sqrt{n}(\hat{\beta}_n - \beta)$ under missing data scenarios (the second and third histograms in the middle row) is that besides the cluster centered around zero, there is another cluster around the right tail with a higher peak. For the cluster, its center corresponds to the value of $\hat{\theta}_n$ being zero. Due to the prevalence of missing values, with higher chance that the estimate $\hat{\theta}_n$ is far from the true value, hence the corresponding $Z(\hat{\theta}_n)$ is not correlated to the survival outcome, and the estimate $\hat{\beta}_n$ tends to be close to zero.

We are also interested in the actual confidence limits for (θ, β, γ) when we estimate these parameters by MPLE method and the true values of (θ, β, γ) in the data generating process is (199/206, -4.97, 0.69). The empirical confidence limits obtained

Figure 5.5: Comparison of the IPW MPL estimators under missing and original MPL estimator without missing, upper row $n^{1/(2H)}(\hat{\theta} - \theta_0)$, middle row $\sqrt{n}(\hat{\beta}_n - \beta_0)$, lower row $\sqrt{n}(\hat{\gamma}_n - \gamma_0)$



from the simulated IPW MPL estimators (middle column in Figure 5.5) provides some insights in this respect.

The 0.025 and 0.975 quantiles for $(n^{1/(2H)}(\hat{\theta}_n - \theta_0), \sqrt{n}(\hat{\beta}_n - \beta), \sqrt{n}(\hat{\gamma}_n - \gamma_0))$ are [-24.4, 0], [-19.6, 30.7], [-4.2, 4.4] respectively. It follows that the .95 confidence intervals for θ, β, γ are $[.127, .966]$,$[-8.24, 0.15]$ and $[-0.01, 1.42]$ respectively.

We find the 95% confidence interval for θ, [.127,.966], covers almost all the domain of θ. In contrast, if there is no missing values in the dataset, by the histogram of θ in the first column of Figure 5.5, the 95% confidence interval is $[0.964, 0.968]$. So missing values in this dataset lead to difficulty for the inference of θ.

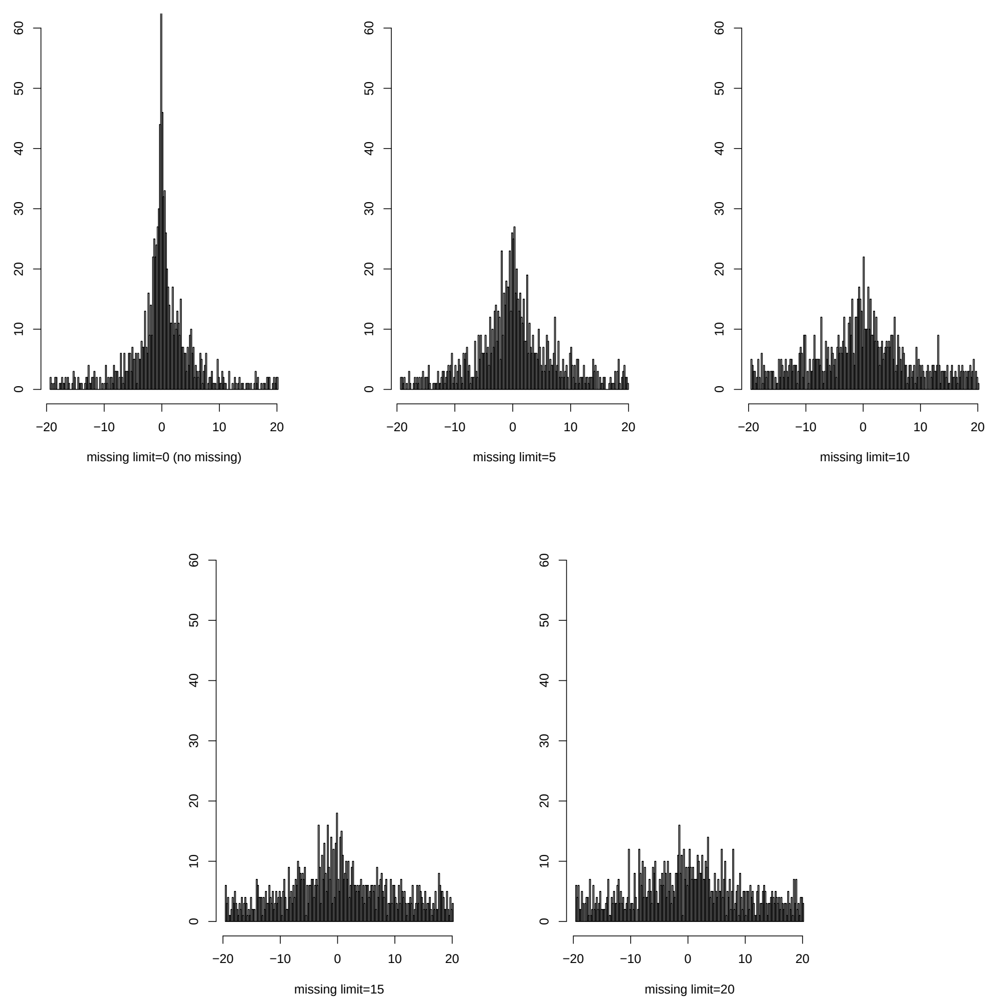
5.2.3 Influence of different proportions of missingness

To understand more about the influence of different proportions of missingness in the data (i.e., count of missing values), simulations are conducted. For $\theta_0 = 0.5$, $\beta_0 = -5$, $\gamma_0 = 0.7$, $J = 200$, $H = 0.5$, $\sigma = 0.3$, $n = 40$, $Z(\theta_0) \sim N(-0.12, 0.22^2)$ and $X \sim N(3, 0.75^2)$, we simulated Rep=1,000 replicates for the missingness size bound equal to 0,1,2,3,...,20. Notice that we adopted similar values to the breast cancer data except for the θ and σ. Here we set $\theta = 0.5$ because we are interested in the length of confidence limits instead of the asymmetry of the confidence interval w.r.t. the point estimate. σ is set to be 0.3 since for larger σ, it requires finer grids, and hence more computing power to obtain confidence limits with reasonable precision.

For missingness count bound equal to $K(0 \leq K \leq 20)$, we generate $n = 40$ random numbers, $M_1, M_2, ..., M_n$ from uniform distribution on the integers from 0 to K and assign these numbers to subjects 1,2,...,40. Each subject i randomly select M_i points on its $J = 200$ grid points and set them as missing values.

Then the IPW MPL estimator is used to estimate (θ, β, γ). In each scenario (i.e. missingness count bound), we have histograms and empirical confidence limits for the estimator's distribution. Obviously the scenario that missingness count bound equal to zero corresponds to the original MPL estimator without missing data. Comparing

Figure 5.6: Comparison of IPW MPL estimator under different proportions of missingness



all these results will show the impact of different levels of missingness.

The adjusted empirical confidence limits come from average of the absolute values of upper-tail and lower-tail ones. Since the distribution of $\hat{\theta}_n$ is expected to be symmetric about zero, taking average can reduce the estimation error of the estimated confidence limits.

Table 5.2 and Figures 5.6, 5.7 show the distributions of the IPW MPL estimator $\hat{\theta}_n$ become flatter with the increase of missing proportions, which means the adjusted empirical confidence limits increase with the increase of missing proportions. The speed of the increase, however, decreases gradually (see Figure 5.7). Eventually, with severer loss of information, the confidence limits will approximate those of the uniform distribution on $[-20, 20]$, 18 and 19 (since the 90% and 95% confidence intervals of the uniform distribution on $[-20, 20]$ are $20 \cdot 0.90 = 18$ and $20 \cdot 0.95 = 19$ respectively).

Note in each scenario of missingness count bounds, the average proportion of missingness is $1/2 \cdot \frac{\text{missing bound}}{\text{full sample size}}$.

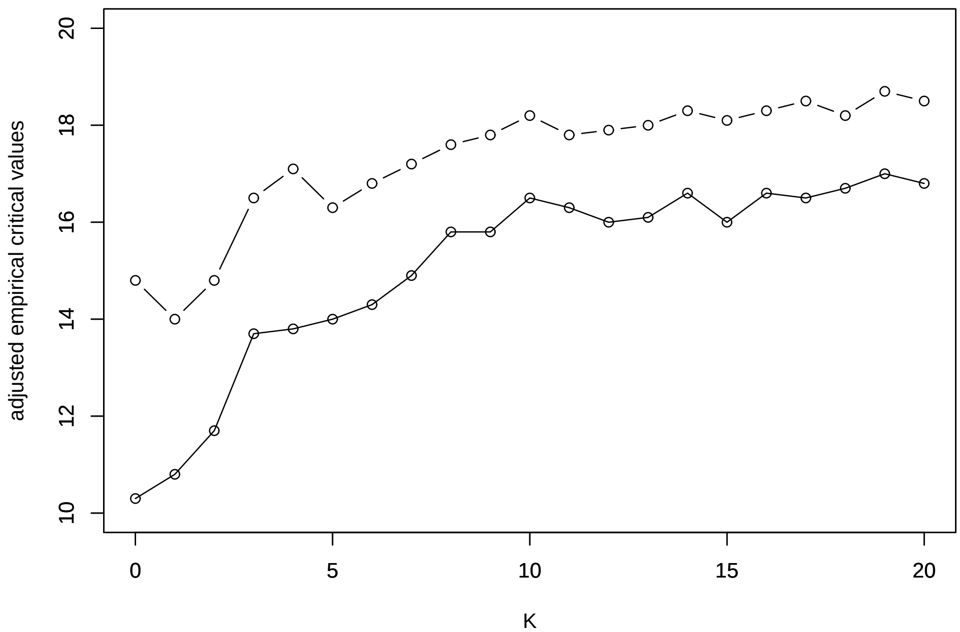
So Figure 5.7 also presents the trend of adjusted empirical confidence limits with the increase of inflation ratio, where inflation ratio is defined as the ratio of full sample size vs. the expected observed sample size under the missing count bound and represent the inflation of sample size caused by the IPW method.

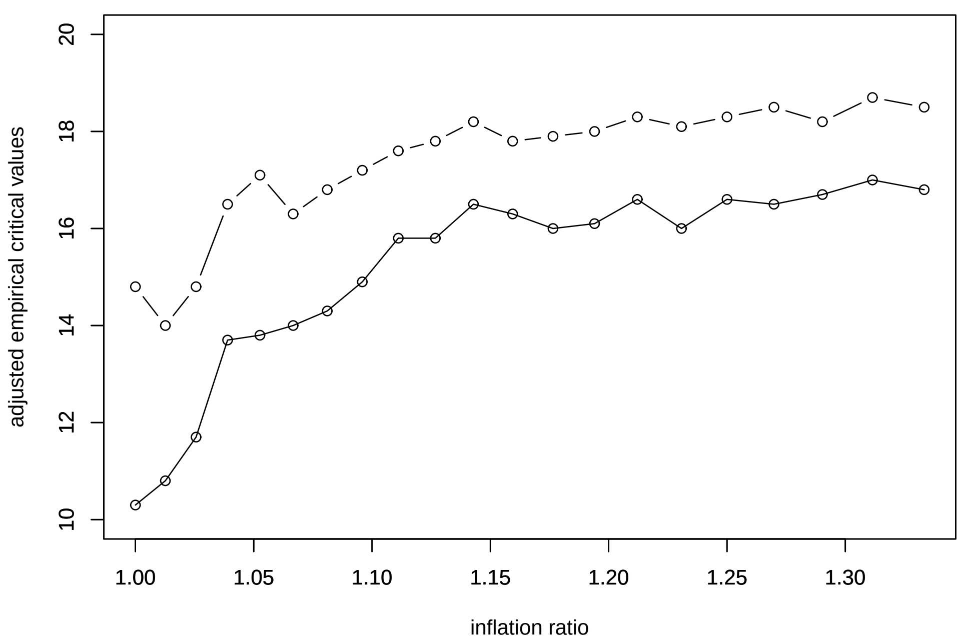
$$\text{inflation ratio} = \frac{\text{full sample size}}{\text{full sample size} - 1/2 \cdot \text{missingness count bound}}.$$

Table 5.2: Empirical confidence limits under different proportions of missingness

K	$EC_{.95}$	$EC_{.975}$	$EC_{.05}$	$EC_{.025}$	adjusted $EC_{.95}$	adjusted $EC_{.975}$
0	10.0	15.2	-10.6	-14.4	10.3	14.8
1	12.0	14.6	-9.6	-13.4	10.8	14.0
2	11.0	15.0	-12.4	-14.6	11.7	14.8
3	13.0	16.2	-14.4	-16.8	13.7	16.5
4	13.8	17.4	-13.8	-16.8	13.8	17.1
5	14.0	16.2	-14.0	-16.4	14.0	16.3
6	14.6	17.0	-14.0	-16.6	14.3	16.8
7	15.8	17.6	-14.0	-16.8	14.9	17.2
8	15.6	17.4	-16.0	-17.8	15.8	17.6
9	16.0	18.0	-15.6	-17.6	15.8	17.8
10	16.4	18.4	-16.6	-18.0	16.5	18.2
11	16.8	18.4	-15.8	-17.2	16.3	17.8
12	16.4	18.2	-15.6	-17.6	16.0	17.9
13	16.2	18.0	-16.0	-18.0	16.1	18.0
14	15.6	17.6	-17.6	-19.0	16.6	18.3
15	16.4	18.4	-15.6	-17.8	16.0	18.1
16	16.6	18.6	-16.6	-18.0	16.6	18.3
17	16.4	19.0	-16.6	-18.0	16.5	18.5
18	16.8	18.4	-16.6	-18.0	16.7	18.2
19	17.4	19.0	-16.6	-18.4	17.0	18.7
20	17.2	18.8	-16.4	-18.2	16.8	18.5

Figure 5.7: Trends of adjusted empirical confidence limits under different missing sizes





Chapter 6

Summary and Discussion

This chapter summarizes the thesis and discusses its applicability and limitations. Possible directions for future research are discussed as well.

6.1 Summary

The thesis starts with an introduction to the field of biomarker discovery in cancer research which motivated the project. Cox models are widely used in cancer research for statistical analysis. Biomarker discovery can be formalized in terms of estimating location parameters in functional predictors for Cox models. Finding "optimal" estimators of such parameters is the goal of this thesis.

Formulating this problem in the framework of M-estimation, we establish large sample properties for the proposed estimator, including consistency, rates of convergence, and asymptotic distributions. The finite sample performance is studied using extensive simulations. Due to the asymmetry of the finite sample distribution of the proposed estimator, we introduce a Domain-Restricted Asymptotic Distribution as a way of providing more accurate calibration for the inferential procedures.

The proposed approach is applied to gene expression data from two cancer mortality studies. To deal with missing gene expression data, an Inverse Proability Weighted

Maximum Partial Likelihood estimator is introduced and its performance is studied.

Besides the approach developed in this thesis, another way to incorporate functional predictor is the extension of the functional linear regression to the Cox model framework. As such an extension models varying effects of the whole predictor process, it may lack interpretability. The approach developed in this thesis captures the most predictive components of the functional covariate, hence it has better interpretability, although it may be less flexible. The proposed approach is more suitable in some applications. For example, in cancer research, instead of estimating the contribution of every gene to the cancer mortality, it is more informative to locate the most sensitive genes to make targeted therapies feasible.

We investigate the proposed method thoroughly from both theoretical and practical perspectives. From the theoretical point of view, the large sample properties of the proposed estimator are studied using empirical processes theory. To construct accurate confidence intervals in finite sample cases, the approach uses a truncated form of the asymptotic distribution, the Domain-Restricted Asymptotic Distribution. The impact of missingness in the functional predictor is evaluated using simulated data.

In this thesis, the proposed method is applied to two genomic data sets as a way of locating biomarkers which are predictive of cancer mortality. Other possible uses could be economics and finance, environmental science and network traffic studies.

6.2 Discussion and future research

This section discusses possible limitations of the proposed approach, and some ideas for future research in this area.

6.2.1 The fBm assumption

A key assumption in developing the large sample theory for the proposed model is that the functional predictor is a 2-sided fractional Brownian motion. FBm is a simple and natural way to represent fractal-like trajectories, and a good starting point. However, we feel the assumption is too restrictive.

There are various ways in which the assumption may be changed or relaxed. The 1-dimensional 2-sided fBm could be extended to higher dimensions, like 2-dimensional or 3-dimensional fBm, or it could be extended to 1-sided fBm.

We can also consider the possibility of extending the model by using other stochastic processes such as the O−U process, the Lévy process or the ARIMA process. However, even if such extensions are feasible, to develop the corresponding theories will be much more challenging.

6.2.2 Sensitive region, independence, missing data, the Bootstrap and others

Restricting attention to using one sensitive point in the proposed model is needed for studying the large sample properties. This is a simplification of the more general situation where a sensitive region is correlated to the time-to-event risk. If a sensitive region is incorporated into the Cox model, what can we do? This was discussed in [24] for the logistic regression setting and similar arguments applies here as well. We can allow multiple covariates in the Cox model, where the multiple covariates form the sensitive region of the functional covariate. For any given length of sensitive region, by maximizing the partial likelihood function, we can estimate the sensitive region of the given length. Since the length is generally unknown in practice, we can add some penalty function to penalize the length of the sensitive region so that we can estimate the length of the sensitive region and the region itself at the same time.

In this thesis, we assume all the subjects in the sample are independent. In practice, they could be correlated. In such scenarios, putting a correlation structure into the model will better describe the data. The *frailty model* is one way to incorporate the correlation information into the model.

Missing data is expected to appear in such high volume data collection. Besides the simple IPW method of handling missing data, it is desirable to study some alternative methods which can be implemented under different missing mechanisms. As mentioned in Chapter 5.2.2.1, various missing data methods surveyed in [52] can be used in our setting as long as they can be used in the classical Cox model setting.

Another aspect worthy of further investigation is the bootstrap. Since our main interest is θ (and β, γ), there are nuisance parameters like (H, σ) involved in this model. Using the bootstrap could help us circumvent the problem of estimating nuisance parameters. However, like in the linear regression setting with the fBm assumption which was discussed in [29], we do not expect Efron's bootstrap to provide valid inference. Whether a martingale residual bootstrap [25] or a weighted bootstrap [12] works deserves further investigation.

For applications, there are other practical considerations.

We only consider the cross sectional measurement of the functional covariate at certain time point, or we assume it won't change over time. In fact, the functional covariate's value could be time varying. For example, the gene expression level would change if measured at different time points. To study the time-varying impact systematically is beyond the scope of the thesis. However, it would be interesting to estimate the sensitive point and its confidence interval sequentially for a series of measurements of the functional covariates at different time, and see how the sensitive point and region change over time.

Another practical consideration is that the sensitive point itself could be missed in the data collection process. If the missing is partial, i.e., not all the subjects missed the measurement of the functional covariate at the sensitive point, the efficiency of the proposed method would be impaired. The degree of impairment depends on the proportion of missingness as shown in Chapter 5.2.3. As for what proportion of missingness would render the proposed method hopeless to capture the signal, it depends on too many factors to have a rule of thumb. For any specific problem, to evaluate the impact of missingness, we can conduct simulation following the example of Chapter

5.2.3. However, if all the subjects missed the measurement of the functional covariate at the sensitive time point, not just our method, essentially there is no way to find out the sensitive time point by statistical techniques.

The last point to be addressed is the computational feasibility of the proposed method. Fitting Cox models is computationally more costly than linear model. To calibrate the quantiles of the DRAD is much more expensive than fitting the Cox model itself. When the functional covariate is observed on a dense grid of points, implementing the proposed method would be challenging computationally. If the number of grid points is J, the number of subjects (i.e., sample size) is n, it seems feasible only if $n \cdot J < 10^7$ on a desktop computer.

Bibliography

[1] P. K. Andersen and R. D. Gill. Cox's regression model for counting processes: A large sample study. *Ann. Statist.*, 10(4):1100–1120, 1982.

[2] P. K. Anderson, O. Borgan, R. D. Gill, and N. Keiding. *Statistical Models based on Counting Processes.* Springer-Verlag, 1993.

[3] M. Banerjee and I. W. McKeague. Confidence sets for split points in decision trees. *Ann. Statist.*, 35(2):543–574, 2007.

[4] M. Banerjee and I. W. McKeague. Estimating optimal step-function approximations to instantaneous hazard rates. *Bernoulli*, 13(1):279–299, 2007.

[5] P. K. Bhattacharya and P. J. Brockwell. The minimum of an additive process with applications to signal estimation and storage theory. *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 37:51–75, 1976.

[6] Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin. Pharmacol. Ther.*, 69(3):89–95, Mar 2001.

[7] N. Breslow. Contribution to the discussion of the paper by D.R. Cox. *J. Roy. Statist. Soc. B.*, 34:187–220, 1972.

[8] N. Breslow. Covariance analysis of censored survival data. *Biometrics*, 30:89–99, 1974.

[9] M. Brooks. Breast cancer screening and biomarkers. *Methods Mol Biol.*, 472:307–21, 2009.

[10] D. R. Cox. Regression models and life tables (with discussion). *J. Roy. Statist. Soc. B.*, 34:187–220, 1972.

[11] D. R. Cox. Partial likelihood. *Biometrika*, 62:269–276, 1975.

[12] W. Q. Cui, Y. N. Yang, and Y. H. Wu. Random weighting method for Cox's proportional hazards model. *Science in China Series A: Mathematics*, 51(10):1843–1854, Oct. 2008.

[13] B. Efron. Efficiency of Cox's likelihood function for censored data. *J. Amer. Statist. Assoc.*, 72:557–565, 1977.

[14] F. B. Furnari, T. Fenton, R. M. Bachoo, and A. Mukasa. Malignant astrocytic glioma: genetics, biology, and paths to treatment. *Genes Development*, 21:2683–2710, 2007.

[15] D. G. Horvitz and D. J. Thompson. A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.*, 47:663–685, 1952.

[16] M.M. Inda, X Fan, J. Munoz, and C. Perot et al. Chromosomal abnormalities in human glioblastomas: gain in chromosome 7p correlating with loss in chromosome 10q. *Molecular Carcinogenesis*, 36(1):6–14, 2003.

[17] G. M. James, J. Wang, and J. Zhu. Functional linear regression that's interpretable. *Ann. Statist.*, 37:2083–2108, 2009.

[18] J. D. Kalbfleisch and R. L. Prentice. Marginal likelihoods based on Cox's regression and life model. *Biometrika*, 60:267–278, 1973.

[19] J. D. Kalbfleisch and R. L. Prentice. *The Statistical Analysis of Failure Time Data*. Wiley-InterScience, 2002.

[20] M.E. Law, K.L. Templeton, G. Kitange, Smith J., A. Misra, Feuerstein B.G., and Jenkins R.B. Molecular cytogenetic analysis of chromosomes 1 and 19 in glioma cell lines. *Cancer Genetics and Cytogenetics*, 160(1):1–14, July 2005.

[21] Chihoon Lee. A geometric drift inequality for a reflected fractional Brownian motion process on the positive orthant. *Journal of Applied Probability*, 48(3), 2011.

[22] Timothy J. Ley, E.R. Mardis, L Ding, and B Fulton et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature*, 456:66–72, 2008.

[23] E. Lieberman-Aiden, N. L. van Berkum, and L.m Williams et al. Comprehensive mapping of long-range interactions reveals folding principles of human genome. *Science*, 326:289–293, 2009.

[24] Martin A. Lindquist and Ian W. McKeague. Logistic regression with Brownian-like predictors. *J. Amer. Statist. Assoc.*, 104(488):1575–1585, December 2009.

[25] Thomas M. Loughin. A residual bootstrap for regression parameters in proportional hazards models. *Journal of Statistical Computation and Simulation*, 52(4):367–384, 1995.

[26] D Maier, Z Zhang, E Taylor, MF Hamou, O Gratzl, EG Van Meir, RJ Scott, and A. Merlo. Somatic deletion mapping on chromosome 10 and sequence analysis of pten/mmac1 point to the 10q25-26 region as the primary target in low-grade and high-grade gliomas. *Oncogene*, 16(25):3331–3335, June 1998.

[27] B. B. Mandelbrot. *The Fractal Geometry of Nature*. W. H. Freeman and Co., New York, 1982.

[28] B. B. Mandelbrot and J. Van Ness. Fractional Brownian motions, fractional noises and applications. *SIAM Rev.*, 10:422–437, 1968.

[29] I. W. McKeague and B. Sen. Fractals with point impact in functional linear regression. *Ann. Statist.*, 38(4):2559–2586, 2010.

[30] National Cancer Institute. Website, Dec. 1, 2010. `http://www.cancer.gov/cancertopics/factsheet/Therapy/targeted`.

[31] A. Novikov and E. Valkeila. On some maximal inequalities for fractional Brownian motions. *Statistics and Probability Letters*, 44(1):47–54, 1999.

[32] D. Oakes. The asymptotic information in censored survival data. *Biometrika*, 64:441–448, 1977.

[33] C.-K. Peng, S.V. Buldyrev, S. Havlin, M. Simons, H.E. Stanley, and A. L. Goldberger. Mosaic organization of dna nucleotides. *Phys. Rev. E*, 49:1685–1689, 1994.

[34] E. D. Pleasance, R.K. Cheetham, P.J. Stephens, and D.J. McBride et al. A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature*, 463:191–196, 2010.

[35] E. D. Pleasance, P.J. Stephens, S. O' Meara, and D.J. McBride et al. A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature*, 463:184–190, 2010.

[36] M. Pugh, J. Robins, S. Lipsitz, and D. Harrington. Inference in the Cox proportional hazards model with missing covariate data. *Technical Report*, 1994. Harvard School of Public Health, Dept. of Biostatistics.

[37] J.O. Ramsay and B.W. Silverman. *Functional Data Analysis*. Springer, second edition, 2005.

[38] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion*. Springer, 2005.

[39] A.L. Richardson, Z.C. Wang, A. De Nicolo, X. Lu, M. Brown, A. Miron, Liao X., J.D. Iglehart, D.M. Livingston, and S. Ganesan. X chromosomal abnormality in basel-like human breast cancer. *Cancer Cell*, 9:121–132, 2006.

[40] J. Robins, A. Rotnitzky, and L. P. Zhao. Estimation of regression coefficients when some regressors are not always observed. *J. Amer. Statist. Assoc.*, 89:846–866, 1994.

[41] Donald B. Rubin and Roderick J. A. Little. *Statistical Analysis with Missing Data*. New York: Wiley, second edition, 2002.

[42] W. Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, third edition, 1976.

[43] S. P. Shah, R.D. Morin, J. Khattra, and L. Prentice et al. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature*, 461:809–813, 2009.

[44] J.M. Simard, S.K. Woo, M.D. Norenberg, C. Tosun, Z. Chen, S. Ivanova, O. T-symbalyuk, J. Bryan, D. Landsman, and V. Gerzanich. Brief suppression of Abcc8 prevents autodestruction of spinal cord after trauma. *Sci Transl Med.*, 2:28ra29, April 2010.

[45] Therese Sorlie, C.M. Perou, R. Tibshirani, and T. Aas et al. Gene expression patterns of breast cancinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. USA*, 98(19):10869–10874, 2001.

[46] H. Stryhn. The location of the maximum of asymmetric two-sided Brownian motion with triangular drift. *Statistics and Probability letters*, 29:279–284, 1996.

[47] Erik Talvila. Necessary and sufficient conditions for differentiating under the integral sign. *American Mathematical Monthly*, 108:544–548, 2001.

[48] The Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*, 455:1061–1068, 2008.

[49] R. Tibshirani. Regression shrinkage and selection via the LASSO. *J. Roy. Statist. Soc. B.*, 58(1):267–288, 1996.

[50] R. Tibshirani. The LASSO method for variable selection in the Cox model. *Statistics in Medicine*, 16(4):385–395, 1997.

[51] Mark R. Trusheim, Ernst R. Berndt, and Frank L. Douglas. Stratified medicine: strategic and economic implications of combining drugs and clinical biomarkers. *Nature Reviews Drug Discovery*, 6:287–293, 2007.

[52] Anastasios Tsiatis. *Semiparametric Theory and Missing Data*. Springer, 2006.

[53] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.

[54] A. W. van der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes*. Springer, 1996.

[55] A. W. van der Vaart and J. A. Wellner. Empirical processes indexed by estimated functions. *Asymptotics: particles, processes and inverse Problems*, IMS Lecture Notes Monograph Series:234–252, April 2007. Institute of Mathematical Statistics.

[56] Q. Xu, M. C. Paik, X. Luo, and W. Y. Tsai. Reweighting estimators for Cox regression with missing covariates. *J. Amer. Statist. Assoc.*, 104:1155–1167, 2009.

Appendix A

Additional proof details for Chapter 2.2

As mentioned in Chapter 2.2.2.3, we will prove the rate of convergence for the simplified model here. To obtain the rate of convergence for the M-estimator of θ, we can use Theorem 3.2.5 in VW [54]. The key step is get the modulus of continuity.

Modulus of continuity

To apply the *rate of convergence* Theorem, we will try to find the modulus of continuity in this section, i.e., to bound $P^* \sup_{d(\theta,\theta_0)<\delta} \sqrt{n}|(\mathbb{M}_n - \mathbb{M})(\theta) - (\mathbb{M}_n - \mathbb{M})(\theta_0)|$, where

$$\mathbb{M}_n(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left[Z_i(\theta)N_i(\tau) - \int_0^\tau \log S^{(0)}(\theta,u)dN_i(u)\right],$$

$$\mathbb{M}(\theta) = P\left[Z(\theta)N(\tau) - \int_0^\tau \log P[Y(u)e^{Z(\theta)}]dN(u)\right].$$

Write $\mathbb{M}_n(\theta)$ in empirical process form,

$$\mathbb{M}_n(\theta) = \mathbb{P}_n\left[Z(\theta)N(\tau) - \int_0^\tau \log \mathbb{P}_n[Y(u)e^{Z(\theta)}]dN(u)\right].$$

$$\sqrt{n}\left[(\mathbb{M}_n - \mathbb{M})(\theta) - (\mathbb{M}_n - \mathbb{M})(\theta_0)\right]$$

$$
\begin{aligned}
= \quad & \sqrt{n}\mathbb{P}_n\left[Z(\theta)N(\tau) - \int_0^\tau \log[\mathbb{P}_n Y(u)e^{Z(\theta)}]dN(u)\right] \\
& \qquad -\sqrt{n}P\left[Z(\theta)N(\tau) - \int_0^\tau \log s^{(0)}(\theta,u)dN(u)\right] \\
& -\sqrt{n}\mathbb{P}_n\left[Z(\theta_0)N(\tau) - \int_0^\tau \log[\mathbb{P}_n Y(u)e^{Z(\theta_0)}]dN(u)\right] \\
& \qquad +\sqrt{n}P\left[Z(\theta_0)N(\tau) - \int_0^\tau \log s^{(0)}(\theta_0,u)dN(u)\right] \\
= \quad & \sqrt{n}(\mathbb{P}_n - P)\left[(Z(\theta) - Z(\theta_0))N(\tau)\right] \\
& \qquad +\sqrt{n}\int_0^\tau \left[\log s^{(0)}(\theta,u) - \log s^{(0)}(\theta_0,u)\right] PdN(u) \\
& \qquad -\sqrt{n}\int_0^\tau [\log \mathbb{P}_n Y(u)e^{Z(\theta)} - \log \mathbb{P}_n Y(u)e^{Z(\theta_0)}]\mathbb{P}_n dN(u) \\
= \quad & \sqrt{n}(\mathbb{P}_n - P)\left[(Z(\theta) - Z(\theta_0))N(\tau)\right] \\
& \qquad -\int_0^\tau (\log s^{(0)}(\theta,u) - \log s^{(0)}(\theta_0,u))\sqrt{n}(\mathbb{P}_n - P)dN(u) \\
& -\sqrt{n}\int_0^\tau \left[\log \mathbb{P}_n Y(u)e^{Z(\theta)} - \log s^{(0)}(\theta,u)\right]\mathbb{P}_n dN(u) \\
& \qquad +\sqrt{n}\int_0^\tau \left[\log \mathbb{P}_n Y(u)e^{Z(\theta_0)} - \log s^{(0)}(\theta_0,u)\right]\mathbb{P}_n dN(u) \\
\equiv\ & I_1 - I_2 - I_3 + I_4,
\end{aligned}
$$

where the third equality follows by subtracting

$$\sqrt{n}\int_0^\tau \left[\log s^{(0)}(\theta,u) - \log s^{(0)}(\theta_0,u)\right]\mathbb{P}_n dN(u)$$

from the second item while adding them into the third item.

By Theorem 5.2 in Banerjee and McKeague [3], a slight extension of Theorem 3.2.5 of [54], to derive an upper bound for the rate of convergence of $\hat{\theta}_n$, it suffices to bound

$$P^* \sup_{d(\theta,\theta_0)<\delta} \sqrt{n}|(\mathbb{M}_n - \mathbb{M})(\theta) - (\mathbb{M}_n - \mathbb{M})(\theta_0)|1_{\Omega_n},$$

where $\{\Omega_n\}_{n\geq 1}$ is a sequence of subsets of the sample space, such that $P^*(\Omega_n) \to 1$ as $n \to \infty$. The Ω_n could be appropriately chosen to make the calculation of the modulus of continuity easier.

$$
\begin{aligned}
& P^* \sup_{d(\theta,\theta_0)<\delta} \sqrt{n}|(\mathbb{M}_n - \mathbb{M})(\theta) - (\mathbb{M}_n - \mathbb{M})(\theta_0)| 1_{\Omega_n} \\
&= P^* \sup_{d(\theta,\theta_0)<\delta} |I_1 - I_2 - I_3 + I_4| 1_{\Omega_n} \\
&\leq P^* \sup_{d(\theta,\theta_0)<\delta} |I_1| + P^* \sup_{d(\theta,\theta_0)<\delta} |I_2| + P^* \sup_{d(\theta,\theta_0)<\delta} |I_3 - I_4| 1_{\Omega_n}
\end{aligned}
$$

Now we deal with each term separately.

The first term

$$P^* \sup_{d(\theta,\theta_0)<\delta} |I_1| = P^* \sup_{d(\theta,\theta_0)<\delta} \left|\mathbb{G}_n \left[W(\theta - \theta_0) N(\tau)\right]\right|,$$

can be written as

$$P^* \sup_{\sqrt{|x|}<\delta} \left|\mathbb{G}_n \left[W(x) N(\tau)\right]\right| \leq 2P^* \sup_{0\leq x<\delta^2} \left|\mathbb{G}_n \left[W(x) N(\tau)\right]\right|$$

Similar to $\mathcal{M}_{ZN,\theta_M}$, $\mathcal{M}_{WN,\delta^2} = \{W(x)N(\tau) : 0 \leq x < \delta^2\}$, also has bounded bracketing entropy, i.e. $J_{[]}(1, \mathcal{M}_{WN,\delta^2}, L_2(P)) < \infty$. Then by Theorem 2.14.2 of [54], it can be further bounded by

$$
\begin{aligned}
& J_{[]}(1, \mathcal{M}_{WN,\delta^2}, L_2(P)) \sqrt{P^* \sup_{0\leq x<\delta^2} W^2(x) N^2(1)} \lesssim \sqrt{P^* \sup_{0\leq x<\delta^2} W^2(x)} \\
& \lesssim \sqrt{4P[W^2(\delta^2)]} \lesssim \sqrt{4\delta^2} \lesssim \delta,
\end{aligned}
$$

where we used $N(1) \leq 1$ in the first inequality and Doob's maximal inequality in the second inequality.

For the second term,

$$
\begin{aligned}
P^* \sup_{d(\theta,\theta_0)<\delta} |I_2| &= P^* \sup_{d(\theta,\theta_0)<\delta} \left| \int_0^\tau \log e^{\frac{|\theta-\theta_0|}{2}} \sqrt{n}(\mathbb{P}_n - P) dN(u) \right| \\
&= P^* \sup_{d(\theta,\theta_0)<\delta} \frac{|\theta - \theta_0|}{2} \left|\sqrt{n}(\mathbb{P}_n - P) N(\tau)\right| \\
&\leq \frac{\delta^2}{2} \left(P^* |\sqrt{n}(\mathbb{P}_n - P) N(\tau)|^2\right)^{\frac{1}{2}} \leq \frac{\delta^2}{2} \left[P^* N^2(\tau)\right]^{\frac{1}{2}} \leq \frac{\delta^2}{2}.
\end{aligned}
$$

Now consider $I_3 = \sqrt{n}\int_0^\tau \left[\log \mathbb{P}_n Y(u)e^{Z(\theta)} - \log s^{(0)}(\theta,u)\right]\mathbb{P}_n dN(u)$ and

$$I_4 = \sqrt{n}\int_0^\tau \left[\log \mathbb{P}_n Y(u)e^{Z(\theta_0)} - \log s^{(0)}(\theta_0,u)\right]\mathbb{P}_n dN(u).$$

Since function $\log(x)$ is continuously differentiable with derivative $\frac{1}{x}$, by Mean Value Theorem (page 108 of [42]), we have

$$I_3 = \int_0^\tau \frac{S^{(0)}(\theta,u) - s^{(0)}(\theta,u)}{\tilde{S}^{(0)}(\theta,u)}\sqrt{n}\mathbb{P}_n dN(u),$$
$$I_4 = \int_0^\tau \frac{S^{(0)}(\theta_0,u) - s^{(0)}(\theta_0,u)}{\tilde{S}^{(0)}(\theta_0,u)}\sqrt{n}\mathbb{P}_n dN(u),$$

where

$$\tilde{S}^{(0)}(\theta,u) \equiv K_{\theta,u}(\omega)S^{(0)}(\theta,u) + (1-K_{\theta,u}(\omega))s^{(0)}(\theta,u),$$
$$\tilde{S}^{(0)}(\theta_0,u) \equiv K_{\theta_0,u}(\omega)S^{(0)}(\theta_0,u) + (1-K_{\theta_0,u}(\omega))s^{(0)}(\theta_0,u),$$

with $K_{\theta,u}(\omega), K_{\theta_0,u}(\omega) \in (0,1)$. Note that $K_{\theta,u}, K_{\theta_0,u}$ are all random variables, and we write out ω in the two preceding displayed equations to stress this point.

$$I_3 - I_4 = \int_0^\tau \left[\frac{S^{(0)}(\theta,u) - s^{(0)}(\theta,u)}{\tilde{S}^{(0)}(\theta,u)} - \frac{S^{(0)}(\theta_0,u) - s^{(0)}(\theta_0,u)}{\tilde{S}^{(0)}(\theta_0,u)}\right]\sqrt{n}\mathbb{P}_n dN(u)$$

$$= \int_0^\tau \frac{[S^{(0)}(\theta,u) - s^{(0)}(\theta,u)] - [S^{(0)}(\theta_0,u) - s^{(0)}(\theta_0,u)]}{\tilde{S}^{(0)}(\theta,u)}\sqrt{n}\mathbb{P}_n dN(u)$$
$$+ \int_0^\tau \left(\frac{1}{\tilde{S}^{(0)}(\theta,u)} - \frac{1}{\tilde{S}^{(0)}(\theta_0,u)}\right)\left[S^{(0)}(\theta_0,u) - s^{(0)}(\theta_0,u)\right]\sqrt{n}\mathbb{P}_n dN(u)$$
$$\equiv I_3' + I_4'$$

Since the numerator of I_3''s integrand can be written as empirical process form,

$$[S^{(0)}(\theta,u) - s^{(0)}(\theta,u)] - [S^{(0)}(\theta_0,u) - s^{(0)}(\theta_0,u)] = (\mathbb{P}_n - P)\left[Y(u)\left(e^{Z(\theta)} - e^{Z(\theta_0)}\right)\right],$$

we now consider the bracketing entropy property of $\{Y(u)\left(e^{Z(\theta)} - e^{Z(\theta_0)}\right) : u \in [0,\tau], \theta \in [0,\theta_M]\}$ in order to apply Theorem 2.14.2 in [54] to bound the $L_1(P)$ norm of $\sqrt{n}(\mathbb{P}_n - P)\left[Y(u)\left(e^{Z(\theta)} - e^{Z(\theta_0)}\right)\right]$.

Now we have a Lemma to present, as the tool to study the bracketing entropy property of $\{Y(u)\left(e^{Z(\theta)} - e^{Z(\theta_0)}\right) : u \in [0,\tau], \theta \in [0,\theta_M]\}$.

Lemma A.0.1. *Under Assumptions 2.2.1, the class of functions* $\mathcal{M} = \{Y(u)e^{Z(\theta)} : u \in [0,\tau], \theta \in [0,\theta_M]\}$ *has finite bracketing entropy.*

The proof can be found in Appendix C.1.

Notice that by Theorem 2.14.2 of VW [54],

$$P^* \sup_{u,\theta} \left|\sqrt{n}(\mathbb{P}_n - P)\left(Y(u)e^{Z(\theta)}\right)\right| \leq J_{[]}(1, \mathcal{M}, L_2(P))\sqrt{P^* \sup_{u,\theta}\left(Y(u)e^{Z(\theta)}\right)^2}. \tag{A.1}$$

By Lemma A.0.1, $\quad J_{[]}(1, \mathcal{M}, L_2(P)) < \infty$.

Considering the independence of $Z(\theta) - Z(\theta_0)$ and $Z(\theta_0)$ by Assumptions 2.2.1, the distributional property of Brownian motion and maximal inequality for submartingales,

$$P^* \sup_{u,\theta}\left(Y(u)e^{Z(\theta)}\right)^2 \leq P\left[e^{2Z(\theta_0)}\right] \cdot 4P\left[e^{2W(\theta_M)}\right] < \infty.$$

Then divide both sides of (A.1) by $n^{1/6}$, we have

$$P^* \sup_{u,\theta}\left|n^{1/3}(\mathbb{P}_n - P)\left(Y(u)e^{Z(\theta)}\right)\right| \lesssim n^{-1/6},$$

$$\lim_{n\to\infty} P^* \sup_{u,\theta}\left|n^{1/3}(\mathbb{P}_n - P)\left(Y(u)e^{Z(\theta)}\right)\right| = 0.$$

By Markov Inequality,

$$P^*\left(\sup_{u,\theta}\left|n^{1/3}(\mathbb{P}_n - P)\left(Y(u)e^{Z(\theta)}\right)\right| \geq 1\right)$$
$$\leq P^* \sup_{u,\theta}\left|n^{1/3}(\mathbb{P}_n - P)\left(Y(u)e^{Z(\theta)}\right)\right| \lesssim n^{-1/6}.$$

If we define

$$\Omega_n \equiv \{\omega : \sup_{u,\theta} n^{\frac{1}{3}} |S^{(0)}(\theta, u) - s^{(0)}(\theta, u)| \leq 1\}$$
$$= \{\omega : \sup_{u,\theta} |S^{(0)}(\theta, u) - s^{(0)}(\theta, u)| \leq n^{-\frac{1}{3}}\},$$

then $P^*(\Omega_n) = 1 - P^*\left(\sup_{u,\theta} \left|n^{1/3}(\mathbb{P}_n - P)\left(Y(u)e^{Z(\theta)}\right)\right| \geq 1\right) \geq 1 - c \cdot n^{-1/6} \to 1$, as $n \to \infty$. We have designed a sequence of subsets, Ω_n, of the sample space such that $P^*(\Omega_n) \to 1$ as $n \to \infty$.

Note here $S^{(0)}(\theta, u)$ depends on n.

Now recall (2.4) and (2.9) for use in the following arguments.

$\forall n > \left(\frac{2}{B(\theta_0)}\right)^3$, we have $n^{-\frac{1}{3}} < \frac{1}{2}B(\theta_0)$, then on Ω_n for all $n > \left(\frac{2}{B(\theta_0)}\right)^3$,

$$S^{(0)}(\theta, u) \geq s^{(0)}(\theta, u) - \tfrac{1}{2}B(\theta_0) \geq \tfrac{1}{2}B(\theta_0). \tag{A.2}$$

We have that Ω_n satisfies $P^*(\Omega_n) \to 1$, and for those $n > \left(\frac{2}{B(\theta_0)}\right)^3 \equiv N_{\theta_0}$, $S^{(0)}(\theta, u)$ has the positive lower bound $\frac{1}{2}B(\theta_0)$ on Ω_n.

$$\begin{aligned}
&P^* \sup_{d(\theta,\theta_0)<\delta} |I_3'| 1_{\Omega_n} 1_{n > N_{\theta_0}} \\
&= P^* \sup_{d(\theta,\theta_0)<\delta} \int_0^\tau \frac{\sqrt{n}|(\mathbb{P}_n - P)Y(u)(e^{Z(\theta)} - e^{Z(\theta_0)})|}{\tilde{S}^{(0)}(\theta, u)} 1_{\Omega_n} 1_{n > N_{\theta_0}} \mathbb{P}_n dN(u) \\
&\leq P^* \sup_{d(\theta,\theta_0)<\delta} \int_0^\tau \frac{\sqrt{n}|(\mathbb{P}_n - P)Y(u)(e^{Z(\theta)} - e^{Z(\theta_0)})|}{\frac{1}{2}B(\theta_0)} \mathbb{P}_n dN(u) \\
&\leq P^* \sup_{d(\theta,\theta_0)<\delta, u \in [0,\tau]} \sqrt{n} \left|(\mathbb{P}_n - P)Y(u)(e^{Z(\theta)} - e^{Z(\theta_0)})\right| \int_0^\tau \frac{2\mathbb{P}_n dN(u)}{B(\theta_0)} \\
&= P^* \sup_{d(\theta,\theta_0)<\delta, u \in [0,\tau]} \left|\mathbb{G}_n Y(u)(e^{Z(\theta)} - e^{Z(\theta_0)})\right| \frac{2\mathbb{P}_n N(\tau)}{B(\theta_0)} \\
&\leq \frac{2}{B(\theta_0)} P^* \sup_{d(\theta,\theta_0)<\delta, u \in [0,\tau]} \left|\mathbb{G}_n Y(u)(e^{Z(\theta)} - e^{Z(\theta_0)})\right|,
\end{aligned}$$

where the first inequality utilizes the lower bound of the denominator $\tilde{S}^{(0)}(\theta, u)$ on Ω_n for large n and the last inequality follows from $N(\tau) \leq 1$.

Denote $\mathcal{M}_{Z,\delta} \equiv \{Y(u)(e^{Z(\theta)} - e^{Z(\theta_0)}) : u \in [0,\tau], d(\theta,\theta_0) < \delta\}$, then by Theorem 2.14.2 of VW [54],

$$\begin{aligned}
&P^* \sup_{d(\theta,\theta_0)<\delta, u\in[0,\tau]} \left|\mathbb{G}_n Y(u)(e^{Z(\theta)} - e^{Z(\theta_0)})\right| \\
&\le J_{[]}(1, \mathcal{M}_{Z,\delta}, L_2(P)) \sqrt{P^* \sup_{d(\theta,\theta_0)<\delta, u\in[0,\tau]} Y^2(u)(e^{Z(\theta)} - e^{Z(\theta_0)})^2} \\
&\le J_{[]}(1, \mathcal{M}_{Z,\delta}, L_2(P)) \sqrt{P^* \sup_{d(\theta,\theta_0)<\delta} (e^{Z(\theta)} - e^{Z(\theta_0)})^2}. \qquad \text{(A.3)}
\end{aligned}$$

We make some transformations for the class of functions $\mathcal{M}_{Z,\delta}$.

$$\begin{aligned}
\mathcal{M}_{Z,\delta} &= \{Y(u)e^{Z(\theta_0)}(e^{Z(\theta)-Z(\theta_0)} - 1) : u \in [0,\tau], d(\theta,\theta_0) < \delta\} \\
&= \{e^{Z(\theta_0)}Y(u)(e^{W(\theta-\theta_0)} - 1) : u \in [0,\tau], d(\theta,\theta_0) < \delta\} \\
&= \{e^{Z(\theta_0)}Y(u)(e^{W(x)} - 1) : u \in [0,\tau], |x| < \delta^2\} \equiv e^{Z(\theta_0)}\mathcal{M}_{W-,\delta},
\end{aligned}$$

where $\mathcal{M}_{W-,\delta} = \{Y(u)(e^{W(x)} - 1) : u \in [0,\tau], |x| < \delta^2\}, W(\cdot) = Z(\theta_0 + \cdot) - Z(\theta_0)$ is a 2-sided S.B.M. starting from 0. Since $P(e^{2Z(\theta_0)}) < \infty$ from Assumptions 2.2.1, by transformation of brackets,

$$J_{[]}\left(1, \mathcal{M}_{Z,\delta}, L_2(P)\right) \lesssim J_{[]}\left(1, \mathcal{M}_{W-,\delta}, L_2(P)\right). \qquad \text{(A.4)}$$

$$\begin{aligned}
\mathcal{M}_{W-,\delta} =& \{Y(u)(e^{B(x)} - 1) : u \in [0,\tau], 0 \le x < \delta^2\} \\
&\cup \{Y(u)(e^{B(x)} - 1) : u \in [0,\tau], -\delta^2 < x \le 0\} \\
&\equiv \mathcal{M}^+_{B-,\delta} \cup \mathcal{M}^-_{B-,\delta}, \quad \text{where } B(x) \text{ is 1-sided S.B.M. starting from 0.}
\end{aligned}$$

By Lemma C.1.1 and symmetry,

$$J_{[]}\left(1, \mathcal{M}_{W-,\delta}, L_2(P)\right) \lesssim J_{[]}\left(1, \mathcal{M}^+_{B-,\delta}, L_2(P)\right). \qquad \text{(A.5)}$$

Since $\mathcal{M} \equiv \{Y(u)e^{B(x)} : u \in [0,\tau], 0 \le x < \theta_M\}$ has bounded bracketing entropy integral by Lemma A.0.1, by the same way of proving Lemma A.0.1, we have $\mathcal{M}_- \equiv \{Y(u)(e^{B(x)} - 1) : u \in [0,\tau], 0 \le x < \theta_M\}$ also has bounded bracketing entropy integral, it follows that $\mathcal{M}^+_{B-,\delta}$ has bounded bracketing entropy integral by $\mathcal{M}^+_{B-,\delta} \subset \mathcal{M}_-$, i.e.,

$$J_{[]}\left(1, \mathcal{M}^+_{B-,\delta}, L_2(P)\right) < \infty. \tag{A.6}$$

Then by (A.4), (A.5) and (A.6), $J_{[]}(1, \mathcal{M}_{Z,\delta}, L_2(P)) < \infty$ has been proved.

$$\begin{aligned}
&P^* \sup_{d(\theta,\theta_0)<\delta} |I_3'| 1_{\Omega_n} 1_{n \geq N_{\theta_0}} \quad \text{(continued)} \\
&\leq J_{[]}(1, \mathcal{M}_{Z,\delta}, L_2(P)) \sqrt{P^* \sup_{d(\theta,\theta_0)<\delta} (e^{Z(\theta)} - e^{Z(\theta_0)})^2} \frac{2}{B(\theta_0)} \\
&\lesssim \sqrt{Ee^{2Z(\theta_0)}} \sqrt{P^* \sup_{d(\theta,\theta_0)<\delta} (e^{Z(\theta)-Z(\theta_0)} - 1)^2} \frac{2}{B(\theta_0)} \\
&= \sqrt{Ee^{2Z(\theta_0)}} \sqrt{P^* \sup_{d(\theta,\theta_0)<\delta} (e^{W(\theta-\theta_0)} - 1)^2} \frac{2}{B(\theta_0)} \\
&= \sqrt{Ee^{2Z(\theta_0)}} \sqrt{P^* \sup_{|x|<\delta^2} (e^{W(x)} - 1)^2} \frac{2}{B(\theta_0)} \\
&\leq \sqrt{P^* \sup_{0 \leq x < \delta^2} (e^{B(x)} - 1)^2} \frac{4E(e^{Z(\theta_0)})}{B(\theta_0)} \quad \text{(by symmetry)},
\end{aligned}$$

where the second inequality holds by the independence of $Z(\theta) - Z(\theta_0)$ and $Z(\theta_0)$ according to Assumptions 2.2.1.

Since $f(x) = (e^x - 1)^2, (x \geq 0)$ is a convex function and $\{B(x) : x \geq 0\}$ is a martingale, $\{(e^{B(x)} - 1)^2\}_{x \geq 0}$ is a submartingale by Jensen's Inequality (and it is nonnegative as well). By Doob's maximal inequality,

$$\begin{aligned}
&\sqrt{P^* \sup_{0 \leq x < \delta^2} (e^{B(x)} - 1)^2} \frac{4\sqrt{Ee^{2Z(\theta_0)}}}{B(\theta_0)} \leq \sqrt{4E(e^{B(\delta^2)} - 1)^2} \frac{4\sqrt{Ee^{2Z(\theta_0)}}}{B(\theta_0)} \\
&= 8\sqrt{e^{2\delta^2} - 2e^{1/2\delta^2} + 1} \frac{\sqrt{Ee^{2Z(\theta_0)}}}{B(\theta_0)} \leq 16\delta \frac{\sqrt{Ee^{2Z(\theta_0)}}}{B(\theta_0)} \lesssim \delta, \quad \text{(for } \delta \text{ small)}.
\end{aligned}$$

And we deal with $I_4' 1_{\Omega_n} 1_{n > N(\theta_0)}$. By (A.2),

$$0 < \frac{1_{\Omega_n} 1_{n>N(\theta_0)}}{\tilde{S}^{(0)}(\theta, u)} \equiv \frac{1_{\Omega_n} 1_{n>N(\theta_0)}}{K_{\theta,u} S^{(0)}(\theta, u) + (1 - K_{\theta,u}) s^{(0)}(\theta, u)} \leq \frac{1}{\frac{1}{2} B(\theta_0)},$$

$$0 < \frac{1_{\Omega_n} 1_{n>N(\theta_0)}}{\tilde{S}^{(0)}(\theta_0, u)} \equiv \frac{1_{\Omega_n} 1_{n>N(\theta_0)}}{K_{\theta_0,u} S^{(0)}(\theta_0, u) + (1 - K_{\theta_0,u}) s^{(0)}(\theta_0, u)} 1_{\Omega_n} \leq \frac{1}{\frac{1}{2} B(\theta_0)};$$

and $\left(\frac{1}{\tilde{S}^{(0)}(\theta,u)} - \frac{1}{\tilde{S}^{(0)}(\theta_0,u)}\right) 1_{\Omega_n} 1_{n>N(\theta_0)} = \frac{\tilde{S}^{(0)}(\theta_0,u) - \tilde{S}^{(0)}(\theta,u)}{\tilde{S}^{(0)}(\theta,u)\tilde{S}^{(0)}(\theta_0,u)} 1_{\Omega_n} 1_{n>N(\theta_0)}$.

It follows that

$$\begin{aligned}
&\left|\frac{1}{\tilde{S}^{(0)}(\theta,u)} - \frac{1}{\tilde{S}^{(0)}(\theta_0,u)}\right| 1_{\Omega_n} 1_{n>N(\theta_0)} = \frac{\left|\tilde{S}^{(0)}(\theta_0,u) - \tilde{S}^{(0)}(\theta,u)\right|}{\tilde{S}^{(0)}(\theta,u)\tilde{S}^{(0)}(\theta_0,u) 1_{\Omega_n} 1_{n>N(\theta_0)}} \\
&\quad = \frac{|K_\theta(S_\theta - s_\theta) + s_\theta - K_{\theta_0}(S_{\theta_0} - s_{\theta_0}) - s_{\theta_0}|}{\tilde{S}^{(0)}(\theta,u)\tilde{S}^{(0)}(\theta_0,u) 1_{\Omega_n} 1_{n>N(\theta_0)}} \\
&\quad = \frac{|K_\theta(S_\theta - s_\theta) + s_\theta - K_{\theta_0}(S_{\theta_0} - s_{\theta_0}) - s_{\theta_0}|}{\frac{1}{4}B^2(\theta_0)} \\
&\quad = \frac{|K_{\theta_0}[(S_\theta - s_\theta) - (S_{\theta_0} - s_{\theta_0})] + s_\theta - s_{\theta_0} + (K_\theta - K_{\theta_0})(S_{\theta_0} - s_{\theta_0})|}{\frac{1}{4}B^2(\theta_0)} \\
&\quad \le \frac{4}{B^2(\theta_0)} \left(|(S_\theta - s_\theta) - (S_{\theta_0} - s_{\theta_0})| + |s_\theta - s_{\theta_0}| + 2|S_{\theta_0} - s_{\theta_0}|\right) \\
&\quad \lesssim |(S_\theta - s_\theta) - (S_{\theta_0} - s_{\theta_0})| + |s_\theta - s_{\theta_0}| + |S_{\theta_0} - s_{\theta_0}|.
\end{aligned}$$

Notice that $S_\theta, s_\theta, S_{\theta_0}$ and s_{θ_0} all depend on u. We omit u in their expressions to make the notation simpler in the displayed inequalities.

Therefore, $P^* \sup_{d(\theta,\theta_0)<\delta} |I_4'| 1_{\Omega_n} 1_{n>N(\theta_0)}$ can be controlled by summation of three terms.

$$\begin{aligned}
&P^* \sup_{d(\theta,\theta_0)<\delta} |I_4'| 1_{\Omega_n} 1_{n>N(\theta_0)} \\
&\quad \lesssim P^* \sup_{d(\theta,\theta_0)<\delta} \int_0^\tau |(S_\theta - s_\theta) - (S_{\theta_0} - s_{\theta_0})| \sqrt{n} \left|(\mathbb{P}_n - P)Y(u)e^{Z(\theta_0)}\right| \mathbb{P}_n dN(u) \\
&\quad + P^* \sup_{d(\theta,\theta_0)<\delta} (e^{|\theta-\theta_0|} - 1) \int_0^\tau s_{\theta_0} \cdot \sqrt{n} \left|(\mathbb{P}_n - P)Y(u)e^{Z(\theta_0)}\right| \mathbb{P}_n dN(u) \\
&\quad + E \int_0^\tau \sqrt{n} \left|(\mathbb{P}_n - P)Y(u)e^{Z(\theta_0)}\right|^2 \mathbb{P}_n dN(u) \\
&\equiv I_{4,a} + I_{4,b} + I_{4,c}.
\end{aligned}$$

We deal with each of them separately.

$$\begin{aligned}
I_{4,a} &\le P^* \left[\sup_{d(\theta,\theta_0)<\delta, u\in[0,\tau]} |(S_\theta - s_\theta) - (S_{\theta_0} - s_{\theta_0})| \cdot \sup_{u\in[0,\tau]} \sqrt{n} \left|(\mathbb{P}_n - P)Y(u)e^{Z(\theta_0)}\right|\right] \\
&\le \sqrt{P^* \sup_{d(\theta,\theta_0)<\delta, u\in[0,\tau]} |(S_\theta - s_\theta) - (S_{\theta_0} - s_{\theta_0})|^2} \sqrt{P^* \sup_{u\in[0,\tau]} n \left|(\mathbb{P}_n - P)Y(u)e^{Z(\theta_0)}\right|^2},
\end{aligned}$$

by Cauchy–Schwartz Inequality.

On one hand, by Theorem 2.14.2 and Theorem 2.14.5 (for $p = 2$) of van der Vaart and Wellner (1996),

$$\begin{aligned}
&\sqrt{P^* \sup_{d(\theta,\theta_0)<\delta, u\in[0,\tau]} |(S_\theta - s_\theta) - (S_{\theta_0} - s_{\theta_0})|^2} \\
&\leq \left(J_{[]}(1, \mathcal{M}_{Z,\delta}, L_2(P)) + 1\right) \sqrt{P^* \sup_{d(\theta,\theta_0)<\delta, u\in[0,\tau]} Y^2(u)(e^{Z(\theta)} - e^{Z(\theta_0)})^2} \\
&\leq \left(J_{[]}(1, \mathcal{M}_{Z,\delta}, L_2(P)) + 1\right) \sqrt{P^* \sup_{d(\theta,\theta_0)<\delta} (e^{Z(\theta)} - e^{Z(\theta_0)})^2} \lesssim \delta,
\end{aligned}$$

where the last inequality is obtained following the steps for $I_3 1_{\Omega_n} 1_{n\geq N_{\theta_0}}$ starting from (A.3);

on the other hand,

$$\begin{aligned}
&\sqrt{P^* \sup_{u\in[0,\tau]} n \left|(\mathbb{P}_n - P)Y(u)e^{Z(\theta_0)}\right|^2} = \sqrt{P^* \sup_{u\in[0,\tau]} \left|\mathbb{G}_n Y(u)e^{Z(\theta_0)}\right|^2} \\
&\leq J_{[]}(1, \mathcal{M}_Y, L_2(P)) \sqrt{P \sup_u \left[Y(u)e^{Z(\theta_0)}\right]^2} \lesssim \sqrt{Pe^{2Z(\theta_0)}} < \infty, \qquad \text{(A.7)}
\end{aligned}$$

where $\mathcal{M}_Y \equiv \{Y(u)e^{Z(\theta_0)} : u \in [0,\tau]\}$, the first inequality follows from Theorem 2.14.1 in VW [54], $J_{[]}(1, \mathcal{M}_Y, L_2(P)) < \infty$ follows from $\mathcal{M}_Y \subset \mathcal{M}$ and $J_{[]}(1, \mathcal{M}, L_2(P)) < \infty$ (by Lemma A.0.1), $Pe^{2Z(\theta_0)} < \infty$ follows from Assumptions 2.2.1.

Thus the product of them is bounded by a constant times δ, and hence $I_{4,a} \lesssim \delta$.

For the other two terms $I_{4,b}$ and $I_{4,c}$, we have

$$\begin{aligned}
I_{4,b} &= P^* \sup_{d(\theta,\theta_0)<\delta} (e^{|\theta-\theta_0|} - 1) \int_0^\tau s_{\theta_0} \cdot \sqrt{n} \left|(\mathbb{P}_n - P)Y(u)e^{Z(\theta_0)}\right| \mathbb{P}_n dN(u) \\
&\lesssim \delta^2 \cdot E \int_0^\tau s_{\theta_0} \cdot \sqrt{n} \left|(\mathbb{P}_n - P)Y(u)e^{Z(\theta_0)}\right| \mathbb{P}_n dN(u) \\
&\lesssim \delta^2 \cdot P(e^{Z(\theta_0)}) \cdot P(e^{W(\theta_M)}) \cdot \sqrt{P^* \sup_{u\in[0,\tau]} n \left|(\mathbb{P}_n - P)Y(u)e^{Z(\theta_0)}\right|^2} \lesssim \delta^2,
\end{aligned}$$

by Taylor expansion $e^{\delta^2} - 1 = \delta^2 + o(\delta^2)$ for small δ in the first inequality and the independence of $Z(\theta) - Z(\theta_0)$ and $Z(\theta_0)$ in the second inequality; the last inequality follows from $Pe^{2Z(\theta_0)} < \infty$ according to Assumptions 2.2.1, the property of 2-sided

Brownian motion $W(\cdot) \equiv Z(\theta) - Z(\theta_0)$ and $\sqrt{P^* \sup_{u\in[0,\tau]} n \left|(\mathbb{P}_n - P)Y(u)e^{Z(\theta_0)}\right|^2} < \infty$ from (A.7).

$$
\begin{aligned}
I_{4,c} &= E\int_0^\tau \sqrt{n}\left|(\mathbb{P}_n - P)Y(u)e^{Z(\theta_0)}\right|^2 \mathbb{P}_n dN(u) \\
&\leq \frac{1}{\sqrt{n}}\left[P^* \sup_{u\in[0,\tau]} n\left|(\mathbb{P}_n - P)Y(u)e^{Z(\theta_0)}\right|^2\right] \lesssim \frac{1}{\sqrt{n}},
\end{aligned}
$$

where we used $\sqrt{P^* \sup_{u\in[0,\tau]} n \left|(\mathbb{P}_n - P)Y(u)e^{Z(\theta_0)}\right|^2} < \infty$ from (A.7).

Summing them up, we have

$$
P^* \sup_{d(\theta,\theta_0)<\delta} |I_4^{'}| 1_{\Omega_n} 1_{n\geq N_{\theta_0}} \lesssim \delta + \delta^2 + \frac{1}{\sqrt{n}}.
$$

So up to now, we have obtained that on Ω_n with $P^*(\Omega_n) \to 1$,

$$
\begin{aligned}
& P^* \sup_{d(\theta,\theta_0)<\delta} |I_1 - I_2 - I_3 + I_4| 1_{\Omega_n} 1_{n\geq N_{\theta_0}} \\
&= P^* \sup_{d(\theta,\theta_0)<\delta} |I_1 - I_2 - I_3^{'} + I_4^{'}| 1_{\Omega_n} 1_{n\geq N_{\theta_0}} \\
&\lesssim \delta + \delta^2 + \frac{1}{\sqrt{n}} \lesssim \delta + \frac{1}{\sqrt{n}} = \phi_n(\delta), \quad \text{for small} \quad \delta.
\end{aligned}
$$

Since we have proved $\hat{\theta}_n \to_{P^*} \theta_0$ in Chapter 2.3.3, then solve $r_n^2 \phi_n(\frac{1}{r_n}) \leq \sqrt{n}$, get $r_n = \sqrt{n}$.

$$
\sqrt{n} d(\hat{\theta}_n, \theta_0) = O_P^*(1) \Rightarrow n(\hat{\theta}_n - \theta_0) = O_P^*(1).
$$

We get an upper bound for the rate of converge of $\hat{\theta}_n$, $\tilde{r}_n = n$.

Appendix B

Additional proof details for Chapter 2.3

As mentioned in Chapter 2.3.4.2 and Chapter 2.3.4.3, we put the lengthy proof details here.

B.1 Proof of the convergence of I_3 (in Chapter 2.3.4.2) to zero

If we can obtain Lipschitz property of g_π, it is easy to bound $P^* \sup_{\pi \in \Phi} |\mathbb{G}_n f_\pi|$. Take the difference of f_π at π_1 and π_2,

$$\begin{aligned} & g_{\pi_1} - g_{\pi_2} \\ &= 1_{(T \le C)} 1_{(0 < T \le \tau)} \log \frac{s^{(0)}(\pi_1, T)}{s^{(0)}(\pi_2, T)} - e^{\beta_0 Z_H(\theta_0) + \gamma_0 X} \int_0^\tau 1_{(T \ge u)} \lambda_0(u) \log \frac{s^{(0)}(\pi_1, u)}{s^{(0)}(\pi_2, u)} du. \end{aligned} \tag{B.1}$$

Since $\forall \pi, \pi_0 \in \Phi$,

$$\begin{aligned} \log \frac{s^{(0)}(\pi, u)}{s^{(0)}(\pi_0, u)} &= -\frac{1}{2}\beta^2 |\theta - \theta_0|^{2H} PN(\tau) \\ & - P\left[\int_0^\tau \log \frac{P \exp(\beta Z_H(\theta_0) + \gamma X - e^{\beta_0 Z_H(\theta_0) + \gamma_0 X} \int_0^u \lambda_0(s) ds)}{P \exp(\beta_0 Z_H(\theta_0) + \gamma_0 X - e^{\beta_0 Z_H(\theta_0) + \gamma_0 X} \int_0^u \lambda_0(s) ds)} dN(u) \right], \end{aligned}$$

$$\begin{aligned}
&\log \frac{s^{(0)}(\pi_1,u)}{s^{(0)}(\pi_2,u)} \\
&= -\frac{1}{2}\left(\beta_1^2|\theta_1-\theta_0|^{2H} - \beta_2^2|\theta_2-\theta_0|^{2H}\right) PN(\tau) \\
&\quad - P\left[\int_0^\tau \log \frac{P\exp(\beta_1 Z_H(\theta_0)+\gamma_1 X - e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\int_0^u \lambda_0(s)ds)}{P\exp(\beta_2 Z_H(\theta_0)+\gamma_2 X - e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\int_0^u \lambda_0(s)ds)} dN(u)\right] \\
&= -\frac{1}{2}\left(\beta_1^2-\beta_2^2\right)|\theta_1-\theta_0|^{2H} PN(\tau) - \frac{1}{2}\left(|\theta_1-\theta_0|^{2H} - |\theta_2-\theta_0|^{2H}\right)\beta_2^2 PN(\tau) \\
&\quad - P\left[\int_0^\tau \log \frac{P\exp(\beta_1 Z_H(\theta_0)+\gamma_1 X - e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\int_0^u \lambda_0(s)ds)}{P\exp(\beta_2 Z_H(\theta_0)+\gamma_2 X - e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\int_0^u \lambda_0(s)ds)} dN(u)\right].
\end{aligned}$$

We will deal with the three terms in the decomposition of $\log \frac{s^{(0)}(\pi_1,u)}{s^{(0)}(\pi_2,u)}$ separately.

For the first term, $\left|\beta_1^2-\beta_2^2\right| |\theta_1-\theta_0|^{2H} PN(\tau) \le 2\beta_M \theta_M^{2H}\cdot|\beta_1-\beta_2|$, by $|\beta_1-\beta_2| \le 2\beta_M, |\theta_1-\theta_0| \le \theta_M$ and $PN(\tau)\le 1$.

For the second term, consider

$$\begin{aligned}
&\frac{|\theta_1-\theta_0|^{2H} - |\theta_2-\theta_0|^{2H}}{|\theta_1-\theta_2|^H} \\
&\quad = \frac{|\theta_1-\theta_0|^{2H} - |\theta_2-\theta_0|^{2H}}{|\theta_1-\theta_0|^{H} - |\theta_2-\theta_0|^{H}} \cdot \frac{|\theta_1-\theta_0|^{H} - |\theta_2-\theta_0|^{H}}{|\theta_1-\theta_2|^H} \\
&\quad = \left(|\theta_1-\theta_0|^{H} + |\theta_2-\theta_0|^{H}\right)\cdot \frac{|\theta_1-\theta_0|^{H} - |\theta_2-\theta_0|^{H}}{|\theta_1-\theta_2|^H}.
\end{aligned} \tag{B.2}$$

We deal with $\frac{|\theta_1-\theta_0|^{H} - |\theta_2-\theta_0|^{H}}{|\theta_1-\theta_2|^H}$ first. Since function $k(x) = x^H (0 < H < 1)$ is concave on $[0,\infty)$ and $k(0)=0$, it is easy to show $\frac{k(a)-k(0)}{a-0} \ge \frac{k(a+b)-k(b)}{(a+b)-b}$ for $\forall a > 0, b \ge 0$. then $k(a)+k(b) \ge k(a+b)+k(0) = k(a+b)$. Take $a = |\theta_2-\theta_0|$ and $b = |\theta_1-\theta_2|$, we have

$$k(|\theta_2-\theta_0|) + k(|\theta_1-\theta_2|) \ge k(|\theta_2-\theta_0| + |\theta_1-\theta_2|). \tag{B.3}$$

By monotonicity of $k(x)$ and Triangle Inequality $|\theta_2-\theta_0| + |\theta_1-\theta_2| \ge |\theta_1-\theta_0|$, we have

$$k(|\theta_2-\theta_0| + |\theta_1-\theta_2|) \ge k(|\theta_1-\theta_0|). \tag{B.4}$$

Then by (B.3) and (B.4),

$$k(|\theta_2 - \theta_0|) + k(|\theta_1 - \theta_2|) \geq k(|\theta_1 - \theta_0|). \tag{B.5}$$

By symmetry between θ_1 and θ_2,

$$k(|\theta_1 - \theta_0|) + k(|\theta_1 - \theta_2|) \geq k(|\theta_2 - \theta_0|). \tag{B.6}$$

By (B.5) and (B.6), we have

$$k(|\theta_1 - \theta_2|) \geq k(|\theta_1 - \theta_0|) - k(|\theta_2 - \theta_0|) \quad \text{and} \quad k(|\theta_1 - \theta_2|) \geq k(|\theta_2 - \theta_0|) - k(|\theta_1 - \theta_0|).$$

It follows that $$|k(|\theta_1 - \theta_0|) - k(|\theta_2 - \theta_0|)| \leq k(|\theta_1 - \theta_2|).$$

Hence $$\frac{\big||\theta_1 - \theta_0|^H - |\theta_2 - \theta_0|^H\big|}{|\theta_1 - \theta_2|^H} \leq 1.$$

Furthermore, $|\theta_1 - \theta_0|^H + |\theta_2 - \theta_0|^H \leq 2\theta_M^H$ by $\theta, \theta_1, \theta_2 \in [0, \theta_M]$. Then from (B.2), we have

$$\big||\theta_1 - \theta_0|^{2H} - |\theta_2 - \theta_0|^{2H}\big| \leq 2\theta_M^H |\theta_1 - \theta_2|^H.$$

Use this inequality for the second term in the decomposition of $\log \frac{s^{(0)}(\pi_1, u)}{s^{(0)}(\pi_2, u)}$ and notice $\beta_2^2 \leq \beta_M^2$, it follows that

$$\big||\theta_1 - \theta_0|^{2H} - |\theta_2 - \theta_0|^{2H}\big| \beta_2^2 PN(\tau) \leq 2\beta_M^2 \theta_M^H |\theta_1 - \theta_2|^H.$$

Note: we can't get $\frac{\big||\theta_1 - \theta_0|^{2H} - |\theta_2 - \theta_0|^{2H}\big|}{|\theta_1 - \theta_2|^{2H}} \leq 1$ in the same way as what we did for $\frac{\big||\theta_1 - \theta_0|^H - |\theta_2 - \theta_0|^H\big|}{|\theta_1 - \theta_2|^H} \leq 1$ since $k^2(x) = x^{2H} (0 < H < 1)$ is not guaranteed a concave function.

To deal with the third term in the decomposition of $\log \frac{s^{(0)}(\pi_1, u)}{s^{(0)}(\pi_2, u)}$, consider

$$\begin{aligned} f(\beta, \gamma, u) &\equiv \log PF(\beta, \gamma, u) \\ &= \log P \exp(\beta Z_H(\theta_0) + \gamma X - e^{\beta_0 Z_H(\theta_0) + \gamma_0 X} \int_0^u \lambda_0(s) ds). \end{aligned}$$

By the same way as that of Chapter 2.3.4.1, we have

$$\frac{\partial^2 f}{\partial^2 \beta} > 0, \quad \frac{\partial^2 f}{\partial^2 \gamma} > 0, \quad \left(\frac{\partial^2 f}{\partial \beta \partial \gamma}\right)^2 - \frac{\partial^2 f}{\partial^2 \beta} \cdot \frac{\partial^2 f}{\partial^2 \gamma} < 0.$$

Now we verify the Lipschitz condition for $f(\beta, \gamma, u)$.

$$|f(\beta_1, \gamma_1, u) - f(\beta_2, \gamma_2, u)| \leq |f(\beta_1, \gamma_1, u) - f(\beta_2, \gamma_1, u)| + |f(\beta_2, \gamma_1, u) - f(\beta_2, \gamma_2, u)|.$$

For the first part, we have

$$\begin{aligned} |f(\beta_1, \gamma_1, u) - f(\beta_2, \gamma_1, u)| &\leq \sup_{|\beta|=\beta_M, \gamma=\gamma_1} \left|\frac{\partial f}{\partial \beta}\right| |\beta_1 - \beta_2| \\ &\leq \sup_{|\beta|=\beta_M, |\gamma|\leq\gamma_M, 0\leq u\leq\tau} \frac{|P(Z_H(\theta_0)F(\beta,\gamma,u))|}{PF(\beta,\gamma,u)} |\beta_1 - \beta_2|, \end{aligned}$$

where the first inequality holds by the monotonic $\frac{\partial^2 f}{\partial^2 \beta} > 0$).

Similarly for the second part,

$$|f(\beta_2, \gamma_1, u) - f(\beta_2, \gamma_2, u)| \leq \sup_{|\gamma|=\gamma_M, |\beta|\leq\beta_M, 0\leq u\leq\tau} \frac{|P(XF(\beta,\gamma,u))|}{PF(\beta,\gamma,u)} |\gamma_1 - \gamma_2|.$$

Summing three terms of the decomposition of $\log \frac{s^{(0)}(\pi_1,u)}{s^{(0)}(\pi_2,u)}$, we have

$$\begin{aligned} \left|\log \frac{s^{(0)}(\pi_1, u)}{s^{(0)}(\pi_2, u)}\right| \leq & \beta_M \theta_M^{2H} |\beta_1 - \beta_2| + \beta_M^2 \theta_M^H |\theta_1 - \theta_2|^H \\ & + \sup_{|\beta|=\beta_M, |\gamma|\leq\gamma_M, 0\leq u\leq\tau} \frac{|P(Z_H(\theta_0)F(\beta,\gamma,u))|}{PF(\beta,\gamma,u)} |\beta_1 - \beta_2| \\ & + \sup_{|\gamma|=\gamma_M, |\beta|\leq\beta_M, 0\leq u\leq\tau} \frac{|P(XF(\beta,\gamma,u))|}{PF(\beta,\gamma,u)} |\gamma_1 - \gamma_2|. \end{aligned}$$

Then we evaluate (B.1),

$$\begin{aligned} & |g_{\pi_1} - g_{\pi_2}| \\ \leq & \left|1_{(T\leq C)} 1_{(0<T\leq\tau)} \log \frac{s^{(0)}(\pi_1, T)}{s^{(0)}(\pi_2, T)}\right| + \left|e^{\beta_0 Z_H(\theta_0)+\gamma_0 X} \int_0^\tau 1_{(T\geq u)} \lambda_0(u) \log \frac{s^{(0)}(\pi_1, u)}{s^{(0)}(\pi_2, u)} du\right| \\ \leq & \left[1 + e^{\beta_0 Z_H(\theta_0)+\gamma_0 X} \Lambda_0(\tau)\right] \cdot \\ & \left(\beta_M \theta_M^{2H} |\beta_1 - \beta_2| + \sup_{|\beta|=\beta_M, |\gamma|\leq\gamma_M, 0\leq u\leq\tau} \frac{|P(Z_H(\theta_0)F(\beta,\gamma,u))|}{PF(\beta,\gamma,u)} |\beta_1 - \beta_2| \right. \\ & \left. + \beta_M^2 \theta_M^H |\theta_1 - \theta_2|^H + \sup_{|\gamma|=\gamma_M, |\beta|\leq\beta_M, 0\leq u\leq\tau} \frac{|P(XF(\beta,\gamma,u))|}{PF(\beta,\gamma,u)} |\gamma_1 - \gamma_2| \right), \end{aligned}$$

where $\Lambda_0(\tau) \equiv \int_0^\tau \lambda_0(u)du$.

The leading term

$$P\left[1 + e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\Lambda_0(\tau)\right]^2 \lesssim 1 + \Lambda_0^2(\tau)P\left[e^{2(\beta_0 Z_H(\theta_0)+\gamma_0 X)}\right] < \infty \tag{B.7}$$

by the conditions from Assumptions 2.3.1.

Then to obtain the Lipschitz property of g_π, it remains to prove that

$$\sup_{|\beta|=\beta_M, |\gamma|\le\gamma_M, 0\le u\le\tau} \frac{|P(Z_H(\theta_0)F(\beta,\gamma,u))|}{PF(\beta,\gamma,u)}|\beta_1 - \beta_2| < \infty \tag{B.8}$$

and
$$\sup_{|\gamma|=\gamma_M, |\beta|\le\beta_M, 0\le u\le\tau} \frac{|P(XF(\beta,\gamma,u))|}{PF(\beta,\gamma,u)}|\gamma_1 - \gamma_2| < \infty. \tag{B.9}$$

It is easy to show that $\frac{P(Z_H(\theta_0)F(\beta,\gamma,u))}{PF(\beta,\gamma,u)}$ and $\frac{P(XF(\beta,\gamma,u))}{PF(\beta,\gamma,u)}$ are both continuous functions of (β,γ,u), so are $\frac{|P(Z_H(\theta_0)F(\beta,\gamma,u))|}{PF(\beta,\gamma,u)}$ and $\frac{|P(XF(\beta,\gamma,u))|}{PF(\beta,\gamma,u)}$. Their supremums on $\{(\beta,\gamma,u) : |\beta| = \beta_M, |\gamma| \le \gamma_M, 0 \le u \le \tau\}$ and $\{(\beta,\gamma,u) : |\beta| \le \beta_M, |\gamma| = \gamma_M, 0 \le u \le \tau\}$ (which are both closed and bounded sets) respectively are both achieved. Hence (B.8) and (B.9) are proved.

Now it follows that g_π is "Lipschitz in parameter" and hence $J_{[]}(1, \mathcal{M}_{g,H}, L_2(P)) < \infty$ (p.294, VW [54]), where $\mathcal{M}_{g,H} = \{g_\pi, \pi \in \Phi\}$.

Now we prove the envelope function of $\mathcal{M}_{g,H}$, $\sup_{\pi\in\Phi} g_\pi^2$, has finite second moment.

By Lipschitz property of the function g_π, $\sup_{\pi\in\Phi} g_\pi \le |g_{\pi_0}| + L_g \cdot d(\pi,\pi_0)$, where $d(\pi,\pi_0) = \sqrt{|\beta-\beta_0|^2 + |\gamma-\gamma_0|^2 + |\theta-\theta_0|^{2H}}$, $L_g = \left[1 + e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\int_0^\tau \lambda_0(u)du\right]$ $\cdot\Big(\beta_M\theta_M^{2H}|\beta_1-\beta_2| + \sup_{|\beta|=\beta_M,|\gamma|\le\gamma_M,0\le u\le\tau}\frac{|P(Z_H(\theta_0)F(\beta,\gamma,u))|}{PF(\beta,\gamma,u)}|\beta_1-\beta_2|$ $+ \beta_M^2\theta_M^H|\theta_1-\theta_2|^H + \sup_{|\gamma|=\gamma_M,|\beta|\le\beta_M,0\le u\le\tau}\frac{|P(XF(\beta,\gamma,u))|}{PF(\beta,\gamma,u)}|\gamma_1-\gamma_2|\Big)$.

Hence
$$P\sup_{\pi\in\Phi} g_\pi^2 \lesssim \sup_{\pi\in\Phi} d^2(\pi,\pi_0)\cdot P[L_g^2] + Pg_{\pi_0}^2. \tag{B.10}$$

On one hand,

$$\sup_{\pi\in\Phi} d^2(\pi,\pi_0)\cdot P[L_g^2] \le \left(\theta_M^{2H} + 4\beta_M^2 + 4\gamma_M^2\right)P[L_g^2] < \infty$$

where $P[L_g^2] < \infty$ follows from (B.7), (B.8) and (B.9);

on the other hand,

$$\begin{aligned}
Pg_{\pi_0}^2 &\le \sup_u \left|\log s^0(\pi_0, u)\right|^2 \cdot P\left(1 + \Lambda_0(\tau)e^{Z_H(\theta_0)}\right)^2 \\
&\le 2\sup_u \left|\log P\left(Y(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\right)\right|^2 \cdot P\left(1 + \Lambda_0^2(\tau)e^{2Z_H(\theta_0)}\right) \\
&\le 2\left[\left|\log Pe^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\right|^2 + \left|\log PY(\tau)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\right|^2\right] P\left(1 + \Lambda_0^2(\tau)e^{2Z_H(\theta_0)}\right) \\
&< \infty,
\end{aligned}$$

where we used $Y(\tau) \le Y(u) \le 1$ in the third inequality and $0 < P(e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}) < \infty$, $0 < P(Y(\tau)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}) < \infty$ which follow from Assumptions 2.3.1 in the last inequality.

Since both $P\sup_{\pi\in\Phi} d^2(\pi, \pi_0) \cdot P[L_g^2]$ and $Pg_{\pi_0}^2$ are finite, by (B.10),

$$P\sup_{\pi\in\Phi} g_\pi^2 < \infty.$$

By Theorem 2.14.2 of VW [54],

$$\frac{3}{\sqrt{n}\epsilon}P^*\sup_{\pi\in\Phi}|\mathbb{G}_n g_\pi| \le \frac{3}{\sqrt{n}\epsilon}J_{[]}(1, \mathcal{M}_{g,H}, L^2(P))\sqrt{P\sup_{\pi\in\Phi} g_\pi^2} \lesssim \frac{1}{\sqrt{n}\epsilon}$$

It follows that, as $n \to \infty$, $I_3 \to 0$.

B.2 Proof of the Rate of Convergence (in Chapter 2.3.4.3)

To obtain the convergence rates of $\hat{\pi}_n = (\hat{\beta}_n, \hat{\gamma}_n, \hat{\theta}_n)$, the next step is to check the modulus of continuity.

$$
\begin{aligned}
& \sqrt{n}\left[(\mathbb{M}_n - \mathbb{M})(\pi) - (\mathbb{M}_n - \mathbb{M})(\pi_0)\right] \\
= \; & \sqrt{n}\mathbb{P}_n\left[(\beta Z_H(\theta) + \gamma X)N(\tau) - \int_0^\tau \log\left[\mathbb{P}_n Y(u) e^{\beta Z_H(\theta)+\gamma X}\right] dN(u)\right] \\
& \quad - \sqrt{n}P\left[(\beta Z_H(\theta) + \gamma X)N(\tau) - \int_0^\tau \log s^{(0)}(\pi, u) dN(u)\right] \\
& - \sqrt{n}\mathbb{P}_n\left[(\beta_0 Z_H(\theta_0) + \gamma_0 X)N(\tau) - \int_0^\tau \log\left[\mathbb{P}_n Y(u) e^{(\beta_0 Z_H(\theta_0)+\gamma_0 X)}\right] dN(u)\right] \\
& \quad + \sqrt{n}P\left[(\beta_0 Z_H(\theta_0) + \gamma_0 X)N(\tau) - \int_0^\tau \log s^{(0)}(\pi_0, u) dN(u)\right] \\
= \; & \sqrt{n}(\mathbb{P}_n - P)\left[\left((\beta Z_H(\theta) + \gamma X) - (\beta_0 Z_H(\theta_0) + \gamma_0 X)\right) N(\tau)\right] \\
& \quad + \sqrt{n}\int_0^\tau \left[\log s^{(0)}(\pi, u) - \log s^{(0)}(\pi_0, u)\right] P dN(u) \\
& \quad - \sqrt{n}\int_0^\tau [\log \mathbb{P}_n Y(u) e^{\beta Z_H(\theta)+\gamma X} - \log \mathbb{P}_n Y(u) e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}] \mathbb{P}_n dN(u) \\
= \; & \sqrt{n}(\mathbb{P}_n - P)\left[\left((\beta Z_H(\theta) + \gamma X) - (\beta_0 Z_H(\theta_0) + \gamma_0 X)\right) N(\tau)\right] \\
& \quad + \sqrt{n}\int_0^\tau \left[\log s^{(0)}(\pi, u) - \log s^{(0)}(\pi_0, u)\right] (\mathbb{P}_n - P) dN(u) \\
& \quad - \sqrt{n}\int_0^\tau \left[\log s^{(0)}(\pi, u) - \log s^{(0)}(\pi_0, u)\right] \mathbb{P}_n dN(u) \\
& \quad + \sqrt{n}\int_0^\tau [\log \mathbb{P}_n Y(u) e^{\beta Z_H(\theta)+\gamma X} - \log \mathbb{P}_n Y(u) e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}] \mathbb{P}_n dN(u) \\
= \; & \sqrt{n}(\mathbb{P}_n - P)\left[\left((\beta Z_H(\theta) + \gamma X) - (\beta_0 Z_H(\theta_0) + \gamma_0 X)\right) N(\tau)\right] \\
& \quad + \sqrt{n}\int_0^\tau \left[\log s^{(0)}(\pi, u) - \log s^{(0)}(\pi_0, u)\right] (\mathbb{P}_n - P) dN(u) \\
& \quad + \sqrt{n}\int_0^\tau \left[\log \frac{\mathbb{P}_n Y(u) e^{\beta Z_H(\theta)+\gamma X}}{\mathbb{P}_n Y(u) e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}} \frac{s^{(0)}(\pi_0, u)}{s^{(0)}(\pi, u)}\right] \mathbb{P}_n dN(u) \\
\equiv \; & I_4 - I_5 - I_6,
\end{aligned}
$$

To bound $P^* \sup_{d(\pi,\pi_0)<\delta} \sqrt{n}\,|(\mathbb{M}_n - \mathbb{M})(\pi) - (\mathbb{M}_n - \mathbb{M})(\pi_0)|$, it suffices to bound the supremum's outer expectations separately for $|I_4|, |I_5|$, and $|I_6|$.

We start with $P^* \sup_{d(\pi,\pi_0)<\delta} |I_4|$.

$$
\begin{aligned}
& P^* \sup_{d(\pi,\pi_0)<\delta} \left|\sqrt{n}(\mathbb{P}_n - P)\left[\left((\beta Z_H(\theta) + \gamma X) - (\beta_0 Z_H(\theta_0) + \gamma_0 X)\right) N(\tau)\right]\right| \\
=& P^* \sup_{d(\pi,\pi_0)<\delta} \left|\mathbb{G}_n\left[\left((\beta(Z_H(\theta) - Z_H(\theta_0)) + (\beta - \beta_0) Z_H(\theta_0) + (\gamma - \gamma_0) X\right) N(\tau)\right]\right| \\
\leq& |\beta| P^* \sup_{d(\pi,\pi_0)<\delta} |\mathbb{G}_n W_H(\theta - \theta_0) N(\tau)| + P^* \sup_{d(\pi,\pi_0)<\delta} |(\gamma - \gamma_0)\mathbb{G}_n X N(\tau)| \\
& + P^* \sup_{d(\pi,\pi_0)<\delta} |(\beta - \beta_0)\mathbb{G}_n Z_H(\theta_0) N(\tau)|\,.
\end{aligned}
$$

The first term

$$P^* \sup_{|\theta-\theta_0|^{2H}<\delta} |\mathbb{G}_n W_H(\theta - \theta_0) N(\tau)| \leq 2P^* \sup_{o\leq x<\delta^{1/2H}} |\mathbb{G}_n W_H(x) N(\tau)|\,.$$

Since $W_H(x)$ is fBm starting from 0 with Hurst parameter H, by Lemma 8.1 in McKeague and Sen [29] and similar argument as that in Chapter 2.2.2.2,

$$\lesssim P^* \sup_{o\leq x<\delta^{1/2H}} W^2(x) N^2(1) \lesssim 2P^* \sup_{o\leq x<\delta^{1/2H}} W^2(x) \lesssim \left(\delta^{1/2H}\right)^{2H} = \delta,$$

where the last inequality follows from Theorem 1.1 in Novikov and Valkeila [31].

The second term

$$P^* \sup_{d(\pi,\pi_0)<\delta} |(\beta - \beta_0)\mathbb{G}_n Z_H(\theta_0) N(\tau)| \leq \delta P\, |\mathbb{G}_n Z_H(\theta_0) N(\tau)| \lesssim \delta.$$

The third term

$$P^* \sup_{d(\pi,\pi_0)<\delta} |(\gamma - \gamma_0)\mathbb{G}_n X N(\tau)| \leq \delta P\, |\mathbb{G}_n X N(\tau)| \lesssim \delta.$$

So summing up the three terms, we obtain $$P^* \sup_{d(\pi,\pi_0)<\delta} |I_4| \lesssim \delta.$$

$$
\begin{aligned}
\text{Consider} \quad & P^* \sup_{d(\pi,\pi_0)<\delta} |I_5| \\
=& P^* \sup_{d(\pi,\pi_0)<\delta} \left| \sqrt{n} \int_0^\tau \left[\log s^{(0)}(\pi,u) - \log s^{(0)}(\pi_0,u) \right] (\mathbb{P}_n - P) dN(u) \right| \\
=& \sup_{d(\pi,\pi_0)<\delta} \left| \log s^{(0)}(\pi,u) - \log s^{(0)}(\pi_0,u) \right| \cdot |P \mathbb{G}_n N(\tau)| \\
\lesssim& \sup_{d(\pi,\pi_0)<\delta} \left| \log s^{(0)}(\pi,u) - \log s^{(0)}(\pi_0,u) \right| \cdot \sqrt{P \left[\mathbb{G}_n N(\tau) \right]^2} \\
\lesssim& \sup_{d(\pi,\pi_0)<\delta} \left| \log s^{(0)}(\pi,u) - \log s^{(0)}(\pi_0,u) \right|,
\end{aligned}
$$

where the second equality holds since $\sup_{d(\pi,\pi_0)<\delta} \left| \log s^{(0)}(\pi,u) - \log s^{(0)}(\pi_0,u) \right|$ is a deterministic function and the last inequality holds by

$$
\sqrt{P \left[\mathbb{G}_n N(\tau) \right]^2} = \sqrt{P(N(\tau) - PN(\tau))^2} \le 1.
$$

We control $\left| \log s^{(0)}(\pi,u) - \log s^{(0)}(\pi_0,u) \right|$ by decomposing it into three parts.

$$
\begin{aligned}
& \left| \log s^{(0)}(\pi,u) - \log s^{(0)}(\pi_0,u) \right| \\
& \quad \le \left| \log PY(u) e^{\beta Z_H(\theta) + \gamma X} - \log PY(u) e^{\beta Z_H(\theta_0) + \gamma X} \right| \\
& \qquad + \left| \log PY(u) e^{\beta Z_H(\theta_0) + \gamma X} - \log PY(u) e^{\beta_0 Z_H(\theta_0) + \gamma X} \right| \\
& \qquad + \left| \log PY(u) e^{\beta_0 Z_H(\theta_0) + \gamma X} - \log PY(u) e^{\beta_0 Z_H(\theta_0) + \gamma_0 X} \right|.
\end{aligned}
$$

The first term $\left| \log PY(u) e^{\beta Z_H(\theta) + \gamma X} - \log PY(u) e^{\beta Z_H(\theta_0) + \gamma X} \right| = 1/2 |\theta - \theta_0|^{2H}$ by (2.11).

The second term

$$
\begin{aligned}
& \left| \log PY(u) e^{\beta Z_H(\theta_0) + \gamma X} - \log PY(u) e^{\beta_0 Z_H(\theta_0) + \gamma X} \right| \\
\le & \sup_{|\beta - \beta_0| \le \delta, |\gamma - \gamma_0| \le \delta} \left| \frac{\partial}{\partial \beta} \log P e^{\beta Z_H(\theta_0) + \gamma X} \right| |\beta - \beta_0| \\
\le & \sup_{|\beta| \le \beta_M, |\gamma| \le \gamma_M} \left| \frac{P \left[Z_H(\theta_0) e^{\beta Z_H(\theta_0) + \gamma X} \right]}{P e^{\beta Z_H(\theta_0) + \gamma X}} \right| |\beta - \beta_0|,
\end{aligned}
$$

where $\sup_{|\beta|\le\beta_M,|\gamma|\le\gamma_M}\left|\frac{P\left[Z_H(\theta_0)e^{\beta Z_H(\theta_0)+\gamma X}\right]}{Pe^{\beta Z_H(\theta_0)+\gamma X}}\right| < \infty$ since the supremum of a continuous function in a closed and bounded region is always achieved. Similar arguments apply to the third term,

$$\begin{aligned}&\left|\log PY(u)e^{\beta_0 Z_H(\theta_0)+\gamma X} - \log PY(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\right|\\ \le& \sup_{|\beta|\le\beta_M,|\gamma|\le\gamma_M}\left|\frac{P\left[Z_H(\theta_0)e^{\beta Z_H(\theta_0)+\gamma X}\right]}{Pe^{\beta Z_H(\theta_0)+\gamma X}}\right|\left|\gamma-\gamma_0\right|.\end{aligned}$$

To sum up these three terms, we get

$$\begin{aligned}\sup_{d(\pi,\pi_0)<\delta}\left|\log s^{(0)}(\pi,u) - \log s^{(0)}(\pi_0,u)\right| \le& \sup_{d(\pi,\pi_0)<\delta} 1/2|\theta-\theta_0|^{2H}\\ &+ \sup_{|\beta|\le\beta_M,|\gamma|\le\gamma_M}\left|\frac{P\left[Z_H(\theta_0)e^{\beta Z_H(\theta_0)+\gamma X}\right]}{Pe^{\beta Z_H(\theta_0)+\gamma X}}\right|\cdot\sup_{d(\pi,\pi_0)<\delta}|\beta-\beta_0|\\ &+ \sup_{|\beta|\le\beta_M,|\gamma|\le\gamma_M}\left|\frac{P\left[Z_H(\theta_0)e^{\beta Z_H(\theta_0)+\gamma X}\right]}{Pe^{\beta Z_H(\theta_0)+\gamma X}}\right|\cdot\sup_{d(\pi,\pi_0)<\delta}\left|\gamma-\gamma_0\right|.\\ &\lesssim \delta^2+\delta\lesssim\delta, \qquad \text{for small } \delta.\end{aligned}$$

Hence $$P^*\sup_{d(\pi,\pi_0)<\delta}|I_5|\lesssim\delta \quad \text{for small } \delta.$$

We control I_6 by the following way.

$$\begin{aligned}\sup_{d(\pi,\pi_0)<\delta}|I_6| =& \sup_{\pi_\delta}\left|\sqrt{n}\int_0^\tau\left[\log\frac{\mathbb{P}_nY(u)e^{\beta Z_H(\theta)+\gamma X}}{\mathbb{P}_nY(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}}\frac{s^{(0)}(\pi_0,u)}{s^{(0)}(\pi,u)}\right]\mathbb{P}_n dN(u)\right|\\ \le& \sup_{\pi_\delta}\sqrt{n}\int_0^\tau\left|\log\frac{\mathbb{P}_nY(u)e^{\beta Z_H(\theta)+\gamma X}}{PY(u)e^{\beta Z_H(\theta)+\gamma X}} - \log\frac{\mathbb{P}_nY(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}}{PY(u)e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}}\right|\mathbb{P}_n dN(u)\\ \equiv& \sup_{\pi_\delta}\sqrt{n}\int_0^\tau\left|\log\frac{\mathbb{P}_nU(\beta,\theta,\gamma,u)}{PU(\beta,\theta,\gamma,u)} - \log\frac{\mathbb{P}_nU(\beta_0,\theta_0,\gamma_0,u)}{PU(\beta_0,\theta_0,\gamma_0,u)}\right|\mathbb{P}_n dN(u),\end{aligned}$$

if we denote $U(\beta,\theta,\gamma,u)\equiv Y(u)e^{\beta Z_H(\theta)+\gamma X}$ and $\pi_\delta\equiv\{\pi : d(\pi,\pi_0)<\delta\}$ in this section.

Since for any continuously differentiable function $g(x), x\in[a,b]$, $|g(x_1)-g(x_2)|\le$

$\sup_{x\in[a,b]}|g'(x)|\cdot|x_1-x_2|$, take $g(x)=\log x$, then

$$\begin{aligned}
&\sup_{d(\pi,\pi_0)<\delta}|I_6| \\
&\le \int_0^\tau \sqrt{n}\sup_{\pi_\delta}\frac{PU(\beta,\theta,\gamma,u)}{\mathbb{P}_nU(\beta,\theta,\gamma,u)}\cdot\sup_{\pi_\delta}\left|\frac{\mathbb{P}_nU(\beta,\theta,\gamma,u)}{PU(\beta,\theta,\gamma,u)}-\frac{\mathbb{P}_nU(\beta_0,\theta_0,\gamma_0,u)}{PU(\beta_0,\theta_0,\gamma_0,u)}\right|\mathbb{P}_n dN(u) \\
&\le \sup_{\pi_\delta,u\in[0,\tau]}\frac{PU(\beta,\theta,\gamma,u)}{\mathbb{P}_nU(\beta,\theta,\gamma,u)}\cdot\sqrt{n}\sup_{\pi_\delta,u\in[0,\tau]}\left|\frac{\mathbb{P}_nU(\beta,\theta,\gamma,u)}{PU(\beta,\theta,\gamma,u)}-\frac{\mathbb{P}_nU(\beta_0,\theta_0,\gamma_0,u)}{PU(\beta_0,\theta_0,\gamma_0,u)}\right|,
\end{aligned}$$

where we used $\mathbb{P}_n(N(\tau)-N(0))=\mathbb{P}_nN(\tau)\le 1$ in the last inequality.

For modulus of continuity, it suffices to prove $\sup_{\pi_\delta}|I_6|\,1_{\Omega_n}$ is bounded by a function of δ, where $P^*(\Omega_n)\to 1$ as $n\to\infty$. We can set

$$\Omega_n=\left\{\omega:\sup_{\pi_\delta,u\in[0,\tau]}\left|\frac{(\mathbb{P}_n-P)Y(u)e^{\beta Z_H(\theta)+\gamma X}}{PY(u)e^{\beta Z_H(\theta)+\gamma X}}\right|\le\frac{1}{2}\right\}.$$

Since $\{Y(u)e^{\beta Z_H(\theta)+\gamma X}: u\in[0,\tau],|\beta|\le\beta_M,|\gamma|\le\gamma_M,\theta\in[0,\theta_M]\}$ is a P-Donsker class by Lemma C.2.1, it is P-Glivenko−Cantelli a.s. (see page 82 of VW [54]),

$$\lim_{n\to\infty}\sup_{\pi_\delta,u\in[0,\tau]}\left|(\mathbb{P}_n-P)Y(u)e^{\beta Z_H(\theta)+\gamma X}\right|=0.\quad P^*\text{-a.s.}.$$

Considering $PY(u)e^{\beta Z_H(\theta)+\gamma X}$ has positive lower bound C_m for $u\in[0,\tau],|\beta|\le\beta_M,|\gamma|\le\gamma_M,\theta\in[0,\theta_M]$ as proved in Chapter 2.3.4.2, it follows

$$\lim_{n\to\infty}\sup_{\pi_\delta,u\in[0,\tau]}\frac{|(\mathbb{P}_n-P)U(\beta,\theta,\gamma,u)|}{PU(\beta,\theta,\gamma,u)}=0,\quad P^*\text{-a.s.},$$

and hence $P^*(\Omega_n)\to 1$ as $n\to\infty$ for Ω_n.

By definition of Ω_n, we have that for samples in Ω_n,

$$\frac{1}{2}\le\frac{\mathbb{P}_nU(\beta,\theta,\gamma,u)}{PU(\beta,\theta,\gamma,u)}\le\frac{3}{2},\quad\text{and hence}\quad\frac{PU(\beta,\theta,\gamma,u)}{\mathbb{P}_nU(\beta,\theta,\gamma,u)}\le 2.$$

$$\text{Then}\quad P^*\sup_{\pi_\delta}|I_6|\,1_{\Omega_n} \le 2P^*\sup_{\pi_\delta,u\in[0,\tau]}\sqrt{n}\left|\frac{\mathbb{P}_nU(\beta,\theta,\gamma,u)}{PU(\beta,\theta,\gamma,u)} - \frac{\mathbb{P}_nU(\beta_0,\theta_0,\gamma_0,u)}{PU(\beta_0,\theta_0,\gamma_0,u)}\right|$$
$$= P^*\sup_{\pi_\delta,u\in[0,\tau]}\sqrt{n}\left|\frac{(\mathbb{P}_n-P)U(\beta,\theta,\gamma,u)}{PU(\beta,\theta,\gamma,u)} - \frac{(\mathbb{P}_n-P)U(\beta_0,\theta_0,\gamma_0,u)}{PU(\beta_0,\theta_0,\gamma_0,u)}\right|$$
$$\le P^*\sup_{\pi_\delta,u\in[0,\tau]}\left|\frac{\mathbb{G}_n\left[U(\beta,\theta,\gamma,u)-U(\beta_0,\theta_0,\gamma_0,u)\right]}{PU(\beta,\theta,\gamma,u)}\right|$$
$$+ P^*\sup_{\pi_\delta,u\in[0,\tau]}\left|\frac{\mathbb{G}_nU(\beta_0,\theta_0,\gamma_0,u)}{PU(\beta_0,\theta_0,\gamma_0,u)}\cdot\frac{P\left[U(\beta,\theta,\gamma,u)-U(\beta_0,\theta_0,\gamma_0,u)\right]}{PU(\beta,\theta,\gamma,u)}\right|.$$

$$\text{Hence}\quad P^*\sup_{\pi_\delta}|I_6|\,1_{\Omega_n}$$
$$\lesssim P^*\sup_{\pi_\delta,u\in[0,\tau]}\left|\mathbb{G}_n\left[U(\beta,\theta,\gamma,u)-U(\beta_0,\theta_0,\gamma_0,u)\right]\right|$$
$$+\sup_{\pi_\delta,u\in[0,\tau]}\left|P\left[U(\beta,\theta,\gamma,u)-U(\beta_0,\theta_0,\gamma_0,u)\right]\right|\cdot P\left|\mathbb{G}_nU(\beta_0,\theta_0,\gamma_0,u)\right|$$
$$\lesssim P^*\sup_{\pi_\delta,u\in[0,\tau]}\left|\mathbb{G}_n\left[U(\beta,\theta,\gamma,u)-U(\beta_0,\theta_0,\gamma_0,u)\right]\right|$$
$$+\sup_{\pi_\delta,u\in[0,\tau]}\left|P\left[U(\beta,\theta,\gamma,u)-U(\beta_0,\theta_0,\gamma_0,u)\right]\right|,$$

where we used that $PU(\beta,\theta,\gamma,u) = PY(u)e^{\beta Z_H(\theta)+\gamma X}$ has a positive lower bound C_m (proved in Chapter 2.3.4.2) in the first inequality and $P\left|\mathbb{G}_nU(\beta_0,\theta_0,\gamma_0,u)\right| < \infty$ in the second inequality.

Now we prove $P\left|\mathbb{G}_nU(\beta_0,\theta_0,\gamma_0,u)\right| < \infty$.

$$P\left|\mathbb{G}_nU(\beta_0,\theta_0,\gamma_0,u)\right| \le \sqrt{P\left|\mathbb{G}_nU(\beta_0,\theta_0,\gamma_0,u)\right|^2}$$
$$\le \sqrt{P[U(\beta_0,\theta_0,\gamma_0,u)-PU(\beta_0,\theta_0,\gamma_0,u)]^2} \le \sqrt{PU^2(\beta_0,\theta_0,\gamma_0,u)}$$
$$\le \sqrt{Pe^{2\beta_0Z_H(\theta_0)+2\gamma_0X}} < \infty \qquad \text{(by Assumptions 2.3.1)}.$$

So to bound $P^*\sup_{\pi_\delta}|I_6|\,1_{\Omega_n}$, we just need to bound

$$P^*\sup_{\pi_\delta,u\in[0,\tau]}\left|\mathbb{G}_n\left[U(\beta,\theta,\gamma,u)-U(\beta_0,\theta_0,\gamma_0,u)\right]\right|$$

and

$$\sup_{\pi_\delta,u\in[0,\tau]}\left|P\left[U(\beta,\theta,\gamma,u)-U(\beta_0,\theta_0,\gamma_0,u)\right]\right|.$$

For the second term,

$$\begin{aligned}
&|P\left[U(\beta,\theta,\gamma,u)-U(\beta_0,\theta_0,\gamma_0,u)\right]| \\
&=\left|PY(u)\left(e^{\beta Z_H(\theta)+\gamma X}-e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\right)\right| \leq P\left|e^{\beta Z_H(\theta)+\gamma X}-e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\right| \\
&=P\left|e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\left(e^{\beta Z_H(\theta)+\gamma X-\beta_0 Z_H(\theta_0)-\gamma_0 X}-1\right)\right| \\
&\leq \sqrt{Pe^{2\beta_0 Z_H(\theta_0)+2\gamma_0 X}P\left(e^{\beta Z_H(\theta)+\gamma X-\beta_0 Z_H(\theta_0)-\gamma_0 X}-1\right)^2}.
\end{aligned}$$

Since $(a+b+c)^2 \leq 3(a^2+b^2+c^2)$,

$$\begin{aligned}
P\left(e^{\beta Z_H(\theta)+\gamma X-\beta_0 Z_H(\theta_0)-\gamma_0 X}-1\right)^2 &\leq 3P\left[\left(e^{\beta(Z_H(\theta)-Z_H(\theta_0))}-1\right)e^{(\beta-\beta_0)Z_H(\theta_0)+(\gamma-\gamma_0)X}\right]^2 \\
&+3P\left[e^{(\beta-\beta_0)Z_H(\theta_0)}\left(e^{(\gamma-\gamma_0)X}-1\right)\right]^2+3P\left(e^{(\beta-\beta_0)Z_H(\theta_0)}-1\right)^2.
\end{aligned}$$

We handle the three components on the right hand side of the inequality separately.

The first component

$$\begin{aligned}
&P\left[\left(e^{\beta(Z_H(\theta)-Z_H(\theta_0))}-1\right)e^{(\beta-\beta_0)Z_H(\theta_0)+(\gamma-\gamma_0)X}\right]^2 \\
&=P\left(e^{\beta(Z_H(\theta)-Z_H(\theta_0))}-1\right)^2 Pe^{2(\beta-\beta_0)Z_H(\theta_0)+2(\gamma-\gamma_0)X} \qquad \text{(by independence)} \\
&\lesssim P\left(e^{2\beta^2|\theta-\theta_0|^{2H}}-2e^{1/2\beta^2|\theta-\theta_0|^{2H}}+1\right) \lesssim |\theta-\theta_0|^{2H},
\end{aligned}$$

by $P\left(e^{2\beta^2|\theta-\theta_0|^{2H}}-2e^{1/2\beta^2|\theta-\theta_0|^{2H}}+1\right)$

$$=2\beta^2|\theta-\theta_0|^{2H}(1+o(1))-1/2\beta^2|\theta-\theta_0|^{2H}(1+o(1))=3/2\beta^2|\theta-\theta_0|^{2H}(1+o(1))$$

and $Pe^{2(\beta-\beta_0)Z_H(\theta_0)+2(\gamma-\gamma_0)X}=Pe^{2(\beta-\beta_0)Z_H(\theta_0)}Pe^{2(\gamma-\gamma_0)X}$

$$\leq P\left[e^{2(\beta_M-\beta_0)Z_H(\theta_0)}+e^{2(-\beta_M-\beta_0)Z_H(\theta_0)}\right]P\left[e^{2(\gamma_M-\gamma_0)X}+e^{2(-\gamma_M-\gamma_0)X}\right]<\infty.$$

(The first inequality holds by monotonicity, the last inequality holds by Conditions B.2.1 (to be presented later in this section).)

The second component

$$\begin{aligned}
&P\left[e^{(\beta-\beta_0)Z_H(\theta_0)}\left(e^{(\gamma-\gamma_0)X}-1\right)\right]^2=Pe^{2(\beta-\beta_0)Z_H(\theta_0)}P\left(e^{(\gamma-\gamma_0)X}-1\right)^2 \\
&\leq P\left(e^{2(\beta_M-\beta_0)Z_H(\theta_0)}+e^{2(-\beta_M-\beta_0)Z_H(\theta_0)}\right)P\left(|(\gamma-\gamma_0)X|e^{0\vee[(\gamma-\gamma_0)X]}\right)^2 \\
&\lesssim (\gamma-\gamma_0)^2 P\left[X^2\left(e^{2(\gamma_M-\gamma_0)X}+e^{2(-\gamma_M-\gamma_0)X}\right)\right] \lesssim (\gamma-\gamma_0)^2,
\end{aligned}$$

where the first inequality holds by monotonity, the second and third inequalities holds by Conditions B.2.1 and Cauchy–Schwartz Inequality.

The third component

$$P\left(e^{(\beta-\beta_0)Z_H(\theta_0)}-1\right)^2$$
$$\le (\beta-\beta_0)^2 P\left[Z_H^2(\theta_0)\left(e^{2(\beta_M-\beta_0)Z_H(\theta_0)}+e^{2(-\beta_M-\beta_0)Z_H(\theta_0)}\right)\right] \lesssim (\beta-\beta_0)^2,$$

where the first inequality holds by monotonity, the second and third inequalities holds by Conditions B.2.1 and Cauchy Schwartz Inequality.

So summing up the three components,

$$|P\left[U(\beta,\theta,\gamma,u)-U(\beta_0,\theta_0,\gamma_0,u)\right]| \lesssim |\theta-\theta_0|^{2H}+(\gamma-\gamma_0)^2+(\beta-\beta_0)^2 \lesssim d^2(\pi,\pi_0).$$

It follows that

$$\sup_{\pi_\delta,u\in[0,\tau]} |P\left[U(\beta,\theta,\gamma,u)-U(\beta_0,\theta_0,\gamma_0,u)\right]| \lesssim \sup_{\pi_\delta} d^2(\pi,\pi_0) \lesssim \delta^2.$$

Now to bound $P^*\sup_{\pi_\delta}|I_6|\,1_{\Omega_n}$, it remains to bound

$$P^*\sup_{\pi_\delta,u\in[0,\tau]} \left|\mathbb{G}_n\left[U(\beta,\theta,\gamma,u)-U(\beta_0,\theta_0,\gamma_0,u)\right]\right|.$$

Since

$$\mathcal{N}_{\delta,-}\equiv\{Y(u)\left(e^{\beta Z_H(\theta)+\gamma X}-e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\right): u\in[0,\tau], d(\pi,\pi_0)<\delta,\} \quad \text{is a subset of}$$

$$\mathcal{N}_{-}\equiv\{Y(u)\left(e^{\beta Z_H(\theta)+\gamma X}-e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\right): u\in[0,\tau], |\beta|\le\beta_M, |\gamma|\le\gamma_M, \theta\in[0,\theta_M]\}$$

which has the same number of bracketing as that of $\mathcal{N}\equiv\{Y(u)e^{\beta Z_H(\theta)+\gamma X}: u\in[0,\tau], |\beta|\le\beta_M, |\gamma|\le\gamma_M, \theta\in[0,\theta_M]\}$, then by Lemma C.2.1 in Appendix C,

$$J_{[]}(1,\mathcal{N}_{\delta,-},L_2(P))<\infty.$$

By Theorem 2.14.2 in VW [54],

$$
\begin{aligned}
&P^* \sup_{\pi_\delta, u\in[0,\tau]} \left|\mathbb{G}_n \left[U(\beta,\theta,\gamma,u) - U(\beta_0,\theta_0,\gamma_0,u)\right]\right| \\
&\le J_{[]}(1, \mathcal{N}_{\delta,-}, L_2(P)) \sqrt{P^* \sup_{\pi_\delta, u\in[0,\tau]} \left[Y(u)\left(e^{\beta Z_H(\theta)+\gamma X} - e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\right)\right]^2} \\
&\lesssim \sqrt{P^* \sup_{\pi_\delta} \left(e^{\beta Z_H(\theta)+\gamma X} - e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\right)^2} \\
&\lesssim \sqrt[4]{P e^{4\beta_0 Z_H(\theta_0)+4\gamma_0 X} P^* \sup_{\pi_\delta} \left(e^{\beta Z_H(\theta)+\gamma X-\beta_0 Z_H(\theta_0)-\gamma_0 X} - 1\right)^4} \\
&\lesssim \sqrt[4]{P^* \sup_{\pi_\delta} \left(e^{\beta Z_H(\theta)+\gamma X-\beta_0 Z_H(\theta_0)-\gamma_0 X} - 1\right)^4} \\
&\lesssim \sqrt[4]{P^* \sup_{\pi_\delta} \left(|\beta Z_H(\theta)+\gamma X-\beta_0 Z_H(\theta_0)-\gamma_0 X| e^{(\beta Z_H(\theta)+\gamma X-\beta_0 Z_H(\theta_0)-\gamma_0 X)\vee 0}\right)^4} \\
&\lesssim \sqrt[8]{P^* \sup_{\pi_\delta} [\beta Z_H(\theta)+\gamma X-\beta_0 Z_H(\theta_0)-\gamma_0 X]^8 P^* \sup_{\pi_\delta} \left(e^{(\beta Z_H(\theta)+\gamma X-\beta_0 Z_H(\theta_0)-\gamma_0 X)\vee 0}\right)^8},
\end{aligned}
$$

where the second inequality holds by $J_{[]}(1, \mathcal{N}_{\delta,-}, L_2(P)) < \infty$ and $Y(u) \le 1$, the third and sixth holds by Cauchy–Schwartz Inequality, the forth holds by Conditions B.2.1 and the fifth holds by $|f(a) - f(b)| \le \sup_{x\in[a,b]} f'(x) \cdot |a-b|$ for continuously differentiable function $f(x)$.

Furthermore,

$$
\begin{aligned}
&P^* \sup_{\pi_\delta} [\beta Z_H(\theta)+\gamma X-\beta_0 Z_H(\theta_0)-\gamma_0 X]^8 \\
=&P^* \sup_{\pi_\delta} \left[\beta(Z_H(\theta)-Z_H(\theta_0)) + (\beta-\beta_0) Z_H(\theta_0) + (\gamma-\gamma_0) X\right]^8 \\
\lesssim&P^* \sup_{\pi_\delta} \left[\beta(Z_H(\theta)-Z_H(\theta_0))\right]^8 + P^* \sup_{\pi_\delta} \left[(\beta-\beta_0) Z_H(\theta_0)\right]^8 + P^* \sup_{\pi_\delta} \left[(\gamma-\gamma_0) X\right]^8, \\
\lesssim&\beta_M^2 \sup_{\pi_\delta} (\theta-\theta_0)^{8H} + \sup_{\pi_\delta} (\beta-\beta_0)^8 P\left[Z_H^8(\theta_0)\right] + \sup_{\pi_\delta} (\gamma-\gamma_0)^8 P(X^8),
\end{aligned}
$$

which is bounded by δ^8 up to a constant using $PZ_H^8(\theta_0) < \infty$ and $PX^8 < \infty$ from Conditions B.2.1 and the maximal inequality for fractional Brownian motion $Z_H(\theta)-$

$Z_H(\theta_0)$ from Novikov and Valkeila [31];

$$\text{and}\quad P^* \sup_{\pi_\delta} \left(e^{(\beta Z_H(\theta)+\gamma X-\beta_0 Z_H(\theta_0)-\gamma_0 X)\vee 0}\right)^8$$
$$= 1 + P\left(e^{8(\beta_M Z_H(\theta)+\gamma_M X-\beta_0 Z_H(\theta_0)-\gamma_0 X)} + e^{8(\beta_M Z_H(\theta)-\gamma_M X-\beta_0 Z_H(\theta_0)-\gamma_0 X)}\right.$$
$$\left.+e^{8(-\beta_M Z_H(\theta)+\gamma_M X-\beta_0 Z_H(\theta_0)-\gamma_0 X)} + e^{8(-\beta_M Z_H(\theta)-\gamma_M X-\beta_0 Z_H(\theta_0)-\gamma_0 X)}\right)$$
$$\lesssim 1 + P\left(e^{8((\beta_M-\beta_0)Z_H(\theta_0)+(\gamma_M-\gamma_0)X)} + e^{8((\beta_M-\beta_0)Z_H(\theta_0)-(\gamma_M+\gamma_0)X)}\right.$$
$$\left.+e^{8(-(\beta_M+\beta_0)Z_H(\theta_0)+(\gamma_M-\gamma_0)X)} + e^{8(-(\beta_M+\beta_0)Z_H(\theta_0)-(\gamma_M X+\gamma_0)X)}\right) < \infty,$$

where we use the monotonity and Conditions B.2.1.

$$\text{It follows that}\quad P^* \sup_{\pi_\delta, u\in[0,\tau]} \left|\mathbb{G}_n\left[U(\beta,\theta,\gamma,u) - U(\beta_0,\theta_0,\gamma_0,u)\right]\right| \leq \delta.$$

$$\text{Hence}\qquad P^* \sup_{\pi_\delta} |I_6|\, 1_{\Omega_n} \lesssim \delta^2 + \delta \lesssim \delta \qquad \text{for small}\quad \delta.$$

In the proof of this section, we used the following moment conditions:

Conditions B.2.1.

1. $P(e^{4\beta_0 Z_H(\theta_0)}) < \infty,\ Pe^{4\gamma_0 X} < \infty.$

2. $P(Z_H^8(\theta_0)) < \infty,\ P(X^8) < \infty.$

3. $P(e^{8(\beta_M-\beta_0)Z_H(\theta_0)}) < \infty, P(e^{-8(\beta_M+\beta_0)Z_H(\theta_0)}) < \infty,$
   $P(e^{8(\gamma_M-\gamma_0)X}) < \infty, P(e^{-8(\gamma_M+\gamma_0)X}) < \infty.$

These conditions are by no means the best conditions, but are sufficient conditions.

Summing up all the results for $P^* \sup_{\pi_\delta} |I_4|, P^* \sup_{\pi_\delta} |I_5|$ and $P^* \sup_{\pi_\delta} |I_6|\, 1_{\Omega_n}$ in this section, we obtain

$$P^* \sup_{d(\pi,\pi_0)<\delta} \sqrt{n}\left|(\mathbb{M}_n - \mathbb{M})(\pi) - (\mathbb{M}_n - \mathbb{M})(\pi_0)\right| 1_{\Omega_n} \lesssim \delta \quad \text{for small}\quad \delta.$$

Then considering Theorem 3.2.5, $\phi_n(\delta) = \delta$. Since $\hat{\pi}_n \to_{P^*} \pi_0$ is proved in the previous section, then solve $r_n^2 \phi_n(\frac{1}{r_n}) \leq \sqrt{n}$, get $r_n = \sqrt{n}$.

$$\sqrt{n} d(\hat{\pi}_n, \pi_0) = O_P^*(1) \Rightarrow n \left[(\beta - \beta_0)^2 + (\gamma - \gamma_0)^2 + (\theta - \theta_0)^{2H}\right] = O_P^*(1).$$

We get the upper bounds $\sqrt{n}, \sqrt{n}, n^{1/(2H)}$ for the rates of convergence of $\hat{\beta}_n, \hat{\gamma}_n, \hat{\theta}_n$ respectively.

Appendix C

Finite entropy integral with bracketing

C.1 Proof of Lemma A.0.1

Now to prove $\mathcal{M} \equiv \{Y(u)e^{Z(\theta)} : u \in [0,\tau], \theta \in [0,\theta_M]\}$ has finite integral of $L_2(P)$ entropy with bracketing, we make a transformation,

$$\begin{aligned}\mathcal{M} &= \{e^{Z(\theta_0)} \cdot Y(u)e^{Z(\theta-\theta_0)} : u \in [0,\tau], \theta \in [0,\theta_M]\} \\ &\equiv \{e^{Z(\theta_0)} \cdot Y(u)e^{W(\theta)} : u \in [0,\tau], \theta \in [-\theta_0, \theta_M - \theta_0]\},\end{aligned}$$

where $W(\cdot) \equiv Z(\cdot + \theta_0) - Z(\theta_0)$ is a 2-sided S.B.M. starting from 0.

Suppose that $\mathcal{M}_{W,0} \equiv \{Y(u)e^{W(\theta)} : u \in [0,\tau], \theta \in [-\theta_0, \theta_M - \theta_0]\}$ has finite integral of $L_2(P)$ entropy with bracketing. We assume the $L_2(P)$ ϵ-sized brackets are $\{(l_i(\epsilon), u_i(\epsilon)) : i = 1, \cdots, N_{[]}(\epsilon, \mathcal{M}_{W,0}, L_2(P))\}$ with $l_i(\epsilon) \geq 0$, then a natural choice of brackets to cover

$$\{Y(u)e^{Z(\theta)} : u \in [0,\tau], \theta \in [0,\theta_M]\}$$

is $\{(e^{Z(\theta_0)} \cdot l_i(\epsilon), e^{Z(\theta_0)} \cdot u_i(\epsilon)) : i = 1, \cdots, N_{[]}(\epsilon, \mathcal{M}_{W,0}, L_2(P))\}$. Its $L_2(P)$ size is

$$\begin{aligned}\sqrt{P(e^{Z(\theta_0)}(u_i(\epsilon) - l_i(\epsilon)))^2} &= \sqrt{P(e^{2Z(\theta_0)})P(u_i(\epsilon) - l_i(\epsilon))^2} \\ &= \sqrt{P(e^{2Z(\theta_0)}) \cdot P(u_i(\epsilon) - l_i(\epsilon))^2} \leq \sqrt{P(e^{2Z(\theta_0)})} \cdot \epsilon,\end{aligned}$$

where the independence of $\{(l_i(\epsilon), u_i(\epsilon)) : i = 1, \cdots, N_{[]}(\epsilon, \mathcal{M}_{W,0}, L_2(P))\}$ and $e^{Z(\theta_0)}$ is used in the first equality. They are independent since the former ones only depend on $W(\cdot)$ which are independent of $Z(\theta_0)$. By the construction of $\mathcal{M}_{W,0}$'s brackets, we have

$$N_{[]}(\sqrt{P(e^{2Z(\theta_0)})} \cdot \epsilon, \mathcal{M}, L_2(P)) \leq N_{[]}(\epsilon, \mathcal{M}_{W,0}, L_2(P)).$$

(Note: Actually equality holds since we can construct the brackets of $\mathcal{M}_{W,0}$ by a reverse transformation from that of $\mathcal{M}$.) Then the entropy integral

$$\begin{aligned}
&\int_0^\infty \sqrt{\log N_{[]}(\tilde{\epsilon}, \mathcal{M}, L_2(P))} d\tilde{\epsilon} \\
=&\frac{1}{\sqrt{P(e^{2Z(\theta_0)})}} \cdot \int_0^\infty \sqrt{\log N_{[]}(\sqrt{P(e^{2Z(\theta_0)})} \cdot \epsilon, \mathcal{M}, L_2(P))} d\epsilon \\
\leq&\frac{1}{\sqrt{P(e^{2Z(\theta_0)})}} \cdot \int_0^\infty \sqrt{\log N_{[]}(\epsilon, \mathcal{M}_{W,0}, L_2(P))} d\epsilon < \infty.
\end{aligned}$$

Now the problem is transformed into proving the integral of $L_2(P)$ entropy with bracketing is finite for $\mathcal{M}_{W,0}$. $\mathcal{M}_{W,0}$ can be further written as $\{Y(u)e^{W(\theta)} : u \in [0,\tau], \theta \in [-\theta_0, 0]\} \cup \{Y(u)e^{W(\theta)} : u \in [0,\tau], \theta \in [0, \theta_M - \theta_0]\}$. Now we introduce a lemma to reduce the problem into prove that for $\{Y(u)e^{W(\theta)} : u \in [0,\tau], \theta \in [-\theta_0, 0]\}$ and $\{Y(u)e^{W(\theta)} : u \in [0,\tau], \theta \in [0, \theta_M - \theta_0]]\}$ separately.

Lemma C.1.1. *If $\mathcal{F}_1$ and $\mathcal{F}_2$ each has finite integral of $L_2(P)$ entropy with bracketing, then $\mathcal{F}_1 \cup \mathcal{F}_2$ has finite integral of $L_2(P)$ entropy with bracketing.*

Proof. Since $\mathcal{F}_1$ and $\mathcal{F}_2$ each has finite integral of $L_2(P)$ entropy with bracketing, then for each of $i = 1, 2$, there exist $\delta_i^* > 0$ and a single bracket (l_i, u_i) covering $\mathcal{F}_i$ such that

$$\int_0^{\delta_i^*} \sqrt{\log N_{[]}(\epsilon, \mathcal{F}_i, L_2(P))} d\epsilon < \infty,$$

$\sqrt{P(u_i^2)} < \infty$, $\sqrt{P(l_i^2)} < \infty$, $\sqrt{P(u_i - l_i)^2} \leq \delta_i^*$, and $N_{[]}(\epsilon, \mathcal{F}_i, L_2(P)) \geq 2$ for $\epsilon < \delta_i^*$.

Firstly we find such a single bracket and δ for $\mathcal{F}_1 \cup \mathcal{F}_2$. Since (l_i, u_i) covers $\mathcal{F}_i$ for $i = 1, 2$, it follows that $(l_1 \wedge l_2, u_1 \vee u_2)$ covers $\mathcal{F}_1 \cup \mathcal{F}_2$, with $\sqrt{P((u_1 \vee u_2)^2)} <$

$\sqrt{2P(u_1^2+u_2^2)} < \infty$, $\sqrt{P((l_1 \wedge l_2)^2)} < \sqrt{2P(l_1^2+l_2^2)} < \infty$, and its $L_2(P)$ size

$$\begin{aligned}\sqrt{P((u_1 \vee u_2 - l_1 \wedge l_2)^2)} &\leq \sqrt{P(|u_1|+|u_2|+|l_1|+|l_2|)^2} \\ &\leq 2\sqrt{P(u_1^2+u_2^2+l_1^2+l_2^2)} < \infty\end{aligned}$$

Then for any $\epsilon \geq 2\sqrt{P(u_1^2+u_2^2+l_1^2+l_2^2)}$, $N_{[]}(\epsilon, \mathcal{F}_1 \cup \mathcal{F}_2, L_2(P)) \leq 1$.

For any $\epsilon > 0$, by definition of bracketing numbers, it is trivial to see

$$N_{[]}(\epsilon, \mathcal{F}_1 \cup \mathcal{F}_2, L_2(P)) \leq N_{[]}(\epsilon, \mathcal{F}_1, L_2(P)) + N_{[]}(\epsilon, \mathcal{F}_2, L_2(P)).$$

Since for any $0 < \epsilon < \delta_1^* \vee \delta_2^*$,

$$N_{[]}(\epsilon, \mathcal{F}_1, L_2(P)) \vee N_{[]}(\epsilon, \mathcal{F}_2, L_2(P)) \geq 2,$$

and $a + b \leq (ab)^2$ for any natural numbers a, b s.t. $a \vee b \geq 2$, then

$$N_{[]}(\epsilon, \mathcal{F}_1 \cup \mathcal{F}_2, L_2(P)) \leq \left(N_{[]}(\epsilon, \mathcal{F}_1, L_2(P)) \cdot N_{[]}(\epsilon, \mathcal{F}_2, L_2(P))\right)^2.$$

For $\epsilon > \delta_1^* \vee \delta_2^*$,

$$N_{[]}(\epsilon, \mathcal{F}_1, L_2(P)) = N_{[]}(\epsilon, \mathcal{F}_2, L_2(P)) = 1,$$

and hence

$$N_{[]}(\epsilon, \mathcal{F}_1 \cup \mathcal{F}_2, L_2(P)) \leq 1 + 1 = 2.$$

To summarize, we have

$$\begin{aligned}&\int_0^\infty \sqrt{\log N_{[]}(\epsilon, \mathcal{F}_1 \cup \mathcal{F}_2, L_2(P))}d\epsilon \\ &\leq \left(\int_0^{\delta_1^* \vee \delta_2^*} + \int_{\delta_1^* \vee \delta_2^*}^{2\sqrt{P(u_1^2+u_2^2+l_1^2+l_2^2)}} + \int_{2\sqrt{P(u_1^2+u_2^2+l_1^2+l_2^2)}}^\infty\right) \sqrt{\log N_{[]}(\epsilon, \mathcal{F}_1 \cup \mathcal{F}_2, L_2(P))}d\epsilon \\ &\leq \int_0^{\delta_1^* \vee \delta_2^*} \sqrt{\log\left(N_{[]}(\epsilon, \mathcal{F}_1, L_2(P)) \cdot N_{[]}(\epsilon, \mathcal{F}_2, L_2(P))\right)^2}d\epsilon \\ &\quad + \int_{\delta_1^* \vee \delta_2^*}^{2\sqrt{P(u_1^2+u_2^2+l_1^2+l_2^2)}} \sqrt{\log 2}d\epsilon + \int_{2\sqrt{P(u_1^2+u_2^2+l_1^2+l_2^2)}}^\infty \sqrt{\log 1}d\epsilon\end{aligned}$$

$$
\begin{aligned}
&\leq \int_0^{\delta_1^* \vee \delta_2^*} \sqrt{2\left(\log N_{[]}(\epsilon, \mathcal{F}_1, L_2(P)) + \log N_{[]}(\epsilon, \mathcal{F}_2, L_2(P))\right)} d\epsilon \\
&\quad + \sqrt{\log 2} \cdot \left(2\sqrt{P(u_1^2 + u_2^2 + l_1^2 + l_2^2)} - \delta_1^* \vee \delta_2^*\right) \\
&\leq \int_0^{\delta_1^* \vee \delta_2^*} \sqrt{2\left(\sqrt{\log N_{[]}(\epsilon, \mathcal{F}_1, L_2(P))} + \sqrt{\log N_{[]}(\epsilon, \mathcal{F}_2, L_2(P))}\right)^2} d\epsilon \\
&\quad + \sqrt{\log 2} \cdot \left(2\sqrt{P(u_1^2 + u_2^2 + l_1^2 + l_2^2)} - \delta_1^* \vee \delta_2^*\right) \\
&= \sqrt{2} \int_0^{\delta_1^*} \sqrt{\log N_{[]}(\epsilon, \mathcal{F}_1, L_2(P))} d\epsilon + \sqrt{2} \int_0^{\delta_2^*} \sqrt{\log N_{[]}(\epsilon, \mathcal{F}_2, L_2(P))} d\epsilon \\
&\quad + \sqrt{\log 2} \cdot \left(2\sqrt{P(u_1^2 + u_2^2 + l_1^2 + l_2^2)} - \delta_1^* \vee \delta_2^*\right) \\
&< \infty.
\end{aligned}
$$

□

Remark C.1.2. *: It is trivial to extend this lemma to finite many sets' union, and whether it holds for $L_r(P)$ norm with $r \geq 1$ can be investigated.*

By symmetry of 2-sided Brownian motion, as long as we prove that for $\{Y(u)e^{W(\theta)} : u \in [0, \tau], \theta \in [0, \theta_0]\}$ and $\{Y(u)e^{W(\theta)} : u \in [0, \tau], \theta \in [0, \theta_M - \theta_0]\}$, we are done. By definition of bracketing numbers, any subset of a functional class with finite integral of $L_2(P)$ entropy with bracketing is a functional class with finite integral of $L_2(P)$ entropy with bracketing; it suffices to prove $\mathcal{M}_{Y,B} \equiv \{Y(u)e^{B(\theta)} : u \in [0, \tau], \theta \in [0, \theta_M]\}$ has finite integral of $L_2(P)$ entropy with bracketing, where $\{B(\theta)\}_{\theta \geq 0}$ is a 1-sided S.B.M. starting from 0.

Lemma C.1.3. *$\mathcal{M}_{Y,B}$ has finite integral of $L_2(P)$ entropy with bracketing.*

To obtain the bracketing number of $\mathcal{M}_{Y,B}$, consider $\mathcal{M}_{Y,B} = \mathcal{F}_Y \cdot \mathcal{G}_B$, where $\mathcal{F}_Y \equiv \{Y(u) : u \in [0, \tau]\}$, $\mathcal{G}_B \equiv \{e^{B(\theta)} : \theta \in [0, \theta_M]\}$. We will try to get brackets and bracketing numbers for $\mathcal{M}_{Y,B}$ from those of $\mathcal{F}_Y$ and $\mathcal{G}_B$.

In order to apply Theorem 2.7.11 of VW [54] to bound the bracketing number of $\mathcal{G}_B$, we verify its Lipschitz property first.

Lemma C.1.4. *The trajectory of $e^{B(\theta)}$ satisfy the Lipschitz condition in Chapter 2.7.4 of VW (1996).*

Proof. For x, y bounded in absolute value by $C > 0$, $|e^x - e^y| \leq e^C |x-y|$, then

$$\forall \theta_1, \theta_2 \in [0, \theta_M], \quad |e^{B(\theta_1)} - e^{B(\theta_2)}| \leq e^{\sup_{\theta\in[0,\theta_M]} B(\theta)} |B(\theta_1) - B(\theta_2)|,$$
$$\text{so} \quad \forall t > 0, \qquad |e^{B(\theta_1)} - e^{B(\theta_2)}|^t \leq \sup_{\theta\in[0,\theta_M]} e^{t\cdot B(\theta)} |B(\theta_1) - B(\theta_2)|^t.$$

Denote $U(\theta) \equiv e^{tB(\theta)}$. Since convex function of a martingale is submartingale under certain conditions (see p.13 of Karatzas and Shreve (1991)), $\{U(\theta)\}_{\theta\in[0,\theta_M]}$ is submartingale (and nonnegative), by Doob's maximal inequality, $P[\sup_{\theta\in[0,\theta_M]} U(\theta)] \leq P[4U(\theta_M)]$. So

$$P\left(\sup_{\theta\in[0,\theta_M]} e^{tB(\theta)}\right) \leq P\left[4e^{tB(\theta_M)}\right] = 4\exp(1/2t^2\theta_M) < \infty.$$

$$\forall t > 0, \quad P\left[\left|e^{B(\theta_1)} - e^{B(\theta_2)}\right|^t\right] \leq \sqrt{P\left(e^{\sup_{\theta\in[0,\theta_M]} B(\theta)}\right)^{2t} \cdot P\left(B(\theta_1) - B(\theta_2)\right)^{2t}}$$
$$\leq \sqrt{4\exp(1/2(2t)^2\theta_M) \cdot C_{2t}|\theta_1 - \theta_2|^t} = 2\sqrt{C_{2t}}\exp(t^2\theta_M)|\theta_1 - \theta_2|^{t/2},$$

where we used $P\left[B(\theta_1) - B(\theta_2)\right]^{2t} = C_{2t}|\theta_1 - \theta_2|^t$ (p.28, Revuz and Yor (2006)) in the second inequality.

By Theorem 1.2.2 of Revuz and Yor ([38]),

$$P\left[\left(\sup_{\theta_1\neq\theta_2} \frac{|e^{B(\theta_1)} - e^{B(\theta_2)}|}{|\theta_1 - \theta_2|^\alpha}\right)^t\right] < \infty, \quad \forall \alpha \in [0, 1/2), \quad \forall t > 0.$$

$$\text{So} \qquad \left|e^{B(\theta_1)} - e^{B(\theta_2)}\right| \leq L|\theta_1 - \theta_2|^\alpha, \quad \forall \alpha \in [0, 1/2),$$
$$\text{where} \quad P[L^t] < \infty, \quad \forall t > 0.$$

So up to now, we have established the Lipschitz property $\left|e^{B(\theta_1)} - e^{B(\theta_2)}\right| \leq L \cdot d(\theta_1, \theta_2)$ for the trajectory of $e^{B(\theta)}$, where $d(\theta_1, \theta_2) = |\theta_1 - \theta_2|^\alpha$ and $P(L^t) < \infty$ for any $t > 0$. This result holds for any $\alpha \in [0, 1/2)$. □

Then by a slight modification of Theorem 2.7.11 of VW [54], for $\mathcal{G}_B$ and $\mathcal{G}_{B-} \equiv \{e^{B(\theta)} - 1 : \theta \in [0, \theta_M]\}$, for the true underlying probability measure P, for any $r \geq 1$, their bracketing numbers

$$N_{[]}(\epsilon \| e^{\sup_{\theta \in [0,\theta_M]} B(\theta)} \|_{P,r}, \mathcal{G}_B, L_r(P)) \leq N(\epsilon, [0, \theta_M], d)$$

$$N_{[]}(\epsilon \| e^{\sup_{\theta \in [0,\theta_M]} B(\theta)} \|_{P,r}, \mathcal{G}_{B-}, L_r(P)) \leq N(\epsilon, [0, \theta_M], d).$$

for $\forall \alpha \in [0, 1/2)$, where $d(\theta_1, \theta_2) = |\theta_1 - \theta_2|^\alpha$. Because $N(\epsilon, [0, \theta_M], d) \leq \lceil \theta_M / \epsilon^{1/\alpha} \rceil$, where $\lceil a \rceil$ is defined as the smallest integer that is no less than a.

$$N_{[]}(\epsilon \| e^{\sup_{\theta \in [0,\theta_M]} B(\theta)} \|_{P,r}, \mathcal{G}_B, L_r(P)) \leq \lceil \theta_M \epsilon^{-1/\alpha} \rceil, \tag{C.1}$$

$$N_{[]}(\epsilon \| e^{\sup_{\theta \in [0,\theta_M]} B(\theta)} \|_{P,r}, \mathcal{G}_{B-}, L_r(P)) \leq \lceil \theta_M \epsilon^{-1/\alpha} \rceil \tag{C.2}$$

Now we can prove Lemma C.1.3.

Proof. Since $\mathcal{F}_Y$ is a class of monotone functions not exceeding 1, by Theorem 2.7.5 of VW (1996), it has finite bracketing entropy integral with envelope 1,

$$\log N_{[]}(\epsilon, \mathcal{F}_Y, L_r(Q)) \leq K \frac{1}{\epsilon}$$

for every probability measure Q, every $r \geq 1$, and a constant K that depends on r only.

Then for any $\epsilon > 0$, we can choose no more than $e^{K/\epsilon}$ brackets (l_i^f, u_i^f), s.t. $0 \leq l_i^f \leq u_i^f \leq 1$ (since otherwise we can take $(l_i^f \vee 0, u_i^f \wedge 1)$, which still forms brackets covering $\mathcal{F}_Y$) and $[P(u_i^f - l_i^f)^r]^{\frac{1}{r}} \leq \epsilon$ for any $r \geq 1$. Here we omitted ϵ in $(l_i^f(\epsilon), u_i^f(\epsilon))$ for notational convenience. We will do the same thing for the brackets of $\mathcal{G}_B$ in the following paragraphs.

Remark C.1.5. *: Actually we can choose no more than K/ϵ brackets satisfying all these conditions since $Y(\cdot)$ is an indicator function and monotone, whose brackets can be constructed similarly as that of C.D.F. (but the bracketing numbers here do* *not need change for different norm $L_r(Q)$). But for the proof of Lemma C.1.3, the bracketing number $e^{K/\epsilon}$ suffices.*

Consider the case $r = 4$ for (C.1), we can choose no more than $\lceil \theta_M \epsilon^{-1/\alpha} \rceil$ brackets (l_j^g, u_j^g) for $\mathcal{G}_B$, s.t. $[P(u_j^g - l_j^g)^4]^{\frac{1}{4}} \le \epsilon \| e^{\sup_{\theta \in [0,\theta_M]} B(\theta)} \|_{P,4}$, where ϵ is the same as that used for brackets of class $\mathcal{F}_Y \equiv \{Y(u) : u \in [0, \tau]\}$. It is obvious that we can choose all the bracket functions for $\mathcal{G}_B$ to fall within $[0, \sup_{\theta \in [0,\theta_M]} e^{B(\theta)}]$ (since otherwise we can take $(l_j^g \vee 0, u_j^g \wedge \sup_{\theta \in [0,\theta_M]} e^{B(\theta)})$, which still forms brackets covering $\mathcal{G}_B$).

Obviously all the brackets formed by $(l_i^f \cdot l_j^g, u_i^f \cdot u_j^g)$ can cover class $\mathcal{M}_{Y,B}$. If the bracketing entropy integral of $\mathcal{M}_{Y,B}$ is finite, then Lemma C.1.3 is done.

The $L_2(P)$ size of the bracket $(l_i^f \cdot l_j^g, u_i^f \cdot u_j^g)$

$$\sqrt{P\left(u_i^f \cdot u_j^g - l_i^f \cdot l_j^g\right)^2} = \sqrt{P\left(u_i^f (u_j^g - l_j^g) + l_j^g (u_i^f - l_i^f)\right)^2}$$

$$\le \sqrt{2P\left(u_i^f (u_j^g - l_j^g)\right)^2 + 2P\left(l_j^g (u_i^f - l_i^f)\right)^2}$$

$$\le \sqrt{2\left[P\left(1 \cdot (u_j^g - l_j^g)\right)^2 + P\left(l_i^g (u_i^f - l_i^f)\right)^2\right]} \qquad (0 \le u_i^f \le 1)$$

$$\le \sqrt{2\left[\left(P(u_j^g - l_j^g)^4\right)^{\frac{1}{2}} + \left(P(l_j^g)^4 P(u_i^f - l_i^f)^4\right)^{\frac{1}{2}}\right]} \quad \text{(Cauchy–Schwartz inequality)}$$

$$\le \sqrt{2\left[\left(\epsilon \| e^{\sup_{\theta \in [0,\theta_M]} B(\theta)} \|_{P,4}\right)^2 + \left(\| e^{\sup_{\theta \in [0,\theta_M]} B(\theta)} \|_{P,4}^4 \cdot \epsilon^4\right)^{\frac{1}{2}}\right]}$$

$$\le \sqrt{2\left[\left(\epsilon \| e^{\sup_{\theta \in [0,\theta_M]} B(\theta)} \|_{P,4}\right)^2 + \| e^{\sup_{\theta \in [0,\theta_M]} B(\theta)} \|_{P,4}^2 \epsilon^2\right]} = 2 \| e^{\sup_{\theta \in [0,\theta_M]} B(\theta)} \|_{P,4} \cdot \epsilon$$

Then the number of brackets $(l_i^f \cdot l_j^g, u_i^f \cdot u_j^g)$ can be written as

$$\begin{aligned} & N_{[]}\left(2 \| e^{\sup_{\theta \in [0,\theta_M]} B(\theta)} \|_{P,4} \epsilon, \mathcal{M}_{Y,B}, \| \cdot \|_{P,2}\right) \\ & \qquad \le N_{[]}(\epsilon, \mathcal{F}_Y, L_4(P)) \cdot N_{[]}(\epsilon \| e^{\sup_{\theta \in [0,\theta_M]} B(\theta)} \|_{P,r}, \mathcal{G}_B, L_4(P)), \\ & \qquad \le e^{\frac{K}{\epsilon}} \lceil \theta_M \epsilon^{-1/\alpha} \rceil \lesssim e^{\frac{K}{\epsilon}} \cdot \epsilon^{-1/\alpha}, \\ & \log N_{[]}\left(2 \| e^{\sup_{\theta \in [0,\theta_M]} B(\theta)} \|_{P,4} \epsilon, \mathcal{M}_{Y,B}, \| \cdot \|_{P,2}\right) \lesssim \frac{K}{\epsilon} + \alpha \log \frac{1}{\epsilon} \lesssim \frac{1}{\epsilon}, \\ & \int_0^1 \sqrt{\log N_{[]}\left(2 \| e^{\sup_{\theta \in [0,\theta_M]} B(\theta)} \|_{P,4} \epsilon, \mathcal{M}_{Y,B}, \| \cdot \|_{P,2}\right)} d\epsilon < \infty. \end{aligned}$$

Then

$$\int_0^{2\|e^{\sup_{\theta\in[0,\theta_M]} B(\theta)}\|_{P,4}} \sqrt{\log N_{[]}\left(\tilde{\epsilon}, \mathcal{M}_{Y,B}, \|\cdot\|_{P,2}\right)}d\tilde{\epsilon} < \infty.$$

So up to now we have proved the finite entropy integral property of $\mathcal{M}_{Y,B}$, and hence that of $\mathcal{M}$. Lemma 2.2.4 is proved. □

C.2 Proof of the functional set (I) for Chapter 2.3

Lemma C.2.1. $\mathcal{N} \equiv \{Y(u)e^{\beta Z_H(\theta)+\gamma X} : u \in [0,\tau], |\beta| \le \beta_M, \theta \in [0,\theta_M], |\gamma| \le \gamma_M\}$ *has finite integral of $L_2(P)$ entropy with bracketing.*

In this section, we adopt the same procedure and will directly apply some results obtained from the previous section. To prove $\mathcal{N}$ has finite integral of $L_2(P)$ entropy with bracketing, we make a transformation,

$$\begin{aligned}\mathcal{N} &= \{e^{\beta Z_H(\theta_0)+\gamma X}\cdot Y(u)e^{\beta(Z_H(\theta-\theta_0))} : u \in [0,\tau], |\beta| \le \beta_M, \theta \in [0,\theta_M], |\gamma| \le \gamma_M\}\\ &\equiv \{e^{\beta Z_H(\theta_0)+\gamma X}\cdot Y(u)e^{\beta W_H(\theta)} : u \in [0,\tau], |\beta| \le \beta_M, \theta \in [-\theta_0,\theta_M-\theta_0], |\gamma| \le \gamma_M\},\end{aligned}$$

where $W_H(\cdot) \equiv Z_H(\cdot+\theta_0) - Z_H(\theta_0)$ is a 2-sided S.B.M.starting from 0. Consider $\mathcal{N} = \mathcal{P}_0 \cdot \mathcal{N}_{W,0}$, where $\mathcal{P}_0 \equiv \{e^{\beta Z_H(\theta_0)+\gamma X} : |\beta| \le \beta_M, |\gamma| \le \gamma_M\}$ and $\mathcal{N}_{W,0} \equiv \{Y(u)e^{\beta W_H(\theta)} : u \in [0,\tau], |\beta| \le \beta_M, \theta \in [-\theta_0, \theta_M-\theta_0]\}$. We aim to get brackets and bracketing numbers of $\mathcal{N}$ from those of $\mathcal{P}_0$ and $\mathcal{N}_{W,0}$.

For $\mathcal{P}_0$, we use the Lipschitz property of $e^{\beta Z_H(\theta_0)+\gamma X}$ to obtain its bracketing number. Since $|e^x - e^y| \le e^{x\vee y}\cdot|x-y|$, it follows that

$$|e^{\beta_1 Z_H(\theta_0)+\gamma_1 X} - e^{\beta_2 Z_H(\theta_0)+\gamma_2 X}| \le e^{\beta_M|Z_H(\theta_0)|+\gamma_M|X|}\cdot|(\beta_1-\beta_2)Z_H(\theta_0)+(\gamma_1-\gamma_2)X|.$$

The bracketing numbers of $\{\beta : |\beta| \le \beta_M\}$ and $\{\gamma : |\gamma| \le \gamma_M\}$ are $N(\epsilon, [-\beta_M, \beta_M], d)$ and $N(\epsilon, [-\gamma_M, \gamma_M], d)$ respectively (with $d(\beta_1,\beta_2) = |\beta_1-\beta_2|$ and $d(\gamma_1,\gamma_2) = |\gamma_1 - \gamma_2|$), then we denote their brackets by $(l_i^\beta, u_i^\beta)(i = 1, ..., N(\epsilon, [-\beta_M, \beta_M], d))$ and $(l_j^\gamma, u_j^\gamma)(j = 1, ..., N(\epsilon, [-\gamma_M, \gamma_M], d))$. Notice that the brackets constructed have $|u_i^\beta - l_i^\beta| \le \epsilon$ and $|u_i^\gamma - l_i^\gamma| \le \epsilon$.

Then we can construct no more than $N(\epsilon, [0, \beta_M], d) \cdot N(\epsilon, [-\gamma_M, \gamma_M], d)$ brackets for $\mathcal{P}_0$ using $e^{\beta_M |Z_H(\theta_0)| + \gamma_M |X|}(l_i^\beta \cdot Z_H(\theta_0) + l_j^\gamma \cdot X, u_i^\beta \cdot Z_H(\theta_0) + u_j^\gamma \cdot X)$ with bracket size bounded by

$$\left|\left| e^{\beta_M |Z_H(\theta_0)| + \gamma_M |X|} \left((u_i^\beta - l_i^\beta) Z_H(\theta_0) + (u_j^\gamma - l_j^\gamma) X \right) \right|\right|$$

for norm $\| \cdot \|$.

In particular, for $L_2(P)$ norm, the bracket size is bounded by

$$\begin{aligned} &\sqrt{P\left(e^{\beta_M |Z_H(\theta_0)| + \gamma_M |X|} \left((u_i^\beta - l_i^\beta) Z_H(\theta_0) + (u_j^\gamma - l_j^\gamma) X \right)\right)^2} \\ \leq &\sqrt{P\left(e^{\beta_M |Z_H(\theta_0)| + \gamma_M |X|} \left(|Z_H(\theta_0)| + |X|\right) \epsilon\right)^2} = \epsilon \|(|Z_H(\theta_0)| + |X|) \cdot e^{\beta_M |Z_H(\theta_0)| + \gamma_M |X|}\|_{P,2}. \end{aligned}$$

Since $P(Z_H^2(\theta_0) e^{2\beta_M |Z_H(\theta_0)|}) < \infty$,$P(X^2 e^{2\gamma_M |X|}) < \infty$, and $Z_H(\theta_0), X$ are independent, it is easy to deduce

$$P_{Z_0,X} \equiv \|(|Z_H(\theta_0)| + |X|) \cdot e^{\beta_M |Z_H(\theta_0)| + \gamma_M |X|}\|_{P,2} < \infty,$$

and hence $\quad N_{[]}(\epsilon \cdot P_{Z_0,X}, \mathcal{P}_0, L_2(P)) \leq N(\epsilon, [-\beta_M, \beta_M], d) \cdot N(\epsilon, [-\gamma_M, \gamma_M], d),$

$$N_{[]}(\epsilon \cdot P_{Z_0,X}, \mathcal{P}_0, L_2(P)) \leq \lceil \tfrac{2\beta_M}{\epsilon} \rceil \lceil \tfrac{2\gamma_M}{\epsilon} \rceil. \tag{C.3}$$

Then we can choose no more than $\lceil \frac{2\beta_M}{\epsilon} \rceil \lceil \frac{2\gamma_M}{\epsilon} \rceil$ brackets $\{(l_i^p(\epsilon), u_i^p(\epsilon)) : i = 1, ..., N_{[]}(\epsilon \cdot P_{Z_0,X}, \mathcal{P}_0, L_2(P))\}$ that cover $\mathcal{P}_0$ with $l_i^p \geq 0$, $u_i^p \leq \sup_{|\beta| \leq \beta_M, |\gamma| \leq \gamma_M} e^{\beta Z(\theta_0) + \gamma X}$ and $\sqrt{P(u_i^p(\epsilon) - l_i^p(\epsilon))^2} \leq \epsilon \cdot P_{Z_0,X}$ for $i = 1, ..., N_{[]}(\epsilon \cdot P_{Z_0,X}, \mathcal{P}_0, L_2(P))$.

Suppose that $\mathcal{N}_{W,0}$ has finite integral of $L_2(P)$ entropy with bracketing. We assume the $L_2(P)$ ϵ-sized brackets are $\{(l_i^n(\epsilon), u_i^n(\epsilon)) : i = 1, \cdots, N_{[]}(\epsilon, \mathcal{N}_{W,0}, L_2(P))\}$ with $l_i^n(\epsilon) \geq 0$, $u_i^n(\epsilon) \leq \sup_{u \in [0,\tau], |\beta| \leq \beta_M, \theta \in [-\theta_0, \theta_M - \theta_0]} Y(u) e^{\beta W_H(\theta)}$, then a natural choice of brackets to cover $\mathcal{N}$ is $\{(l_i^p(\epsilon) \cdot l_i^n(\epsilon), u_i^p(\epsilon) \cdot u_i^n(\epsilon)) : i = 1, \cdots, N_{[]}(\epsilon, \mathcal{N}_{W,0}, L_2(P))\}$. Its $L_2(P)$ size is

$$\begin{aligned} &\sqrt{P(u_i^p \cdot u_i^n - l_i^p \cdot l_i^n)^2} = \sqrt{P\left((u_i^p - l_i^p) u_i^n + l_i^p (u_i^n - l_i^n)\right)^2} \\ &\leq \sqrt{2\left[P\left((u_i^p - l_i^p) u_i^n\right)^2 + P\left(l_i^p (u_i^n - l_i^n)\right)^2\right]} \\ &= \sqrt{2\left[P(u_i^p - l_i^p)^2 P(u_i^n)^2 + P(l_i^p)^2 P(u_i^n - l_i^n)^2\right]} \end{aligned}$$

$$\leq \sqrt{2\left[\left(\epsilon \cdot P_{Z_0,X}\right)^2 \cdot P \sup_{u,\beta,\theta}\left(Y(u)e^{\beta W_H(\theta)}\right)^2 + P\sup_{\beta,\gamma}\left(e^{\beta Z_H(\theta_0)+\gamma X}\right)^2 \cdot \epsilon^2\right]}$$
$$= \sqrt{2\left[\left(P_{Z_0,X}\right)^2 \cdot P \sup_{u,\beta,\theta}\left(Y(u)e^{\beta W_H(\theta)}\right)^2 + P\sup_{\beta,\gamma}\left(e^{\beta Z_H(\theta_0)+\gamma X}\right)^2\right]} \cdot \epsilon$$
$$\equiv P_{Z_0,X,W} \cdot \epsilon$$

where we abbreviated the notation of supremums since they are evident here. The independence of $\{(l_i(\epsilon), u_i(\epsilon)) : i = 1, \cdots, N_{[]}(\epsilon, \mathcal{N}_{W,0}, L_2(P))\}$ and $e^{Z_H(\theta_0)}$ are used in the second equality. They are independent since the brackets of $\mathcal{N}_{W,0}$ only depend on $W_H(\cdot)$, the brackets of $\mathcal{P}_0$ only depend on $(Z_H(\theta_0), X)$, and $W_H(\cdot)$ are independent of $(Z_H(\theta_0), X)$.

It is easy to prove $P_{Z_0,X,W} < \infty$.

By the construction of $\mathcal{N}$'s brackets, we have

$$N_{[]}(P_{Z_0,X,W} \cdot \epsilon, \mathcal{N}, L_2(P)) \leq N_{[]}(\epsilon \cdot P_{Z_0,X}, \mathcal{P}_0, L_2(P)) \cdot N_{[]}(\epsilon, \mathcal{N}_{W,0}, L_2(P)).$$

Then the entropy integral

$$\begin{aligned}
&\int_0^\infty \sqrt{\log N_{[]}(\tilde{\epsilon}, \mathcal{N}, L_2(P))}d\tilde{\epsilon} \\
=&\frac{1}{P_{Z_0,X,W}} \cdot \int_0^\infty \sqrt{\log N_{[]}(P_{Z_0,X,W} \cdot \epsilon, \mathcal{N}, L_2(P))}d\epsilon \\
\leq&\frac{1}{P_{Z_0,X,W}} \cdot \int_0^\infty \sqrt{\log N_{[]}(\epsilon \cdot P_{Z_0,X}, \mathcal{P}_0, L_2(P)) + \log N_{[]}(\epsilon, \mathcal{N}_{W,0}, L_2(P))}d\epsilon \\
\leq&\frac{1}{P_{Z_0,X,W}} \cdot \int_0^\infty \left(\sqrt{\log N_{[]}(\epsilon \cdot P_{Z_0,X}, \mathcal{P}_0, L_2(P))} + \sqrt{\log N_{[]}(\epsilon, \mathcal{N}_{W,0}, L_2(P))}\right) d\epsilon \\
\leq&\frac{1}{P_{Z_0,X,W}} \cdot \int_0^\infty \left(\sqrt{\log\left(\lceil\frac{2\beta_M}{\epsilon}\rceil\lceil\frac{2\gamma_M}{\epsilon}\rceil\right)} + \sqrt{\log N_{[]}(\epsilon, \mathcal{N}_{W,0}, L_2(P))}\right) d\epsilon.
\end{aligned}$$

Since $\int_0^\infty \sqrt{\log\left(\lceil\frac{2\beta_M}{\epsilon}\rceil\lceil\frac{2\gamma_M}{\epsilon}\rceil\right)}d\epsilon < \infty$, now the problem is transformed into proving the integral of $L_2(P)$ entropy with bracketing is finite for $\mathcal{N}_{W,0}$. Following the lines of Appendix C.1, it suffices to prove $\mathcal{N}_{Y,B} \equiv \{Y(u)e^{\beta B(\theta)} : u \in [0,\tau], |\beta| \leq \beta_M, \theta \in [0,\theta_M]\}$ has finite integral of $L_2(P)$ entropy with bracketing, where $\{B(\theta)\}_{\theta\geq 0}$ is a 1-sided fBm starting from 0 with Hurst parameter H.

Lemma C.2.2. *$\mathcal{N}_{Y,B}$ has finite integral of $L_2(P)$ entropy with bracketing.*

To obtain the bracketing number of $\mathcal{N}_{Y,B}$, consider $\mathcal{N}_{Y,B} = \mathcal{F}_Y \cdot \mathcal{H}_B$, where $\mathcal{F}_Y \equiv \{Y(u) : u \in [0,\tau]\}$, $\mathcal{H}_B \equiv \{e^{\beta B(\theta)} : |\beta| \leq \beta_M, \theta \in [0, \theta_M]\}$. We will try to get brackets and bracketing numbers for $\mathcal{N}_{Y,B}$ from those of $\mathcal{F}_Y$ and $\mathcal{H}_B$.

In order to apply Theorem 2.7.11 of VW [54] to bound the bracketing number of $\mathcal{H}_B$, we verify its Lipschitz property first.

Lemma C.2.3. *The trajectory of $e^{\beta B(\theta)}$ satisfy the Lipschitz condition in chapter 2.7.4 of VW (1996).*

Proof. For x, y bounded in absolute value by $C > 0$, $|e^x - e^y| \leq e^C|x-y|$, then

$$\forall\theta_1,\theta_2 \in [0,\theta_M], \beta_1,\beta_2 \in [-\beta_M, \beta_M],$$
$$|e^{\beta_1 B(\theta_1)} - e^{\beta_2 B(\theta_2)}| \leq e^{\beta_1 B(\theta_1)\vee\beta_2 B(\theta_2)}|\beta_1 B(\theta_1) - \beta_2 B(\theta_2)|,$$
$$\leq e^{\sup_{|\beta|\leq\beta_M,\theta\in[0,\theta_M]}\beta B(\theta)}|(\beta_1-\beta_2)B(\theta_1) + \beta_2(B(\theta_1) - B(\theta_2))|.$$

By Theorem 1.2.2 of Revuz and Yor ([38]) and fBm's property,

$$|B(\theta_1) - B(\theta_2)| \leq L|\theta_1-\theta_2|^\alpha, \quad \forall\alpha\in[0,H), \quad \text{where} \quad P[L^t] < \infty, \quad \forall t > 0.$$

Then

$$|e^{\beta_1 B(\theta_1)} - e^{\beta_2 B(\theta_2)}| \leq \sup_\theta |B(\theta)| e^{\sup_{\beta,\theta}\beta B(\theta)}|\beta_1-\beta_2| + \beta_M e^{\sup_{\beta,\theta}\beta B(\theta)} L|\theta_1-\theta_2|^\alpha$$
$$\equiv L_\beta \cdot d_\beta(\beta_1,\beta_2) + L_\theta \cdot d_\theta(\theta_1,\theta_2).$$
$$\forall\alpha\in[0,H), \quad \text{where} \quad P[L^t] < \infty, \quad \forall t > 0.$$

Since $\forall t > 0$, $P\sup_\theta |B(\theta)|^t \leq \sqrt{P\sup_\theta B^{2t}(\theta)} < \infty$ by Theorem 1.1 of Novikov and Valkeila [31], and

$$P\left(e^{\sup_{\beta,\theta}\beta B(\theta)}\right)^t \leq Pe^{\sup_\theta t\beta_M|B(\theta)|} \leq Pe^{\sup_\theta t\beta_M B(\theta)} + Pe^{\sup_\theta t\beta_M(-B(\theta))}$$
$$\leq 2Pe^{\sup_\theta t\beta_M B(\theta)} \leq 2\cdot 4e^{1/2(t\beta_M)^2\beta_M^{2H}},$$

by Lemma 3.2 of Lee [21] for $\forall H \in (1/2, 1)$ (see Lemma C.2.4 in the end of Appendix C.2) and Doob's maximal inequality for $H = 1/2$.

Then by Cauchy–Schwartz Inequality, for $\forall t > 0$, $PL^t_\beta < \infty$ and $PL^t_\theta < \infty$.

So up to now, we have established the Lipschitz property $\left|e^{\beta_1 B(\theta_1)} - e^{\beta_2 B(\theta_2)}\right| \leq L_\beta \cdot d_\beta(\beta_1, \beta_2) + L_\theta \cdot d_\theta(\theta_1, \theta_2)$ for the trajectory of $e^{\beta B(\theta)}$, where $d_\theta(\theta_1, \theta_2) = |\theta_1 - \theta_2|^\alpha$, $d_\beta(\beta_1, \beta_2) = |\beta_1 - \beta_2|$ and $P(L^t_\beta) < \infty$, $PL^t_\theta < \infty$ for any $t > 0$. This result holds for any $\alpha \in [0, H)$ and $H \in [1/2, 1)$. □

Then by the same way we did for $\mathcal{P}_0$, for $\mathcal{H}_B$ and $\mathcal{H}_{B-} \equiv \{e^{\beta B(\theta)} - 1 : |\beta| \leq \beta_M, \theta \in [0, \theta_M]\}$, for the true underlying probability measure P, for any $r \geq 1$, their bracketing numbers

$$N_{[]}(\epsilon^{1/\alpha}\|L_\beta + L_\theta\|_{P,r}, \mathcal{H}_B, L_r(P)) \leq N(\epsilon, [0, \theta_M], d_\theta) \cdot N(\epsilon, [-\beta_M, \beta_M], d_\beta),$$

$$N_{[]}(\epsilon\|L_\beta + L_\theta\|_{P,r}, \mathcal{H}_B, L_r(P)) \leq N(\epsilon^\alpha, [0, \theta_M], d_\theta) \cdot N(\epsilon^\alpha, [-\beta_M, \beta_M], d_\beta),$$

for an $\alpha \in [0, H)$. Because $N(\epsilon^\alpha, [0, \theta_M], d_\theta) \leq \lceil \theta_M/\epsilon \rceil$, and $N(\epsilon^\alpha, [-\beta_M, \beta_M], d_\beta) \leq \lceil 2\beta_M/\epsilon^\alpha \rceil$,

$$N_{[]}(\epsilon\|L_\beta + L_\theta\|_{P,r}, \mathcal{H}_B, L_r(P)) \leq \lceil \theta_M \epsilon \rceil \lceil 2\beta_M/\epsilon^\alpha \rceil, \tag{C.4}$$

The same result holds for $\mathcal{H}_{B-}$ as well.

Now we can prove Lemma C.1.3.

Consider the case $r = 4$ for (C.1), we can choose no more than $\lceil \theta_M \epsilon^{-1/\alpha} \rceil \lceil 2\beta_M/\epsilon^\alpha \rceil$ brackets (l^g_j, u^g_j) for $\mathcal{H}_B$, s.t. $[P(u^g_j - l^g_j)^4]^{\frac{1}{4}} \leq \epsilon^{1/\alpha}\|L_\beta + L_\theta\|_{P,4}$, where ϵ is the same as that used for brackets of class $\mathcal{F}_Y \equiv \{Y(u) : u \in [0, \tau]\}$. It is obvious that we can choose all the bracket functions for $\mathcal{G}_B$ to fall within $[0, \sup_{|\beta| \leq \beta_M, \theta \in [0, \theta_M]} e^{\beta B(\theta)}]$ (since otherwise we can take $(l^g_j \vee 0, u^g_j \wedge \sup_{|\beta| \leq \beta_M, \theta \in [0, \theta_M]} e^{\beta B(\theta)})$, which still forms brackets covering $\mathcal{H}_B$).

Obviously all the brackets formed by $(l^f_i \cdot l^g_j, u^f_i \cdot u^g_j)$ can cover class $\mathcal{N}_{Y,B}$. If the bracketing entropy integral of $\mathcal{N}_{Y,B}$ is finite, then Lemma C.1.3 is done.

The $L_2(P)$ size of the bracket $(l_i^f \cdot l_j^g, u_i^f \cdot u_j^g)$

$$\sqrt{P\left(u_i^f \cdot u_j^g - l_i^f \cdot l_j^g\right)^2} = \sqrt{P\left(u_i^f(u_j^g - l_j^g) + l_j^g(u_i^f - l_i^f)\right)^2}$$
$$\leq \sqrt{2P\left(u_i^f(u_j^g - l_j^g)\right)^2 + 2P\left(l_j^g(u_i^f - l_i^f)\right)^2}$$
$$\leq \sqrt{2\left[P\left(1 \cdot (u_j^g - l_j^g)\right)^2 + P\left(l_i^g(u_i^f - l_i^f)\right)^2\right]} \qquad (0 \leq u_i^f \leq 1)$$
$$\leq \sqrt{2\left[\left(P(u_j^g - l_j^g)^4\right)^{\frac{1}{2}} + \left(P(l_j^g)^4 P(u_i^f - l_i^f)^4\right)^{\frac{1}{2}}\right]} \quad \text{(Cauchy–Schwartz inequality)}$$
$$\leq \sqrt{2\left[\left(\epsilon^{1/\alpha}\|L_\beta + L_\theta\|_{P,4}\right)^2 + \left(\|L_\beta + L_\theta\|_{P,4}^4 \cdot \epsilon^4\right)^{\frac{1}{2}}\right]} \leq 2\|L_\beta + L_\theta\|_{P,4} \cdot \epsilon^{1/\alpha},$$

where we used the fact that $0 < \alpha < H < 1$.

Then the number of brackets $(l_i^f \cdot l_j^g, u_i^f \cdot u_j^g)$ can be written as

$$N_{[]}\left(2\|L_\beta + L_\theta\|_{P,4}\epsilon^{1/\alpha}, \mathcal{N}_{Y,B}, \|\cdot\|_{P,2}\right)$$
$$\leq N_{[]}(\epsilon, \mathcal{F}_Y, L_4(P)) \cdot N_{[]}(\epsilon^{1/\alpha}\|L_\beta + L_\theta\|_{P,4}, \mathcal{H}_B, L_4(P)),$$
$$N_{[]}\left(2\|L_\beta + L_\theta\|_{P,4}\epsilon, \mathcal{N}_{Y,B}, \|\cdot\|_{P,2}\right)$$
$$\leq N_{[]}(\epsilon^\alpha, \mathcal{F}_Y, L_4(P)) \cdot N_{[]}(\epsilon\|L_\beta + L_\theta\|_{P,4}, \mathcal{H}_B, L_4(P)),$$
$$N_{[]}\left(2\|L_\beta + L_\theta\|_{P,4}\epsilon, \mathcal{N}_{Y,B}, \|\cdot\|_{P,2}\right) \leq e^{\frac{K}{\epsilon^\alpha}}\lceil\theta_M\epsilon^{-1}\rceil\lceil 2\beta_M/\epsilon^\alpha\rceil \lesssim e^{\frac{K}{\epsilon^\alpha}} \cdot \epsilon^{-1-\alpha},$$
$$\log N_{[]}\left(2\|L_\beta + L_\theta\|_{P,4}\epsilon, \mathcal{N}_{Y,B}, \|\cdot\|_{P,2}\right) \lesssim \frac{K}{\epsilon^\alpha} + (1+\alpha)\log\frac{1}{\epsilon} \lesssim \frac{1}{\epsilon},$$
$$\int_0^1 \sqrt{\log N_{[]}\left(2\|L_\beta + L_\theta\|_{P,4}\epsilon, \mathcal{N}_{Y,B}, \|\cdot\|_{P,2}\right)}d\epsilon < \infty.$$

Then
$$\int_0^{2\|L_\beta + L_\theta\|_{P,4}} \sqrt{\log N_{[]}\left(\tilde{\epsilon}, \mathcal{N}_{Y,B}, \|\cdot\|_{P,2}\right)}d\tilde{\epsilon} < \infty.$$

So up to now we have proved the finite entropy integral property of $\mathcal{N}_{Y,B}$, and hence that of $\mathcal{N}$.

Lemma 3.2 of Lee [21]

Before presenting Lemma 3.2 of Lee [21], we quote the definition of multi-dimensional fBm from Lee [21].

Let $d \in \mathbb{N}$. A stochastic process $B_H = \{B_H(t) = (B_H^{(1)}(t), ..., B_H^{(d)}(t)), t \geq 0\}$ defined on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, is called a d-dimensional fBm of Hurst parameter $H \in (0,1)$, starting from $B_H(0) \in \mathbb{R}^d$, and associated matrix Λ, if it satisfies the following conditions: The process B_H is a continuous Gaussian process with initial condition $B_H(0) \quad \mathbb{P} - a.s.$ and its covariance function is given by

$$\mathrm{Cov}(B_H(t), B_H(s)) = \mathbb{P}((B_H(t) - B_H(0)(B_H(t) - B_H(0))^T)) = \Lambda_H(s,t)\Lambda,$$

for any $s, t \geq 0$, where Λ is a $d \times d$ positive definite matrix and

$$\Lambda_H(s,t) \equiv \frac{1}{2}(t^{2H} + s^{2H} - |t-s|^{2H}).$$

Without loss of generality, we assume that the diagonal entries of Λ are all ones. Also, it is assumed that B_H is adapted to the filtration $(\mathcal{F}_t)_{t\geq 0}$. We will say that B_H is a d-dimensional fBm with associated data $(B_H(0), H, \Lambda)$.

Lemma C.2.4 (Lemma 3.2 of Lee [21])**.** *Let $h \in (0,\infty)$ be a constant. For $n \in \mathbb{N}$, let ν_n be defined as follows:*

$$\nu_n = \sup\{|B_H(s) - B_H((n-1)h)| : (n-1)h \leq s \leq nh\},$$

where $B_H(\cdot)$ is a d-dimensional fBm with data $(0, H, \Lambda)$ and Hurst parameter $H \in (1/2, 1)$. Then, for any $\gamma \in (0,\infty)$ and $n \in \mathbb{N}$,

$$\mathbb{P}[e^{\gamma\nu_n}] \leq 4de^{1/2\gamma^2 d^2 h^{2H}},$$

where d is the dimension of the fBm B_H.

C.3 Proof of the functional set (II) for Chapter 2.3

Lemma C.3.1. $\mathcal{Q} \equiv \{\beta Z_H(\theta) + \gamma X : |\beta| \leq \beta_M, \theta \in [0,\theta_M], |\gamma| \leq \gamma_M\}$ *has finite integral of $L_2(P)$ entropy with bracketing.*

To prove $\mathcal{Q}$ has finite integral of $L_2(P)$ entropy with bracketing, we will use the Lipschitz property of $\beta Z_H(\theta) + \gamma X$.

$$\begin{aligned}
&|(\beta_1 Z_H(\theta_1) + \gamma_1 X) - (\beta_2 Z_H(\theta_2) + \gamma_2 X)| \\
&\qquad = |(\beta_1 - \beta_2) Z_H(\theta_1) + \beta_2 (Z_H(\theta_1) - Z_H(\theta_2)) + (\gamma_1 - \gamma_2) X| \\
&\qquad \leq |\beta_1 - \beta_2||Z_H(\theta_1)| + |\beta_2||Z_H(\theta_1) - Z_H(\theta_2)| + |\gamma_1 - \gamma_2||X| \\
&\qquad \leq |\beta_1 - \beta_2||Z_H(\theta_1)| + \beta_M L|\theta_1 - \theta_2|^\alpha + |\gamma_1 - \gamma_2||X|
\end{aligned}$$

Same as in the previous section, the bracketing numbers of $\{\beta : |\beta| \leq \beta_M\}$,$\{\theta : \theta \in [0, \beta_M]\}$ and $\{\gamma : |\gamma| \leq \gamma_M\}$ are $N(\epsilon, [-\beta_M, \beta_M], d_\beta)$, $N(\epsilon, [0, \theta_M], d_\theta)$ and $N(\epsilon, [-\gamma_M, \gamma_M], d_\gamma)$ respectively (with $d_\beta(\beta_1, \beta_2) = |\beta_1 - \beta_2|$, $d_\theta(\theta_1, \theta_2) = |\theta_1 - \theta_2|^\alpha$ and $d_\gamma(\gamma_1, \gamma_2) = |\gamma_1 - \gamma_2|$).

Then we can denote each of their brackets by $(l_i^\beta, u_i^\beta)(i = 1, ..., N(\epsilon, [-\beta_M, \beta_M], d_\beta))$, $(l_k^\theta, u_k^\theta)(k = 1, ..., N(\epsilon, [-\theta_M, \theta_M], d_\theta))$ and $(l_j^\gamma, u_j^\gamma)(j = 1, ..., N(\epsilon, [-\gamma_M, \gamma_M], d_\gamma))$. Notice that the brackets constructed have $|u_i^\beta - l_i^\beta| \leq \epsilon$, $|u_k^\theta - l_k^\theta|^\alpha \leq \epsilon$ and $|u_j^\gamma - l_j^\gamma| \leq \epsilon$. We also have $N(\epsilon, [-\beta_M, \beta_M], d_\beta) = \lceil \frac{2\beta_M}{\epsilon} \rceil$, $N(\epsilon, [0, \theta_M], d_\theta) = \lceil \frac{\theta_M}{\epsilon^{1/\alpha}} \rceil$ and $N(\epsilon, [-\gamma_M, \gamma_M], d_\gamma) = \lceil \frac{2\gamma_M}{\epsilon} \rceil$.

Then we can construct no more than

$$N(\epsilon, [-\beta_M, \beta_M], d_\beta) \cdot N(\epsilon, [0, \theta_M], d_\theta) \cdot N(\epsilon, [-\gamma_M, \gamma_M], d_\gamma)$$

brackets $\left(l_i^\beta (Z_H(\theta_0) + L(l_k^\theta - \theta_0)) + l_j^\gamma X, u_i^\beta (Z_H(\theta_0) + L(u_k^\theta - \theta_0)) + u_j^\gamma X\right)$ to cover $\mathcal{Q}$, and the $L_2(P)$ size of the bracket

$$\begin{aligned}
&\sqrt{P\left(u_i^\beta (Z_H(\theta_0) + L(u_k^\theta - \theta_0)) + u_j^\gamma X - \left(l_i^\beta (Z_H(\theta_0) + L(l_k^\theta - \theta_0)) + l_j^\gamma X\right)\right)^2} \\
\leq& \sqrt{3}\sqrt{P\left((u_i^\beta - l_i^\beta)(Z_H(\theta_0) - L\theta_0)\right)^2 + P\left(L(u_i^\beta u_k^\theta - l_i^\beta l_k^\theta)\right)^2 + P\left((u_j^\gamma - l_j^\gamma) X\right)^2} \\
\leq& \sqrt{3}\sqrt{P\left(\epsilon^2 (Z_H(\theta_0) - L\theta_0)^2\right) + P\left(L(u_i^\beta (u_k^\theta - l_k^\theta) + (u_i^\beta - l_i^\beta) l_k^\theta)\right)^2 + P\left(\epsilon^2 X^2\right)}
\end{aligned}$$

Since

$$P(Z_H(\theta_0) - L\theta_0)^2 \leq 2\left(PZ_H^2(\theta_0) + PL^2\theta_0^2)\right) \leq 2\left(PZ_H^2(\theta_0) + PL^2\theta_M^2)\right),$$

$$P\left(L(u_i^\beta(u_k^\theta - l_k^\theta) + (u_i^\beta - l_i^\beta)l_k^\theta)\right)^2 \leq 2P\left(L^2(u_i^\beta)^2(u_k^\theta - l_k^\theta)^2\right) + 2P\left(L^2(u_i^\beta - l_i^\beta)^2(l_k^\theta)^2\right)$$
$$\leq 2PL^2\beta_M^2\epsilon^{2/\alpha} + 2PL^2\theta_M^2\epsilon^2,$$

The $L_2(P)$ size of the bracket

$$\begin{aligned}
&\leq\sqrt{6}\sqrt{\left(PZ_H^2(\theta_0) + 2\theta_M^2 PL^2 + PX^2\right)\epsilon^2 + \beta_M^2 PL^2\epsilon^{2/\alpha}} \\
&\leq\sqrt{6}\sqrt{\left(PZ_H^2(\theta_0) + 2\theta_M^2 PL^2 + PX^2 + \beta_M^2 PL^2\right)\epsilon^{2/\alpha}} \\
&\equiv L_Q\cdot\epsilon^{1/\alpha}.
\end{aligned}$$

by the fact that $\alpha < H < 1$.

Then we have

$$\begin{aligned}
&N_{[]}\left(L_Q\cdot\epsilon^{1/\alpha}, \mathcal{Q}, \|\cdot\|_{P,2}\right) \leq \lceil\frac{2\beta_M}{\epsilon}\rceil\lceil\frac{\theta_M}{\epsilon^{1/\alpha}}\rceil\lceil\frac{2\gamma_M}{\epsilon}\rceil, \\
&N_{[]}\left(L_Q\cdot\epsilon, \mathcal{Q}, \|\cdot\|_{P,2}\right) \leq \lceil\frac{2\beta_M}{\epsilon^\alpha}\rceil\lceil\frac{\theta_M}{\epsilon}\rceil\lceil\frac{2\gamma_M}{\epsilon^\alpha}\rceil, \\
&\log N_{[]}\left(L_Q\cdot\epsilon, \mathcal{Q}, \|\cdot\|_{P,2}\right) \lesssim \log\frac{1}{\epsilon} \lesssim \frac{1}{\epsilon}, \\
&\int_0^1 \sqrt{\log N_{[]}\left(L_Q\cdot\epsilon, \mathcal{Q}, \|\cdot\|_{P,2}\right)}d\epsilon < \infty.
\end{aligned}$$

Then

$$\int_0^{L_Q} \sqrt{\log N_{[]}\left(\tilde{\epsilon}, \mathcal{Q}, \|\cdot\|_{P,2}\right)}d\tilde{\epsilon} < \infty.$$

So up to now we have proved $\mathcal{Q}$ has finite entropy integral of $L_2(P)$ bracketing.

Appendix D

Exchangeability of differentiation and expectation

The exchangeability of differentiation and expectation is to be justified when we take derivatives of $\mathbb{M}(\eta)$ w.r.t. β. Those derivatives w.r.t. γ and second derivatives w.r.t. β and γ can be handled in exactly the same way.

The derivative of $\mathbb{M}(\eta)$ w.r.t. β can be broken into two parts, that of $P(\beta Z_H(\theta)+\gamma X)N(\tau)$ and that of $\int_0^\tau \log\left[PY(u)\exp(\beta Z_H(\theta)+\gamma X)\right]dPN(u)$. We deal with them separately.

For the first item, consider any $\beta_1, \beta_2 \in [0, \beta_M]$,

$$
\begin{aligned}
&\left(P(\beta_1 Z_H(\theta)+\gamma X)N(\tau) - P(\beta_2 Z_H(\theta)+\gamma X)N(\tau)\right)/(\beta_1-\beta_2) \\
&\qquad\qquad = P(\beta_1-\beta_2)Z_H(\theta)N(\tau)/(\beta_1-\beta_2) = PZ_H(\theta)N(\tau),
\end{aligned}
$$

For the second item,

$$
\begin{aligned}
&\int_0^\tau \log\left[PY(u)\exp(\beta Z_H(\theta)+\gamma X)\right]dPN(u) \\
&\quad = \int_0^\tau \log\left[P\exp(\beta Z_H(\theta)+\gamma X - e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\int_0^u \lambda_0(s)ds)\right]dPN(u). \\
&\quad = \int_0^\tau \log\left[P\exp(\beta Z_H(\theta)+\gamma X - e^{\beta_0 Z_H(\theta_0)+\gamma_0 X}\int_0^u \lambda_0(s)ds)\right]\lambda_0(u)s^{(0)}(\theta_0,u)du,
\end{aligned}
$$

where we used (2.7) in the last equality.

For continuous type random variables $Z_H(\theta_0)$ with P.D.F. f_Z and X with P.D.F. f_X, the integrand can be expressed as a double integral by f_Z and f_X,

$$\log\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\exp(\beta z+\gamma x-e^{\beta_0 z+\gamma_0 x}\int_0^u\lambda_0(s)ds)f_Z(z)f_X(x)dzdx\right]\lambda_0(u)s^{(0)}(\theta_0,u).$$

All the functions involved here are continuous and has continuous derivatives w.r.t. β and γ, it follows that the exchangeability of differentiation and expectation is justified by [47] as long as $P[Z_H(\theta_0)Y(u)\exp(\beta Z_H(\theta_0)+\gamma X)]<\infty$ for any $(\beta,\gamma,u)\in[-\beta_M,\beta_M]\times[-\gamma_M,\gamma_M]\times[0,\tau]$.

For discrete type random variable $Z_H(\theta_0)$ with P.M.F. (probability mass function) $P(Z_H(\theta_0)=z_i)=p_{Z,i}, i=1,2,...$ and continuous type random variable X with P.D.F. f_X, the integrand can be expressed as

$$\log\left[\int_{-\infty}^{\infty}\sum_{i=1}^{\infty}\exp(\beta z_i+\gamma x-e^{\beta_0 z_i+\gamma_0 x}\int_0^u\lambda_0(s)ds)p_{Z,i}f_X(x)dx\right]\lambda_0(u)s^{(0)}(\theta_0,u).$$

By Theorem 7.9 and Theorem 7.11 of Rudin [42], the exchangeability holds if

$$\lim_{n\to\infty}\sup_{|\beta|\le\beta_M}\left|\sum_{i=n+1}^{\infty}z_i\exp(\beta z_i+\gamma x-e^{\beta_0 z_i+\gamma_0 x}\int_0^u\lambda_0(s)ds)p_{Z,i}\right|=0.$$

Since $z_i\exp(\beta z_i+\gamma x-e^{\beta_0 z_i+\gamma_0 x}\int_0^u\lambda_0(s)ds)p_{Z,i}$ is monotonically increasing w.r.t. β, it suffices to require that

$$\lim_{n\to\infty}\sum_{i=n+1}^{\infty}z_i\exp(\beta z_i+\gamma x-e^{\beta_0 z_i+\gamma_0 x}\int_0^u\lambda_0(s)ds)p_{Z,i}=0$$

for $|\beta|=\beta_M$ and any $|\gamma|\le\gamma_M$. It is guaranteed also by $P[Z_H(\theta_0)Y(u)\exp(\beta Z_H(\theta_0)+\gamma X)]<\infty$ for any $(\beta,\gamma,u)\in[-\beta_M,\beta_M]\times[-\gamma_M,\gamma_M]\times[0,\tau]$.

For discrete type random variables $Z_H(\theta_0)$ with P.M.F. $P(Z_H(\theta_0)=z_i)=p_{Z,i}, i=1,2,...$ and X with P.M.F. $P(X=x_j)=p_{X,j}, j=1,2,...$, the integrand can be expressed as

$$\log\left[\sum_{j=1}^{\infty}\sum_{i=1}^{\infty}\exp(\beta z_i+\gamma x_j-e^{\beta_0 z_i+\gamma_0 x_j}\int_0^u\lambda_0(s)ds)p_{Z,i}p_{X,j}\right]\lambda_0(u)s^{(0)}(\theta_0,u).$$

The exchangeability follows similarly as the previous paragraph, as long as $P[Z_H(\theta_0)Y(u)\exp(\beta Z_H(\theta_0)+\gamma X)] < \infty$ for any $(\beta, \gamma, u) \in [-\beta_M, \beta_M] \times [-\gamma_M, \gamma_M] \times [0, \tau]$, which follows by Cauchy–Schwartz Inequality using Assumptions 2.3.1 (3).