UNIVERSITY OF CALIFORNIA,
IRVINE

Statistical Considerations for Healthspan-Focused Research: A Conditional Modeling Approach

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Statistics

by

Spencer Matthews

Dissertation Committee:
Bin Nan, Ph.D., Chair
Daniel Gillen, PhD.
Arkajyoti Saha, PhD.

2026

DEDICATION

To Isabelle, and our baby boy Lewis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Bin Nan for his tireless hours of patience as I worked my way through many complex topics. I additionally am grateful for input from and collaboration with Dr. Daniel Gillen, Dr. Joshua Grill, and Dr. Arkajyoti Saha. Further, the work in this document would not have been possible without the lessons learned while working with fantastic collaborators, including Dr. Michelle McDonnell, Dr. Tahseen Mozaffar, Dr. Namita Goyal, Isela Hernandez, and Dr. Pawel Jankowski.

I am grateful to Dr. Maria Corrada for sharing The 90+ Study dataset. I am also grateful to Dr. Christina Wolfson for sharing the CHSA dataset.

My wife, Isabelle has supported me every step of the way. I wouldn't be here without her. And countless friends and family members who have also done the same. I stand on the shoulders of the amazing people I'm privileged to know and associate with.

The work is supported in part by NIH grant RF1 AG075107 and UCI Alzheimer's Disease Research Center Grant P30 AG066519, and by NSF grant DMS 2412746.

The CSHA was supported by the Seniors Independence Research Program, through the National Health Research and Development Program (NHRDP) of Health Canada (project 6606-3954-MC[S]). The progression of dementia project within the CSHA was supported by Pfizer Canada through the Health Activity Program o f the Medical Research Council of Canada and the Pharmaceutical Manufacturers Association of Canada; by the NHRDP (project 6603-1417-302[RJ); by Bayer; and by the British Columbia Health Research Foundation (projects 38 [93-2] and 34 [96-1]).

The NACC database is funded by NIA/NIH Grant U24 AG072122. NACC data are contributed by the NIA-funded ADRCs: P30 AG062429 (PI James Brewer, MD, PhD), P30 AG066468 (PI Oscar Lopez, MD), P30 AG062421 (PI Bradley Hyman, MD, PhD), P30 AG066509 (PI Thomas Grabowski, MD), P30 AG066514 (PI Mary Sano, PhD), P30 AG066530 (PI Helena Chui, MD), P30 AG066507 (PI Marilyn Albert, PhD), P30 AG066444 (PI David Holtzman, MD), P30 AG066518 (PI Lisa Silbert, MD, MCR), P30 AG066512 (PI Thomas Wisniewski, MD), P30 AG066462 (PI Scott Small, MD), P30 AG072979 (PI David Wolk, MD), P30 AG072972 (PI Charles DeCarli, MD), P30 AG072976 (PI Andrew Saykin, PsyD), P30 AG072975 (PI Julie A. Schneider, MD, MS), P30 AG072978 (PI Ann McKee, MD), P30 AG072977 (PI Robert Vassar, PhD), P30 AG066519 (PI Frank LaFerla, PhD), P30 AG062677 (PI Ronald Petersen, MD, PhD), P30 AG079280 (PI Jessica Langbaum, PhD), P30 AG062422 (PI Gil Rabinovici, MD), P30 AG066511 (PI Allan Levey, MD, PhD), P30 AG072946 (PI Linda Van Eldik, PhD), P30 AG062715 (PI Sanjay Asthana, MD, FRCP), P30 AG072973 (PI Russell Swerdlow, MD), P30 AG066506 (PI Glenn Smith, PhD, ABPP), P30 AG066508 (PI Stephen Strittmatter, MD, PhD), P30 AG066515 (PI Victor Henderson, MD, MS), P30 AG072947 (PI Suzanne Craft, PhD), P30 AG072931 (PI Henry Paulson, MD, PhD), P30 AG066546 (PI Sudha Seshadri, MD), P30 AG086401 (PI Erik Roberson, MD, PhD), P30 AG086404 (PI Gary Rosenberg, MD), P20 AG068082 (PI Angela Jefferson,

PhD), P30 AG072958 (PI Heather Whitson, MD), P30 AG072959 (PI James Leverenz, MD).

Also, thank you to Wiley Online Publishing for allowing me to reproduce my work "Regression for Left-Truncated and Right-Censored Data: A Semiparametric Sieve Likelihood Approach" in this document.

VITA

Spencer Matthews

EDUCATION

Doctor of Philosophy in Statistics — **2026**
University of California, Irvine — *Irvine, California*

Masters of Science in Statistics — **2023**
University of California, Irvine — *Irvine, California*

Bachelor of Science in Statistics — **2020**
Brigham Young University — *Provo, Utah*

RESEARCH EXPERIENCE

Graduate Research Assistant (Dr. Bin Nan) — **2022-2026**
University of California, Irvine — *Irvine, California*

Research Assistant (Dr. Brian Hartman) — **2019-2021**
Brigham Young University — *Provo, Utah*

TEACHING EXPERIENCE

Course Instructor, STATS 67 — **Spring 2024**
University of California, Irvine — *Irvine, California*

Teaching Assistant — **2021-2023**
University of California, Irvine — *Irvine, California*

Course Assistant, STAT 121 — **2020**
Brigham Young University — *Provo, Utah*

Teaching Assistant, STAT 121 — **2019**
Brigham Young University — *Provo, Utah*

REFEREED JOURNAL PUBLICATIONS

Regression for Left-Truncated and Right-Censored Data: A Semiparametric Sieve Likelihood Approach **2026**
Statistics in Medicine

Is More Better? Multicenter Analysis of the Incidence and Mechanisms of Multiple Pelvic Fixation in Adult Spinal Deformity Surgery **2025**
The Spine Journal

Machine Learning in Ratemaking, an Application in Commercial Auto Insurance **2022**
Risks

mSHAP: SHAP values for Two-Part Models **2022**
Risks

SOFTWARE

ltrc `https://github.com/srmatth/ltrc`
R Programming package for fitting semiparametric linear models to left-truncated and right-censored data.

mSHAP `https://github.com/srmatth/mshap`
R Programming package that extends the SHAP methodology to two-part multiplicative models.

ABSTRACT OF THE DISSERTATION

Statistical Considerations for Healthspan-Focused Research: A Conditional Modeling Approach

By

Spencer Matthews

Doctor of Philosophy in Statistics

University of California, Irvine, 2026

Bin Nan, Ph.D., Chair

Cohort studies of aging-related diseases often produce event-time data that are subject to left-truncation, since participants enroll after some baseline age while being free of the disease of interest, in addition to right-censoring due to variable follow-up. This dissertation develops a suite of semiparametric methods, organized around a two-part conditional modeling framework, for drawing valid inference from such data and for quantifying healthspan (the portion of life spent in good health) as an endpoint in aging research. The first project proposes a semiparametric sieve likelihood approach for fitting a linear regression model to a response subject to both left-truncation and right-censoring; the resulting estimators are shown to be consistent, asymptotically normal, and semiparametrically efficient, and the method is illustrated on data from the Canadian Study of Health and Aging and The 90+ Study. The second project introduces the Golden Health Index (GHI), the expected proportion of post-baseline life lived disease-free; we estimate the GHI and its covariate associations via a two-part model that accounts for left-truncation, and establish its large-sample properties, and applied to the National Alzheimer's Coordinating Center data, the framework shows that APOE $\varepsilon 4$ allele status is negatively associated with the GHI, whereas lifespan is positively associated with the GHI after an initial dip. The third project closes a theoretical gap by establishing the asymptotic normality of sieve estimates in a broad range of non-

and semi-parametric regression models in which a smooth unknown function is among the parameters of interest, thereby justifying likelihood-based inference in the settings of the first two projects.

Chapter 1

Introduction

The diagram in Figure 1.1 represents the traditional trajectory of most age-related diseases. A person is healthy at some age A_0, and eventually dies at time T. Between this baseline age and death, the person either develops the disease of interest at time S, or does not have it before death. We let Δ_S be an indicator of whether or not the person contracted the disease during lifetime.

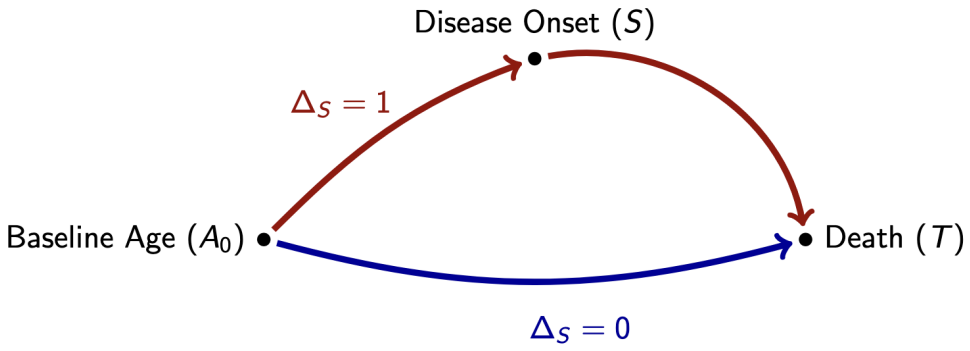


Figure 1.1: A visual representation of the conditional modeling approach employed for age-related diseases.

This work is motivated by studying data generated by this type of process. To do so, we employ a conditional modeling approach composed of two parts. The first part is to model

the probability that an individual develops the disease of interest before dying at T, i.e., $P(\Delta_S = 1|T, X)$, here X denotes a set of covariates. The second part is to model the distribution of S conditional on an individual developing the disease during their lifetime, i.e., $P(S \leq s|T, X, \Delta_S = 1)$.

Studies that consider a fixed baseline age A_0, such as a prospective cohort study with a disease-free requirement for entry, are subject to left-truncation. To aid in the analysis of left-truncated data, Chapter 2 presents a novel method for fitting semiparametric linear models to data that is both left-truncated and right-censored. This work extends the prior work of Ding and Nan [2011a] to include the left-truncated case. As a part of this work, we developed a code package, and it is available online for easy implementation, allowing users to rapidly fit this type of model. In addition to a comparison data analysis on The Canadian Study of Health and Aging, we preview Chapter 3 by modeling a derived quantity called the Healthspan Proportion on the left-truncated 90+ Dataset from the University of California, Irvine [Corrada et al., 2010].

Using the work in Chapter 2, we are able to model the time to disease given that the disease occurs under left-truncation. This is the second part of our two-part conditional modeling framework mentioned above. Chapter 3 addresses the overall modeling framework and tackles the challenge of deriving an interpretable metric from this two-part conditional modeling framework. In this chapter, we introduce the Golden Health Index, which provides an interpretation of the average proportion of disease-free life after some baseline age A_0. We motivate this development with a detailed data analysis of the National Alzheimer's Disease Coordinating Center Uniform Data Set (NACC UDS), focused on the impact of the APOE $\varepsilon 4$ allele on the Golden Health Index. We propose a logistic regression model for Δ_S and a semiparametric linear regression model for $\text{logit}(S/T)$, respectively, in the two-part modeling framework, and derive an appropriate estimator of the Golden Health Index. This estimator takes into account left-truncation and is shown to have desirable properties, both

theoretically and in simulation. When applied to NACC UDS, the results indicate that the Golden Health Index is strongly associated with the number of APOE $\varepsilon 4$ alleles. This association comes from both model parts and remains in effect across a variety of sensitivity analyses.

Chapter 4 returns to the work of Chapter 2 to complete the theory of joint asymptotic normality of both finite- and infinite-dimensional parameters. In Chapter 2 the asymptotic normality was only obtained for the finite-dimensional parameter, where the asymptotic variance is derived from the efficient score function. In practice, however, a sieve likelihood estimation problem is an approximation of the finite-dimensional maximum likelihood estimation problem. It is tempting to instead estimate the asymptotic variance by inverting the information matrix as is standard in parametric likelihood estimation. Such an estimator is presented for comparison in the simulation study results in Chapter 2. While it is similar to the efficient score function based estimate, its validity needs a theoretical justification. Additionally, the asymptotic normality of the infinite-dimensional parameter estimator is desirable because it can provide a justified uncertainty quantification for the estimator.

The heart of Chapter 4 is Theorem 4.1, which gives the asymptotic normality for contrasts of sieve likelihood parameter estimates. Under standard conditions that are typically satisfied in these sort of estimation problems, we show that the parametric parameters and sieve coefficient parameters are jointly normally distributed, and use this to derive the pointwise normality of the infinite-dimensional parameter estimator. Having developed this theorem, we are able to derive an estimator for the variance of all coefficients in a sieve likelihood problem, as well as give pointwise confidence bands on the infinite-dimensional parameter. It also paves the way for future work to allow for more flexible modeling methods in our two-part framework.

Together, these three chapters form a tight body of work around the two-part modeling framework described at the beginning of the chapter. They allow researchers to ask and

answer questions about healthspan-based issues, giving them the appropriate modeling tools and the corresponding estimating procedures.

Chapter 2

Regression for Left-Truncated and Right-Censored Data: A Semiparametric Sieve Likelihood Approach

2.1 Introduction

There has been significant research effort in understanding age-related diseases, such as Alzheimer's disease and dementia, in recent decades. A key method in understanding these diseases is to design a prospective cohort study, where study participants are recruited and followed, with data collected at regular intervals. Oftentimes the ultimate outcome of interest is the time at which the disease occurs. Such studies typically require that participants who enroll in the study meet an age requirement while being disease free; for instance, The 90+ Study only enrolled dementia-free individuals at age 90 or older [Corrada et al., 2010].

This method of participant enrollment induces left-truncation into the resulting dataset in addition to right-censoring.

Left-truncation, often called delayed entry in longitudinal studies, occurs when individuals who have already developed the condition of interest prior to the commencement of the study are not observed or are excluded from the analysis. Simply ignoring left-truncation introduces bias, as individuals who develop the disease earlier are systematically underrepresented, potentially skewing the findings toward a healthier or longer-lived population. Appropriately accounting for left-truncation is therefore essential to obtain valid inference, particularly in longitudinal studies of aging and chronic disease.

Another common feature of event time data is right-censoring, which occurs when some participants in the study do not experience the event of interest within the specified study time period, or are lost to follow-up before the event can be recorded. As a result, the exact time of the event for such an individual remains unknown, and the available information is limited to the fact that the event has not occurred up to the last observed time.

While fitting parametric models or semiparametric hazard regression models, e.g., the Cox model, to data subject to left-truncation and right-censoring has been studied extensively [Chen and Yi, 2021], the case of fitting a semiparametric linear model to such data is underdeveloped, and a linear model is a natural choice for the transformed outcome of interest in the example of The 90+ Study. For some reference time A_0, often referred to as time zero in survival analysis which can be either deterministic or random, define $(\tilde{Y}, \tilde{T}, \tilde{C})$ as the underlying event time, left-truncation time (i.e., study entry time), and right-censoring time since A_0, respectively, where $\tilde{Y}$ can be left-truncated by $\tilde{T}$ and right-censored by $\tilde{C}$. Given a well-defined continuous and monotonic transformation $\xi(\cdot)$, let $Y^* = \xi(\tilde{Y})$, $T = \xi(\tilde{T})$ and $C = \xi(\tilde{C})$, then Y^* can be left-truncated by T and right-censored by C. Define $Y = Y^* \wedge C$, where $a \wedge b = \min(a, b)$, and $\Delta = 1(Y^* \leq C)$, where $1(\cdot)$ is an indicator function. We also define $a \vee b = \max(a, b)$. Let X denote a set of covariates. Variables (Y, Δ, X, T) become

observed when $Y > T$ and otherwise unobservable, giving the so-called left-truncated data that are also subject to right-censoring, which follow a conditional distribution given $Y > T$.

For the underlying transformed event time of interest Y^*, consider the following linear regression model given covariates X:

$$Y^* = X^{\mathrm{T}}\beta + e, \tag{2.1}$$

where $\beta \in R^d$, d is a fixed integer, and the error e follows some unknown distribution F_e with density f_e. Thus model (2.1) is a semiparametric model. If $\xi(\cdot) = \log(\cdot)$, then this becomes the so-called accelerated failure time model as in Kalbfleisch and Prentice [2002] Chapter 7. We extend the terminology to model (2.1) with any known transformation ξ.

Another model of potential interest similar to the above model (2.1) is the linear transformation model, where the function ξ is unknown but the distribution of e is given, see Wang and Wang [2015] and references of earlier work therein. Incorporating left-truncation in this setting would be an interesting future work. We consider a known transformation ξ here for a more straightforward interpretation of β.

One method for fitting a semiparametric model like (2.1) is rank regression. Adichie [1967] was among the first who used residual rank statistics for fitting regression parameters, and Bhattacharya et al. [1983] extended the approach to left-truncated data in their work in astronomy. Lai and Ying [1991b, 1992b, 1994] used these ideas in exploring rank statistics with left-truncated and right-censored data as they studied the behavior and asymptotic properties of such a regression. Kim and Lai [2000] pointed out various issues with practical implementations of rank regression methods, the most important being the large computational cost associated with fitting these models. In an attempt to overcome some of these issues, they introduced an adaptive M-estimation method using efficient scores for censored and truncated regression models in which they used splines to approximate their estimat-

ing equations. While they showed their method to have nice theoretical properties, the implementation is dense and no code package is readily available. Also note that another estimation method for model (2.1) under left-truncation and right-censoring is the length-biased sample approach, where it is assumed that the left-truncation time follows a uniform distribution [Chen and Qiu, 2023, Ning et al., 2014, Chen, 2024]. Following Wang [1989], in contrast, the method we consider in this work does not make any distributional assumption about the left-truncation time.

A more computationally efficient approach, and statistically efficient as well, is to fit model (2.1) using a sieve likelihood. The method of sieves provides a general framework for estimating infinite-dimensional parameters by approximating them with a sequence of finite-dimensional spaces whose complexity increases with the sample size. First explored by Geman and Hwang [1982], sieve estimation has developed into an effective tool for semiparametric and nonparametric inference. It allows the flexibility of infinite-dimensional models while preserving the tractability of likelihood-based estimation. In the sieve likelihood approach, the likelihood function is maximized over a sequence of finite-dimensional approximations of infinite-dimensional parameters with dimension growing at an appropriate rate to ensure both consistency and asymptotic efficiency [Shen and Wong, 1994, Chen, 2007]. This framework has been successfully applied in a variety of semiparametric contexts, including censored and truncated data models, [van der Vaart, 2023] where traditional likelihood methods can be difficult to implement due to the presence of infinite-dimensional nuisance parameters.

In this work, we consider a sieve likelihood approach for fitting model (2.1) to left-truncated and right-censored data where the unknown infinite-dimensional nuisance parameter, the logarithm of the hazard function of e in our case, is estimated using splines. Our method builds on the work of Ding and Nan [2011a] for bundled parameters where only right-censoring was considered. This approach allows for a semiparametric likelihood estimation for a frequently

encountered problem in practice, which not only provides a simple numerical implementation but also yields desirable asymptotic properties for the regression coefficient estimates, namely asymptotical normality and semiparametric efficiency. We demonstrate this approach through simulation studies and applications to the Canadian Study of Health and Aging [The Canadian Study of Health and Aging Working Group, 1994a,b] and The 90+ Study [Corrada et al., 2010].

2.2 Sieve Likelihood Approach

A sieve approach approximates the true unknown infinite-dimensional parameter with a sequence of increasingly complex parametric models. In our case, we propose to use a sequence of B-spline basis expansions to approximate the logarithm of the unknown hazard function of e in model (2.1).

Denote the hazard function of e by λ_e, and let $g = \log \lambda_e$. For a finite interval $[a, b]$ and K knots placed within the interval: $a < \kappa_1 < \ldots < \kappa_K < b$, we approximate g by

$$g(s) = \log \lambda_e(s) \approx \sum_{k=1}^{K} \gamma_k B_k(s),$$

where $B_k, k = 1, \ldots, K$, are cubic B-spline basis functions. We chose cubic B-splines so that the second derivative is continuous, but other splines may be considered as well. For further discussion see De Boor [2001] Chapter 9.

Since $e = Y^* - X^{\mathrm{T}}\beta$, g is a composite function of β. Thus (β, g) form a pair of bundled parameters. [Ding and Nan, 2011a] Denote $\epsilon_{\beta,i} = Y_i - X_i^{\mathrm{T}}\beta$ and $\tau_{\beta,i} = T_i - X_i^{\mathrm{T}}\beta$ that denotes the left-truncation time on the residual scale. Assume Y^* is independent of (C, T) given X. Then for independent and identically distributed observations $\{(y_i, \delta_i, x_i, t_i) : y_i > t_i, i =$

$1, \ldots, n\}$, we have the following likelihood function:

$$\prod_{i=1}^{n} [\exp\{g(\epsilon_{\beta,i})\}]^{\Delta_i} \exp\left[-\int_{\tau_{\beta,i}}^{\epsilon_{\beta,i}} \exp\{g(u)\}du\right],$$

which is then approximated by the sieve likelihood:

$$L_n(\beta,\gamma) = \prod_{i=1}^{n}\left[\exp\left\{\sum_{k=1}^{K}\gamma_k B_k(\epsilon_{\beta,i})\right\}\right]^{\Delta_i} \exp\left[-\int_{\tau_{\beta,i}}^{\epsilon_{\beta,i}} \exp\left\{\sum_{k=1}^{K}\gamma_k B_k(u)\right\}du\right].$$

See Appendix B for the detailed construction of the density function of the observed data (Y, Δ, X, T) given $Y > T$ which is obtained by using the left-truncation observation condition and leads to the above likelihood function for data with left-truncation and right-censoring. Then we obtain the following sieve log-likelihood function:

$$\ell_n(\beta,\gamma) = \sum_{i=1}^{n}\left(\Delta_i \sum_{k=1}^{K}\gamma_k B_k\left(\epsilon_{\beta,i}\right) - \int_{\tau_{\beta,i}}^{\epsilon_{\beta,i}} \exp\left\{\sum_{k=1}^{K}\gamma_k B_k(s)\right\}ds\right). \tag{2.2}$$

Denote $\dot{B}_k(s) = dB_k(s)/ds$ and $\ddot{B}_k = d^2B_k(s)/ds^2$. Taking partial derivatives of the above sieve log-likelihood function (2.2), we arrive at the following sieve score functions:

$$S_n(\beta,\gamma) = \left(\frac{\partial\ell_n(\beta,\gamma)}{\partial\beta}, \frac{\partial\ell_n(\beta,\gamma)}{\partial\gamma}\right)^{\mathrm{T}}, \tag{2.3}$$

with

$$\begin{aligned}\frac{\partial\ell_n(\beta,\gamma)}{\partial\beta} = \sum_{i=1}^{n} X_i^{\mathrm{T}}\Bigg(&-\Delta_i\sum_{k=1}^{K}\gamma_k\dot{B}_k\left(\epsilon_{\beta,i}\right) + \exp\left\{\sum_{k=1}^{K}\gamma_k B_k\left(\epsilon_{\beta,i}\right)\right\} \\ &- \exp\left\{\sum_{k=1}^{K}\gamma_k B_k\left(\tau_{\beta,i}\right)\right\}\Bigg)\end{aligned} \tag{2.4}$$

and

$$\frac{\partial \ell_n(\beta,\gamma)}{\partial \gamma_k} = \sum_{i=1}^{n} \left(\Delta_i B_k\left(\epsilon_{\beta,i}\right) - \int_{\tau_{\beta,i}}^{\epsilon_{\beta,i}} B_k(s) \exp\left\{ \sum_{k=1}^{K} \gamma_k B_k(s) \right\} ds \right). \tag{2.5}$$

Furthermore, the observed sieve information matrix is

$$I_n(\beta,\gamma) = \begin{pmatrix} \frac{\partial^2 \ell_n(\beta,\gamma)}{\partial\beta\partial\beta^{\mathrm{T}}} & \frac{\partial^2 \ell_n(\beta,\gamma)}{\partial\beta\partial\gamma^{\mathrm{T}}} \\ \frac{\partial^2 \ell_n(\beta,\gamma)}{\partial\gamma\partial\beta^{\mathrm{T}}} & \frac{\partial^2 \ell_n(\beta,\gamma)}{\partial\gamma\partial\gamma^{\mathrm{T}}} \end{pmatrix}, \tag{2.6}$$

with

$$\begin{aligned}
\frac{\partial^2 \ell_n(\beta,\gamma)}{\partial\beta\partial\beta^{\mathrm{T}}} &= \sum_{i=1}^{n} X_i X_i^{\mathrm{T}} \left(\Delta_i \sum_{k=1}^{K} \gamma_k \ddot{B}_k\left(\epsilon_{\beta,i}\right) - \exp\left\{ \sum_{k=1}^{K} \gamma_k B_k\left(\epsilon_{\beta,i}\right) \right\} \sum_{k=1}^{K} \gamma_k \dot{B}_k(\epsilon_{\beta,i}) \right. \\
&\qquad \left. + \exp\left\{ \sum_{k=1}^{K} \gamma_k B_k\left(\tau_{\beta,i}\right) \right\} \sum_{k=1}^{K} \gamma_k \dot{B}_k\left(\tau_{\beta,i}\right) \right), \\
\frac{\partial^2 \ell_n(\beta,\gamma)}{\partial\beta\partial\gamma_k} &= \sum_{i=1}^{n} X_i^{\mathrm{T}} \left(-\Delta_i \dot{B}_k\left(\epsilon_{\beta,i}\right) + B_k\left(\epsilon_{\beta,i}\right) \exp\left\{ \sum_{k=1}^{K} \gamma_k B_k\left(\epsilon_{\beta,i}\right) \right\} \right. \\
&\qquad \left. - B_k\left(\tau_{\beta,i}\right) \exp\left\{ \sum_{k=1}^{K} \gamma_k B_k\left(\tau_{\beta,i}\right) \right\} \right), \\
\frac{\partial^2 \ell_n(\beta,\gamma)}{\partial\gamma_k\partial\gamma_l} &= \sum_{i=1}^{n} \left(- \int_{\tau_{\beta,i}}^{\epsilon_{\beta,i}} B_k(s) B_l(s) \exp\left\{ \sum_{m=1}^{K} \gamma_m B_m(s) \right\} ds \right).
\end{aligned}$$

We consider a gradient search algorithm by setting (2.4) and (2.5) equal to zero and solving for β and γ, which leads to the sieve maximum likelihood estimates $(\hat{\beta}_n, \hat{\gamma}_n)$. Similar to Ding and Nan [2011a] we can see that, under the regularity conditions given in the next section, the sieve log-likelihood function (2.2) is locally concave in a neighborhood of true parameters (β_0, g_0). Thus, we start the search algorithm from multiple randomly selected initial values and set $(\hat{\beta}_n, \hat{\gamma}_n)$ to be the converging result that gives the maximum value of the sieve

log-likelihood (2.2). Specifically in our numerical examples, we generate 10 random initial values each obtained by adding a vector of random noises to the naive ordinary least squares estimator of β obtained from complete observations, where the standard deviation of each random noise is chosen to be three times the standard error of the corresponding estimated coefficient from the naive model. We multiply the standard error by three to avoid getting caught too close to the naive estimate for all random starts. We sample initial γ values from the standard normal distribution. To avoid singular information matrix during search iterations, we only use diagonal elements of the sieve information matrix (2.6), with the final full information matrix evaluated at $(\hat{\beta}_n, \hat{\gamma}_n)$. The method works well in our simulation studies presented in Section 2.4.

2.3 Asymptotic Results

Define $\epsilon_0 = Y - X^{\mathrm{T}}\beta_0$ and $\tau_0 = T - X^{\mathrm{T}}\beta_0$, where β_0 is the true parameter. Let λ_0 be the hazard function of $e_0 = Y^* - X^{\mathrm{T}}\beta_0$ and Λ_0 be the corresponding cumulative hazard function. Following Lai and Ying [1991b] we assume e_0 is independent of (C, T, X), a reasonable assumption for a linear regression model, which implies that Y^* is independent of (C, T) given X, an assumption of conditional independent left-truncation and right-censoring that we need for obtaining the joint density function of the observed data (Y, Δ, X, T) with $Y > T$ in Appendix B. To obtain desirable asymptotic properties of the proposed estimator, we introduce the same regularity conditions in Ding and Nan [2011a] with a few additional assumptions for the left-truncation.

(A.1) The true parameter β_0 belongs to the interior of a compact set $\mathcal{B} \subset R^d$.

(A.2) (a) The covariate X takes values in a bounded subset $\mathcal{X} \subset R^d$; (b) $E(XX^{\mathrm{T}})$ is non-singular.

(A.3) (a) There is a constant $b < \infty$ such that, for some constant δ, $\text{pr}(\epsilon_0 > b \mid X) \geq \delta > 0$ almost surely with respect to the probability measure of X. This implies that $\Lambda_0(b) \leq -\log \delta < \infty$; (b) We also assume $\text{pr}(\tau_0 < b \mid X) \geq 1 - \delta_1$ for some constant $\delta_1 \in (0, 1)$ almost surely, and $\text{pr}(\tau_0 > \epsilon_0 \mid X) \leq \delta_2$ for some constant $\delta_2 \in [0, 1)$ almost surely.

(A.4) The error e_0's density f_0 and its derivative $\dot{f}_0$ are bounded and

$$\int \left\{ \dot{f}_0(s)/f_0(s) \right\}^2 f_0(s) ds < \infty.$$

(A.5) (a) The conditional density of C given (T, X) and its derivative are uniformly bounded for all possible values of $T \in \mathcal{T} \subseteq R$ and $X \in \mathcal{X}$, that is,

$$\sup_{t \in \mathcal{T}, x \in \mathcal{X}} f_{C|T,X}(y \mid t, x) \leq K_1, \qquad \sup_{t \in \mathcal{T}, x \in \mathcal{X}} \left| \dot{f}_{C|T,X}(y \mid t, x) \right| \leq K_2$$

for all $y \in \mathcal{Y} \subseteq R$ with some constants $K_1, K_2 > 0$; (b) Similarly, the conditional density of T given X and its derivative are uniformly bounded for all possible values of $X \in \mathcal{X}$, that is,

$$\sup_{x \in \mathcal{X}} f_{T|X}(t \mid x) \leq K_3, \qquad \sup_{x \in \mathcal{X}} \left| \dot{f}_{T|X}(t \mid x) \right| \leq K_4$$

for all $t \in \mathcal{T}$ with some constants $K_3, K_4 > 0$; (c) Both $f_{C|T,X}$ and $f_{T,X|Y>T}$ are free of parameters β and λ_e.

(A.6) Let $\mathcal{G}^p$ denote the collection of bounded functions g on $[a, b]$ with bounded derivatives $g^{(j)}, j = 1, \ldots, k$. The kth derivative $g^{(k)}$ satisfies the following Lipschitz continuity condition:

$$\left| g^{(k)}(s) - g^{(k)}(t) \right| \leq L|s - t|^m \qquad \text{for } s, t \in [a, b],$$

where k is a positive integer and $m \in (0, 1]$ such that $p = k + m \geq 3$, $L < \infty$ is an

unknown constant, $a = \inf_{y,x}(y - x^{\mathrm{T}}\beta_0) \vee \inf_{t,x}(t - x^{\mathrm{T}}\beta_0) > -\infty$ and b is defined in condition (A.3). The true log hazard function of the error e_0, $g_0(\cdot) = \log \lambda_0(\cdot)$, belongs to $\mathcal{G}^p$.

(A.7) For some $\eta \in (0,1)$, $u^{\mathrm{T}}\mathrm{Var}(X \mid \epsilon_0)u \geq \eta u^{\mathrm{T}} E(XX^{\mathrm{T}} \mid \epsilon_0)u$ almost surely for all $u \in R^d$.

The modified regularity conditions to those given in Section 3 of Ding and Nan [2011a] are in (A.3)(b) and (A.5). Condition (A.3)(b) guarantees that we observe left-truncated data almost surely. Condition (A.5)(a) is a natural extension of the condition in Ding and Nan [2011a] for the conditional density function of C. Condition (A.5)(b) plays a similar role as (A.5)(a). Condition (A.5)(c) contains a noninformative censoring assumption for right-censored data and a similar but less stringent noninformative truncation assumption made by Lai and Ying [1992a] for left-truncated data. Condition (A.5)(c) simplifies the likelihood function to the desirable form we use in this article. Note that while $f_{T,X|Y>T}$ being free of β and λ_e holds in our setup (as shown in Wang [1989]), we include this assumption in our conditions for clarity. When $\sup\{t \in \mathcal{T}\} \leq \inf\{y \in \mathcal{Y}\}$, left-truncation does not play any role so we only deal with right-censored data and the problem reduces to that in Ding and Nan [2011a]. On the other hand, when $\mathrm{pr}(Y^* \leq C) = 1$, the problem reduces to left-truncation only.

Consider the collection of functions

$$\mathcal{H}^p = \{\zeta(\cdot,\beta)\colon \zeta(s,x,\beta) = g(\psi(s,x,\beta)), g \in \mathcal{G}^p, s \in [a,b], x \in \mathcal{X}, \beta \in \mathcal{B}\},$$

where $\psi(s,x,\beta) = s - x^{\mathrm{T}}(\beta - \beta_0)$ and $\mathcal{G}^p$ is defined in Condition (A.6). Here ζ is a composite function of g composed with ψ. Note that $\zeta(s,x,\beta_0) = g(s)$. Then for $\zeta(\cdot,\beta) \in \mathcal{H}^p$ we define

the following norm:

$$\|\zeta(\cdot,\beta)\|_2 = \left[\int_{\mathcal{X}}\int_a^b \left\{g\left(s - x^{\mathrm{T}}(\beta-\beta_0)\right)\right\}^2 d\Lambda_0(s)dF_X(x)\right]^{1/2}.$$

For any $\theta_1 = (\beta_1, \zeta_1(\cdot,\beta_1))$ and $\theta_2 = (\beta_2, \zeta_2(\cdot,\beta_2))$ in the space of $\Theta^p = \mathcal{B} \times \mathcal{H}^p$, define the following distance:

$$d(\theta_1,\theta_2) = \left\{|\beta_1-\beta_2|^2 + \|\zeta_1(\cdot,\beta_1) - \zeta_2(\cdot,\beta_2)\|_2^2\right\}^{1/2}. \tag{2.7}$$

This is the same distance used by Ding and Nan [2011a] in their treatment of the right-censored case.

Let $\mathcal{G}_n^p$ be the space of polynomial splines of order $p \geq 1$. For details on this space see Schumaker [2007] Chapter 4. Denote

$$\mathcal{H}_n^p = \{\zeta(\cdot,\beta)\colon \zeta(s,x,\beta) = g(\psi(s,x,\beta)), g \in \mathcal{G}_n^p, s \in [a,b], x \in \mathcal{X}, \beta \in \mathcal{B}\}$$

and $\Theta_n^p = \mathcal{B} \times \mathcal{H}_n^p$. Clearly, $\mathcal{H}_n^p \subseteq \mathcal{H}_{n+1}^p \subseteq \cdots \subseteq \mathcal{H}^p$ for all $n \geq 1$. The sieve estimator $\hat{\theta}_n = (\hat{\beta}_n, \hat{\zeta}_n(\cdot,\hat{\beta}_n))$, where $\hat{\zeta}_n(s,x,\hat{\beta}_n) = \hat{g}_n(s - x^{\mathrm{T}}(\hat{\beta}_n - \beta_0))$, is the maximizer of the log-likelihood over the sieve space Θ_n^p. The following theorem gives the convergence rate of the proposed estimator $\hat{\theta}_n$ to the true parameter $\theta_0 = (\beta_0, \zeta_0(\cdot,\beta_0)) = (\beta_0, g_0)$.

Theorem 2.1. *Let K_n be the number of internal knots in the spline space and $K_n = O(n^\nu)$, where ν satisfies $1/\{2(1+p)\} < \nu < 1/(2p)$ with p being the smoothness parameter defined in condition (A.6). Suppose conditions (A.1)–(A.7) hold. Then*

$$d(\hat{\theta}_n, \theta_0) = O_p\left\{n^{-\min(p\nu,(1-\nu)/2)}\right\},$$

where $d(\cdot,\cdot)$ is defined in (2.7).

Remark 2.1. *The sieve space $\mathcal{G}_n^p$ does not have to be restricted to the B-spline space. It can be any sieve space as long as the estimator $\hat{\theta}_n \in \mathcal{B} \times \mathcal{H}_n^p$ satisfies the conditions of Theorem 1 in Shen and Wong [1994]. Our choice of the cubic B-spline space is primarily motivated by its simplicity of numerical implementation, which is a tremendous advantage of the proposed approach over existing numerical methods for fitting accelerated failure time models for data subject to right-censoring and left-truncation.*

The proof of Theorem 2.1 follows that of Ding and Nan [2011b] by verifying the conditions of Shen and Wong [1994], and is given in Appendix C.3. The main difference is controlling the additional complication brought by $\tau_\beta = T - X^{\mathrm{T}}\beta$ in the likelihood function, which is managed by Condition (A.3)(b) as well as desirable properties of linear and indicator functions. Theorem 2.1 implies that if $\nu = 1/(1+2p)$, $d(\hat{\theta}_n, \theta_0) = O_p(n^{-p/(1+2p)})$ which is the optimal convergence rate in the nonparametric regression setting. Although the overall convergence rate is slower than $n^{-1/2}$, the next theorem states that the proposed estimator of the regression parameter is still asymptotically normal and semiparametrically efficient.

Theorem 2.2. *Given the following efficient score function obtained by Lai and Ying [1992a] for the censored and truncated linear model:*

$$l^*_{\beta_0}(Y, \Delta, X, T) = \int \{X - E[X \mid \epsilon_0 \geq s, \tau_0 \leq s]\} \left\{ \frac{-\dot{\lambda}_0}{\lambda_0}(s) \right\} dM(s),$$

where

$$M(s) = \Delta 1(\epsilon_0 \leq s) 1(\epsilon_0 \geq \tau_0) - \int_{-\infty}^{s} 1(\epsilon_0 \geq u) 1(\tau_0 \leq u) \lambda_0(u) du$$

is the failure counting process martingale, suppose that the conditions in Theorem 2.1 hold and $I(\beta_0) = P\{l^*_{\beta_0}(Y, \Delta, X, T)^{\otimes 2}\}$ is nonsingular. Then

$$n^{1/2}(\hat{\beta}_n - \beta_0) = n^{-1/2} I^{-1}(\beta_0) \sum_{i=1}^{n} l^*_{\beta_0}(Y_i, \Delta_i, X_i, T_i) + o_p(1) \to N(0, I^{-1}(\beta_0))$$

in distribution.

The proof of Theorem 2.2 closely follows the work of Ding and Nan [2011a,b], and Zhao et al. [2017] by checking Conditions (A1) through (A6) of Ding and Nan [2011a], and is given in Appendix C.4. Verifying their Condition (A3) requires searching for the least-favorable submodel, which leads to the same efficient score function found by Lai and Ying [1992a]. The main difference to Ding and Nan [2011a] is again the inclusion of the left-truncation indicator, which has minimum impact under the additional regularity conditions (A.3)(b) and (A.5)(b) and desirable properties of indicator and linear functions.

The following theorem provides the consistency result of the variance estimator based on the efficient score function given in Theorem 2.2. Denote $Pf = \int f(y, \delta, x, t) dP(y, \delta, x, t)$ and $P_n f = n^{-1} \sum_{i=1}^{n} f(Y_i, \Delta_i, X_i, T_i)$, where P is a probability measure, and P_n is an empirical probability measure.

Theorem 2.3. *Suppose the conditions in Theorem 2.2 hold. Denote*

$$\hat{l}^*_{\hat{\beta}_n}(Y, \Delta, X, T) = \int \{X - \bar{X}(s; \hat{\beta}_n)\}\{-\dot{\hat{g}}_n(s)\} d\hat{M}(s),$$

where

$$\bar{X}(s; \hat{\beta}_n) = \frac{P_n\{X 1(Y - X^{\mathrm{T}}\hat{\beta}_n \geq s) 1(T - X^{\mathrm{T}}\hat{\beta}_n \leq s)\}}{P_n\{1(Y - X^{\mathrm{T}}\hat{\beta}_n \geq s) 1(T - X^{\mathrm{T}}\hat{\beta}_n \leq s)\}}$$

and

$$\begin{aligned}\hat{M}(s) = \Delta 1(T - X^{\mathrm{T}}\hat{\beta}_n \leq Y - X^{\mathrm{T}}\hat{\beta}_n \leq s) \\ - \int_{-\infty}^{s} 1(Y - X^{\mathrm{T}}\hat{\beta}_n \geq u) 1(T - X^{\mathrm{T}}\hat{\beta}_n \leq u) \exp\{\hat{g}_n(u)\} du.\end{aligned}$$

Then

$$P_n\left\{\hat{l}^*_{\hat{\beta}_n}(Y, \Delta, X, T)^{\otimes 2}\right\} \to P\left\{l^*_{\beta_0}(Y, \Delta, X, T)^{\otimes 2}\right\}$$

in probability.

In Theorem 2.3, $\bar{X}(s, \hat{\beta}_n)$ estimates $E[X \mid \epsilon_0 \geq s, \tau_0 \leq s]$ from Theorem 2.2. Once again, the proof of Theorem 2.3 follows Ding and Nan [2011b] with a major difference of including the truncation indicator, which can be easily dealt with. The details of this proof are given in Appendix C.5.

2.4 A Simulation Study

2.4.1 Simulation Setup

We perform an extensive simulation study to assess the validity of the proposed methodology. To compare with existing methods, we simulate datasets following the exact same setup as used in Ning et al. [2014] from the following linear model:

$$\log(\tilde{Y}) = 1 + \beta_1 X_1 + \beta_2 X_2 + e, \tag{2.8}$$

where $\beta_1 = 0.5$ and $\beta_2 = 1$. We generate binary covariate X_1 from a Bernoulli(0.5) distribution, continuous covariate X_2 from a Uniform(0,1) distribution, truncation time $\tilde{T}$ from a Uniform(0, c) distribution, censoring times $\tilde{C}$ from a Uniform(0, c) distribution, and the error term e from either a Uniform(-0.5, 0.5) distribution or a $N(0, 1/12)$ distribution. Then $\tilde{Y}$ is determined from the above equation (2.8). The constant c is chosen to give a 15%, or 30%, or 50% censoring rate. The simulated censoring and truncation distributions reflect a real study setup where censoring or study entry could occur at any point during the fixed duration of the study. Simulations are run with sample sizes of 100 and 200 at the three previously mentioned levels of censoring. Simulation results are directly compared to those reported in Ning et al. [2014].

We further expand the simulation to demonstrate the robustness of the sieve likelihood method under different error distributions. Specifically, the error term comes from one of three distributions: (a) generalized extreme value distribution, or Gumbel distribution, (b) a 50% even mixture of $N(0, 1/24)$ and $N(0, 1/8)$ distributions, and (c) a 50% even mixture of $N(0, 1/12)$ and $N(-0.3, 1/24)$ distributions. These distributions were chosen to have variances the same as or similar to those in Ning et al. [2014], while exhibiting more complex nuisance parameter curves.

In both parts, the number of internal knots, K, used to estimate the nuisance parameter for each setup was determined via cross-validation and ranged between 1 and 4. Since left-truncation removes data, we keep simulating independent data until accumulated non-truncated data reach the desired sample size, then treat the sample size as a known quantity. For each simulation setup, we generate 1,000 datasets. We use 10 randomly selected initial values as described in Section 2.2 when fitting each model.

2.4.2 Simulation Results

Table 2.1 shows a comparison of the results of the first part of our simulation study to those presented in Section 4 of Ning et al. [2014]. The comparison methods are the semi-rank-based estimator U_{II} proposed in Ning et al. [2014], the inverse weighted estimating equation estimator U_{IW} proposed in Shen et al. [2009], the Buckley-James-type estimator U_{BJ} from Ning et al. [2011], and the rank-based estimator U_{LT} proposed by Lai and Ying [1991b]. Among these methods, U_{II}, U_{IW}, and U_{BJ} all use the length-biased approach, meaning that they assume a uniform truncation distribution, whereas U_{LT} applies to general left-truncated data.

The bias of our approach is similar to that of the other presented approaches, and in almost all cases the variance of our estimator is smaller than that of all comparison approaches

presented in Ning et al. [2014]. The one clear exception is in the 50% censoring case when the sample size is 100. This is expected, as the high censoring rate leads to a much smaller effective sample size, undermining the effectiveness of the method of sieves. In Ning et al. [2014], the authors state that based on the outcome of their simulation study, there is "no uniform best estimation method in terms of statistical efficiency among the estimation methods considered." We see that our sieve likelihood approach does indeed yield a uniformly most efficient estimator, as long as the number of observed events is large enough for splines to be used effectively. This aligns with the semiparametric efficiency theory from which the estimator was derived.

Results from the second part of our simulation study are presented in Table 2.2. For both β_1 and β_2 the biases are negligibly low, although the magnitude of the bias is often larger for β_2 which likely stems from the lower signal-to-noise ratio for that covariate.

The variance estimates across the three methods of estimation are very similar, especially at larger sample sizes. While the theory presented in this work only guarantees that the variance estimator Var^1 obtained from the efficient score function converges to the truth, we see that the variance estimator Var^2 obtained by inverting the information matrix of (β, γ) is close to the value of Var^1, and they both are close to the empirical variance Var^3. Further, all three variances are reduced approximately by half as the sample size is doubled, which is expected from the large sample theory. The confidence intervals created using Var^2 provide proper coverages. The same pattern holds irrespective of error distributions.

Recall that we use 10 random starting points to fit the model to each dataset, and select the result with the highest log-likelihood value. Given the known truth, we can examine the convergence of these 10 random starts for each simulated dataset. For both β_1 and β_2, we find about $80 \sim 95\%$ of the randomly selected initial values lead to convergence somewhere close to the truth.

Table 2.1: Comparison of the Sieve Likelihood method to other approaches as reported in Ning et al. [2014] in simulations. Results of other approaches are taken directly from Ning et al. [2014], where U_{II} is the estimator proposed in Ning et al. [2014]; U_{IW} is the inverse weighted estimating equation from Shen et al. [2009]; U_{BJ} is the Buckley-James-type estimating equation from Ning et al. [2011]; U_{LT} is the rank-based approach presented in Lai and Ying [1991b]. For further details on these approaches see Ning et al. [2014]. All variances and biases are multiplied by 10^3.

		Sieve Likelihood				U_{II}				U_{IW}				U_{BJ}				U_{LT}			
N	Cens. %	β_1		β_2		β_1		β_2		β_1		β_2		β_1		β_2		β_1		β_2	
		Bias	Var[1]	Bias	Var[1]	Bias	Var	Bias	Var	Bias	Var	Bias	Var	Bias	Var	Bias	Var	Bias	Var	Bias	Var
									$e \sim \text{Unif}(-0.5, 0.5)$												
100	15%	1.6	1.4	5.2	5.0	-4.0	1.9	-6.0	6.2	-7.0	4.9	-5.0	14.9	0.0	4.6	2.0	2.6	1.0	8.5	-11.0	4.1
100	30%	2.2	2.1	0.7	7.5	8.0	2.6	27.0	9.0	-7.0	5.3	-6.0	18.5	-5.0	4.5	-1.0	14.4	0.0	3.5	-6.0	13.7
100	50%	-0.7	3.1	9.5	11.1	-18.0	4.9	-33.0	16.1	-8.0	7.6	-5.0	24.7	15.0	11.3	36.0	36.0	0.0	6.1	-60.0	35.0
200	15%	1.3	0.6	-1.1	2.1	-3.0	0.7	-7.0	2.4	-3.0	2.5	0.0	7.6	1.0	2.3	-1.0	6.4	1.0	0.9	-1.0	2.9
200	30%	-0.2	1.0	-1.9	3.5	-11.0	1.0	-25.0	3.9	-3.0	2.5	-5.0	8.8	0.0	2.3	2.0	7.4	1.0	1.3	-3.0	4.5
200	50%	-3.6	1.6	0.5	5.5	-13.0	1.8	-28.0	6.1	0.0	3.4	0.0	11.4	15.0	4.4	36.0	14.4	1.0	2.1	-14.0	10.4
									$e \sim N(0, 1/12)$												
100	15%	-0.5	4.1	8.8	13.4	-6.0	5.3	-19.0	14.6	-2.0	5.3	-11.0	15.4	-1.0	5.2	-2.0	13.9	4.0	7.4	-4.0	19.0
100	30%	-1.4	4.4	6.5	14.6	-18.0	5.8	-43.0	18.5	-2.0	5.8	-13.0	18.0	1.0	4.5	-5.0	15.2	1.0	8.6	-14.0	22.8
100	50%	29.8	3331.6	4.0	18.6	-29.0	7.6	-44.0	26.2	-3.0	7.6	-12.0	25.6	-30.0	9.4	-65.0	34.6	-2.0	10.8	-60.0	33.5
200	15%	-1.4	2.2	2.6	6.8	-10.0	2.4	-24.0	7.7	-2.0	2.6	-11.0	7.4	-1.0	2.4	-1.0	6.6	0.0	3.6	-3.0	9.4
200	30%	0.7	2.5	1.8	7.8	-21.0	2.6	-44.0	9.2	-2.0	2.9	-12.0	8.6	-1.0	2.5	2.0	7.2	-1.0	4.0	-6.0	11.2
200	50%	-0.2	3.0	6.7	9.9	-28.0	3.6	-31.0	13.5	-2.0	3.5	-9.0	12.3	-30.0	4.2	-52.0	14.6	-3.0	4.9	-12.0	15.6

Table 2.2: Simulation results, where $\beta_1 = 0.5$ and $\beta_2 = 1$. Error distributions: (a) Generalized Extreme Value Distribution, (b) $0.5N(0, 1/24) + 0.5N(0, 1/8)$, (c) $0.5N(0, 1/12) + 0.5N(-0.3, 1/24)$. Truncation probability was 60% to 90%, depending on error and censoring distributions. Var^1 comes from the efficient score function estimator in Theorem 2.3, Var^2 comes from inverting the information matrix for β and γ, and Var^3 is the empirical variance. The coverage probability is obtained using Var^2.

Error	Sample Size	Censoring %	Bias$\times 10^3$		$\text{Var}^1 \times 10^3$		$\text{Var}^2 \times 10^3$		$\text{Var}^3 \times 10^3$		95% Coverage	
			β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2	β_1	β_2
(a)	200	15%	0.9	12.7	1.5	4.8	1.8	5.5	1.7	5.3	93.4	93.3
		30%	0.5	13.4	1.6	5.1	1.9	6.1	1.7	6.1	93.8	91.5
		50%	2.7	15.6	1.7	5.9	2.0	7.3	2.2	6.9	91.2	92.4
	400	15%	-0.1	10.1	0.8	2.5	0.8	2.6	1.0	2.9	93.1	93.2
		30%	1.4	11.6	0.8	2.6	0.9	2.7	1.0	2.9	93.7	92.7
		50%	5.7	13.9	0.9	2.9	1.0	3.1	1.0	3.3	93.4	92.8
	800	15%	-0.7	10.0	0.4	1.2	0.4	1.3	0.4	1.4	94.5	93.0
		30%	1.3	9.4	0.4	1.3	0.4	1.3	0.4	1.5	94.6	93.2
		50%	5.3	15.2	0.4	1.4	0.5	1.5	0.5	1.6	92.7	91.2
(b)	200	15%	1.5	9.1	2.1	6.6	2.5	8.6	2.8	7.9	91.5	92.3
		30%	4.1	5.6	2.3	7.0	2.7	9.0	3.0	9.1	92.1	90.9
		50%	0.4	2.7	2.6	8.4	3.4	11.6	3.6	10.8	92.6	92.2
	400	15%	-4.3	-5.8	1.0	3.2	1.2	4.0	1.3	4.2	93.1	92.2
		30%	-0.9	-4.9	1.1	3.3	1.2	4.0	1.4	4.2	92.4	92.4
		50%	-3.6	-12.3	1.2	3.9	1.4	4.9	1.7	5.4	92.0	91.4
	800	15%	-1.8	-6.0	0.5	1.6	0.5	1.7	0.6	1.9	93.2	93.9
		30%	-3.4	-8.7	0.6	1.7	0.6	2.0	0.7	2.2	93.1	92.2
		50%	-3.4	-9.6	0.6	2.0	0.7	2.3	0.8	2.7	92.1	91.0
(c)	200	15%	1.7	2.2	2.1	6.4	3.0	10.4	2.5	9.2	92.5	89.8
		30%	2.0	3.3	2.2	6.9	4.4	13.3	3.0	9.5	91.2	90.3
		50%	3.4	3.4	2.5	8.5	5.0	13.2	3.3	12.6	91.8	89.8
	400	15%	-0.2	-2.1	1.1	3.3	1.4	4.5	1.4	4.0	92.1	92.7
		30%	-1.3	-4.2	1.1	3.6	1.4	5.2	1.4	4.9	92.4	91.2
		50%	0.0	-7.4	1.3	4.2	1.9	6.5	1.7	5.3	92.0	91.5
	800	15%	-0.1	-2.9	0.6	1.7	0.6	2.6	0.6	2.1	95.4	92.6
		30%	-0.2	-2.5	0.6	1.8	0.6	2.0	0.6	2.1	94.9	93.0
		50%	-1.0	-5.5	0.7	2.1	0.8	2.7	0.8	2.7	94.0	91.3

Along with the regression coefficient estimates, we also examine the estimates of the infinite-dimensional nuisance parameter, the logarithm of the hazard function, in simulations with a sample size of 800. Figure 2.1 summarizes the simulation results for each error distribution, including a sample of 100 individual estimates plotted with light-blue solid lines, the point-wise average over all 1,000 datasets plotted with a red solid line, and the ground truth plotted with a black dashed line. The bounds of the x-axis represent the middle 95% of the theoretical error distribution. We see from Figure 2.1 that the spline estimates of $\log(\lambda_0)$ perform reasonably well in simulations.

2.5 Real Data Applications

2.5.1 The Canadian Study of Health and Aging

The Canadian Study of Health and Aging (CSHA) is a multicenter epidemiologic study of dementia and related conditions among Canadians aged 65 years and older. [The Canadian Study of Health and Aging Working Group, 1994a,b] In the first phase (1991–1992), 10,263 participants were enrolled from a stratified random sample of community and institutional residents across Canada. Community-dwelling participants completed a brief cognitive screening, and those with lower scores (together with a random sample of higher scorers) underwent a more detailed clinical assessment following common evaluation practices. Participants were classified into major dementia subtypes using standardized clinical guidelines. In the follow-up phase (1996–1997), information on survival and date of death was obtained for all participants diagnosed with dementia at baseline.

Because participants entered the study only after dementia onset had occurred, the data are left-truncated, as individuals who developed dementia but died before enrollment were not observed. This means that the observed sample overrepresents longer survivors and under-

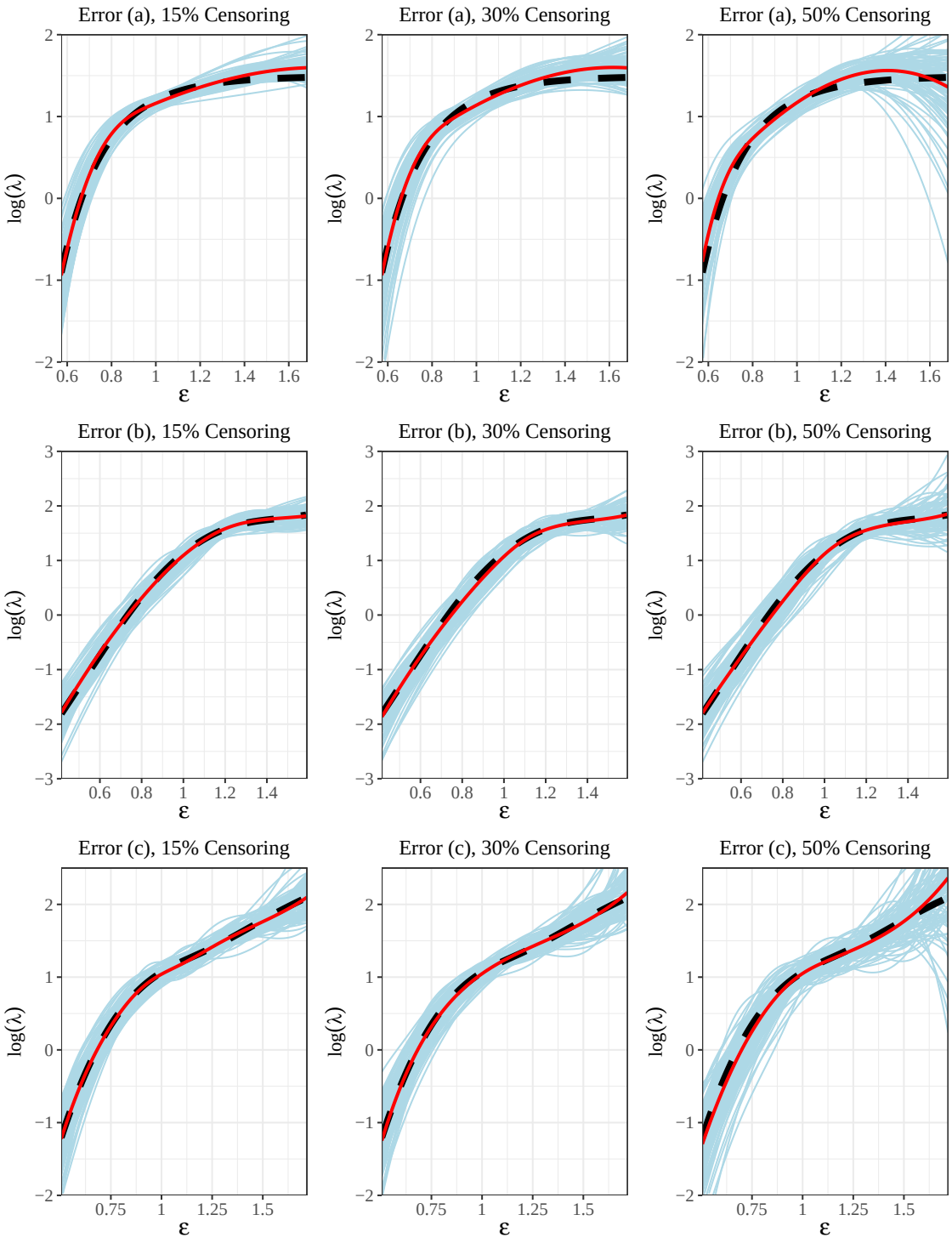


Figure 2.1: Plots showing the spline estimator for the log hazard function of each error distribution at 15%, 30%, and 50% censoring. 100 samples are plotted in light-blue solid lines, the point-wise average of all 1000 estimates is plotted in a red solid line and the true function plotted in a black dashed line. Error distributions: (a) Generalized Extreme Value Distribution, (b) $0.5N(0, 1/24) + 0.5N(0, 1/8)$, (c) $0.5N(0, 1/12) + 0.5N(-0.3, 1/24)$.

Table 2.3: Descriptive statistics of the 821 patients analyzed from the CHSA data. AD means Alzheimer's Disease, SD denotes the standard deviation, and * indicates only computed from non-censored subjects.

Variable	Mean (SD) or N (%)		
	Probable AD	Possible AD	Vascular Dementia
Age of Dementia Onset	82.3 (7.7)	81.0 (8.9)	78.5 (9.0)
Survival (post-diagnosis, in years)*	6.3 (3.9)	7.2 (5.6)	6.4 (6.4)
Censored Patients	85 (21.5%)	60 (23.8%)	34 (19.5%)
Sex			
Female	312 (79.0%)	171 (67.9%)	100 (57.5%)
Male	83 (21.0%)	81 (32.1%)	74 (42.5%)
Education (years)	8.6 (3.8)	8.6 (3.5)	8.6 (3.9)
≤ 8 years	166 (42.0%)	119 (47.2%)	83 (47.7%)
> 8 years	149 (37.7%)	93 (36.9%)	67 (38.5%)
Missing	80 (20.3%)	40 (15.9%)	24 (13.8%)

represents those with shorter post-dementia lifespans. In addition, some subjects were still alive at the end of follow-up, resulting in right-censored observations. Fitting a standard survival model to the observed data would yield biased estimates of survival time and covariate effects due to the left-truncation, underscoring the need for statistical methods that explicitly account for both truncation and censoring.

The present analysis focuses on analyzing the time from dementia onset, the reference age A_0, to death and parallels work done in Wolfson et al. [2001] and Ning et al. [2014] under the length-biased sampling paradigm. After excluding subjects with missing onset dates or other dementia types, 821 individuals remained (395 probable Alzheimer's, 252 possible Alzheimer's, and 174 vascular dementia). Demographic information for each group is shown in Table 2.3.

We fit a semiparametric linear model using the sieve likelihood approach, adjusting for age of dementia onset and sex. To avoid dropping data (and since the values are similar across dementia types) we did not include education in our model.

$$\begin{aligned}\log(\text{Time to Death (in days)}) &= \beta_1 1(\text{Diagnosis} = \text{Probable AD}) \\ &\quad + \beta_2 1(\text{Diagnosis} = \text{Vascular Dementia}) \\ &\quad + \beta_3(\text{Dementia Onset Age}) \\ &\quad + \beta_4 1(\text{Sex} = \text{Female}) + \text{Error}\end{aligned} \tag{2.9}$$

The form of the model is given in Equation 2.9 and parameter estimates are presented in Table 2.4. Five-fold cross-validation was used to determine the optimal number of knots, which was one in the final model.

From the fitted results, we conclude that even after adjusting for age at disease onset and sex, there is no evidence of a difference in survival times between the different dementia categories considered here. As a comparison, we also ran an unadjusted analysis and obtained estimates within the confidence intervals estimated by Ning, Qin, and Yu Ning et al. [2014]. In other words, we reached the same conclusions as in Ning, Qin, and Yu, Ning et al. [2014] where no difference was seen between the groups in the unadjusted analysis.

Table 2.4: Results of the sieve method for fitting model (2.9) to the left-truncated right-censored CHSA data.

Variable	Estimate	95% CI	P-Value
Type of Dementia			
Possible Alzheimer's	Referent	–	–
Probable Alzheimer's	0.094	(-0.043, 0.231)	0.177
Vascular Dementia	0.055	(-0.063, 0.175)	0.362
Age at Disease Onset	0.004	(-0.003, 0.011)	0.342
Sex			
Male	Referent	–	–
Female	0.044	(-0.077, 0.164)	0.477

2.5.2 The 90+ Study of Dementia

To further illustrate the proposed methodology, we also apply our method to the dataset of The 90+ Study of Alzheimer's disease and dementia. [Corrada et al., 2010] This dataset comes from an ongoing study at the Alzheimer's Disease Research Center at the University of California, Irvine. The study enrolled participants who were at least 90 years old and did not have Alzheimer's disease or dementia at baseline. These individuals were followed via yearly visits. The main feature of interest is a diagnosis of the cognitive state of each participant, which comes from a panel of experts after reviewing the results of neurologic examination and a battery of neuropsychological tests.

Understanding dementia onset during lifespan is of great interest. Because dementia may or may not occur during one's lifetime, a clear description of such a problem involves two components: the lifetime dementia-free probability and the age distribution of dementia onset during one's lifetime given dementia has occurred. These two components can be modeled separately conditional on lifetime, where lifetime becomes a covariate in both conditional models. It is well-known that complete-case analysis when a covariate is subject to right censoring is a valid approach to estimate regression coefficients, see Kong and Nan [2016], Kong et al. [2018], and references therein. Thus we apply complete-case analysis to the second component in the analysis of The 90+ Study where only individuals with available death time are included in the analysis. Specifically, we are interested in understanding the impact of demographic variables on the age of dementia onset relative to death time among participants who have developed dementia at some point during their lifetime by modeling the proportion of life after age 90 lived in a healthy state. For this purpose, we restrict our analysis to 349 available participants with complete information of age at death and age at dementia, out of the total of 921 patients available in the dataset.

These original 921 participants were from two main cohorts recruited at two different time

Table 2.5: Descriptive statistics of the 349 participants of The 90+ Study. SD, standard deviation; HP_{90}, the proportion of lifespan after 90 years of age that is lived without dementia.

Variable	Mean (SD) or N (%)
Enrollment Age	93.201 (2.745)
Dementia Diagnosis Age	96.343 (3.222)
Death Age	98.514 (3.374)
HP_{90}	0.737 (0.201)
Cohort	
First	278 (79.7%)
Second	71 (20.3%)
Sex	
Female	251 (71.9%)
Male	98 (28.1%)
Education	
Did Not Graduate College	208 (59.6%)
Graduated College	141 (40.4%)
Has Had a Stroke?	
No	211 (60.5%)
Yes	138 (39.5%)
Has Had Heart Disease?	
No	142 (40.7%)
Yes	207 (59.3%)

periods: The first cohort (599 out of 921 patients; 278 out of 349 in our sample of interest) is made up of survivors from the Leisure World Cohort Study (LCWS) and is primarily Caucasian, well-educated, upper middle-class, and mostly female. [Corrada et al., 2008, 2010] The LCWS is a health survey study of the residents of the Leisure World (now known as Laguna Woods Village), a large retirement community in Orange County, California. This cohort was recruited into the study between 2003 and 2006. The second cohort (322 out of 921 patients; 71 out of 349 in our sample of interest) was recruited between 2014 and 2017, using a variety of methods including community outreach, earned media, direct mailings, and referrals. [Melikyan et al., 2019] The second cohort has a higher proportion of males, a higher proportion of subjects who graduate college, and similar rates of stroke and heart disease. Since participants enrolled in the study at different ages that range from 90.1 to 103.0 years old, the data we consider in this analysis are left-truncated. We do not have right-censoring

in the dataset because ages at dementia onset are observed for all participants. Some details of the dataset are given in Table 2.5, where age at dementia ranges from 90.9 to 109.8 years and age at death ranges from 91.3 to 111.2 years.

To better answer the question of interest, we derive a new variable called the Healthspan Proportion for Age 90, HP_{90}, that is the number of healthy years after the reference age $A_0 = 90$ divided by the number of total years lived after age 90. In other words, HP_{90} is defined as [(Age at Dementia) - A_0] / [(Age at Death) - A_0]. In order to correctly fit the model, we must also scale the truncation time in this same way. Higher HP_{90} implies better quality of life for an individual in this population of oldest old. We consider the following linear regression model for the logit transformed HP_{90} and leave the error distribution unspecified:

$$\begin{aligned} \log\left(\frac{\text{HP}_{90}}{1-\text{HP}_{90}}\right) &= \beta_1(\text{Death Age - 90}) + \beta_2 1(\text{Cohort} = 2) \\ &\quad + \beta_3 1(\text{Sex} = \text{Female}) \\ &\quad + \beta_4 1(\text{Education} = \text{Graduated College}) \\ &\quad + \beta_5 1(\text{Stroke} = \text{Yes or Heart Disease} = \text{Yes}) + \text{Error}. \end{aligned} \tag{2.10}$$

Since HP_{90} potentially ranges from 0 to 1, the logit transformed HP_{90} can take any real value, and it is left-truncated because age at dementia is left-truncated. We apply our proposed method to fit a left-truncated regression model to the data. We use five-fold cross-validation to select the number of internal knots for the considered B-spline basis expansion, and ultimately choose two internal knots. From the regression results given in Table 2.6 we see that increased longevity is associated with higher HP_{90}, and the second cohort has higher HP_{90} than the first cohort.

Table 2.6: Results of the sieve method for fitting model (2.10) to the left-truncated data of The 90+ Study.

Variable	Estimate	95% CI	P-Value
Death Age	0.0578	(0.0221, 0.936)	0.0015
Cohort			
First	Referent	–	–
Second	0.3312	(0.0158, 0.6465)	0.0396
Sex			
Male	Referent	–	–
Female	-0.0660	(-0.3431, 0.2109)	0.6403
Education			
Did Not Graduate College	Referent	–	–
Graduated College	0.1638	(-0.0882, 0.4157)	0.2026
Comorbidity: Stroke or CHD			
No	Referent	–	–
Yes	0.0200	(-0.3332, 0.3733)	0.9116

2.6 Discussion

Left-truncation and right-censoring are often present in prospective cohort studies. Although linear regression models with unspecified error distributions are appealing due to simplicity, their applications to censored or truncated event time data are rather limited, mainly because of lacking easily implementable estimating methods with desirable statistical properties. In this work we show that the semiparametric sieve maximum likelihood approach for bundled parameters possesses nice properties in dealing with left-truncated and right-censored data in linear regression models. The method yields asymptotically efficient and normally distributed estimates for regression coefficients, with estimators obtained using an easily implementable gradient search algorithm. We expect this work would encourage the proper use of linear regression models in analyzing event time data that are subject to left-truncation and right-censoring.

Chapter 3

Modeling the Golden Health Index

3.1 Introduction

While lifespan has long been the dominant endpoint in aging research, increasing attention is now given to healthspan, the duration of life spent free from chronic disease or disability [Olshansky, 2018, Garmany et al., 2021]. Healthspan better reflects quality of life in older age and is a core focus of geroscience, an emerging field aimed at delaying age-related functional decline [Vaiserman and Lushchak, 2023, Sierra, 2016].

Researchers currently work with a range of measures to quantify patient health. These include objective event times, biomarker measurements, and physical performance metrics, as well as subjective patient-reported outcome measures (PROMs) that reflect lived experiences of mobility, memory, and well-being [Cella et al., 2007, Krogsgaard et al., 2021, Benson, 2023]. Several existing approaches integrate survival with health status, most notably quality-adjusted life years (QALYs) and quality-adjusted progression-free survival (QAPFS) [Weinstein and Stason, 1977, Bravo Vergel and Sculpher, 2008, Diaby et al., 2014]. These measures weight survival time by health-related quality-of-life utilities, typically derived from

PROMs. While valuable in many contexts, this formulation poses challenges for healthspan research. Utility weights are subjective and may change independently of clinically defined disease processes. Consequently, quality-adjusted metrics can blur the distinction between time lived with versus without disease. Many healthspan questions instead focus on the timing and duration of clinically defined disease-free states, motivating estimands that are directly tied to disease incidence and survival.

On the other hand, these endpoints are often analyzed using survival models that focus on relative hazards across covariates, such as the Cox model [Cox, 1972]. While such approaches yield important insights into the risk and timing of adverse events, they offer limited information about how long individuals remain in good health relative to their total lifespan.

To bridge these gaps, we introduce the Golden Health Index (GHI), a healthspan-oriented metric defined as the expected proportion of post-baseline life lived without a given disease. We develop a two-part semiparametric framework to estimate the GHI while accounting for left-truncation. Left-truncation is a common challenge in aging research, where individuals are often enrolled after some baseline age with a disease-free requirement. This selection criterion conditions on disease-free survival at entry and failure to account for it leads to biased results [Wang et al., 1986, Matthews and Nan, 2026].

We apply this framework to estimate dementia-free healthspan using longitudinal data from the National Alzheimer's Coordinating Center. In particular, we examine how healthspan varies by APOE genotype, a well-established genetic marker for dementia risk. The $\varepsilon 4$ allele has been associated with higher lifetime risk for Alzheimer's disease [Safieh et al., 2019, Fortea et al., 2024], but its impact on the duration of dementia-free life remains less well understood. This application illustrates how the GHI can uncover new dimensions of age-related disease trajectories in genetically diverse populations.

3.2 Motivation and Data Description

Understanding not only whether, but also when, age-related diseases occur is central to assessing health in an aging population. In this paper, we focus on dementia as a prototypical disease of aging and consider healthspan through the lens of the proportion of remaining life spent free from dementia. We are interested in the following questions. Of primary scientific interest, among U.S. senior adults who reach age 65 without a diagnosis of dementia, how is the number of apolipoprotein E (APOE) $\varepsilon 4$ alleles associated with the proportion of remaining lifespan spent dementia-free? Of secondary interest, how does APOE $\varepsilon 4$ status affect the probability of developing late-onset dementia during one's lifetime; does APOE $\varepsilon 4$ status also influence the timing of dementia onset if it were to occur?

To answer these questions and explore variation in dementia-free healthspan by APOE genotype, we use data from the National Alzheimer's Coordinating Center (NACC) [Center, 2023], a centralized data repository for studies at over 40 Alzheimer's Disease Research Centers (ADRCs) across the United States since 1984. The NACC Uniform Data Set (UDS) includes longitudinal clinical and neuropsychological data collected between 2005 and the present via yearly in-office visits [Beekly et al., 2007, Morris et al., 2006]. It captures individuals across the spectrum of cognitive function and includes APOE genotype data for a large number of participants, which serves as our primary predictor. For the analyses presented here, we use data from the NACC UDS as of the December 2024 freeze.

In the NACC UDS there are 30,122 individuals who have at least one visit at or after the age of 65 and who are dementia-free at the age of 65. Participants without a recorded date of death as of December 2024 (24,376 individuals) are excluded. Note that excluding individuals whose death times are censored leads to a complete case analysis that is a valid approach when time to death is treated as a covariate [Kong et al., 2018, Wang et al., 2024]. Because our analysis centers on APOE $\varepsilon 4$ status and includes education as one of the adjust-

ment variables, we also exclude individuals with missing APOE genotype (687) or education information (23). These criteria yield a final analytic sample of 5,036 participants from 38 ADRCs. Table 3.1 contains descriptive statistics for this cohort. A detailed comparison of included and excluded participants is provided in Table F.5 in Appendix F and further discussion is given in Section 3.5.2.

We define age 65 as the baseline for analysis to align with Medicare eligibility, which often marks the onset of more consistent health care utilization and systematic clinical assessment among older adults in the United States. Because participants may enroll in the study well after this baseline age, the resulting data are subject to substantial left-truncation. In the analytic sample, the average age at enrollment exceeds 78 years, underscoring the importance of accounting for delayed study entry.

3.3 Concept and Statistical Modeling

Consider a typical setup encountered when studying age-related diseases such as dementia. Let A_0 denote a fixed baseline age (e.g. 65 years), which we take as time zero. Define S as the number of years after A_0 at which an individual is diagnosed with the disease of interest, and T as the number of years after A_0 at which the individual dies. Note that S may be unobserved if the individual dies disease-free.

Let Δ_S be an indicator of whether the disease was diagnosed before death, and let L denote the number of years after A_0 at which an individual enters the study (i.e., the left-truncation time), so that the individual is included in the study only if $S > L$ when $\Delta_S = 1$ and $T > L$. We assume T is always observed (i.e., non-missing), and we observe a set of baseline covariates Z for each individual. For notational simplicity and modeling purposes, we also include T in Z.

Table 3.1: Descriptive statistics of the 5,036 patients from the NACC UDS stratified by the number of APOE $\varepsilon 4$ alleles. The values shown are Mean (SD) or N (%), where SD stands for standard deviation. Comorbidities included are traumatic brain injury, stroke, diabetes, arthritis, and heart attack at baseline.

Variable	Number of APOE $\varepsilon 4$ Alleles		
	Zero ($N = 3,291$)	One ($N = 1,525$)	Two ($N = 220$)
Enrollment Age	80.5 (7.6)	77.7 (7.2)	73.8 (6.1)
Dementia Before Death			
No	2,194 (66.7%)	757 (49.6%)	73 (33.2%)
Yes	1,097 (33.3%)	768 (50.4%)	147 (66.8%)
Death Age	87.9 (8.0)	85.6 (7.4)	81.3 (6.7)
Sex			
Female	1,753 (53.3%)	741 (48.6%)	89 (40.5%)
Male	1,538 (46.7%)	784 (51.4%)	131 (59.5%)
Years of Education	15.5 (3.1)	15.7 (3.1)	16.1 (3.1)
Married			
No	1,510 (45.9%)	574 (37.7%)	56 (25.5%)
Yes	1,781 (54.1%)	951 (62.4%)	164 (74.5%)
Comorbidity			
No	2,345 (71.3%)	1,117 (73.2%)	169 (76.8%)
Yes	946 (28.7%)	408 (26.8%)	51 (23.2%)
Race			
White	2,893 (87.9%)	1,314 (86.2%)	190 (86.3%)
Black	314 (9.5%)	192 (12.6%)	25 (11.4%)
Other	84 (2.6%)	19 (1.2%)	5 (2.3%)

To support a healthspan-oriented analysis and to answer the scientific questions of interest proposed in Section 3.2, we define the healthspan proportion (HP), an individual-level outcome indexed by disease status and baseline age, as the following:

$$\mathrm{HP}_{\mathrm{Disease},A_0} = \begin{cases} 1 & \text{if } \Delta_S = 0, \\ S/T & \text{if } \Delta_S = 1. \end{cases}$$

This quantity can be interpreted as the proportion of years lived after age A_0 that are free from the disease. Given T, a higher value of $\mathrm{HP}_{\mathrm{Disease},A_0}$ is more desirable.

We propose a two-part modeling approach to estimate the expectation of $\mathrm{HP}_{\mathrm{Disease},A_0}$, which

gives the GHI. Let $X^{(1)}$ be a subset of covariates from Z, which includes T and L. The first part models the conditional probability of receiving a disease diagnosis before death given $X^{(1)}$, denoted by $\mu_{X^{(1)}} = P(\Delta_S = 1|X^{(1)})$, using a logistic regression:

$$\log\left(\frac{\mu_{X^{(1)}}}{1-\mu_{X^{(1)}}}\right) = \beta_0^T X^{(1)}, \tag{3.1}$$

where β_0 is the vector of true regression coefficients. Since data with $S < L$ are not observable, which implies that the observed disease status Δ_S depends on L, we model the effect of L parametrically in (3.1) for the conditional probability $\mu_{X^{(1)}}$ by treating L as another covariate.

The second part models the expected value of a logit-transformed healthspan proportion conditional on $\Delta_S = 1$ using a linear regression. Specifically, let $Y = S/T$ when $\Delta_S = 1$, and we have

$$\log\left(\frac{Y}{1-Y}\right) = \omega_0^T X^{(2)} + \epsilon, \tag{3.2}$$

where $X^{(2)}$ is a subset of Z, ω_0 is the vector of true regression coefficients, and $\epsilon \sim F$ for some unspecified distribution F. Covariate vector $X^{(2)}$ should include T but not L. We do not include L in $X^{(2)}$ because left-truncation can be handled in a more robust way for model (3.2) under conditional independence of Y and L given Z (see Appendix D for regularity conditions). Specifically, under left-truncation, regression coefficients in model (3.2) can be estimated using a semiparametric sieve-likelihood approach for bundled parameters [Ding and Nan, 2011a, Matthews and Nan, 2026], where the logarithm of the hazard function of ϵ is approximated using spline functions.

Based on both models (3.1) and (3.2), for a given set of covariate values $Z = z$ with respective subsets $x^{(1)}$ and $x^{(2)}$, where L is set at 0 in $x^{(1)}$ which corresponds to the baseline time A_0,

we obtain the GHI as follows:

$$
\begin{aligned}
R &= (1 - \mu_{x^{(1)}}) + \mu_{x^{(1)}} E\left[Y \mid X^{(2)} = x^{(2)}, \Delta_S = 1\right] \qquad (3.3)\\
&= \left(1 - \frac{\exp\{\beta_0^T x^{(1)}\}}{1 + \exp\{\beta_0^T x^{(1)}\}}\right) + \left(\frac{\exp\{\beta_0^T x^{(1)}\}}{1 + \exp\{\beta_0^T x^{(1)}\}}\right)\left(\int \frac{\exp\{\omega_0^T x^{(2)} + u\}}{1 + \exp\{\omega_0^T x^{(2)} + u\}}\, dF(u)\right).
\end{aligned}
$$

The GHI R in the above (3.3) is our primary quantity of interest, which summarizes the expected proportion of life after A_0 spent in good health during one's lifetime given a set of baseline covariate values z that contains APOE $\varepsilon 4$ status. This directly answers our first question whether the number of APOE $\varepsilon 4$ alleles is associated with the proportion of advanced years spent free from dementia. Further, defining R in this way permits a decomposition of effects that allows us to determine if both the occurrence of dementia and the timing play a role in the association.

3.4 Estimation and Properties

Due to left-truncation, it may be impossible to estimate $E\left[Y \mid X^{(2)} = x^{(2)}, \Delta_S = 1\right]$ in (3.3). Similarly to the techniques used in Wang et al. [1986] and Ding [2010], we define two constant values $0 < a^* < b^* < 1$ to instead estimate the "trimmed" mean

$$
E\left[Y \mid X^{(2)} = x^{(2)}, \Delta_S = 1, a^* < Y < b^*\right],
$$

which also satisfies the requirement of bounded support for the splines used in the sieve-likelihood estimation procedure [Matthews and Nan, 2026] for model (3.2). We choose a^* close to 0 and b^* close to 1 and consider a fixed z (with $x^{(1)}$, and $x^{(2)}$) as above. Note that a^* is on the scale of the healthspan proportion, and thus is a truncation time scaled by death time. This means $a^* = l^*/t$ for a fixed truncation time l^* and death time t (both in years after A_0). For a consistent interpretation, we also use the value of l^* corresponding to a^*,

instead of 0, as the left-truncation time in $x^{(1)}$ for evaluating the GHI when predicting with model (3.1). We can then define the "trimmed" version of the GHI as

$$R^* = (1 - \mu_{x^{(1)}}) + \mu_{x^{(1)}} E\left[Y \mid X^{(2)} = x^{(2)}, \Delta_S = 1, a^* < Y < b^*\right], \tag{3.4}$$

which is estimable under left-truncation.

Let $\hat{\beta}_n$ be the maximum likelihood estimator of the logistic regression parameters in model (3.1) and $\hat{\omega}_m$ be the sieve likelihood estimator of the regression parameters in model (3.2), where $m = \sum_{i=1}^{n} \Delta_{S,i}$. We define $\hat{\epsilon}_{\hat{\omega}_m,m}$ as the observed residuals for a given $\hat{\omega}_m$. Denote $\hat{F}_{m,\hat{\omega}_m}(u)$ as the product-limit estimator of the cumulative distribution function F using observed residuals, taking into account left-truncation [Wang et al., 1986]. Also, let $a = \log(a^*/(1-a^*)) - \omega_0^T x^{(2)}$, $b = \log(b^*/(1-b^*)) - \omega_0^T x^{(2)}$, $\hat{a} = \log(a^*/(1-a^*)) - \hat{\omega}_m^T x^{(2)}$ and $\hat{b} = \log(b^*/(1-b^*)) - \hat{\omega}_m^T x^{(2)}$. Denote

$$\begin{aligned}
\hat{\mu}_{x^{(1)}} &= \frac{\exp\{\hat{\beta}_n^T x^{(1)}\}}{1 + \exp\{\hat{\beta}_n^T x^{(1)}\}}, \\
\theta &= E\left[Y \mid X^{(2)} = x^{(2)}, \Delta_S = 1, a^* < Y < b^*\right] = \int_a^b \frac{\exp\{\omega_0^T x^{(2)} + u\}}{1 + \exp\{\omega_0^T x^{(2)} + u\}}\, dF(u), \\
\hat{\theta} &= \int_{\hat{a}}^{\hat{b}} \frac{\exp\{\hat{\omega}_m^T x^{(2)} + u\}}{1 + \exp\{\hat{\omega}_m^T x^{(2)} + u\}}\, d\hat{F}_{m,\hat{\omega}_m}(u).
\end{aligned}$$

We estimate R^* in (3.4) by

$$\begin{aligned}
\hat{R}^* &= (1 - \hat{\mu}_{x^{(1)}}) + \hat{\mu}_{x^{(1)}} \hat{\theta} \\
&= \left(1 - \frac{\exp\{\hat{\beta}_n^T x^{(1)}\}}{1 + \exp\{\hat{\beta}_n^T x^{(1)}\}}\right) \\
&\qquad + \left(\frac{\exp\{\hat{\beta}_n^T x^{(1)}\}}{1 + \exp\{\hat{\beta}_n^T x^{(1)}\}}\right) \left(\int_{\hat{a}}^{\hat{b}} \frac{\exp\{\hat{\omega}_m^T x^{(2)} + u\}}{1 + \exp\{\hat{\omega}_m^T x^{(2)} + u\}}\, d\hat{F}_{m,\hat{\omega}_m}(u)\right).
\end{aligned} \tag{3.5}$$

The following theorem gives the consistency and asymptotic normality of $\hat{R}^*$ under a set of mild regularity conditions that are provided in Appendix D.1, with a proof sketch given in Appendix D.3.

Theorem 3.1. *Suppose Conditions (C1)–(C7) hold. For R^* given in (3.4) and $\hat{R}^*$ given in (3.5), we have*

(i) $\hat{R}^* \to R^*$ in probability, and

(ii) $\sqrt{n}(\hat{R}^* - R^*) \to N(0, \sigma^2_{R^*})$ in distribution with

$$\sigma^2_{R^*} = (1-\theta)^2\sigma_1^2 + 2(1-\theta)(1-\mu_{x^{(1)}})\sigma_{12} + (1-\mu_{x^{(1)}})^2\sigma_2^2, \tag{3.6}$$

where σ_1^2 is the asymptotic variance of $\sqrt{n}(\hat{\mu}_{x^{(1)}} - \mu_{x^{(1)}})$, σ_2^2 is the asymptotic variance of $\sqrt{n}(\hat{\theta} - \theta)$, and σ_{12} is their asymptotic covariance. See Appendix D.4 for further details on the form of this asymptotic variance.

Although the expression (3.6) appears relatively simple, the term σ_2^2 involves derivatives of the distribution function with respect to β, making it highly complex to evaluate in practice. This complexity is illustrated explicitly in the right-censored case studied by Ding [2010]. Due to the computational burden associated with calculating σ_2^2, we recommend using a bootstrap estimator for the variance $\sigma^2_{R^*}$ [Efron, 1992].

We also conduct a simulation study to evaluate the finite-sample performance of the proposed methodology. Data are generated from models for death time, disease onset time, and probability of disease occurrence consistent with the proposed framework and designed to reflect key characteristics of the observed NACC UDS data. Across a range of covariate configurations spanning much of the $(0, 1)$ GHI scale, the estimator exhibit negligible bias. Further, variances decrease at the expected $1/n$ rate and bootstrapped variance estimates closely match empirical variances. Coverage probabilities are near the nominal 95% level

and improve with increasing sample size. Full details of the simulation design and complete numerical results are provided in Appendix E.

3.5 NACC UDS Data Application

3.5.1 Results

We now apply our proposed metric and estimation procedure to NACC UDS. For the part-one model, we fit the logistic regression model

$$\begin{aligned}\log\left(\frac{\mu}{1-\mu}\right) \;=\; & \beta_1(\text{Death Age - 65}) + \beta_2(\text{Years of Education}) \\ & + \beta_3 1(\text{Is Female}) + \beta_4 1(\text{Is Married}) \\ & + \beta_5 1(\text{Has a Comorbidity}) + \beta_6 1(\text{Race = Black}) \\ & + \beta_7 1(\text{Race = Other}) + \beta_8 1(\varepsilon 4 \text{ Alleles} = 1) \\ & + \beta_9 1(\varepsilon 4 \text{ Alleles} = 2) + \beta_{10}(\text{Age at Study Entry - 65}).\end{aligned} \tag{3.7}$$

A comorbidity is defined as any history at baseline of diabetes, arthritis, traumatic brain injury, a heart attack, or a stroke.

We consider the following linear regression model for our part-two model, modeling the logit transformed $\text{HP}_{\text{Dementia},65}$ (denoted by Y) given that dementia occurred, and leave the error distribution unspecified:

$$\begin{aligned}\log\left(\frac{Y}{1-Y}\right) &= \omega_1(\text{Death Age - 65}) + \omega_2(\text{Years of Education}) \\ &\quad + \omega_3 1(\text{Is Female}) + \omega_4 1(\text{Is Married}) \\ &\quad + \omega_5 1(\text{Has a Comorbidity}) + \omega_6 1(\text{Race = Black}) \\ &\quad + \omega_7 1(\text{Race = Other}) + \omega_8 1(\varepsilon 4 \text{ Alleles = 1}) \\ &\quad + \omega_9 1(\varepsilon 4 \text{ Alleles = 2}) + \text{Error}.\end{aligned} \tag{3.8}$$

Both models are fitted using R, with (3.7) fitted using the `glm` function and (3.8) fitted using the `ltrc` package [Matthews and Nan, 2026]. Regression results are given in Table 3.2, and a set of predicted Golden Health Index values across covariate levels is given in Figures 3.1 and 3.2. For the predictions in Figures 3.1 and 3.2, we used $a^* = 0.25/T$, so that these values can be interpreted as the GHI after age 65 given that dementia or death does not occur within three months after turning 65.

Living longer is associated with a higher probability of being diagnosed with dementia before death, but it is also associated with a larger healthspan proportion among those who develop dementia. These opposing effects create the "J"-shaped curves observed in Figures 3.1 and 3.2, where predicted GHI values first decline and then rise with increasing survival.

Focusing on the part-one model in Table 3.2, the covariates of sex, marital status, and race show associations with dementia probability that differ from expectations based on prior literature [Sommerlad et al., 2018, Kornblith et al., 2022, Aggarwal and Mielke, 2023, Merrick and Brayne, 2024]. Similar counterintuitive patterns have been reported by others using the NACC dataset [Wang and Li, 2021, Lennon et al., 2022, Karakose et al., 2025].

Table 3.2: Results of fitting models (3.7) and (3.8) to the left-truncated data from the NACC UDS, using a baseline age $A_0 = 65$.

Variable	Part 1 Model (3.7)			Part 2 Model (3.8)		
	Estimate	95% CI	P-Value	Estimate	95% CI	P-Value
Death Age	0.083	(0.069, 0.098)	< 0.001	0.083	(0.070, 0.095)	< 0.001
Yrs of School	-0.013	(-0.033, 0.006)	0.186	-0.020	(-0.054, 0.013)	0.231
Sex						
Male	Referent	–	–	Referent	–	–
Female	-0.155	(-0.288, -0.022)	0.023	-0.057	(-0.265, 0.150)	0.589
Is Married						
No	Referent	–	–	Referent	–	–
Yes	0.443	(0.305, 0.581)	< 0.001	-0.083	(-0.307, 0.141)	0.466
Comorbidity						
No	Referent	–	–	Referent	–	–
Yes	-0.135	(-0.273, 0.002)	0.054	0.316	(0.103, 0.528)	0.004
Race						
White	Referent	–	–	Referent	–	–
Black	-0.588	(-0.803, -0.374)	< 0.001	-0.194	(-0.548, 0.160)	0.283
Other	0.234	(-0.170, 0.638)	0.256	-0.373	(-1.216, 0.469)	0.385
$\varepsilon 4$ Alleles						
Zero	Referent	–	–	Referent	–	–
One	0.696	(0.567, 0.825)	< 0.001	-0.460	(-0.683, -0.238)	< 0.001
Two	1.389	(1.087, 1.691)	< 0.001	-0.363	(-0.838, 0.113)	0.135
Study Entry	-0.077	(-0.093, -0.062)	< 0.001	–	–	–

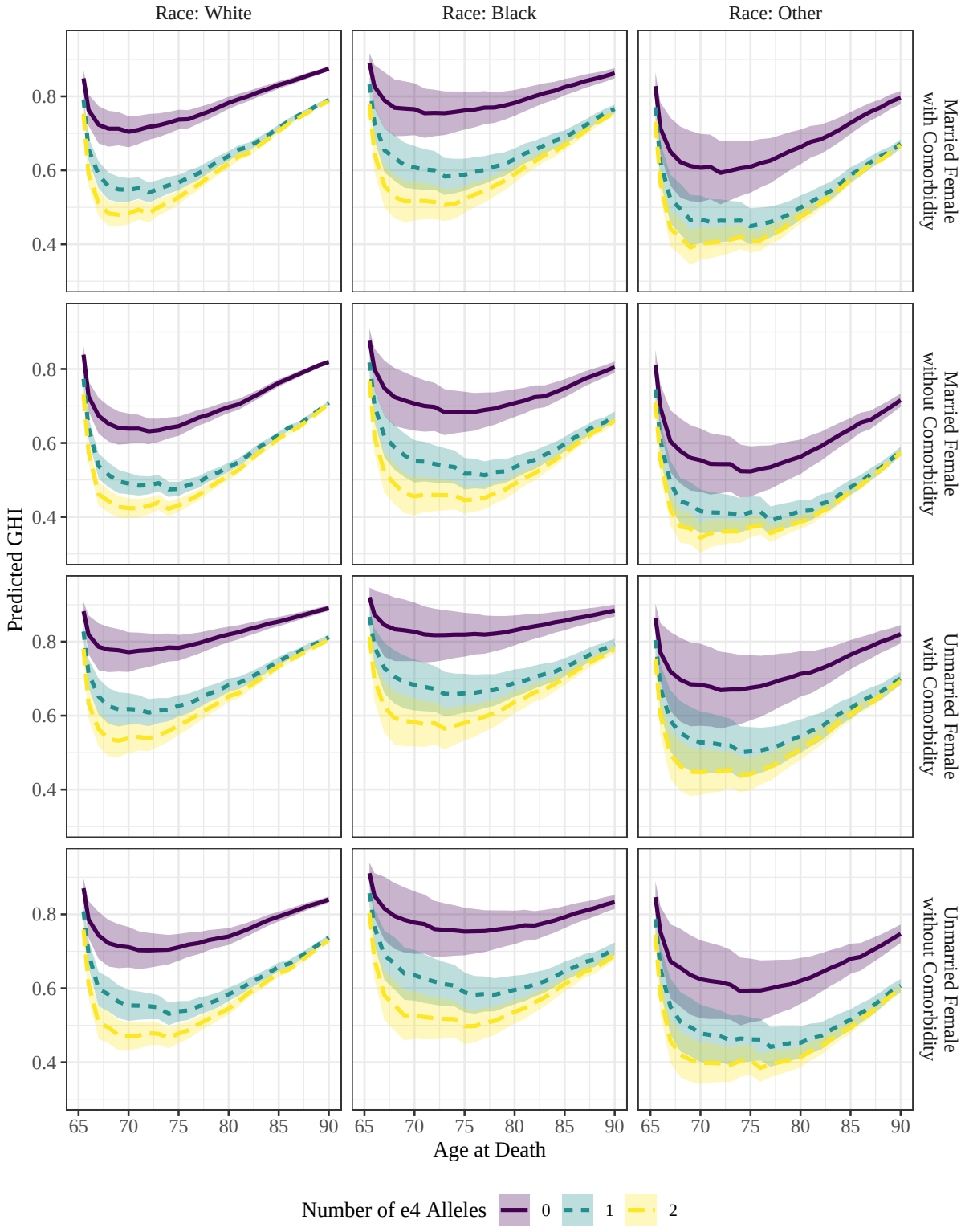


Figure 3.1: Predicted values of GHI based on our two-part fitted models for female patients with a college degree (16 years of education). Shaded bands indicate pointwise 95% confidence intervals for the predictions, with the variance estimated using the bootstrap.

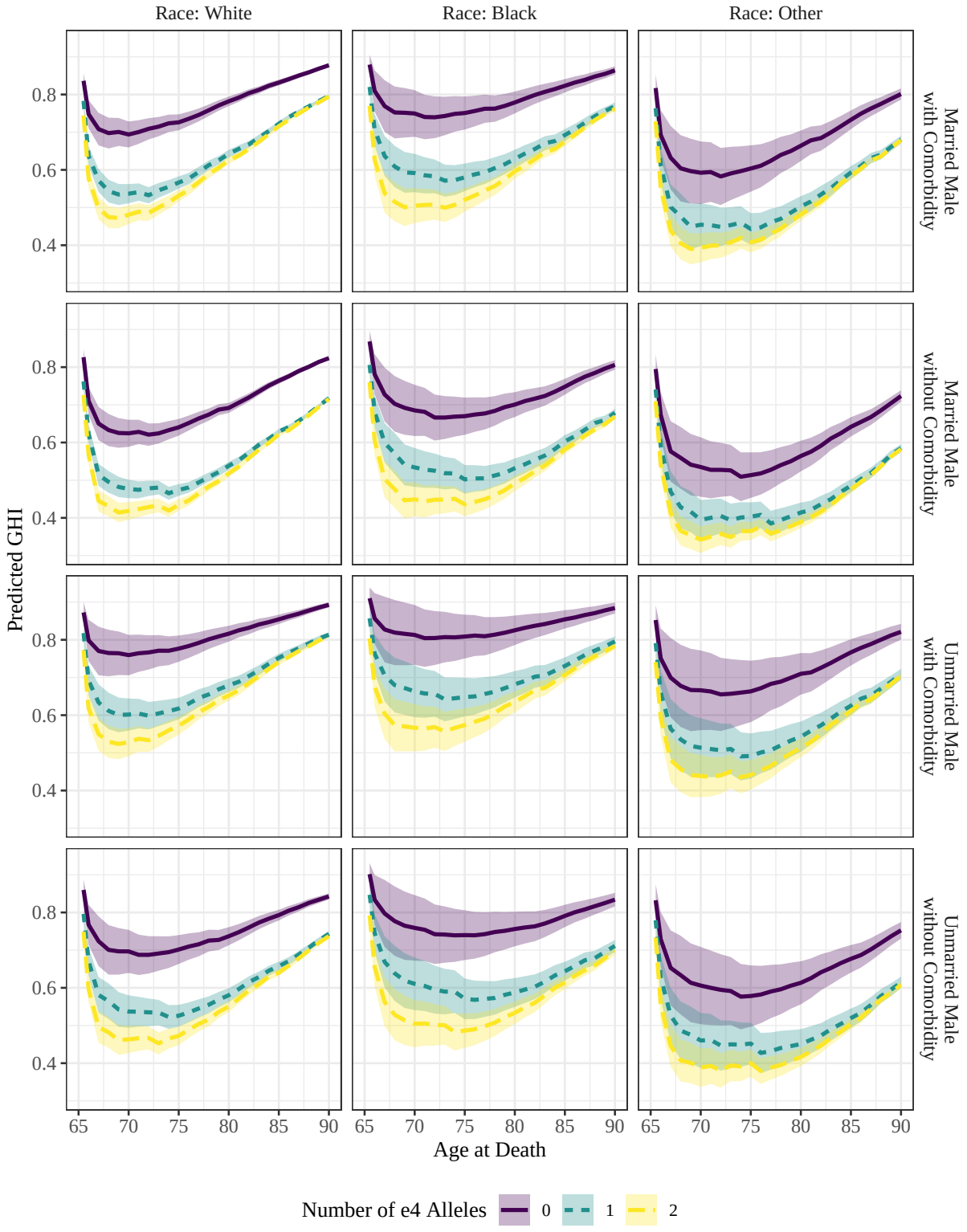


Figure 3.2: Predicted values of GHI based on our two-part fitted models for male patients with a college degree (16 years of education). Shaded bands indicate pointwise 95% confidence intervals for the predictions, with the variance estimated using the bootstrap.

Our main predictor of interest is the number of APOE $\varepsilon 4$ alleles. Figures 3.1 and 3.2 show clear separation in predicted GHI values by APOE $\varepsilon 4$ status. This answers our primary question of interest, highlighting the association between dementia-free proportion of life after age 65 and the APOE $\varepsilon 4$ allele. Specifically, APOE $\varepsilon 4$ carriers have lower GHI comparing to non-carriers, and carriers with two copies have the lowest GHI.

Table 3.2 documents how differences arise in both modeling components, particularly for $\varepsilon 4$ allele carriers, answering the secondary questions of interest. Having a copy of the $\varepsilon 4$ allele increases the probability of dementia before death, consistent with prior findings [Safieh et al., 2019, Fortea et al., 2024]. Notably, the part-two model shows that the $\varepsilon 4$ allele is also associated with a lower expected healthspan proportion among those diagnosed with dementia. In other words, $\varepsilon 4$ carriers not only face a higher risk of developing dementia, but once diagnosed, they also tend to spend a greater fraction of their remaining life with the disease even after adjusting for other covariates. This pattern is consistent with earlier onset of a low-morbidity condition, resulting in a longer duration of life lived with disease.

The result of the part-two model is less clear for individuals with two copies of the $\varepsilon 4$ allele. Although the estimated effect among homozygotes is negative with a similar size, the effect is not statistically significant. This may reflect limited sample size rather than a true absence of association, or could be related to the accelerated cognitive decline seen in this group [Craft et al., 1998].

3.5.2 Model Checking and Sensitivity Analyses

We have conducted extensive model diagnostics and sensitivity analyses to assess the validity and robustness of these results. Full diagnostic figures and tables are provided in Appendix F.

For the part-one logistic regression model (Figures F.1 and F.2), surrogate residual diagnostics show strong agreement with theoretical expectations [Liu and Zhang, 2018]. The Q–Q plot closely follow the reference line and residuals are centered at zero with no systematic patterns. Additionally, a Hosmer-Lemeshow test based on deciles of predicted risk yields a non-significant result ($p = 0.08$), indicating no evidence of lack of fit of the model [Hosmer and Lemesbow, 1980]. Linearity assumptions for continuous covariates are evaluated using (a) surrogate residual plots and (b) likelihood ratio comparisons with spline-based alternatives. Both approaches support the assumed linear functional forms. Multicollinearity is modest (maximum VIF ≈ 3.8). Influence diagnostics based on leverage and Cook's distance have identified several observations above typical thresholds [Pregibon, 1981], but refitting the model after removing potentially influential subsets produces nearly identical estimates.

For the part-two semiparametric model (Figures F.3–F.5), martingale residual diagnostics accounting for left-truncation support the assumed linear covariate effects [Barlow and Prentice, 1988]. Residual-versus-covariate plots and cumulative martingale residual processes largely follow their theoretical mean-zero behavior, and alternative nonlinear transformations of death time do not improve model fit or alter other parameter estimates [Therneau et al., 1990, Lin et al., 1993]. Removing high-leverage observations or extreme residuals has negligible impact. Comparison with a parametric truncated regression model indicates that the normal error assumption is inappropriate, supporting the semiparametric specification.

Enrollment practices differ considerably across ADRCs, leading to heterogeneity in variable distributions. For example, some centers enroll participants at much older ages, while others target specific demographic subgroups, resulting in variation in sex, marital status, and racial composition across centers. In addition, the analytic sample slightly over-represents APOE $\varepsilon 4$ carriers [Rasmussen et al., 2019, Lumsden et al., 2020]. Tables F.1 and F.2 illustrate center variations in participants' characteristics in the analytic sample and the target population, respectively. We perform sensitivity analyses addressing the impact of heterogeneity across

ADRCs. In particular, we find that introducing random effects (either random intercept or random intercept plus random slopes, Table F.3) in the part-one model produces parameter estimates consistent with the primary analysis. For the part-two component, analogous comparisons using a parametric model with normal error are provided merely as an illustration, since including random effects in the current semiparametric modeling methodology for truncated data has not yet been developed. While not a perfect solution, nonetheless the addition of random effects to the truncated parametric model does not seriously impact those model estimates (Table F.4). These findings indicate that between-center heterogeneity does not materially influence the reported associations.

Because over 80% of individuals in the target population are excluded due to censored death times, one would expect different demographic characteristics between included and excluded participants, which are illustrated in Table F.5. We evaluate potential impacts of the selection biases. To assess sensitivity to temporal sampling, we re-fit the models using earlier data, effectively analyzing the data as if at prior time points, and compare the estimating results with the current results obtained from December 2024 data freeze. We find these models have consistent effect directions and overlapping confidence intervals with our main analysis, see Table F.6. Notably, APOE $\varepsilon 4$ estimates remain stable across historical analyses performed at many earlier time points, see Figure F.6.

Although changing covariate distribution does not necessarily alter the association between covariates and the response variable in a regression model, we look into the impact of potential oversampling APOE $\varepsilon 4$ carriers in the NACC UDS by conducting a weighted analysis. Weights for APOE $\varepsilon 4$ status are obtained from the UK Biobank [Lumsden et al., 2020], stratified on sex and race with their proportions determined from the US Census data. Since APOE $\varepsilon 4$ allele frequencies are only available for white race that is also the majority in the analytic data set (see Table 3.1), we assign the same allele frequencies to other race groups for the sake of sensitivity analysis. Each participant is assigned a weight equal to

the target proportion for their stratum divided by the corresponding observed proportion in the analytic data, so that overrepresented strata are downweighted and underrepresented strata are upweighted. Weighted and unweighted results are nearly identical for both model components (see Table F.7), suggesting that differential covariate sampling has little impact on our main conclusions and supporting the robustness of our findings.

3.6 Discussion

While this study demonstrates the utility of the GHI for characterizing dementia-free healthspan, several limitations warrant discussion. First, the generalizability of our findings may be constrained by the nature of the study and the overall recruitment efforts in the NACC. Certain populations, including individuals from underrepresented racial and ethnic groups and those with lower educational attainment, are not proportionally represented in the NACC data. Moreover, participants who engage in regular clinic visits and who have APOE gene information available may differ systematically in health behavior or disease risk compared to the broader population. These factors limit the extent to which estimates from the NACC sample can be directly extrapolated to the general aging population. However, the sensitivity analyses discussed in Section 3.5.2 indicate that these results are reasonably robust to these sampling issues and the modeling choice.

Second, in this work, we focus on individuals with recorded death dates to avoid complications related to censoring. However, many longitudinal studies (including the NACC UDS) feature a substantial proportion of participants who are still alive at their last follow-up. Developing methods to appropriately account for censoring in both disease onset and death times remains an important area for future work. A related consideration is using this method to predict the GHI for living individuals. Since death time is a covariate in our models, all predictions must make some assumption about what the age of death will be.

One other note on the topic of death time is that any computation of the GHI must be conditional on death time, as implemented here. Otherwise, comparisons of the GHI do not take into account absolute lifespan and are thus less useful.

Third, the choice of integration limits a^* and b^* plays a nuanced role in interpreting the predicted GHI. In our analysis, we set a^* and b^* to be small and fixed durations from the boundaries (typically between one-quarter year and one year, scaled by death time) to satisfy the assumptions needed for asymptotic validity. These values strike a practical balance between theoretical requirements and data limitations. Although it is computationally feasible to predict the GHI down to the smallest observed residual, doing so complicates interpretability. For example, a lower limit corresponding to 0.1598 years is less intuitive than a clean cutoff like 0.25 years. We also note that as a^* increases, the directional relationships between covariates and GHI remain stable, though the predicted GHI generally increases as well. This aligns with intuition: conditioning on longer disease-free survival implies a healthier subset of individuals, leading to higher GHI estimates.

Despite these limitations, the proposed framework offers a novel and flexible approach to studying healthspan in aging populations. By explicitly modeling the time spent in disease-free states, our method provides new insights into the lived experience of aging, complementing traditional survival-based analyses.

Chapter 4

Asymptotic Normality of Sieve M-Estimators

4.1 Introduction

Semiparametric models provide a flexible framework for statistical inference when the parameter of interest contains both finite- and infinite-dimensional components. In many applications, the infinite-dimensional component is modeled nonparametrically, while the finite-dimensional component retains a direct scientific interpretation. This combination allows researchers to avoid overly restrictive assumptions while still performing interpretable inference on key model features. A common estimation approach in this setting is sieve estimation, which approximates the infinite-dimensional parameter by a sequence of spaces whose dimension grows with the sample size. This turns the original infinite-dimensional problem into a sequence of problems with parameters estimated via standard parametric methods. Note that when there is no finite-dimensional component, the model reduces to a nonparametric model.

The setting considered in this work is motivated by semiparametric models with bundled parameters [Ding and Nan, 2011a], where the infinite-dimensional component may depend on the finite-dimensional parameter. This structure arises naturally in models where the infinite-dimensional component is tied to the finite-dimensional component in some way, and reduces to a standard semiparametric model when finite- and infinite-dimensional parameters are separated. Existing results for sieve M-estimation with bundled parameters provide general tools for establishing consistency and asymptotic normality of finite-dimensional components [Ding and Nan, 2011a, Matthews and Nan, 2026]. However, for many applications, it is also important to conduct inference on the infinite-dimensional component itself, for example, through pointwise confidence intervals for the estimated function.

In this work we develop an asymptotic normality result for sieve M-estimators in a bundled-parameter setting, with particular attention to pointwise inference on the nonparametric component. The estimator maximizes an empirical criterion over a sieve space, and the proof uses an expansion at the sieve population maximizer, or pseudo-truth, rather than at the full population truth. This allows the argument to separate the stochastic estimation error from the deterministic sieve approximation bias. Our main result has three parts. First, it establishes asymptotic normality for arbitrary contrasts of the sieve coefficient vector. Second, it shows that the corresponding sandwich variance estimator is consistent for such contrasts, leading to feasible inference. Third, it derives a pointwise limiting distribution for the nonparametric component, including the contribution of the sieve approximation bias. The result therefore provides a direct route to pointwise inference for basis-expansion-based sieve estimators in semiparametric models with bundled parameters.

4.2 Notation and Problem Setup

Let $Z_1, \ldots, Z_n \sim P_0$ be independent and identically distributed observations taking values in a measurable space $\mathcal{Z}$, where P_0 is the true probability law. Let $m : \mathcal{Z} \times \Theta \to \mathbb{R}$ be a measurable criterion function indexed by $\theta \in \Theta$. We write the population objective as

$$S(\theta) = Pm(\theta) = \int m(z;\theta)\, dP_0(z),$$

and the empirical objective as

$$S_n(\theta) = P_n m(\theta) = \frac{1}{n}\sum_{i=1}^{n} m(Z_i;\theta).$$

Our goal is to study the asymptotic behavior of a maximizer of S_n in a semiparametric model with bundled parameters, in the sense of Ding and Nan [2011a].

The parameter takes the form

$$\theta = (\beta, \zeta(\cdot,\beta)),$$

where $\beta \in \mathcal{B} \subset \mathbb{R}^d$ is finite-dimensional and $\zeta(\cdot,\beta)$ is an infinite-dimensional parameter that depends on β. Note that in the case where ζ does not depend on β, the results in this work still hold as a special case. Let $\mathcal{H}$ be a functional space equipped with norm $\|\cdot\|_\zeta$. The full parameter space is

$$\Theta = \{(\beta, \zeta(\cdot,\beta)) : \beta \in \mathcal{B},\ \zeta(\cdot,\beta) \in \mathcal{H}\}.$$

For $\theta_1 = (\beta_1, \zeta_1)$ and $\theta_2 = (\beta_2, \zeta_2)$, define the distance

$$d(\theta_1, \theta_2) = \left\{|\beta_1 - \beta_2|^2 + \|\zeta_1(\cdot,\beta_1) - \zeta_2(\cdot,\beta_2)\|_\zeta^2\right\}^{1/2},$$

where $|\cdot|$ denotes the Euclidean norm on $\mathbb{R}^d$.

Let

$$\theta_0 = (\beta_0, \zeta_0(\cdot, \beta_0)) = \arg\max_{\theta \in \Theta} S(\theta)$$

denote the true parameter, defined as the unique maximizer of S over Θ, and assume that θ_0 lies in the interior of Θ.

We adopt the notation of Ding and Nan [2011a] for Fréchet derivatives. For the criterion function $m(z;\theta)$ at $\theta = (\beta, \zeta)$, define

$$\begin{aligned} \dot{m}_\beta(z;\theta) &= \frac{\partial}{\partial \beta} m(z; \beta, \zeta(\cdot, \beta)), \\ \dot{m}_\zeta(z;\theta)[h] &= \lim_{t \to 0} \frac{m(z; \beta, \zeta + th) - m(z;\theta)}{t}, \end{aligned}$$

where $h \in \mathcal{H}$ is a direction in the functional space. Similarly, denote respective second derivatives as the following:

$$\ddot{m}_{\beta\beta}, \quad \ddot{m}_{\beta\zeta}[h], \quad \ddot{m}_{\zeta\beta}[h], \quad \ddot{m}_{\zeta\zeta}[h_1, h_2].$$

Under Assumption (B1) below, which permits the interchange of differentiation and integration up to third order, the population score operators

$$\dot{S}_\beta(\theta) = P\dot{m}_\beta(\cdot;\theta), \qquad \dot{S}_\zeta(\theta)[h] = P\dot{m}_\zeta(\cdot;\theta)[h]$$

are well-defined, with empirical counterparts

$$\dot{S}_{\beta,n}(\theta) = P_n\dot{m}_\beta(\cdot;\theta), \qquad \dot{S}_{\zeta,n}(\theta)[h] = P_n\dot{m}_\zeta(\cdot;\theta)[h].$$

Second-order population operators $\ddot{S}_{\beta\beta}(\theta)$, $\ddot{S}_{\beta\zeta}(\theta)[h]$, $\ddot{S}_{\zeta\beta}(\theta)[h]$, and $\ddot{S}_{\zeta\zeta}(\theta)[h_1, h_2]$ are defined analogously under (B1), along with their empirical counterparts.

For any continuous j-linear form T acting on the parameter space, we use the induced operator norm

$$\|T\|_{\mathrm{op}} = \sup\Big\{ |T[v_1, \ldots, v_j]| : \|v_1\|_\Theta, \ldots, \|v_j\|_\Theta \le 1 \Big\},$$

where $\|v\|_\Theta^2 = |a|^2 + \|h\|_\zeta^2$ for $v = (a, h) \in \mathbb{R}^d \times \mathcal{H}$. For a matrix $A \in \mathbb{R}^{p\times p}$ viewed as a bilinear form, this reduces to the usual spectral norm $\|A\|_{\mathrm{op}} = \sup_{\|x\|=1} \|Ax\|$.

The estimation of the infinite-dimensional parameter ζ is carried out in a sequence of growing-dimensional sieve spaces. Let $\Theta_n \subset \Theta$ be an increasing sequence of subsets such that $\bigcup_{n=1}^\infty \Theta_n$ is dense in Θ. This sieve approximates θ_0 at rate ϵ_n in the sense that there exists a sequence $\epsilon_n \to 0$ and $\tilde{\theta}_n \in \Theta_n$ such that $d(\tilde{\theta}_n, \theta_0) = O(\epsilon_n)$. Throughout, we consider sieve spaces Θ_n formed by basis expansions of ζ. For each n, the function $\zeta \in \mathcal{H}$ is approximated by a basis expansion of the form

$$\zeta_n(\cdot, \beta; \gamma) = \sum_{j=1}^{k_n} \gamma_j \, B_{k_n,j}(\phi(\cdot; \beta)), \qquad \gamma = (\gamma_1, \ldots, \gamma_{k_n})^\top \in \mathbb{R}^{k_n},$$

where $\{B_{k_n,j}\}_{j=1}^{k_n}$ is a set of k_n basis functions, and ϕ is a function mapping the dependence of ζ on β.

Since β has fixed dimension d, the corresponding sieve parameter has total dimension

$$p_n = d + k_n, \qquad p_n \to \infty.$$

It is convenient to work with the growing-dimensional sieve coordinate

$$\eta = (\beta^\top, \gamma^\top)^\top \in \mathbb{R}^{p_n},$$

and with the induced parameter map

$$\theta(\eta) = (\beta, \zeta_n(\cdot, \beta; \gamma)) \in \Theta_n,$$

where $\Theta_n = \mathbb{R}^d \times \mathcal{H}_n$, and $\mathcal{H}_n = \text{span}\{B_{k_n,1}(\cdot, \beta), \ldots, B_{k_n,k_n}(\cdot, \beta)\}$. The map $\theta : \mathbb{R}^{p_n} \to \Theta_n$ embeds the sieve coordinate into the full parameter space; we will use $\theta(\eta)$ and η interchangeably where the context is clear.

With a slight abuse of notation, we write

$$S(\eta) = S(\theta(\eta)) = P\, m(\theta(\eta)), \qquad S_n(\eta) = S_n(\theta(\eta)) = P_n\, m(\theta(\eta)).$$

Their gradient and negative Hessian are denoted by

$$U^{(n)}(\eta) = \frac{\partial}{\partial \eta} S(\eta), \qquad U_n^{(n)}(\eta) = \frac{\partial}{\partial \eta} S_n(\eta),$$

and

$$J^{(n)}(\eta) = -\frac{\partial^2}{\partial \eta \partial \eta^T} S(\eta), \qquad J_n^{(n)}(\eta) = -\frac{\partial^2}{\partial \eta \partial \eta^T} S_n(\eta).$$

Because both the score vector and the information matrix depend on the sieve basis $\{B_{k_n,j}\}$ and hence on n, we use the superscript $^{(n)}$ to make this explicit throughout. Define the

sieve-projected score at any $\theta \in \Theta$:

$$u_i^{(n)}(\theta) = \begin{pmatrix} \dot{m}_\beta(Z_i;\theta) \\ \dot{m}_\zeta(Z_i;\theta)[B_{k_n,1}] \\ \vdots \\ \dot{m}_\zeta(Z_i;\theta)[B_{k_n,k_n}] \end{pmatrix} \in \mathbb{R}^{p_n},$$

with population and empirical versions $\dot{S}^{(n)}(\theta) = P\, u_i^{(n)}(\theta)$ and $\dot{S}_n^{(n)}(\theta) = P_n\, u_i^{(n)}(\theta)$. Similarly, the sieve-projected information matrix at θ is

$$J^{(n)}(\theta) = -\begin{pmatrix} \ddot{S}_{\beta\beta}(\theta) & \ddot{S}_{\beta\zeta}(\theta)[B_{k_n,1}] & \cdots & \ddot{S}_{\beta\zeta}(\theta)[B_{k_n,k_n}] \\ \ddot{S}_{\zeta\beta}(\theta)[B_{k_n,1}] & \ddot{S}_{\zeta\zeta}(\theta)[B_{k_n,1}, B_{k_n,1}] & \cdots & \ddot{S}_{\zeta\zeta}(\theta)[B_{k_n,1}, B_{k_n,k_n}] \\ \vdots & \vdots & \ddots & \vdots \\ \ddot{S}_{\zeta\beta}(\theta)[B_{k_n,k_n}] & \ddot{S}_{\zeta\zeta}(\theta)[B_{k_n,k_n}, B_{k_n,1}] & \cdots & \ddot{S}_{\zeta\zeta}(\theta)[B_{k_n,k_n}, B_{k_n,k_n}] \end{pmatrix}.$$

Define $J_n^{(n)}(\theta)$ in a similar way for the empirical second derivatives, and define $J_{n,i}^{(n)}(\theta)$ as the individual-level matrix for observation i. When $\theta = \theta(\eta) \in \Theta_n$, these coincide with the coordinate gradient and Hessian: $u_i^{(n)}(\theta(\eta)) = \nabla_\eta m(Z_i;\theta(\eta))$, $\dot{S}^{(n)}(\theta(\eta)) = U^{(n)}(\eta)$, and $J^{(n)}(\theta(\eta)) = -\nabla_\eta^2 S(\theta(\eta)) = J^{(n)}(\eta)$.

Although the true parameter θ_0 need not belong to the growing-dimensional sieve space Θ_n, the projected information matrix $J^{(n)}(\theta_0)$ is still well-defined. It is obtained by evaluating the full population second-derivative bilinear form $-\ddot{S}(\theta_0)$ on the collection of directions generated by perturbing the coordinates of β and by perturbing ζ along the basis functions $B_{k_n,1}, \ldots, B_{k_n,k_n}$. Thus $J^{(n)}(\theta_0)$ should be understood as a sieve projection of the full information operator at the truth, not as the Hessian of a growing-dimensional criterion evaluated at a point of Θ_n.

Define the sieve population maximizer by

$$\theta_{0,n} = (\beta_{0,n}, \zeta_{0,n}(\cdot, \beta_{0,n}; \gamma_{0,n})) = \arg\max_{\theta \in \Theta_n} S(\theta),$$

or equivalently, in coordinates,

$$\eta_{0,n} = \arg\max_{\eta:\, \theta(\eta) \in \Theta_n} S(\eta).$$

The sieve M-estimator is

$$\hat{\theta}_n = (\hat{\beta}_n, \hat{\zeta}_n(\cdot, \hat{\beta}_n)) = \arg\max_{\theta \in \Theta_n} S_n(\theta),$$

with coordinate representation $\hat{\eta}_n$ defined by

$$\hat{\theta}_n = \theta(\hat{\eta}_n), \qquad \theta_{0,n} = \theta(\eta_{0,n}).$$

When $m(z;\theta) = \log p_\theta(z)$ is the log-likelihood, $\hat{\theta}_n$ becomes the semiparametric or nonparametric sieve maximum likelihood estimator.

Define the sieve-projected score variance

$$V_n = E_0\big[u_i^{(n)}(\theta_{0,n})\, u_i^{(n)}(\theta_{0,n})^T\big],$$

the sandwich covariance $\Sigma_n = \{J^{(n)}(\theta_{0,n})\}^{-1} V_n \{J^{(n)}(\theta_{0,n})\}^{-1}$, and the empirical score outer product

$$\hat{V}_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} u_i^{(n)}(\theta)\, u_i^{(n)}(\theta)^T.$$

Since $\theta_{0,n} \in \Theta_n$, the score $u_i^{(n)}(\theta_{0,n}) = \nabla_\eta m(Z_i; \theta_{0,n})$ and the information matrix $J^{(n)}(\theta_{0,n}) =$

$-\nabla^2_\eta S(\eta_{0,n})$ are the usual gradient and Hessian evaluated at $\eta_{0,n}$.

4.3 Assumptions

We assume the following regularity conditions:

(B1) (*Criterion smoothness and interchangeability*) For P-almost every z, $m(z;\theta)$ is three-times continuously Fréchet differentiable in a neighborhood of θ_0, with uniformly bounded third derivatives on compact sets; in other words, for each $j = 1, 2, 3$, there exists a measurable function M_j satisfying $PM_j < \infty$ such that

$$\sup_{\theta:\, d(\theta,\theta_0)\leq\delta} \left\| D^j m(z;\theta) \right\|_{\text{op}} \leq M_j(z)$$

for some $\delta > 0$, and for P-almost every z, where $D^j m(z;\theta)$ denotes the jth Fréchet derivative of $m(z;\theta)$ viewed as a continuous j-linear form on the parameter space. Moreover, for each $j = 1, 2, 3$, differentiation with respect to θ and integration with respect to P_0 may be interchanged in a neighborhood of θ_0, so that $D^j P m(\cdot;\theta) = P\, D^j m(\cdot;\theta)$.

(B2) (*Basis Requirement*) The expansion basis $\{B_{k_n,j}\}_{j=1}^{k_n}$ is normalized.

(B3) (*Rate of Convergence for Non-Bundled Parameters*) There exists $\xi > 1/4$ such that $d(\hat{\theta}_n, \theta_0) = O_p(n^{-\xi})$, the sieve approximation rate satisfies $\epsilon_n = O(n^{-\xi})$, and the sieve dimension satisfies $p_n = o(n^{2\xi-1/2})$.

(B3*) (*Rate of Convergence for Bundled Parameters*) There exists $\xi > 3/8$ such that $d(\hat{\theta}_n, \theta_0) = O_p(n^{-\xi})$, the sieve approximation rate satisfies $\epsilon_n = O(n^{-\xi})$, and the sieve dimension satisfies $p_n = o(n^{2\xi-1/2})$.

(B4) *(Uniqueness)* $\theta_0 \in \Theta$ is the unique maximizer of S over Θ, well-separated in the sense that for every $\varepsilon > 0$, there exists a $\delta > 0$ such that $\sup_{\theta \in \Theta:\, d(\theta,\theta_0) \geq \varepsilon} S(\theta) < S(\theta_0) - \delta$.

(B5) *(Positive Definite Information Operator)* There exist constants $0 < c_J < C_J < \infty$ not depending on n, such that for every direction $v = (a, h) \in \mathbb{R}^d \times \mathcal{H}$ with $\|v\|_\Theta < \infty$,

$$c_J \|v\|_\Theta^2 \leq -\ddot{S}(\theta_0)[v, v] \leq C_J \|v\|_\Theta^2,$$

where $\|v\|_\Theta^2 = |a|^2 + \|h\|_\zeta^2$. Further, there exist constants $0 < c_V < C_V < \infty$ not depending on n such that, for all $c \in \mathbb{R}^{p_n}$,

$$c_V \|c\|^2 \leq c^\top V_n c \leq C_V \|c\|^2.$$

(B6) (*Approximate score equations*) The estimator $\hat{\theta}_n$ satisfies

$$\|\dot{S}_{\beta,n}(\hat{\theta}_n)\| = o_p(n^{-1/2})$$

and

$$\sup_{h \in \mathcal{H}_n,\, \|h\|_\zeta \leq 1} \left| \dot{S}_{\zeta,n}(\hat{\theta}_n)[h] \right| = o_p(n^{-1/2}).$$

(B7) (*Pointwise bias scaling*) For each fixed $x \in [a, b]$, $\sqrt{n/p_n}\{\zeta_{0,n}(x, \beta_{0,n}) - \zeta_0(x, \beta_0)\} \to b(x)$ where $|b(x)| < \infty$.

Condition (B1) is the infinite-dimensional analogue of the usual smoothness and dominated-differentiation conditions imposed in classical maximum likelihood theory [Lehmann and Casella, 1998], and additionally requires that differentiation and integration may be interchanged up to third order. Here the condition is imposed to an arbitrary objective function that includes likelihood as a special case. The interchangeability holds automatically when

the sample space $\mathcal{Z}$ is compact, and can be satisfied more generally by many standard models with unbounded response. Condition (B2) establishes a basic condition on the basis functions. Conditions (B3)–(B6) follow (A1)–(A4) of Ding and Nan [2011a] closely, with some minor adjustments required to ensure that normality can be extended to the nonparametric portion of θ. Specifically, Condition (B3) is standard and can be verified by, for example, checking conditions (C1)–(C3) in Shen and Wong [1994]. The convergence rate needs to be achieved before obtaining asymptotic normality. Condition (B3*) is an alternate version of (B3) necessary in the bundled parameter case. Condition (B4) is common for maximum likelihood and usually holds. Condition (B5) provides local identifiability and coercivity of the population information bilinear form on $\mathbb{R}^d \times \mathcal{H}$. For a likelihood model this reduces to that the Fisher information is bounded above and away from zero uniformly over normalized directions. For global maximizers over the sieve, the first-order conditions hold exactly, so (B6) is trivially satisfied. The $o_p(n^{-1/2})$ form accommodates approximate optimizers (e.g., from gradient descent with finite termination criteria). Condition (B7) defines the rate that balances the bias and variance from the sieve estimation, and also allows for the undersmoothing case where $b(x) = 0$.

Remark 4.1. *Since θ_0 lies in the interior of Θ and is the maximizer of S, the smoothness of m in (B1) along with (B4) implies that the population score equations*

$$\dot{S}_\beta(\theta_0) = 0, \qquad \dot{S}_\zeta(\theta_0)[h] = 0 \text{ for all } h \in \mathcal{H},$$

hold as first-order conditions.

Remark 4.2. *In the bundled-parameter case, one may ask that the sieve approximation of ζ be uniformly Lipschitz in β, i.e., for each fixed x in a neighborhood of θ_0,*

$$|\zeta_n(x, \beta_1; \gamma) - \zeta_n(x, \beta_2; \gamma)| \leq L_\zeta(x) \, \|\beta_1 - \beta_2\|,$$

for some $L_\zeta(x)$ not depending on n. The smoothness and dominating-function bounds in (B1), applied along β-directions to the basis expansion $\sum_j \gamma_j B_{k_n,j}(\phi(\cdot;\beta))$, yield such a Lipschitz bound in a neighborhood of θ_0.

Remark 4.3. *Assumption (B3) defines consistency as a prerequisite for the asymptotic normality results derived below while putting an upper bound on the sieve dimension growth. If the sieve approximation error satisfies $\epsilon_n = O(p_n^{-r})$ and $p_n = O(n^\kappa)$, then Assumption (B3) is satisfied whenever $\frac{\xi}{r} \leq \kappa < 2\xi - \frac{1}{2}$. This interval is nonempty provided $\xi > r/(4r-2)$, which is possible for $r > 1$.*

Remark 4.4. *Assumption (B3*) is the replacement for (B3) in the bundled-parameter case, where the basis representation of the nuisance component depends on the finite-dimensional parameter. In this setting, pointwise evaluation of $\hat{\zeta}_n(x, \hat{\beta}_n)$ introduces an additional remainder term from replacing $\beta_{0,n}$ by $\hat{\beta}_n$ inside the basis functions. Under Remark 4.2, this bundled remainder is of order $O_p(n^{-\xi})$. Since the pointwise nonparametric component is normalized by $\sqrt{n/p_n}$, we require $n^{-\xi}\sqrt{n/p_n} \to 0$. Together with the upper bound $p_n = o(n^{2\xi-1/2})$, this compatibility is guaranteed under the stronger rate condition in (B3*). When the nuisance component is not bundled with β, this additional remainder is absent, and (B3*) is not required.*

Remark 4.5. *The constants c_J and C_J are uniform in n because (B5) is stated on the full function space $\mathcal{H}$, not on a dimension-dependent subspace. When the coordinate map $v \mapsto (a, \sum_j v_{\gamma,j} B_{k_n,j})$ is norm-preserving (as for orthonormal bases), the eigenvalues of the growing-dimensional matrix $J^{(n)}(\theta_0) \in \mathbb{R}^{p_n \times p_n}$ are bounded below by c_J and above by C_J, uniformly in n.*

Remark 4.6. *Assumption (B7) is stated pointwise (for each fixed x) because Theorem 4.1(iii) delivers a pointwise result; uniform confidence bands would require stronger uniformity conditions on $b(x)$ and are deferred to future work. The quantity $b(x)$ represents the renormalized sieve approximation bias at x; it is generally non-zero and is the reason the limiting distri-*

bution in Theorem 4.1(iii) is not centered.

Remark 4.7. *Our conditions differ from those of Ding and Nan [2011a] in scope. Their analysis targets the Euclidean parameter along the least favorable direction, and consequently relies on stochastic equicontinuity of the score together with smoothness of the least favorable submodel. Our aim is joint asymptotic normality across the full sieve coordinate vector, and this stronger conclusion is obtained through the uniform third-derivative control in (B1). The bounded third derivative delivers direct control of the Hessian variation and hence of the Taylor remainder in every sieve direction, which in turn removes the need for the score equicontinuity and least-favorable-direction smoothness conditions of Ding and Nan [2011a].*

4.4 Main Results

Lemma 4.1 (Sieve pseudo-truth convergence and rate)**.**

(i) *Under* (B1) *and* (B4), $d(\theta_{0,n}, \theta_0) \to 0$ *as* $n \to \infty$.

(ii) *Under additionally* (B5), $d(\theta_{0,n}, \theta_0) = O(\epsilon_n)$.

Lemma 4.2 (Local properties of J)**.** *Under* (B1) *and* (B3)–(B5)*, the following hold:*

(i) *For* $\bar{\theta}_n$ *and* θ_n^* *in the neighborhood of* θ_0*, with* $d(\bar{\theta}_n, \theta_n^*) = O_p(n^{-\xi})$,

$$\|J_n^{(n)}(\bar{\theta}_n) - J_n^{(n)}(\theta_n^*)\|_{\mathrm{op}} = O_p(p_n n^{-\xi}) \quad \textit{and} \quad \|J^{(n)}(\bar{\theta}_n) - J^{(n)}(\theta_n^*)\|_{\mathrm{op}} = O_p(p_n n^{-\xi}).$$

(ii) *For any* θ_n^* *with* $d(\theta_n^*, \theta_0) \le C n^{-\xi}$,

$$\|J_n^{(n)}(\theta_n^*) - J^{(n)}(\theta_n^*)\|_{\mathrm{op}} = O_p\left(\max\left\{p_n n^{-\xi}, p_n\sqrt{\frac{\log p_n}{n}}\right\}\right).$$

(iii) *For θ_n^* with $d(\theta_n^*, \theta_0) \le Cn^{-\xi}$, there exists $c_J' > 0$ not depending on n such that for all $v_n \in \mathbb{R}^{p_n}$ and all n sufficiently large,*

$$c_J' \|v_n\|^2 \le v_n^\top J^{(n)}(\theta_n^*)\, v_n \le C_J' \|v_n\|^2.$$

In other words, (B5) transfers to the sieve in the neighborhood of the truth, and $J^{(n)}(\theta_n^*)$ is uniformly invertible in n.

Lemma 4.3 (Sieve variance consistency)**.** *Under assumptions* (B1) *and* (B3)–(B5)*,*

$$\|J_n^{(n)}(\hat\theta_n) - J^{(n)}(\theta_{0,n})\|_{\mathrm{op}} = o_p(1) \qquad \textit{and} \qquad \|\hat V_n(\hat\theta_n) - V_n\|_{\mathrm{op}} = o_p(1).$$

Theorem 4.1. *Let $a_n \in \mathbb{R}^{p_n}$ be an arbitrary vector of real values satisfying $a_n \neq 0$.*

(i) ***(Directional asymptotic normality.)*** *Under* (B1)–(B6)*,*

$$\frac{\sqrt{n}\, a_n^T(\hat\eta_n - \eta_{0,n})}{(a_n^T \Sigma_n a_n)^{1/2}} \;\to_d\; N(0,1).$$

(ii) ***(Feasible inference.)*** *Under the same conditions, the sandwich covariance estimator*

$$\hat\Sigma_n = \{J_n^{(n)}(\hat\theta_n)\}^{-1}\, \hat V_n(\hat\theta_n)\, \{J_n^{(n)}(\hat\theta_n)\}^{-1}$$

satisfies $(a_n^T \hat\Sigma_n a_n)/(a_n^T \Sigma_n a_n) \to_p 1$, and

$$\frac{\sqrt{n}\, a_n^T(\hat\eta_n - \eta_{0,n})}{(a_n^T \hat\Sigma_n a_n)^{1/2}} \;\to_d\; N(0,1).$$

(iii) ***(Pointwise inference for the infinite-dimensional parameter.)*** *If, in addition,* (B7) *holds, and in the bundled-parameter case* (B3*) *holds, then for each fixed*

x,

$$\sqrt{\frac{n}{p_n}}\Big(\hat{\zeta}_n(x,\hat{\beta}_n) - \zeta_0(x,\beta_0)\Big) \;\to_d\; b(x) + \sigma_2(x)\, Z,$$

where $Z \sim N(0,1)$ *and the variance* $\sigma_2^2(x) = \lim_{n\to\infty} p_n^{-1} e_n(x)^\top \Sigma_n e_n(x)$, $e_n(x) = (0_d^\top, B_{k_n,1}(\phi(x;\beta_{0,n})), \ldots, B_{k_n,k_n}(\phi(x;\beta_{0,n})))^\top$.

Proof of Theorem 4.1. The proof centers the Taylor expansion at the sieve pseudo-truth $\theta_{0,n} \in \Theta_n$ and works in the growing-dimensional sieve coordinate $\eta \in \mathbb{R}^{p_n}$. To avoid questions about o_p in growing-dimensional spaces, we project onto the direction a_n at the outset.

Scalar linearization. Let $a_n \in \mathbb{R}^{p_n}$ be arbitrary. Define $c_n = \big[J^{(n)}(\theta_{0,n})\big]^{-1} a_n$. Since $\theta_{0,n}$ maximizes S over Θ_n, the population sieve score vanishes meaning $U^{(n)}(\eta_{0,n}) = \dot{S}^{(n)}(\theta_{0,n}) = 0$. By (B6), $\|U_n^{(n)}(\hat{\eta}_n)\| = o_p(n^{-1/2})$, so $c_n^T U_n^{(n)}(\hat{\eta}_n) = o_p(n^{-1/2}\|c_n\|)$.

A Taylor expansion of $U_n^{(n)}$ around $\eta_{0,n}$ gives

$$U_n^{(n)}(\hat{\eta}_n) = U_n^{(n)}(\eta_{0,n}) - J_n^{(n)}(\eta_{0,n})(\hat{\eta}_n - \eta_{0,n}) + R_n^{(n)}, \tag{4.1}$$

where the remainder is

$$R_n^{(n)} = \big[J_n^{(n)}(\eta_{0,n}) - J_n^{(n)}(\tilde{\eta}_n)\big](\hat{\eta}_n - \eta_{0,n})$$

for some $\tilde{\eta}_n$ on the segment between $\hat{\eta}_n$ and $\eta_{0,n}$. Projecting (4.1) by c_n^T and using $c_n^T U_n^{(n)}(\hat{\eta}_n) = o_p(n^{-1/2}\|c_n\|)$:

$$c_n^T J_n^{(n)}(\eta_{0,n})(\hat{\eta}_n - \eta_{0,n}) = c_n^T U_n^{(n)}(\eta_{0,n}) + c_n^T R_n^{(n)} + o_p(n^{-1/2}\|c_n\|). \tag{4.2}$$

Since $\tilde{\eta}_n$ lies on the segment between $\hat{\eta}_n$ and $\eta_{0,n}$, we have $\|\tilde{\eta}_n - \eta_{0,n}\| \le \|\hat{\eta}_n - \eta_{0,n}\|$. By (B3)

and the triangle inequality,

$$\|\hat{\eta}_n - \eta_{0,n}\| \leq d(\hat{\theta}_n, \theta_{0,n}) \leq d(\hat{\theta}_n, \theta_0) + d(\theta_0, \theta_{0,n}) = O_p(n^{-\xi}) + O(n^{-\xi}) = O_p(n^{-\xi}),$$

where the first equality uses Lemma 4.1(ii) combined with (B3). Therefore $\|\tilde{\eta}_n - \eta_{0,n}\| = O_p(n^{-\xi})$. By Lemma 4.2(i),

$$\|J_n^{(n)}(\tilde{\eta}_n) - J_n^{(n)}(\eta_{0,n})\|_{\text{op}} = O_p(p_n n^{-\xi}).$$

Combining,

$$|c_n^T R_n^{(n)}| \leq \|c_n\| \, \|J_n^{(n)}(\eta_{0,n}) - J_n^{(n)}(\tilde{\eta}_n)\|_{\text{op}} \, \|\hat{\eta}_n - \eta_{0,n}\| = \|c_n\| \cdot O_p(p_n n^{-2\xi}).$$

By Lemma 4.2(ii),

$$\|J_n^{(n)}(\eta_{0,n}) - J^{(n)}(\theta_{0,n})\|_{\text{op}} = O_p\left(\max\left\{p_n n^{-\xi}, p_n\sqrt{\frac{\log p_n}{n}}\right\}\right),$$

so

$$\begin{aligned}\left|c_n^T\{J_n^{(n)}(\eta_{0,n}) - J^{(n)}(\theta_{0,n})\}(\hat{\eta}_n - \eta_{0,n})\right| &\leq \|c_n\| \cdot O_p\left(\max\left\{p_n n^{-\xi}, p_n\sqrt{\frac{\log p_n}{n}}\right\}\right) \cdot O_p(n^{-\xi}) \\ &= \|c_n\| \cdot O_p\left(\max\left\{p_n n^{-2\xi}, n^{-\xi} p_n\sqrt{\frac{\log p_n}{n}}\right\}\right).\end{aligned}$$

To normalize these remainder terms, first note that by (B5) and the definition of c_n we have

$$c_V\|c_n\|^2 \leq c_n^T V_n c_n = a_n^T \Sigma_n a_n \leq C_V\|c_n\|^2. \tag{4.3}$$

Normalizing the remainder terms by $\sqrt{a_n^T \Sigma_n a_n}$ and using $c_n^T V_n c_n \geq c_V \|c_n\|^2$ from (4.3),

$$\frac{|c_n^T R_n^{(n)}|}{\sqrt{a_n^T \Sigma_n a_n}} \leq \frac{\|c_n\| \cdot O_p(p_n n^{-2\xi})}{\sqrt{c_V \|c_n\|^2}} = \frac{O_p(p_n n^{-2\xi})}{\sqrt{c_V}} = o_p(n^{-1/2}),$$

by (B3), and similarly

$$\begin{aligned}\frac{\left|c_n^T \{J_n^{(n)}(\eta_{0,n}) - J^{(n)}(\theta_{0,n})\}(\hat{\eta}_n - \eta_{0,n})\right|}{\sqrt{a_n^T \Sigma_n a_n}} &\leq \frac{\|c_n\| \cdot O_p\left(n^{-\xi} p_n \sqrt{\frac{\log p_n}{n}}\right)}{\sqrt{c_V \|c_n\|^2}} \\ &= \frac{O_p\left(n^{-\xi} p_n \sqrt{\frac{\log p_n}{n}}\right)}{\sqrt{c_V}} = o_p(n^{-1/2}),\end{aligned}$$

since $n^{-\xi} p_n \sqrt{\log p_n} \to 0$.

By symmetry of $J^{(n)}(\theta_{0,n})$, we have $c_n^T J^{(n)}(\theta_{0,n}) = a_n^T$, so combining (4.2) with the above remainder bounds gives

$$\begin{aligned}\frac{a_n^T(\hat{\eta}_n - \eta_{0,n})}{\sqrt{a_n^T \Sigma_n a_n}} &= \frac{c_n^T}{\sqrt{a_n^T \Sigma_n a_n}} U_n^{(n)}(\eta_{0,n}) + o_p(n^{-1/2}) \\ &= \frac{a_n^T \left[J^{(n)}(\theta_{0,n})\right]^{-1}}{\sqrt{a_n^T \Sigma_n a_n}} \frac{1}{n} \sum_{i=1}^n u_i^{(n)}(\theta_{0,n}) + o_p(n^{-1/2}). \qquad (4.4)\end{aligned}$$

Part (i). Define $W_{n,i} = c_n^T u_i^{(n)}(\theta_{0,n})$. Multiplying (4.4) by $\sqrt{n}$ and plugging in,

$$\frac{\sqrt{n}\, a_n^T(\hat{\eta}_n - \eta_{0,n})}{\sqrt{a_n^T \Sigma_n a_n}} = \frac{1}{\sqrt{n}} \sum_{i=1}^n W_{n,i} + o_p(1).$$

We have $a_n^T \Sigma_n a_n = c_n^T V_n c_n$, so

$$\frac{\sqrt{n}\, a_n^T(\hat{\eta}_n - \eta_{0,n})}{\sqrt{a_n^T \Sigma_n a_n}} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{W_{n,i}}{\sqrt{c_n^T V_n c_n}} + o_p(1).$$

Since $E[u_i^{(n)}(\theta_{0,n})] = 0$, we have $E[W_{n,i}] = 0$. Define the normalizing constant $\sigma_n^2 = c_n^T V_n c_n = a_n^T \Sigma_n a_n$. By assumption, $\liminf_{n\to\infty} \sigma_n^2 > 0$. Define normalized variables

$$Z_{n,i} = \frac{W_{n,i}}{\sigma_n}, \qquad E[Z_{n,i}] = 0, \qquad \mathrm{Var}(Z_{n,i}) = \frac{c_n^T V_n c_n}{\sigma_n^2} = 1.$$

We verify the Lindeberg condition for $X_{n,i} = Z_{n,i}/\sqrt{n}$ with row variances $s_n^2 = \sum_{i=1}^n \mathrm{Var}(X_{n,i}) = 1$. For any $\varepsilon > 0$,

$$L_n(\varepsilon) = \sum_{i=1}^n E\big[X_{n,i}^2 \, \mathbf{1}\{|X_{n,i}| > \varepsilon\}\big] \to 0. \tag{4.5}$$

By coordinatewise boundedness of the observation-level score, $\|u_i^{(n)}(\theta_{0,n})\| \le A_1\sqrt{p_n}$ and we have $|W_{n,i}| = |c_n^T u_i^{(n)}(\theta_{0,n})| \le A_1\sqrt{p_n}\|c_n\|$ a.s. Using $\sigma_n^2 = c_n^T V_n c_n \ge c_V\|c_n\|^2$ from (4.3),

$$|Z_{n,i}| = \frac{|W_{n,i}|}{\sigma_n} \le \frac{A_1\sqrt{p_n}\|c_n\|}{\sqrt{c_V\|c_n\|^2}} = \frac{A_1\sqrt{p_n}}{\sqrt{c_V}}.$$

Therefore $|X_{n,i}| = |Z_{n,i}|/\sqrt{n} \le A_1\sqrt{p_n}/(\sqrt{c_V n}) \to 0$ by (B3), so for large n, $|X_{n,i}| < \varepsilon$ and the indicator in (4.5) vanishes. By the Lindeberg–Feller CLT,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n Z_{n,i} \to_d N(0,1),$$

which gives the desired result.

Part (ii). Define $\hat{c}_n = \Big[J_n^{(n)}(\hat{\theta}_n)\Big]^{-1} a_n$ and $\hat{V}_n = \hat{V}_n(\hat{\theta}_n)$ so that $a_n^\top \hat{\Sigma}_n a_n = \hat{c}_n^\top \hat{V}_n \hat{c}_n$. We will show

$$\frac{\hat{c}_n^\top \hat{V}_n \hat{c}_n}{c_n^\top V_n c_n} = \frac{c_n^\top V_n c_n + (\hat{c}_n^\top \hat{V}_n \hat{c}_n - c_n^\top V_n c_n)}{c_n^\top V_n c_n} = 1 + o_p(1)$$

Decompose

$$\hat{c}_n^\top \hat{V}_n \hat{c}_n - c_n^\top V_n c_n = \hat{c}_n^\top (\hat{V}_n - V_n)\hat{c}_n + \{\hat{c}_n^\top V_n \hat{c}_n - c_n^\top V_n c_n\}.$$

For the first term, we have

$$\left| \frac{\hat{c}_n^\top (\hat{V}_n - V_n)\hat{c}_n}{c_n^\top V_n c_n} \right| \leq \frac{\|\hat{V}_n - V_n\|_{\text{op}} \|\hat{c}_n\|^2}{c_V \|c_n\|^2}.$$

Let

$$\begin{aligned} r_{J,n} = \hat{c}_n - c_n &= \left(\left[J_n^{(n)}(\hat{\theta}_n) \right]^{-1} - \left[J^{(n)}(\theta_{0,n}) \right]^{-1} \right) a_n \\ &= \left[J_n^{(n)}(\hat{\theta}_n) \right]^{-1} \left(J_n^{(n)}(\hat{\theta}_n) - J^{(n)}(\theta_{0,n}) \right) c_n \end{aligned}$$

where the last equality holds by the standard equivalence $A^{-1} - B^{-1} = A^{-1}(B - A)B^{-1}$. By Lemma 4.3, we have $\|J_n^{(n)}(\hat{\theta}_n) - J^{(n)}(\theta_{0,n})\|_{\text{op}} = o_p(1)$ and from Lemma 4.2(iii) we have $\left[J_n^{(n)}(\hat{\theta}_n) \right]^{-1} = O_p(1)$. Thus,

$$\begin{aligned} \|r_{J,n}\| &\leq \left\| \left[J_n^{(n)}(\hat{\theta}_n) \right]^{-1} \right\|_{\text{op}} \left\| \left(J_n^{(n)}(\hat{\theta}_n) - J^{(n)}(\theta_{0,n}) \right) \right\|_{\text{op}} \|c_n\| \\ &= O_p(1) \cdot o_p(1) \cdot \|c_n\| \\ &= o_p(1) \cdot \|c_n\|, \end{aligned}$$

and applying Lemma 4.3 with $\|\hat{V}_n - V_n\|_{\text{op}} = o_p(1)$ it follows that

$$\left| \frac{\hat{c}_n^\top (\hat{V}_n - V_n)\hat{c}_n}{c_n^\top V_n c_n} \right| \leq \frac{\|\hat{V}_n - V_n\|_{\text{op}} \|\hat{c}_n\|^2}{c_V \|c_n\|^2} = \frac{o_p(1) \cdot O_p(1) \cdot \|c_n\|^2}{c_V \|c_n\|^2} = o_p(1)$$

For the second part of the decomposition, write

$$\hat{c}_n^\top V_n \hat{c}_n - c_n^\top V_n c_n = 2c_n^\top V_n r_{J,n} + r_{J,n}^\top V_n r_{J,n}.$$

Using the uniform upper and lower bounds on V_n, we know $c_n^\top V_n c_n \geq c_V \|c_n\|^2$, and $\|V_n\|_{\text{op}} \leq C_V$. Thus we obtain

$$\frac{|2c_n^\top V_n r_{J,n}|}{c_n^\top V_n c_n} \leq \frac{2\|c_n\|\|V_n\|_{\text{op}}\|r_{J,n}\|}{c_V\|c_n\|^2} \leq \frac{2C_V}{c_V}\frac{\|r_{J,n}\|}{\|c_n\|} = o_p(1),$$

and similarly

$$\frac{|r_{J,n}^\top V_n r_{J,n}|}{c_n^\top V_n c_n} \leq \frac{C_V\|r_{J,n}\|^2}{c_V\|c_n\|^2} = o_p(1).$$

Therefore

$$\frac{\hat{c}_n^\top V_n \hat{c}_n - c_n^\top V_n c_n}{c_n^\top V_n c_n} = o_p(1).$$

Combining this with the first part gives

$$\frac{a_n^\top \hat{\Sigma}_n a_n}{a_n^\top \Sigma_n a_n} = \frac{\hat{c}_n^\top \hat{V}_n \hat{c}_n}{c_n^\top V_n c_n} \to_p 1.$$

By Slutsky's theorem,

$$\frac{\sqrt{n}\, a_n^\top(\hat{\eta}_n - \eta_{0,n})}{(a_n^\top \hat{\Sigma}_n a_n)^{1/2}} \to_d N(0,1).$$

Part (iii). For each fixed x, let

$$e_n(x) = \left(0_d^\top, B_{k_n,1}(\phi(x;\beta_{0,n})), \ldots, B_{k_n,k_n}(\phi(x;\beta_{0,n}))\right)^\top \in \mathbb{R}^{p_n}.$$

We decompose

$$\hat{\zeta}_n(x,\hat{\beta}_n) - \zeta_0(x,\beta_0) = \underbrace{e_n(x)^T(\hat{\eta}_n - \eta_{0,n})}_{\text{(I)}} + \underbrace{\big(\zeta_{0,n}(x,\beta_{0,n}) - \zeta_0(x,\beta_0)\big)}_{\text{(II)}} + \underbrace{r_n(x)}_{\text{(III)}},$$

where $r_n(x) = \hat{\zeta}_n(x,\hat{\beta}_n) - \hat{\zeta}_n(x,\beta_{0,n})$ arises from the β-dependence of the basis, and is only non-zero in the bundled parameter setting.

For term (I), set $a_n = e_n(x)$ and define

$$\sigma_n^2(x) = e_n(x)^T \Sigma_n \, e_n(x).$$

By assumption (B5) and Lemma 4.2(iii), Σ_n is bounded in operator norm, so $p_n^{-1}\,\sigma_n^2(x) \to \sigma_2^2(x)$.

Applying part (i) with $a_n = e_n(x)$ we have

$$\frac{\sqrt{n}\, e_n(x)^T(\hat{\eta}_n - \eta_{0,n})}{\sigma_n(x)} \;\to_d\; N(0,1).$$

Rescaling by $\sqrt{n/p_n}$ and using $\sigma_n^2(x)/p_n \to \sigma_2^2(x)$,

$$\sqrt{\frac{n}{p_n}}\, e_n(x)^T(\hat{\eta}_n - \eta_{0,n}) = \frac{\sigma_n(x)}{\sqrt{p_n}} \cdot \frac{\sqrt{n}\, e_n(x)^\top(\hat{\eta}_n - \eta_{0,n})}{\sigma_n(x)} \;\to_d\; N\big(0,\, \sigma_2^2(x)\big),$$

by Slutsky's.

Term (II) is the sieve approximation bias. By (B7), $\sqrt{n/p_n}(\zeta_{0,n}(x,\beta_{0,n}) - \zeta_0(x,\beta_0)) \to b(x)$.

Term (III) is identically zero in the case of non-bundled parameters. Otherwise, by Remark 4.2 we have $r_n(x) = \hat{\zeta}_n(x,\hat{\beta}_n) - \hat{\zeta}_n(x,\beta_{0,n}) \le L_\zeta(x)\|\hat{\beta}_n - \beta_{0,n}\|$. Since $\|\hat{\eta}_n - \eta_{0,n}\| = O_p(n^{-\xi})$ by (B3) and the triangle inequality (applying Lemma 4.1), we have $|r_n(x)| = O_p(n^{-\xi})$, and thus $\sqrt{n/p_n}|r_n(x)| = O_p(n^{-\xi}\sqrt{n/p_n}) = o_p(1)$ by (B3*).

Combining terms (I)–(III) and multiplying through by $\sqrt{n/p_n}$, we have

$$\sqrt{\frac{n}{p_n}}\big(\hat{\zeta}_n(x,\hat{\beta}_n) - \zeta_0(x,\beta_0)\big) \;\to_d\; b(x) + \sigma_2(x)\, Z.$$

□

Bibliography

J. N. Adichie. Estimates of Regression Parameters Based on Rank Tests. *Ann. Math. Statist.*, 38(3):894–904, June 1967. ISSN 0003-4851. doi: 10.1214/aoms/1177698883. URL http://projecteuclid.org/euclid.aoms/1177698883.

Neelum T. Aggarwal and Michelle M. Mielke. Sex Differences in Alzheimer's Disease. *Neurologic Clinics*, 41(2):343–358, May 2023. ISSN 07338619. doi: 10.1016/j.ncl.2023.01.001. URL https://linkinghub.elsevier.com/retrieve/pii/S0733861923000014.

William E. Barlow and Ross L. Prentice. Residuals for relative risk regression. *Biometrika*, 75(1):65–74, 1988. ISSN 0006-3444, 1464-3510. doi: 10.1093/biomet/75.1.65. URL https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/75.1.65.

Duane L. Beekly, Erin M. Ramos, William W. Lee, Woodrow D. Deitrich, Mary E. Jacka, Joylee Wu, Janene L. Hubbard, Thomas D. Koepsell, John C. Morris, and Walter A. Kukull. The National Alzheimer's Coordinating Center (NACC) Database: The Uniform Data Set. *Alzheimer Disease & Associated Disorders*, 21(3):249–258, July 2007. ISSN 0893-0341. doi: 10.1097/WAD.0b013e318142774e. URL https://journals.lww.com/00002093-200707000-00009.

Tim Benson. Why it is hard to use PROMs and PREMs in routine health and care. *BMJ Open Quality*, 12(4):e002516, December 2023. ISSN 2399-6641. doi: 10.1136/bmjoq-2023-002516. URL https://qir.bmj.com/lookup/doi/10.1136/bmjoq-2023-002516.

P. K. Bhattacharya, Herman Chernoff, and S. S. Yang. Nonparametric Estimation of the Slope of a Truncated Regression. *Ann. Statist.*, 11(2):505–514, 1983. ISSN 00905364. URL http://www.jstor.org/stable/2240565.

Yolanda Bravo Vergel and Mark Sculpher. Quality-adjusted life years. *Practical Neurology*, 8(3):175–182, June 2008. ISSN 1474-7758, 1474-7766. doi: 10.1136/pn.2007.140186. URL https://pn.bmj.com/lookup/doi/10.1136/pn.2007.140186.

David Cella, Susan Yount, Nan Rothrock, Richard Gershon, Karon Cook, Bryce Reeve, Deborah Ader, James F. Fries, Bonnie Bruce, and Mattias Rose. The Patient-Reported Outcomes Measurement Information System (PROMIS): Progress of an NIH Roadmap Cooperative Group During its First Two Years. *Medical Care*, 45(5):S3–S11, May 2007.

ISSN 0025-7079. doi: 10.1097/01.mlr.0000258615.42478.55. URL https://journals.lww.com/00005650-200705001-00002.

National Alzheimer's Coordinating Center. NACC data and resources, 2023. URL https://www.naccdata.org.

Li-Pang Chen. Feature screening via concordance indices for left-truncated and right-censored survival data. *Journal of Statistical Planning and Inference*, 232:106153, September 2024. ISSN 03783758. doi: 10.1016/j.jspi.2024.106153. URL https://linkinghub.elsevier.com/retrieve/pii/S0378375824000107.

Li-Pang Chen and Bangxu Qiu. Analysis of Length-Biased and Partly Interval-Censored Survival Data with Mismeasured Covariates. *Biometrics*, 79(4):3929–3940, December 2023. ISSN 0006-341X, 1541-0420. doi: 10.1111/biom.13898. URL https://academic.oup.com/biometrics/article/79/4/3929/7587556.

Li-Pang Chen and Grace Y. Yi. Semiparametric methods for left-truncated and right-censored survival data with covariate measurement error. *Annals of the Institute of Statistical Mathematics*, 73(3):481–517, June 2021. ISSN 0020-3157, 1572-9052. doi: 10.1007/s10463-020-00755-2. URL https://link.springer.com/10.1007/s10463-020-00755-2.

Xiaohong Chen. Large Sample Sieve Estimation of Semi-Nonparametric Models. In *Handbook of Econometrics*, volume 6, pages 5549–5632. Elsevier, 2007. ISBN 978-0-444-53200-8. doi: 10.1016/S1573-4412(07)06076-X. URL https://linkinghub.elsevier.com/retrieve/pii/S157344120706076X.

M. M. Corrada, R. Brookmeyer, D. Berlau, A. Paganini-Hill, and C. H. Kawas. Prevalence of dementia after age 90: Results from The 90+ Study. *Neurology*, 71(5):337–343, July 2008. ISSN 0028-3878, 1526-632X. doi: 10.1212/01.wnl.0000310773.65918.cd. URL https://www.neurology.org/doi/10.1212/01.wnl.0000310773.65918.cd.

María M. Corrada, Ron Brookmeyer, Annlia Paganini-Hill, Daniel Berlau, and Claudia H. Kawas. Dementia incidence continues to increase with age in the oldest old: The 90+ study. *Ann. Neuro.*, 67(1):114–121, January 2010. ISSN 0364-5134, 1531-8249. doi: 10.1002/ana.21915. URL https://onlinelibrary.wiley.com/doi/10.1002/ana.21915.

D. R. Cox. Regression Models and Life-Tables. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 34(2):187–202, January 1972. ISSN 1369-7412, 1467-9868. doi: 10.1111/j.2517-6161.1972.tb00899.x. URL https://academic.oup.com/jrsssb/article/34/2/187/7027194.

S. Craft, L. Teri, S. D. Edland, W. A. Kukull, G. Schellenberg, W. C. McCormick, J. D. Bowen, and E. B. Larson. Accelerated decline in apolipoprotein E-e4 homozygotes with Alzheimer's disease. *Neurology*, 51(1):149–153, July 1998. ISSN 0028-3878, 1526-632X. doi: 10.1212/WNL.51.1.149. URL https://www.neurology.org/doi/10.1212/WNL.51.1.149.

Carl De Boor. *A practical guide to splines*. Springer, New York, rev. edition, 2001. ISBN 978-0-387-95366-3.

Vakaramoko Diaby, Georges Adunlin, Askal Ayalew Ali, and Rima Tawk. Using quality-adjusted progression-free survival as an outcome measure to assess the benefits of cancer drugs in randomized-controlled trials: case of the BOLERO-2 trial. *Breast Cancer Research and Treatment*, 146(3):669–673, August 2014. ISSN 0167-6806, 1573-7217. doi: 10.1007/s10549-014-3047-y. URL http://link.springer.com/10.1007/s10549-014-3047-y.

Ying Ding. *Some New Insights about the Accelerated Failure Time Model*. PhD thesis, The University of Michigan, Michigan, U.S.A., 2010.

Ying Ding and Bin Nan. A sieve M-theorem for bundled parameters in semiparametric models, with application to the efficient estimation in a linear model for censored data. *Ann. Statist.*, 39(6):2795–3443, 2011a.

Ying Ding and Bin Nan. Supplement to "A sieve M-theorem for bundled parameters in semiparametric models, with application to the efficient estimation in a linear model for censored data.". *Ann. Statist.*, 39(6), 2011b. doi: 10.1214/11-AOS934SUPP.

Bradley Efron. Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics: Methodology and distribution*, pages 569–593. Springer, 1992.

Juan Fortea, Jordi Pegueroles, Daniel Alcolea, Olivia Belbin, Oriol Dols-Icardo, Lídia Vaqué-Alcázar, Laura Videla, Juan Domingo Gispert, Marc Suárez-Calvet, Sterling C. Johnson, Reisa Sperling, Alexandre Bejanin, Alberto Lleó, and Víctor Montal. APOE4 homozygosity represents a distinct genetic form of Alzheimer's disease. *Nature Medicine*, 30(5):1284–1291, May 2024. ISSN 1078-8956, 1546-170X. doi: 10.1038/s41591-024-02931-w. URL https://www.nature.com/articles/s41591-024-02931-w.

Armin Garmany, Satsuki Yamada, and Andre Terzic. Longevity leap: mind the healthspan gap. *npj Regenerative Medicine*, 6(1):57, September 2021. ISSN 2057-3995. doi: 10.1038/s41536-021-00169-5. URL https://www.nature.com/articles/s41536-021-00169-5.

Stuart Geman and Chii-Ruey Hwang. Nonparametric Maximum Likelihood Estimation by the Method of Sieves. *The Annals of Statistics*, 10(2):401–414, 1982. ISSN 00905364, 21688966. URL http://www.jstor.org/stable/2240675.

David W. Hosmer and Stanley Lemesbow. Goodness of fit tests for the multiple logistic regression model. *Communications in Statistics - Theory and Methods*, 9(10):1043–1069, January 1980. ISSN 0361-0926. doi: 10.1080/03610928008827941. URL https://www.tandfonline.com/doi/abs/10.1080/03610928008827941.

John D. Kalbfleisch and Ross L. Prentice. *The statistical analysis of failure time data*. J. Wiley, Hoboken (N.J.), 2nd ed edition, 2002. ISBN 978-0-471-36357-6.

Selin Karakose, Martina Luchetti, Yannick Stephan, Angelina R. Sutin, and Antonio Terracciano. Marital status and risk of dementia over 18 years: Surprising findings

from the National Alzheimer's Coordinating Center. *Alzheimer's & Dementia*, 21(3): e70072, March 2025. ISSN 1552-5260, 1552-5279. doi: 10.1002/alz.70072. URL https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.70072.

Chul-Ki Kim and Tze Leung Lai. Efficient score estimation and adaptive M-estimators in censored and truncated regression models. *Stat. Sin*, 10(3):731–749, 2000. ISSN 1017-0405.

Shengchun Kong and Bin Nan. Semiparametric approach to regression with a covariate subject to a detection limit. *Biometrika*, 103(1):161–174, March 2016. ISSN 0006-3444, 1464-3510. doi: 10.1093/biomet/asv055. URL https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/asv055.

Shengchun Kong, Bin Nan, John D. Kalbfleisch, Rajiv Saran, and Richard Hirth. Conditional Modeling of Longitudinal Data With Terminal Event. *Journal of the American Statistical Association*, 113(521):357–368, January 2018. ISSN 0162-1459, 1537-274X. doi: 10.1080/01621459.2016.1255637. URL https://www.tandfonline.com/doi/full/10.1080/01621459.2016.1255637.

Erica Kornblith, Amber Bahorik, W. John Boscardin, Feng Xia, Deborah E. Barnes, and Kristine Yaffe. Association of Race and Ethnicity With Incidence of Dementia Among Older Adults. *JAMA*, 327(15):1488, April 2022. ISSN 0098-7484. doi: 10.1001/jama.2022.3550. URL https://jamanetwork.com/journals/jama/fullarticle/2791223.

Michael R. Krogsgaard, John Brodersen, Jonas Jensen, Christian Fugl Hansen, and Jonathan D. Comins. Potential problems in the use of patient reported outcome measures (PROMs) and reporting of PROM data in sports science. *Scandinavian Journal of Medicine & Science in Sports*, 31(6):1249–1258, June 2021. ISSN 0905-7188, 1600-0838. doi: 10.1111/sms.13888. URL https://onlinelibrary.wiley.com/doi/10.1111/sms.13888.

Tze Leung Lai and Zhiliang Ying. Large Sample Theory of a Modified Buckley-James Estimator for Regression Analysis with Censored Data. *The Annals of Statistics*, 19(3): 1370–1402, 1991a. ISSN 00905364, 21688966. URL http://www.jstor.org/stable/2241954.

Tze Leung Lai and Zhiliang Ying. Rank Regression Methods for Left-Truncated and Right-Censored Data. *Ann. Statist.*, 19(2):531–556, 1991b. ISSN 00905364. URL http://www.jstor.org/stable/2242073.

Tze Leung Lai and Zhiliang Ying. Asymptotically efficient estimation in censored and truncated regression models. *Stat. Sin.*, 2(1):17–46, 1992a. ISSN 1017-0405. URL https://www.jstor.org/stable/24304118.

Tze Leung Lai and Zhiliang Ying. Linear rank statistics in regression analysis with censored or truncated data. *J. of Multivar. Anal.*, 40(1):13–45, January 1992b. ISSN 0047259X. doi: 10.1016/0047-259X(92)90088-W. URL https://linkinghub.elsevier.com/retrieve/pii/0047259X9290088W.

Tze Leung Lai and Zhiliang Ying. A Missing Information Principle and M-Estimators in Regression Analysis with Censored and Truncated Data. *Ann. Statist.*, 22(3):1222–1255, 1994. ISSN 00905364. URL http://www.jstor.org/stable/2242224.

E. L. Lehmann and George Casella. *Theory of point estimation*. Springer texts in statistics. Springer, New York, 2nd ed edition, 1998. ISBN 978-0-387-98502-2.

Jack C. Lennon, Stephen L. Aita, Victor A. Del Bene, Tasha Rhoads, Zachary J. Resch, Janelle M. Eloi, and Keenan A. Walker. Black and White individuals differ in dementia prevalence, risk factors, and symptomatic presentation. *Alzheimer's & Dementia*, 18(8): 1461–1471, August 2022. ISSN 1552-5260, 1552-5279. doi: 10.1002/alz.12509. URL https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.12509.

D. Y. Lin, L. J. Wei, and Z. Ying. Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika*, 80(3):557–572, 1993. ISSN 0006-3444, 1464-3510. doi: 10.1093/biomet/80.3.557. URL https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/80.3.557.

Dungang Liu and Heping Zhang. Residuals and Diagnostics for Ordinal Regression Models: A Surrogate Approach. *Journal of the American Statistical Association*, 113(522):845–854, April 2018. ISSN 0162-1459, 1537-274X. doi: 10.1080/01621459.2017.1292915. URL https://www.tandfonline.com/doi/full/10.1080/01621459.2017.1292915.

Amanda L. Lumsden, Anwar Mulugeta, Ang Zhou, and Elina Hyppönen. Apolipoprotein E (APOE) genotype-associated disease risks: a phenome-wide, registry-based, case-control study utilising the UK Biobank. *eBioMedicine*, 59:102954, September 2020. ISSN 23523964. doi: 10.1016/j.ebiom.2020.102954. URL https://linkinghub.elsevier.com/retrieve/pii/S2352396420303303.

Spencer Matthews and Bin Nan. Regression for Left-Truncated and Right-Censored Data: A Semiparametric Sieve Likelihood Approach. *Statistics in Medicine*, 45(8-9):e70509, April 2026. ISSN 0277-6715, 1097-0258. doi: 10.1002/sim.70509. URL https://onlinelibrary.wiley.com/doi/10.1002/sim.70509.

Zarui A Melikyan, Dana E Greenia, Maria M Corrada, Marilyn M Hester, Claudia H Kawas, and Joshua D Grill. Recruiting the oldest-old for clinical research. *Alzheimer Disease & Associated Disorders*, 33(2):160–162, 2019. ISSN 0893-0341.

Richard Merrick and Carol Brayne. Sex Differences in Dementia, Cognition, and Health in the Cognitive Function and Ageing Studies (CFAS). *Journal of Alzheimer's Disease*, 100 (s1):S3–S12, August 2024. ISSN 13872877, 18758908. doi: 10.3233/JAD-240358. URL https://journals.sagepub.com/doi/full/10.3233/JAD-240358.

John C. Morris, Sandra Weintraub, Helena C. Chui, Jeffrey Cummings, Charles DeCarli, Steven Ferris, Norman L. Foster, Douglas Galasko, Neill Graff-Radford, Elaine R. Peskind, Duane Beekly, Erin M. Ramos, and Walter A. Kukull. The Uniform Data Set (UDS): Clinical and Cognitive Variables and Descriptive Data From Alzheimer Disease Centers. *Alzheimer Disease & Associated Disorders*, 20(4):210–216, October 2006. ISSN

0893-0341. doi: 10.1097/01.wad.0000213865.09806.92. URL https://journals.lww.com/00002093-200610000-00007.

Jing Ning, Jing Qin, and Yu Shen. Buckley-James-Type Estimator with Right-Censored and Length-Biased Data. *Biometrics*, 67(4):1369–1378, December 2011. ISSN 0006341X. doi: 10.1111/j.1541-0420.2011.01568.x. URL https://academic.oup.com/biometrics/article/67/4/1369-1378/7381158.

Jing Ning, Jing Qin, and Yu Yu. Semiparametric accelerated failure time model for length-biased data with application to dementia study. *Statistica Sinica*, 2014. ISSN 10170405. doi: 10.5705/ss.2011.197. URL http://www3.stat.sinica.edu.tw/statistica/J24N1/J24N116/J24N116.html.

S. Jay Olshansky. From Lifespan to Healthspan. *JAMA*, 320(13):1323, October 2018. ISSN 0098-7484. doi: 10.1001/jama.2018.12621. URL http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2018.12621.

Daryl Pregibon. Logistic Regression Diagnostics. *The Annals of Statistics*, 9 (4), July 1981. ISSN 0090-5364. doi: 10.1214/aos/1176345513. URL https://projecteuclid.org/journals/annals-of-statistics/volume-9/issue-4/Logistic-Regression-Diagnostics/10.1214/aos/1176345513.full.

Katrine L Rasmussen, Anne Tybjærg-Hansen, Børge G Nordestgaard, and Ruth Frikke-Schmidt. Plasma levels of apolipoprotein E, *APOE* genotype, and all-cause and cause-specific mortality in 105 949 individuals from a white general population cohort. *European Heart Journal*, 40(33):2813–2824, September 2019. ISSN 0195-668X, 1522-9645. doi: 10.1093/eurheartj/ehz402. URL https://academic.oup.com/eurheartj/article/40/33/2813/5522543.

Mirna Safieh, Amos D. Korczyn, and Daniel M. Michaelson. ApoE4: an emerging therapeutic target for Alzheimer's disease. *BMC Medicine*, 17(1):64, March 2019. ISSN 1741-7015. doi: 10.1186/s12916-019-1299-4. URL https://doi.org/10.1186/s12916-019-1299-4.

Larry L. Schumaker. *Spline functions: basic theory*. Cambridge university press, Cambridge, 3rd ed edition, 2007. ISBN 978-0-521-70512-7.

Xiaotong Shen and Wing Hung Wong. Convergence Rate of Sieve Estimates. *Ann. Statist.*, 22(2):580–615, 1994. ISSN 00905364. URL http://www.jstor.org/stable/2242281.

Yu Shen, Jing Ning, and Jing Qin. Analyzing Length-Biased Data With Semiparametric Transformation and Accelerated Failure Time Models. *Journal of the American Statistical Association*, 104(487):1192–1202, September 2009. ISSN 0162-1459, 1537-274X. doi: 10.1198/jasa.2009.tm08614. URL http://www.tandfonline.com/doi/abs/10.1198/jasa.2009.tm08614.

Felipe Sierra. The Emergence of Geroscience as an Interdisciplinary Approach to the Enhancement of Health Span and Life Span. *Cold Spring Harbor Perspectives in Medicine*, 6 (4):a025163, April 2016. ISSN 2157-1422. doi: 10.1101/cshperspect.a025163. URL http://perspectivesinmedicine.cshlp.org/lookup/doi/10.1101/cshperspect.a025163.

Andrew Sommerlad, Joshua Ruegger, Archana Singh-Manoux, Glyn Lewis, and Gill Livingston. Marriage and risk of dementia: systematic review and meta-analysis of observational studies. *Journal of Neurology, Neurosurgery & Psychiatry*, 89(3):231–238, March 2018. ISSN 0022-3050, 1468-330X. doi: 10.1136/jnnp-2017-316274. URL https://jnnp.bmj.com/lookup/doi/10.1136/jnnp-2017-316274.

The Canadian Study of Health and Aging Working Group. The Canadian Study of Health and Aging: Risk factors for Alzheimer's disease in Canada. *Neurology*, 44(11):2073–2073, November 1994a. ISSN 0028-3878, 1526-632X. doi: 10.1212/WNL.44.11.2073. URL https://www.neurology.org/doi/10.1212/WNL.44.11.2073.

The Canadian Study of Health and Aging Working Group. Canadian study of health and aging: study methods and prevalence of dementia. *CMAJ: Canadian Medical Association journal = journal de l'Association medicale canadienne*, 150(6):899–913, March 1994b. ISSN 0820-3946.

T. M. Therneau, P. M. Grambsch, and T. R. Fleming. Martingale-based residuals for survival models. *Biometrika*, 77(1):147–160, March 1990. ISSN 0006-3444, 1464-3510. doi: 10.1093/biomet/77.1.147. URL https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/77.1.147.

Alexander M. Vaiserman and Oleh V. Lushchak. Geroscience—the concept. In *Anti-Aging Pharmacology*, pages 1–12. Elsevier, 2023. ISBN 978-0-12-823679-6. doi: 10.1016/B978-0-12-823679-6.00004-7. URL https://linkinghub.elsevier.com/retrieve/pii/B9780128236796000047.

A. W. van der Vaart. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer International Publishing AG, Cham, 2nd ed edition, 2023. ISBN 978-3-031-29038-1 978-3-031-29040-4.

Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. *Compressed Sensing*, pages 210–268, 2012.

Ge Wang and Wei Li. Sex as a Risk Factor for Developing Cognitive Impairments in National Alzheimer's Coordinating Center Participants. *Journal of Alzheimer's Disease Reports*, 5(1):1–6, March 2021. ISSN 2542-4823, 2542-4823. doi: 10.3233/ADR-200275. URL https://journals.sagepub.com/doi/10.3233/ADR-200275.

Mei-Cheng Wang. A Semiparametric Model for Randomly Truncated Data. *Journal of the American Statistical Association*, 84(407):742–748, September 1989. ISSN 0162-1459, 1537-274X. doi: 10.1080/01621459.1989.10478828. URL https://www.tandfonline.com/doi/full/10.1080/01621459.1989.10478828.

Mei-Cheng Wang, Nicholas P Jewell, and Wei-Yann Tsai. Asymptotic properties of the product limit estimate under random truncation. *The Annals of Statistics*, 14(4):1597–1605, 1986. ISSN 0090-5364.

Xuan Wang and Qihua Wang. Semiparametric linear transformation model with differential measurement error and validation sampling. *Journal of Multivariate Analysis*, 141:67–80, October 2015. ISSN 0047259X. doi: 10.1016/j.jmva.2015.05.017. URL https://linkinghub.elsevier.com/retrieve/pii/S0047259X15001505.

Yue Wang, Bin Nan, and John D. Kalbfleisch. Kernel Estimation of Bivariate Time-Varying Coefficient Model for Longitudinal Data with Terminal Event. *Journal of the American Statistical Association*, 119(546):1102–1111, April 2024. ISSN 0162-1459, 1537-274X. doi: 10.1080/01621459.2023.2169702. URL https://www.tandfonline.com/doi/full/10.1080/01621459.2023.2169702.

Milton C. Weinstein and William B. Stason. Foundations of Cost-Effectiveness Analysis for Health and Medical Practices. *New England Journal of Medicine*, 296(13):716–721, March 1977. ISSN 0028-4793, 1533-4406. doi: 10.1056/NEJM197703312961304. URL http://www.nejm.org/doi/abs/10.1056/NEJM197703312961304.

Jon A. Wellner and Ying Zhang. Two likelihood-based semiparametric estimation methods for panel count data with covariates. *Ann. Statist.*, 35(5), October 2007. ISSN 0090-5364. doi: 10.1214/009053607000000181. URL https://projecteuclid.org/journals/annals-of-statistics/volume-35/issue-5/Two-likelihood-based-semiparametric-estimation-methods-for-panel-count-data/10.1214/009053607000000181.full.

Christina Wolfson, David B. Wolfson, Masoud Asgharian, Cyr Emile M'Lan, Truls Østbye, Kenneth Rockwood, and D.B. Hogan. A Reevaluation of the Duration of Survival after the Onset of Dementia. *New England Journal of Medicine*, 344(15):1111–1116, April 2001. ISSN 0028-4793, 1533-4406. doi: 10.1056/NEJM200104123441501. URL http://www.nejm.org/doi/abs/10.1056/NEJM200104123441501.

Xingqiu Zhao, Yuanshan Wu, and Guosheng Yin. Sieve maximum likelihood estimation for a general class of accelerated hazards models with bundled parameters. *Bernoulli*, 23(4B):3385–3411, November 2017. ISSN 1350-7265. doi: 10.3150/16-BEJ850. URL https://projecteuclid.org/journals/bernoulli/volume-23/issue-4B/Sieve-maximum-likelihood-estimation-for-a-general-class-of-accelerated/10.3150/16-BEJ850.full.

Appendix A

Code Availability

The `R` code used for the numerical implementations in Chapter 2, along with an `R` package for fitting models where the response is subject to both left-truncation and right-censoring, can be found on my GitHub page at `https://github.com/srmatth/ltrc`.

The `R` code used for the numerical implementations in Chapter 3 can be found on my GitHub page at `https://github.com/srmatth/ghi`.

Appendix B

Derivation of the Likelihood under Left-Truncation and Right-Censoring

Let $\bar{F}_A(a)$ denote the survival function of an arbitrary random variable A. We proceed with looking at two sub-distributions that correspond to $\Delta = 1$ and $\Delta = 0$, respectively. We have

$$\begin{aligned}
&\text{pr}(Y \leq y, \Delta = \delta \mid Y > T, T = t, X = x) \\
&\quad = \frac{\text{pr}(Y \leq y, \Delta = \delta, Y > t \mid T = t, X = x)}{\text{pr}(Y > t \mid T = t, X = x)} 1(y > t). \qquad \text{(B.1)}
\end{aligned}$$

From our assumption, Y^* is independent of (C, T) given X. Consider first the denominator in (B.1). We have

$$\begin{aligned}
\text{pr}(Y > t \mid T = t, X = x) &= \text{pr}(Y^* > t, C > t \mid T = t, X = x) \\
&= \bar{F}_{Y^*|X}(t \mid x) \bar{F}_{C|T,X}(t \mid t, x). \qquad \text{(B.2)}
\end{aligned}$$

Now consider the numerator in (B.1). For the case where $\Delta = 1$, we have

$$\begin{aligned}
&\text{pr}(Y \leq y, \Delta = 1, Y > t \mid T = t, X = x) \\
&\quad = \text{pr}(Y^* \leq y, Y^* < C, Y^* > t \mid T = t, X = x) \\
&\quad = \int_t^y \int_u^\infty f_{Y^*,C|T,X}(u, v \mid t, x) dv du \\
&\quad = \int_t^y \int_u^\infty f_{Y^*|X}(u \mid x) f_{C|T,X}(v \mid t, x) dv du \\
&\quad = \int_t^y f_{Y^*|X}(u \mid x) \bar{F}_{C|T,X}(u \mid t, x) du.
\end{aligned}$$

When $\Delta = 0$ we have

$$\begin{aligned}
&\text{pr}(Y \leq y, \Delta = 0, Y > t \mid T = t, X = x) \\
&\quad = \text{pr}(C \leq y, Y^* > C, C > t \mid T = t, X = x) \\
&\quad = \int_t^y \int_v^\infty f_{Y^*,C|T,X}(u, v \mid t, x) du dv \\
&\quad = \int_t^y \int_v^\infty f_{Y^*|X}(u \mid x) f_{C|T,X}(v \mid t, x) du dv \\
&\quad = \int_t^y f_{C|T,X}(v \mid t, x) \bar{F}_{Y^*|X}(v \mid x) dv.
\end{aligned}$$

Thus we obtain the following conditional density function:

$$\begin{aligned}
&f_{Y,\Delta|Y>T,T,X}(y, \delta \mid y > t, t, x) \\
&\quad = \frac{\left\{ f_{Y^*|X}(y \mid x) \bar{F}_{C|T,X}(y \mid t, x) \right\}^\delta \left\{ f_{C|T,X}(y \mid t, x) \bar{F}_{Y^*|X}(y \mid x) \right\}^{1-\delta}}{\bar{F}_{Y^*|X}(t \mid x) \bar{F}_{C|T,X}(t \mid t, x)} 1(y > t).
\end{aligned}$$

Then the joint density function of left-truncated data is

$$
\begin{aligned}
&f_{Y,\Delta,T,X|Y>T}(y,\delta,t,x \mid y>t)1(y>t) \\
&\quad = f_{Y,\Delta|T,X,Y>T}(y,\delta \mid t,x,y>t) f_{T,X|Y>T}(t,x \mid y>t)1(y>t) \\
&\quad = \frac{\left\{f_{Y^*|X}(y \mid x)\bar{F}_{C|T,X}(y \mid t,x)\right\}^{\delta}\left\{f_{C|T,X}(y \mid t,x)\bar{F}_{Y^*|X}(y \mid x)\right\}^{1-\delta}}{\bar{F}_{Y^*|X}(t \mid x)\bar{F}_{C|T,X}(t \mid t,x)} \\
&\qquad \times f_{T,X|Y>T}(t,x \mid y>t)1(y>t). \qquad \text{(B.3)}
\end{aligned}
$$

Dropping the factors that are free of (β, λ_e) from (B.3), we obtain the likelihood:

$$
L(\beta,\lambda_e \mid Y,\Delta,T,X) = \frac{\left\{f_{Y^*|X}(Y \mid X)\right\}^{\Delta}\left\{\bar{F}_{Y^*|X}(Y \mid X)\right\}^{1-\Delta}}{\bar{F}_{Y^*|X}(T \mid X)}.
$$

Using hazard functions, we can rewrite the above likelihood function as the following:

$$
\begin{aligned}
L(\beta,\lambda_e \mid Y,\Delta,T,X) &= \frac{\left\{\lambda_{Y^*|X}(Y \mid X)\right\}^{\Delta}\exp\left\{-\int_{-\infty}^{Y}\lambda_{Y^*|X}(u \mid X)du\right\}}{\exp\left\{-\int_{-\infty}^{T}\lambda_{Y^*|X}(u \mid X)du\right\}} \\
&= \left\{\lambda_{Y^*|X}(Y \mid X)\right\}^{\Delta}\exp\left\{-\int_{T}^{Y}\lambda_{Y^*|X}(u \mid X)du\right\} \\
&= \left\{\lambda_e(Y-X^{\mathrm{T}}\beta)\right\}^{\Delta}\exp\left\{-\int_{T-X^{\mathrm{T}}\beta}^{Y-X^{\mathrm{T}}\beta}\lambda_e(u)du\right\}.
\end{aligned}
$$

Appendix C

Proof of Theorems 2.1, 2.2, and 2.3

C.1 Additional Notation for Proofs

Most of the notation in this section will be consistent with that of Chapter 2, with a few different symbols used now that we will be focused in the theoretical space:

- Y is the observed survival time, and $Y = Y^* \wedge C$ where Y^* is the true survival time and C is the censoring time
- X are covariates
- T is the left truncation time
- $\Delta = 1(Y^* < C)$ is the indicator that censoring time is greater than the survival time.
- $\epsilon_\beta = Y - X'\beta$, $\epsilon_0 = Y - X'\beta_0$
- $\tau_\beta = T - X'\beta$, $\tau_0 = T - X'\beta_0$
- a and b are the limits of integration for the sieve space

- Z includes all observed data: (Y, Δ, X, T)
- $\psi(\cdot, \beta) = \cdot - X'\beta$
- $g_0(\cdot) = \log \lambda_0(\psi(\cdot, \beta)) = \zeta(\cdot, \beta)$
- $\hat{g}(s) = \sum_{k=1}^{K} \hat{\gamma}_k B_k(s)$

Now let us define the norm we will be using. This is the same as the norm defined in Equation 4.2 of Ding and Nan [2011a]. First define a collection of functions $\mathcal{H}^P$ as follows:

$$\mathcal{H}^P = \{\zeta(\cdot, \beta) \colon \zeta(s, x, \beta) = g(\psi(s, x, \beta)), g \in \mathcal{G}^P, s \in [a, b], x \in \mathcal{X}, \beta \in \mathcal{B}\},$$

where

$$\psi(s, x, \beta) = s - x'(\beta - \beta_0)$$

and $\mathcal{G}^P$ is defined in our assumptions A.6 in the main work. Here ζ is a composite function of g composed with ψ. Note that $\zeta(s, x, \beta_0) = g(s)$. Then for $\zeta(\cdot, \beta) \in \mathcal{H}^P$ we define the following norm:

$$\|\zeta(\cdot, \beta)\|_2 = \left\{ \int_{\mathcal{X}} \int_a^b \{g(s - x'(\beta - \beta_0))\}^2 \, d\Lambda_0(s) dF_X(x) \right\}^{1/2}. \tag{C.1}$$

We also have the following collection of scores:

$$\mathbb{H} = \left\{ h \colon h(\cdot, \beta) = \left. \frac{\partial \zeta_n(\cdot, \beta)}{\partial \eta} \right|_{\eta=0} = w(\psi(\cdot, \beta)), \zeta_n \in \mathcal{H}^P \right\},$$

in which $h(s, x, \beta) = w(\psi(s, x, \beta)) = w(s - x'(\beta - \beta_0))$.

For any $\theta_1 = (\beta_1, \zeta_1(\cdot, \beta_1))$ and $\theta_2 = (\beta_2, \zeta_2(\cdot, \beta_2))$ in the space of $\Theta^P = \mathcal{B} \times \mathcal{H}^P$, define the

following distance:

$$d(\theta_1, \theta_2) = \{|\beta_1 - \beta_2|^2 + \|\zeta_1(\cdot, \beta_1) - \zeta_2(\cdot, \beta_2)\|_2^2\}^{1/2}. \tag{C.2}$$

We will use this distance frequently in the following proofs.

Furthermore, up to this point we have mostly discussed the likelihood of the data, as

$$\ell(\beta, \gamma \mid X, Y, \Delta, T) = \sum_{i=1}^{n} \left(\delta_i \sum_{k=1}^{K} \gamma_k B_k \left(\epsilon_{\beta,i}\right) - \int_{\tau_{\beta,i}}^{\epsilon_{\beta,i}} \exp\left\{ \sum_{k=1}^{K} \gamma_k B_k(s) \right\} ds \right).$$

In the proofs, we will switch over to discussing the theoretical likelihood, using $\dot{l}$ and $\ddot{l}$ to denote first and second derivatives, respectively, as in Ding and Nan [2011a]. We have the following:

$$\begin{aligned}
l(\beta, \zeta(\cdot, \beta); Z) &= \Delta g(\epsilon_\beta) - \int_a^b 1(s \le \epsilon_\beta) 1(s \ge \tau_\beta) \exp\{g(s)\} ds \\
\dot{l}_\beta(\beta, \zeta(\cdot, \beta); Z) &= -X \left\{ \Delta \dot{g}(\epsilon_\beta) - \int_a^b 1(s \le \epsilon_\beta) 1(s \ge \tau_\beta) \exp\{g(s)\} \dot{g}(s) ds \right\} \\
\dot{l}_\zeta(\beta, \zeta(\cdot, \beta); Z)[h] &= \Delta h(\epsilon_\beta) - \int_a^b 1(s \le \epsilon_\beta) 1(s \ge \tau_\beta) \exp\{g(s)\} h(s) ds \\
\ddot{l}_{\beta\beta}(\beta, \zeta(\cdot, \beta); Z) &= XX' \left\{ \Delta \ddot{g}(\epsilon_\beta) \right. \\
&\qquad \left. - \int_a^b 1(s \le \epsilon_\beta) 1(s \ge \tau_\beta) \exp\{g(s)\} [\dot{g}^2(s) + \ddot{g}(s)] ds \right\} \\
\ddot{l}_{\beta\zeta}(\beta, \zeta(\cdot, \beta); Z)[h] &= -X \left\{ \Delta \dot{h}(\epsilon_\beta) \right. \\
&\qquad \left. - \int_a^b 1(s \le \epsilon_\beta) 1(s \ge \tau_\beta) \exp\{g(s)\} [\dot{h}(s) + \dot{g}(s) h(s)] ds \right\} \\
\ddot{l}_{\zeta\zeta}(\beta, \zeta(\cdot, \beta); Z)[h_1, h_2] &= - \int_a^b 1(s \le \epsilon_\beta) 1(s \ge \tau_\beta) \exp\{g(s)\} h_1(s) h_2(s) ds
\end{aligned}$$

These will be used extensively in the following proofs.

C.2 Technical Lemmas and their Proofs

Lemma C.1. *Under conditions (A.1)–(A.3) and (A.6), the log-likelihood*

$$l(\beta, \zeta(\cdot,\beta); Z) = \Delta g(\epsilon_0 - X'(\beta - \beta_0)) \\ - \int_a^b 1(s \le \epsilon_0) 1(s \ge \tau_0) \exp\{g(s - X'(\beta - \beta_0))\} ds$$

has bounded and continuous first and second derivatives with respect to $\beta \in \mathcal{B}$ and $\zeta(\cdot,\beta) \in \mathcal{H}^P$.

Proof. This result follows by direct calculation. See all the first and second derivatives given above. By Conditions (A.1)–(A.3) and (A.6), these are continuous and bounded. □

Lemma C.2. *For $g_0 \in \mathcal{G}^p$, there exists a function $g_{0,n} \in \mathcal{G}_n^p$ such that*

$$\|g_{0,n} - g_0\|_\infty = O\left(n^{-p\nu}\right)$$

Proof. This is a direct result of Corollary 6.21 in Schumaker [2007], that is, there exists a $g_{0,n} \in \mathcal{G}_n^p$ such that $\zeta_{0,n}(s, x, \beta_0) = g_{0,n}(s)$ and

$$\|\zeta_{0,n}(\cdot,\beta_0) - \zeta_0(\cdot,\beta_0)\|_\infty = \|g_{0,n} - g_0\|_\infty = O(q_n^{-p}) = O(n^{-p\nu}).$$

□

Lemma C.3. *Let $\theta_{0,n} = (\beta_0, \zeta_{0,n}(\cdot,\beta_0))$ with $\zeta_{0,n}(\cdot,\beta_0) \equiv g_{0,n}$ defined in Lemma C.2. Denote $\mathcal{F}_n = \{l(\theta; z) - l(\theta_{0,n}; z) \colon \theta \in \Theta_n^P\}$. Assume that conditions (A.1)–(A.3) and (A.6) hold, then the ε-bracketing number associated with $\|\cdot\|_\infty$ norm for $\mathcal{F}_n$ is bounded by $(1/\varepsilon)^{cq_n+d}$, that is, $N_{[]}(\varepsilon, \mathcal{F}_n, \|\cdot,\|_\infty) \lesssim (1/\varepsilon)^{cq_n+d}$ for some constant $c > 0$.*

Proof. By the calculation in Shen and Wong [1994] on page 597, denote the ceiling of x by $\lceil x \rceil$, then for any $\varepsilon > 0$, there exists a set of brackets $\{[g_i^L, g_i^U] \colon i = 1, 2, \ldots, \lceil (1/\varepsilon)^{c_1 q_n} \rceil\}$ such that for any $g \in \mathcal{G}_n^P, g_i^L(s) \leq g(s) \leq g_i^U(s)$ for some $1 \leq i \leq \lceil (1/\varepsilon)^{c_1 q_n} \rceil$ and all $s \in [a, b]$, where $\|g_i^U - g_i^L\|_\infty \leq \varepsilon$. Since $\mathcal{B} \subseteq \mathbb{R}^d$ is compact, $\mathcal{B}$ can be covered by $\lceil c_2 (1/\varepsilon)^d \rceil$ balls with radius ε; that is, for any $\beta \in \mathcal{B}$, there exists β_r, $1 \leq r \leq \lceil c_2 (1/\varepsilon)^d \rceil$, such that $|\beta - \beta_r| \leq \varepsilon$, i.e., $|(\beta - \beta_0) - (\beta_r - \beta_0)| \leq \varepsilon$, and hence $|x'(\beta - \beta_0) - x'(\beta_r - \beta_0)| \leq C\varepsilon$ for any $x \in \mathcal{X}$ because of Condition (A.2), where $C > 0$ is a constant. This indicates that $s - x'(\beta - \beta_0) \in [s - x'(\beta_r - \beta_0) - C\varepsilon, s - x'(\beta_r - \beta_0) + C\varepsilon]$ for any x and s. Assume $g_i^L(s - x'(\beta_r - \beta_0) + c_1^{i,s}\varepsilon)$ and $g_i^U(s - x'(\beta_r - \beta_0) + c_2^{i,s}\varepsilon)$ are the minimum and maximum values of g_i^L and g_i^U within the interval $[s - x'(\beta_r - \beta_0) - C\varepsilon, s - x'(\beta_2 - \beta_0) + C\varepsilon]$, where $c_1^{i,s}$ and $c_2^{i,s}$ are two constants that only depend on g_i^L, g_i^U and s with $|c_1^{i,s}|, |c_2^{i,s}| \leq C$. So we have

$$\begin{aligned} g_i^L(s - x'(\beta_r - \beta_0) + c_1^{i,s}\varepsilon) &\leq g_i^L(s - x'(\beta - \beta_0)) \\ &\leq g(s - x'(\beta - \beta_0)) \\ &\leq g_i^U(s - x'(\beta - \beta_0)) \leq g_i^U(s - x'(\beta_r - \beta_0) + c_2^{i,s}\varepsilon). \end{aligned}$$

Hence we can construct a set of brackets

$$\left\{[m_{i,r}^L(Z), m^U U_{i,r}(Z)] \colon i = 1, \ldots, \lceil (1/\varepsilon)^{c_1 q_n} \rceil; r = 1, \ldots, \lceil c_2 (1/\varepsilon)^d \rceil\right\}$$

such that for any $m(\theta; Z) \in \mathcal{F}_n$, there exists a pair (i, r) such that for any sample point Z, $m(\theta; Z) \in [m_{i,r}^L(Z), m_{i,r}^U(Z)]$, where

$$\begin{aligned} m_{i,r}^L = \big\{ &\Delta g_i^L(\epsilon_0 - X'(\beta_r - \beta_0) + c_1^{i,\epsilon_0}\varepsilon) \\ &- \int_a^b 1(s \leq \epsilon_0) 1(s \geq \tau_0) \exp\{g_i^U(s - x'(\beta_r - \beta_0) + c_2^{i,s}\varepsilon)\} ds \Big\} \\ &- l(\theta_{0,n}; Z), \end{aligned}$$

and

$$
\begin{aligned}
m_{i,r}^U = \big\{ & \Delta g_i^U(\epsilon_0 - X'(\beta_r - \beta_0) + c_2^{i,\epsilon_0}\varepsilon) \\
& - \int_a^b 1(s \le \epsilon_0) 1(s \ge \tau_0) \exp\{g_i^L(s - x'(\beta_r - \beta_0) + c_1^{i,s}\varepsilon)\} ds \Big\} \\
& - l(\theta_{0,n}; Z),
\end{aligned}
$$

It then follows that

$$
\begin{aligned}
& |m_{i,r}^U(Z) - m_{i,r}^L(Z)| \\
& \quad \le |g_i^U(\epsilon_0 - X'(\beta_r - \beta_0) + c_2^{i,\epsilon_0}\varepsilon) - g_i^L(\epsilon_0 - X'(\beta_r - \beta_0) + c_1^{i,\epsilon_0}\varepsilon)| \\
& \qquad + \int_a^b |\exp\{g_i^U(s - x'(\beta_2 - \beta_0) + c_2^{i,s}\varepsilon)\} - \exp\{g_i^L(s - x'(\beta_r - \beta_0) + c_1^{i,s}\varepsilon)\}| ds \\
& \quad = A_1 + A_2
\end{aligned}
$$

For A_1, by subtracting and adding the terms of $g(\epsilon_0 - X'(\beta_r - \beta_0) + c_2^{i,\epsilon_0}\varepsilon)$ and $g(\epsilon_0 - X'(\beta_r - \beta_0) + c_1^{i,\epsilon_0}\varepsilon)$ and applying the Taylor expansion to g at $\epsilon_0 - X'(\beta_r - \beta_0) + c_1^{i,\epsilon_0}\varepsilon$, we have

$$
\begin{aligned}
A_1 \le & |g_i^U(\epsilon_0 - X'(\beta_r - \beta_0) + c_2^{i,\epsilon_o}\varepsilon) - g(\epsilon_0 - X'(\beta_r - \beta_0) + c_2^{i,\epsilon_0}\varepsilon)| \\
& + |g(\epsilon_0 - X'(\beta_r - \beta_0) + c_2^{i,\epsilon_0}\varepsilon) - g(\epsilon_0 - X'(\beta_r - \beta_0) + c_1^{i,\epsilon_0}\varepsilon)| \\
& + |g(\epsilon_0 - X'(\beta_r - \beta_0) + c_1^{i,\epsilon_0}\varepsilon) - g_i^L(\epsilon_0 - X'(\beta_r - \beta_0) + c_1^{i,\epsilon_0}\varepsilon)| \\
\le & \|g_i^U - g\|_\infty + |\dot{g}(\epsilon_0 - X'(\beta_r - \beta_0) + \tilde{c}\varepsilon)(c_2^{i,\epsilon_0} - c_1^{i,\epsilon_0})\varepsilon| + \|g - g_i^L\|_\infty \\
\le & \|g_i^U - g_i^L\|_\infty + C_1|(c_2^{i,\epsilon_0} - c_1^{i,\epsilon_0})|\varepsilon + \|g_i^U - g_i^L\|_\infty \\
\le & 2\varepsilon + 2C_1C_2\varepsilon \lesssim \varepsilon
\end{aligned}
$$

where the third inequality holds because $\|g_i^U - g\|_\infty, \|g - g_i^L\|_\infty \le \|g_i^U - g_i^L\|_\infty$ and $\dot{g}$ is bounded by C_1. The constant C_1 may be proportional to c_n that is allowed to grow with n slowly enough, but it does not affect the later calculations on convergence rate (see Shen and

Wong [1994] page 591 for their constant l_n), thus we drop c_n for simplicity. For A_2, by using the similar arguments as for A_1 and denoting $s - X'(\beta_r - \beta_0) = s_r$ for notational simplicity, we have

$$\begin{aligned}
A_2 \leq & \int_a^b \{|\exp\{g_i^U(s_r + c_2^{i,s}\varepsilon)\} - \exp\{g(s_s + c_2^{i,s}\varepsilon)\}| \\
& + |\exp\{g(s_r + c_2^{i,s}\varepsilon)\} - \exp\{g(s_r + c_1^{i,s}\varepsilon)\}| \\
& + |\exp\{g(s_r + c_1^{i,s}\varepsilon)\} - \exp\{g_i^L(s_r + c_1^{i,s}\varepsilon)\}|\}ds \\
= & \int_a^b \{|\exp\{\tilde{g}_i^U(s_r + c_2^{i,s}\varepsilon)\}(g_i^U - g)(s_r + c_2^{i,s}\varepsilon)| \\
& + |\exp\{g(s_r + \tilde{c}\varepsilon)\}(c_2^{i,s} - c_1^{i,s})\varepsilon| \\
& + |\exp\{\tilde{g}_i^L(s_r + c_1^{i,s}\varepsilon)\}(g_i^L - g)(s_r + c_1^{i,s}\varepsilon)|\}ds \\
\lesssim & \|g_i^U - g\|_\infty + |(c_2^{i,s} - c_1^{i,s})\varepsilon| + \|g - g_i^L\|_\infty \lesssim \varepsilon.
\end{aligned}$$

The above equality is from the Taylor expansion, where $\tilde{g}_i^U = g + \xi(g_i^U - g)$ for some $0 < \xi < 1$ and thus $|\tilde{g}_i^U(\cdot)| \leq |g(\cdot)| + \varepsilon$, which is bounded in $[a, b]$; similarly $|\tilde{g}_i^L|$ is also bounded in $[a, b]$. Hence $\|m_i^U - m_i^L\|_\infty \lesssim \varepsilon$ and the ε-bracketing number associated with $\|\cdot\|_\infty$ norm for the class $\mathcal{F}_n$ follows

$$N_{[]}(\varepsilon, \mathcal{F}_n, \|\cdot\|_\infty) \leq (1/\varepsilon)^{c_1 q_n} c_2 (1/\varepsilon)^d \lesssim (1/\varepsilon)^{c_1 q_n + d}.$$

□

Lemma C.4. *Let $h_j^*(s, x, \beta) = w_j^*(\psi(s, X, \beta))$, where $h_j^*(s, x, \beta_0) = w_j^*(s) = -\dot{g}_0(s)P(X_j \mid \epsilon_0 \geq s, \tau_0 \leq s), j = 1, \ldots, d$. Assume conditions (A.1)–(A.6) hold, then there exists $h_{j,n}^*(s, x, \beta) = w_{j,n}^*(\psi(s, x, \beta)) \in \mathcal{H}_n^2$ such that $\|h_{j,n}^* - h_j^*\|_\infty = O(n^{-2\nu})$, or equivalently, $\|w_{j,n}^* - w_j^*\|_\infty = O(n^{-2\nu})$ where $w_{j,n}^* \in \mathcal{G}_n^2$*

Proof. In the proof of Theorem 2.2 we show that such defined $(h_1^*, \ldots, h_d^*)$ determines the least favorable submodel for β. Now, by conditions (A.4)–(A.5), the following conditional

density of ϵ_0 given T and X

$$f_{\epsilon_0|T,X}(s|t,x) = f_{Y^*|X}(s+x'\beta_0|x)\bar{F}_{C|T,X}(s+x'\beta_0 \mid t,x)+f_{C|T,X}(s+x'\beta_0 \mid t,x)\bar{F}_{Y^*|X}(s+x'\beta_0 \mid x)$$

is uniformly bounded for all $t \in \mathcal{T}$ and $x \in \mathcal{X}$, and its derivative with respect to s

$$\begin{aligned}\dot{f}_{\epsilon_0|T,X}(s \mid t,x) = \dot{f}_{Y^*|X}(s+x'\beta_0 \mid x)\bar{F}_{C|T,X}(s+x'\beta_0 \mid t,x) \\ - f_{Y^*|X}(s+x'\beta_0 \mid x)f_{C|T,X}(s+x'\beta_0 \mid t,x) \\ + \dot{f}_{C|T,X}(s+x'\beta_0 \mid t,x)\bar{F}_{Y^*|X}(s+x'\beta_0 \mid x) \\ - f_{C|T,X}(s+x'\beta_0 \mid t,x)f_{Y^*|X}(s+x'\beta_0 \mid x)\end{aligned}$$

is also uniformly bounded. Hence the density of ϵ_0

$$f_{\epsilon_0} = \int_{\mathcal{X}}\int_{\mathcal{T}} f_{\epsilon_0|T,X}(s \mid t,x)f_{T|X}(t \mid x)f_X(x)dtdx$$

and its derivative

$$\dot{f}_{\epsilon_0}(s) = \int_{\mathcal{X}}\int_{\mathcal{T}} \dot{f}_{\epsilon_0|T,X}(s \mid t,x)f_{T|X}(t \mid x)f_X(x)dtdx$$

are bounded. Thus the first and second derivatives of $P(\epsilon_0 \geq s)$, i.e., $-f_{\epsilon_0}(s)$ and $-\dot{f}_{\epsilon_0}(s)$ are both bounded. In addition, under Condition (A.2), the first and second derivatives of $P[X1(s \leq \epsilon_0)1(s \geq \tau_0)]$ with respect to s

$$\frac{dP[X1(s \leq \epsilon_0)1(s \geq \tau_0)]}{ds} = -\int_{\mathcal{X}}\int_{\mathcal{T}} xf_X(x)f_{T|X}(t \mid x)f_{\epsilon_0|T,X}(s \mid t,x)dtdx$$

and

$$\frac{d^2P[X1(s \leq \epsilon_0)1(s \geq \tau_0)]}{ds^2} = -\int_{\mathcal{X}}\int_{\mathcal{T}} sf_X(x)f_{T|X}(t \mid x)\dot{f}_{\epsilon_0|T,X}(s \mid t,x)dtdx$$

are also bounded. Therefore, $P[X \mid \epsilon_0 \geq s, \tau_0 \leq s] = P[X1(\epsilon_0 \geq s)1(\tau_0 \leq s)]/P[1(\epsilon_0 \geq s)1(\tau_0 \leq s)]$ has a bounded second derivative with respect for s for $s \leq b$, where b is the truncation time defined in Condition (A.3). Thus $P[X \mid \epsilon_0 \geq s, \tau_0 \leq s] \in \mathcal{G}^2$. Moreover, since $g_0 \in \mathcal{G}^P$ for $p \geq 3$, we have $\dot{g}_0 \in \mathcal{G}^{p-1}$ with $p - 1 \geq 2$. Thus according to Corollary 6.21 of Schumaker [2007], there exists an $h^*_{j,n} \in \mathcal{H}_n^{\min(p-1,2)} = \mathcal{H}_n^2$ such that $h^*_{j,n}(s, x, \beta_0) = w^*_{j,n}(\psi(s, x, \beta_0)) = w^*_{j,n}(s)$ and $\|h^*_{j,n}(\cdot, \beta_0) - h^*_j(\cdot, \beta_0)\|_\infty = \|w^*_{j,n} - w^*_j\|_\infty = O(q_n^{-2}) = O(n^{-2\nu})$. □

Lemma C.5. *For h^*_j defined in Lemma C.4, denote the class of functions*

$$\mathcal{F}^j_n(\eta) = \{\dot{l}_\zeta(\theta; z)[h^*_j - h_j] \colon \theta \in \Theta^P_n, h_j \in \mathcal{H}^2_n, d(\theta, \theta_0) \leq \eta, \|h_j - h^*_j\|_\infty \leq \eta\}.$$

Assume conditions (A.1)–(A.6) hold, then $N_{[]}(\varepsilon, \mathcal{F}^j_n(\eta), \|\cdot\|_\infty) \lesssim (\eta/\varepsilon)^{cq_n+d}$ for some constant $c > 0$.

Proof. We shall use the similar argument for the bracketing number of the class $\mathcal{F}_n$ in Lemma C.3. First, define the classes of functions

$$\begin{aligned}
\mathcal{H}^p_n(\eta) &= \{g \in \mathcal{H}^p_n, \|g - g_0\|_2 \leq \eta\}, \\
\mathcal{H}^2_{n,j}(\eta) &= \{h \in \mathcal{H}^2_n, \|h - h^*_j\|_\infty \leq \eta\}, \quad \text{and} \\
\mathcal{B}(\eta) &= \{\beta \in \mathcal{B} \subseteq \mathbb{R}^d, |\beta - \beta_0| \leq \eta\},
\end{aligned}$$

then it follows by the calculation of Shen and Wong [1994] in page 597 that

$$N_{[]}(\varepsilon, \mathcal{H}^p_n(\eta), \|\cdot\|_\infty) \leq (\eta/\varepsilon)^{c_1 q_n} \quad \text{and} \quad N_{[]}(\varepsilon, \mathcal{H}^2_{n,j}(\eta, \|\cdot\|_\infty) \leq (\eta/\varepsilon)^{c_2 q_n}$$

for some constants $c_1, c_2 > 0$. In addition, since $\mathcal{B} \subseteq \mathbb{R}^d$ is compact, the covering number for $\mathcal{B}(\eta)$ follows $N(\varepsilon, \mathcal{B}(\eta), \|\cdot\|_\infty) \leq c_3(\eta/\varepsilon)^d$.

Then as in the proof of Lemma C.3, let $\epsilon_r = Y - X'\beta_r$, $\epsilon_0 = Y - X'\beta_0$, and $s_r = s - X'\beta_r$ for notational simplicity. g_i^L and g_i^U are functions that bracket g with $\|g_i^U - g_i^L\|_\infty \leq \varepsilon$; h_k^L and h_k^U are functions that bracket h with $\|h_k^U - h_k^L\| \leq \varepsilon$; and β_r satisfies $s - x'\beta \in [s - x'\beta_2 - C\varepsilon, s - x'\beta_r + C\varepsilon] = [s_r - C\varepsilon, s_r + C\varepsilon]$ for any x, s and some constant $C > 0$. Moreover, let $c_1^{j,k,s}$ and $c_2^{j,k,s}$ be two constants such that $(h_j^* - h_k^U)(s_r + c_i^{j,k,s}\varepsilon)$ and $(h_j^* - h_k^L)(s_r + c_2^{j,k,s}\varepsilon)$ are the minimum and maximum values of $(h_j^* - h_k^U)$ and $(h_j^* - h_k^L)$ in $[s_r - C\varepsilon, s_r + C\varepsilon]$, respectively; and let $c_3^{i,s}$ and $c_4^{i,s}$ be two constants such that $g_i^L(s_r + c_3^{i,s}\varepsilon)$ and $g_i^U(s_r + c_4^{i,s}\varepsilon)$ are the minimum and maximum values of g_i^L and g_i^U in $[s_r - C\varepsilon, s_r + C\varepsilon]$, respectively. Then we can construct a set of brackets for $\mathcal{F}_n^j(\eta)$ as

$$\left\{[d_{i,k,r}^L(Z), d_{i,k,r}^U(Z)] \colon 1 < i < \lceil(\eta/\varepsilon)^{c_1 q_n}\rceil; 1 \leq k \leq \lceil(\eta/\varepsilon)^{c_2 q_n}\rceil; 1 \leq s \leq \lceil c_3(\eta/\varepsilon)^d\rceil\right\}$$

that for any $\dot{l}_g(\theta; Z)[h_j^* - h] \in \mathcal{F}_n^j(\eta)$, there exists a triplet (i, k, r) such that

$$\dot{l}_g(\theta; Z)[h_j^* - h] \in [d_{i,k,r}^L(Z), d_{i,k,r}^U(Z)]$$

for any sample point Z, where

$$\begin{aligned} d_{i,k,s}^L(Z) &= \Delta(h_j^* - h_k^U)(\epsilon_r + c_1^{j,k,Y}\varepsilon) \\ &\quad - \int_a^b 1(\epsilon_0 \geq s)1(\tau_0 \leq s)e^{g_i^U(s_r + c_4^{i,s}\varepsilon)}(h_j^* - h_k^L)(s_r + c_2^{j,k,s}\varepsilon)1\{(h_j^* - h)(s_r + c_2^{j,k,s}\varepsilon) \geq 0\}ds \\ &\quad - \int_a^b 1(\epsilon_0 \geq s)1(\tau_0 \leq s)e^{g_i^L(s_r + c_3^{i,s}\varepsilon)}(h_j^* - h_k^L)(s_r + c_2^{j,k,s}\varepsilon)1\{(h_j^* - h)(s_r + c_2^{j,k,s}\varepsilon) < 0\}ds \end{aligned}$$

and

$$
\begin{aligned}
d^U_{i,k,s}(Z) &= \Delta(h^*_j - h^L_k)(\epsilon_r + c_2^{j,k,Y}\varepsilon) \\
&\quad - \int_a^b 1(\epsilon_0 \geq s)1(\tau_0 \leq s)e^{g_i^L(s_r + c_3^{i,s}\varepsilon)}(h^*_j - h^U_k)(s_r + c_1^{j,k,s}\varepsilon)1\{(h^*_j - h)(s_r + c_1^{j,k,s}\varepsilon) \geq 0\}ds \\
&\quad - \int_a^b 1(\epsilon_0 \geq s)1(\tau_0 \leq s)e^{g_i^U(s_r + c_4^{i,s}\varepsilon)}(h^*_j - h^U_k)(s_r + c_1^{j,k,s}\varepsilon)1\{(h^*_j - h)(s_r + c_1^{j,k,s}\varepsilon) < 0\}ds
\end{aligned}
$$

Then it follows that

$$
\begin{aligned}
|d^U_{i,k,r}(Z) - d^L_{i,k,r}(Z)| &\leq |(h^*_j - h^L_k)(\epsilon_r + c_2^{j,k,Y}\varepsilon) - (h^*_j - h^U_k)(\epsilon_r + c_1^{j,k,Y}\varepsilon)| \\
&\quad + \int_a^b |e^{g_i^U(s_r + c_4^{i,s}\varepsilon)}(h^*_j - h^L_k)(s_r + c_2^{j,k,s}\varepsilon)1\{(h^*_j - h)(s_r + c_2^{j,k,s}\varepsilon) \geq 0\} \\
&\qquad - e^{g_i^L(s_r + c_3^{i,s}\varepsilon)}(h^*_j - h^U_k)(s_r + c_1^{j,k,s}\varepsilon)1\{(h^*_j - h)(s_r + c_1^{j,k,s}\varepsilon) \geq 0\}|ds \\
&\quad + \int_a^b |e^{g_i^L(s_r + c_3^{i,s}\varepsilon)}(h^*_j - h^L_k)(s_r + c_2^{j,k,s}\varepsilon)1\{(h^*_j - h)(s_r + c_2^{j,k,s}\varepsilon) < 0\} \\
&\qquad - e^{g_i^U(s_r + c_4^{i,s}\varepsilon)}(h^*_j - h^U_k)(s_r + c_1^{j,k,s}\varepsilon)1\{(h^*_j - h)(s_r + c_1^{j,k,s}\varepsilon) < 0\}|ds \\
&= A_1 + A_2 + A_3.
\end{aligned}
$$

For term A_1, by subtracting and adding the terms $h^L_k(\epsilon_r + c_1^{j,k,s}\varepsilon)$, $h(\epsilon_r + c_1^{j,k,s}\varepsilon)$ and $h(\epsilon + c_2^{j,k,s}\varepsilon)$, and applying the Taylor expansions for h^*_j and h, we have

$$
\begin{aligned}
A_1 &\leq |h^*_j(\epsilon_r + c_2^{j,k,Y}\varepsilon) - h^*_j(\epsilon_r + c_1^{j,k,Y}\varepsilon)| + |h^U_k(\epsilon_r + c_1^{j,k,Y}\varepsilon) - h^L_k(\epsilon_r + c_1^{j,k,Y}\varepsilon)| \\
&\quad + |h^L_k(\epsilon_r + c_1^{j,k,Y}\varepsilon) - h(\epsilon_r + c_1^{j,k,Y}\varepsilon)| + |h(\epsilon_r + c_1^{j,k,Y}\varepsilon) - h(\epsilon_r + c_2^{j,k,s}\varepsilon)| \\
&\quad + |h(\epsilon_r + c_2^{j,k,Y}\varepsilon) - h^L_k(\epsilon_r + c_2^{j,k,Y}\varepsilon)| \\
&\lesssim |c_2^{j,k,Y} - c_1^{j,k,Y}|\varepsilon + \|h^U_k - h^L_k\|_\infty + \|h^L_k - h\|_\infty + |c_2^{j,k,Y} - c_1^{j,k,Y}|\varepsilon + \|h - h^L_k\|_\infty \\
&\lesssim |c_2^{j,k,Y} - c_1^{j,k,Y}|\varepsilon + \|h^U_k - h^L_k\|_\infty \lesssim \varepsilon
\end{aligned}
$$

Next for A_2, first, by subtracting and adding the term $e^{g_i^U(s_r + c_4^{k,s}\varepsilon)}(h^*_j - h^U_k)(s_r + c_1^{j,k,s}\varepsilon)$ we

have

$$
\begin{aligned}
A_2' &= \int_a^b |e^{g_i^U(s_r+c_4^{i,s}\varepsilon)}(h_j^*-h_k^L)(s_r+c_2^{j,k,s}\varepsilon) - e^{g_i^L(s_r+c_3^{i,s}\varepsilon)}(h_j^*-h_k^U(s_r+c_1^{j,k,s}\varepsilon)|ds \\
&\leq \int_a^b e^{g_i^U(s_r+c_4^{i,s}\varepsilon)} \left|(h_j^*-h_k^L)(s_r+c_s^{j,k,s}\varepsilon) - (h_j^*-h_k^U)(s_r+c_1^{j,k,s}\varepsilon)\right| ds \\
&\qquad \int_a^b \left|\left\{e^{g_i^U(s_r+c_4^{i,s}\varepsilon)} - e^{g_i^L(s_r+c_3^{i,s}\varepsilon)}\right\}(h_j^*-h_k^U)(s_r+c_1^{j,k,s}\varepsilon)\right| ds \\
&= A_4 + A_5
\end{aligned}
$$

By the above result for A_1, since $\|g_i^U\|_\infty \leq \|g\|_\infty + \|g_i^U - g\|_\infty \leq \|g\|_\infty + \varepsilon$, which is bounded, we have

$$
A_4 \leq \|A_1\|_\infty \int_a^b e^{g_i^U(s_r+c_4^{i,s}\varepsilon)} ds \lesssim \varepsilon.
$$

Then for A_5, using the similar argument for A_1, it follows that

$$
\begin{aligned}
A_5 &\leq \|h_j^* - h_k^U\|_\infty \int_a^b \left\{\left|e^{g_i^U(s_r+c_4^{i,s}\varepsilon)} - e^{g(s_r+c_4^{i,s}\varepsilon)}\right| + \left|e^{g(s_r+c_4^{i,s}\varepsilon)} - e^{g(s_r+c_3^{i,s}\varepsilon)}\right| \right. \\
&\qquad\qquad \left. + \left|e^{g(s_r+c_3^{i,s}\varepsilon)} - e^{g_i^L(s_r+c_3^{i,s}\varepsilon)}\right|\right\} ds \\
&\leq (\|h_j^* - h\|_\infty + \|h - h_k^U\|_\infty) \left\{\int_a^b \left[e^{\tilde{g}_i^U(s_r+c_4^{i,s}\varepsilon)} \left|(g_i^U - g)(s_r+c_4^{i,s}\varepsilon)\right| \right.\right. \\
&\qquad \left.\left. + e^{g(s_r+\tilde{c}\varepsilon)} \left|\left(c_4^{i,s} - c_3^{i,s}\right)\varepsilon\right| + e^{\tilde{g}_i^L(s_r+c_3^{i,s}\varepsilon)} \left|(g - g_i^L)\left(s_r + c_3^{i,s}\varepsilon\right)\right|\right] ds\right\} \\
&\lesssim (\eta+\varepsilon)\left\{\|g_i^U - g\|_\infty + |c_4^{i,s} - c_3^{i,s}|\varepsilon + \|g - g_i^L\|_\infty\right\} \\
&\lesssim (\eta+\varepsilon)\varepsilon \lesssim \varepsilon,
\end{aligned}
$$

where $\tilde{g}_i^U = g + \xi_1(g_i^U - g)$ and $\tilde{g}_i^L = g + \xi_2(g_i^L - g)$ for some constants $\xi_1, \xi_2 \in (0,1)$, and

$\tilde{c} = c_3^{i,s} + \xi(c_4^{i,s} - c_3^{i,s})$ for some $\xi]in(0,1)$. Thus $A_2' \lesssim \varepsilon$. Therefore,

$$
\begin{aligned}
A_2 \leq \int_a^b & \left| e^{g_i^U(s_r + c_4^{i,s}\varepsilon)}(h_j^* - h_k^L)(s_r + c_2^{i,k,s}\varepsilon) - e^{g_i^L(s_r + c_3^{i,s}\varepsilon)}(h_j^* - h_k^U)(s_r + c_1^{j,k,s}\varepsilon) \right| \\
& \qquad \cdot 1\left\{(h_j^* - h)(s_r + c_2^{j,k,s}\varepsilon) \geq 0\right\} ds \\
& + \int_a^b e^{g_i^L(s_r + c_3^{i,s}\varepsilon)} \left| (h_j^* - h_k^U)(s_r + c_1^{j,k,s}\varepsilon) \right| \\
& \qquad \cdot \left| 1\left\{(h_j^* - h)(s_r + c_2^{j,k,s}\varepsilon) \geq 0\right\} - 1\left\{(h_j^* - h)(s_r + c_1^{j,k,s}\varepsilon) \geq 0\right\} \right| ds \\
\lesssim A_2' + & \int_a^b \left| 1\left\{(h_j^* - h)(s_r + c_2^{j,k,s}\varepsilon) \geq 0\right\} - 1\left\{(h_j^* - h)(s_r + c_1^{j,k,s}\varepsilon) \geq 0\right\} \right| ds \\
\lesssim \varepsilon + & \left| \left(c_2^{j,k,s} - c_1^{j,k,s}\right)\varepsilon \right| \lesssim \varepsilon,
\end{aligned}
$$

where the second inequality holds since $\|g_i^L\|_\infty \leq \|g\|_\infty + \|g_i^L - g\|_\infty \leq \|g\|_\infty + \varepsilon$ and thus is bounded, then

$$
\|e^{g_i^L(s_r + c_3^{i,s}\varepsilon)}(h_j^* - h_k^U)(s_r + c_1^{j,k,s}\varepsilon)\|_\infty \leq \|e^{g_i^L}\|_\infty(\|h_j^* - h\|_\infty + \|h - h_k^U\|_\infty) \lesssim (\eta + \varepsilon),
$$

which is bounded. Finally, by using the same argument as for A_2, we can also show that $A_3 \lesssim \varepsilon$. Hence $\|d_{i,k,r}^U(Z) - d_{i,k,r}^L(Z)\|_\infty \lesssim \varepsilon$. Therefore, the ε-bracketing number for the class $\mathcal{F}_n^j(\eta)$ is bounded by $(\eta/\varepsilon)^{c_1 q_n}(\eta/\varepsilon)^{c_2 q_n} c_3(\eta/\varepsilon)^d$, that is,

$$
N_{[]}(\varepsilon, \mathcal{F}_n^j(\eta), \|\cdot\|_\infty) \lesssim (\eta/\varepsilon)^{(c_1+c_2)q_n + d}.
$$

□

Lemma C.6. *For $j = 1, \ldots, d$, define the following two classes of functions:*

$$
\mathcal{F}_{n,j}^\beta(\eta) = \{\dot{l}_{\beta_j}(\theta; z) - \dot{l}_{\beta_j}(\theta_0; z) \colon \theta \in \Theta_n^P, d(\theta, \theta_0) \leq \eta, \|\dot{g}(\psi(\cdot, \beta)) - \dot{g}_0(\psi(\cdot, \beta_0))\|_2 \leq \eta\}
$$

and

$$\mathcal{F}_{n,j}^{\zeta}(\eta) = \{\dot{l}_{\zeta}(\theta; z)[h_j^*(\cdot, \beta)] - \dot{l}_{\zeta}(\theta_0; z)[h_j^*(\cdot, \beta_0)] \colon \theta \in \Theta_n^P, d(\theta, \theta_0) \leq \eta\}$$

where $\dot{l}_{\beta_j}(\theta; Z)$ *is the* j*th element of* $\dot{l}_{\beta}(\theta; Z)$*,* $\dot{g}(\cdot)$ *denotes the derivative of* $g(\cdot)$ *and* h_j^* *is defined in Lemma C.5. Assume conditions (A.1)–(A.6) hold, then* $N_{[]}(\varepsilon, \mathcal{F}_{j,n}^{\beta}(\eta), \|\cdot\|_{\infty}) \lesssim (\eta/\varepsilon)^{c_1 q_n + d}$ *and* $N_{[]}(\varepsilon, \mathcal{F}_{j,n}^{\zeta}(\eta), \|\cdot\|_{\infty}) \lesssim (\eta/\varepsilon)^{c_2 q_n + d}$ *for some constants* $c_1, c_2 > 0$*.*

Proof. First, for the following classes of functions

$$\begin{aligned}
\mathcal{H}_n^P(\eta) &= \left\{g \in \mathcal{H}_n^P, \|g - g_0\|_2 \leq \eta\right\}, \\
\mathcal{H}_n^{p-1}(\eta) &= \left\{\dot{g} \in \mathcal{H}_n^{p-1}, \|\dot{g} - \dot{g}_0\|_2 \leq \eta\right\} \quad \text{and} \\
\mathcal{B}(\eta) &= \left\{\beta \in \mathcal{B} \subseteq \mathbb{R}^d, |\beta - \beta_0| \leq \eta\right\},
\end{aligned}$$

by the same argument in the proof of Lemma C.5, we have

$$N_{[]}(\varepsilon, \mathcal{H}_n^p(\eta), \|\cdot\|_{\infty}) \leq (\eta/\varepsilon)^{c_1 q_n}, \quad N_{[]}(\varepsilon, \mathcal{H}_n^{p-1}(\eta), \|\cdot\|_{\infty}) \leq (\eta/\varepsilon)^{c_2 q_n},$$

and $N_{[]}(\varepsilon, \mathcal{B}(\eta), \|\cdot, \|_{\infty}) \leq c_3(\eta/\varepsilon)^d$ for some constants $c_1, c_2, c_3 > 0$.

Then using a similar argument as in the proof of Lemma C.5, let g_i^L and g_i^U be functions that bracket g with $\|g_i^U - g_i^L\|_{\infty} \leq \varepsilon$; $\dot{g}_k^L$ and $\dot{g}_k^U$ be functions that bracket $\dot{g}$ with $\|\dot{g}_k^U - \dot{g}_k^L\|_{\infty} \leq \varepsilon$; and let β_r satisfy $s - x'\beta \in [s - x'\beta_r - C\varepsilon, x - x'\beta_r + C\varepsilon] = [s_r - C\varepsilon, s_r + C\varepsilon]$ for any x, s and some constant $C > 0$. Moreover, let $c_1^{k,s}$ and $c_2^{k,s}$ be two constants such that $\dot{g}_k^L(s_r + c_1^{k,s}\varepsilon)$ and $\dot{g}_k^U(s_r + c_2^{k,s}\varepsilon)$ are the minimum and maximum values of $\dot{g}_k^L$ and $\dot{g}_k^U$ in $[s_r - C\varepsilon, s_r + C\varepsilon]$, respectively; let $c_3^{i,s}$ and $c_4^{i,s}$ be two constants such that $g_i^L(s_r + c_3^{i,s}\varepsilon)$ and $g_i^U(s_r + c_4^{i,s}\varepsilon)$ are the minimum and maximum values of g_i^L and g_i^U in $[s_r - C\varepsilon, s_r + C\varepsilon]$, respectively; and let $c_5^{j,s}$ and $c_6^{j,s}$ be two constants such that $h_j^*(s_r + c_5^{j,s}\varepsilon)$ and $h_j^*(s_r + c_6^{j,s}\varepsilon)$ are the minimum and maximum values of h_j^* in $[s_r - C\varepsilon, s_r + C\varepsilon]$. Then we can similarly construct a set of

brackets for $\mathcal{F}^{\beta}_{n,j}(\eta)$ as

$$\left\{\left[u^L_{i,k,r}(Z), u^U_{i,k,r}(Z)\right] : 1 \le i \le \lceil(\eta/\varepsilon)^{c_1 q_n}\rceil; 1 \le k \le \lceil(\eta/\varepsilon)^{c_2 q_n}\rceil; 1 \le r \le \lceil c_3(\eta/\varepsilon)^d\rceil\right\}$$

that for any element in $\mathcal{F}^{\beta}_{n,j}(\eta)$, there exists a triplet (i,k,r) such that

$$\dot{l}_{\beta_j}(\theta; Z) - \dot{l}_{\beta_j}(\theta_0; Z) \in \left[u^L_{i,k,r}(Z), u^U_{i,k,r}(Z)\right]$$

for any sample point Z (without loss of generality, assume X_j is nonnegative; for negative X_j, just switch the lower and upper brackets), where

$$\begin{aligned} u^L_{i,k,r}(Z) = -X_j \Big\{ & \Delta \dot{g}^U_k(\epsilon_r + c_2^{k,Y}\varepsilon) \\ & - \int_a^b 1(\epsilon_0 \ge s) 1(\tau_0 \le s) e^{g^L_i(s_r + c_3^{i,s}\varepsilon)} \dot{g}^L_k(s_r + c_1^{k,s}\varepsilon) 1\{\dot{g}(s_r + c_1^{k,s}\varepsilon) \ge 0\} ds \\ & - \int_a^b 1(\epsilon_0 \ge s) 1(\tau_0 \le s) e^{g^U_i(s_r + c_4^{i,s}\varepsilon)} \dot{g}^L_k(s_r + c_1^{k,s}\varepsilon) 1\{\dot{(}s_r + c_1^{k,s}\varepsilon) < 0\} ds \Big\} - \dot{l}_{\beta_j}(\beta_0, g_0; Z) \end{aligned}$$

and

$$\begin{aligned} u^U_{i,k,r}(Z) = -X_j \Big\{ & \Delta \dot{g}^L_k(\epsilon_r + c_1^{k,Y}\varepsilon) \\ & - \int_a^b 1(\epsilon_0 \ge s) 1(\tau_0 \le s) e^{g^U_i(s_r + c_4^{i,s}\varepsilon)} \dot{g}^U_k(s_r + c_2^{k,s}\varepsilon) 1\{\dot{g}(s_r + c_2^{k,s}\varepsilon) \ge 0\} ds \\ & - \int_a^b 1(\epsilon_0 \ge s) 1(\tau_0 \le s) e^{g^L_i(s_r + c_3^{i,s}\varepsilon)} \dot{g}^U_k(s_r + c_2^{k,s}\varepsilon) 1\{\dot{(}s_r + c_2^{k,s}\varepsilon) < 0\} ds \Big\} - \dot{l}_{\beta_j}(\beta_0, g_0; Z) \end{aligned}$$

with ϵ_r and s_r being defined as in the proof of Lemma C.4. Also we can construct a bracket for $\mathcal{F}^{g}_{n,j}(\eta)$ as

$$\left\{\left[v^L_{i,r}(Z), v^U_{i,r}(Z)\right] : 1 \le i \le \lceil(\eta/\varepsilon)^{c_1 q_n}\rceil; 1 \le r \le \lceil c_3(\eta/\varepsilon)^d\rceil\right\}$$

that for any element in $\mathcal{F}^g_{n,j}(\eta)$, there exists a pair (i,r) such that

$$\dot{l}_g(\theta;Z)[h)j^*] - \dot{l}_g(\theta_0;Z)[h^*_j] \in \left[v^L_{i,r}(Z), v^U_{i,r}(Z)\right]$$

for any sample point Z, where

$$\begin{aligned}
v^L_{i,r} &= \Delta h^*_j(\epsilon_r + c^{j,Y}_5\varepsilon) \\
&\quad - \int_a^b 1(\epsilon_0 \geq s)1(\tau_0 \leq s)e^{g^U_i(s_r + c^{i,s}_4\varepsilon)}h^*_j(s_r + c^{j,s}_6\varepsilon)1\{h^*_j(s_r + c^{j,s}_6\varepsilon) \geq 0\}ds \\
&\quad - \int_a^b 1(\epsilon_0 \geq s)1(\tau_0 \leq s)e^{g^L_i(s_r + c^{i,s}_3\varepsilon)}h^*_j(s_r + c^{j,s}_6\varepsilon)1\{h^*_j(s_r + c^{j,s}_6\varepsilon) < 0\}ds - \dot{l}_g(\theta_0;Z)[h^*_j]
\end{aligned}$$

and

$$\begin{aligned}
v^U_{i,r} &= \Delta h^*_j(\epsilon_r + c^{j,Y}_5\varepsilon) \\
&\quad - \int_a^b 1(\epsilon_0 \geq s)1(\tau_0 \leq s)e^{g^L_i(s_r + c^{i,s}_3\varepsilon)}h^*_j(s_r + c^{j,s}_5\varepsilon)1\{h^*_j(s_r + c^{j,s}_5\varepsilon) \geq 0\}ds \\
&\quad - \int_a^b 1(\epsilon_0 \geq s)1(\tau_0 \leq s)e^{g^U_i(s_r + c^{i,s}_4\varepsilon)}h^*_j(s_r + c^{j,s}_5\varepsilon)1\{h^*_j(s_r + c^{j,s}_5\varepsilon) < 0\}ds - \dot{l}_g(\theta_0;Z)[h^*_j].
\end{aligned}$$

By using a similar argument for the proof of Lemma C.5, we can show that

$$\|u^U_{i,k,r}(Z) - u^L_{i,k,r}(Z)\|_\infty \lesssim \varepsilon \quad \text{and} \quad \|v^U_{i,r}(Z) - v^L_{i,r}(Z)\|_\infty \lesssim \varepsilon.$$

Therefore,

$$N_{[]}(\varepsilon, \mathcal{F}^\beta_{n,j}(\eta), \|\cdot\|_\infty) \lesssim (\eta/\varepsilon^{c_1 q_n}(\eta/\varepsilon)^{c_2 q_n}(\eta/\varepsilon)^d = (\eta/\varepsilon)^{(c_1+c_2)q_n + d}$$

and

$$N_{[]}(\varepsilon, \mathcal{F}^g_{n,j}(\eta), \|\cdot\|_\infty) \lesssim (\eta/\varepsilon)^{c_1 q_n}(\eta/\varepsilon)^d = (\eta/\varepsilon)^{c_1 q_n + d}.$$

□

C.3 Proof of Theorem 2.1

Theorem C.1. *Let $K_n = O(n^\nu)$ where ν satisfies the restriction $\frac{1}{2(1+p)} < \nu < \frac{1}{2p}$ with p being the smoothness parameter defined in condition (A.6). Suppose conditions (A.1) - (A.7) hold, and the failure time Y follows the typical AFT model subject to left-truncation and right-censoring. Then*

$$d(\hat{\theta}_n, \theta_0) = O_p\left\{n^{-\min(p\nu,(1-\nu)/2)}\right\},$$

where $d(\cdot,\cdot)$ is defined in (C.2) and $\theta = \begin{pmatrix} \beta^T & \gamma^T \end{pmatrix}^T$.

Proof. We shall apply Theorem 1 of Shen and Wong [1994] to derive the convergence rate. First, we verify their Condition C1. Since $Pl(\beta, g; Z)$ is maximized at (β_0, g_0), its first derivative at (β_0, g_0) is equal to 0. By Lemma C.1 that all the second derivatives of $l(\beta, g; Z)$ are continuous and bounded, the Taylor expansion yields

$$\begin{aligned} Pl(\beta, g; Z) - & Pl(\beta_0, g_0; Z) \\ &= \frac{1}{2} P \Big\{ (\beta-\beta_0)' \ddot{l}_{\beta\beta}(\beta_0, g_0; Z)(\beta-\beta_0) + 2(\beta-\beta_0)' \ddot{l}_{\beta g}(\beta_0.g_0; Z)[g-g_0] \\ &\qquad + \ddot{l}_{gg}(\beta_0, g_0; Z)[g-g_0, g-g_0] \Big\} + o(d^2(\theta, \theta_0)) \\ &= A + o(d^2(\theta, \theta_0)), \end{aligned} \tag{C.3}$$

where $\theta = (\beta, g) \in \Theta_n^p$. By the model assumption we have $E\{\dot{l}_g(\beta_0, g_0; Z)[h]|X\} = 0$ for all $h \in \{h = \frac{\partial g_\eta}{\partial \eta}|_{\eta=0}, g_\eta \in \mathcal{H}^p\}$. The result holds in conditional expectation because the log-likelihood $l(\beta, g; Z)$ is from the conditional density of (Y, Δ, X, T) with conditioning on

X. Taking h to be $\ddot{g}_0$ and $\dot{g} - \dot{g}_0$ respectively, we have

$$E\left\{\Delta\ddot{g}_0(\epsilon_0) - \int 1(\epsilon_0 \geq s)1(\tau_0 \leq s)\exp\{g_0(s)\}\ddot{g}_0(s)ds \mid X\right\} = 0$$

and

$$E\left\{\Delta(\dot{g} - \dot{g}_0)(\epsilon_0) - \int 1(\epsilon_0 \geq s)1(\tau_0 \leq s)\exp\{g_0(s)\}(\dot{g} - \dot{g}_0)(s)ds \mid X\right\} = 0.$$

Then it follows

$$\begin{aligned}
A &= P\left\{\frac{1}{2}\int_a^b 1(\epsilon_0 \geq s)1(\tau_0 \leq s)\exp\{g_0(s)\}\left\{-[\dot{g}_0(s)(\beta - \beta_0)'X]^2\right.\right. \\
&\qquad \left.\left. + 2\dot{g}_0(s)(\beta - \beta_0)'X(g - g_0)(s) - (g - g_0)(s)\right\}ds\right\} \\
&= -P\left\{\frac{1}{2}\int_a^b 1(\epsilon_0 \geq s)1(\tau_0 \leq s)\exp\{g_0(s)\}\left[\dot{g}_0(s)(\beta - \beta_0)'X - (g - g_0)(s)\right]^2 ds\right\} \\
&= -\frac{1}{2}\int_a^b \exp\{g_0(s)\}P\left\{1(\epsilon_0 \geq s)1(\tau_0 \leq s)\left[\dot{g}_0(s)(\beta - \beta_0)'X - (g - g_0)(s)\right]^2\right\}ds
\end{aligned} \tag{C.4}$$

The integrand is from a to b because of Condition (A.3) and (A.6). Denote $a_0(s) = -\dot{g}_0(s), a_1(s; \epsilon_0, \tau_0, X) = 1(\epsilon_0 \geq s)1(\tau_0 \leq s)(\beta - \beta_0)'X$, and $a_2(s; \epsilon_0, \tau_0) = 1(\epsilon_0 \geq s)1(\tau_0 \leq s)$, then

$$\begin{aligned}
&P\left\{1(\epsilon_0 \geq s)1(\tau_0 \leq s)[\dot{g}_0(s)(\beta - \beta_0)'X - (g - g_0)(s)]^2\right\} \\
&\quad = P\left\{[a_0(s)a_1(s; \epsilon_0, \tau_0, X) + (g - g_0)(s)a_2(s; \epsilon_0, \tau_0)]^2\right\} \\
&\quad \geq a_0^2(s)P(a_1^2) + (g - g_0)^2(s)P(a_2^2) - 2|a_0(s)(g - g_0)(s)P(a_1a_2)|
\end{aligned} \tag{C.5}$$

Using the same argument in Wellner and Zhang [2007], page 2126, under Condition (A.7),

we have $[P(a_1a_2)]^2 \leq (1-\eta)P(a_1^2)P(a_2^2)$ for some $\eta \in (0,1)$. It follows that

$$
\begin{aligned}
&a_0^2(s)P(a_1^2) + (g-g_0)^2(s)P(a_2^2) - 2|a_0(s)(g-g_0)(s)P(a_1a_2)| \\
&\quad \geq a_0(s)P(s_1^2) + (g-g_0)^2(s)P(a_2^2) \\
&\qquad\quad - (1-\eta)^{\frac{1}{2}} \cdot 2|a_0(s)\{P(a_1^2)\}^{\frac{1}{2}}| \cdot |(g-g_0)(s)\{P(a_2^2)\}^{\frac{1}{2}}| \\
&\quad \geq \{1-(1-\eta)^{\frac{1}{2}}\}\{a_0^2(s)P(a_1^2) + (g-g_0)^2(s)P(a_2^2)\} \\
&\quad \gtrsim \dot{g}_0^2(s)(\beta-\beta_0)'P\left[1(\epsilon_0 \geq s)1(\tau_0 \leq s)XX'\right](\beta-\beta_0) \\
&\qquad\quad + (g-g_0)^2(s)P\left[1(\epsilon_0 \geq s)1(\tau_0 \leq s)\right]
\end{aligned}
$$

Hence, we have

$$
\begin{aligned}
A &= -\frac{1}{2}\int_a^b \exp\{g_0(s)\}P\left\{1(\epsilon_0 \geq s)1(\tau_0 \leq s)\left[\dot{g}_0(s)(\beta-\beta_0)'X - (g-g_0)(s)\right]^2\right\}ds \\
&\lesssim -\left\{(\beta-\beta_0)'\left[\int_a^b \exp\{g_0(s)\}\dot{g}_0^2(s)P\left[1(\epsilon_0 \geq s)1(\tau_0 \leq s)XX'\right]ds\right](\beta-\beta_0)\right. \\
&\qquad \left. + \int_a^b \exp\{g_0(s)\}(g-g_0)^2(s)P\left[1(\epsilon_0 \geq s)1(\tau_0 \leq s)\right]ds\right\} \\
&= -(A_1 + A_2)
\end{aligned}
$$

For A_1, Condition (A.3) implies that

$$
\begin{aligned}
P\left[1(\epsilon_0 \geq s)1(\tau_0 \leq s)XX'\right] &= P[XX'E(\epsilon_0 \geq t, \tau_0 \leq t|X)] \\
&\geq P[XX'E(\epsilon_0 \geq b, \tau_0 \leq b|X)] \\
&\geq \delta(1-\delta_1)P(XX').
\end{aligned}
$$

Then Condition (A.2) yields that $P(XX')$ is positive definite and thus its smallest eigenvalue $\lambda_1 > 0$. In addition, $\int_a^b \exp\{g_0(s)\}\dot{g}_0^2(s)ds$ is bounded away from zero since $\exp\{g_0(s)\}, \dot{g}_0^2(s) \geq$

0 but not $\equiv 0$ for $s \in [a, b]$. Hence it follows that

$$A_1 \gtrsim (\beta - \beta_0)' P(XX')(\beta - \beta_0) \geq \lambda_1 |\beta - \beta_0|^2 \gtrsim |\beta - \beta_0|^2$$

For A_2, Condition (A.3) with the fact that $\exp\{g_0(s)\} = d\Lambda_{\epsilon_0}(s)$ yields

$$A_2 \geq P(\epsilon_0 \geq b) P(\tau_0 \leq a) \int_a^b (g - g_0)^2(s) d\Lambda_{\epsilon_0}(s) \gtrsim \|g - g_0\|_2^2.$$

Therefore

$$A \lesssim -(|\beta - \beta_0|^2 + \|g - g_0\|_2^2) = -d^2(\theta, \theta_0)$$

and thus

$$Pl(\beta, g; Z) - Pl(\beta_0, g_0; Z) \lesssim -d^2(\theta, \theta_0) + o(d^2(\theta, \theta_0)) \lesssim -d^2(\theta, \theta_0),$$

i.e., $P(l(\theta_0; Z) - l(\theta; Z)) \gtrsim d^2(\theta, \theta_0)$ for any $\theta \in \Theta_n^p$, which implies that

$$\inf_{\{d(\theta, \theta_0) \geq \varepsilon, \theta \in \Theta_n^p\}} P(l(\theta_0, Z) - l(\theta; Z)) \gtrsim \varepsilon^2$$

Hence Condition C1 of Shen and Wong [1994] on page 583 holds with the constant $\alpha = 1$ in their notation.

Next we verify the Condition C2 of Shen and Wong [1994]. Denote $\epsilon_\beta = Y - X'\beta, \tau_\beta =$

$T - X'\beta$ and $s_\beta = s - X'(\beta - \beta_0)$ for notational simplicity (as done before). It follows that

$$
\begin{aligned}
&[L(\theta; Z) - l(\theta_0; Z)]^2 \\
&= \left\{ \Delta g(\epsilon_\beta) - \int_a^b 1(\epsilon_0 \geq s) 1(\tau_0 \leq s) e^{g(s_\beta)} ds - \Delta g_0(\epsilon_0) + \int_a^b 1(\epsilon_0 \geq s) 1(\tau_0 \leq s) e^{g_0(s)} ds \right\}^2 \\
&\lesssim \Delta [g(\epsilon_\beta) - g_0(\epsilon_0)]^2 + \left\{ \int_a^b 1(\epsilon_0 \geq s) 1(\tau_0 \leq s) [e^{g(s_\beta)} - e^{g_0(s)}] ds \right\}^2 \\
&\lesssim \Delta [g(\epsilon_\beta) - g_0(\epsilon_0)]^2 + \int_a^b [e^{g(s_\beta)} - e^{g_0(s)}]^2 ds \\
&\lesssim \Delta [g(\epsilon_\beta) - g(\epsilon_0)]^2 + \Delta [g(\epsilon_0) - g_0(\epsilon_0)]^2 \\
&\qquad + \int_a^b [g^{g(s_\beta)} - e^{g(s)}]^2 ds + \int_a^b [e^{g(s)} - e^{g_0(s)}]^2 ds \\
&= I_1 + I_2 + I_3 + I_4
\end{aligned}
$$

where the second inequality holds because of the Cauchy-Schwarz inequality

$$
\begin{aligned}
&\left\{ \int_a^b 1(\epsilon_0 \geq s) 1(\tau_0 \leq s) [e^{g(s_\beta)} - e^{g_0(s)}] ds \right\}^2 \\
&\qquad \leq \left\{ \int_a^b 1(\epsilon_0 \geq s) 1(\tau_0 \leq s) ds \right\} \left\{ \int_a^b [g^{g(s_\beta)} - e^{g_0(s)}]^2 ds \right\} \\
&\qquad \leq (b - a) \int_a^b [g^{g(s_\beta)} - e^{g_0(s)}]^2 ds
\end{aligned}
$$

and the third inequality holds by subtracting and adding the terms $g(\epsilon_0)$ and $e^{g(s)}$. For I_1, since $\dot{g} \in \mathcal{H}_n^{p-1}$ is bounded, applying the Taylor expansion of g at ϵ_0 we get

$$
\begin{aligned}
PI_1 &= P\{\Delta [g(\epsilon_0 - X'(\beta - \beta_0)) - g(\epsilon_0)]^2\} \\
&\leq P[\dot{g}(\epsilon_0 - X'(\tilde{\beta} - \beta_0)) X'(\beta - \beta_0)]^2 \\
&\lesssim P[X'(\beta - \beta_0)]^2 = (\beta - \beta_0)' P(XX')(\beta - \beta_0) \\
&\leq \lambda_d |\beta - \beta_0|^2 \lesssim |\beta - \beta_0|^2
\end{aligned}
$$

where λ_d is the largest eigenvalue of $P(XX')$. For I_2, since the density function for $(Y, \Delta =$

$1, X, T)$ is

$$f_{Y,\Delta,X,T}(y,1,x,t) = \lambda_{\epsilon_0}(y - x'\beta_0)e^{-\Lambda_{\epsilon_0}(y-x'\beta_0)}\bar{F}_{C|T,X}(y \mid t,x)f_{T|X}(t \mid x)f_X(x)[P(Y \geq T)]^{-1}$$

it follows that

$$\begin{aligned}
PI_2 &= P[\Delta(g - g_0)^2(\epsilon_0)] \\
&= \int_{\mathcal{X}} \left\{ \int_a^b (g(s) - g_0(s))^2 \lambda_{\epsilon_0}(s) e^{-\Lambda_{\epsilon_0}(s)} \right. \\
&\qquad \left. \cdot \bar{F}_{C|T,X}(s + x'\beta_0 \mid t,x) f_{T|X}(s \mid x)[P(Y \geq T)]^{-1} ds \right\} f_X(x)dx \\
&\lesssim \int_X \left\{ \int_a^b (g(s) - g_0(s))^2 \lambda_{\epsilon_0}(s) ds \right\} f_X(x)dx \\
&= \int_X \|g - g_0\|_2^2 \cdot f_X(x)dx = \|g - g_0\|_2^2
\end{aligned}$$

Where the $\lesssim$ holds by conditions (A.3) and (A.5). Then for I_3, since $g \in \mathcal{H}_n^p$ is bounded, it follows that

$$\begin{aligned}
PI_3 &= P \int_a^b \left[e^{g(s - X'(\beta - \beta_0))} - e^{g(s)} \right]^2 ds \\
&= P \int_a^b e^{2g(s - X'(\tilde{\beta} - \beta_0))} \left[X'(\beta - \beta_0) \right]^2 ds \\
&\lesssim \int_a^b P \left[X'(\beta - \beta_0) \right]^2 ds \\
&\lesssim (\beta - \beta_0) P[XX'](\beta - \beta_0) \lesssim |\beta - \beta_0|^2
\end{aligned}$$

Finally for I_4, by the Taylor expansion for $e^{g(s)}$ at g_0, we have

$$\begin{aligned} PI_4 &= \int_a^b \left[e^{g(s)} - e^{g_0(s)}\right]^2 ds \\ &\leq \int_a^b e^{2\tilde{g}(s)}(g(s) - (g_0(s))^2 ds \\ &= \int_a^b e^{2\tilde{g}(s) - g_0(s)}(g(s) - g_0(s))^2 d\Lambda_{\epsilon_0}(s) \\ &\lesssim \int_a^b (g(s) - g_0(s))^2 d\Lambda_{\epsilon_0}(s) = \|g - g_0\|_2^2, \end{aligned}$$

where $\tilde{g}(s) = g_0(s) + \xi(g - g_0)(s)$ for some $0 < \xi < 1$ and hence is bounded. Therefore we have

$$P(l(\theta; Z) - l(\theta_0; Z))^2 \lesssim |\beta - \beta_0|^2 + \|g - g_0\|_2^2 = d^2(\theta, \theta_0)$$

for any $\theta \in \Theta_n^p$ which implies that

$$\sup_{\{d(\theta,\theta_0)\leq\varepsilon, \theta\in\Theta_n^p\}} \mathrm{Var}(l(\theta_0; Z) - l(\theta; Z)) \leq \sup_{\{d(\theta,\theta_0)\leq\varepsilon, \theta\in\Theta_n^p\}} P(l(\theta_0; Z) - l(\theta; Z))^2 \lesssim \varepsilon^2.$$

So Condition C2 of Shen and Wong [1994] on page 583 holds with the constant $\beta = 1$ in their notation.

Finally, we verify the Condition C3 in Shen and Wong [1994]. By Lemma C.3, for $\mathcal{F}_n = \{l(\theta; Z) - l(\theta_{0,n}; Z) \colon \theta \in \Theta_n^p\}$, we have $N_{[]}(\varepsilon, \mathcal{F}_n, \|\cdot\|_\infty) \lesssim (1/\varepsilon)^{cq_n + d}$. Then by the fact that the covering number is bounded by the bracketing number, it follows that

$$H(\varepsilon, \mathcal{F}_n, \|\cdot\|_\infty) = \log N_{[]}(\varepsilon, \mathcal{F}_n, \|\cdot\|_\infty) \lesssim (cq_n + d)\log(1/\varepsilon) \lesssim n^\eta \log(1/\varepsilon).$$

So Condition C3 of Shen and Wong [1994] on page 583 holds with the constants $2r_0 = \nu$ and $r = 0^+$ in their notation.

Therefore, the constant τ in Theorem 1 of Shen and Wong [1994], page 584, is $\frac{1-\nu}{2} - \frac{\log\log n}{2\log n}$. Since $\frac{\log\log n}{2\log n} \to 0$ as $n \to \infty$, we can pick a $\tilde{\nu}$ slightly greater than ν such that $\frac{1-\tilde{\nu}}{2} \leq \frac{1-\nu}{2} - \frac{\log\log n}{2\log n}$ for large n. We still denote the $\tilde{\nu}$ as ν and then $\tau = \frac{1-\nu}{2}$. Then by definition $\hat{\theta}_n$ maximizes the empirical log-likelihood $\mathbb{P}_n l(\theta; Z)$ over the sieve space Θ_n^p, so $\hat{\theta}_n$ satisfies the inequality (1.1) in Shen and Wong [1994] with $\eta_n = 0$. By Lemma C.2, there exists an $g_{0,n} \in \mathcal{H}_n^p$ such that $\|g_{0,n} - g_0\|_\infty = O(n^{-p\nu})$. Moreover, by the Taylor expansion for $P[l(\beta_0, g_0; Z) - l(\beta, g; Z)]$ in C.3 and plugging in $\theta = \theta_{0,n} = (\beta_0, g_{0,n})$, the Kullback-Leilber pseudodistance of $\theta_{0,n} = (\beta_0, g_{0,n})$ and $\theta_0 = (\beta_0, g_0)$ follows

$$
\begin{aligned}
K(\theta_{0,n}, \theta_0) &= P[l(\theta_0; Z) - l(\theta_{0,n}; Z)] \\
&= -P\{\ddot{l}_{gg}(\beta_0, g_0)[g_{0,n} - g_0, g_{0,n} - g_0]\} + o(\|g_{0,n} - g_0\|_2^2) \\
&= P\left\{\int_a^b 1(\epsilon_0 \geq s) 1(\tau_0 \leq s) \exp\{g_0(s)\}(g_{0,n}(s) - g_0(s))^2 ds\right\} + o(\|g_{0,n} - g_0\|_2^2) \\
&\leq P\left\{\int_a^b (g_{0,n}(s) - g_0(s))^2 d\Lambda_0(s)\right\} + o(\|g_{0,n} - g_0\|_2^2) \\
&= \|g_{0,n} - g_0\|_2^2 + o(\|g_{0,n} - g_0\|_2^2) = O(n^{-2p\nu})
\end{aligned}
$$

where the last equality holds because $\|g_{0,n} - g_0\|_2 \leq \|g_{0,n} - g_0\|_\infty = O(n^{-p\nu})$. Therefore $K^{1/2}(\theta_{0,n}, \theta_0) = O(n^{-p\nu})$. Thus by Theorem 1 of Shen and Wong [1994], we obtain the convergence rate for $\hat{\theta}_n$ as

$$
d(\hat{\theta}_n, \theta_0) = O_p\left\{\max\left(n^{-(1-\nu)/2}, n^{-p\nu}, n^{-p\nu}\right)\right\} = O_p\left\{n^{-\min(p\nu, (1-\nu)/2)}\right\}.
$$

This completes the proof. □

C.4 Proof of Theorem 2.2

Theorem C.2. *Suppose that the conditions in Theorem 1 hold, and let $I(\beta_0)$ be the information matrix evaluated at the true parameters. Then,*

$$n^{1/2}(\hat{\beta}_n - \beta_0) \to N(0, I^{-1}(\beta_0))$$

Proof. To prove this theorem, we use Theorem 2.1 of Ding and Nan [2011a]. Thus, we must check their conditions (A1) - (A6) in the aforementioned work. Note that in our case the criterion function of a single observation is the log-likelihood function $l(\beta, \zeta(\cdot, \beta); Z)$. So instead of m, we will use l to denote the criterion function. We repeat the assumptions we need to check here for ease of reading:

(A1) (Rate of convergence) For an estimator $\hat{\theta}_n = (\hat{\beta}_n, \hat{\zeta}_n(\cdot, \hat{\beta}_n)) \in \Theta_n$ and the true parameter $\theta_0 = (\beta_0, \zeta_0(\cdot, \beta_0)) \in \Theta$, $d(\hat{\theta}_n, \theta_0) = O_p(n^{-\xi})$ for some $\xi > 0$.

(A2) $\dot{S}_\beta(\beta_0, \zeta_0(\cdot, \beta_0)) = 0$ and $\dot{S}_\zeta(\beta_0, \zeta_0(\cdot, \beta_0))[h] = 0$ for all $h \in \mathbb{H}$

(A3) (Positive Information) There exists an $\mathbf{h}^* = (h_1^*, \ldots, h_d^*)'$, where $h_j^* \in \mathbb{H}$ for $j = 1, \ldots, d$, such that

$$\ddot{S}_{\beta\zeta}(\beta_0, \zeta_0(\cdot, \beta_0))[h] - \ddot{S}_{\zeta\zeta}(\beta_0, \zeta_0(\cdot, \beta_0))[\mathbf{h}^*, h] = 0$$

for all $h \in \mathbb{H}$. Furthermore, the matrix

$$\begin{aligned} A &= -\ddot{S}_{\beta\beta}(\beta_0, \zeta_0(\cdot, \beta_0)) + \ddot{S}_{\zeta\beta}(\beta_0, \zeta_0(\cdot, \beta_0))[\mathbf{h}^*] \\ &= -P\{\ddot{m}_{\beta\beta}(\beta_0, \zeta_0(\cdot, \beta_0); Z) - \ddot{m}_{\zeta\beta}(\beta_0, \zeta_0(\cdot, \beta_0); Z)[\mathbf{h}^*]\} \end{aligned}$$

is nonsingular.

(A4) The estimator $(\hat{\beta}_n, \hat{\zeta}_n(\cdot, \hat{\beta}_n))$ satisfies

$$\dot{S}_{\beta,n}(\hat{\beta}_n, \hat{\zeta}_n(\cdot, \hat{\beta}_n)) = o_p(n^{-1/2}) \quad \text{and} \quad \dot{S}_{\zeta,n}(\hat{\beta}_n, \hat{\zeta}_n(\cdot, \hat{\beta}_n))[\mathbf{h}^*] = o_p(n^{-1/2}).$$

(A5) (Stochastic equicontinuity) For some $C > 0$,

$$\sup_{d(\theta,\theta_0)\leq Cn^{-\xi},\theta\in\Theta_n} \mid \sqrt{n}(\dot{S}_{\beta,n} - \dot{S}_\beta)(\beta, \zeta(\cdot,\beta)) - \sqrt{n}(\dot{S}_{\beta,n} - \dot{S}_\beta)(\beta_0, \zeta(\cdot,\beta_0)) \mid = o_p(1)$$

and

$$\begin{aligned}\sup_{d(\theta,\theta_0)\leq Cn^{-\xi},\theta\in\Theta_n} &|\sqrt{n}(\dot{S}_{\zeta,n} - \dot{S}_\zeta)(\beta, \zeta(\cdot,\beta))[\mathbf{h}^*(\cdot,\beta)] \\ &- \sqrt{n}(\dot{S}_{\zeta,n} - \dot{S}_\zeta)(\beta_0, \zeta_0(\cdot,\beta_0))[\mathbf{h}^*(\cdot,\beta_0)] \mid = o_p(1).\end{aligned}$$

(A6) (Smoothness of the model) For some $\alpha > 1$ satisfying $\alpha\xi > 1/2$, and for θ in a neighborhood of θ_0: $\{\theta\colon d(\theta,\theta_0) \leq Cn^{-\xi}, \theta \in \Theta_n\}$,

$$\begin{aligned}&\mid \dot{S}_\beta(\beta, \zeta(\cdot,\beta)) - \dot{S}_\beta(\beta_0, \zeta_0(\cdot,\beta_0)) - \ddot{S}_{\beta\beta}(\beta_0, \zeta_0(\cdot,\beta_0))(\beta - \beta_0) \\ &\qquad - \ddot{S}_{\beta\zeta}(\beta_0, \zeta_0(\cdot,\beta_0))[\zeta(\cdot,\beta) - \zeta_0(\cdot,\beta_0)] \mid \\ &= O(d^\alpha(\theta,\theta_0))\end{aligned}$$

and

$$\begin{aligned}&\mid \dot{S}_\zeta(\beta, \zeta(\cdot,\beta))[\mathbf{h}^*(\cdot,\beta)] - \dot{S}_\zeta(\beta_0, \zeta_0(\cdot,\beta_0))[\mathbf{h}^*(\cdot,\beta)] - \ddot{S}_{\zeta\beta}(\beta_0, \zeta_0(\cdot,\beta_0))[\mathbf{h}^*(\cdot,\beta_0)](\beta - \beta_0) \\ &\qquad - \ddot{S}_{\zeta\zeta}(\beta_0, \zeta_0(\cdot,\beta_0))[\mathbf{h}^*(\cdot,\beta_0), \zeta(\cdot,\beta) - \zeta_0(\cdot,\beta_0)] \mid \\ &= O(d^\alpha(\theta,\theta_0))\end{aligned}$$

By Theorem 1, we know that assumption (A1) holds with $\xi = \min(p\nu, (1-\nu)/2)$ and the norm defined in (C.2). Assumption (A2) automatically holds for scores.

For assumption (A3) consider that

$$\begin{aligned}
&\ddot{S}_{\beta\zeta}(\beta_0,\zeta_0(\cdot,\beta_0))[h] - \ddot{S}_{\zeta\zeta}(\beta_0,\zeta_0(\cdot,\beta_0))[\mathbf{h}^*,h] \\
&\qquad = P\{\ddot{l}_{\beta\zeta}(\beta_0,\zeta_0(\cdot,\beta_0);Z)[h] - \ddot{l}_{\zeta\zeta}(\beta_0,\zeta_0(\cdot,\beta_0);Z)[\mathbf{h}^*,h]\}
\end{aligned}$$

for all $h \in \mathbb{H}$ with $h(t,x,\beta) = w(t - x'(\beta - \beta_0))$. Note that

$$\begin{aligned}
&P\{\ddot{l}_{\beta\zeta}(\beta_0,\zeta_0(\cdot,\beta_0);Z)[h] - \ddot{l}_{\zeta\zeta}(\beta_0,\zeta_0(\cdot,\beta_0);Z)[\mathbf{h}^*,h]\} \\
&\qquad = P\left\{-X\left[\Delta\dot{w}(\epsilon_0) - \int_a^b 1(\epsilon_0 \geq s)1(\tau_0 \leq s)\exp\{g_0(s)\}\dot{w}(s)ds\right]\right. \\
&\qquad\qquad \left. + \int_a^b 1(\epsilon_0 \geq s)1(\tau_0 \leq s)\exp\{g_0(s)\}w(s)[X\dot{g}_0(s) + \mathbf{w}^*(s)]ds\right\}.
\end{aligned}$$

Since $P\{\dot{l}_\zeta(\beta_0,\zeta_0(\cdot,\beta_0);Z)[h]|X\} = 0$ for all $h \in \mathbb{H}$, replacing $h(\cdot,\beta_0)$ by $\dot{w}$ we have

$$\begin{aligned}
&P\left\{-X\left[\Delta\dot{w}(\epsilon_0) - \int_a^b 1(\epsilon_0 \geq s)1(\tau_0 \leq s)\exp\{g_0(s)\}\dot{w}(s)ds\right]\right\} \\
&\qquad = P\left\{-X \cdot P\left[\Delta\dot{w}(\epsilon_0) - \int_a^b 1(\epsilon_0 \geq s)1(\tau_0 \leq s)\exp\{g_0(s)\}\dot{w}(s)ds\middle| X\right]\right\} \\
&\qquad = P\{-X \cdot 0\} = 0.
\end{aligned}$$

It follows that we only need to find a $\mathbf{w}^*$ such that

$$P\left\{\int_a^b 1(\epsilon_0 \geq s)1(\tau_0 \leq s)\exp\{g_0(s)\}w(s)[X\dot{g}_0(s) + \mathbf{w}^*(s)]ds\right\} = 0.$$

In other words,

$$\dot{g}_0(s)P[1(\epsilon_0 \geq s)1(\tau_0 \leq s)X] + \mathbf{w}^*(s)P[1(\epsilon_0 \geq s)1(\tau_0 \leq s)] = 0.$$

The obvious choice is

$$\mathbf{h}^*(s,t,x,\beta_0) = \mathbf{w}^*(s) = -\dot{g}(s)\frac{P[1(\epsilon_0 \geq s)1(\tau_0 \leq s)X]}{P[1(\epsilon_0 \geq s)1(\tau_0 \leq s)]} = -\dot{g}(s)P(X \mid \epsilon_0 \geq s, \tau_0 \leq s). \quad \text{(C.6)}$$

Then it follows that

$$\begin{aligned}
&\dot{l}_\beta(\beta_0,\zeta_0(\cdot,\beta_0);Z) - \dot{l}_\zeta(\beta_0,\zeta_0(\cdot,\beta_0);Z)[\mathbf{h}^*] \\
&\quad = \Delta\{-\dot{g}_0(\epsilon_0)\}\{X - P(X \mid \epsilon_0 \geq Y - X'\beta_0, \tau_0 \leq T - X'\beta_0)\} \\
&\qquad - \int 1(\epsilon_0 \geq s)1(\tau_0 \leq s)\{X - P(X \mid \epsilon_0 \geq s, \tau_0 \leq s)\}\{-\dot{g}_0(s)\}\exp\{g_0(s)\}ds \\
&\quad = \int \{X - P(X|\epsilon_0 \geq s, \tau_0 \leq s)\}\{-\dot{g}_0(s)\}dM(s) \\
&\quad = l^*_{\beta_0}(Y,\Delta,X,T)
\end{aligned}$$

which is the efficient score functions for β_0 as mentioned in Section 3 of Lai and Ying [1992a], where

$$M(s) = \Delta 1(\tau_0 \leq \epsilon_0 \leq s) - \int_{-\infty}^{s} 1(\epsilon_0 \geq u)1(\tau_0 \leq u)\exp\{g_0(u)\}du$$

as in Equation 2.38 of Lai and Ying [1991b].

By the fact of zero-mean for a score function, it is straightforward to verify the following equalities:

$$\begin{aligned}
P\ddot{l}_{\beta\zeta}(\beta,\zeta(\cdot,\beta);Z)[h] &= -P\{\dot{l}_\beta(\beta,\zeta(\cdot,\beta);Z)\dot{l}'_\zeta(\beta,\zeta(\cdot,\beta);Z)[h]\} \\
P\ddot{l}_{\zeta\beta}(\beta,\zeta(\cdot,\beta);Z)[h] &= -P\{\dot{l}_\zeta(\beta,\zeta(\cdot,\beta);Z)[h]\dot{l}'_\beta(\beta,\zeta(\cdot,\beta);Z)\} \\
P\ddot{l}_{\beta\beta}(\beta,\zeta(\cdot,\beta);Z) &= -P\{\dot{l}_\beta(\beta,\zeta(\cdot,\beta);Z)\dot{l}'_\beta(\beta,\zeta(\cdot,\beta);Z)\} \\
P\ddot{l}_{\zeta\zeta}(\beta,\zeta(\cdot,\beta);Z)[h_1,h_2] &= -P\{\dot{l}_\zeta(\beta,\zeta(\cdot,\beta);Z)[h_1]\dot{l}'_\zeta(\beta,\zeta(\cdot,\beta);Z)[h_2]\}
\end{aligned}$$

Then together with the fact that

$$P\{\ddot{l}_{\beta\zeta}(\beta_0, \zeta_0(\cdot, \beta_0); Z)[\mathbf{h}^*] - \ddot{l}_{\zeta\zeta}(\beta_0, \zeta_0(\cdot, \beta_0); Z)[\mathbf{h}^*, \mathbf{h}^*]\} = 0$$

the matrix A in assumption (A3) of Theorem 2.1 is given by

$$\begin{aligned}
A &= P\{-\ddot{l}_{\beta\beta}(\beta_0, \zeta_0(\cdot, \beta_0); Z) + \ddot{l}_{\zeta\beta}(\beta_0, \zeta_0(\cdot, \beta_0); Z)[\mathbf{h}^*, \mathbf{h}^*] \\
&\qquad + \ddot{l}_{\beta\zeta}(\beta_0, \zeta_0(\cdot, \beta_0); Z)[\mathbf{h}^*] - \ddot{l}_{\zeta\zeta}(\beta_0, \zeta_0(\cdot, \beta_0); Z)[\mathbf{h}^*, \mathbf{h}^*]\} \\
&= P\{\dot{l}_\beta(\beta, \zeta(\cdot, \beta); Z)\dot{l}'_\beta(\beta, \zeta(\cdot, \beta); Z) - \dot{l}_\zeta(\beta, \zeta(\cdot, \beta); Z)[\mathbf{h}^*]\dot{l}'_\beta(\beta, \zeta(\cdot, \beta); Z) \\
&\qquad - \dot{l}_\beta(\beta, \zeta(\cdot, \beta); Z)\dot{l}'_\zeta(\beta, \zeta(\cdot, \beta); Z)[\mathbf{h}^*] + \dot{l}_\zeta(\beta, \zeta(\cdot, \beta); Z)[\mathbf{h}^*]\dot{l}'_\zeta(\beta, \zeta(\cdot, \beta); Z)[\mathbf{h}^*] \\
&= P\{\dot{l}_\beta(\beta_0, \zeta_0(\cdot, \beta_0); Z) - \dot{l}_\zeta(\beta_0, \zeta_0(\cdot, \beta_0); Z)[\mathbf{h}^*]\}^{\otimes 2} \\
&= Pl^*_{\beta_0}(Y, \Delta, X, T)^{\otimes 2}
\end{aligned}$$

which is the information matrix for β_0.

To verify (A4) we note that the first part automatically holds since $\hat{\beta}_n$ satisfies the score equation $\dot{S}_{\beta,n}(\hat{\beta}_n, \hat{\zeta}_n(\cdot, \hat{\beta}_n)) = \mathbb{P}_n \dot{l}_\beta(\hat{\beta}_n, \hat{\zeta}_n(\cdot, \hat{\beta}_n); Z) = 0$. Next we shall show that

$$\begin{aligned}
&\dot{S}_{\zeta,n}(\hat{\beta}_n, \hat{\zeta}_n(\cdot, \hat{\beta}_n))[h^*_j] \\
&\quad = \mathbb{P}\left\{\Delta w^*_j(Y - X'\hat{\beta}_n) - \int 1(Y \geq s)1(T \leq s)\exp\{\hat{\zeta}_n(s, X, \hat{\beta}_n)\}w^*_j(s - X'\hat{\beta}_n)ds\right\} \\
&\quad = o_p(n^{-1/2})
\end{aligned}$$

where $w^*_j(s) = -\dot{g}_0(s)P(X_j|\epsilon_0 \geq s, \tau_0 \leq s)$, $j = 1, \ldots, d$ is the jth component of $\mathbf{w}^*(s)$ given in C.6. According to Lemma C.4, there exists $h^*_{j,n} \in \mathcal{H}^2_n$ such that $\|h^*_j - h^*_{j,n}\|_\infty = O(n^{-2\nu})$. Then by the score equation for γ: $\dot{S}_{\gamma,n}(\hat{\beta}_n, \hat{\gamma}_n) = \mathbb{P}_n \dot{l}_\gamma(\hat{\beta}_n, \hat{\gamma}_n; Z) = 0$ and the fact that $w^*_{j,n}(s)$ can be written as $w^*_{j,n}(s) = \sum_{k=1}^{q_n} \gamma^*_{j,k} B_k(s)$ for some coefficients $\{\gamma^*_{j,1}, \ldots, \gamma^*_{j,q_n}\}$ and the basis

functions $B_k(s)$ of the spline space, it follows that

$$\mathbb{P}_n \left\{ \Delta w^*_{j,n}(Y - X'\hat{\beta}_n) - \int 1(Y \geq s) 1(T \leq s) \exp\{\hat{\zeta}_n(s, X, \hat{\beta}_n)\} w^*_{j,n}(s - x'\hat{\beta}_n) ds \right\} = 0$$

So it suffices to show that for each $1 \leq j \leq d$,

$$I_n = \mathbb{P}_n \dot{l}_\zeta(\hat{\beta}_n, \hat{\zeta}_n(\cdot, \hat{\beta}_n); Z)[h^*_j - h^*_{j,n}] = o_p(n^{-1/2}).$$

Since $P\{\dot{l}_\zeta(\beta_0, \zeta_0(\cdot, \beta_0); Z)[h^*_j - h^*_{j,n}]\} = 0$, we decompose I_n into $I_n = I_{1n} + I_{2n}$, where

$$I_{1n} = (\mathbb{P}_n - P)\dot{l}_\zeta(\hat{\beta}_n, \hat{\zeta}_n(\cdot, \hat{\beta}_n); Z)[h^*_j - h^*_{j,n}]$$

and

$$I_{2n} = P\{\dot{l}_\zeta(\hat{\beta}_n, \hat{\zeta}_n(\cdot, \hat{\beta}_n); Z)[h^*_j - h^*_{j,n}] - \dot{l}_\zeta(\beta_0, \zeta_0(\cdot, \beta_0); Z)[h^*_j - h^*_{j,n}]\}.$$

We will show that I_{1n} and I_{2n} are both $o_p(n^{-1/2})$.

First, consider I_{1n}. According to Lemma C.5, the ε-bracketing number associated with $\|\cdot\|_\infty$ norm for the class $\mathcal{F}^j_n(\eta)$ defined in Lemma C.5 is bounded by $(\eta/\varepsilon)^{cq_n+d}$. This implies that

$$\log N_{[]}(\varepsilon, \mathcal{F}^j_n(\eta), L_2(P)) \leq \log N_{[]}(\varepsilon, \mathcal{F}^j_n(\eta), \|\cdot\|_\infty) \lesssim q_n \log(\eta/\varepsilon),$$

which leads to the bracketing integral

$$\begin{aligned} J_{[]}(\eta, \mathcal{F}^j_n(\eta), L_2(P)) &= \int_0^\eta \sqrt{1 + \log N_{[]}(\varepsilon, \mathcal{F}^j_n(\eta), L_2(P))} d\varepsilon \\ &\lesssim q_n^{1/2} \eta. \end{aligned}$$

Now we pick η to be $\eta_n = O\{n^{-\min(2\nu,(1-\nu)/2)}\}$, then

$$\|h_j^* - h_{j,n}^*\|_\infty = O(n^{-2\nu} \leq O(n^{-\min(2\nu,(1-\nu)/2)}) = \eta_n,$$

and since $p \geq 3$,

$$d(\hat{\theta}_n, \theta_0) = O_p\{n^{-\min(p\nu,(1-\nu)/2)}\} \leq O_p\{n^{-\min(2\nu,(1-\nu)/2)}\} = \eta_n.$$

Therefore, $\dot{l}_\zeta(\hat{\beta}_n, \hat{\zeta}_n(\cdot, \hat{\beta}_n); z)[h_j^* - h_{j,n}^*] \in \mathcal{F}_n^j(\eta_n)$. Denote $s_\beta = s - X'(\beta - \beta_0)$ for notational simplicity, for any $\dot{l}_\zeta(\theta; Z)[h_j^* - h] \in \mathcal{F}_n^j(\eta_n)$, it follows that

$$\begin{aligned}
&P\{\dot{l}_\zeta(\theta; Z)[h_j^* - h]\}^2 \\
&\quad = P\left\{\Delta(w_j^* - w)(\epsilon_\beta) + \int_a^b 1(\epsilon_0 \geq s)1(\tau_0 \leq s)\exp\{g(s_\beta)\}(w_j^* - w)(s_\beta)ds\right\}^2 \\
&\quad \lesssim \|w_j^* - w\|_\infty^2 + P\left\{\int_a^b \exp\{2g(s_\beta)\}(w_j^* - w)^2(s_\beta)ds\right\} \\
&\quad \lesssim \|w_j^* - w\|_\infty^2 + \|w_j^* - w\|_\infty^2 \int_a^b P[\exp\{2g(s_\beta)\}]ds,
\end{aligned}$$

where the first inequality holds because of the Cauchy-Schwarz inequality. Since $\|w_j^* - w\|_\infty \leq \eta_n$, by the same argument as Shen and Wong [1994] (page 591), for slowly growing c_n (their l_n), for example, $c_n = o(\log(\eta_n^{-1}))$, we know that $\|\dot{l}_\zeta(\theta; Z)[h_j^* - h]\|_\infty$ is bounded by some constant $0 < M < \infty$ and $P\{\dot{l}_\zeta(\theta; Z)[h_j^* - h]\}^2 \lesssim \eta_n$ for a slightly enlarged η_n obtained by a fine adjustment of ν. Then by the maximal inequality in Lemma 3.4.2 of van der Vaart

[2023], it follows that

$$
\begin{aligned}
E_P\|\mathbb{G}_n\|_{\mathcal{F}_n^j(\eta_n)} &\lesssim J_{[]}(\eta_n, \mathcal{F}_n^j(\eta_n), L_2(P)) \left(1 + \frac{J_{[]}(\eta_n, \mathcal{F}_n^j(\eta_n), L_2(P))}{\eta_n^2\sqrt{n}} M\right) \\
&\lesssim q_n^{1/2}\eta_n + q_n n^{-1/2} \\
&= O\left\{n^{\nu/2-\min(2\nu,(1-\nu)/2)}\right\} + O(n^{\nu-1/2}) \\
&= O\left\{n^{-\min(3\nu/2,1/2-\nu)}\right\} + O(n^{\nu-1/2}) = o(1),
\end{aligned}
$$

where the last equality holds because $0 < \nu < 1/2$. Thus by Markov's inequality, $I_{1n} = n^{-1/2}\mathbb{G}_n\dot{l}_\zeta(\hat{\theta}_n; Z)[h_j^* - h_{j,n}^*] = o_p(n^{-1/2})$.

Next, for I_{2n}, the Taylor expansion for $\dot{l}_\zeta(\hat{\theta}_n; Z)[h_j^* - h_{j,n}^*]$ at θ_0 yields

$$
\begin{aligned}
&\dot{l}_\zeta(\hat{\beta}_n,\hat{\zeta}_n(\cdot,\hat{\beta}_n); Z)[h_j^* - h_{j,n}^*] - \dot{l}_\zeta(\beta_0, \zeta_0(\cdot,\beta_0); Z)[h_j^* - h_{j,n}^*] \\
&\quad = (\hat{\beta} - \beta_0)'\ddot{l}_{\beta\zeta}(\tilde{\beta}_n, \tilde{\zeta}_n(\cdot,\tilde{\beta}_n); Z)[h_j^* - h_{j,n}^*] \\
&\qquad + \ddot{l}_{\zeta\zeta}(\tilde{\beta}_n, \tilde{\zeta}_n(\cdot,\tilde{\beta}_n); Z)[h_j^* - h_{j,n}^*, \hat{\zeta}_n - \zeta_0],
\end{aligned}
$$

where $(\tilde{\beta}_n, \tilde{\zeta}_n(\cdot, \tilde{\beta}_n))$ is between $(\beta_0, \zeta_0(\cdot, \beta_0))$ and $(\hat{\beta}_n, \hat{\zeta}_n(\cdot, \hat{\beta}_n))$. Then it follows that

$$
\begin{aligned}
&|\ddot{l}_{\beta\zeta}(\tilde{\beta}_n, \tilde{\zeta}_n(\cdot, \tilde{\beta}_n); Z)[h_j^* - h_{j,n}^*]| \\
&\quad = \left| X \left\{ \Delta(\dot{w}_j^* - \dot{w}_{j,n}^*)(\epsilon_{\tilde{\beta}_n}) \right.\right. \\
&\qquad - \int_a^b 1(\epsilon_0 \geq s) 1(\tau_0 \leq s) \exp\{\tilde{g}_n(s_{\tilde{\beta}_n})\}[(\dot{w}_j^* - \dot{w}_{j,n}^*)(s_{\tilde{\beta}_n}) \\
&\qquad\quad \left.\left. + \dot{\tilde{g}}_n(s_{\tilde{\beta}_n})(w_j^* - w_{j,n}^*)(s_{\tilde{\beta}_n})] ds \right\} \right| \\
&\quad \lesssim \|\dot{w}_j^* - \dot{w}_{j,n}^*\|_\infty + \|\dot{w}_j^* - \dot{w}_{j,n}^*\|_\infty \left\{ \int_a^b \exp\{\tilde{g}_n(s_{\tilde{\beta}_n})\} ds \right\} \\
&\qquad + \|w_j^* - w_{j,n}^*\|_\infty \left\{ \int_a^b \exp\{\tilde{g}_n(s_{\tilde{\beta}_n})\} \dot{\tilde{g}}(s_{\tilde{\beta}_n}) ds \right\} \\
&\quad \lesssim \|\dot{w}_j^* - \dot{w}_{j,n}^*\|_\infty + \|w_j^* - w_{j,n}^*\|_\infty \\
&\quad = O(n^{-\nu}) + O(n^{-2\nu}) = O(n^{-\nu})
\end{aligned}
$$

where the second inequality holds because $\tilde{g}_n$ and its first derivative $\dot{\tilde{g}}_n$ are bounded (or growing with n slowly enough so it can be effectively treated as bounded based on the same argument of Shen and Wong [1994] on page 591), and the last equality holds due to the Corollary 6.21 of Schumaker [2007] that $\|\dot{w}_j^* - \dot{w}_{j,n}^*\|_\infty = O(n^{-(2-1)\nu}) = O(n^{-\nu})$. Thus,

$$
\begin{aligned}
&P|(\hat{\beta}_n - \beta_0)' \ddot{l}_{\beta\zeta}(\tilde{\beta}_n, \tilde{\zeta}_n(\cdot, \tilde{\beta}_n); Z)[h_j^* - h_{j,n}^*]| \\
&\qquad = |\hat{\beta}_n - \beta_0| \cdot O(n^{-\nu}) \\
&\qquad = O_p\{n^{-\min(p\nu, (1-\nu)/2)}\} \cdot O(n^{-\nu}) \\
&\qquad = O_p\{n^{-\min((p_1)\nu, (1+3\nu)/2)}\}.
\end{aligned}
$$

Also,

$$
\begin{aligned}
&|\ddot{l}_{\zeta\zeta}(\tilde{\beta}_n, \tilde{\zeta}_n(\cdot, \tilde{\beta}_n); Z)[h_j^* - h_{j,n}^*, \hat{\zeta}_n - \zeta_0]| \\
&\quad = \left| \int_a^b 1(\epsilon_0 \geq s) 1(\tau_0 \leq s) \exp\{\tilde{g}_n(s_{\tilde{\beta}_n})\}(w_j^* - w_{j,n}^*)(s_{\tilde{\beta}_n})(\hat{g}_n - g_0)(s_{\tilde{\beta}_n}) ds \right| \\
&\quad \leq \|w_j^* - w_{j,n}^*\|_\infty \cdot \left\{ \int_a^b \exp\{\tilde{g}_n(s_{\tilde{\beta}_n})\}(\hat{g}_n - g_0)(s_{\tilde{\beta}_n}) ds \right\} \\
&\quad = \|w_j^* - w_{j,n}^*\|_\infty \cdot I_{3n}.
\end{aligned}
$$

By the Cauchy-Schwarz inequality and the boundedness of $\tilde{g}_n$, we have

$$
\begin{aligned}
P\{I_{3n}\}^2 &= P\left\{ \int_a^b \exp\{\tilde{g}_n(s_{\tilde{\beta}_n})\}(\hat{g}_n - g_0)(s_{\tilde{\beta}_n}) ds \right\}^2 \\
&\lesssim \int_{\mathcal{X}} \int_a^b (\hat{g}_n - g_0)^2 (s_{\tilde{\beta}_n}) d\Lambda_{\epsilon_0}(s) dF_X(x) \\
&= \|\hat{\zeta}_n(\cdot, \tilde{\beta}_n) - \zeta_0(\cdot, \tilde{\beta}_n)\|_2^2 \\
&\lesssim |\tilde{\beta}_n - \hat{\beta}_n|^2 + \|\hat{\zeta}_n(\cdot, \tilde{\beta}_n) - \zeta_0(\cdot, \beta_0)\|_2^2 + |\beta_0 - \tilde{\beta}_n|^2 \\
&\lesssim |\hat{\beta}_n - \beta_0|^2 + \|\hat{\zeta}_n(\cdot, \tilde{\beta}_n) - \zeta_0(\cdot, \beta_0)\|_2^2 \\
&= d(\hat{\theta}_n, \theta_0)^2.
\end{aligned}
$$

Hence $P|I_{3n}| \lesssim d(\hat{\theta}_n, \theta_0)$ and

$$
\begin{aligned}
&P|\ddot{l}_{\zeta\zeta}(\tilde{\beta}_n, \tilde{g}_n; Z)[h_j^* - h_{j,n}^*, \hat{\zeta}_n - \zeta_0]| \\
&\quad \lesssim \|w_j^* - w_{j,n}^*\|_\infty \cdot d(\hat{\theta}_n, \theta_0) \\
&\quad = O(n^{-2\nu}) \cdot O_p\{n^{-\min(p\nu, (1-\nu)/2)}\} \\
&\quad = O_p\{n^{-\min((p+2)\nu, (1+3\nu)/2)}\}.
\end{aligned}
$$

Since $(2(1+p)^{-1} < \nu < (1+2p)^{-1}$, it follows that $I_{2n} = O\{n^{-\min((p+2)\nu, (1+3\nu)/2)}\} = o(n^{-1/2})$. Thus $I_n = I_{1n} + I_{2n} = o_p(n^{-1/2})$, and condition (A4) holds.

Now we verify assumption (A5). First by Lemma C.6, the ε-bracketing numbers for the classes of functions $\mathcal{F}^{\beta}_{n,j}(\eta)$ and $\mathcal{F}^{\zeta}_{n,j}(\eta)$ are both bounded by $(\eta/\varepsilon)^{cq_n+d}$, which implies that the corresponding ε-bracketing integrals are both bounded by $q_n^{1/2}\eta$, that is,

$$J_{[]}(\eta, \mathcal{F}^{\beta}_{n,j}(\eta), L_2(P)) \lesssim q_n^{1/2}\eta \quad \text{and} \quad J_{[]}(\eta, \mathcal{F}^{\zeta}_{n,j}(\eta), L_2(P)) \lesssim q_n^{1/2}\eta.$$

Then for $\dot{l}_{\beta_j}(\theta; z) - \dot{l}_{\beta_j}(\theta_0; z)$, by applying the Cauchy-Schwarz inequality, together with subtracting and adding the terms $\dot{g}(\epsilon_0), e^{g_0(s_\beta)}\dot{g}(s_\beta), e^{g_0(s)}\dot{g}(s_\beta)$ and $e^{g_0(s)}\dot{g}_0(s_\beta)$ we have

$$\begin{aligned}
&\{\dot{l}_{\beta_j}(\theta; Z) - \dot{l}_{\beta_j}(\theta_0; Z)\}^2 \\
&\quad = \left\{-\Delta X_j[\dot{g}(\epsilon_\beta) - \dot{g}_0(\epsilon_0)] + X_j \int_a^b 1(\epsilon_0 \geq s)1(\tau_0 \leq s)[e^{t(s_\beta)}\dot{g}(s_\beta) - e^{g_0(s)}\dot{g}_0(s)]ds\right\}^2 \\
&\quad \lesssim \{\Delta[\dot{g}(\epsilon_\beta) - \dot{g}_0(\epsilon_0)]^2 + \left\{\int_a^b [e^{g(s_\beta)}\dot{g}(s_\beta) - e^{g_0(s)}\dot{g}_0(s)]^2 ds\right\} \\
&\quad \lesssim \{\Delta[\dot{g}(\epsilon_\beta) - \dot{g}(\epsilon_0)]^2\} + \{\Delta[\dot{g}(\epsilon_0) - \dot{g}_0(\epsilon_0)]^2\} \\
&\qquad + \int_a^b \left\{\left[e^{g(s_\beta)} - e^{g_0(s_\beta)}\right]^2 + \left[e^{g_0(s_\beta)} - e^{g_0(s)}\right]^2\right\} \dot{g}^2(s_\beta) ds \\
&\qquad + \int_a^b e^{2g_0(s)} \left\{[\dot{g}(s_\beta) - \dot{g}_0(s_\beta)]^2 + e^{2g_0(s)}[\dot{g}_0(s_\beta) - \dot{g}_0(s)]^2\right\} ds \\
&\quad = B_1 + B_2 + B_3 + B_4.
\end{aligned}$$

For B_1, since $\ddot{g}$ is bounded and the largest eigenvalue of $P(XX')$ satisfies $0 < \lambda_d < \infty$ by condition (A.2), it follows that

$$\begin{aligned}
PB_1 &\leq P[\ddot{(}Y - X'\tilde{\beta})X'(\beta - \beta_0)]^2 \lesssim P[X'(\beta - \beta_0)]^2 \\
&\leq \lambda_d |\beta - \beta_0|^2 \lesssim |\beta - \beta_0|^2 \leq \eta_2.
\end{aligned}$$

For B_2, we have

$$\begin{aligned} PB_2 &\leq \int_{\mathcal{X}} \left\{ \int_a^b \left(\dot{g}(s) - \dot{g}_0(s) \right)^2 d\Lambda_{\epsilon_0}(s) \right\} dF_X(x) \\ &= \| \dot{g}(\psi(\cdot, \beta_0)) - \dot{g}_0(\psi(\cdot, \beta_0)) \|_2^2 \\ &\lesssim |\beta - \beta_0|^2 + \| \dot{g}(\psi(\cdot, \beta)) - \dot{g}_0(\psi(\cdot, \beta_0)) \|_2^2 \lesssim \eta^2 \end{aligned}$$

For B_3, by using the mean value theorem, it follows that

$$\begin{aligned} PB_3 &= P \left\{ \int_a^b \left\{ \left[e^{\tilde{g}(s_\beta)} (g - g_0)(s_\beta) \right]^2 + \left[e^{g_0(s_{\tilde{\beta}})} X'(\beta - \beta_0) \right]^2 \right\} \dot{g}^2(s_\beta) ds \right\} \\ &\lesssim \int_{\mathcal{X}} \int_a^b (g - g_0)^2(s_\beta) d\Lambda_{\epsilon_0}(s) dF_X(x) + P[X'(\beta - \beta_0)]^2 \\ &\lesssim \| \zeta(\cdot, \beta) - \zeta_0(\cdot, \beta_0) \|_2^2 + |\beta - \beta_0|^2 \leq \eta^2, \end{aligned}$$

where $\tilde{g} = g_0 + \xi(g - g_0)$ for some $0 < \xi < 1$ and thus is bounded. Finally, for B_4, by the mean value theorem, it follows that

$$\begin{aligned} PB_4 &= P \left\{ \int_a^b e^{2g_0(s)} \left\{ [\dot{g}(s_\beta) - \dot{g}_0(s_\beta)]^2 + e^{2g_0(s)} [\dot{g}_0(s_\beta) - \dot{g}_0(s)]^2 \right\} ds \right\} \\ &\lesssim \int_{\mathcal{X}} \int_a^b (\dot{g} - \dot{g}_0)^2(s_\beta) d\Lambda_{\epsilon_0}(s) dF_X(x) + P \int_a^b [\ddot{g}_0(s_{\tilde{\beta}}) X'(\beta - \beta_0)]^2 ds \\ &\lesssim \| \dot{g}(\psi(\cdot, \beta)) - \dot{g}_0(\psi(\cdot, \beta)) \|_2^2 + P[X'(\beta - \beta_0)]^2 \\ &\lesssim \| \dot{g}(\psi(\cdot, \beta)) - \dot{g}_0(\psi(\cdot, \beta_0)) \|_2^2 + |\beta - \beta_0|^2 \lesssim \eta^2 \end{aligned}$$

Therefore we have $P\{\dot{l}_{\beta_j}(\theta; Z) - \dot{l}_{\beta_j}(\theta_0; Z)\}^2 \lesssim \eta^2$. Using the similar argument, we can show that $P\{\dot{l}_\zeta(\theta; Z)[h_j^*] - \dot{l}_\zeta(\theta_0; Z)[h_j^*]\}^2 \lesssim \eta^2$. By Lemma C.1, we also have $\|\dot{l}_{\beta_j}(\theta; Z) - \dot{l}_{\beta_j}(\theta_0; Z)\|_\infty$ and $\|\dot{l}_\zeta(\theta; Z)[h_j^*] - \dot{l}_\zeta(\theta_0; Z)[h_j^*]\|_\infty$ are both bounded. Now we pick η as $\eta_n = O\{n^{-\min((p-1)\nu, (1-\nu)/2)}\}$, then by the maximal inequality in Lemma 3.4.2 of van der Vaart

[2023], it follows that

$$E_p\|\mathbb{G}_n\|_{\mathcal{F}^{\beta}_{n,j}(\eta_n)} \lesssim q_n^{1/2}\eta_n + q_n n^{-1/2}]$$
$$= O\{n^{\max((3/2-p)\nu,\nu-1/2)}\} + O(n^{\nu-1/2}) = o(1),$$

where the last equality holds since $p \geq 3$ and $\nu < \frac{1}{2}$. Similarly, we have $E_P\|\mathbb{G}_n\|_{\mathcal{F}^{\zeta}_{n,j}(\eta_n)} = o(1)$. Thus for $\xi = \min(p\nu, (1-\nu)/2)$ and

$$Cn^{-\xi} = O\{n^{-\min(p\nu,(1-\nu)/2)}\}$$

by Markov's inequality,

$$\sup_{d(\theta,\theta_0)\leq Cn^{-\xi}} \mathbb{G}_n\{\dot{l}_{\beta_j}(\beta,\zeta(\cdot,\beta);Z) - \dot{l}_{\beta_j}(\beta_0,\zeta_0(\cdot,\beta_0);Z)\} = o_p(1),$$
$$\sup_{d(\theta,\theta_0)\leq Cn^{-\xi}} \mathbb{G}_n\{\dot{l}_{\zeta}(\beta,\zeta(\cdot,\beta);Z)[h_j^*] - \dot{l}_{\zeta}(\beta_0,\zeta_0(\cdot,\beta_0);Z)[h_j^*]\} = o_p(1).$$

This completes the verification of assumption (A5).

Finally, assumption (A6) can be verified by using the Taylor expansion. Since the proofs for the two equations in (A6) are essentially identical, we just prove the first equation. In a neighborhood of θ_0: $\{\theta\colon d(\theta,\theta_0) \leq Cn^{-\xi}, \theta \in \Theta_n^p\}$ with $\xi = \min(p\nu, (1-\nu)/2)$, the Taylor expansion for $\dot{l}_\beta(\theta, Z)$ yields

$$\dot{l}_\beta(\theta;Z) = \dot{l}_\beta(\theta_0;Z) + \ddot{l}_{\beta\beta}(\tilde{\theta};Z)(\beta-\beta_0) + \ddot{l}_{\beta g}(\tilde{\theta};Z)[\zeta(\cdot,\beta)-\zeta_0(\cdot,\beta_0)]$$
$$= \dot{l}_\beta(\theta_0;Z) + \ddot{l}_{\beta\beta}(\theta_0;Z)(\beta-\beta_0) + \ddot{l}_{\beta\zeta}(\theta_0;Z)[\zeta(\cdot,\beta)-\zeta_0(\cdot,\beta_0)]$$
$$+ \{\ddot{l}_{\beta\beta}(\tilde{\theta};Z)(\beta-\beta_0) - \ddot{l}_{\beta\beta}(\theta_0;Z)(\beta-\beta_0)\}$$
$$+ \{\ddot{l}_{\beta\zeta}(\tilde{\theta};Z)[\zeta(\cdot,\beta)-\zeta_0(\cdot,\beta_0)] - \ddot{l}_{\beta\zeta}(\theta_0;Z)[\zeta(\cdot,\beta)-\zeta_0(\cdot,\beta_0)]\},$$

where $\tilde{\theta} = (\tilde{\beta}, \tilde{\zeta}(\cdot, \tilde{\beta}))$ is a midpoint between θ_0 and θ. So

$$
\begin{aligned}
&P\{\dot{l}_\beta(\theta; Z) - \dot{l}_\beta(\theta_0; Z) - \ddot{l}_{\beta\beta}(\theta_0; Z)(\beta - \beta_0) - \ddot{l}_{\beta\zeta}(\theta_0; Z)[\zeta(\cdot, \beta) - \zeta_0(\cdot, \beta_0)]\} \\
&\quad = P\{\ddot{l}_{\beta\beta}(\tilde{\theta}; Z) - \ddot{l}_{\beta\beta}(\theta_0; Z)\}(\beta - \beta_0) \\
&\qquad + P\{\ddot{l}_{\beta\zeta}(\tilde{\theta}; Z)[\zeta(\cdot, \beta) - \zeta_0(\cdot, \beta_0)] - \ddot{l}_{\beta\zeta}(\theta_0; Z)[\zeta(\cdot, \beta) - \zeta_0(\cdot, \beta_0)]\}
\end{aligned}
$$

Then by direct calculation we have

$$
\begin{aligned}
&P|\ddot{l}_{\beta\beta}(\tilde{\theta}; Z) - \ddot{l}_{\beta\beta}(\theta_0; Z)| \\
&\quad \le P|XX'\Delta\{\dddot{\tilde{g}}(\epsilon_{\tilde{\beta}}) - \dddot{g}_0(\epsilon_0)\}| \\
&\qquad + P\left\{XX'\left|\int_a^b 1(\epsilon_0 \ge s)1(\tau_0 \le s)\{\exp\{\tilde{g}(s_{\tilde{\beta}})\}\ddot{\tilde{g}}(s_{\tilde{\beta}}) - \exp\{g_0(s)\}\ddot{g}_0(s)\}ds\right.\right. \\
&\qquad\qquad \left.\left. + \int_a^b 1(\epsilon_0 \ge s)1(\tau_0 \le s)\left\{\exp\{\tilde{g}(s_{\tilde{\beta}})\}\dot{\tilde{g}}^2(s_{\tilde{\beta}}) - \exp\{g_0(s)\}\dot{g}_0^2(s)\right\}ds\right|\right\} \\
&\quad \lesssim P|\Delta\{\dddot{\tilde{g}}(\epsilon_{\tilde{\beta}}) - \dddot{g}_0(\epsilon_0)\}| \\
&\qquad + P\left\{\int_a^b |\exp\{\tilde{g}(s_{\tilde{\beta}})\}\ddot{\tilde{g}}(s_{\tilde{\beta}}) - \exp\{g_0(s)\}\ddot{g}_0(s)|ds\right\} \\
&\qquad + P\left\{\int_a^b |\exp\{\tilde{g}(s_{\tilde{\beta}})\}\dot{\tilde{g}}^2(s_{\tilde{\beta}}) - \exp\{g_0(s)\}\dot{g}_0^2(s)|ds\right\} \\
&\quad = C_1 + C_2 + C_3.
\end{aligned}
$$

By applying a similar argument that we used before for verifying (A5) and condition (A.6), we can show

$$
\begin{aligned}
C_1 &\lesssim |\beta - \beta_0| + \|\ddot{g}(\psi(\cdot, \beta)) - \ddot{g}_0(\psi(\cdot, \beta_0))\|_2 \\
&= O(n^{-\xi}) + O\left\{n^{-\min((p-2)\nu, (1-\nu)/2)}\right\}.
\end{aligned}
$$

Similarly, we can show

$$\begin{aligned} C_2 &\lesssim |\beta - \beta_0| + \|\ddot{g}(\psi(\cdot, \beta)) - \ddot{g}_0(\psi(\cdot, \beta_0))\|_2 \\ &= O(n^{-\xi}) + O\left\{n^{-\min((p-2)\nu, (1-\nu)/2)}\right\} \end{aligned}$$

and

$$\begin{aligned} C_3 &\lesssim |\beta - \beta_0| + \|\dot{g}(\psi(\cdot, \beta)) - \dot{g}_0(\psi(\cdot, \beta_0))\|_2 \\ &= O(n^{-\xi}) + O\left\{n^{-\min((p-2)\nu, (1-\nu)/2)}\right\} \end{aligned}$$

where $\xi = \min(p\nu, (1-\nu)/2)$. Therefore,

$$P|\ddot{l}_{\beta\beta}(\tilde{\theta}; Z) - \ddot{l}_{\beta\beta}(\theta_0; Z)| = O\left\{n^{-\min((p-2)\nu, (1-\nu)/2)}\right\}$$

and thus

$$\begin{aligned} &P|\ddot{l}_{\beta\beta}(\tilde{\theta}; Z) - \ddot{l}_{\beta\beta}(\theta_0; Z)|(\beta - \beta_0) \\ &\quad = O\left\{n^{-\min((p-2)\nu, (1-\nu)/2)}\right\} \cdot O\left\{n^{-\min(p\nu, (1-\nu)/2)}\right\} \\ &\quad = O\left\{n^{-\min(2(p-1)\nu, 1/2+(p-5/2)\nu, 1-\nu)}\right\} \\ &\quad = o(n^{-1/2}), \end{aligned}$$

where the last equality holds since $p \geq 3$ so $2(p-1)\nu > \frac{p-1}{p+1} \geq \frac{1}{2}, \frac{1}{2} + (p - \frac{5}{2})\nu > \frac{1}{2}$, and $1 - \nu > \frac{1}{2}$. Similarly, we can show

$$\begin{aligned} &P|\ddot{l}_{\beta\zeta}(\tilde{\theta}; Z)[\zeta(\cdot, \beta) - \zeta_0(\cdot, \beta_0)] - \ddot{l}_{\beta\zeta}(\theta_0; Z)[\zeta(\cdot, \beta) - \zeta_0(\cdot, \beta_0)]| \\ &\quad = O\left\{n^{-\min(2(p-1)\nu, 1/2+(p-5/2)\nu, 1-\nu)}\right\} \\ &\quad = o(n^{-1/2}). \end{aligned}$$

Therefore, we have

$$
\begin{aligned}
&|P\{\dot{l}_\beta(\theta;Z) - \dot{l}_\beta(\theta_0;Z) - \ddot{l}_{\beta\beta}(\theta_0;Z)(\beta-\beta_0) - \ddot{l}_{\beta\zeta}(\tilde{\theta};Z)[\zeta(\cdot,\beta)-\zeta_0(\cdot,\beta_0)]\}| \\
&\quad = O\left\{n^{-\min(2(p-1)\nu,1/2+(p-5/2)\nu,1-\nu)}\right\} \\
&\quad = O(n^{-\alpha\xi}),
\end{aligned}
$$

where $\alpha = \min(2(p-1)\nu, \frac{1}{2}+(p-\frac{5}{2})\nu, 1-\nu)/\min(p\nu, \frac{1-\nu}{2}) > 1$ and $\alpha\xi > \frac{1}{2}$.

Therefore, we have verified all six assumptions, and thus we have

$$
\sqrt{n}(\hat{\beta}_n - \beta_0) = A^{-1}\sqrt{n}\mathbb{P}_n l^*_{\beta_0}(\beta_0, \zeta_0(\cdot,\beta_0); Z) + o_p(1) \to N(0, A^{-1}B(A^{-1})'),
$$

where $l^*_{\beta_0}(\theta_0;Z) = \dot{l}_\beta(\theta_0;Z) - \dot{l}_\zeta(\theta_0;Z)[\mathbf{h}^*]$ is the efficient score function for β_0 and $A = P\{l^*_{\beta_0}(Y,\Delta,X,T)\}^{\otimes 2} = I(\beta_0)$, which is shown when verifying (A3). Hence, $A = B$ and $A^{-1}B(A^{-1})' = A^{-1} = I^{-1}(\beta_0)$, and

$$
\sqrt{n}\mathbb{P}_n l^*_{\beta_0}(\theta_0;Z) = n^{-1/2}\sum_{i=1}^{n} l^*_{\beta_0}(Y_i,\Delta_i,X_i,T_i).
$$

Thus we complete the proof of Theorem 2.2. □

C.5 Proof of Theorem 2.3

Define $\mathbf{w}^*_n(s) = -\dot{\hat{g}}_n(s)\bar{X}(s;\hat{\beta}_n)$. Then we have

$$
l^*_{\hat{\beta}_n}(Y,\Delta,X,T) = \dot{l}_\beta(\hat{\theta}_n;Z) - \dot{l}_\zeta(\hat{\theta}_n;Z)[\mathbf{h}^*_n].
$$

Define

$$
\begin{aligned}
I^{jk}(\beta_0) &= P\left[\left\{\dot{l}_{\beta_j}(\theta_0; Z) - \dot{l}_{\zeta}(\theta_0; Z)[h_j^*]\right\}\right. \\
&\qquad \left.\times \left\{\dot{l}_{\beta_k}(\theta_0; Z) - \dot{l}_{\zeta}(\theta_0; Z)[h_k^*]\right\}\right] = PA^{jk}(\theta_0; Z), \\
\hat{I}_n^{jk}(\hat{\beta}_n) &= \mathbb{P}_n\left[\left\{\dot{l}_{\beta_j}(\hat{\theta}_n; Z) - \dot{l}_{\zeta}(\hat{\theta}_n; Z)[h_j^*]\right\}\right. \\
&\qquad \left.\times \left\{\dot{l}_{\beta_k}(\hat{\theta}_n; Z) - \dot{l}_{\zeta}(\hat{\theta}_n; Z)[h_{k,n}^*]\right\}\right] = \mathbb{P}_n A_n^{jk}(\hat{\theta}_n; Z),
\end{aligned}
$$

where h_j^* is defined in Lemma C.4. We will prove $\mathbb{P}_n A_n^{jk}(\hat{\theta}_n; Z) \to PA^{jk}(\theta_0; Z)$ in probability for all $j, k = 1, \ldots, d$. Let

$$
\begin{aligned}
\mathbb{P}_n A_n^{jk}(\hat{\theta}_n; Z) - PA^{jk}(\theta_0; Z) &= (\mathbb{P}_n - P)A_n^{jk}(\hat{\theta}_n; Z) \\
&\qquad + P\left\{A_n^{jk}(\hat{\theta}_n; Z) - A^{jk}(\theta_0; Z)\right\} \\
&= I_{1n} + I_{2n}
\end{aligned}
$$

For I_{1n} we first define the class of functions

$$
\begin{aligned}
\mathcal{F}_{n,j}^{\beta,\zeta}(\eta) = \Big\{ &\dot{l}_{\beta_j}(\theta; z) - \dot{l}_{\zeta}(\theta; z)[h_j] \colon \theta \in \Theta_n^p, d(\theta, \theta_0) \le \eta, h_j \in \mathcal{H}_n^2 \\
&\|\dot{g}(\psi(\cdot, \beta)) - \dot{g}_0(\psi(\cdot, \beta_0))\|_2 \le \eta, \|h_j - h_j^*\|_\infty \le \eta \Big\}.
\end{aligned}
$$

Then by Lemmas C.5 and C.6 we have

$$
N_{[]}(\varepsilon, \mathcal{F}_{n,j}^{\beta,\zeta}, \|\cdot\|_\infty) \lesssim (1/\varepsilon)^{cq_n + d}
$$

for some constant $c > 0$. This is because for any function $\dot{l}_{\beta_j}(\theta; z) - \dot{l}_{\zeta}(\theta; z)[h_j] \in \mathcal{F}_{n,j}^{\beta,\zeta}(\eta)$, it

can be written as

$$
\begin{aligned}
&\dot{l}_{\beta_j}(\theta; z) - \dot{l}_{\zeta}(\theta; z)[h_j] \\
&\quad = \left\{\dot{l}_{\beta_j}(\theta; z) - \dot{l}_{\beta_j}(\theta_0; z)\right\} - \left\{\dot{l}_{\zeta}(\theta; z)[h_j^*] - \dot{l}_{\zeta}(\theta_0; z)[h_j^*]\right\} \\
&\qquad\quad + \left\{\dot{l}_{\zeta}(\theta; z)[h_j^*] - \dot{l}_{\zeta}(\theta; z)[h_j]\right\} + \left\{\dot{l}_{\beta_j}(\theta_0; z) - \dot{l}_{\zeta}(\theta_0; z)[h_j^*]\right\} \\
&\quad = A_1 + A_2 + A_3 + A_4,
\end{aligned}
$$

where $A_1 \in \mathcal{F}_{n,j}^{\beta}(\eta)$ and $A_2 \in \mathcal{F}_{n,j}^{\zeta}(\eta)$ defined in Lemma C.6, $A_3 \in \mathcal{F}_{n}^{j}(\eta)$ defined in Lemma C.5, and A_4 is a fixed function (the efficient score function).

Assume $A_m^L \le A_m \le A_m^U$ with $\|A_m^U - A_m^L\|_\infty \lesssim \varepsilon, m = 1, 2, 3$. Then $A_m^L + A_{m'}^L \le A_m + A_{m'} \le A_m^U + A_{m'}^U$ with $\|(A_m^U + A_{m'}^U) - (A_m^L A_{m'}^L)\|_\infty \le \|A_m^U - A_m^L\|_\infty + \|A_{m'}^U - A_{m'}^L\|_\infty \lesssim \varepsilon$. Therefore the ε-bracketing number associated with $\|\cdot\|_\infty$ for $\mathcal{F}_{n,j}^{\beta,\zeta}(\eta)$ is also bounded by $(\eta/\varepsilon)^{cq_n+\alpha}$.

We next define the class of functions

$$
\begin{aligned}
\mathcal{F}_{n,jk}^{\beta,\zeta}(\eta) = \Big\{ &(\dot{l}_{\beta_j}(\theta; z) - \dot{l}_{\zeta}(\theta; z)[h_j])(\dot{l}_{\zeta}(\theta; z)[h_k]) \colon \theta \in \Theta_n^p, \\
&h_j, h_k \in \mathcal{H}_n^{p-1}, d(\theta, \theta_0) \le \eta, \|\dot{g}(\psi(\cdot, \beta)) - \dot{g}_0(\psi(\cdot, \beta_0))\|_2 \le \eta, \\
&\|h_j - h_j^*\|_\infty \le \eta, \|h_k - h_k^*\|_\infty \le \eta \Big\}.
\end{aligned}
$$

Then if $B_j^L \le \dot{l}_{\beta_j}(\theta; z) - \dot{l}_{\zeta}(\theta; z)[h_j] \le B_j^U$ and $B_k^L \le \dot{l}_{\beta_k}(\theta; z) - \dot{l}_{\zeta}(\theta; z)[h_k] \le B_k^U$ with $\|B_j^U - B_j^L\|_\infty \le \varepsilon$ and $\|B_k^U - B_k^L\|_\infty \le \varepsilon$, we have $B_j^* B_k^* \le (\dot{l}_{\beta_j}(\theta; z) - \dot{l}_{\zeta}(\theta; z)[h_j])(\dot{l}_{\beta_k}(\theta; z) - \dot{l}_{\zeta}(\theta; z)[h_k]) \le B_j^{**} B_k^{**}$, where B_j^*, B_j^{**} take values of either B_j^L or B_j^U, and the same for

B_k^*, B_k^{**}. Thus

$$\begin{aligned}\|B_j^{**}B_k^{**} - B_j^*B_k^*\|_\infty &= \|(B_j^{**} - B_j^*)B_k^{**} + (B_k^{**} - B_k^*)B_j^*\|_\infty \\ &= \|B_j^{**} - B_j^*\|_\infty\|B_k^{**}\|_\infty + \|B_k^{**} - B_k^*\|_\infty\|B_j^*\|_\infty \\ &\lesssim \|B_j^U - B_j^L\|_\infty + \|B_k^U - B_k^L\|_\infty \\ &\lesssim \varepsilon,\end{aligned}$$

which yields

$$N_{[]}(\varepsilon, \mathcal{F}_{n,jk}^{\beta,\zeta}, \|\cdot\|_\infty) \lesssim (1/\varepsilon)^{cq_n+d}$$

for some constant $c > 0$.

Finally, similar to the verification of Assumption (A4) in the proof of Theorem 2.2 and together with the following fact:

$$\begin{aligned}&\|h_j^*(\cdot,\beta_0) - h_{j,n}^*(\cdot,\hat{\beta}_n)\|_\infty \\ &\quad= \|\dot{g}_0(s)E[X \mid \tau_0 \le t \le \epsilon_0] - \dot{\hat{g}}_n(s_{\tilde{\beta}_n};\hat{\beta}_n)\|_\infty \\ &\quad\le \|\dot{g}_0(s)E[X \mid \tau_0 \le s \le \epsilon_0] - \dot{\hat{g}}_n(s)\bar{X}(s;\hat{\beta}_n)\|_\infty \\ &\qquad+\|\dot{\hat{g}}_n(s_{\hat{\beta}_n})\bar{X}(s_{\hat{\beta}_n};\hat{\beta}_n) - \dot{\hat{g}}_n(s_{\hat{\beta}_n})\bar{X}(s;\hat{\beta}_n)\|_\infty \\ &\qquad+ \|\dot{\hat{g}}_n(s_{\hat{\beta}_n})\bar{X}(s;\hat{\beta}_n) - \dot{\hat{g}}_n(s)\bar{X}(s;\hat{\beta}_n)\|_\infty,\end{aligned} \tag{C.7}$$

where the first term on the right-hand side of inequality (C.7) is

$$\begin{aligned}
&\|\dot{g}_0(s)E[X \mid \tau_0 \le s \le \epsilon_0] - \dot{\hat{g}}_n(s)\bar{X}(s;\hat{\beta}_n)\|_\infty \\
&\quad \le \|\dot{g}_0(s) - \dot{\hat{g}}_n(s)\|_\infty \|E[X \mid \tau_0 \le s \le \epsilon_0]\|_\infty \\
&\qquad\qquad + \|E[X \mid \tau_0 \le s \le \epsilon_0] - \bar{X}(s;\hat{\beta}_n)\|_\infty \|\dot{\hat{g}}(s)\|_\infty \\
&\quad = O_p(n^{-2\nu}) + O_p(n^{-1/2}) = O_p(n^{-2/nu})
\end{aligned}$$

by Lemma C.4 and Corollary 6.21 in Schumaker [2007] for the first term and straightforward argument using empirical process theory for Donsker classes for the second term, together with the boundedness of $\|E[X \mid \tau_0 \le s \le \epsilon_0]\|_\infty$ and $\|\dot{\hat{g}}_n(s)\|_\infty$, and it is straightforward to see that the remaining two terms on the right-hand side of inequality (C.7) is $O_p(n^{-1/2})$. Thus, we have $I_{1n} = o_p(1)$.

That $I_{2n} = o_p(1)$ can be argued directly by the dominated convergence theorem. We now have proved the theorem.

Appendix D

Proof of Theorem 3.1

D.1 Regularity Conditions

Let $Y^* = \log(Y/(1-Y))$ and $L^* = \log\left(\frac{L}{T} \Big/ \left(1 - \frac{L}{T}\right)\right)$. Note that $\epsilon = Y^* - X^{(2)}\omega_0$ and define $\tau = L^* - X^{(2)}\omega_0$. Let λ be the hazard function of ϵ and Λ be the corresponding cumulative hazard function. Following Lai and Ying [1991b] we assume ϵ is independent of (L, Z), which implies that Y is independent of L given Z, an assumption we need for obtaining the joint density function of (Y, Z, L) given $Y > L/T$. To obtain desirable asymptotic properties of the proposed estimator, we introduce the same regularity conditions in Ding and Nan [2011a] with a few additional assumptions for the left-truncation as described in Matthews and Nan [2026], with the conditions for censoring removed.

(C1) (a) The true parameter β_0 belongs to the interior of a compact set $\mathcal{B} \subseteq R^{d_1}$; (b) the true parameter ω_0 belongs to the interior of a compact set $\mathcal{W} \subseteq R^{d_2}$.

(C2) (a) The covariate Z takes values in a bounded subset $\mathcal{Z} \subseteq R^d$; (b) $E(ZZ^T)$ is non-singular.

(C3) (a) There is a constant $b < \infty$ such that, for some constant δ, $\text{pr}(\epsilon > b \mid Z) \geq \delta > 0$ almost surely with respect to the probability measure of Z. This implies that $\Lambda(b) \leq -\log\delta < \infty$; (b) We also assume $\text{pr}(\tau < b \mid Z) \geq 1 - \delta_1$ for some constant $\delta_1 \in (0,1)$ almost surely, and $\text{pr}(\tau > \epsilon \mid Z) \leq \delta_2$ for some constant $\delta_2 \in [0,1)$ almost surely.

(C4) The error ϵ's density f and its derivative $\dot{f}$ are bounded and

$$\int \left\{\dot{f}(s)/f(s)\right\}^2 f(s)ds < \infty.$$

(C5) (a) The conditional density of L^* given Z and its derivative are uniformly bounded for all possible values of $Z \in \mathcal{Z}$, that is,

$$\sup_{z\in\mathcal{Z}} f_{L^*|Z}(l \mid z) \leq K_1, \qquad \sup_{z\in\mathcal{Z}} \left|\dot{f}_{L^*|Z}(l \mid z)\right| \leq K_2$$

for all $l \in \mathcal{L}$ with some constants $K_1, K_2 > 0$; (c) $f_{L^*,Z|Y^*>L^*}$ is free of parameters ω and λ.

(C6) Let $\mathcal{G}^p$ denote the collection of bounded functions g on $[a,b]$ with bounded derivatives $g^{(j)}, j = 1,\ldots,k$. The kth derivative $g^{(k)}$ satisfies the following Lipschitz continuity condition:

$$\left|g^{(k)}(s) - g^{(k)}(t)\right| \leq K_3|s-t|^m \qquad \text{for } s,t \in [a,b],$$

where k is a positive integer and $m \in (0,1]$ such that $p = k + m \geq 3$, $K_3 < \infty$ is an unknown constant, $a = \inf_{y^*,x}(y^* - x^T\beta_0) \vee \inf_{l^*,x}(l^* - x^T\beta_0) > -\infty$ and b is defined in condition (A.3). The true log hazard function of the error ϵ, $g_0(\cdot) = \log\lambda(\cdot)$, belongs to $\mathcal{G}^p$.

(C7) For some $\eta \in (0,1)$, $u^T\text{Var}(Z \mid \epsilon)u \geq \eta u^T E(ZZ^T \mid \epsilon)u$ almost surely for all $u \in R^d$.

D.2 Preliminary Lemma and Proof

Lemma D.1. *Let*

$$f(\omega; X, Y, x) = \frac{\exp\{x\omega + Y - X\omega\}}{1 + \exp\{x\omega + Y - X\omega\}} = \frac{\exp\left\{Y + \sum_{j=1}^{d}(x_j - X_j)\omega_j\right\}}{1 + \exp\left\{Y + \sum_{j=1}^{d}(x_j - X_j)\omega_j\right\}}$$

and define $\mathcal{F} = \{f(\omega; X, Y, x) \colon \omega \in \mathcal{W}; X, x \in \mathcal{X}; Y \in \mathcal{Y}\}$. *Under conditions (C1)-(C2),* $\mathcal{F}$ *is a Donsker class.*

Proof. Consider first just the numerator:

$$f_1(\omega; X, Y, x) = \exp\left\{Y + \sum_{j=1}^{d}(x_j - X_j)\omega_j\right\} = \exp\{Y\}\prod_{j=1}^{d}\exp\{(x_j - X_j)\omega_j\}$$

We will start by showing that f_1 is Donsker. By Condition (A1) and (A2) we have that $|x_j - X_j| \leq K_1$ and $|\omega_j| \leq K_2$ for all $j = 1, \ldots, d$ and some $K_1, K_2 \leq \infty$. Thus it follows that $\phi_j(\omega_j; X_j, x_j) = \exp\{(x_j - X_j)\omega_j\}$ is uniformly bounded. Note that $\phi_j(\omega_j; X_j, x_j)$ is also monotone in $\omega_{n,j}$ and so by Theorem 2.7.9 in van der Vaart [2023] we have that ϕ_j is Donsker for all $j = 1, \ldots, d$. Then by Theorem 2.10.8 and example 2.10.10 in van der Vaart [2023] we have that $\zeta(\omega; X, x) = \prod_{j=1}^{d}\phi_j(\omega_j; X_j, x_j) = \prod_{j=1}^{d}\exp\{(x_j - X_j)\omega_j\}$ is Donsker. By the same theorems we have that $f_1(\omega; X, Y, x)$ is Donsker since $\exp\{Y\}$ is bounded. Thus the numerator is Donsker.

The denominator is simply one plus the numerator, which also forms a Donsker class. Note that the denominator is bounded away from 0. Thus again by Theorem 2.10.8 of van der Vaart [2023] we have that $\mathcal{F}$ is Donsker. □

D.3 Sketch Proof of Theorem 3.1

Through maximum likelihood theory and the continuous mapping theorem, we have

$$\hat{\mu}_{x^{(1)}} = \frac{\exp\{\hat{\beta}_n^T x^{(1)}\}}{1+\exp\{\hat{\beta}_n^T x^{(1)}\}} \to \frac{\exp\{\beta_0^T x^{(1)}\}}{1+\exp\{\beta_0^T x^{(1)}\}} = \mu_{x^{(1)}}$$

in probability. Using the asymptotic properties of $\hat{\omega}_m$ as derived in Matthews and Nan [2026] and applying a similar technique as in Section 3 of Lai and Ying [1991a], we have

$$\hat{\theta} = \int_{\hat{a}}^{\hat{b}} \frac{\exp\{\hat{\omega}_m^T x^{(2)} + u\}}{1+\exp\{\hat{\omega}_m^T x^{(2)} + u\}} \, d\hat{F}_{m,\hat{\omega}_m}(u) \to E\left[Y \mid X^{(2)} = x^{(2)}, \Delta_S = 1, a^* < Y < b^*\right] = \theta$$

in probability. The two main differences between our work and that of Lai and Ying [1991a] is the "trimmed" nature of the mean estimation and the additional left-truncation indicator function. The former simplifies the proof by helping control the tail probabilities, and the latter does not bring much complication due to its nice Donsker properties (see Lemma D.1). Consistency of $\hat{R}^*$ follows through the application of Slutsky's theorem.

Standard maximum likelihood theory also gives the asymptotic linearity that leads to the asymptotic normality of $\hat{\beta}_n$, which further leads to the asymptotic linearity and the asymptotic normality given below:

$$\sqrt{n}(\hat{\mu}_{x^{(1)}} - \mu_{x^{(1)}}) = \sqrt{n}\left(\frac{\exp\{\hat{\beta}_n^T x^{(1)}\}}{1+\exp\{\hat{\beta}_n^T x^{(1)}\}} - \frac{\exp\{\beta_0^T x^{(1)}\}}{1+\exp\{\beta_0^T x^{(1)}\}}\right) \to N(0, \sigma_1^2)$$

in distribution with $\sigma_1^2 = \left(\mu_{x^{(1)}}(1-\mu_{x^{(1)}})\right)^2 \left(x^{(1)}\right)^T I^{-1}(\beta_0) x^{(1)}$, where $I(\beta_0)$ is the information matrix.

For the estimator of the conditional expectation portion, we can write

$$
\begin{aligned}
&\sqrt{n}\left(\int_{\hat{a}}^{\hat{b}} \frac{\exp\{\hat{\omega}_m^T x^{(2)}+u\}}{1+\exp\{\hat{\omega}_m^T x^{(2)}+u\}}\, d\hat{F}_{m,\hat{\omega}_m}(u) - \int_a^b \frac{\exp\{\omega_0^T x^{(2)}+u\}}{1+\exp\{\omega_0^T x^{(2)}+u\}}\, dF(u)\right) \\
&\quad = \sqrt{n}\left(\int_{\hat{a}}^{\hat{b}} \frac{\exp\{\hat{\omega}_m^T x^{(2)}+u\}}{1+\exp\{\hat{\omega}_m^T x^{(2)}+u\}}\, d\hat{F}_{m,\hat{\omega}_m}(u) - \int_{\hat{a}}^{\hat{b}} \frac{\exp\{\omega_0^T x^{(2)}+u\}}{1+\exp\{\omega_0^T x^{(2)}+u\}}\, d\hat{F}_{m,\hat{\omega}_m}(u)\right) \\
&\qquad + \sqrt{n}\left(\int_{\hat{a}}^{\hat{b}} \frac{\exp\{\omega_0^T x^{(2)}+u\}}{1+\exp\{\omega_0^T x^{(2)}+u\}}\, d\hat{F}_{m,\hat{\omega}_m}(u) - \int_{\hat{a}}^{\hat{b}} \frac{\exp\{\omega_0^T x^{(2)}+u\}}{1+\exp\{\omega_0^T x^{(2)}+u\}}\, dF_{\hat{\omega}_m}(u)\right) \\
&\qquad + \sqrt{n}\left(\int_{\hat{a}}^{\hat{b}} \frac{\exp\{\omega_0^T x^{(2)}+u\}}{1+\exp\{\omega_0^T x^{(2)}+u\}}\, dF_{\hat{\omega}_m}(u) - \int_{\hat{a}}^{\hat{b}} \frac{\exp\{\omega_0^T x^{(2)}+u\}}{1+\exp\{\omega_0^T x^{(2)}+u\}}\, dF_{\omega_0}(u)\right) \\
&\qquad + \sqrt{n}\left(\int_{\hat{a}}^{\hat{b}} \frac{\exp\{\omega_0^T x^{(2)}+u\}}{1+\exp\{\omega_0^T x^{(2)}+u\}}\, dF_{\omega_0}(u) - \int_a^b \frac{\exp\{\omega_0^T x^{(2)}+u\}}{1+\exp\{\omega_0^T x^{(2)}+u\}}\, dF_{\omega_0}(u)\right) \\
&\quad = I_1 + I_2 + I_3 + I_4, \qquad \text{(D.1)}
\end{aligned}
$$

where F_ω is the distribution function of $Y^* - \omega^T X^{(2)}$ so $F_{\omega_0} = F$. The asymptotic linearity of I_1 can be shown by applying Theorem 2 of Matthews and Nan [2026] and the delta method. The asymptotic linearity of I_2 and I_3 can be argued in the same way as in the proof of Theorem 2.2.3 of Ding [2010], and the asymptotic linearity of I_4 comes directly from that of $\hat{\omega}_m$. Some modifications to the technical details are made to accomodate left-truncation, but again without much added technical difficulty. We then obtain the influence function for (D.1) that yields the asymptotic normality of our estimator $\hat{\theta}$ of the conditional expectation $\theta = E\left[Y \mid X^{(2)} = x^{(2)}, \Delta_S = 1, a^* < Y < b^*\right]$, i.e., $\sqrt{n}(\hat{\theta} - \theta) \to N(0, \sigma_2^2)$ in distribution.

Since we have obtained asymptotic linearity for both $\sqrt{n}(\hat{\mu}_{x^{(1)}} - \mu_{x^{(1)}})$ and $\sqrt{n}(\hat{\theta} - \theta)$, we know that they have an asymptotic joint bivariate normal distribution with a covariance σ_{12}. We can then apply the bivariate delta method to obtain the asymptotic normality result in Theorem 3.1.

D.4 Detailed Derivation of the Form of the Asymptotic Variance

We work here in the absence of censoring and truncation, and thus recommend fitting the part 1 model through the standard Newton-Raphson algorithm and fitting the part two model via least squares. This will give estimates for β and ω, which we will refer to as β_n and ω_n, respectively. For simplicity and without a loss of generality, assume $X^{(1)} = X^{(2)}$ and that there is only one covariate.

Now, consider some fixed covariate value x, at which we want to evaluate the quantity R (feasible since we assume no left truncation). Define

$$g_n(\beta_n; x) = \frac{\exp\{x\beta_n\}}{1 + \exp\{x\beta_n\}},$$
$$f_m(\omega_m; x) = E_\epsilon \left[\frac{\exp\{x\omega_n + \epsilon\}}{1 + \exp\{x\omega_n + \epsilon\}} \right],$$

where $m = \sum_{i=1}^{n} \Delta_{S,i}$.

Step one is to show that

$$\sqrt{n} \begin{pmatrix} g_n - g \\ f_m - f \end{pmatrix} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \begin{pmatrix} A(\Delta_{S,i}, X_i; x, \beta) \\ B(\Delta_{S,i}, Y_i, X_i; x, \omega) \end{pmatrix} = \mathcal{N}_2(0, \Sigma) + o_p(1).$$

First we derive A. Via a first-order Taylor expansion we obtain

$$\begin{aligned}
g(\beta_n; x) - g(\beta; x) &= \frac{\partial g(\beta; x)}{\partial \beta}(\beta_n - \beta) + O_p\left(n^{-1}\right) \\
&= xg(\beta; x)[1 - g(\beta; x)]I^{-1}(\beta)U(\beta) + o_p\left(n^{-1/2}\right) + O_p\left(n^{-1}\right) \\
&= xg(\beta; x)[1 - g(\beta; x)]I^{-1}(\beta)\frac{1}{n}\sum_{i=1}^{n} X_i\left(\Delta_{S,i} - \frac{\exp\{X_i\beta\}}{1 + \exp\{X_i\beta\}}\right) + o_p\left(n^{-1/2}\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} xg(\beta; x)[1 - g(\beta; x)]I^{-1}(\beta)X_i\left(\Delta_{S,i} - \frac{\exp\{X_i\beta\}}{1 + \exp\{X_i\beta\}}\right) + o_p\left(n^{-1/2}\right) \\
&= \mathbb{P}_n A(\Delta_{S,i}, X_i; x, \beta) + o_p\left(n^{-1/2}\right)
\end{aligned}$$

where the equality on the second line holds via the Newton-Raphson algorithm and

$$A(\Delta_{S,i}, X_i; x, \beta) = xg(\beta; x)[1 - g(\beta; x)]I^{-1}(\beta)X_i\left(\Delta_{S_i} - \frac{\exp\{X_i\beta\}}{1 + \exp\{X_i\beta\}}\right).$$

For B we again use a first-order Taylor expansion to get

$$f_n(\omega_n; x) - f(\omega; x) = \frac{\partial f(\omega; x)}{\partial \omega}(\omega_n - \omega) + O_p(\|\omega_n - \omega\|^2)$$

Under standard regularity conditions we have that

$$\frac{\partial f(\omega; x)}{\partial \omega} = \frac{\partial}{\partial \omega}E_\epsilon\left[\frac{\exp\{x\omega + \epsilon\}}{1 + \exp\{x\omega + \epsilon\}}\right] = E_\epsilon\left[\frac{\partial}{\partial \omega}\frac{\exp\{x\omega + \epsilon\}}{1 + \exp\{x\omega + \epsilon\}}\right]$$

and

$$\begin{aligned}
\omega_n - \omega &= I^{-1}(\omega)U(\omega) + o_p(n^{-1/2}) \\
&= I^{-1}(\omega)\frac{1}{m}\sum_{j=1}^{m} X_j\left(\log\frac{Y_j}{1 - Y_j} - X_j^T\omega\right) + o_p(n^{-1/2}) \\
&= I^{-1}(\omega)\frac{1}{n}\sum_{i=1}^{n}\frac{n}{m}\Delta_{S,i}X_i\left(\log\frac{Y_i}{1 - Y_i} - X_i^T\omega\right) + o_p(n^{-1/2}),
\end{aligned}$$

giving us

$$f_n(\omega_n; x) - f(\omega; x) = E_\epsilon \left[\left(\frac{\exp\{x\omega + \epsilon\}}{1 + \exp\{x\omega + \epsilon\}} \right) \left(1 - \frac{\exp\{x\omega + \epsilon\}}{1 + \exp\{x\omega + \epsilon\}} \right) \right] x^T I^{-1}(\omega) \times \frac{1}{n} \sum_{i=1}^{n} E[\Delta_S]^{-1} \Delta_{S,i} X_i \left(\log \frac{Y_i}{1 - Y_i} - X_i^T \omega \right) + o_p(n^{-1/2})$$

$$= \mathbb{P}_n B(\Delta_{S,i}, Y_i, X_i; x, \omega)$$

where

$$B(\Delta_{S,i}, Y_i, X_i; x, \omega) = E_\epsilon \left[\left(\frac{\exp\{x\omega + \epsilon\}}{1 + \exp\{x\omega + \epsilon\}} \right) \left(1 - \frac{\exp\{x\omega + \epsilon\}}{1 + \exp\{x\omega + \epsilon\}} \right) \right] x^T I^{-1}(\omega) \times E[\Delta_S]^{-1} \Delta_{S,i} X_i \left(\log \frac{Y_i}{1 - Y_i} - X_i^T \omega \right).$$

Thus, we have that

$$\sqrt{n} \begin{pmatrix} g_n - g \\ f_m - f \end{pmatrix} \to \mathcal{N}_2(0, \Sigma)$$

where Σ is a 2×2 matrix, with $\sigma_{11} = P(A - PA)^2 = PA^2$, $\sigma_{22} = P(B - PB)^2 = PB^2$ and $\sigma_{12} = \sigma_{21} = P(A - PA)(B - PB) = PAB$. These equalities hold since $PA = PB = 0$.

Without loss of generality, assume that both models contain an intercept term. We compute

the elements of the variance matrix as follows. First, for σ_{11} we have

$$
\begin{aligned}
\sigma_{11} &= E\left[A(\Delta_{S,i}, X_i; x, \beta)^2\right] \\
&= E\left[\left(xg(\beta; x)[1 - g(\beta; x)]I^{-1}(\beta)X_i\left(\Delta_{S_i} - \frac{\exp\{X_i\beta\}}{1 + \exp\{X_i\beta\}}\right)\right)^2\right] \\
&= E\left[\left(g(\beta; x)[1 - g(\beta; x)]\right)^2 x^T I^{-1}(\beta)X_i\left(\Delta_{S_i} - \frac{\exp\{X_i\beta\}}{1 + \exp\{X_i\beta\}}\right)^2 X_i^T I^{-1}(\beta)x\right] \\
&= \left(g(\beta; x)[1 - g(\beta; x)]\right)^2 x^T I^{-1}(\beta) \\
&\qquad \times E\left\{E\left[X_i\left(\Delta_{S_i} - \frac{\exp\{X_i\beta\}}{1 + \exp\{X_i\beta\}}\right)^2 X_i^T \mid X_i\right]\right\} I^{-1}(\beta)x \\
&= \left(g(\beta; x)[1 - g(\beta; x)]\right)^2 x^T I^{-1}(\beta)E\left[g(\beta; X_i)[1 - g(\beta; X_i)]X_i X_i^T\right] I^{-1}(\beta)x \\
&= \left(g(\beta; x)[1 - g(\beta; x)]\right)^2 x^T I^{-1}(\beta)x.
\end{aligned}
$$

Then for σ_{22} we can compute

$$\begin{aligned}
\sigma_{22} &= E[B(\Delta_{S,i}, Y_i, X_i; x, \omega)^2] \\
&= E\left[\left(E_\epsilon\left[\left(\frac{\exp\{x\omega+\epsilon\}}{1+\exp\{x\omega+\epsilon\}}\right)\left(1-\frac{\exp\{x\omega+\epsilon\}}{1+\exp\{x\omega+\epsilon\}}\right)\right]x^T I^{-1}(\omega)\right.\right. \\
&\qquad\left.\left.\times E[\Delta_S]^{-1}\Delta_{S,i}X_i\left(\log\frac{Y_i}{1-Y_i}-X_i^T\omega\right)\right)^2\right] \\
&= E\left[D(x,\omega_n)^2 x^T I^{-1}(\omega)E[\Delta_S]^{-2}\Delta_{S,i}X_i\epsilon_i^2 X_i^T I^{-1}(\omega)x\right] \\
&= E\left(E\left[D(x,\omega)^2 x^T I^{-1}(\omega)E[\Delta_S]^{-2}\Delta_{S,i}X_i\epsilon_i^2 X_i^T I^{-1}(\omega)x \mid \Delta_{S,i}\right]\right) \\
&= P(\Delta_S=0)\cdot 0 \\
&\qquad + P(\Delta_S=1)E\left[D(x,\omega)^2 x^T I^{-1}(\omega)E[\Delta_S]^{-2}X_i\epsilon_i^2 X_i^T I^{-1}(\omega)x \mid \Delta_S=1\right] \\
&= \frac{1}{E[\Delta_S]}E\left[D(x,\omega)^2 x^T I^{-1}(\omega)X_i\epsilon_i^2 X_i^T I^{-1}(\omega)x \mid \Delta_S=1\right] \\
&= \frac{1}{E[\Delta_S]}E\left[D(x,\omega)^2 x^T I^{-1}(\omega)X_i E[\epsilon_i^2 \mid X_i]X_i^T I^{-1}(\omega)x \mid \Delta_S=1\right] \\
&= \frac{1}{E[\Delta_S]}D(x,\omega)^2 x^T I^{-1}(\omega)\sigma_\epsilon^2 E\left[X_i X_i^T \mid \Delta_S=1\right]I^{-1}(\omega)x \\
&= \frac{1}{E[\Delta_S]}D(x,\omega)^2 x^T I^{-1}(\omega)x,
\end{aligned}$$

where

$$D(x,\omega) = E_\epsilon\left[\left(\frac{\exp\{x\omega+\epsilon\}}{1+\exp\{x\omega+\epsilon\}}\right)\left(1-\frac{\exp\{x\omega+\epsilon\}}{1+\exp\{x\omega+\epsilon\}}\right)\right] \quad \text{and} \quad \sigma_\epsilon^2 = \text{var}(\epsilon).$$

For the cross-terms, we get

$$\begin{aligned}
\sigma_{12} = \sigma_{21} &= E[A(\Delta_{S,i}, X_i; x, \beta)B(\Delta_{S,i}, Y_i, X_i; x, \omega)] \\
&= E\left[x^T g(\beta; x)[1 - g(\beta; x)]I^{-1}(\beta)X_i \left(\Delta_{S,i} - \frac{\exp\{X_i\beta\}}{1 + \exp\{X_i\beta\}}\right) D(x, \omega)\right. \\
&\qquad \left. \times E[\Delta_S]^{-1}\Delta_{S,i}\left(\log\frac{Y_i}{1 - Y_i} - X_i^T\omega\right) X_i^T I^{-1}(\omega)x\right] \\
&= D(x, \omega)g(x; \beta)[1 - g(x; \beta)]x^T I^{-1}(\beta) \\
&\qquad \times E\left[X_i\left(\Delta_{S,i} - \frac{\exp\{X_i\beta\}}{1 + \exp\{X_i\beta\}}\right)\left(\log\frac{Y_i}{1 - Y_i} - X_i^T\omega\right) X_i^T \mid \Delta_S = 1\right] I^{-1}(\omega)x
\end{aligned}$$

Now, we can use the bivariate Delta method to obtain the asymptotic distribution of R. We just derived the asymptotic distribution of (g_n, f_n). Now we define a transformation $\phi = g_n + (1 - g_n)f_n$. We have

$$\nabla\phi = \begin{pmatrix} \frac{\partial\phi}{\partial g_n} \\ \frac{\partial\phi}{\partial f_n} \end{pmatrix} = \begin{pmatrix} 1 - f_n \\ 1 - g_n \end{pmatrix}$$

and so by the multivariate delta method we have

$$\sqrt{n}(\phi_n - \phi) = \sqrt{n}(g_n + (1 - g_n)f_n - g + (1 - g)f) \to \mathcal{N}(0, \nabla\phi^T\Sigma\nabla\phi)$$

in distribution, where

$$\begin{aligned}
\nabla\phi^T\Sigma\nabla\phi &= \begin{pmatrix} 1 - f_n & 1 - g_n \end{pmatrix} \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \begin{pmatrix} 1 - f_n \\ 1 - g_n \end{pmatrix} \\
&= \begin{pmatrix} (1 - f_n)\sigma_{11} + (1 - g_n)\sigma_{21} & (1 - f_n)\sigma_{12} + (1 - g_n)\sigma_{22} \end{pmatrix} \begin{pmatrix} 1 - f_n \\ 1 - g_n \end{pmatrix} \\
&= (1 - f_n)^2\sigma_{11} + 2(1 - f_n)(1 - g_n)\sigma_{21} + (1 - g_n)^2\sigma_{22}.
\end{aligned}$$

Appendix E

Simulation Study for Modeling the Golden Health Index

E.1 Simulation Setup

In order to assess the validity of the proposed methodology and demonstrate its asymptotic properties, we perform a simulation study with the following setup. We first generate left truncation time L from a Uniform(0, 5) distribution and two covariates: X_1 from a Uniform(-3, 3) distribution, and X_2 from a Bernoulli(0.4) distribution. Death time is simulated from the model

$$\log(T) = 1.7 + 0.05X_1 + 0.05X_2 + \zeta, \quad \zeta \sim N(0, 0.3^2),$$

with coefficients chosen to ensure the simulated data approximates the observed death time distribution in the NACC UDS. The healthspan proportion Y conditional on getting the

disease before death, i.e., $\Delta_S = 1$, is simulated from the model

$$\log\left(\frac{Y}{1-Y}\right) = -0.1T + X_1 - X_2 + \epsilon, \quad \epsilon \sim N(0,1). \tag{E.1}$$

The small coefficient of T yields a signal strength that is comparable to those of X_1 and X_2. The underlying disease onset time is then computed as $S = T \times Y$. Left truncation affects the simulated data by only keeping observations where $S > L$ and $T > L$. The setup we have described here causes about 60% of observations to be truncated. Finally, we generate Δ_S from a Bernoulli(μ) distribution with

$$\log\left(\frac{\mu}{1-\mu}\right) = -0.1T - 0.5X_1 + 2X_2 + 0.2L,$$

and set $Y = 1$ if $\Delta_S = 0$.

Simulations are carried out with sample sizes of 400, 800, and 1600. Since left-truncation removes data, we keep simulating independent data until accumulated non-truncated data reach the desired sample size, then treat the sample size as a known quantity. For each sample size, we perform 250 replications and use 300 bootstrap samples in each replication to estimate the variance of the predicted GHI, as suggested in Section 4 of the main text.

As noted in Section 4 of the main text, we need to estimate the trimmed mean due to left-truncation. In the simulation, we choose a^* by adding 1 to A_0, which leads to the lower bound

$$\hat{a} = \log\left(\frac{1/T}{1-1/T}\right) - \hat{\omega}_1 T - \hat{\omega}_2 X_1 - \hat{\omega}_3 X_2.$$

In real life, this can be interpreted as the predicted GHI given that a patient is still alive and disease-free in one year after the baseline age A_0. The upper bound $\hat{b}$ is set to be the largest observed residual for practical convenience.

E.2 Simulation Results

We present a summary of simulation results in Table E.1, where covariate values are chosen to yield GHI values that span a large portion of the interval $(0, 1)$. Biases are negligibly small in most cases, and the variances are approximately inversely proportional to the sample size. The bootstrap variance estimator matches the corresponding empirical variance, and the coverage is close to the nominal value of 95%, especially as the sample size increases.

Table E.1: Results from numerical simulations, sorted by ascending GHI value. Var^1 is the variance estimated from 300 bootstrap replications, and Var^2 is the empirical variance.

T	X_1	X_2	True GHI	n	Bias $\times 10^3$	$\text{Var}^1 \times 10^3$	$\text{Var}^2 \times 10^3$	95% Cov.
10	-1.5	1	0.263	400	17.7	8.6	7.0	92.2
				800	-2.2	3.4	3.1	93.2
				1600	5.8	1.6	1.5	94.0
7	-1.5	1	0.290	400	11.3	6.4	5.6	94.7
				800	-4.8	2.6	2.6	91.6
				1600	4.6	1.2	1.2	94.0
10	0.0	1	0.406	400	9.8	7.0	5.9	94.3
				800	-5.3	2.6	2.6	94.4
				1600	3.5	1.2	1.2	93.6
7	0.0	1	0.420	400	6.4	3.9	3.6	93.1
				800	-6.4	1.4	1.5	91.2
				1600	2.9	0.6	0.7	92.8
7	-1.5	0	0.582	400	5.3	4.5	3.6	95.2
				800	-3.3	2.0	1.8	94.8
				1600	1.1	0.9	1.0	94.0
10	-1.5	0	0.608	400	7.6	7.7	6.8	94.0
				800	-3.1	3.6	3.2	95.2
				1600	-0.1	1.7	1.8	96.4
10	1.5	1	0.642	400	2.9	3.3	3.3	94.7
				800	-4.7	1.5	1.4	94.0
				1600	1.2	0.7	0.7	95.2
7	1.5	1	0.647	400	3.9	1.3	1.4	92.7
				800	-3.3	0.5	0.6	94.4
				1600	1.2	0.3	0.3	92.8
7	0.0	0	0.774	400	2.6	1.0	0.8	98.4

Table E.1: Results from numerical simulations, sorted by ascending GHI value. Var^1 is the variance estimated from 300 bootstrap replications, and Var^2 is the empirical variance.

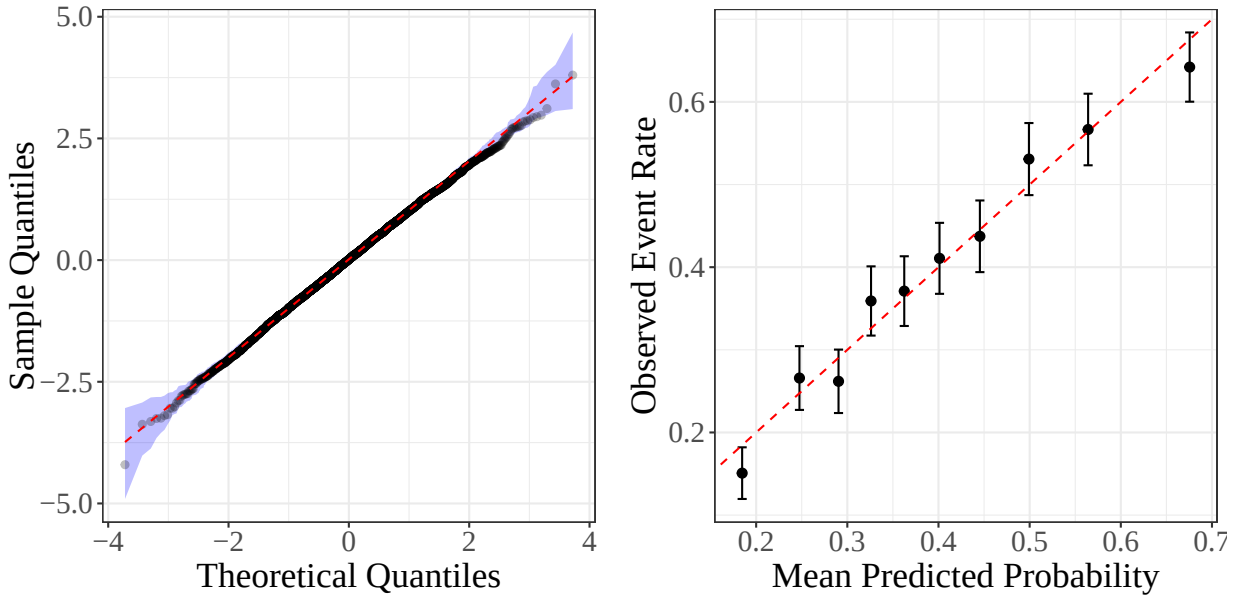
T	X_1	X_2	True GHI	n	Bias $\times 10^3$	$\text{Var}^1 \times 10^3$	$\text{Var}^2 \times 10^3$	95% Cov.
				800	-0.6	0.5	0.4	96.0
				1600	-0.2	0.2	0.3	90.8
10	0.0	0	0.794	400	1.1	2.5	2.4	93.2
				800	-1.9	1.2	1.1	95.2
				1600	-1.2	0.6	0.6	94.8
7	1.5	0	0.925	400	-0.1	0.2	0.2	92.4
				800	0.8	0.1	0.1	91.6
				1600	-0.4	0.1	0.1	94.8
10	1.5	0	0.931	400	-1.5	0.4	0.4	91.6
				800	-0.5	0.2	0.2	95.2
				1600	-0.9	0.1	0.1	95.2

Appendix F

Model Diagnostics and Sensitivity Analyses for Models 3.7 and 3.8

Supporting figures and tables for the information in Section 2.5. For notational ease, we denote the two parts of our modeling strategy by Model (1) for the logistic regression portion and Model (2) for the semiparametric conditional portion.

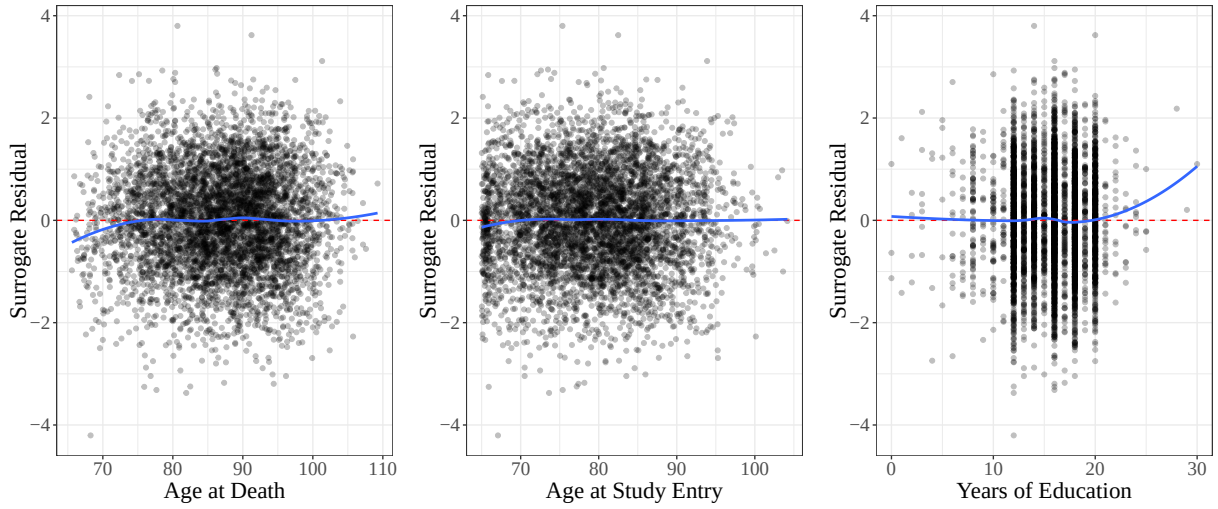
F.1 Model Diagnostics



(a) Surrogate Residual Q-Q Plot

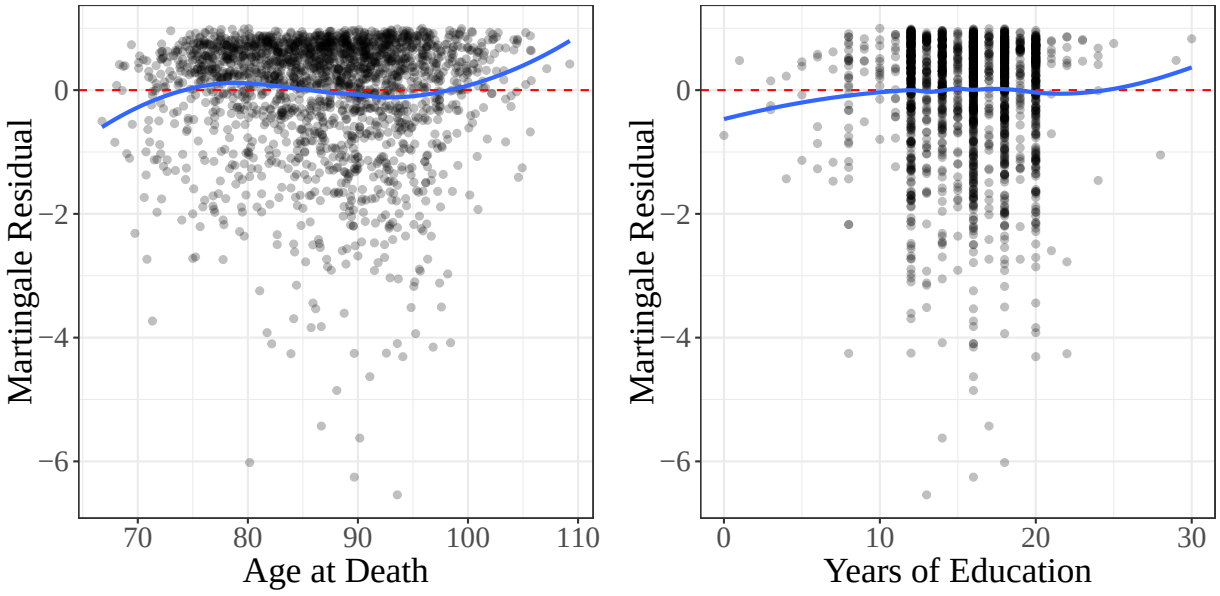
(b) Observed vs. Predicted Probability

Figure F.1: (a) shows a Q-Q plot based on surrogate residuals for Model (1), with the blue band representing the envelope of possible values after 100 simulated sets, the black points representing a single example set of surrogate residuals, and the red dashed line showing the theoretical trend [Liu and Zhang, 2018]. (b) shows the observed event rate against the average predicted probability from Model (1), where the data are split into deciles as in the Hosmer-Lemeshow test; error bars indicate binomial confidence bands [Hosmer and Lemesbow, 1980].



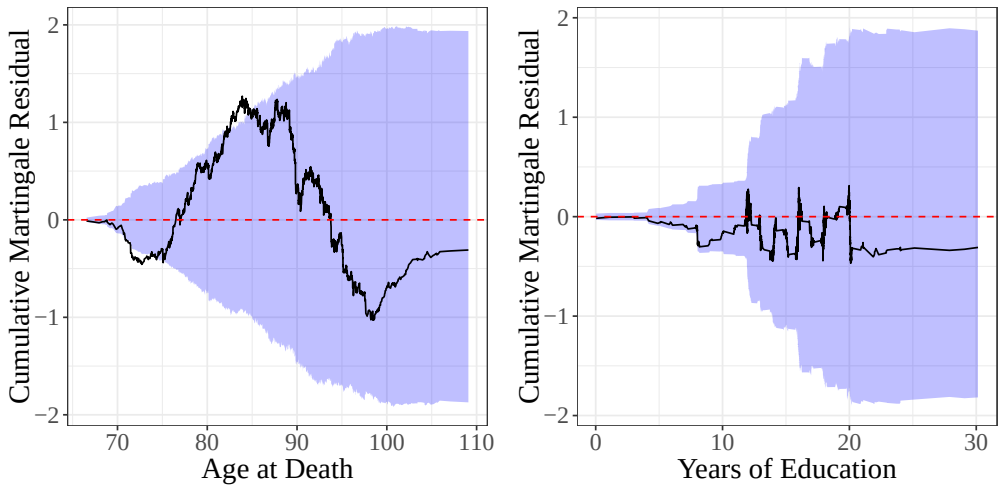
(a) Age at Death (b) Age at Study Entry (c) Years of Education

Figure F.2: Surrogate residual plots for numerical predictors in Model (1), assessing the linear form assumed in the model [Liu and Zhang, 2018]. The red dashed line shows 0, and the blue solid line indicates a smoother fit of the observed points. In all cases, the overall trend is consistent with a linear functional form; for Years of Education, the slight upward trend at high values reflects sparse data in that range.



(a) Age at Death (b) Years of Education

Figure F.3: Martingale residual vs. the numerical predictors in Model (2) [Therneau et al., 1990]. The dashed red line indicates zero, and the blue solid line is the smoothed fit. Variations in the tails of covariate values (where there is sparse data) are expected. These plots exhibit the characteristics we expect for a linear functional form.



(a) Age at Death (b) Years of Education

Figure F.4: Cumulative martingale residual plots for the numerical predictors in Model (2) [Lin et al., 1993] show 95% simulation-based confidence envelopes (1,000 null simulations assuming a zero-mean process or linear functional form). Age at Death exhibits slight departure from the null, but alternative functional forms did not improve model fit.

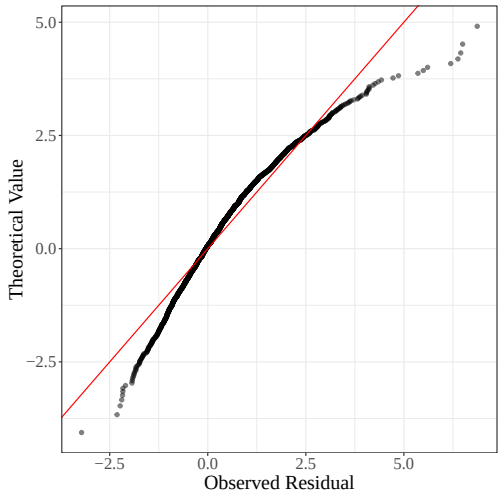


Figure F.5: A Q–Q plot of the observed residuals in Model (2), as compared to a truncated normal assumption. The plot shows significant deviation from the theoretical quantiles and gives evidence for the semiparametric model choice.

F.2 Sensitivity Analyses

Table F.1: Comparison of select demographic variables by ADRC for patients who are included in our analytical sample (N = 5,036).

ADRC	N	Entry Age	Got Dementia	Is Female	Is Married	Is Black	APOE ε4 0	APOE ε4 1	APOE ε4 2
186	2	75.1	50.0%	0.0%	100.0%	0.0%	100.0%	0.0%	0.0%
289	144	74.6	35.4%	42.4%	65.3%	0.0%	62.5%	30.6%	6.9%
354	105	78.2	39.0%	45.7%	63.8%	14.3%	65.7%	27.6%	6.7%
490	141	75.4	44.0%	39.7%	71.6%	15.6%	56.0%	37.6%	6.4%
911	15	76.7	13.3%	66.7%	40.0%	0.0%	73.3%	26.7%	0.0%
943	17	81.8	29.4%	58.8%	58.8%	23.5%	52.9%	47.1%	0.0%
1018	6	76.4	16.7%	50.0%	66.7%	0.0%	66.7%	33.3%	0.0%
1354	177	81.2	43.5%	48.6%	49.2%	3.4%	68.9%	28.2%	2.8%
1416	16	77.2	12.5%	37.5%	87.5%	6.2%	62.5%	37.5%	0.0%
2096	377	77.3	40.8%	46.9%	68.2%	29.2%	52.0%	39.3%	8.8%
2125	28	79.3	32.1%	21.4%	71.4%	10.7%	42.9%	46.4%	10.7%
2289	100	78.7	28.0%	52.0%	48.0%	6.0%	70.0%	25.0%	5.0%
2578	217	80.6	41.5%	47.9%	62.2%	0.0%	64.1%	29.0%	6.9%
2793	8	76.8	50.0%	37.5%	87.5%	0.0%	87.5%	12.5%	0.0%
2958	95	80.8	41.1%	46.3%	56.8%	1.1%	68.4%	27.4%	4.2%
3630	81	78.2	39.5%	42.0%	64.2%	4.9%	66.7%	30.9%	2.5%
3697	20	78.8	25.0%	40.0%	65.0%	5.0%	65.0%	35.0%	0.0%
4032	21	77.6	33.3%	19.0%	66.7%	0.0%	71.4%	28.6%	0.0%
4347	160	80.3	31.2%	66.9%	43.8%	10.0%	74.4%	25.0%	0.6%
4935	89	80.6	43.8%	46.1%	39.3%	14.6%	67.4%	29.2%	3.4%
4967	220	77.0	53.2%	43.6%	69.1%	8.6%	57.3%	34.1%	8.6%
5310	115	80.3	19.1%	55.7%	52.2%	3.5%	75.7%	21.7%	2.6%
5452	202	76.9	43.6%	51.5%	69.8%	4.5%	69.8%	27.7%	2.5%
5783	531	86.6	29.6%	63.7%	38.6%	6.6%	77.0%	21.7%	1.3%
5897	12	74.4	25.0%	33.3%	83.3%	8.3%	66.7%	8.3%	25.0%
6061	98	80.5	34.7%	55.1%	40.8%	1.0%	74.5%	22.4%	3.1%
6499	66	75.7	59.1%	37.9%	69.7%	7.6%	53.0%	40.9%	6.1%
6518	291	78.7	46.4%	51.2%	58.1%	20.3%	62.9%	34.4%	2.7%
6713	66	75.4	24.2%	69.7%	36.4%	84.8%	56.1%	37.9%	6.1%
8354	97	76.1	45.4%	42.3%	68.0%	10.3%	61.9%	29.9%	8.2%
8361	167	81.2	44.3%	55.7%	52.7%	0.6%	64.7%	28.7%	6.6%
8646	379	78.1	49.6%	50.7%	59.4%	12.9%	66.2%	31.1%	2.6%
8658	451	78.7	36.4%	56.8%	60.5%	6.4%	65.2%	31.5%	3.3%
8660	6	77.5	50.0%	66.7%	83.3%	0.0%	16.7%	83.3%	0.0%
8683	200	80.6	44.0%	49.5%	57.0%	14.0%	67.0%	29.5%	3.5%
8974	162	78.0	53.7%	48.8%	64.8%	1.2%	58.6%	33.3%	8.0%

9637	37	77.3	35.1%	32.4%	75.7%	5.4%	67.6%	29.7%	2.7%
9661	117	78.5	35.0%	57.3%	47.0%	16.2%	66.7%	31.6%	1.7%

Table F.2: Comparison of select demographic variables by ADRC for patients who are in our target population in the NACC UDS (N = 30,122). Note that APOE percentages may not add up to 100% due to missing values.

ADRC	N	Entry Age	Got Dementia	Is Female	Is Married	Is Black	APOE $\varepsilon 4$ 0	1	2
186	390	75.0	3.6%	63.6%	64.9%	1.3%	43.8%	11.3%	1.8%
238	58	74.8	12.1%	67.2%	60.3%	1.7%	44.8%	19.0%	6.9%
289	1173	72.4	12.6%	52.9%	65.9%	2.5%	58.7%	24.4%	4.0%
354	977	73.6	14.5%	62.8%	58.0%	29.1%	56.2%	28.5%	5.1%
490	741	73.7	14.4%	56.5%	63.2%	33.6%	55.7%	27.8%	5.3%
740	54	72.2	0.0%	59.3%	68.5%	7.4%	74.1%	20.4%	0.0%
911	529	72.9	8.3%	57.3%	54.1%	21.6%	55.6%	22.1%	2.8%
943	795	72.9	5.4%	66.4%	59.0%	39.0%	55.2%	28.6%	3.5%
1018	610	72.8	5.9%	66.9%	62.5%	21.5%	45.4%	15.9%	1.8%
1354	1079	75.7	15.3%	59.5%	57.4%	4.4%	44.9%	21.6%	3.0%
1416	756	74.4	8.2%	58.7%	64.6%	9.8%	47.5%	19.0%	1.6%
1553	112	73.2	5.4%	53.6%	72.3%	28.6%	49.1%	31.2%	4.5%
2096	1155	74.6	22.9%	51.7%	67.1%	34.3%	41.3%	28.4%	4.8%
2125	280	73.7	18.2%	56.1%	63.2%	16.8%	37.5%	39.6%	12.5%
2289	1001	74.3	13.9%	60.4%	45.4%	14.8%	47.1%	16.9%	2.3%
2578	983	75.6	13.8%	62.9%	64.6%	0.9%	53.6%	22.8%	5.1%
2793	135	72.8	8.1%	49.6%	68.1%	13.3%	49.6%	35.6%	2.2%
2933	51	77.2	2.0%	51.0%	60.8%	13.7%	0.0%	0.0%	0.0%
2958	644	74.7	16.3%	52.3%	62.6%	5.6%	49.2%	25.2%	3.7%
3630	905	72.9	8.8%	62.4%	67.1%	8.0%	39.0%	19.6%	2.4%
3697	366	73.9	14.5%	53.3%	60.4%	12.3%	48.1%	25.4%	5.7%
3964	13	75.7	7.7%	53.8%	76.9%	7.7%	0.0%	0.0%	0.0%
3966	211	75.4	3.8%	70.6%	61.1%	25.1%	0.0%	0.0%	0.0%
4032	434	73.9	6.2%	49.5%	72.1%	1.6%	66.6%	24.9%	5.1%
4347	602	76.7	16.3%	67.9%	47.0%	18.3%	65.6%	23.3%	1.7%
4935	1156	76.5	14.7%	49.5%	44.2%	18.6%	39.7%	17.7%	2.2%
4967	1099	73.3	20.8%	57.5%	68.5%	14.8%	56.7%	28.4%	5.6%
5310	412	75.0	16.5%	59.2%	62.9%	16.7%	62.9%	31.3%	5.8%
5452	1045	74.1	21.7%	61.0%	64.8%	10.8%	51.7%	19.2%	1.6%
5783	1120	82.1	20.9%	65.4%	40.4%	13.7%	61.0%	18.3%	1.4%
5897	383	72.8	14.1%	54.6%	66.8%	17.0%	50.1%	26.9%	5.7%
6061	1297	73.7	9.6%	67.9%	41.4%	14.2%	57.1%	22.7%	2.1%
6499	616	70.4	14.4%	59.4%	68.5%	17.5%	56.3%	32.1%	6.0%
6518	1224	75.5	19.7%	62.2%	52.0%	23.6%	50.7%	26.9%	2.2%
6713	344	73.3	10.2%	82.3%	33.1%	93.0%	44.5%	24.4%	2.0%
8073	97	72.5	2.1%	63.9%	58.8%	57.7%	51.5%	24.7%	4.1%
8354	657	72.5	17.2%	59.7%	65.8%	24.5%	49.2%	30.7%	4.9%
8361	675	75.4	19.4%	57.3%	62.5%	0.9%	64.0%	29.9%	3.7%

8646	1440	73.6	26.7%	56.6%	62.2%	17.8%	55.1%	28.7%	3.1%
8658	1516	74.8	16.8%	61.7%	64.7%	15.7%	56.2%	24.9%	2.9%
8660	135	75.3	23.7%	54.1%	72.6%	11.9%	7.4%	4.4%	0.0%
8683	839	75.4	21.6%	61.1%	58.6%	23.7%	37.2%	18.2%	2.3%
8734	87	71.0	6.9%	59.8%	70.1%	20.7%	13.8%	16.1%	9.2%
8974	665	74.6	22.6%	56.2%	68.6%	1.2%	60.2%	28.0%	4.1%
9637	516	72.0	11.0%	52.7%	73.6%	13.8%	59.9%	30.8%	3.9%
9661	745	73.5	14.6%	60.1%	58.9%	18.8%	54.2%	23.2%	2.6%

Table F.3: Comparison of Model (1) with two mixed effects models. Mixed effects model A includes only a random intercept, while mixed effects model B includes a random intercept and random effects for sex, marital status, comorbidity status, race, and $\varepsilon 4$ alleles.

Variable	Standard GLM		Mixed Effects Model A		Mixed Effects Model B	
	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
Death Age	0.083	(0.069, 0.098)	0.082	(0.067, 0.097)	0.084	(0.068, 0.010)
Yrs of School	-0.013	(-0.033, 0.006)	-0.012	(-0.032, 0.008)	-0.013	(-0.033, 0.007)
Sex						
Male	Referent	–	Referent	–	Referent	–
Female	-0.155	(-0.288, -0.022)	-0.129	(-0.262, 0.004)	-0.139	(-0.288, -0.010)
Is Married						
No	Referent	–	Referent	–	Referent	–
Yes	0.443	(0.305, 0.581)	0.434	(0.294, 0.573)	0.421	(0.241, 0.601)
Comorbidity						
No	Referent	–	Referent	–	Referent	–
Yes	-0.135	(-0.273, 0.002)	-0.159	(-0.298, -0.020)	-0.158	(-0.305, -0.011)
Race						
White	Referent	–	Referent	–	Referent	–
Black	-0.588	(-0.803, -0.374)	-0.611	(-0.834, -0.388)	-0.522	(-0.794, -0.250)
Other	0.234	(-0.170, 0.638)	0.211	(-0.199, 0.621)	0.213	(-0.210, 0.636)
$\varepsilon 4$ Alleles						
Zero	Referent	–	Referent	–	Referent	–
One	0.696	(0.567, 0.825)	0.690	(0.559, 0.821)	0.681	(0.540, 0.822)
Two	1.389	(1.087, 1.691)	1.389	(1.083, 1.695)	1.390	(1.064, 1.715)
Study Entry	-0.077	(-0.093, -0.062)	-0.072	(-0.088, -0.056)	-0.074	(-0.090, -0.058)

Table F.4: Comparison of Model (2) with two mixed effects models, which are fitted under a truncated normal error assumption due to a lack of available methods for fitting random effects in the considered semiparametric model. Since the normal error assumption is incorrect (see Figure S.5), the results are just for showing robustness against center effects and should not be interpreted. Mixed effects model A includes only a random intercept, while mixed effects model B includes a random intercept and random effects for sex, marital status, comorbidity status, race, and $\varepsilon 4$ alleles.

Variable	Standard Trunc. Reg.		Mixed Effects Model A		Mixed Effects Model B	
	Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
Death Age	0.070	(0.051, 0.089)	0.053	(0.033, 0.073)	0.055	(0.035, 0.075)
Yrs of School	-0.039	(-0.083, 0.004)	-0.043	(-0.086, -0.001)	-0.042	(-0.084, 0.001)
Sex						
Male	Referent	–	Referent	–	Referent	–
Female	-0.030	(-0.331, 0.270)	-0.061	(-0.352, 0.229)	-0.066	(-0.358, 0.226)
Is Married						
No	Referent	–	Referent	–	Referent	–
Yes	-0.315	(-0.635, 0.004)	-0.244	(-0.554, 0.066)	-0.340	(-0.707, 0.027)
Comorbidity						
No	Referent	–	Referent	–	Referent	–
Yes	0.172	(-0.147, 0.491)	0.172	(-0.137, 0.480)	0.162	(-0.146, 0.470)
Race						
White	Referent	–	Referent	–	Referent	–
Black	-0.267	(-0.784, 0.250)	-0.141	(-0.649, 0.366)	-0.148	(-0.652, 0.356)
Other	-0.794	(-1.767, 0.180)	-0.675	(-1.616, 0.265)	-0.664	(-1.597, 0.269)
$\varepsilon 4$ Alleles						
Zero	Referent	–	Referent	–	Referent	–
One	-0.643	(-0.936, -0.349)	-0.605	(-0.888, -0.321)	-0.601	(-0.948, -0.254)
Two	-1.018	(-1.600, -0.436)	-0.999	(-1.561, -0.436)	-1.016	(-1.577, -0.455)

Table F.5: Comparison of demographics between included and excluded study participants. Participants in the analytic data set have older age at enrollment, longer follow-up length, and higher dementia incidence. Variations in sex, comorbidity status, and race likely reflect evolving ADRC recruitment strategies and known sex differences in lifespan. No differences are observed in education, marital status, or APOE $\varepsilon 4$ status.

Variable	65+ without Dementia (N = 30,122)	
	Included (N = 5,036)	Excluded (N = 25,086)
Enrollment Age	79.34 (7.61)	73.41 (6.40)
Length of Follow-Up	7.53 (4.29)	3.69 (4.19)
Dementia Observed		
No	3,024 (60.1%)	22,474 (89.6%)
Yes	2,012 (40.0%)	2,612 (10.4%)
Sex		
Female	2,583 (51.3%)	15,363 (61.2%)
Male	2,453 (48.7%)	9,273 (38.8%)
Years of Education	15.60 (3.12)	15.62 (3.23)
Missing	—	*134 (0.5% of total)*
Married		
No	2,140 (42.5%)	10,062 (40.1%)
Yes	2,896 (57.5%)	15,024 (59.9%)
Comorbidity		
No	3,631 (72.1%)	14,754 (58.8%)
Yes	1,405 (27.9%)	10,332 (41.2%)
Race		
White	4,397 (87.3%)	18,942 (75.5%)
Black	531 (10.5%)	4,547 (18.1%)
Other	108 (2.1%)	1,597 (6.4%)
APOE $\varepsilon 4$ Alleles		
None	3,291 (65.4%)	12,198 (65.2%)
One	1,525 (30.3%)	5,695 (30.5%)
Two	220 (4.4%)	802 (4.3%)
Missing	—	*6,391 (25.5% of total)*

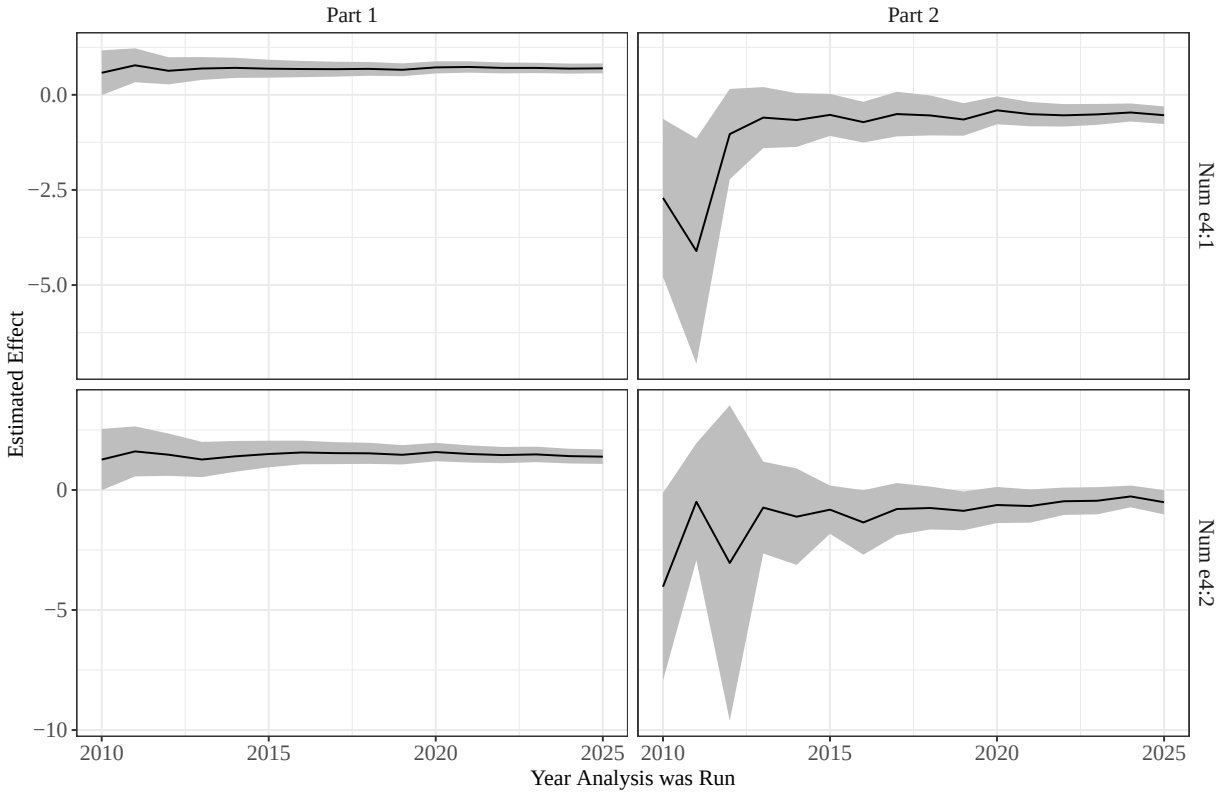


Figure F.6: The evolution of the results of proposed analysis if it had been done at the beginning of each year starting from 2010. Each panel shows the effect of the specified number of APOE $\varepsilon 4$ alleles for the specified model. Unstable estimates for the Part 2 Model in the early 2010s are due to small sample sizes in those subsets.

Table F.6: Model parameter estimates for a temporal robustness analysis. The values presented below would have been the estimates if the analysis had been run in 2015 ($N = 1,854$). Estimates are similar to those presented in the main analysis of the paper, and our main conclusions do not change.

Variable	Part 1 Model			Part 2 Model		
	Estimate	95% CI	P-Value	Estimate	95% CI	P-Value
Death Age	0.290	(0.241, 0.339)	< 0.001	0.055	(0.026, 0.084)	< 0.001
Yrs of School	-0.045	(-0.079, -0.011)	0.009	-0.069	(-0.138, 0.001)	0.051
Sex						
Male	Referent	–	–	Referent	–	–
Female	-0.235	(-0.474, 0.003)	0.053	-0.121	(-0.636, 0.394)	0.645
Is Married						
No	Referent	–	–	Referent	–	–
Yes	0.405	(0.155, 0.656)	0.002	-0.511	(-1.068, 0.046)	0.072
Comorbidity						
No	Referent	–	–	Referent	–	–
Yes	-0.027	(-0.263, 0.207)	0.819	0.375	(-0.103, 0.853)	0.124
Race						
White	Referent	–	–	Referent	–	–
Black	-0.477	(-0.899, -0.074)	0.023	-0.457	(-1.904, 0.988)	0.534
Other	0.168	(-0.749, 0.997)	0.701	0.403	(-1.908, 2.714)	0.732
$\varepsilon 4$ Alleles						
Zero	Referent	–	–	Referent	–	–
One	0.687	(0.451, 0.923)	< 0.001	-0.639	(-1.282, 0.005)	0.052
Two	1.499	(0.952, 2.064)	< 0.001	-1.077	(-2.428, 0.274)	0.118
Study Entry	-0.277	(-0.327, -0.228)	< 0.001	–	–	–

Table F.7: Results from fitting Models (1) and (2) to the left-truncated NACC UDS data using baseline age $A_0 = 65$ and inverse-probability weights based on population distributions of APOE $\varepsilon 4$ allele prevalence, sex, and race. Weights for APOE $\varepsilon 4$ status are obtained from the UK Biobank [Lumsden et al., 2020], stratified on sex and race with their proportions determined from the US Census data. Since APOE $\varepsilon 4$ allele frequencies are only available for white race that is also the majority in the analytic data set (see Table 1 in the main text), we assign the same allele frequencies to other race groups for the sake of sensitivity analysis. Weights are defined as the ratio of the population proportion to the corresponding observed sample proportion. For Model (1), a weighted logistic regression is used, and the variance is estimated using a sandwich estimator. For Model (2), the sieve likelihood approach can computationally support a weighting scheme, and variance is estimated using the bootstrap. The weighted estimates are similar to those from the main analysis.

Parameter	Original Estimate (95% CI)	Weighted Estimate (95% CI)
β_1	0.083 (0.069, 0.098)	0.091 (0.074, 0.110)
β_2	-0.013 (-0.033, 0.006)	-0.016 (-0.043, 0.012)
β_3	-0.155 (-0.288, -0.022)	-0.173 (-0.331, -0.015)
β_4	0.443 (0.305, 0.581)	0.486 (0.320, 0.651)
β_5	-0.135 (-0.273, 0.002)	-0.181 (-0.353, -0.008)
β_6	-0.588 (-0.803, -0.374)	-0.567 (-0.810, -0.325)
β_7	0.234 (-0.17, 0.638)	0.220 (-0.194, 0.634)
β_8	0.696 (0.567, 0.825)	0.618 (0.449, 0.787)
β_9	1.389 (1.087, 1.691)	1.400 (1.035, 1.765)
β_{10}	-0.077 (-0.093, -0.062)	-0.079 (-0.097, -0.062)
ω_1	0.083 (0.070, 0.095)	0.079 (0.066, 0.092)
ω_2	-0.020 (-0.054, 0.013)	-0.069 (-0.103, -0.035)
ω_3	-0.057 (-0.265, 0.150)	-0.129 (-0.354, 0.096)
ω_4	-0.083 (-0.307, 0.141)	-0.094 (-0.334, 0.146)
ω_5	0.316 (0.103, 0.528)	0.301 (0.071, 0.530)
ω_6	-0.194 (-0.548, 0.160)	-0.190 (0.569, 0.188)
ω_7	-0.373 (-1.216, 0.469)	-0.441 (-1.368, 0.486)
ω_8	-0.460 (-0.683, -0.238)	-0.504 (-0.735, -0.273)
ω_9	-0.363 (-0.838, 0.113)	-0.452 (-0.955, 0.052)

Appendix G

Proofs of Lemmas 4.1, 4.2, and 4.3

G.1 Proof of Lemma 4.1

Proof. Part (i). By our definition of the sieve space, $\epsilon_n \to 0$. Choose $\tilde{\theta}_n \in \Theta_n$ with $d(\tilde{\theta}_n, \theta_0) \leq 2\epsilon_n$. Since $\theta_{0,n}$ maximizes S over Θ_n and $\tilde{\theta}_n \in \Theta_n$, we have $S(\theta_{0,n}) \geq S(\tilde{\theta}_n)$.

By continuity of S, $d(\tilde{\theta}_n, \theta_0) \to 0$ implies $S(\tilde{\theta}_n) \to S(\theta_0)$. Since θ_0 maximizes S over $\Theta \supset \Theta_n$:

$$0 \leq S(\theta_0) - S(\theta_{0,n}) \leq S(\theta_0) - S(\tilde{\theta}_n) \to 0.$$

Now suppose for contradiction that $d(\theta_{0,n}, \theta_0) \not\to 0$. Then there exists $\varepsilon > 0$ and a subsequence n_k with $d(\theta_{0,n_k}, \theta_0) \geq \varepsilon$ for all k. By the well-separation condition in (B4), $\sup_{\theta \in \Theta:\, d(\theta,\theta_0) \geq \varepsilon} S(\theta) < S(\theta_0) - \delta$. But $S(\theta_{0,n_k}) \to S(\theta_0)$ along this subsequence, a contradiction. Therefore $d(\theta_{0,n}, \theta_0) \to 0$.

Part (ii). By part (i), $d(\theta_{0,n}, \theta_0) \to 0$. Choose $\tilde{\theta}_n \in \Theta_n$ such that $d(\tilde{\theta}_n, \theta_0) \leq 2\epsilon_n$. Since

$\dot{S}(\theta_0) = 0$ by Remark 4.1, (B1) gives us the following expansion for θ sufficiently close to θ_0:

$$S(\theta_0) - S(\theta) = \frac{1}{2}\{-\ddot{S}(\theta_0)[\theta - \theta_0, \theta - \theta_0]\} + O\{d(\theta, \theta_0)^3\}.$$

By (B5), there exist constants $0 < c_J < C_J < \infty$ such that

$$c_J d(\theta, \theta_0)^2 \leq -\ddot{S}(\theta_0)[\theta - \theta_0, \theta - \theta_0] \leq C_J d(\theta, \theta_0)^2.$$

Applying the lower bound with $\theta = \theta_{0,n}$, together with $d(\theta_{0,n}, \theta_0) \to 0$, yields

$$S(\theta_0) - S(\theta_{0,n}) \geq c_1 d(\theta_{0,n}, \theta_0)^2$$

for all sufficiently large n, for some constant $c_1 > 0$.

Since $\theta_{0,n}$ maximizes S over Θ_n and $\tilde{\theta}_n \in \Theta_n$,

$$S(\theta_0) - S(\theta_{0,n}) \leq S(\theta_0) - S(\tilde{\theta}_n).$$

Applying the upper bound with $\theta = \tilde{\theta}_n$, and using $d(\tilde{\theta}_n, \theta_0) \leq 2\epsilon_n$, gives

$$S(\theta_0) - S(\tilde{\theta}_n) \leq C_1 d(\tilde{\theta}_n, \theta_0)^2 \leq 4C_1\epsilon_n^2$$

for all sufficiently large n, for some constant $C_1 < \infty$.

Combining the preceding displays,

$$c_1 d(\theta_{0,n}, \theta_0)^2 \leq 4C_1\epsilon_n^2.$$

Therefore

$$d(\theta_{0,n}, \theta_0) \leq 2\sqrt{C_1/c_1}\,\epsilon_n = O(\epsilon_n).$$

□

G.2 Proof of Lemma 4.2

Proof. Part (i). To bound $\|J_n^{(n)}(\bar{\theta}_n) - J_n^{(n)}(\theta_n^*)\|_{\text{op}}$, we use that fact that $\|A\|_{\text{op}} \leq \|A\|_F$ for any matrix A, where $\|\cdot\|_F$ is the Frobenius norm. First, we note that by (B1), $J_n^{(n)}(\theta_n)$ is element-wise Lipschitz, meaning,

$$\left[J_n^{(n)}(\bar{\theta}_n) - J_n^{(n)}(\theta_n^*)\right]_{k\ell} \leq L_1\|\bar{\theta}_n - \theta_n^*\|$$

for $k, \ell = 1, \ldots, p_n$ and some constant L_1 that does not depend on n. Now, using the Frobenius norm, we have

$$\begin{aligned}
\|J_n^{(n)}(\bar{\theta}_n) - J_n^{(n)}(\theta_n^*)\|_F &= \sqrt{\sum_{k=1}^{p_n}\sum_{\ell=1}^{p_n}\left|\left[J_n^{(n)}(\bar{\theta}_n)\right]_{k\ell} - \left[J_n^{(n)}(\theta_n^*)\right]_{k\ell}\right|^2} \\
&\leq \sqrt{\sum_{k=1}^{p_n}\sum_{\ell=1}^{p_n} L_1^2\|\bar{\theta}_n - \theta_n^*\|^2} \\
&= p_n L_1\|\bar{\theta}_n - \theta_n^*\| \\
&= O_p(p_n n^{-\xi})
\end{aligned}$$

Since $\|J_n^{(n)}(\bar{\theta}_n) - J_n^{(n)}(\theta_n^*)\|_{\text{op}} \leq \|J_n^{(n)}(\bar{\theta}_n) - J_n^{(n)}(\theta_n^*)\|_F$, the result follows. Since expectation is a linear operator, the element-wise Lipschitz argument passes through expectation and the same argument yields the same rate for $\|J^{(n)}(\bar{\theta}_n) - J^{(n)}(\theta_n^*)\|_{\text{op}}$.

Part (ii). Decompose

$$\begin{aligned}\|J_n^{(n)}(\theta_n^*) - J^{(n)}(\theta_n^*)\|_{\text{op}} \leq \|J_n^{(n)}(\theta_n^*) - J_n^{(n)}(\theta_0)\|_{\text{op}} \\ + \|J^{(n)}(\theta_0) - J^{(n)}(\theta_n^*)\|_{\text{op}} \\ + \|J_n^{(n)}(\theta_0) - J^{(n)}(\theta_0)\|_{\text{op}} \\ = I_1 + I_2 + I_3\end{aligned}$$

By part (i) of this lemma, we have $I_1 = O_p(p_n n^{-\xi})$ and $I_2 = O_p(p_n n^{-\xi})$. For I_3, we define $X_i = n^{-1}\{J_{n,i}^{(n)}(\theta_0) - J^{(n)}(\theta_0)\}$. Thus $\sum_{i=1}^n X_i = J_n^{(n)}(\theta_0) - J^{(n)}(\theta_0)$. By (B1), we know that each element of $J_n^{(n)}(\theta_0) - J^{(n)}(\theta_0)$ is bounded by some constant $M_2 < \infty$. Thus the bounded entry Frobenius argument (similar to above) gives

$$\|J_n^{(n)}(\theta_0) - J^{(n)}(\theta_0)\|_{\text{op}} \leq M_2 p_n \quad \text{and thus} \quad \|X_i\|_{\text{op}} \leq \frac{M_2 p_n}{n}.$$

Similarly, we have

$$\begin{aligned}\left\|\sum_{i=1}^n E\, X_i^2\right\|_{\text{op}} &= \left\|\frac{1}{n^2}\sum_{i=1}^n E\,(J_{n,i}^{(n)}(\theta_0) - J^{(n)}(\theta_0))^2\right\|_{\text{op}} \\ &= \left\|\frac{1}{n} E\,(J_{n,i}^{(n)}(\theta_0) - J^{(n)}(\theta_0))^2\right\|_{\text{op}} \\ &\leq \frac{1}{n} E\,\|J_{n,i}^{(n)}(\theta_0) - J^{(n)}(\theta_0)\|_{\text{op}}^2 \\ &\leq \frac{M_2^2 p_n^2}{n}\end{aligned}$$

Applying Vershynin [2012], Theorem 5.29, we have

$$P\left\{\|J_{n,i}^{(n)}(\theta_0) - J^{(n)}(\theta_0)\|_{\text{op}} \geq t\right\} \leq 2p_n \exp\left(\frac{-t^2/2}{M_2^2 p_n^2/n + M_2 p_n t/n}\right),$$

and choosing $t = M_3 p_n \sqrt{\log p_n / n}$ and plugging in we have

$$P\left\{ \|J_{n,i}^{(n)}(\theta_0) - J^{(n)}(\theta_0)\|_{\text{op}} \geq M_3 p_n \sqrt{\frac{\log p_n}{n}} \right\} \leq 2p_n \exp\left(-K M_3^2 \log p_n\right) \to 0,$$

for a choice of M_3 large enough so that $KM_3^2 > 1$ (note here that K absorbs all other constants). Thus, $I_3 = O_p(p_n \sqrt{(\log p_n)/n})$. Putting the three parts together, we obtain the result.

Part (iii). Define $v_n = (v_{n,\beta}^\top, v_{n,\gamma}^\top)^\top \in \mathbb{R}^{p_n}$. We denote the functional direction defined by v_n in the sieve space as $h_{v_n} = \sum_{j=1}^{k_n} v_{n,\gamma,j} B_j$. By definition, the vector $v_n^\top J^{(n)}(\theta_0) v_n = -\ddot{S}(\theta_0)[(v_{n,\beta}, h_{v_n}), (v_{n,\beta}, h_{v_n})]$. By assumption (B5),

$$c_J \|v_n\|^2 \leq v_n^\top J^{(n)}(\theta_0) v_n \leq C_J \|v_n\|^2.$$

Now, we decompose,

$$v_n^\top J^{(n)}(\theta_n^*) v_n = v_n^\top J^{(n)}(\theta_0) v_n + v_n^\top \left\{ J^{(n)}(\theta_n^*) - J^{(n)}(\theta_0) \right\} v_n,$$

and have

$$\left| v_n^\top \left\{ J^{(n)}(\theta_n^*) - J^{(n)}(\theta_0) \right\} v_n \right| \leq \|J^{(n)}(\theta_n^*) - J^{(n)}(\theta_0)\|_{\text{op}} \|v_n\|^2.$$

By part (i), $\|J^{(n)}(\theta_n^*) - J^{(n)}(\theta_0)\| = O_p(p_n n^{-\xi}) = o_p(1)$, so

$$\left| v_n^\top \left\{ J^{(n)}(\theta_n^*) - J^{(n)}(\theta_0) \right\} v_n \right| = o_p(1) \|v_n\|^2.$$

Combining this with the curvature bound at θ_0, we obtain

$$v_n^\top J^{(n)}(\theta_n^*) v_n \geq c_J \|v_n\|^2 - o_p(1)\|v_n\|^2 = \{c_J - o_p(1)\} \|v_n\|^2$$

and

$$v_n^\top J^{(n)}(\theta_n^*)v_n \leq C_J\|v_n\|^2 + o_p(1)\|v_n\|^2 = \{C_J + o_p(1)\}\|v_n\|^2.$$

By setting $c'_J = c_J/2$ and $C'_J = 2C_J$, the result follows. □

G.3 Proof of Lemma 4.3

Proof. For the first rate, consider the decomposition

$$J_n^{(n)}(\hat{\theta}_n) - J^{(n)}(\theta_{0,n}) = \left[J_n^{(n)}(\hat{\theta}_n) - J_n^{(n)}(\theta_{0,n})\right] + \left[J_n^{(n)}(\theta_{0,n}) - J^{(n)}(\theta_{0,n})\right] = I_1 + I_2$$

By the triangle inequality, $\|\hat{\theta}_n - \theta_{0,n}\| = O_p(n^{-\xi})$. Thus by Lemma 4.2(i), $\|I_1\|_{\text{op}} = O_p(p_n n^{-\xi})$. Since $d(\theta_{0,n}, \theta_0) = O(n^{-\xi})$ by (B3) and Lemma 4.1, we can apply Lemma 4.2(ii) to I_2, giving

$$\begin{aligned}\|J_n^{(n)}(\hat{\theta}_n) - J^{(n)}(\theta_{0,n})\|_{\text{op}} &\leq \|I_1\|_{\text{op}} + \|I_2\|_{\text{op}} \\ &= O_p(p_n n^{-\xi}) + O_p\left(\max\left\{p_n n^{-\xi}, p_n\sqrt{\frac{\log p_n}{n}}\right\}\right) \\ &= o_p(1)\end{aligned}$$

where the last equality holds by the rate conditions on p_n in (B3).

For the second rate, we consider a similar decomposition

$$\hat{V}_n(\hat{\theta}_n) - V_n = \left[\hat{V}_n(\hat{\theta}_n) - \hat{V}_n(\theta_{0,n})\right] + \left[\hat{V}_n(\theta_{0,n}) - V_n\right] = I_3 + I_4$$

First consider I_3. By (B1), we have that the score vector $u_i^{(n)}(\theta)$ is element-wise Lipschitz in the neighborhood of θ_0. Thus, the observation level matrices $\hat{V}_{n,i}(\theta) = u_i^{(n)}(\theta)u_i^{(n)}(\theta)^T$ are

element-wise Lipschitz. We can then use the same strategy from the proof of Lemma 4.2(i) of bounding the Frobenius norm. We omit the details here due to large duplication of the work. In the end, we derive the same result, namely that $\|I_3\|_{\text{op}} = O_p(p_n n^{-\xi})$. Note that in many cases, $\hat{V}_n(\theta) = J_n^{(n)}(\theta)$, so this result can be obtained by direct application of 4.2(i).

We now bound I_4. Applying the same technique as in the proof of Lemma 4.2(ii), we have that the operator norm of $\hat{V}_{n,i}(\theta_{0,n}) = O_p(p_n)$. Applying Theorem 5.29 of Vershynin [2012], we obtain the rate

$$\|I_4\|_{\text{op}} = O_p\left(p_n\sqrt{\frac{\log p_n}{n}}\right).$$

Combining and applying (B3), we get

$$\|\hat{V}_n(\hat{\theta}_n) - V_n\|_{\text{op}} \leq \|I_3\|_{\text{op}} + \|I_4\|_{\text{op}} = O_p(p_n n^{-\xi}) + O_p\left(p_n\sqrt{\frac{\log p_n}{n}}\right) = o_p(1)$$

□