Censored Quantile Regression with Auxiliary Information

by

© **Chithran Vadaverkkot Vasudevan**

A thesis submitted to the School of Gradate Studies in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Department of Mathematics and Statistics

Memorial University

September 2016

St. John's, Newfoundland and Labrador, Canada

Abstract

In Survival analysis, it is vital to understand the effect of the covariates on the survival time. Commonly studied models are the Cox [1972] proportional hazards model and the accelerated failure time model. These methods mainly focus on one characteristic of the survival time. In reality, the association between the response and risk factors is not homogeneous always. This leads to the use of quantile regression [Koenker and Basset, 1978] models, which provide a global description of the association. In quantile regression modeling of the survival data, the problem of estimating the regression coefficients for extreme quantiles can be affected by severe censoring [Portnoy, 2003], especially when the sample size is small. In epidemiological studies, however, there are often times when only a subset of the whole study cohort is accurately observed. The rest of the cohort has only some auxiliary covariate available. The naive use of the auxiliary covariate in the model without the accurately measured covariate could lead to biased estimates. To deal with this problem in censored quantile regression, we propose a regression calibration based method when there is a linear relationship between the auxiliary covariate and the accurately measured covariate. When the relationship is non-linear, we propose a non-parametric kernel smoothing technique. We also propose an empirical likelihood [Owen, 1998, 2001] based weighted censored quantile regression to improve the efficiency of the censored quantile regression estimation by utilizing the auxiliary information about the target population parameters available through scientific facts/previous studies. The proposed estimators are consistent and have asymptotically Gaussian distributions. The efficiency gain compared to the existing methods is remarkable. These methods provide the possibilities of looking into extreme quantiles of the survival distribution. We also applied our proposed methods in real case examples.

I dedicate this work to my family, teachers and friends

Acknowledgements

By the blessings of God, it is my privilege to express my deepest gratitude to my supervisors Dr. Asokan Mulayath Variyath and Dr. Zhao Zhi Fan for their immense support in every way for my PhD program. I am thankful to them for introducing me to the wonderful world of survival analysis, censored quantile regression and empirical likelihood. With their encouragement, endless patience and constant support, I was able to improve my research skills. I was blessed to have them as my supervisors and am grateful for their guidance in the preparation of my thesis. I am also thankful to my supervisory committee members Dr. Hong Wang and Dr. Alwell Oyet for their support and suggestions throughout the program.

I want to sincerely acknowledge the financial support provided by the School of Graduate Studies, the Department of Mathematics and Statistics, the Department of Applied Science, Dr. Variyath and Dr. Fan in the form of a graduate assistantship and teaching assistantships. Further, I express my sincere thanks to the helpful staff at our departmental office for providing necessary facilities.

I am grateful to Dr. P.G. Sankaran, Dr. N. Balakrishna (Cochin University of Science and Technology, India), and Prof. P.K. Venugopalan (Sree Kerala Varma College, Thrissur, India) for referring me to Memorial University of Newfoundland. I take this opportunity to thank Dr. Hong Wang, Dr. Asokan Mulayath Variyath and Dr. J.C. Loredo-Osti for introducing me to various fields in statistics through their courses.

I thank all the graduate students in mathematics and statistics for helping me to complete this program. The calm and quiet atmosphere of St. John's helped me to perform well throughout my program. I am grateful for the support of my friends here in St. John's and also back home.

Last but not least I would like to thank my parents and other family members for their unconditional love and support.

Statement of contribution

This thesis is my joint work with my supervisors Dr. Asokan Mulayath Variyath and Dr. Zhao Zhi Fan.

I received the financial support from school of graduate studies, department of mathematics and statistics, department of applied science and my supervisors during my PhD program.

Table of contents

List of tables

List of figures

Chapter 1

Introduction

Survival analysis deals with the analysis of time to event data. In a survival study, an individual is followed until the occurrence of a specific event from a starting point like date of birth, experimental study entry time, hospital admission, etc. This time interval is known as the failure time (or the time to event).

One common feature of survival data is censoring. During the experimental/study period, a subject's failure time is censored when its follow-up is lost due to some cause. The cause of the censoring must be independent of the event of interest to enable us to perform the standard methods of analysis. There are different kinds of censoring: right censoring, left censoring and interval censoring. In right censoring, the failure times are not observed/followed after a specific time. In left censoring, the failure times are not observed/followed before a specific time. In the case of interval censoring, the failure times are observed/followed only between two specific time points. For a subject, we observe the survival time (Y) as either the censoring time (C) or the failure time (T), whichever occurs first for right censoring. In general, a subject's survival time is right censored if the 'event of interest' for this particular subject did not happen before censoring. The observed data are the triplet $(Y, \delta, \boldsymbol{X})$, where $\delta = \mathbb{I}(T \leq C)$ is the censoring indicator and $\boldsymbol{X}$ is the vector of covariates. Here $\mathbb{I}(\cdot)$ denotes the indicator function.

The survival, $S(t)$ and the hazard, $h(t)$ functions are the two main functions based on which survival analysis is mainly conducted. Let T be a non-negative and continuous failure time with probability density function, $f(t)$ and distribution function, $F(t)$. Then the probability of an individual surviving beyond a specific time t is

$$S(t) = P(T \geq t) = 1 - F(t) = \int_t^{\infty} f(u)du.$$

The hazard function (or hazard rate) is defined as the instantaneous failure rate or as the probability that the failure occurs for a subject in a short period of time, $[t,\ t+\Delta t)$ conditional on the fact that the subject survived until the time t. The hazard function is defined as,

$$h(t) = \lim_{\Delta t \to 0} \frac{P(T \in [t, t+\Delta t) \mid T \geq t)}{\Delta t} = \frac{f(t)}{S(t)} = -\frac{d}{dt} \log S(t).$$

There are parametric, non-parametric and semi-parametric approaches for modeling the survival and the hazard functions. In parametric methods, we assume that the underlying survival distribution is known, up to a few unknown parameters, such as exponential, Weibull and log-normal distributions. Under the parametric model framework, it is common to have the model parameters estimated by the maximum likelihood method.

Without distributional assumptions, survival analysis can be conducted non-parametrically. The Kaplan–Meier estimator [Kaplan and Meier, 1958] also known as the product limit estimator, is a non-parametric estimator of the survival function. If there are no tied event times, the Kaplan–Meier estimator of the survival function is

$$\widehat{S}(y) = \begin{cases} 1, & \text{if } y < y_{(1)} \\ \displaystyle\prod_{y_{(i)} \leq y} \frac{n_i - d_i}{n_i}, & \text{otherwise,} \end{cases} \tag{1.1}$$

where $y_{(i)};\ i = 1, 2, \ldots, m$ are the ordered survival times, n_i is the number of subjects at risk at time $y_{(i)}$ and d_i is the number of events at time $y_{(i)}$.

The Nelson–Aalen estimator is a non-parametric estimator of the cumulative hazard function, $H(y) = -\log S(y) = \int_0^y h(u)du$ [Nelson, 1972; Aalen, 1978],

$$\widehat{H}(y) = \sum_{y_{(i)} \leq y} \frac{d_i}{n_i}. \tag{1.2}$$

The statistical properties of the Nelson–Aalen estimator are discussed in a counting process framework by Fleming and Harrington [2011]. Both of these methods are good for the comparison between two groups of survival data. Other functions (e.g., quantiles of survival time) can also be estimated from the estimated survival function or the hazard function.

Generally, the main interest in survival analysis is to describe the relationship of a factor of interest (e.g., treatment) to the time to event, in the presence of several covariates ($\boldsymbol{X}$), under censoring. Since the survival times are non-negative and often involve censored observations, a standard linear regression model may not be appropriate to explain the relationship. There are a number of models that can be used in analyzing the effects of covariates, such as blood pressure, body temperature, age, weight, etc. over the survival time, Y. Survival models are often partially parameterized, which leads to the so called semi-parametric models. The two most popular semi-parametric models are Cox's proportional hazards and accelerated failure time models.

The proportional hazards (PH) model [Cox, 1972] is widely used in survival analysis to analyze the effect of the explanatory variables on the survival time by modeling the hazard function. The hazard function at time t, conditional on the vector of covariates, $\boldsymbol{X}$, can be modeled using the regression parameters, $\boldsymbol{\beta}$ as

$$h(t \mid \boldsymbol{X}) = h_0(t)\; e^{\boldsymbol{X}^\top \boldsymbol{\beta}},$$

where $h_0(t)$ is the baseline hazard function, the hazard function at $\boldsymbol{X} = 0$. The PH model is semi-parametric with the baseline hazard function completely unspecified. The inferences based on the PH model are asymptotically efficient, but it is difficult to interpret the regression parameters explicitly.

The accelerated failure time (AFT) model is another semi-parametric model which defines a linear relationship between the logarithm of the failure time and the covariates. The AFT model is widely used because of the possibility of interpreting the regression parameters explicitly. It is semi-parametric in the sense that the distribution of the error term in the linear regression model has not been specified.

Let T_i $(i = 1, 2, \ldots, n)$ be the logarithm of failure time of the i^{th} subject and let $\boldsymbol{X}_i$ be the p-vector covariate. The AFT model with the regression parameters, $\boldsymbol{\beta}$ is

$$T_i = \boldsymbol{X}_i{}^\top \boldsymbol{\beta} + \epsilon_i, \tag{1.3}$$

where ϵ_i are the iid error random variables from a distribution function, F, such as normal distribution, extreme value distribution or log logistic distribution.

Inference procedures for the AFT model include the work of Prentice [1978], Buckley and James [1979], Tsiatis [1990], Ritov [1990], Wei, Ying and Lin [1990], among others. These procedures have been derived with F completely unspecified. However, the independent error terms are required to be homogeneous. For more details see Cox and Oakes [1984]; Kalbfleisch and Prentice [2002]; Klein and Moeschberger [2003].

In practice, the distribution of the response, hence the regression model, could vary with different stages of the observation process or when the distribution of the response approaches its boundary, such as in the longevity of the Mediterranean fruit fly study [Carey et al., 1992]. The mortality rate of Mediterranean fruit flies decline at older ages, which is a contradiction to the fact that survival rate generally decreases with age. This phenomenon occurs because there is a shift in the upper tail of the distribution of the survival times of Mediterranean fruit flies. A quantile regression [Koenker and Basset, 1978] model provides an alternative way to investigate this kind of change [Koenker and Geling, 2001]. The quantile regression model assumes that, for a specific $0 < \tau < 1$, the τ^{th} quantile of the random error term is equal to zero. Cox PH and AFT models focus on one characteristic of the survival time. They are not capable of estimating the effect of the covariates over different quantiles of the failure time. In general, all the mean-based regression models are vulnerable to outliers. But the quantile regression models are not only robust to the outliers, they are also robust to misspecification of the error distribution, heteroscedasticity, scale transform of the variables, etc. [Koenker, 2005].

1.1 Quantile Regression

In quantile regression, the conditional quantiles of the response variable for a given set of predictor variables are modeled. The regression parameters are estimated by minimizing a check loss function at a specific quantile, τ, instead of the square loss function as in the standard linear regression.

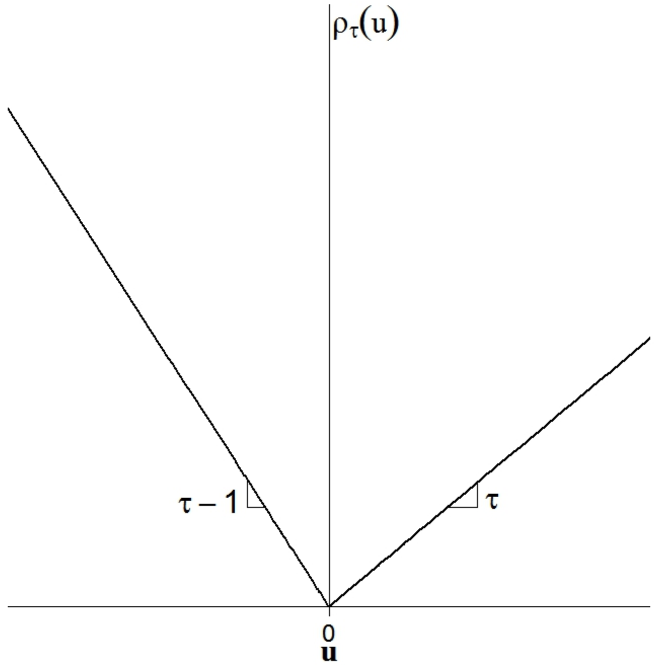

Figure 1.1: Check-loss function, $\rho_\tau(u) = u[\tau - \mathbb{I}(u < 0)]$

A quantile regression model based on properly selected quantiles could provide a global assessment of the covariate effects on the response, which is often ignored by the standard linear regression model, such as the model for the plant self-thinning phenomenon [Cade and Guo, 2000].

For a given response random variable, Y, the τ^{th} quantile can be defined as

$$Q_Y(\tau \mid \boldsymbol{X} = \boldsymbol{x}) = \inf\{y : P(Y \leq y \mid \boldsymbol{X} = \boldsymbol{x}) \geq \tau\},$$

where $\boldsymbol{X}$ is the vector of explanatory variables. Consider a linear conditional quantile function, say $Q_Y(\tau \mid \boldsymbol{X} = \boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{\beta}(\tau)$. Let $h(\cdot)$ be a monotonically non-decreasing function, then we have

$$Q_{h(Y)}(\tau \mid \boldsymbol{x}) = h\left(Q_Y(\tau \mid \boldsymbol{x})\right).$$

This equivariance property of the conditional quantile function allows us to tackle the model parameter interpretation issues involved with variable transformations.

The covariate effect at the τ^{th} quantile of the response can be estimated as the minimizer of an objective function, say

$$\widehat{\boldsymbol{\beta}}(\tau) = \underset{\beta \in \Re^p}{\arg\min} \sum_{i=1}^{n} \rho_\tau \left(y_i - \boldsymbol{x}_i{}^\top \boldsymbol{\beta} \right) \tag{1.4}$$

where $\rho_\tau(u) = u[\tau - \mathbb{I}(u < 0)]$, is the check loss function.

The minimization problem in (1.4) can be solved by using a linear programming algorithm [Koenker, 2005]. However, if we consider the quantile regression model for the survival data, the inferences of the covariate effect over the survival time become more complicated due to censoring.

1.1.1 Censored Quantile Regression

Recently, censored quantile regression has been studied extensively. Powell [1984] introduced the least absolute deviation (LAD) estimator, also called the median regression model for the left censored survival data, using the censored Tobit model [Tobin, 1958]. Powell [1986] generalized the LAD estimation to any quantile. Consider the linear latent variable model with the regression parameters, $\boldsymbol{\beta}$,

$$T_i = \boldsymbol{X}_i{}^\top \boldsymbol{\beta} + u_i,$$

where T_i is the latent variable (Not directly completely observed) and u_i's are assumed to be iid error random variables with distribution function, F. Powell [1984, 1986] considered a case when all left censoring values $C_i,\ i = 1, 2, \ldots n$ are observed (fixed censoring). For the observed survival time, $Y_i = \max(T_i, C_i)$, the covariate vector, $\boldsymbol{X}_i$ and for the τ^{th} $(0 < \tau < 1)$ quantile, the linear conditional quantile function is

$$Q_{T_i}(\tau \mid \boldsymbol{X}_i = \boldsymbol{x}_i) = F^{-1}(\tau) + \boldsymbol{x}_i^\top \boldsymbol{\beta}(\tau),$$

and we can estimate $\boldsymbol{\beta}$ at the τ^{th} quantile as

$$\widehat{\boldsymbol{\beta}}(\tau) = \underset{\beta \in \Re^p}{\arg\min} \sum_{i=1}^{n} \rho_\tau \Big(Y_i - \max\big\{ C_i, \boldsymbol{X}_i{}^\top \boldsymbol{\beta} \big\} \Big), \tag{1.5}$$

where $\rho_\tau(u) = u[\tau - \mathbb{I}(u < 0)]$ is the check loss function. The LAD estimator by Powell [1984] is a special case of (1.5) when $\tau = 1/2$. Chernozhukov and Hong [2002] developed a three-step censored quantile regression under left censoring with a separation restriction on the censoring probability.

Let T_i be the logarithm of the failure times, C_i the logarithm of right censoring time and let $Y_i = \min(T_i, C_i)$ be the logarithm of the survival time for the i^{th} subject. Define an event indicator, $\delta_i = \mathbb{I}(T_i \leq C_i)$ ($\delta_i = 1$, if the event has occurred for the i^{th} subject and $\delta_i = 0$, when the failure time is censored for the i^{th} subject). We assume that conditional on the p-vector covariate, $\boldsymbol{X}_i$, C_i is independent of T_i. If we relax the assumptions on F and the iid assumption on the errors, the conditional quantile regression model for the τ^{th} $(0 < \tau < 1)$ quantile is

$$Q_{T_i \,|\, \boldsymbol{X}_i = \boldsymbol{x}_i}(\tau \mid \boldsymbol{x}_i) = {\boldsymbol{x}_i}^\top \boldsymbol{\beta}(\tau), \tag{1.6}$$

with the assumption that $F^{-1}(\tau) = 0$. Then for a given value of τ, the censored quantile regression parameter, $\boldsymbol{\beta}(\tau)$ can be obtained as,

$$\widehat{\boldsymbol{\beta}}(\tau) = \underset{\beta_\tau \in \Re^p}{\arg\min} \sum_{i=1}^{n} \rho_\tau \Big(Y_i - \min\big\{ C_i, {\boldsymbol{X}_i}^\top \boldsymbol{\beta} \big\} \Big), \tag{1.7}$$

where $\rho_\tau(u) = u[\tau - \mathbb{I}(u < 0)]$, is the check loss function.

Newey and Powell [1990] introduced the optimally weighted censored LAD (CLAD) estimators under fixed censoring. Honore, Khan and Powell [2002] proposed a method which extends the censored quantile regression estimator under fixed censoring to the models with random censoring using the Kaplan-Meier estimator in (1.1). They applied this methodology to Powell [1984, 1986] estimators. Portnoy [2003] introduced a censored quantile regression model under random censoring as a generalization of the Kaplan-Meier estimator recursively using the Kaplan-Meier estimator. Peng and Huang [2008] developed a censored quantile regression model based on the Nelson-Aalen estimator in (1.2), using counting processes and martingale theory. The methods of Powell [1986], Portnoy [2003] and Peng and Huang [2008] are implemented in the "quantreg" package with statistical software SAS and R.

Recently, Ying, Jung and Wei [1995] introduced a semi-parametric inference procedures for median regression models with censored observations. Fitzenberger [1997] studied a censored quantile regression model under fixed censoring in more detail with some applications. Lindgren [1997] proposed a method to estimate the parametric quantile function for censored failure times using asymmetric L_1 minimization. Buchinsky and Hahn [1998] developed a censored quantile regression model under fixed censoring by minimizing a globally convex objective function. Yang [1999] introduced two semi-parametric regression estimators for the censored median regression model using weighted empirical hazard and survival functions.

McKeague, Subramanian and Sun [2001] proposed a censored median regression model based on the missing information principle by replacing the LAD estimating equation [Ying et al., 1995] with its estimated conditional expectation. Neocleous, Vanden Branden and Portnoy [2006] corrected the consistency proof of the default "grid" algorithm provided in Portnoy [2003]. Fitzenberger and Winker [2007] developed a new algorithm for estimating the linear censored quantile regression parameters using the heuristic optimization approach based on threshold accepting (TA). Koenker [2008] described three censored quantile regression methods of Powell [1986], Portnoy [2003] and Peng and Huang [2008] with the applications using the "quantreg" package in R software. Yin, Zeng and Li [2008] proposed a class of power-transformed linear censored quantile regression models. Wang and Fygenson [2009] developed a censored quantile regression model for the longitudinal studies. Neocleous and Portnoy [2009] extended the censored quantile regression model of Portnoy [2003] to a partially linear censored quantile regression model by assuming that one or more explanatory variables have a non-linear effect on the response. Wang and Wang [2009] developed a locally weighted censored quantile regression model to relax the assumptions of global linearity at all quantile levels and unconditional independence between the failure time and the censoring time. Portnoy and Lin [2010] provided the asymptotic distribution theory for the censored quantile regression model of Portnoy [2003].

Wagener, Volgushev and Dette [2012] developed a quantile process under random censoring with the assumption of linearity at all the quantiles and a censored quantile process in sparse regression models. Leng and Tong [2013] generalized the median regression model of Ying et al. [1995] to all the quantiles based on an unbiased estimating equation. Wu and Yin [2013] proposed a mixture cure rate censored quantile regression model with a survival fraction in the population. Yin, Zeng and Li [2014] developed a varying coefficient censored quantile regression model. Chernozhukov, Fernandez-Val and Kowalski [2015] developed a censored quantile instrumental variable estimator which incorporates both the censored quantile regression model of Powell [1986] and endogenous covariates. Yin and Cai [2005] investigated the quantile regression models with clustered or correlated failure time data.

1.2 Auxiliary Information

Covariate measurement error problems in the quantile regression model have attracted growing interest among researchers recently. Due to the financial/time constraints or because of the impracticability of precise measurement, it is very common to carry out studies with surrogate measurements in economics, clinical trials, etc. The use of covariates with measurement error (surrogate variables) for the analysis could lead to significant estimation bias. Carroll et al. [2006] extensively studied measurement error problems in mean-based linear/non-linear regression models.

In quantile regression models, the distribution of the response is not specified and the quantiles do not have the additive property unlike the mean. Because of these problems, it is very difficult to correct the bias in the quantile regression model induced by covariate measurement error. He and Liang [2000] discussed the quantile regression model with covariate measurement error by minimizing the check loss function of orthogonal residuals. Since the error distribution is unknown, they assumed that the errors of both the response and the surrogate variables have a joint spherically symmetric distribution, which leads to consistent estimators. Chesher [2001] considered a small measurement error variance approximation approach, which does not require knowledge of the response distribution. However, it fails to provide consistent estimators, and the computation is difficult under heteroscedasticity.

Angrist, Chernozhukov and Fernandez-Val [2006] developed a quantile regression model by minimizing a weighted sum of squared specification errors when the linearity of conditional quantiles has been misspecified. They also developed a bias formula for the quantile regression when the subset of covariates is not available. This formula enables us to determine the bias from measurement error in the covariates. With the presence of an instrument variable, Schennach [2008] discussed a non-parametric method of quantile regression model identification in the presence of measurement error in the predictors and provided consistent estimators for non-parametric quantile functions. Wei and Carroll [2009] introduced an EM algorithm-type iterative quantile regression model identification by estimating the density of the latent variable conditional on the response and the surrogate variables simultaneously for all the quantile levels when the covariates are measured with error. Montes-Rojas [2011] extended the estimation procedure of Angrist et al. [2006] to classical additive measurement error models. Wang, Stefanski and Zhu [2012] developed a corrected-loss estimator for a particular quantile of interest when there is covariate measurement error. It requires only the assumption of linearity of the quantile function and the knowledge of regression error distribution is not required.

In the case of random censoring, it is often difficult to estimate the regression parameters for extreme quantiles. This makes the covariate measurement error problem in the censored quantile regression model more challenging. Ma and Yin [2011] studied a censored quantile regression model with covariate measurement errors for a range of quantiles rather than a given quantile using composite quantile regression [Zou and Yuan, 2008] for a randomly censored data. Ma and Yin [2011] proposed an objective function based on the inverse censoring probability weights. This method requires the assumption that the errors of both the response and the surrogate variables have a joint spherically symmetric distribution and are independent of the covariates. For a given quantile, Wu, Ma and Yin [2015] introduced the censored quantile regression model with covariate measurement errors under random censoring using a smoothed and corrected estimating equation as an extension of Peng and Huang [2008]'s censored quantile regression model. So far, the literature developed in this area is still limited.

Müller and Keilegom [2014] developed an efficient parametric quantile regression estimator in which the responses may be missing at random and the covariates are always observed completely. They estimated a particular conditional quantile when the auxiliary information such as the parametric models of the mean regression or the variance function regarding that quantile are available.

In some studies, accurately measured variables can be obtained together with the surrogate variables (considered as the 'auxiliary information' in general) for a subcohort. In other cases, a validation sample with accurately measured variables is collected, additional to the auxiliary covariates. In some scenarios, the auxiliary information is available from previous experimental studies/records. These various forms of auxiliary information can be used in the censored quantile regression model to improve the efficiency of the estimators and help us to examine the covariate effect over extreme quantiles of the response as well.

In the application of quantile regression models with the survival data under random right censoring, we encountered scenarios where the estimation of regression parameters corresponding to high quantiles fails, especially when the censoring rate is high. The regression parameters are not identifiable at high quantiles when large failure times are all censored [Portnoy, 2003]. A way to deal with this problem could be the extension of the experimental period until a sufficient number of large failure times is recorded, which may be very expensive in terms of time, cost, etc. In observational studies, however, other options are often required. In large scale epidemiological studies, for example, there could be only a limited number of subjects with some key exposures measured accurately, due to technical, financial, or other limitations. This accurately measured data subset forms a validation sample, while the other data subset of the study cohort has been measured only through auxiliary/surrogate covariates, which are easier and cheaper to observe but can only provide partial information about the key exposures. The latter forms the non-validation sample. If one focuses only on the accurately measured validation sample, the identifiability problem of the regression coefficients at high quantiles might very likely occur because of the relatively small sample size.

For the effective incorporation of the accurately measured variable in the model, the presence of the auxiliary variable in relation to those accurately measured variables can be considered as an asset. Auxiliary information is often available in various forms such as additional covariates, known relationships between some covariates or some established relationship between the covariates and the response available from experience/previous records. In this thesis, we develop methodologies to utilize the auxiliary information to improve the efficiency of censored quantile regression parameter estimation. Our simulation studies reveal that utilizing the auxiliary information in the censored quantile regression model could improve the efficiency of the parameter estimation. It may even enable us to investigate the effect of the explanatory variables on the response's higher quantiles under heavy censoring and reduce the possible loss of information.

1.2.1 Regression Calibration in Censored Quantile Regression

The regression calibration type estimation method was introduced by Prentice [1982] in a failure time regression model with normal covariate measurement errors under rare disease assumptions. They introduced a partial likelihood estimator based on the induced relative risk function after correcting for covariate measurement error. Pepe, Self and Prentice [1989] extended the regression calibration method to parametric settings. Wang et al. [1997] applied the regression calibration method to predict the unavailable variables of interest in the non-validation sample using the validation sample and the surrogate covariates. Yu and Nan [2010] introduced the regression calibration method in a semi-parametric accelerated failure time model. In this thesis, we would like to use the regression calibration methodology in the censored quantile regression model with auxiliary covariates to avoid information loss. This approach is straightforward to implement, but it is challenging to provide the theoretical justification. We considered the classic additive covariate measurement error model as a special case of our model. The proposed methodology is implemented in two steps. In the first step, we predict the unobserved covariate in the non-validation sample using the regression calibration method. For this prediction, we use other explanatory variables and the auxiliary covariate by assuming that they are linearly related to the accurately measured variable. And in the second step, we combine the predicted observations in the non-validation sample with the accurately measured observations in the validation sample and apply them to the censored quantile regression model.

1.2.2 Kernel Smoothing in Censored Quantile Regression

In most scenarios, the form of the relationship between the accurately measured covariate and the auxiliary covariate and the measurement error distribution are unknown. To utilize the auxiliary information effectively, Zhou and Pepe [1995] proposed an estimated partial likelihood method for the censored failure time relative risk regression with categorical auxiliary covariates using the validation sample. Zhou and Wang [2000] extended the idea to handle the continuous auxiliary covariates based on a kernel smoothing method using the validation sample. Fan and Wang [2009] further extended this approach to multivariate correlated failure time. Granville and Fan [2014] used a non-parametric kernel smoothing method instead of a regression calibration method to predict the unobserved observations in the non-validation sample. We propose to apply this non-parametric prediction concept to censored quantile regression models. Similar to the application of the regression calibration procedure discussed in the previous section, the implementation is straightforward but the theoretical justification is more challenging.

1.2.3 Empirical Likelihood based Weighted Censored Quantile Regression

In survey sampling, one often has auxiliary information about the target population from previous surveys or records. The information could be used to improve the efficiency of estimation. See Kuk and Mak [1989]; Rao, Kovar and Mantel [1990]; Chen and Qin [1993], among others. Tang and Leng [2012] introduced an empirical likelihood [Owen, 1998, 2001] approach to quantile regression with auxiliary information. We would like to adapt this idea to the censored quantile regression model to improve the efficiency of the estimator. The incorporation of auxiliary information can be tricky because of the presence of censoring time in the observed survival time. The idea is to convert the auxiliary information into empirical likelihood based data driven probabilities and apply them as the weights in the censored quantile regression model.

1.3 Outline of the Thesis

The remainder of the thesis is organized as follows. In Chapter 2, we propose the regression calibration based approach in the non-validation sample. First, we briefly describe the censored quantile regression model and its estimation procedure from the work of Peng and Huang [2008]. Then we introduce the regression calibration method to estimate the unobserved key exposure measurements using the validation sample and the auxiliary covariate. We present a new estimating equation for the estimation of the regression parameters and investigate their asymptotic properties. In the simulation studies, we compare the performance of our proposed method with the results using the validation sample at different quantile levels. We illustrate our proposed method by analyzing the primary biliary cirrhosis (PBC) data, and also by predicting the unobserved copper content in urine measurements.

In Chapter 3, we introduce a non-parametric kernel based prediction for the unavailable key exposure in a censored quantile regression model. This proposed method does not require linearity or Gaussian assumptions. We develop an estimating equation by considering both validation and non-validation samples and investigate its large sample properties as well. In the simulation studies, we compare the performance of our proposed method with the results using only the validation sample. We applied our proposed method to Colorado plateau uranium miners data by assuming the radon exposure measurements are unavailable. For illustration, we applied our proposed method to PBC data as well.

In Chapter 4, we investigate the censored quantile regression with auxiliary information through a weighted censored quantile regression model. We detail the methodology of estimating the weights using the empirical likelihood for both known and unknown target population parameters. We discuss the estimation procedure of the proposed weighted censored quantile regression model parameters and their asymptotic properties based on Peng and Huang [2008] censored quantile regression model. We perform simulation studies of four different models, with correlated and uncorrelated covariates. In the first numerical study we consider the auxiliary information coming from a known linear relationship between the failure time and the covariates. In the second numerical study, we consider the observed survival time instead of the failure time. In both numerical studies, we compare the performance of our proposed method with the standard censored quantile regression results at various quantiles. We illustrate our proposed method by analyzing the North Central Cancer Treatment Group (NCCTG) lung cancer data as well.

Our overall concluding remarks are provided in Chapter 5. From our simulation studies, we observe that our proposed regression calibration and kernel smoothing based methods have a remarkable efficiency gain compared to using only the validation sample in the censored quantile regression model at all quantile levels. The proposed empirical likelihood based weighted censored quantile regression is more efficient than standard censored quantile regression. We present options for some future work in the following section. We would like to extend our proposed methods to non-continuous variables, a variable selection procedure for censored quantile regression, goodness of fit test etc.

Chapter 2

Regression Calibration in Censored Quantile Regression

2.1 Introduction

For the i^{th} $(i = 1, 2, \ldots, n)$ subject, let T_i be the failure time, C_i the right censoring time, $\boldsymbol{X}_i$ the p-vector covariate, $Y_i = T_i \wedge C_i$ the time of failure or censoring, whichever occurs first, and let the event indicator be $\delta_i = \mathbb{I}(T_i \leq C_i)$, where $\wedge$ is the minimum operator and $\mathbb{I}(\cdot)$ is the indicator function. Then for a given quantile level, τ $(0 < \tau < 1)$, the censored quantile regression model parameter, $\boldsymbol{\beta}(\tau)$ can be estimated as,

$$\widehat{\boldsymbol{\beta}}(\tau) = \underset{\beta \in \Re^p}{\arg\min} \sum_{i=1}^{n} \rho_\tau \Big(Y_i - \min \left\{ C_i, {\boldsymbol{X}_i}^\top \boldsymbol{\beta} \right\} \Big).$$

In the application of censored quantile regression models with survival data, we encountered scenarios where the quantile regression parameters are not identifiable at extreme quantiles due to censoring. We may be able to resolve this identification problem by extending the experimental period until the larger failure times have been recorded. This is however often not feasible in practice.

In large scale epidemiological studies, due to technical, financial or other limitations, there could be only a limited number of subjects with key exposures measured accurately. The remaining subjects in the study cohort have been measured only through the auxiliary covariates, which are easier and cheaper to observe but can only provide partial information about the key exposures. The subjects with exact measurements form the validation sample. The remaining subjects form the non-validation sample. If one focuses only on the accurately measured validation sample, the identifiability problem of the censored quantile regression coefficients at higher quantiles may very likely occur. If we exclude the non-validation sample from the analysis, it leads to a loss of information as well.

In this chapter, we propose a regression calibration based method which accommodates both the validation and non-validation samples in the censored quantile regression model. Using this method, we can efficiently estimate the censored quantile regression coefficients corresponding to some high quantiles, whereas we may fail to estimate the regression parameters when we use only the validation sample. At the same time, the efficiency gain of our proposed method as compared to the method which uses only the validation sample is remarkable.

The rest of this chapter is organized as follows. In Section 2.2, we briefly discuss the censored quantile regression model of Peng and Huang [2008] and the estimation procedure of our proposed method, followed by discussing its asymptotic properties. Results of the simulation studies and application of the proposed method to real data are discussed in Section 2.3 and concluding remarks are presented in Section 2.4.

2.2 Estimation

2.2.1 Censored Quantile Regression Model of Peng and Huang [2008]

Accelerated failure time (AFT) models are a family of semi-parametric models which define a linear relationship between the logarithm of the failure time and the covariates. For the i^{th} $(i = 1, 2, \ldots, n)$ subject, let T_i be the logarithm of the failure time and $\boldsymbol{X}_i$ the p-vector covariate. Then the AFT model with regression parameters, $\boldsymbol{\beta}$ is

$$T_i = {\boldsymbol{X}_i}^T \boldsymbol{\beta} + \epsilon_i,$$

where ϵ_i's are the iid error random variables with a distribution, F, such as normal distribution, extreme value distribution or log logistic distribution.

For the i^{th} observation, let C_i be the logarithm of the right censoring time and let $Y_i = \min(T_i, C_i)$ be the logarithm of the observed survival time. Define an event indicator, $\delta_i = \mathbb{I}(T_i \le C_i)$. In this model, they assumed that the independent censoring mechanism, that is, conditional on $\boldsymbol{X}_i$, C_i is independent of T_i .

Because of the monotonicity of the quantile function, the quantile regression model based on the AFT model is a special case of the more flexible model used by Peng and Huang [2008]:

$$Q_{T_i}(\tau \mid \boldsymbol{X}_i) = e^{\boldsymbol{X}_i^\top \boldsymbol{\beta}(\tau)},$$

for a specific quantile, $\tau \in (0,1)$.

Define the counting process, $\mathbb{N}_i(t) = \mathbb{I}(Y_i \leq t, \delta_i = 1)$. Let $H(u) = -\log(1-u)$ for $0 \leq u < 1$. When the $\boldsymbol{X}_i$'s are available for the entire cohort, the observed data are the triplet $\{Y_i, \boldsymbol{X}_i, \delta_i\}$. Peng and Huang [2008] introduced the censored quantile regression estimator as a generalization of the Nelson-Aalen estimator of the cumulative hazard function of T_i. For a fixed $\tau \in (0,1)$, the regression coefficient, $\boldsymbol{\beta}(\tau)$ can be estimated by solving the estimating equation

$$\sqrt{n}\ S_n(\boldsymbol{\beta}, \tau) = \mathbf{0},$$

where

$$S_n(\boldsymbol{\beta}, \tau) = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i \left(\mathbb{N}_i \left(e^{\boldsymbol{X}_i^\top \boldsymbol{\beta}(\tau)} \right) - \int_0^{\tau} \mathbb{I}\left[Y_i \geq e^{\boldsymbol{X}_i^\top \boldsymbol{\beta}(u)} \right] dH(u) \right). \quad (2.1)$$

Here $\mathbb{N}_i \left(e^{\boldsymbol{X}_i^\top \boldsymbol{\beta}(\tau)} \right) - \int_0^{\tau} \mathbb{I}\left[Y_i \geq e^{\boldsymbol{X}_i^\top \boldsymbol{\beta}(u)} \right] dH(u)$ is a martingale associated with the counting process, $\mathbb{N}_i(t)$. The martingale property ensures that $E\{S_n(\boldsymbol{\beta}_0, \tau)\} = 0$, where $\boldsymbol{\beta}_0(\tau)$ is the true value of the censored quantile regression parameter.

Peng and Huang [2008] suggested a grid-based estimation procedure for $\boldsymbol{\beta}_0(\tau)$ because of the stochastic integral representation of $S_n(\boldsymbol{\beta}, \tau)$. $\widehat{\boldsymbol{\beta}}(\tau)$, the estimator of $\boldsymbol{\beta}_0(\tau)$, is a right-continuous piecewise constant function which jumps only on the grid, $\mathbb{S}_{L(n)} = \{0 = \tau_0 < \tau_1 < \cdots < \tau_{L(n)} = \tau_U < 1\}$. The size of $\mathbb{S}_{L(n)}$ is defined as $\left\|\mathbb{S}_{L(n)}\right\| = \max_k \{\tau_k - \tau_{k-1};\ k = 1, \ldots, L(n)\}$. For simplicity, we will use L for $L(n)$. To obtain $\widehat{\boldsymbol{\beta}}(\tau_k);\ k = 1, \ldots, L$, they proposed a sequential solution to the following monotone estimating equation, which is based on (2.1), for $\boldsymbol{\beta}(\tau_k)$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \boldsymbol{X}_i \left(\mathbb{N}_i \left(e^{\boldsymbol{X}_i^\top \boldsymbol{\beta}(\tau_k)} \right) - \sum_{r=0}^{k-1} \mathbb{I}\left[Y_i \geq e^{\boldsymbol{X}_i^\top \widehat{\boldsymbol{\beta}}(\tau_r)} \right] \{H(\tau_{r+1}) - H(\tau_r)\} \right) = 0. \quad (2.2)$$

They defined the estimators, $\widehat{\boldsymbol{\beta}}(\tau_k)$ as generalized solutions [Fygenson and Ritov, 1994] to equation (2.2), because this equation is not continuous and its solution may not be unique. (Fygenson and Ritov [1994] defined a generalized estimating equation, $W(\boldsymbol{\beta})$, as a monotone nondecreasing field, if for any $\boldsymbol{\beta}$ and $\boldsymbol{\xi}$ in $\mathfrak{R}^p$, $\boldsymbol{\xi}^\top W(\boldsymbol{\beta} + x\boldsymbol{\xi})$ is monotone non-decreasing in the scalar x. Tao [2016] stated that "*Generalized solution to an equation such as $Lu(x) = f(x)$ for all $x \in \Omega$ is to allow for the existence of some singular set $S \subset \Omega$ in which the solution u is allowed to be singular or undefined, but require that u be smooth outside of S (or at least smooth enough that it is clear how to define Lu), and only require that the equation $Lu(x) = f(x)$ be true outside of S. Typically the set S will be closed and suitably "small" (e.g. zero measure, or having positive codimension, or being contained in a finite union of hypersurfaces.)*")

Because of the monotone non-deceasing property of equation (2.2) on $\boldsymbol{\beta}(\tau_k)$ [Peng and Huang, 2008; Koenker, 2008], all the generalized solutions belong to a convex set and the left hand side of equation (2.2) is the gradient of a convex function. Peng and Huang reformulated it to the following L_1-type convex objective function (2.3) to obtain the minimizer which is equivalent to the generalized solution of equation (2.2),

$$\begin{aligned} l_k(\mathfrak{b}) = \sum_{i=1}^{n} \left| \delta_i \log Y_i - \delta_i \mathfrak{b}^\top \boldsymbol{X}_i \right| + \left| R^* - \mathfrak{b}^\top \sum_{i=1}^{n} (-\delta_i \boldsymbol{X}_i) \right| \\ + \left| R^* - \mathfrak{b}^\top \sum_{i=1}^{n} \left(2\boldsymbol{X}_i \sum_{r=0}^{k-1} \mathbb{I} \left[Y_i \geq e^{\boldsymbol{X}_i^\top \widehat{\boldsymbol{\beta}}(\tau_r)} \right] \{H(\tau_{r+1}) - H(\tau_r)\} \right) \right|, \end{aligned} \tag{2.3}$$

where R^* is a very large number and $k = 1, 2, \ldots, L$. The solutions are obtained by minimizing this function using the Barrodale-Roberts algorithm [Barrodale and Roberts, 1974]. This has been converted to a linear programming problem and implemented in the "quantreg" package.

2.2.2 Regression Calibration in Censored Quantile Regression

In many epidemiological or other medical studies, the main exposure, say X_1, is not accurately measured for a subcohort. However, an auxiliary covariate, W, is available for the entire study cohort which is linearly related to the unobserved main exposure variable. For the partially unobserved X_1, we assume the linear regression model:

$$X_1 = W\theta_1 + \boldsymbol{X}_2\boldsymbol{\theta}_2 + \xi = \mathbb{X}\boldsymbol{\theta} + \xi, \tag{2.4}$$

where ξ is the random error and $\boldsymbol{X}_2$ are the explanatory variables which are completely available. Here $\boldsymbol{\theta} = (\theta_1, \boldsymbol{\theta}_2)$ is a p-vector and $\boldsymbol{X}_2$ is a matrix with dimension $n \times (p-1)$.

A special case of (2.4) is the classic measurement error model:

$$W = X_1 + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\varepsilon}$ is the measurement error. The subcohort with all the information available is the validation subset. Assume that X_1 is the covariate subset which is partly available and that $\boldsymbol{X}_2$ is the covariate subset which is completely available in the covariate, $\boldsymbol{X} = (X_1, \boldsymbol{X}_2)$. In a sample (with size, n), we completely observe only $\{W, \boldsymbol{X}_2\}$, where W is the auxiliary variable or the error prone observation of X_1. For a subset of m_v subjects randomly selected from the sample ($m_v < n$), X_1 is also completely available, which form the validation sample. X_1 is not available for the remaining $m_n = n - m_v$ subjects, which form the non-validation sample. The observed data are composed of $\{Y_j, W_j, X_{1j}, \boldsymbol{X}_{2j}, \delta_j\},\ j \in \mathbb{V}$, the validation sample and $\{Y_l, W_l, \boldsymbol{X}_{2l}, \delta_l\},\ l \in \overline{\mathbb{V}}$, the non-validation sample.

Assume that the conditional expectation of X_1 given $\{W, \boldsymbol{X}_2\}$ is a function of W and $\boldsymbol{X}_2$, say $\Phi(W, \boldsymbol{X}_2, \boldsymbol{\theta}) = E(X_1 \,|\, W, \boldsymbol{X}_2)$. We propose a regression calibration based approach to predict the unobserved values of X_1 in the non-validation sample. The regression parameter, $\boldsymbol{\theta}$ can be estimated using the validation data.

Let $\Phi_l = \Phi(W_l, \boldsymbol{X}_{2l}, \boldsymbol{\theta})$ and

$$\widehat{\Phi}_l = \widehat{\Phi}(W_l, \boldsymbol{X}_{2l}, \widehat{\boldsymbol{\theta}}) = \mathbb{X}_l^\top \widehat{\boldsymbol{\theta}}, \quad \forall\, l \in \overline{\mathbb{V}}, \tag{2.5}$$

where $\widehat{\boldsymbol{\theta}} = \left(\mathbb{X}_{\mathbb{V}}^\top \mathbb{X}_{\mathbb{V}}\right)^{-1} \mathbb{X}_{\mathbb{V}}^\top X_{1\mathbb{V}}$, is obtained by regressing X_{1j} on $(W_j, \boldsymbol{X}_{2j}),\ j \in \mathbb{V}$. Here $\mathbb{X}_{\mathbb{V}}$ and $X_{1\mathbb{V}}$ are the validation sample subset of $\mathbb{X}$ and X_1 respectively. The dimension of the $\mathbb{X}_{\mathbb{V}}$ matrix is $m_v \times p$.

Denote $\boldsymbol{Z} = (\Phi, \boldsymbol{X}_2)$ and $\widehat{\boldsymbol{Z}} = (\widehat{\Phi}, \boldsymbol{X}_2)$; then our estimating function is

$$\sqrt{n}\; S_n(\boldsymbol{\beta}, \tau)$$

and $\widehat{\boldsymbol{\beta}}$ is the generalized solution of $\sqrt{n}\; S_n(\boldsymbol{\beta}, \tau) = \mathbf{0}$, where

$$\begin{aligned} S_n(\boldsymbol{\beta}, \tau) = & \frac{\rho_n}{m_v} \sum_{j \in \mathbb{V}} \boldsymbol{X}_j \left\{ \mathbb{N}_j \left(e^{\boldsymbol{X}_j^\top \boldsymbol{\beta}(\tau)} \right) - \int_0^\tau \mathbb{I}\left[Y_j \geq e^{\boldsymbol{X}_j^\top \boldsymbol{\beta}(u)} \right] dH(u) \right\} \\ & + \frac{1 - \rho_n}{m_n} \sum_{l \in \overline{\mathbb{V}}} \widehat{\boldsymbol{Z}}_l \left\{ \mathbb{N}_l \left(e^{\widehat{\boldsymbol{Z}}_l^\top \boldsymbol{\beta}(\tau)} \right) - \int_0^\tau \mathbb{I}\left[Y_l \geq e^{\widehat{\boldsymbol{Z}}_l^\top \boldsymbol{\beta}(u)} \right] dH(u) \right\} \\ = & \rho_n\, \Omega_{m_v}^{\mathbb{V}}(\boldsymbol{\beta}, \tau) + (1 - \rho_n)\, \widehat{\Omega}_{m_n}^{\overline{\mathbb{V}}}(\boldsymbol{\beta}, \tau), \end{aligned} \tag{2.6}$$

where $\rho_n = m_v/n$. The first part on the right hand side of the equation (2.6) comes from the validation sample and the second part is from the non-validation sample. For a particular quantile, τ_k, the estimator of $\boldsymbol{\beta}_0(\tau_k)$ is $\widehat{\boldsymbol{\beta}}(\tau_k)$, which is the generalized solution of $\sqrt{n}\ S_n(\boldsymbol{\beta}, \tau_k) = 0$.

$$\sqrt{n}\ S_n(\widehat{\boldsymbol{\beta}}, \tau_k) = \sqrt{n}\left\{\rho_n\ \Omega^{\mathbb{V}}_{m_v}(\widehat{\boldsymbol{\beta}}, \tau_k) + (1-\rho_n)\ \widehat{\Omega}^{\overline{\mathbb{V}}}_{m_n}(\widehat{\boldsymbol{\beta}}, \tau_k)\right\} + \xi_{n,k}, \tag{2.7}$$

for $k = 1, \ldots, L$. Here by the definition of a generalized solution, $\max_{k=1,2,\ldots,L} \|\xi_{n,k}\| \leq \sup_i \|\boldsymbol{X}_i\|/\sqrt{n}$, and

$$\Omega^{\mathbb{V}}_{m_v}(\widehat{\boldsymbol{\beta}}, \tau_k) = \frac{1}{m_v}\sum_{j\in\mathbb{V}} \boldsymbol{X}_j \left\{\mathbb{N}_j\left(e^{\boldsymbol{X}_j^\top \widehat{\boldsymbol{\beta}}(\tau_k)}\right) - \int_0^{\tau_k} \mathbb{I}\left[Y_j \geq e^{\boldsymbol{X}_j^\top \widehat{\boldsymbol{\beta}}(u)}\right] dH(u)\right\},$$

$$\widehat{\Omega}^{\overline{\mathbb{V}}}_{m_n}(\widehat{\boldsymbol{\beta}}, \tau_k) = \frac{1}{m_n}\sum_{l\in\overline{\mathbb{V}}} \widehat{\boldsymbol{Z}}_l \left\{\mathbb{N}_l\left(e^{\widehat{\boldsymbol{Z}}_l^\top \widehat{\boldsymbol{\beta}}(\tau_k)}\right) - \int_0^{\tau_k} \mathbb{I}\left[Y_l \geq e^{\widehat{\boldsymbol{Z}}_l^\top \widehat{\boldsymbol{\beta}}(u)}\right] dH(u)\right\}.$$

Let $s(\boldsymbol{\beta}, \tau) = E\left\{\rho_n\ \Omega^{\mathbb{V}}_{m_v}(\boldsymbol{\beta}, \tau) + (1-\rho_n)\ \widehat{\Omega}^{\overline{\mathbb{V}}}_{m_n}(\boldsymbol{\beta}, \tau)\right\}$. Define

$$\Omega^{\overline{\mathbb{V}}}_{m_n}(\widehat{\boldsymbol{\beta}}, \tau_k) = \frac{1}{m_n}\sum_{l\in\overline{\mathbb{V}}} \boldsymbol{Z}_l \left\{\mathbb{N}_l\left(e^{\boldsymbol{Z}_l^\top \widehat{\boldsymbol{\beta}}(\tau_k)}\right) - \int_0^{\tau_k} \mathbb{I}\left[Y_l \geq e^{\boldsymbol{Z}_l^\top \widehat{\boldsymbol{\beta}}(u)}\right] dH(u)\right\}.$$

Using the martingale property, we have $E\{\Omega^{\mathbb{V}}_{m_v}(\boldsymbol{\beta}_0, \tau)\} = \mathbf{0}$ and $E\{\Omega^{\overline{\mathbb{V}}}_{m_n}(\boldsymbol{\beta}_0, \tau)\} = \mathbf{0}$. By the equation (2.12) and the martingale property, $s(\boldsymbol{\beta}_0, \tau) = \mathbf{0}$, where $\boldsymbol{\beta}_0(\cdot)$ denotes the true $\boldsymbol{\beta}(\cdot)$. Now we have,

$$\begin{aligned}
&\widehat{\Omega}^{\overline{\mathbb{V}}}_{m_n}(\widehat{\boldsymbol{\beta}}, \tau_k) - \Omega^{\overline{\mathbb{V}}}_{m_n}(\widehat{\boldsymbol{\beta}}, \tau_k) \\
&\qquad = \frac{1}{m_n}\left(\sum_{l\in\overline{\mathbb{V}}} \widehat{\boldsymbol{Z}}_l \left\{\mathbb{N}_l\left(e^{\widehat{\boldsymbol{Z}}_l^\top \widehat{\boldsymbol{\beta}}(\tau_k)}\right) - \int_0^{\tau_k} \mathbb{I}\left[Y_l \geq e^{\widehat{\boldsymbol{Z}}_l^\top \widehat{\boldsymbol{\beta}}(u)}\right] dH(u)\right\}\right. \\
&\qquad\qquad \left. - \sum_{l\in\overline{\mathbb{V}}} \boldsymbol{Z}_l \left\{\mathbb{N}_l\left(e^{\boldsymbol{Z}_l^\top \widehat{\boldsymbol{\beta}}(\tau_k)}\right) - \int_0^{\tau_k} \mathbb{I}\left[Y_l \geq e^{\boldsymbol{Z}_l^\top \widehat{\boldsymbol{\beta}}(u)}\right] dH(u)\right\}\right) \\
&\qquad = \frac{1}{m_n}\left(\sum_{l\in\overline{\mathbb{V}}} \widehat{\boldsymbol{Z}}_l \mathbb{N}_l\left(e^{\widehat{\boldsymbol{Z}}_l^\top \widehat{\boldsymbol{\beta}}(\tau_k)}\right) - \sum_{l\in\overline{\mathbb{V}}} \widehat{\boldsymbol{Z}}_l \int_0^{\tau_k} \mathbb{I}\left[Y_l \geq e^{\widehat{\boldsymbol{Z}}_l^\top \widehat{\boldsymbol{\beta}}(u)}\right] dH(u)\right.
\end{aligned}$$

$$
\begin{aligned}
&\quad - \sum_{l\in\overline{\mathbb{V}}} \boldsymbol{Z}_l \mathbb{N}_l\left(e^{\boldsymbol{Z}_l^\top \widehat{\boldsymbol{\beta}}(\tau_k)}\right) - \sum_{l\in\overline{\mathbb{V}}} \boldsymbol{Z}_l \int_0^{\tau_k} \mathbb{I}\left[Y_l \geq e^{\boldsymbol{Z}_l^\top \widehat{\boldsymbol{\beta}}(u)}\right] dH(u)\Bigg) \\
&= \frac{1}{m_n}\left(\sum_{l\in\overline{\mathbb{V}}}\left\{\widehat{\boldsymbol{Z}}_l \mathbb{N}_l\left(e^{\widehat{\boldsymbol{Z}}_l^\top \widehat{\boldsymbol{\beta}}(\tau_k)}\right) - \boldsymbol{Z}_l \mathbb{N}_l\left(e^{\boldsymbol{Z}_l^\top \widehat{\boldsymbol{\beta}}(\tau_k)}\right)\right\}\right. \\
&\quad \left. - \sum_{l\in\overline{\mathbb{V}}} \int_0^{\tau_k}\left\{\widehat{\boldsymbol{Z}}_l \mathbb{I}\left[Y_l \geq e^{\widehat{\boldsymbol{Z}}_l^\top \widehat{\boldsymbol{\beta}}(u)}\right] - \boldsymbol{Z}_l \mathbb{I}\left[Y_l \geq e^{\boldsymbol{Z}_l^\top \widehat{\boldsymbol{\beta}}(u)}\right]\right\} dH(u)\right) \\
&= \frac{1}{m_n}\sum_{l\in\overline{\mathbb{V}}}\left(\widehat{\boldsymbol{Z}}_l\left[\mathbb{N}_l\left(e^{\widehat{\boldsymbol{Z}}_l^\top \widehat{\boldsymbol{\beta}}(\tau_k)}\right) - \mathbb{N}_l\left(e^{\boldsymbol{Z}_l^\top \widehat{\boldsymbol{\beta}}(\tau_k)}\right)\right]\right. \\
&\quad + \left(\widehat{\boldsymbol{Z}}_l - \boldsymbol{Z}_l\right)\mathbb{N}_l\left(e^{\boldsymbol{Z}_l^\top \widehat{\boldsymbol{\beta}}(\tau_k)}\right) \\
&\quad - \int_0^{\tau_k}\left(\widehat{\boldsymbol{Z}}_l\left\{\mathbb{I}\left[Y_l \geq e^{\widehat{\boldsymbol{Z}}_l^\top \widehat{\boldsymbol{\beta}}(u)}\right] - \mathbb{I}\left[Y_l \geq e^{\boldsymbol{Z}_l^\top \widehat{\boldsymbol{\beta}}(u)}\right]\right\}\right. \\
&\quad \left.\left. - \left(\widehat{\boldsymbol{Z}}_l - \boldsymbol{Z}_l\right)\mathbb{I}\left[Y_l \geq e^{\boldsymbol{Z}_l^\top \widehat{\boldsymbol{\beta}}(u)}\right]\right) dH(u)\right) \\
&= \frac{1}{m_n}\sum_{l\in\overline{\mathbb{V}}}\left(\widehat{\boldsymbol{Z}}_l\left[\mathbb{N}_l\left(e^{\widehat{\boldsymbol{Z}}_l^\top \widehat{\boldsymbol{\beta}}(\tau_k)}\right) - \mathbb{N}_l\left(e^{\boldsymbol{Z}_l^\top \widehat{\boldsymbol{\beta}}(\tau_k)}\right)\right]\right. \\
&\quad - \int_0^{\tau_k}\widehat{\boldsymbol{Z}}_l\left\{\mathbb{I}\left[Y_l \geq e^{\widehat{\boldsymbol{Z}}_l^\top \widehat{\boldsymbol{\beta}}(u)}\right] - \mathbb{I}\left[Y_l \geq e^{\boldsymbol{Z}_l^\top \widehat{\boldsymbol{\beta}}(u)}\right]\right\} dH(u) \\
&\quad + \left(\widehat{\boldsymbol{Z}}_l - \boldsymbol{Z}_l\right)\mathbb{N}_l\left(e^{\boldsymbol{Z}_l^\top \widehat{\boldsymbol{\beta}}(\tau_k)}\right) \\
&\quad \left. - \int_0^{\tau_k}\left(\widehat{\boldsymbol{Z}}_l - \boldsymbol{Z}_l\right)\mathbb{I}\left[Y_l \geq e^{\boldsymbol{Z}_l^\top \widehat{\boldsymbol{\beta}}(u)}\right] dH(u)\right) \\
&= \frac{1}{m_n}\sum_{l\in\overline{\mathbb{V}}}\left(\widehat{\boldsymbol{Z}}_l\left[\mathbb{M}_l\left\{\tau_k, \widehat{\boldsymbol{Z}}_l, \widehat{\boldsymbol{\beta}}(\tau_k)\right\} - \mathbb{M}_l\left\{\tau_k, \boldsymbol{Z}_l, \widehat{\boldsymbol{\beta}}(\tau_k)\right\}\right]\right. \\
&\quad \left. + \left(\widehat{\boldsymbol{Z}}_l - \boldsymbol{Z}_l\right)\mathbb{M}_l\left\{\tau_k, \boldsymbol{Z}_l, \widehat{\boldsymbol{\beta}}(\tau_k)\right\}\right)
\end{aligned}
$$

$$
\begin{aligned}
&\sqrt{n}(1-\rho_n)\left(\widehat{\Omega}_{m_n}^{\overline{\mathbb{V}}}(\widehat{\boldsymbol{\beta}}, \tau_k) - \Omega_{m_n}^{\overline{\mathbb{V}}}(\widehat{\boldsymbol{\beta}}, \tau_k)\right) \\
&\quad = \frac{\sqrt{n}(1-\rho_n)}{m_n}\sum_{l\in\overline{\mathbb{V}}}\left(\widehat{\boldsymbol{Z}}_l\left[\mathbb{M}_l\left\{\tau_k, \widehat{\boldsymbol{Z}}_l, \widehat{\boldsymbol{\beta}}(\tau_k)\right\} - \mathbb{M}_l\left\{\tau_k, \boldsymbol{Z}_l, \widehat{\boldsymbol{\beta}}(\tau_k)\right\}\right]\right.
\end{aligned}
$$

$$+\left(\widehat{\boldsymbol{Z}}_l-\boldsymbol{Z}_l\right)\mathbb{M}_l\left\{\tau_k,\boldsymbol{Z}_l,\widehat{\boldsymbol{\beta}}(\tau_k)\right\}\Bigg), \tag{2.8}$$

where

$$\mathbb{M}_i\left\{\tau_k,\boldsymbol{U}_i,\widehat{\boldsymbol{\beta}}(\tau_k)\right\}=\mathbb{N}_i\left(e^{\boldsymbol{U}_i^\top\widehat{\boldsymbol{\beta}}(\tau_k)}\right)-\int_0^{\tau_k}\mathbb{I}\left[Y_i\geq e^{\boldsymbol{U}_i^\top\widehat{\boldsymbol{\beta}}(u)}\right]dH(u).$$

Using Appendix B (proof of Theorem 2) of [Peng and Huang, 2008, p. 647], since $\sqrt{m_n}\,\|\mathbb{S}_L\|\to 0$, we have

$$\frac{1}{\sqrt{m_n}}\sum_{l\in\overline{\mathbb{V}}}\widehat{\boldsymbol{Z}}_l\ \mathbb{M}_l\{\tau_k,\widehat{\boldsymbol{Z}}_l,\widehat{\boldsymbol{\beta}}(\tau)\}=o_{(0,\tau_U])}(\mathbf{1}),\ \ \text{a.s.}.$$

Using similar arguments as in Appendix B (proof of Theorem 2) of [Peng and Huang, 2008, p. 647] and because of the boundedness of $\widehat{\boldsymbol{Z}}_l$, we have

$$\frac{1}{\sqrt{m_n}}\sum_{l\in\overline{\mathbb{V}}}\widehat{\boldsymbol{Z}}_l\ \mathbb{M}_l\{\tau_k,\boldsymbol{Z}_l,\widehat{\boldsymbol{\beta}}(\tau)\}=o_{(0,\tau_U])}(\mathbf{1}),\ \ \text{a.s.}.$$

So we have

$$\frac{1}{\sqrt{m_n}}\sum_{l\in\overline{\mathbb{V}}}\widehat{\boldsymbol{Z}}_l\left[\mathbb{M}_l\left\{\tau_k,\widehat{\boldsymbol{Z}}_l,\widehat{\boldsymbol{\beta}}(\tau_k)\right\}-\mathbb{M}_l\left\{\tau_k,\boldsymbol{Z}_l,\widehat{\boldsymbol{\beta}}(\tau_k)\right\}\right]=o_{(0,\tau_U])}(\mathbf{1})\ \text{a.s.}.$$

Then (2.8) becomes,

$$\begin{aligned}\sqrt{n}(1-\rho_n)&\left(\widehat{\Omega}^{\overline{\mathbb{V}}}_{m_n}(\widehat{\boldsymbol{\beta}},\tau_k)-\Omega^{\overline{\mathbb{V}}}_{m_n}(\widehat{\boldsymbol{\beta}},\tau_k)\right)\\ &=\frac{\sqrt{n}(1-\rho)}{m_n}\sum_{l\in\overline{\mathbb{V}}}\left(\widehat{\boldsymbol{Z}}_l-\boldsymbol{Z}_l\right)\mathbb{M}_l\left\{\tau_k,\boldsymbol{Z}_l,\widehat{\boldsymbol{\beta}}(\tau_k)\right\}+o_{(0,\tau_U])}(\mathbf{1}),\end{aligned} \tag{2.9}$$

where $\rho=\lim_{n\to\infty}\rho_n$.

Now consider,

$$\widehat{\boldsymbol{Z}}_l-\boldsymbol{Z}_l=\begin{pmatrix}\widehat{\Phi}_l-\Phi_l\\ \mathbf{0}\end{pmatrix}=\begin{pmatrix}\mathbb{X}_l^\top\widehat{\boldsymbol{\theta}}-\Phi_l\\ \mathbf{0}\end{pmatrix}$$

where $\mathbf{0}$ is a $(p-1)$ zero-vector.

Under the conditions **C1** and **C2** (at page 24), asymptotically we have,

$$\left\|\widehat{\boldsymbol{Z}}_l-\boldsymbol{Z}_l\right\|=\left|\mathbb{X}_l^\top\boldsymbol{\theta}_0-\Phi_l\right|+O_p\left(\frac{1}{\sqrt{n}}\|\mathbb{X}\|\right)$$

$$= O_p\left(\frac{1}{\sqrt{n}}\|\mathbb{X}\|\right),$$

which acts in (2.9) as

$$\begin{aligned}
&\sqrt{n}(1-\rho_n)\left(\widehat{\Omega}^{\overline{\mathbb{V}}}_{m_n}(\widehat{\boldsymbol{\beta}},\tau_k) - \Omega^{\overline{\mathbb{V}}}_{m_n}(\widehat{\boldsymbol{\beta}},\tau_k)\right) \\
&\qquad = O_p\left(\frac{1}{n}\|\mathbb{X}\|\sum_{l\in\overline{\mathbb{V}}}\mathbb{M}_l\{\tau_k,\boldsymbol{Z}_l,\widehat{\boldsymbol{\beta}}(\tau_k)\}\right) + o_{(0,\tau_U])}(\mathbf{1}) \\
&\qquad = O_p\left(\frac{\mathbf{1}}{\sqrt{n}}\right) + o_{(0,\tau_U])}(\mathbf{1}). \qquad (2.10)
\end{aligned}$$

The equation (2.10) is due to the martingale central limit theorem. So by (2.10), (2.7) becomes

$$\begin{aligned}
&\sqrt{n}\ S_n(\widehat{\boldsymbol{\beta}},\tau_k) \\
&= \sqrt{n}\left\{\rho\ \Omega^{\mathbb{V}}_{m_v}(\widehat{\boldsymbol{\beta}},\tau_k) + (1-\rho)\ \Omega^{\overline{\mathbb{V}}}_{m_n}(\widehat{\boldsymbol{\beta}},\tau_k)\right\} + \xi_{n,k} + O_p\left(\frac{\mathbf{1}}{\sqrt{n}}\right) + o_{(0,\tau_U])}(\mathbf{1}) \\
&= \sqrt{n}\left\{\frac{\rho}{m_v}\sum_{j\in\mathbb{V}}\boldsymbol{X}_j\mathbb{M}_j\left\{\tau_k,\boldsymbol{X}_j,\widehat{\boldsymbol{\beta}}(\tau_k)\right\} + \frac{1-\rho}{m_n}\sum_{l\in\overline{\mathbb{V}}}\boldsymbol{Z}_l\mathbb{M}_l\left\{\tau_k,\boldsymbol{Z}_l,\widehat{\boldsymbol{\beta}}(\tau_k)\right\}\right\} \\
&\qquad + \xi_{n,k} + O_p\left(\frac{\mathbf{1}}{\sqrt{n}}\right) + o_{(0,\tau_U])}(\mathbf{1}) \\
&= \sqrt{n}\left\{\frac{\rho}{m_v}\sum_{j\in\mathbb{V}}\boldsymbol{X}_j\mathbb{M}_j\left\{\tau_k,\boldsymbol{X}_j,\widehat{\boldsymbol{\beta}}(\tau_k)\right\} + \frac{1-\rho}{m_n}\sum_{l\in\overline{\mathbb{V}}}\boldsymbol{Z}_l\mathbb{M}_l\left\{\tau_k,\boldsymbol{Z}_l,\widehat{\boldsymbol{\beta}}(\tau_k)\right\}\right\} \\
&\qquad + O_p\left(\frac{\mathbf{1}}{\sqrt{n}}\right) + o_{(0,\tau_U])}(\mathbf{1}). \qquad (2.11)
\end{aligned}$$

Because of the boundedness of $\widehat{\boldsymbol{Z}}_l$ and $\boldsymbol{Z}_l$ and the martingale property,

$$\begin{aligned}
E\left\{\widehat{\Omega}^{\overline{\mathbb{V}}}_{n-m}(\boldsymbol{\beta}_0,\tau) - \Omega^{\overline{\mathbb{V}}}_{n-m}(\boldsymbol{\beta}_0,\tau)\right\} &= \frac{1}{m_n}E\left\{\sum_{l\in\overline{\mathbb{V}}}\left(\widehat{\boldsymbol{Z}}_l - \boldsymbol{Z}_l\right)\mathbb{M}_l\left\{\tau,\boldsymbol{Z}_l,\boldsymbol{\beta}_0(\tau)\right\}\right\} \\
&= \mathbf{0}. \qquad (2.12)
\end{aligned}$$

2.2.3 Large Sample Theory

Define $F(t\,|\,\cdot) = Pr(Y \le t\,|\,\cdot)$, $\overline{F}(t\,|\,\cdot) = Pr(Y > t\,|\,\cdot)$, $\widetilde{F}(t\,|\,\cdot) = Pr(Y \le t, \delta = 1\,|\,\cdot)$, $\overline{f}(y\,|\,\cdot) = -f(y\,|\,\cdot) = -dF(y\,|\,\cdot)/dy$, $\widetilde{f}(y\,|\,\cdot) = d\widetilde{F}(y\,|\,\cdot)/dy$. (For a vector g, $g^{\otimes 2} = gg^\top$,

$g^{(l)} = l^{\text{th}}$ component of g, $\|g\|$ is the Euclidean norm of g.)

Regularity Conditions:

C1: For any given set $\boldsymbol{A} \subset \Re^p$,

(a) The conditional mean, $\Phi(W, \boldsymbol{X}_2, \boldsymbol{\theta}) = E(X_1 \mid W, \boldsymbol{X}_2)$ is continuous with respect to $\boldsymbol{\theta} \in \boldsymbol{A}$ and uniformly bounded.

(b) The class $\{\Phi(W, \boldsymbol{X}_2, \boldsymbol{\theta}), \boldsymbol{\theta} \in \boldsymbol{A}\}$ forms a P-Donsker class.

C2: The true value of $\boldsymbol{\theta}$, $\boldsymbol{\theta}_0$ is an interior point of $\boldsymbol{A}$ such that $\sqrt{n}\left\{\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right\}$ is asymptotically normal with mean 0 and finite variance.

C3: $\sup_i \|\boldsymbol{X}_i\| < \infty$ and $\sup_i \|\boldsymbol{Z}_i\| < \infty$.

C4: (a) Each component of $E\left[\boldsymbol{X}\mathbb{N}\left(e^{\boldsymbol{X}^\top \boldsymbol{\beta}_0(\tau)}\right)\right]$ and $E\left[\boldsymbol{Z}\mathbb{N}\left(e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau)}\right)\right]$ is a Lipschitz function of τ.

(b) $\widetilde{f}(t \mid \mathbf{x})$ and $f(t \mid \mathbf{x})$ are bounded above uniformly in t and $\mathbf{x}$.

(c) $\widetilde{f}(t \mid \mathbf{z})$ and $f(t \mid \mathbf{z})$ are bounded above uniformly in t and $\mathbf{z}$.

C5: (a) $\widetilde{f}\left(e^{\boldsymbol{X}^\top \boldsymbol{b}} \,\middle|\, \boldsymbol{X}\right) > 0$ and $\widetilde{f}\left(e^{\boldsymbol{Z}^\top \boldsymbol{b}} \,\middle|\, \boldsymbol{Z}\right) > 0$ for all $\boldsymbol{b} \in \mathscr{B}(d_0)$.

(b) To have positive definiteness, $E\left\{\boldsymbol{X}^{\otimes 2}\right\} > 0$ and $E\left\{\boldsymbol{Z}^{\otimes 2}\right\} > 0$.

(c) Each component of $E\left[\boldsymbol{X}^{\otimes 2}\ \overline{f}\left(e^{\boldsymbol{X}^\top \boldsymbol{b}} \,\middle|\, \boldsymbol{X}\right) e^{\boldsymbol{X}^\top \boldsymbol{b}}\right] \times \left(E\left[\boldsymbol{X}^{\otimes 2}\ \widetilde{f}\left(e^{\boldsymbol{X}^\top \boldsymbol{b}} \,\middle|\, \boldsymbol{X}\right) e^{\boldsymbol{X}^\top \boldsymbol{b}}\right]\right)^{-1}$ and $E\left[\boldsymbol{Z}^{\otimes 2}\ \overline{f}\left(e^{\boldsymbol{Z}^\top \boldsymbol{b}} \,\middle|\, \boldsymbol{Z}\right) e^{\boldsymbol{Z}^\top \boldsymbol{b}}\right] \times \left(E\left[\boldsymbol{Z}^{\otimes 2}\ \widetilde{f}\left(e^{\boldsymbol{Z}^\top \boldsymbol{b}} \,\middle|\, \boldsymbol{Z}\right) e^{\boldsymbol{Z}^\top \boldsymbol{b}}\right]\right)^{-1}$ is uniformly bounded in $\boldsymbol{b} \in \mathscr{B}(d_0)$; $\mathscr{B}(d_0)$ is a neighborhood containing $\{\boldsymbol{\beta}_0(\tau), \tau \in (0, \tau_U]\}$, defined in Appendix A.

C6: For any $\nu \in (0, \tau_U]$, $\inf_{\tau \in [\nu, \tau_U]} \text{eigmin } E\left[\boldsymbol{X}^{\otimes 2}\, \widetilde{f}\left(e^{\boldsymbol{X}^\top \boldsymbol{\beta}_0(\tau)} \,\middle|\, \boldsymbol{X}\right) \quad e^{\boldsymbol{X}^\top \boldsymbol{\beta}_0(\tau)}\right] > 0$ and $\inf_{\tau \in [\nu, \tau_U]} \text{eigmin } E\left[\boldsymbol{Z}^{\otimes 2}\, \widetilde{f}\left(e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau)} \,\middle|\, \boldsymbol{Z}\right) \quad e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau)}\right] > 0$, where eigmin$(\cdot)$ denotes the minimum eigenvalue of a matrix.

Theorem 2.2.1. *Assume that the regularity conditions* ***C1-C6*** *hold. If* $\lim_{n\to\infty} \|\mathbb{S}_L\| = 0$, *then* $\sup_{\tau \in [\nu, \tau_U]} \left\|\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\right\| \xrightarrow{\Pr} 0$, *where* $0 < \nu < \tau_U$.

Theorem 2.2.2. *Assume that the regularity conditions **C1-C6** hold. If* $\lim_{n\to\infty} \sqrt{n}\ \|\mathbb{S}_L\| = 0$*, then* $\sqrt{n}\ \left\{\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\right\}$ *weakly converges to a zero-mean Gaussian process for* $\tau \in [\nu, \tau_U]$*, where* $0 < \nu < \tau_U$*.*

Proofs of *Theorems* 2.2.1 and 2.2.2 are deferred to Appendices A and B respectively.

2.3 Numerical Studies

We conduct simulation studies to compare the performance of our proposed method with that based only on the validation sample, as well as the complete case, when the X_1 values are all known. We use the simulation models similar to those in Koenker [2008].

The logarithmic event times are generated from the following linear model:

$$T_i = \beta_0 + \beta_1 X_{1i} + u_i;\ i = 1, 2, \ldots, n$$

and the logarithmic censoring times are also generated from a linear model:

$$C_i = \gamma_0 + \gamma_1 X_{1i} + v_i;\ i = 1, 2, \ldots, n.$$

Here X_{1i}'s are iid $U[0, 5]$ and u_i's and v_i's are iid $N(0, 1)$. The parameters, $\boldsymbol{\beta}^\top = (5, 1)$ and $\boldsymbol{\gamma}^\top = (6.4, 0.75)$ were selected to maintain approximately 30% of the censoring proportion. We assumed that 50% of the observations are in the validation sample. We applied the estimator of Peng and Huang [2008] to the simulated data for the purpose of comparison. We compared our proposed method with the one assuming all X_1 are known ('Complete') and the one using only the validation sample. We generated W from an additive model:

$$W = X_1 + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\varepsilon} \sim N(0, \sigma^2_{\boldsymbol{\varepsilon}})$. In the simulation study, we chose $\sigma_{\boldsymbol{\varepsilon}} = 0.2,\ 0.8$ and sample sizes 200 and 500. We reported the mean bias and the root mean squared error (RMSE) based on 1000 simulations and used 250 bootstrap samples for estimating the standard error (SE) of the estimates and the calculation of the coverage probability (CP) of a 95% confidence interval of the model parameters. The quantiles considered in the simulation study are 0.25, 0.5 and 0.75 and the results are reported in Tables 2.1 - 2.3 respectively.

	Intercept				Slope			
	Bias	RMSE	SE	CP	Bias	RMSE	SE	CP
	$n = 200, \sigma_{\varepsilon} = 0.2$							
Complete	0.0173	0.1891	0.1996	96.00	0.0021	0.0682	0.0711	96.40
Proposed	0.0099	0.1973	0.2032	95.60	0.0035	0.0715	0.0725	96.70
Validation	0.0179	0.2687	0.2917	96.10	-0.0005	0.0966	0.1040	96.30
	$n = 200, \sigma_{\varepsilon} = 0.8$							
Complete	0.0173	0.1891	0.1996	96.00	0.0021	0.0682	0.0711	96.40
Proposed	-0.0519	0.2297	0.2413	95.80	0.0072	0.0811	0.0862	95.30
Validation	0.0179	0.2687	0.2917	96.10	-0.0005	0.0966	0.1040	96.30
	$n = 500, \sigma_{\varepsilon} = 0.2$							
Complete	0.0248	0.1246	0.1248	94.80	-0.0017	0.0429	0.0441	94.80
Proposed	0.0206	0.1268	0.1273	94.40	-0.0019	0.0444	0.0449	94.90
Validation	0.0263	0.1756	0.1783	94.80	0.0001	0.0615	0.0632	95.90
	$n = 500, \sigma_{\varepsilon} = 0.8$							
Complete	0.0248	0.1246	0.1248	94.80	-0.0017	0.0429	0.0441	94.80
Proposed	-0.0451	0.1524	0.1494	94.50	0.0037	0.0512	0.0531	95.40
Validation	0.0263	0.1756	0.1783	94.80	0.0001	0.0615	0.0632	95.90

Table 2.1: Comparison between regression calibration based approach and validation sample approach using the Bias, RMSE, SE and CP of regression parameters at $\tau = 0.25$

	Intercept				Slope			
	Bias	RMSE	SE	CP	Bias	RMSE	SE	CP
	$\boldsymbol{n = 200, \sigma_{\mathcal{E}} = 0.2}$							
Complete	0.0221	0.1835	0.1909	95.20	0.0018	0.0671	0.0700	95.40
Proposed	0.0225	0.1878	0.1939	94.49	0.0026	0.0685	0.0712	94.79
Validation	0.0290	0.2558	0.2784	95.40	0.0002	0.0925	0.1027	97.00
	$\boldsymbol{n = 200, \sigma_{\mathcal{E}} = 0.8}$							
Complete	0.0221	0.1835	0.1909	95.20	0.0018	0.0671	0.0700	95.40
Proposed	0.0317	0.2204	0.2308	95.50	0.0097	0.0827	0.0850	96.00
Validation	0.0290	0.2558	0.2784	95.40	0.0002	0.0925	0.1027	97.00
	$\boldsymbol{n = 500, \sigma_{\mathcal{E}} = 0.2}$							
Complete	0.0266	0.1249	0.1186	94.20	-0.0011	0.0427	0.0429	94.90
Proposed	0.0267	0.1264	0.1211	94.70	-0.0002	0.0435	0.0440	94.90
Validation	0.0285	0.1753	0.1707	94.30	0.0017	0.0609	0.0620	95.50
	$\boldsymbol{n = 500, \sigma_{\mathcal{E}} = 0.8}$							
Complete	0.0266	0.1249	0.1186	94.20	-0.0011	0.0427	0.0429	94.90
Proposed	0.0299	0.1452	0.1443	94.40	0.0079	0.0519	0.0526	94.70
Validation	0.0285	0.1753	0.1707	94.30	0.0017	0.0609	0.0620	95.50

Table 2.2: Comparison between regression calibration based approach and validation sample approach using the Bias, RMSE, SE and CP of regression parameters at $\tau = 0.5$

	Intercept				Slope			
	Bias	RMSE	SE	CP	Bias	RMSE	SE	CP
	$\boldsymbol{n = 200, \sigma_{\mathcal{E}} = 0.2}$							
Complete	0.0321	0.2105	0.2329	96.10	0.0062	0.0832	0.0976	97.10
Proposed	0.0434	0.2202	0.2397	96.50	0.0073	0.0852	0.0967	97.30
Validation	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	$\boldsymbol{n = 200, \sigma_{\mathcal{E}} = 0.8}$							
Complete	0.0321	0.2105	0.2329	96.10	0.0062	0.0832	0.0976	97.10
Proposed	0.1226	0.2907	0.2973	94.40	0.0229	0.1111	0.1195	96.70
Validation	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	$\boldsymbol{n = 500, \sigma_{\mathcal{E}} = 0.2}$							
Complete	0.0351	0.1455	0.1439	95.60	-0.0003	0.0526	0.0563	95.90
Proposed	0.0441	0.1533	0.1479	95.00	-0.0002	0.0551	0.0577	95.90
Validation	0.0450	0.2064	0.2083	95.30	0.0028	0.0767	0.0829	96.60
	$\boldsymbol{n = 500, \sigma_{\mathcal{E}} = 0.8}$							
Complete	0.0351	0.1455	0.1439	95.60	-0.0003	0.0526	0.0563	95.90
Proposed	0.1136	0.1997	0.1807	93.70	0.0172	0.0693	0.0710	95.20
Validation	0.0450	0.2064	0.2083	95.30	0.0028	0.0767	0.0829	96.60

Table 2.3: Comparison between regression calibration based approach and validation sample approach using the Bias, RMSE, SE and CP of regression parameters at $\tau = 0.75$

Tables 2.1, 2.2 and 2.3, show that our proposed estimators are asymptotically unbiased. The measures of variation (SE and RMSE) of the proposed method always stay between those in the complete case and using only the validation sample. Our proposed method is very efficient compared to using the validation sample only. When the $\sigma^2_{\mathcal{E}}$ is small, it works as well as the 'Complete' case. For a larger sample size, the coverage probability of the proposed method for the 95% confidence interval is approximately 95%.

We identified that using only the validation sample fails to provide estimates for higher quantiles when the sample size is small, as in Table 2.3.

2.3.1 PBC Data

As an illustration of our regression calibration based approach in the censored quantile regression model, we analyze the data from a clinical trial on Primary Biliary Cirrhosis (PBC) of the liver, conducted at the Mayo Clinic between 1974 and 1984 (data is available in Appendix D of Fleming and Harrington [2011] or the "survival" package in R software). PBC is a chronic autoimmune disease which affects the liver and is generally found among women aged between 40 and 60. The exact cause of PBC is still unknown and liver transplantation is the only possible way to be clear of it (National Digestive Diseases Information Clearinghouse and American Liver Foundation). It is a slow destruction of the bile ducts, which move bile, a fluid produced by the liver, to the intestines, aiding in the digestion of food and the disposal of worn out red blood cells, cholesterol and toxins from the human body. PBC causes the liver to function improperly because of inflammation and scarring.

In the Mayo clinic trial, 418 observations, each with 20 variables were available. The censoring rate was 0.615. Information regarding 310 observations was completely available and the remaining 108 observations were partially available. The incomplete subjects did not participate in the clinical trial, but provided their basic measurements and agreed to be followed to record survival. We considered only 5 covariates for the analysis, age of the patient (in years), serum albumin content in blood (in mg/dl), copper content in urine (ug/day), standardized blood clotting time and edema, the inflammation caused by excess fluid trapped in the body's tissues. Edema takes the following values:

$$\text{edema} = \begin{cases} 0, & \text{If no edema} \\ 0.5, & \text{If edema untreated or successfully treated} \\ 1, & \text{If there exists edema despite diuretic therapy.} \end{cases}$$

We considered the model using copper content in urine as our X_1 and the other covariates, age, albumin, blood clotting time and edema were complete and included in $\boldsymbol{X}_2$.

We defined two dummy variables for 'edema' as follows.

$$\text{Edema1} = \begin{cases} 1, & \text{If there exists edema despite diuretic therapy} \\ 0, & \text{Otherwise} \end{cases}$$

$$\text{Edema0.5} = \begin{cases} 1, & \text{If edema untreated or successfully treated} \\ 0, & \text{Otherwise} \end{cases}$$

As in the work of Granville and Fan [2014], to make marginal distributions closer to normal, we took the log transformation of age, albumin, copper and blood clotting time. The logarithm of the serum bilirubin content in blood (in mg/dl) is chosen as the auxiliary covariate (W) because of the high correlation (≈ 0.6) with the logarithm of the copper content in urine. We removed 2 subjects because of the missing $\boldsymbol{X}_2$ values. Finally, we considered that $n = 416$ and that the validation sample size is $m_v = 310$. The results are reported in Table 2.4.

In Table 2.4, we reported the parameter estimates, their standard error and 95% confidence limits and compared our proposed method with the estimates based only on the validation sample. We obtained the standard error and confidence limits using 250 bootstrap samples. The standard error of the estimates based on the validation sample is high compared to our proposed method, which shows that our method is more efficient. The widths of the confidence intervals are small for our proposed method.

		Validation			**Proposed**		
	$\tau \longrightarrow$	0.25	0.50	0.75	0.25	0.50	0.75
$\log(\hat{\beta})$	Intercept	19.5599	21.2413	23.6345	19.3461	19.8870	23.3248
	Age	-0.6552	-1.3863	-1.4283	-0.6321	-1.3289	-1.9788
	Albumin	2.1459	2.4975	2.1497	1.9857	2.3005	2.3675
	Copper	-0.5672	-0.6215	-0.7266	-0.5849	-0.6582	-0.8488
	Protime	-4.0750	-3.4839	-3.6980	-3.9234	-2.8374	-2.4402
	Edema1	-0.9777	-0.5987	-0.9373	-0.9970	-0.6912	-1.1445
	Edema0.5	-0.6736	0.0496	-0.2515	-0.6559	-0.0169	-0.3463
SE	Intercept	1.9879	4.5143	5.9999	2.1626	4.0633	5.2636
	Age	0.3019	0.5948	0.8755	0.2888	0.5872	0.8859
	Albumin	0.4453	0.8927	1.0479	0.4205	0.7893	0.8391
	Copper	0.0762	0.1566	0.2606	0.0861	0.1538	0.2499
	Protime	0.7167	1.6214	2.1205	0.7823	1.3471	1.7755
	Edema1	0.3720	0.4194	0.4133	0.3760	0.3775	0.4426
	Edema0.5	0.2130	0.4667	0.5737	0.1851	0.3537	0.4589
CI	Intercept	(15.45,23.25)	(11.77,29.47)	(9.49,33.01)	(15.16,23.64)	(13.61,29.54)	(12.57,33.2)
	Age	(-1.3,-0.12)	(-2.44,-0.1)	(-2.96,0.47)	(-1.37,-0.24)	(-2.64,-0.34)	(-3.42,0.05)
	Albumin	(1.31,3.06)	(0.81,4.31)	(0.26,4.37)	(1.21,2.86)	(0.79,3.89)	(0.41,3.7)
	Copper	(-0.71,-0.41)	(-0.91,-0.3)	(-1.13,-0.11)	(-0.77,-0.44)	(-1.01,-0.4)	(-1.26,-0.28)
	Protime	(-5.33,-2.52)	(-6.61,-0.26)	(-7.44,0.87)	(-5.18,-2.11)	(-5.81,-0.53)	(-6.28,0.68)
	Edema1	(-1.71,-0.25)	(-1.54,0.1)	(-1.82,-0.2)	(-1.72,-0.25)	(-1.46,0.02)	(-1.93,-0.2)
	Edema0.5	(-1.02,-0.19)	(-1.13,0.7)	(-1.26,0.98)	(-0.99,-0.26)	(-0.9,0.48)	(-1.11,0.69)

Table 2.4: Estimates, SE and 95% CI for regression parameters of PBC data analysis

2.4 Discussion

In this Chapter, we proposed the use of the regression calibration method to accommodate the auxiliary covariates in estimating the censored quantile regression parameters. We first applied the regression calibration method to predict the unavailable covariates in the non-validation sample using the auxiliary covariate, by assuming that they are linearly related. Then we applied Peng and Huang's censored quantile regression method to the whole study cohort for identifying the covariate effect on the observed survival time under heavy censoring at various quantile levels. Our proposed method is efficient compared to that using only the validation sample. The proposed estimators are consistent and have asymptotic normality.

Our proposed method is effective when the accurately measured covariate has a strong linear relationship with the auxiliary covariate. i.e, a high correlation between W and X_1. Our numerical studies show that our proposed method works as well as the 'Complete' case if σ^2_{ε} is small. But, our proposed method works always better than that using only the validation sample irrespective of the value of σ^2_{ε}. We applied our proposed method in an unobserved variable scenario of PBC data as an illustration in Section 2.3.1.

For application purposes, we should use only the auxiliary covariate which has a strong linear relationship with the accurately measured main exposure variable. We have to be very cautious when applying this method to the data with a very small validation sample size, m_v, compared to n.

Chapter 3

Kernel Smoothing in Censored Quantile Regression

3.1 Introduction

In the previous chapter, we discussed the regression calibration based method for estimating the censored quantile regression parameters with the auxiliary covariates, by assuming a linear association between the partially unavailable accurately measured covariate and the auxiliary covariates, as well as other available covariates. In this chapter, we would like to relax the restriction of the linearity and parametric assumptions between unavailable and auxiliary covariates. We introduce a non-parametric method to accommodate the auxiliary covariates in a general setup. Zhou and Wang [2000] investigated the failure time regression with error prone covariates based on kernel smoothing. Granville and Fan [2014] investigated Buckley-James estimator of AFT model with the auxiliary covariates in a semi-parametric setting using kernel smoothing.

In reality, it is very difficult to know the type of the association and distributional assumptions between unobserved and auxiliary covariates. For example, consider the Colorado Plateau uranium miners cohort data [Lubin et al., 1994; Langholz and Goldstein, 1996]. The study was undertaken to assess the risk of lung cancer due to radon exposure. We use miners' working time as the auxiliary covariate to predict the unavailable radon exposure. The scatter diagram of the working time and the logarithm of radon exposure is given in Figure 3.1 (page 49) and we can see that the relationship is not linear. To deal with this scenario, we propose a non-parametric estimation procedure to predict the unobserved data. We predict the unavailable covariates in the non-validation sample by using the validation sample with kernel smoothing and accommodating the auxiliary covariates. Other options, such as local linear approximations, can be applied with a similar level of technical difficulty.

The rest of the chapter is organized as follows. In Section 3.2, we present the estimation procedure, which is developed based on the Watson-Nadaraya estimator [Watson, 1964; Nadaraya, 1964] and the censored quantile regression approach of Peng and Huang [2008]. We establish the asymptotic properties of the proposed method in Section 3.2.1. Performance analysis using simulation studies and application of the proposed method to Colorado Plateau uranium miners cohort data are presented in Section 3.3. In Section 3.3.3, we apply the proposed method to PBC data.

3.2 Estimation

For a given τ ($0 < \tau < 1$), the quantile regression coefficient, $\boldsymbol{\beta}(\tau)$ can be estimated by solving the following estimating equation,

$$\sqrt{n}\; S_n(\boldsymbol{\beta}, \tau) = \mathbf{0},$$

where

$$S_n(\boldsymbol{\beta}, \tau) = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i \left(\mathbb{N}_i \left(e^{\boldsymbol{X}_i^\top \boldsymbol{\beta}(\tau)} \right) - \int_0^\tau \mathbb{I}\left[Y_i \geq e^{\boldsymbol{X}_i^\top \boldsymbol{\beta}(u)} \right] dH(u) \right).$$

The accurate measurements of the main exposure are not available for a subcohort. The auxiliary covariate, W which has a relationship with the partially unobserved covariate, X_1, is available for the entire study cohort. In the previous chapter, we assumed that X_1 has a linear relationship with W and other explanatory variables, $\boldsymbol{X}_2$. To relax this restriction, we consider a general relationship between X_1 and $(W, \boldsymbol{X}_2)$ as

$$X_1 = g(W, \boldsymbol{X}_2, \xi). \tag{3.1}$$

Some special cases are, as examples:

- Classical additive measurement error model: $W = X_1 + \boldsymbol{\varepsilon}$
- Berkson measurement error model: $X_1 = W + \xi$
- $E(X_1 \mid W, \boldsymbol{X}_2) = \Phi(W, \boldsymbol{X}_2)$
- $X_1 = \alpha W + \boldsymbol{\beta} \boldsymbol{X}_2 + \boldsymbol{\varepsilon}$,

where the form of $g(\cdot)$, $\Phi(\cdot)$ and the distribution of the random errors ξ and $\boldsymbol{\varepsilon}$ are not specified.

Similar to the previous chapter, we assume that a subcohort was randomly selected as the validation sample. In the covariate, $\boldsymbol{X} = (X_1, \boldsymbol{X}_2)$, we assume that X_1 is the accurately measured covariate subset which is partially available and $\boldsymbol{X}_2$ is the covariate subset which is available for the whole study cohort. We observe only $\{W, \boldsymbol{X}_2\}$ for the entire study cohort, where W is the auxiliary covariate of X_1. The observed data in the validation sample are $\{Y_j, W_j, X_{1j}, \boldsymbol{X}_{2j}, \delta_j\},\ j \in \mathbb{V}$ and in the non-validation sample, they are $\{Y_l, W_l, \boldsymbol{X}_{2l}, \delta_l\},\ l \in \overline{\mathbb{V}}$. If we consider that the entire study cohort sample size is n and that the validation sample size is m_v, then the non-validation sample size is $m_n = n - m_v$.

Assume that the expectation of X_1 conditional on $\{W, \boldsymbol{X}_2\}$ is a function of W and $\boldsymbol{X}_2$, say $\Phi(W, \boldsymbol{X}_2) = E(X_1 \mid W, \boldsymbol{X}_2)$. We propose a local polynomial approximation based approach to predict the unobserved values of X_1 in the non-validation sample. Kernel smoothing is a special case of the local polynomial approximation.

Remark 3.2.1. *The function $g(\cdot)$ in (3.1) can take a very general form as long as the model is informative. If it takes a linear form, then we can apply the regression calibration method introduced in the previous chapter.*

If W is only defined as a categorical variable, the methods of Zhou and Pepe [1995], Liu, Zhou and Cai [2009] and Liu et al. [2012] would be used.

Our estimating function is

$$\sqrt{n}\ S_n(\boldsymbol{\beta}, \tau)$$

and $\widehat{\boldsymbol{\beta}}$ is the generalized solution of $\sqrt{n}\ S_n(\boldsymbol{\beta}, \tau) = \mathbf{0}$. Let $\Phi_l = \Phi(W_l, \boldsymbol{X}_{2l})$ and

$$\widehat{\Phi}_l = \widehat{\Phi}(W_l, \boldsymbol{X}_{2l}) = \frac{\sum_{j\in\mathbb{V}} \mathcal{K}_{\mathfrak{h}}\Big(W_j - W_l, \boldsymbol{X}_{2j} - \boldsymbol{X}_{2l}\Big) X_{1j}}{\sum_{j\in\mathbb{V}} \mathcal{K}_{\mathfrak{h}}\Big(W_j - W_l, \boldsymbol{X}_{2j} - \boldsymbol{X}_{2l}\Big)}, \tag{3.2}$$

where $\mathcal{K}_{\mathfrak{h}}(\cdot) = \mathcal{K}(\cdot/\mathfrak{h})$. $\mathcal{K}(\cdot)$ is the Gaussian kernel on $\Re^p$ with bandwidth $\mathfrak{h}$. Let $\boldsymbol{Z} = (\Phi, \boldsymbol{X}_2)$ and $\widehat{\boldsymbol{Z}} = (\widehat{\Phi}, \boldsymbol{X}_2)$, then

$$\begin{aligned} S_n(\boldsymbol{\beta}, \tau) = {}& \frac{\rho_n}{m_v} \sum_{j\in\mathbb{V}} \boldsymbol{X}_j \left\{ \mathbb{N}_j\left(e^{\boldsymbol{X}_j^\top \boldsymbol{\beta}(\tau)}\right) - \int_0^\tau \mathbb{I}\left[Y_j \geq e^{\boldsymbol{X}_j^\top \boldsymbol{\beta}(u)}\right] dH(u) \right\} \\ & + \frac{1-\rho_n}{m_n} \sum_{l\in\overline{\mathbb{V}}} \widehat{\boldsymbol{Z}}_l \left\{ \mathbb{N}_l\left(e^{\widehat{\boldsymbol{Z}}_l^\top \boldsymbol{\beta}(\tau)}\right) - \int_0^\tau \mathbb{I}\left[Y_l \geq e^{\widehat{\boldsymbol{Z}}_l^\top \boldsymbol{\beta}(u)}\right] dH(u) \right\} \end{aligned}$$

$$= \rho_n \; \Omega^{\mathbb{V}}_{m_v}(\boldsymbol{\beta}, \tau) + (1 - \rho_n) \; \widehat{\Omega}^{\overline{\mathbb{V}}}_{m_n}(\boldsymbol{\beta}, \tau).$$

where $\rho_n = m_v/n$. Here the first part on the right hand side of the equation comes from the validation sample and the second part is from the non-validation sample. For a particular quantile, τ_k, the estimator of $\boldsymbol{\beta}_0(\tau_k)$ is $\widehat{\boldsymbol{\beta}}(\tau_k)$, which is the generalized solution of $\sqrt{n}\; S_n(\boldsymbol{\beta}, \tau_k) = \mathbf{0}$.

$$\sqrt{n}\; S_n(\widehat{\boldsymbol{\beta}}, \tau_k) = \sqrt{n}\left\{\rho_n \; \Omega^{\mathbb{V}}_{m_v}(\widehat{\boldsymbol{\beta}}, \tau_k) + (1 - \rho_n) \; \widehat{\Omega}^{\overline{\mathbb{V}}}_{m_n}(\widehat{\boldsymbol{\beta}}, \tau_k)\right\} + \xi_{n,k}, \tag{3.3}$$

for $k = 1, \ldots, L(n)$, where,

$$\Omega^{\mathbb{V}}_{m_v}(\widehat{\boldsymbol{\beta}}, \tau_k) = \frac{1}{m_v} \sum_{j \in \mathbb{V}} \boldsymbol{X}_j \left\{ \mathbb{N}_j \left(e^{\boldsymbol{X}_j^\top \widehat{\boldsymbol{\beta}}(\tau_k)} \right) - \int_0^{\tau_k} \mathbb{I}\left[Y_j \geq e^{\boldsymbol{X}_j^\top \widehat{\boldsymbol{\beta}}(u)} \right] dH(u) \right\},$$

$$\widehat{\Omega}^{\overline{\mathbb{V}}}_{m_n}(\widehat{\boldsymbol{\beta}}, \tau_k) = \frac{1}{m_n} \sum_{l \in \overline{\mathbb{V}}} \widehat{\boldsymbol{Z}}_l \left\{ \mathbb{N}_l \left(e^{\widehat{\boldsymbol{Z}}_l^\top \widehat{\boldsymbol{\beta}}(\tau_k)} \right) - \int_0^{\tau_k} \mathbb{I}\left[Y_l \geq e^{\widehat{\boldsymbol{Z}}_l^\top \widehat{\boldsymbol{\beta}}(u)} \right] dH(u) \right\}.$$

For simplicity, denote $L = L(n)$. Here, by the definition of a generalized solution (defined in Section 2.2), $\max_{k=1,2,\ldots,L} \|\xi_{n,k}\| \leq \sup_i \|\boldsymbol{X}_i\|/\sqrt{n}$. $\widehat{\boldsymbol{\beta}}(\tau)$ is a right- continuous piecewise constant function which jumps only on a grid, $\mathbb{S}_L = \{0 = \tau_0 < \tau_1 < \cdots < \tau_L = \tau_U < 1\}$. The size of $\mathbb{S}_L$ is defined as $\|\mathbb{S}_L\| = \max_k \{\tau_k - \tau_{k-1};\; k = 1, \ldots, L\}$.

Let $s(\boldsymbol{\beta}, \tau) = E\left\{\rho_n \; \Omega^{\mathbb{V}}_{m_v}(\boldsymbol{\beta}, \tau) + (1 - \rho_n) \; \widehat{\Omega}^{\overline{\mathbb{V}}}_{m_n}(\boldsymbol{\beta}, \tau)\right\}$. Define

$$\Omega^{\overline{\mathbb{V}}}_{m_n}(\boldsymbol{\beta}, \tau_k) = \frac{1}{m_n} \sum_{l \in \overline{\mathbb{V}}} \boldsymbol{Z}_l \left\{ \mathbb{N}_l \left(e^{\boldsymbol{Z}_l^\top \boldsymbol{\beta}(\tau_k)} \right) - \int_0^{\tau_k} \mathbb{I}\left[Y_l \geq e^{\boldsymbol{Z}_l^\top \widehat{\boldsymbol{\beta}}(u)} \right] dH(u) \right\}.$$

Using the martingale property, we have $E\{\Omega^{\mathbb{V}}_{m_v}(\boldsymbol{\beta}_0, \tau)\} = \mathbf{0}$ and $E\{\Omega^{\overline{\mathbb{V}}}_{m_n}(\boldsymbol{\beta}_0, \tau)\} = \mathbf{0}$. By the equation (3.8) and the martingale property, $s(\boldsymbol{\beta}_0, \tau) = \mathbf{0}$, where $\boldsymbol{\beta}_0(\tau_k)$ denotes the true $\boldsymbol{\beta}(\tau_k)$. Now we have,

$$\begin{aligned}
&\widehat{\Omega}^{\overline{\mathbb{V}}}_{m_n}(\widehat{\boldsymbol{\beta}}, \tau_k) - \Omega^{\overline{\mathbb{V}}}_{m_n}(\widehat{\boldsymbol{\beta}}, \tau_k) \\
&\quad = \frac{1}{m_n} \left(\sum_{l \in \overline{\mathbb{V}}} \widehat{\boldsymbol{Z}}_l \left\{ \mathbb{N}_l \left(e^{\widehat{\boldsymbol{Z}}_l^\top \widehat{\boldsymbol{\beta}}(\tau_k)} \right) - \int_0^{\tau_k} \mathbb{I}\left[Y_l \geq e^{\widehat{\boldsymbol{Z}}_l^\top \widehat{\boldsymbol{\beta}}(u)} \right] dH(u) \right\} \right. \\
&\qquad \left. - \sum_{l \in \overline{\mathbb{V}}} \boldsymbol{Z}_l \left\{ \mathbb{N}_l \left(e^{\boldsymbol{Z}_l^\top \widehat{\boldsymbol{\beta}}(\tau_k)} \right) - \int_0^{\tau_k} \mathbb{I}\left[Y_l \geq e^{\boldsymbol{Z}_l^\top \widehat{\boldsymbol{\beta}}(u)} \right] dH(u) \right\} \right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{m_n}\left(\sum_{l\in\overline{\mathbb{V}}}\widehat{\boldsymbol{Z}}_l\mathbb{N}_l\left(e^{\widehat{\boldsymbol{Z}}_l^\top\widehat{\boldsymbol{\beta}}(\tau_k)}\right)-\sum_{l\in\overline{\mathbb{V}}}\widehat{\boldsymbol{Z}}_l\int_0^{\tau_k}\mathbb{I}\left[Y_l\geq e^{\widehat{\boldsymbol{Z}}_l^\top\widehat{\boldsymbol{\beta}}(u)}\right]dH(u)\right.\\
&\qquad\left.-\sum_{l\in\overline{\mathbb{V}}}\boldsymbol{Z}_l\mathbb{N}_l\left(e^{\boldsymbol{Z}_l^\top\widehat{\boldsymbol{\beta}}(\tau_k)}\right)-\sum_{l\in\overline{\mathbb{V}}}\boldsymbol{Z}_l\int_0^{\tau_k}\mathbb{I}\left[Y_l\geq e^{\boldsymbol{Z}_l^\top\widehat{\boldsymbol{\beta}}(u)}\right]dH(u)\right)\\
&= \frac{1}{m_n}\left(\sum_{l\in\overline{\mathbb{V}}}\left\{\widehat{\boldsymbol{Z}}_l\mathbb{N}_l\left(e^{\widehat{\boldsymbol{Z}}_l^\top\widehat{\boldsymbol{\beta}}(\tau_k)}\right)-\boldsymbol{Z}_l\mathbb{N}_l\left(e^{\boldsymbol{Z}_l^\top\widehat{\boldsymbol{\beta}}(\tau_k)}\right)\right\}\right.\\
&\qquad\left.-\sum_{l\in\overline{\mathbb{V}}}\int_0^{\tau_k}\left\{\widehat{\boldsymbol{Z}}_l\mathbb{I}\left[Y_l\geq e^{\widehat{\boldsymbol{Z}}_l^\top\widehat{\boldsymbol{\beta}}(u)}\right]-\boldsymbol{Z}_l\mathbb{I}\left[Y_l\geq e^{\boldsymbol{Z}_l^\top\widehat{\boldsymbol{\beta}}(u)}\right]\right\}dH(u)\right)\\
&= \frac{1}{m_n}\sum_{l\in\overline{\mathbb{V}}}\left(\widehat{\boldsymbol{Z}}_l\left[\mathbb{N}_l\left(e^{\widehat{\boldsymbol{Z}}_l^\top\widehat{\boldsymbol{\beta}}(\tau_k)}\right)-\mathbb{N}_l\left(e^{\boldsymbol{Z}_l^\top\widehat{\boldsymbol{\beta}}(\tau_k)}\right)\right]\right.\\
&\qquad+\left(\widehat{\boldsymbol{Z}}_l-\boldsymbol{Z}_l\right)\mathbb{N}_l\left(e^{\boldsymbol{Z}_l^\top\widehat{\boldsymbol{\beta}}(\tau_k)}\right)\\
&\qquad-\int_0^{\tau_k}\left(\widehat{\boldsymbol{Z}}_l\left\{\mathbb{I}\left[Y_l\geq e^{\widehat{\boldsymbol{Z}}_l^\top\widehat{\boldsymbol{\beta}}(u)}\right]-\mathbb{I}\left[Y_l\geq e^{\boldsymbol{Z}_l^\top\widehat{\boldsymbol{\beta}}(u)}\right]\right\}\right.\\
&\qquad\left.\left.-\left(\widehat{\boldsymbol{Z}}_l-\boldsymbol{Z}_l\right)\mathbb{I}\left[Y_l\geq e^{\boldsymbol{Z}_l^\top\widehat{\boldsymbol{\beta}}(u)}\right]\right)dH(u)\right)\\
&= \frac{1}{m_n}\sum_{l\in\overline{\mathbb{V}}}\left(\widehat{\boldsymbol{Z}}_l\left[\mathbb{N}_l\left(e^{\widehat{\boldsymbol{Z}}_l^\top\widehat{\boldsymbol{\beta}}(\tau_k)}\right)-\mathbb{N}_l\left(e^{\boldsymbol{Z}_l^\top\widehat{\boldsymbol{\beta}}(\tau_k)}\right)\right]\right.\\
&\qquad-\int_0^{\tau_k}\widehat{\boldsymbol{Z}}_l\left\{\mathbb{I}\left[Y_l\geq e^{\widehat{\boldsymbol{Z}}_l^\top\widehat{\boldsymbol{\beta}}(u)}\right]-\mathbb{I}\left[Y_l\geq e^{\boldsymbol{Z}_l^\top\widehat{\boldsymbol{\beta}}(u)}\right]\right\}dH(u)\\
&\qquad+\left(\widehat{\boldsymbol{Z}}_l-\boldsymbol{Z}_l\right)\mathbb{N}_l\left(e^{\boldsymbol{Z}_l^\top\widehat{\boldsymbol{\beta}}(\tau_k)}\right)\\
&\qquad\left.-\int_0^{\tau_k}\left(\widehat{\boldsymbol{Z}}_l-\boldsymbol{Z}_l\right)\mathbb{I}\left[Y_l\geq e^{\boldsymbol{Z}_l^\top\widehat{\boldsymbol{\beta}}(u)}\right]dH(u)\right)\\
&= \frac{1}{m_n}\sum_{l\in\overline{\mathbb{V}}}\left(\widehat{\boldsymbol{Z}}_l\left[\mathbb{M}_l\left\{\tau_k,\widehat{\boldsymbol{Z}}_l,\widehat{\boldsymbol{\beta}}(\tau_k)\right\}-\mathbb{M}_l\left\{\tau_k,\boldsymbol{Z}_l,\widehat{\boldsymbol{\beta}}(\tau_k)\right\}\right]\right.\\
&\qquad\left.+\left(\widehat{\boldsymbol{Z}}_l-\boldsymbol{Z}_l\right)\mathbb{M}_l\left\{\tau_k,\boldsymbol{Z}_l,\widehat{\boldsymbol{\beta}}(\tau_k)\right\}\right)
\end{aligned}
$$

$$
\sqrt{n}(1-\rho_n)\left(\widehat{\Omega}_{m_n}^{\overline{\mathbb{V}}}(\widehat{\boldsymbol{\beta}},\tau_k)-\Omega_{m_n}^{\overline{\mathbb{V}}}(\widehat{\boldsymbol{\beta}},\tau_k)\right)
$$

$$= \frac{\sqrt{n}(1-\rho_n)}{m_n} \sum_{l \in \overline{\mathbb{V}}} \Bigg(\widehat{\boldsymbol{Z}}_l \Big[\mathbb{M}_l \Big\{ \tau_k, \widehat{\boldsymbol{Z}}_l, \widehat{\boldsymbol{\beta}}(\tau_k) \Big\} - \mathbb{M}_l \Big\{ \tau_k, \boldsymbol{Z}_l, \widehat{\boldsymbol{\beta}}(\tau_k) \Big\} \Big]$$
$$+ \Big(\widehat{\boldsymbol{Z}}_l - \boldsymbol{Z}_l \Big) \mathbb{M}_l \Big\{ \tau_k, \boldsymbol{Z}_l, \widehat{\boldsymbol{\beta}}(\tau_k) \Big\} \Bigg), \tag{3.4}$$

where

$$\mathbb{M}_i \Big\{ \tau_k, \boldsymbol{U}_i, \widehat{\boldsymbol{\beta}}(\tau_k) \Big\} = \mathbb{N}_i \left(e^{\boldsymbol{U}_i^\top \widehat{\boldsymbol{\beta}}(\tau_k)} \right) - \int_0^{\tau_k} \mathbb{I} \left[Y_i \geq e^{\boldsymbol{U}_i^\top \widehat{\boldsymbol{\beta}}(u)} \right] dH(u).$$

According to Appendix B (proof of Theorem 2) of [Peng and Huang, 2008, p. 647], since $\sqrt{m_n}\,\|\mathbb{S}_L\| \to 0$, we have

$$\frac{1}{\sqrt{m_n}} \sum_{l \in \overline{\mathbb{V}}} \widehat{\boldsymbol{Z}}_l \ \mathbb{M}_l \{ \tau_k, \widehat{\boldsymbol{Z}}_l, \widehat{\boldsymbol{\beta}}(\tau) \} = o_{(0,\tau_U])}(\mathbf{1}), \ \text{ a. s.}\,.$$

Using similar arguments as in Appendix B (proof of Theorem 2) of [Peng and Huang, 2008, p. 647] and because of the boundedness of $\widehat{\boldsymbol{Z}}_l$, we have

$$\frac{1}{\sqrt{m_n}} \sum_{l \in \overline{\mathbb{V}}} \widehat{\boldsymbol{Z}}_l \ \mathbb{M}_l \{ \tau_k, \boldsymbol{Z}_l, \widehat{\boldsymbol{\beta}}(\tau) \} = o_{(0,\tau_U])}(\mathbf{1}), \ \text{ a. s.}\,.$$

So we have

$$\frac{1}{\sqrt{m_n}} \sum_{l \in \overline{\mathbb{V}}} \widehat{\boldsymbol{Z}}_l \Big[\mathbb{M}_l \Big\{ \tau_k, \widehat{\boldsymbol{Z}}_l, \widehat{\boldsymbol{\beta}}(\tau_k) \Big\} - \mathbb{M}_l \Big\{ \tau_k, \boldsymbol{Z}_l, \widehat{\boldsymbol{\beta}}(\tau_k) \Big\} \Big] = o_{(0,\tau_U])}(\mathbf{1}) \text{ a. s.}\,.$$

Then (3.4) becomes,

$$\sqrt{n}(1-\rho_n) \left(\widehat{\Omega}_{m_n}^{\overline{\mathbb{V}}}(\widehat{\boldsymbol{\beta}}, \tau_k) - \Omega_{m_n}^{\overline{\mathbb{V}}}(\widehat{\boldsymbol{\beta}}, \tau_k) \right)$$
$$= \frac{\sqrt{n}(1-\rho)}{m_n} \sum_{l \in \overline{\mathbb{V}}} \Big(\widehat{\boldsymbol{Z}}_l - \boldsymbol{Z}_l \Big) \mathbb{M}_l \Big\{ \tau_k, \boldsymbol{Z}_l, \widehat{\boldsymbol{\beta}}(\tau_k) \Big\} + o_{(0,\tau_U])}(\mathbf{1}), \tag{3.5}$$

where $\rho = \lim_{n \to \infty} \rho_n$.

Now consider,

$$\widehat{\boldsymbol{Z}}_l - \boldsymbol{Z}_l = \frac{\sum_{j \in \mathbb{V}} \mathcal{K}_{\mathfrak{h}} \Big(W_j - W_l, \boldsymbol{X}_{2j} - \boldsymbol{X}_{2l} \Big) (\boldsymbol{X}_j - \boldsymbol{Z}_l)}{\sum_{j \in \mathbb{V}} \mathcal{K}_{\mathfrak{h}} \Big(W_j - W_l, \boldsymbol{X}_{2j} - \boldsymbol{X}_{2l} \Big)}$$

$$= \frac{\sum_{j\in\mathbb{V}} \mathcal{K}_{\mathfrak{h}}\Big(W_j - W_l, \boldsymbol{X}_{2j} - \boldsymbol{X}_{2l}\Big)(\boldsymbol{X}_j - \boldsymbol{Z}_l)}{\sum_{l\in\overline{\mathbb{V}}} \mathcal{K}_{\mathfrak{h}}\Big(W_j - W_l, \boldsymbol{X}_{2j} - \boldsymbol{X}_{2l}\Big)} \frac{\sum_{l\in\overline{\mathbb{V}}} \mathcal{K}_{\mathfrak{h}}\Big(W_j - W_l, \boldsymbol{X}_{2j} - \boldsymbol{X}_{2l}\Big)}{\sum_{j\in\mathbb{V}} \mathcal{K}_{\mathfrak{h}}\Big(W_j - W_l, \boldsymbol{X}_{2j} - \boldsymbol{X}_{2l}\Big)}.$$

Asymptotically,

$$\frac{m_n}{m_v} \frac{\sum_{l\in\overline{\mathbb{V}}} \mathcal{K}_{\mathfrak{h}}\Big(W_j - W_l, \boldsymbol{X}_{2j} - \boldsymbol{X}_{2l}\Big) \Big/ m_n}{\sum_{j\in\mathbb{V}} \mathcal{K}_{\mathfrak{h}}\Big(W_j - W_l, \boldsymbol{X}_{2j} - \boldsymbol{X}_{2l}\Big) \Big/ m_v} \longrightarrow \varrho,$$

where $\varrho = \lim_{n\to\infty} \dfrac{1-\rho_n}{\rho_n}$. Then, (3.5) can be rewritten as,

$$\begin{aligned} &\sqrt{n}(1-\rho_n)\left(\widehat{\Omega}^{\overline{\mathbb{V}}}_{m_n}(\widehat{\boldsymbol{\beta}}, \tau_k) - \Omega^{\overline{\mathbb{V}}}_{m_n}(\widehat{\boldsymbol{\beta}}, \tau_k)\right) \\ &\quad = \frac{\sqrt{n}(1-\rho)}{m_n} \varrho \sum_{l\in\overline{\mathbb{V}}} \frac{\sum_{j\in\mathbb{V}} \mathcal{K}_{\mathfrak{h}}\Big(W_j - W_l, \boldsymbol{X}_{2j} - \boldsymbol{X}_{2l}\Big)(\boldsymbol{X}_j - \boldsymbol{Z}_l)}{\sum_{l\in\overline{\mathbb{V}}} \mathcal{K}_{\mathfrak{h}}\Big(W_j - W_l, \boldsymbol{X}_{2j} - \boldsymbol{X}_{2l}\Big)} \mathbb{M}_l\left\{\tau_k, \boldsymbol{Z}_l, \widehat{\boldsymbol{\beta}}(\tau_k)\right\} \\ &\qquad + O_p\left(\frac{\mathbf{1}}{\sqrt{n}}\right) + o_{(0,\tau_U])}(\mathbf{1}). \end{aligned}$$

The denominator will become a function of 'j' after summing over 'l' and changing the order of summation,

$$\begin{aligned} &\sqrt{n}(1-\rho_n)\left(\widehat{\Omega}^{\overline{\mathbb{V}}}_{m_n}(\widehat{\boldsymbol{\beta}}, \tau_k) - \Omega^{\overline{\mathbb{V}}}_{m_n}(\widehat{\boldsymbol{\beta}}, \tau_k)\right) \\ &\quad = \sqrt{n}\, \frac{m_n}{n} \frac{1}{m_n} \frac{m_v}{m_v} \varrho \sum_{j\in\mathbb{V}} \frac{\sum_{l\in\overline{\mathbb{V}}} \mathcal{K}_{\mathfrak{h}}\Big(W_j - W_l, \boldsymbol{X}_{2j} - \boldsymbol{X}_{2l}\Big)(\boldsymbol{X}_j - \boldsymbol{Z}_l)\ \mathbb{M}_l\left\{\tau_k, \boldsymbol{Z}_l, \widehat{\boldsymbol{\beta}}(\tau_k)\right\}}{\sum_{l\in\overline{\mathbb{V}}} \mathcal{K}_{\mathfrak{h}}\Big(W_j - W_l, \boldsymbol{X}_{2j} - \boldsymbol{X}_{2l}\Big)} \\ &\qquad + O_p\left(\frac{\mathbf{1}}{\sqrt{n}}\right) + o_{(0,\tau_U])}(\mathbf{1}) \\ &\quad = \frac{\sqrt{n}\ \rho}{m_v} \varrho \sum_{j\in\mathbb{V}} \Big[\boldsymbol{X}_j \mathbb{W}_{qM\overline{\mathbb{V}}}\Big(\mathbb{M}_l\left\{\tau_k, \boldsymbol{Z}_l, \widehat{\boldsymbol{\beta}}(\tau_k)\right\}\Big) \\ &\qquad - \mathbb{W}_{qM\overline{\mathbb{V}}}\Big(\boldsymbol{Z}_l\ \mathbb{M}_l\left\{\tau_k, \boldsymbol{Z}_l, \widehat{\boldsymbol{\beta}}(\tau_k)\right\}\Big)\Big] + O_p\left(\frac{\mathbf{1}}{\sqrt{n}}\right) + o_{(0,\tau_U])}(\mathbf{1}), \qquad (3.6) \end{aligned}$$

where $\mathbb{W}_{qM\overline{\mathbb{V}}}(G)$ denotes $\dfrac{\sum_{l\in\overline{\mathbb{V}}} \mathcal{K}_{\mathfrak{h}}\Big(W_j - W_l, \boldsymbol{X}_{2j} - \boldsymbol{X}_{2l}\Big) G_l}{\sum_{l\in\overline{\mathbb{V}}} \mathcal{K}_{\mathfrak{h}}\Big(W_j - W_l, \boldsymbol{X}_{2j} - \boldsymbol{X}_{2l}\Big)}$, the weighted mean of terms for the non-validation sample.

By (3.6), (3.3) becomes

$$
\begin{aligned}
&\sqrt{n}\ S_n(\widehat{\boldsymbol{\beta}}, \tau_k) \\
&= \sqrt{n}\Bigg\{\rho\ \Omega^{\mathbb{V}}_{m_v}(\widehat{\boldsymbol{\beta}}, \tau_k) + (1-\rho)\ \Omega^{\overline{\mathbb{V}}}_{m_n}(\widehat{\boldsymbol{\beta}}, \tau_k) \\
&+ \frac{\rho}{m_v}\ \varrho \sum_{j\in\mathbb{V}} \Big[\boldsymbol{X}_j \mathbb{W}_{qM\overline{\mathbb{V}}}\Big(\mathbb{M}_l\left\{\tau_k, \boldsymbol{Z}_l, \widehat{\boldsymbol{\beta}}(\tau_k)\right\}\Big) - \mathbb{W}_{qM\overline{\mathbb{V}}}\Big(\boldsymbol{Z}_l\ \mathbb{M}_l\left\{\tau_k, \boldsymbol{Z}_l, \widehat{\boldsymbol{\beta}}(\tau_k)\right\}\Big)\Big]\Bigg\} \\
&\qquad + \xi_{n,k} + O_p\left(\frac{\mathbf{1}}{\sqrt{n}}\right) + o_{(0,\tau_U])}(\mathbf{1}) \\
&= \sqrt{n}\Bigg\{\frac{\rho}{m_v}\sum_{j\in\mathbb{V}}\Bigg(\boldsymbol{X}_j \mathbb{M}_j\left\{\tau_k, \boldsymbol{X}_j, \widehat{\boldsymbol{\beta}}(\tau_k)\right\} \\
&\qquad + \varrho\Big[\boldsymbol{X}_j \mathbb{W}_{qM\overline{\mathbb{V}}}\Big(\mathbb{M}_l\left\{\tau_k, \boldsymbol{Z}_l, \widehat{\boldsymbol{\beta}}(\tau_k)\right\}\Big) - \mathbb{W}_{qM\overline{\mathbb{V}}}\Big(\boldsymbol{Z}_l\ \mathbb{M}_l\left\{\tau_k, \boldsymbol{Z}_l, \widehat{\boldsymbol{\beta}}(\tau_k)\right\}\Big)\Big]\Bigg) \\
&\qquad + \frac{(1-\rho)}{m_n}\sum_{l\in\overline{\mathbb{V}}} \boldsymbol{Z}_l \mathbb{M}_l\left\{\tau_k, \boldsymbol{Z}_l, \widehat{\boldsymbol{\beta}}(\tau_k)\right\}\Bigg\} + \xi_{n,k} + O_p\left(\frac{\mathbf{1}}{\sqrt{n}}\right) + o_{(0,\tau_U])}(\mathbf{1}) \\
&= \sqrt{n}\Bigg\{\frac{\rho}{m_v}\sum_{j\in\mathbb{V}}\Bigg(\boldsymbol{X}_j \mathbb{M}_j\left\{\tau_k, \boldsymbol{X}_j, \widehat{\boldsymbol{\beta}}(\tau_k)\right\} \\
&\qquad + \varrho\Big[\boldsymbol{X}_j \mathbb{W}_{qM\overline{\mathbb{V}}}\Big(\mathbb{M}_l\left\{\tau_k, \boldsymbol{Z}_l, \widehat{\boldsymbol{\beta}}(\tau_k)\right\}\Big) - \mathbb{W}_{qM\overline{\mathbb{V}}}\Big(\boldsymbol{Z}_l\ \mathbb{M}_l\left\{\tau_k, \boldsymbol{Z}_l, \widehat{\boldsymbol{\beta}}(\tau_k)\right\}\Big)\Big]\Bigg) \\
&\qquad + \frac{(1-\rho)}{m_n}\sum_{l\in\overline{\mathbb{V}}} \boldsymbol{Z}_l \mathbb{M}_l\left\{\tau_k, \boldsymbol{Z}_l, \widehat{\boldsymbol{\beta}}(\tau_k)\right\}\Bigg\} + O_p\left(\frac{\mathbf{1}}{\sqrt{n}}\right) + o_{(0,\tau_U])}(\mathbf{1}).
\end{aligned}
\tag{3.7}
$$

$$
\begin{aligned}
&(1-\rho_n)E\left\{\widehat{\Omega}^{\overline{\mathbb{V}}}_{m_n}(\boldsymbol{\beta}_0, \tau) - \Omega^{\overline{\mathbb{V}}}_{m_n}(\boldsymbol{\beta}_0, \tau)\right\} \\
&= \frac{\rho\ \varrho}{m_v}\ E\Bigg\{\sum_{j\in\mathbb{V}}\Big[\boldsymbol{X}_j \mathbb{W}_{qM\overline{\mathbb{V}}}\Big(\mathbb{M}_l\left\{\tau, \boldsymbol{Z}_l, \boldsymbol{\beta}_0(\tau)\right\}\Big) - \mathbb{W}_{qM\overline{\mathbb{V}}}\Big(\boldsymbol{Z}_l\ \mathbb{M}_l\left\{\tau, \boldsymbol{Z}_l, \boldsymbol{\beta}_0(\tau)\right\}\Big)\Big] \\
&\qquad + O_p\left(\frac{\mathbf{1}}{n}\right)\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{\rho}{m_v}\ \varrho\ E\left(\sum_{j\in\mathbb{V}} \frac{\sum_{l\in\overline{\mathbb{V}}} \mathcal{K}_{\mathfrak{h}}(\,\cdot\,)(\boldsymbol{X}_j - \boldsymbol{Z}_l)\ \mathbb{M}_l\left\{\tau, \boldsymbol{Z}_l, \boldsymbol{\beta}_0(\tau)\right\}}{\sum_{l\in\overline{\mathbb{V}}} \mathcal{K}_{\mathfrak{h}}(\,\cdot\,)} + O_p\left(\frac{\mathbf{1}}{n}\right)\right) \\
&= \frac{1-\rho}{m_n}\ \varrho\ E\left(\sum_{l\in\overline{\mathbb{V}}} \frac{\sum_{j\in\mathbb{V}} \mathcal{K}_{\mathfrak{h}}(\,\cdot\,)(\boldsymbol{X}_j - \boldsymbol{Z}_l)}{\sum_{l\in\overline{\mathbb{V}}} \mathcal{K}_{\mathfrak{h}}(\,\cdot\,)} \mathbb{M}_l\left\{\tau, \boldsymbol{Z}_l, \boldsymbol{\beta}_0(\tau)\right\} + O_p\left(\frac{\mathbf{1}}{n}\right)\right) \\
&= \frac{1-\rho_n}{m_n}\ E\left(\sum_{l\in\overline{\mathbb{V}}} \frac{\sum_{j\in\mathbb{V}} \mathcal{K}_{\mathfrak{h}}(\,\cdot\,)(\boldsymbol{X}_j - \boldsymbol{Z}_l)}{\sum_{l\in\overline{\mathbb{V}}} \mathcal{K}_{\mathfrak{h}}(\,\cdot\,)} \frac{\sum_{l\in\overline{\mathbb{V}}} \mathcal{K}_{\mathfrak{h}}(\,\cdot\,)}{\sum_{j\in\mathbb{V}} \mathcal{K}_{\mathfrak{h}}(\,\cdot\,)} \mathbb{M}_l\left\{\tau, \boldsymbol{Z}_l, \boldsymbol{\beta}_0(\tau)\right\}\right) \\
&= \frac{1-\rho_n}{m_n}\ E\left(\sum_{l\in\overline{\mathbb{V}}} \frac{\sum_{j\in\mathbb{V}} \mathcal{K}_{\mathfrak{h}}(\,\cdot\,)(\boldsymbol{X}_j - \boldsymbol{Z}_l)}{\sum_{j\in\mathbb{V}} \mathcal{K}_{\mathfrak{h}}(\,\cdot\,)} \mathbb{M}_l\left\{\tau, \boldsymbol{Z}_l, \boldsymbol{\beta}_0(\tau)\right\}\right) \\
&= \frac{1-\rho_n}{m_n}\ E\left(\sum_{l\in\overline{\mathbb{V}}} \left[\widehat{\boldsymbol{Z}}_l - \boldsymbol{Z}_l\right] \mathbb{M}_l\left\{\tau, \boldsymbol{Z}_l, \boldsymbol{\beta}_0(\tau)\right\}\right) \\
&= \mathbf{0}. \qquad (3.8)
\end{aligned}
$$

This result is due to the boundedness of $\widehat{\boldsymbol{Z}}_l$ and $\boldsymbol{Z}_l$ and the martingale property.

3.2.1 Asymptotic Properties

Define $F(t\,|\,\cdot) = Pr(Y \le t\,|\,\cdot)$, $\overline{F}(t\,|\,\cdot) = Pr(Y > t\,|\,\cdot)$, $\widetilde{F}(t\,|\,\cdot) = Pr(Y \le t, \delta = 1\,|\,\cdot)$, $\overline{f}(y\,|\,\cdot) = -f(y\,|\,\cdot) = -dF(y\,|\,\cdot)/dy$, $\widetilde{f}(y\,|\,\cdot) = d\widetilde{F}(y\,|\,\cdot)/dy$. (For a vector g, $g^{\otimes 2} = gg^\top$, $g^{(l)} = l^{\text{th}}$ component of g, $\|g\|$ is the Euclidean norm of g.)

Regularity Conditions:

R1: $\sup_i \|\boldsymbol{X}_i\| < \infty$ and $\sup_i \|\boldsymbol{Z}_i\| < \infty$.

R2: (a) Each component of $E\left[\boldsymbol{X}\mathbb{N}\left(e^{\boldsymbol{X}^\top\boldsymbol{\beta}_0(\tau)}\right)\right]$, $E\left[\boldsymbol{Z}\mathbb{N}\left(e^{\boldsymbol{Z}^\top\boldsymbol{\beta}_0(\tau)}\right)\right]$ and $E\left[\boldsymbol{X}\mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\mathbb{N}\left(e^{\boldsymbol{Z}^\top\boldsymbol{\beta}_0(\tau)}\right)\right\} - \mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\boldsymbol{Z}\ \mathbb{N}\left(e^{\boldsymbol{Z}^\top\boldsymbol{\beta}_0(\tau)}\right)\right\}\right]$ is a Lipschitz function of τ.

(b) $\widetilde{f}(t\,|\,\mathbf{x})$ and $f(t\,|\,\mathbf{x})$ are bounded above uniformly in t and $\mathbf{x}$.

(c) $\widetilde{f}(t\,|\,\mathbf{z})$ and $f(t\,|\,\mathbf{z})$ are bounded above uniformly in t and $\mathbf{z}$.

R3: (a) $\widetilde{f}\left(e^{\boldsymbol{X}^\top \boldsymbol{b}} \,\middle|\, \boldsymbol{X}\right) > 0$ and $\widetilde{f}\left(e^{\boldsymbol{Z}^\top \boldsymbol{b}} \,\middle|\, \boldsymbol{Z}\right) > 0$ for all $\boldsymbol{b} \in \mathscr{B}(d_0)$.

(b) To have the positive definiteness, $E\left\{\boldsymbol{X}^{\otimes 2}\right\} > 0$ and $E\left\{\boldsymbol{Z}^{\otimes 2}\right\} > 0$.

(c) Each component of

$$E\left[\boldsymbol{X}^{\otimes 2}\ \overline{f}\left(e^{\boldsymbol{X}^\top \boldsymbol{b}} \,\middle|\, \boldsymbol{X}\right)\ e^{\boldsymbol{X}^\top \boldsymbol{b}}\right] \times \left(E\left[\boldsymbol{X}^{\otimes 2}\ \widetilde{f}\left(e^{\boldsymbol{X}^\top \boldsymbol{b}} \,\middle|\, \boldsymbol{X}\right)\ e^{\boldsymbol{X}^\top \boldsymbol{b}}\right]\right)^{-1},$$

$$E\left[\boldsymbol{Z}^{\otimes 2}\ \overline{f}\left(e^{\boldsymbol{Z}^\top \boldsymbol{b}} \,\middle|\, \boldsymbol{Z}\right)\ e^{\boldsymbol{Z}^\top \boldsymbol{b}}\right] \times \left(E\left[\boldsymbol{Z}^{\otimes 2}\ \widetilde{f}\left(e^{\boldsymbol{Z}^\top \boldsymbol{b}} \,\middle|\, \boldsymbol{Z}\right)\ e^{\boldsymbol{Z}^\top \boldsymbol{b}}\right]\right)^{-1} \text{ and}$$

$$E\left[\boldsymbol{X}^{\otimes 2}\ \mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\overline{f}\left(e^{\boldsymbol{Z}^\top \boldsymbol{b}} \,\middle|\, \boldsymbol{Z}\right)\ e^{\boldsymbol{Z}^\top \boldsymbol{b}}\right\} - \mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\boldsymbol{Z}^{\otimes 2}\ \overline{f}\left(e^{\boldsymbol{Z}^\top \boldsymbol{b}} \,\middle|\, \boldsymbol{Z}\right)\ e^{\boldsymbol{Z}^\top \boldsymbol{b}}\right\}\right] \times$$
$$\left(E\left[\boldsymbol{X}^{\otimes 2}\ \mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\widetilde{f}\left(e^{\boldsymbol{Z}^\top \boldsymbol{b}} \,\middle|\, \boldsymbol{Z}\right)\ e^{\boldsymbol{Z}^\top \boldsymbol{b}}\right\} - \mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\boldsymbol{Z}^{\otimes 2}\ \widetilde{f}\left(e^{\boldsymbol{Z}^\top \boldsymbol{b}} \,\middle|\, \boldsymbol{Z}\right)\ e^{\boldsymbol{Z}^\top \boldsymbol{b}}\right\}\right]\right)^{-1}$$

is uniformly bounded in $\boldsymbol{b} \in \mathscr{B}(d_0)$; $\mathscr{B}(d_0)$ is a neighborhood containing $\{\boldsymbol{\beta}_0(\tau), \tau \in (0, \tau_U]\}$, defined in appendix C.

R4: For any $\nu \in (0, \tau_U]$, $\inf_{\tau \in [\nu, \tau_U]} \text{eigmin}\, E\left[\boldsymbol{X}^{\otimes 2}\ \widetilde{f}\left(e^{\boldsymbol{X}^\top \boldsymbol{\beta}_0(\tau)} \,\middle|\, \boldsymbol{X}\right)\ e^{\boldsymbol{X}^\top \boldsymbol{\beta}_0(\tau)}\right] > 0$,

$\inf_{\tau \in [\nu, \tau_U]} \text{eigmin}\, E\left[\boldsymbol{Z}^{\otimes 2}\ \widetilde{f}\left(e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau)} \,\middle|\, \boldsymbol{Z}\right)\ e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau)}\right] > 0$ and

$\inf_{\tau \in [\nu, \tau_U]} \text{eigmin}\, E\left[\boldsymbol{X}^{\otimes 2}\ \mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\widetilde{f}\left(e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau)} \,\middle|\, \boldsymbol{Z}\right)\ e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau)}\right\}\right.$

$\left. - \mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\boldsymbol{X}^{\otimes 2}\ \widetilde{f}\left(e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau)} \,\middle|\, \boldsymbol{Z}\right)\ e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau)}\right\}\right] > 0$ where eigmin$(\cdot)$ denotes the minimum eigenvalue of a matrix.

Theorem 3.2.1. *Assume that the regularity conditions **R1**-**R4** hold. If* $\lim_{n\to\infty} \|\mathbb{S}_L\| = 0$, *then* $\sup_{\tau \in [\nu, \tau_U]} \left\|\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\right\| \xrightarrow{\Pr} 0$, *where* $0 < \nu < \tau_U$.

Theorem 3.2.2. *Assume that the regularity conditions **R1**-**R4** hold. If* $\lim_{n\to\infty} \sqrt{n}\ \|\mathbb{S}_L\| = 0$, *then* $\sqrt{n}\ \left\{\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\right\}$ *weakly converges to a zero-mean Gaussian process for* $\tau \in [\nu, \tau_U]$, *where* $0 < \nu < \tau_U$.

Proofs of *Theorems* 3.2.1 and 3.2.2 are deferred to Appendices C and D respectively.

3.3 Numerical Studies

We conduct a series of simulation studies to compare the performance of our proposed method with the results of using only the validation sample and the complete case. We used the same simulation models as in the simulation study of the previous chapter.

The logarithmic event times are generated from

$$T_i = \beta_0 + \beta_1 X_{1i} + u_i;\ i = 1, 2, \ldots, n$$

and the logarithmic censoring times are generated from

$$C_i = \gamma_0 + \gamma_1 X_{1i} + v_i;\ i = 1, 2, \ldots, n.$$

where X_{1i}'s are generated from $U[0, 5]$ and u_i's and v_i's are from standard normal distribution. The parameters, $\boldsymbol{\beta}^\top = (5, 1)$ and $\boldsymbol{\gamma}^\top = (6.4, 0.75)$ were selected to maintain a censoring rate of approximately 30%. We assumed 50% of the observations are in the validation sample. We applied the Peng and Huang [2008] estimator for comparison purposes. The performance of our proposed method is compared with the one using only the validation sample as well as with the complete cohort. We used the optimal bandwidth, $\mathfrak{h} = (4/3)^{0.2}\sigma_{\mathbb{V}} n^{-1/5} \approx 1.06\sigma_{\mathbb{V}} n^{-1/5}$ for the Gaussian kernel [Silverman, 1986, p. 45], where $\sigma_{\mathbb{V}}$ is the standard deviation of the residuals from the cubic spline fit between W and X_1 from the validation sample. We generated W from an additive model:

$$W = X_1 + \boldsymbol{\varepsilon},$$

where the error term is generated from $\boldsymbol{\varepsilon} \sim N(0, \sigma^2_{\boldsymbol{\varepsilon}})$ with different $\sigma_{\boldsymbol{\varepsilon}} = 0.2$ and 0.8. We conducted the simulation study with different sample sizes, 200 and 500, and reported the mean bias and root mean squared error (RMSE) measures of the parameters based on 1000 simulations. We used 250 bootstrap samples to estimate the standard error (SE) of the parameter estimates and to compute the coverage probability (CP) of the 95% confidence interval of the model parameters. We conducted the simulations for the 25$^{\text{th}}$ percentile (Table 3.1), 50$^{\text{th}}$ percentile (Table 3.2) and 75$^{\text{th}}$ percentile (Table 3.3).

	Intercept				Slope			
	Bias	RMSE	SE	CP	Bias	RMSE	SE	CP
	$n = 200, \sigma_{\mathcal{E}} = 0.2$							
Complete	0.0173	0.1891	0.1996	96.00	0.0021	0.0682	0.0711	96.40
Proposed	0.0077	0.1948	0.2076	96.30	0.0036	0.0709	0.0738	97.20
Validation	0.0179	0.2687	0.2917	96.10	-0.0005	0.0966	0.1040	96.30
	$n = 200, \sigma_{\mathcal{E}} = 0.8$							
Complete	0.0173	0.1891	0.1996	96.00	0.0021	0.0682	0.0711	96.40
Proposed	-0.0590	0.2335	0.2423	97.00	0.0103	0.0826	0.0876	96.70
Validation	0.0179	0.2687	0.2917	96.10	-0.0005	0.0966	0.1040	96.30
	$n = 500, \sigma_{\mathcal{E}} = 0.2$							
Complete	0.0248	0.1246	0.1248	94.80	-0.0017	0.0429	0.0441	94.80
Proposed	0.0209	0.1270	0.1285	94.60	-0.0021	0.0446	0.0452	95.80
Validation	0.0263	0.1756	0.1783	94.80	0.0001	0.0615	0.0632	95.90
	$n = 500, \sigma_{\mathcal{E}} = 0.8$							
Complete	0.0248	0.1246	0.1248	94.80	-0.0017	0.0429	0.0441	94.80
Proposed	-0.0500	0.1526	0.1490	94.60	0.0071	0.0514	0.0527	96.00
Validation	0.0263	0.1756	0.1783	94.80	0.0001	0.0615	0.0632	95.90

Table 3.1: Comparison between kernel smoothing based approach and validation sample approach using the Bias, RMSE, SE and CP of regression parameters at $\tau = 0.25$

	Intercept				Slope			
	Bias	RMSE	SE	CP	Bias	RMSE	SE	CP
	$\boldsymbol{n = 200, \sigma_{\varepsilon} = 0.2}$							
Complete	0.0221	0.1835	0.1909	95.20	0.0018	0.0671	0.0700	95.40
Proposed	0.0226	0.1868	0.1981	95.60	0.0024	0.0681	0.0725	95.30
Validation	0.0290	0.2558	0.2784	95.40	0.0002	0.0925	0.1027	97.00
	$\boldsymbol{n = 200, \sigma_{\varepsilon} = 0.8}$							
Complete	0.0221	0.1835	0.1909	95.20	0.0018	0.0671	0.0700	95.40
Proposed	0.0159	0.2154	0.2346	94.80	0.0151	0.0815	0.0860	96.80
Validation	0.0290	0.2558	0.2784	95.40	0.0002	0.0925	0.1027	97.00
	$\boldsymbol{n = 500, \sigma_{\varepsilon} = 0.2}$							
Complete	0.0266	0.1249	0.1186	95.20	-0.0011	0.0427	0.0429	94.90
Proposed	0.0281	0.1293	0.1224	94.50	-0.0005	0.0437	0.0445	95.00
Validation	0.0285	0.1753	0.1707	94.30	0.0017	0.0609	0.0620	95.50
	$\boldsymbol{n = 500, \sigma_{\varepsilon} = 0.8}$							
Complete	0.0266	0.1249	0.1186	95.20	-0.0011	0.0427	0.0429	94.90
Proposed	0.0157	0.1435	0.1434	94.20	0.0127	0.0530	0.0521	95.30
Validation	0.0285	0.1753	0.1707	94.30	0.0017	0.0609	0.0620	95.50

Table 3.2: Comparison between kernel smoothing based approach and validation sample approach using the Bias, RMSE, SE and CP of regression parameters at $\tau = 0.5$

	Intercept				Slope			
	Bias	RMSE	SE	CP	Bias	RMSE	SE	CP
	$n = 200, \sigma_{\mathcal{E}} = 0.2$							
Complete	0.0321	0.2105	0.2329	96.10	0.0062	0.0832	0.0976	97.10
Proposed	0.0437	0.2204	0.2454	97.20	0.0079	0.0857	0.0992	97.40
Validation	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	$n = 200, \sigma_{\mathcal{E}} = 0.8$							
Complete	0.0321	0.2105	0.2329	96.10	0.0062	0.0832	0.0976	97.10
Proposed	0.1082	0.2826	0.3186	93.10	0.0268	0.1077	0.1268	97.90
Validation	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
	$n = 500, \sigma_{\mathcal{E}} = 0.2$							
Complete	0.0351	0.1455	0.1439	94.60	-0.0003	0.0526	0.0563	95.90
Proposed	0.0466	0.1560	0.1498	94.20	-0.0012	0.0556	0.0584	95.90
Validation	0.0450	0.2064	0.2083	95.30	0.0028	0.0767	0.0829	96.60
	$n = 500, \sigma_{\mathcal{E}} = 0.8$							
Complete	0.0351	0.1455	0.1439	94.60	-0.0003	0.0526	0.0563	95.90
Proposed	0.0908	0.2028	0.1811	93.50	0.0194	0.0702	0.0711	95.80
Validation	0.0450	0.2064	0.2083	95.30	0.0028	0.0767	0.0829	96.60

Table 3.3: Comparison between kernel smoothing based approach and validation sample approach using the Bias, RMSE, SE and CP of regression parameters at $\tau = 0.75$

From Tables 3.1, 3.2 and 3.3, we can observe that our proposed estimators are asymptotically unbiased. From the values of RMSE and SE, as the measures of dispersion for all three estimates, we can see that our proposed method is very efficient compared to the one using only the validation sample. When the $\sigma^2_{\mathcal{E}}$ is small, our proposed method works almost as well as the 'Complete' case. For $n = 500$, our proposed method provides approximately 95% coverage for the 95% confidence interval. The coverage probability is also competitive as compared to the 'Complete' case, when $n = 200$.

We also observed that using only the validation sample fails to provide estimates for higher quantiles, as in Table 3.3.

In this Section and for the regression calibration based approach in Section 2.3 (at page 25), we considered a linear relationship between W and X_1. From the results, we see that both methods are performing equally well.

3.3.1 Non-Linear Auxiliary Covariate

We conduct simulation studies to compare the performance of the regression calibration based approach and the kernel smoothing based approach when there is a non-linear relationship between X_1 and W. We also used the same simulation models as in Section 3.3 in this simulation study. W is generated from a power model:

$$W = {X_1}^{\mathbf{5}} + \boldsymbol{\varepsilon},$$

where the error term is generated from $\boldsymbol{\varepsilon} \sim N(0, \sigma^2_{\boldsymbol{\varepsilon}})$ with different $\sigma_{\boldsymbol{\varepsilon}} = 0.2$ and 0.8.

We used the bandwidth for the Gaussian kernel, $\mathfrak{h} = 1.06\sigma_{\mathbb{V}} n^{-1/5}$, where $\sigma_{\mathbb{V}}$ is the standard deviation of the residuals from the cubic spline fit between W and X_1 available from the validation sample. We reported the mean bias and root mean squared error (RMSE) measures of the parameters based on 1000 simulations. We used 250 bootstrap samples to estimate the standard error (SE) of the parameter estimates and to compute the coverage probability (CP) of the 95% confidence interval of the model parameters. We conducted simulations for the 25$^{\text{th}}$ percentile (Table 3.4), 50$^{\text{th}}$ percentile (Table 3.5) and 75$^{\text{th}}$ percentile (Table 3.6).

	Intercept				**Slope**			
	Bias	RMSE	SE	CP	Bias	RMSE	SE	CP
	$n = 200, \sigma_{\varepsilon} = 0.2$							
Calibration	-0.0992	0.2747	0.2722	95.40	0.0212	0.0948	0.0978	95.90
Smoothing	0.0113	0.2053	0.2236	96.00	0.0025	0.0842	0.0879	95.60
	$n = 200, \sigma_{\varepsilon} = 0.8$							
Calibration	-0.0993	0.2743	0.2722	95.20	0.0211	0.0946	0.0978	95.80
Smoothing	0.0054	0.2038	0.2223	96.20	0.0048	0.0809	0.0876	96.20
	$n = 500, \sigma_{\varepsilon} = 0.2$							
Calibration	-0.0862	0.1848	0.1683	92.90	0.0157	0.0611	0.0602	94.30
Smoothing	0.0250	0.1328	0.1351	94.60	-0.0006	0.0513	0.0536	96.40
	$n = 500, \sigma_{\varepsilon} = 0.8$							
Calibration	-0.0863	0.1847	0.1683	92.90	0.0157	0.0611	0.0602	94.20
Smoothing	0.0174	0.1325	0.1348	94.40	0.0003	0.0483	0.0523	96.10

Table 3.4: Comparison between regression calibration based approach and kernel smoothing based approach using the Bias, RMSE, SE and CP of regression parameters at $\tau = 0.25$

	Intercept				Slope			
	Bias	RMSE	SE	CP	Bias	RMSE	SE	CP
	$\boldsymbol{n = 200, \sigma_{\varepsilon} = 0.2}$							
Calibration	0.0398	0.2438	0.2600	96.30	0.0197	0.0940	0.0983	95.50
Smoothing	0.0238	0.1924	0.2140	96.70	0.0017	0.0791	0.0870	96.80
	$\boldsymbol{n = 200, \sigma_{\varepsilon} = 0.8}$							
Calibration	0.0402	0.2436	0.2600	96.30	0.0197	0.0938	0.0983	95.60
Smoothing	0.0277	0.1960	0.2124	96.20	-0.0002	0.0772	0.0867	96.90
	$\boldsymbol{n = 500, \sigma_{\varepsilon} = 0.2}$							
Calibration	0.0388	0.1637	0.1604	94.30	0.0182	0.0619	0.0595	93.60
Smoothing	0.0336	0.1364	0.1275	94.50	-0.0015	0.0517	0.0517	95.20
	$\boldsymbol{n = 500, \sigma_{\varepsilon} = 0.8}$							
Calibration	0.0388	0.1637	0.1604	94.20	0.0182	0.0619	0.0595	93.70
Smoothing	0.0344	0.1368	0.1283	94.10	-0.0031	0.0489	0.0507	95.10

Table 3.5: Comparison between regression calibration based approach and kernel smoothing based approach using the Bias, RMSE, SE and CP of regression parameters at $\tau = 0.5$

	Intercept				Slope			
	Bias	RMSE	SE	CP	Bias	RMSE	SE	CP
	$n = 200, \sigma_{\mathcal{E}} = 0.2$							
Calibration	0.1395	0.3457	0.3467	96.00	0.0386	0.1365	0.1462	96.40
Smoothing	0.0524	0.2341	0.2716	97.80	0.0002	0.1032	0.1362	97.70
	$n = 200, \sigma_{\mathcal{E}} = 0.8$							
Calibration	0.1403	0.3467	0.3467	96.10	0.0385	0.1370	0.1462	96.20
Smoothing	0.0566	0.2296	0.2681	97.50	-0.0006	0.0959	0.1267	97.80
	$n = 500, \sigma_{\mathcal{E}} = 0.2$							
Calibration	0.1286	0.2415	0.2133	91.80	0.0345	0.0879	0.0878	94.30
Smoothing	0.0535	0.1607	0.1557	94.40	-0.0039	0.0634	0.0688	96.60
	$n = 500, \sigma_{\mathcal{E}} = 0.8$							
Calibration	0.1285	0.2416	0.2133	92.00	0.0344	0.0881	0.0878	94.40
Smoothing	0.0637	0.1633	0.1553	94.10	-0.0091	0.0602	0.0665	96.80

Table 3.6: Comparison between regression calibration based approach and kernel smoothing based approach using the Bias, RMSE, SE and CP of regression parameters at $\tau = 0.75$

From Tables 3.4, 3.5 and 3.6, we can observe that the kernel smoothing based approach clearly outperforms the regression calibration based approach. The kernel smoothing based approach has a smaller bias, smaller RMSE and smaller SE compared to the regression calibration based approach when there is a non-linear relationship between W and X_1. Since the regression calibration based approach has a high bias, the confidence interval is meaningless. We ignore the comparison between their coverage probabilities.

3.3.2 Colorado Plateau Uranium Miners data

As an illustration, we apply our proposed method to the Colorado Plateau uranium miners cohort data. The major interest of this study was to assess the effect of radon exposure to the observed survival time. This data set consists of 3347 male miners who worked underground for at least one month in the uranium mines of the four-state Colorado Plateau area and who were examined at least once by physicians between 1950 and 1960. For convenience, we removed three individuals with missing ‘status’. The censoring rate of this data is 0.624. Apart from the failure time, the miners’ age, the cumulative radon exposure, cumulative smoking in number of packs and miners' working duration are available. In our study, we randomly chose 1672 miners (50% of total observations) as the validation sample. We assumed that the remaining 1672 individuals belonged to the non-validation sample and assumed that a radon exposure measurement is not available for them.

Similarly to the work of Leng and Tong [2013], we considered three covariates such as the logarithm of the cumulative radon exposure (in 100 WLM), X_1; cumulative smoking in 1000 packs, X_2 and age at entry to the study, X_3. Leng and Tong [2013] pointed out that the log survival time is approximately linear with the covariates only for median regression. To predict the unobserved X_1's, we considered the miners' working duration as the auxiliary covariate (W). The scatter plot in Figure 3.1 shows that it is not certainly linear.

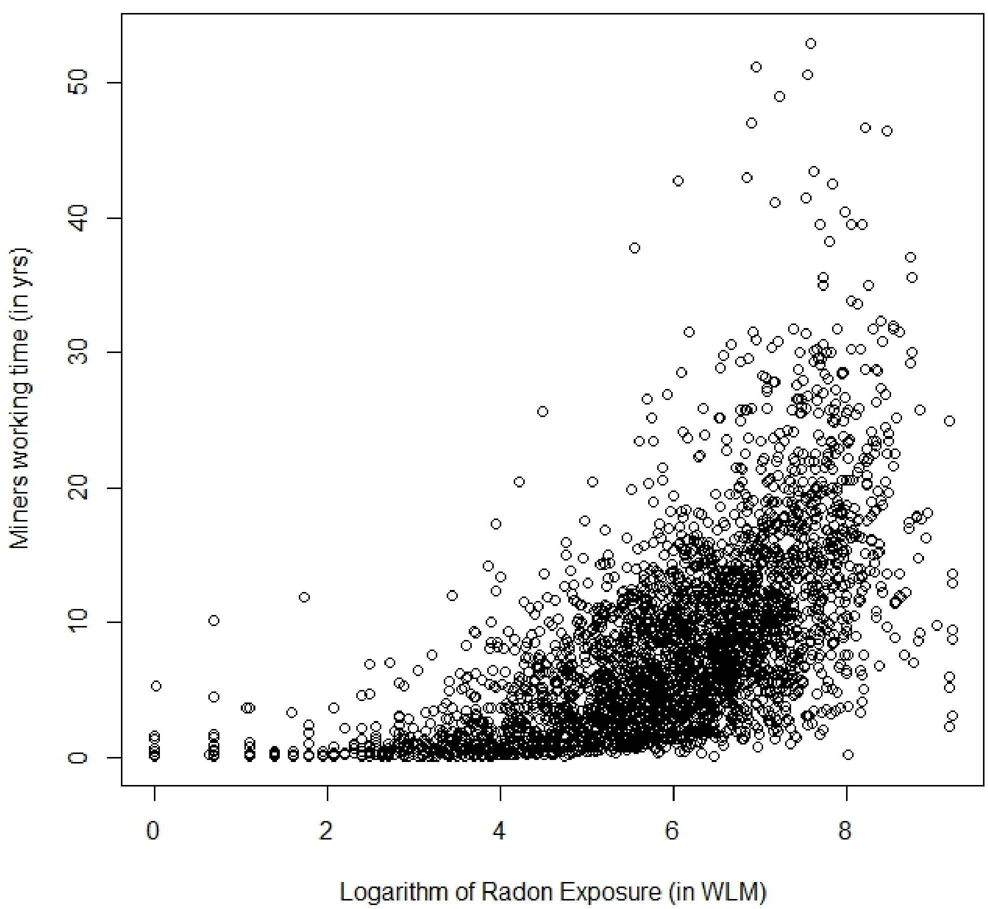

Figure 3.1: Scatter plot of miners' working time and radon exposure

To smooth the unobserved X_1's, we used equation (3.2) with the bandwidth for the Gaussian kernel, $\mathfrak{h} = 1.06\sigma_{\mathbb{V}} n^{-1/5}$, where $\sigma_{\mathbb{V}} \approx 5.44$. After estimating the unobserved X_1's, we fitted the AFT model:

$$\log T = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$

for $\tau = 0.5$ using Peng and Huang [2008] censored quantile regression method. We compared our proposed method with those based on the complete case and the validation sample as in our simulation studies. The results are provided in Table 3.7 and apart from the estimates we produced 95% confidence limits and standard error using the 250 bootstrap samples.

	Estimate	**95% CI**	**SE**
	Intercept		
Complete	4.2670	(4.1752,4.3448)	0.0433
Proposed	4.2540	(4.1625,4.3363)	0.0443
Validation	4.2788	(4.1697,4.3835)	0.0545
	log(Radon)		
Complete	-0.0204	(-0.0297,-0.0114)	0.0047
Proposed	-0.0189	(-0.0307,-0.0095)	0.0054
Validation	-0.0195	(-0.0333,-0.0075)	0.0066
	Smoking		
Complete	-1.6×10^{-5}	(-0.0001,0.0001)	0.0001
Proposed	-2.1×10^{-5}	(-0.0001,0.0001)	0.0001
Validation	-0.6×10^{-5}	(-0.0002,0.0001)	0.0001
	Age		
Complete	0.0024	(0.0014,0.0038)	0.0006
Proposed	0.0024	(0.0015,0.0039)	0.0006
Validation	0.0021	(0.0009,0.0035)	0.0008

Table 3.7: Estimates, SE and 95% CI for regression parameters of Colorado Plateau uranium miners' data at median

From Table 3.7, we can observe that our proposed method has a smaller standard error than that of the validation sample and hence narrower confidence intervals. In the following section, we conduct a study of PBC data discussed in Section 2.3.1.

3.3.3 PBC data

We also used the PBC data set mentioned in the previous chapter to illustrate the performance of the kernel smoothing based method. We used the same model described in Chapter 2. But we smoothed the unobserved values of the logarithm of the copper content in urine, using equation (3.2) with optimum bandwidth for the Gaussian kernel, $\mathfrak{h} = 1.06\sigma_{\mathbb{V}} n^{-1/5}$, where $\sigma_{\mathbb{V}} \approx 0.81$ and fitted the model using Peng and Huang [2008] censored quantile regression method. Results with the kernel smoothing based approach are provided in Table 3.8.

		Validation			**Proposed**		
	$\tau \longrightarrow$	0.25	0.50	0.75	0.25	0.50	0.75
$\log(\hat{\beta})$	Intercept	19.5599	21.2413	23.6345	19.3681	22.0858	20.8564
	Age	-0.6552	-1.3863	-1.4283	-0.6279	-1.5445	-1.6993
	Albumin	2.1459	2.4975	2.1497	1.9140	2.3535	2.3016
	Copper	-0.5672	-0.6215	-0.7266	-0.6454	-0.6528	-0.9076
	Protime	-4.0750	-3.4839	-3.6980	-3.7848	-3.4432	-1.7419
	Edema1	-0.9777	-0.5987	-0.9373	-1.0859	-0.6112	-1.2016
	Edema0.5	-0.6736	0.0496	-0.2515	-0.6592	0.0570	-0.4044
SE	Intercept	1.9879	4.5143	5.9999	2.1339	3.8685	5.6396
	Age	0.3019	0.5948	0.8755	0.2878	0.5647	0.8232
	Albumin	0.4453	0.8927	1.0479	0.4205	0.7920	0.8655
	Copper	0.0762	0.1566	0.2606	0.0875	0.1472	0.2526
	Protime	0.7167	1.6214	2.1205	0.7402	1.3307	1.8631
	Edema1	0.3720	0.4194	0.4133	0.3797	0.3871	0.4361
	Edema0.5	0.2130	0.4667	0.5737	0.1854	0.3460	0.4632
CI	Intercept	(15.45,23.25)	(11.77,29.47)	(9.49,33.01)	(15.41,23.78)	(14.04,29.2)	(11.67,33.78)
	Age	(-1.3,-0.12)	(-2.44,-0.1)	(-2.96,0.47)	(-1.37,-0.24)	(-2.56,-0.35)	(-3.3,-0.08)
	Albumin	(1.31,3.06)	(0.81,4.31)	(0.26,4.37)	(1.2,2.84)	(0.79,3.9)	(0.34,3.74)
	Copper	(-0.71,-0.41)	(-0.91,-0.3)	(-1.13,-0.11)	(-0.78,-0.44)	(-0.99,-0.41)	(-1.25,-0.26)
	Protime	(-5.33,-2.52)	(-6.61,-0.26)	(-7.44,0.87)	(-5.16,-2.26)	(-5.87,-0.65)	(-6.39,0.91)
	Edema1	(-1.71,-0.25)	(-1.54,0.1)	(-1.82,-0.2)	(-1.72,-0.23)	(-1.46,0.06)	(-1.94,-0.23)
	Edema0.5	(-1.02,-0.19)	(-1.13,0.7)	(-1.26,0.98)	(-0.99,-0.26)	(-0.89,0.47)	(-1.14,0.68)

Table 3.8: Estimates, SE and 95% CI for regression parameters of PBC data analysis using kernel smoothing

From the results in Tables 3.8 and 2.4 (page 30), we can observe that the values are almost equal for both regression calibration and kernel smoothing methods. Kernel smoothing based method has smaller standard errors and narrower confidence intervals compared to using only the validation sample. This non-parametric method works as well as the regression calibration based method.

3.4 Discussion

In this chapter, we proposed a semi-parametric method to estimate the censored quantile regression parameters with the auxiliary covariates. We applied the kernel smoothing method to estimate the unobserved covariates using the auxiliary covariates. Then we applied Peng and Huang [2008] censored quantile regression method to the whole study cohort to identify the covariate effect over the observed survival time under heavy censoring for various quantile levels. Our proposed method is more efficient compared to the one using only the validation sample. If the auxiliary covariate and the partially available covariate are closely related, then the performance of our proposed method is close to the one using the completely known study cohort.

Our proposed method performs well for a general relationship between the unobserved covariates and the auxiliary covariates. Numerical results also show that our proposed method works as well as the 'Complete' case if $\sigma^2_{\boldsymbol{\varepsilon}}$ is small. It always outperforms the method using only the validation sample irrespective of the value of $\sigma^2_{\boldsymbol{\varepsilon}}$. We applied our proposed method to the Colorado Plateau uranium miners data with the scenario of variables randomly unavailable as described in Section 3.3.2.

Based on our simulation studies, we suggest the use of the regression calibration based method if the auxiliary covariate has a very strong linear relationship with the unobserved covariate. Zhou and Wang [2000] mentioned that the kernel smoothing based method does not provide stable inference when the dimension of the auxiliary covariate (W) and the covariates correlated with X_1 together are higher than 2. A regression calibration based approach is needed when the dimension of the auxiliary covariate (W) and the covariates correlated with X_1 are higher than 2 and if they are linearly related. But in the general scenario, we would suggest the semi-parametric method which accommodates a more general relationship between the auxiliary covariate and the unobserved variable. We have to be cautious when applying this method to data, especially when we have an extremely small validation sample size, because it may lead to biased estimates.

Chapter 4

Empirical Likelihood based Weighted Censored Quantile Regression

4.1 Introduction

In many studies, auxiliary information about the target population is available from previous studies. For example, in survey sampling, information about the population mean and variance could be available from previous surveys or records. The auxiliary information could be used to improve the efficiency of the statistical inference [Kuk and Mak, 1989; Rao, Kovar and Mantel, 1990; Chen and Qin, 1993].

Consider a known relationship between the survival time, Y (or the failure time, T) and a subset of covariates, say $\boldsymbol{X}_d$,

$$Y = f\left(\boldsymbol{X}_d; \boldsymbol{\theta}\right). \tag{4.1}$$

The knowledge about this relationship can be treated as auxiliary information. In a more general case, the auxiliary information can be expressed as $E\{g(\boldsymbol{Z}; \boldsymbol{\theta})\} = 0$ for some d-dimensional parameter, $\boldsymbol{\theta} \in \mathfrak{R}^d$, where $\boldsymbol{Z}$ is the observed data from the present study and $g(\boldsymbol{Z}; \boldsymbol{\theta}) \in \mathfrak{R}^q$ in some function with $q \geq d$.

The parameter, $\boldsymbol{\theta}$ could be unknown, and estimated using the information available from previous studies.

Chen and Qin [1993] introduced the use of auxiliary information to improve the efficiency of estimators in the context of survey sampling using empirical likelihood [Owen, 1998, 2001]. Li and Wang [2003] accommodated the auxiliary information to the censored linear regression model using empirical likelihood by defining a synthetic variable [Koul, Susarla and Ryzin, 1981]. Fang et al. [2013] proposed the effective use of auxiliary information in the linear regression model with right censored data using empirical likelihood, by utilizing the Buckley-James [Buckley and James, 1979] estimating equation. Tang and Leng [2012] introduced an empirical likelihood (EL) based linear quantile regression model using auxiliary information. In this chapter, we propose an empirical likelihood based approach to accommodate auxiliary information to the censored quantile regression. We utilize the EL based data driven probabilities as the weights by using the estimating function, $g(\boldsymbol{Z};\boldsymbol{\theta})$ and incorporate those weights into the censored quantile regression model.

For the i^{th} $(i = 1, 2, \ldots, n)$ subject, let T_i be the logarithm of the failure time, C_i the logarithm of right censoring time, $\boldsymbol{X}_i$ the p-vector covariate and let $Y_i = \min(T_i, C_i)$ be the logarithm of the survival time. As an extension to the censored quantile regression model in (1.7), for a given quantile, τ, the regression coefficients, $\boldsymbol{\beta}(\tau)$ in the weighted censored quantile regression can be estimated as

$$\widehat{\boldsymbol{\beta}}(\tau) = \underset{\beta \in \Re^p}{\arg\min} \sum_{i=1}^{n} \omega_i \ \rho_\tau \left(Y_i - \min\left\{C_i, {\boldsymbol{X}_i}^\top \boldsymbol{\beta}\right\}\right), \tag{4.2}$$

where ω_i's are the weights. We propose to use the EL based data driven probabilities as the weights. Our simulation results show that the EL based weighted censored quantile regression performs more efficiently than the standard linear censored quantile regression.

The rest of the chapter is organized as follows. In Section 4.2, we present the estimation procedure of the EL based data driven probabilities. In Section 4.3, we introduce the EL based weighted censored quantile regression and investigate the asymptotic properties of the estimators. In Section 4.4, performance analysis of the proposed method is conducted using the simulations. The application to the north central cancer treatment lung cancer data is presented in Section 4.4.4 as an illustration. A brief discussion is provided in Section 4.5.

4.2 Estimation of Weights using Empirical Likelihood

In this section, we develop a method that converts the auxiliary information to the EL based data driven probabilities, which are further used in the weighted censored quantile regression as the weights.

Qin and Lawless [1994] developed the EL based on general estimating equations. For a random sample, $\{T_i, Y_i, \delta_i, \boldsymbol{X}_{di}\}_{i=1}^n$, denote it as $\{\boldsymbol{Z}_i\}_{i=1}^n$ and for an estimating function, $g(\boldsymbol{Z}_i; \boldsymbol{\theta})$ with parameter, $\boldsymbol{\theta}$, the maximum empirical likelihood is given by

$$L_{\text{EL}}(\boldsymbol{\theta}) = \sup \left\{ \prod_{i=1}^n p_i : p_i \geq 0, \sum_{i=1}^n p_i = 1, \sum_{i=1}^n p_i \; g\left(\boldsymbol{Z}_i; \boldsymbol{\theta}\right) = 0 \right\}, \tag{4.3}$$

where $p_i = Pr(Y_i = y_i)$ and $\boldsymbol{\theta}$ is the parameter in the auxiliary information which can be assumed to be known. The parameter, $\boldsymbol{\theta}$ could be any parametric information available from the previous studies which has an influence on the model parameter, $\boldsymbol{\beta}(\tau)$. For a given $g(\boldsymbol{Z}_i; \boldsymbol{\theta})$, $\boldsymbol{\theta}$ should satisfy $E\{g(\boldsymbol{Z}_i; \boldsymbol{\theta})\} = 0$ to avoid the convex hull issues. (This is the scenario for when zero is not an inner point of the convex hull of the $g(\boldsymbol{Z}_i; \boldsymbol{\theta}),\ i = 1, 2, \ldots, n$, which will fail to provide positive p_i's). For a given value of $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, using the Lagrange multiplier method, $L_{\text{EL}}(\boldsymbol{\theta}_0)$ attains its maximum at

$$p_i(\boldsymbol{\theta}_0) = \frac{1}{n\left\{1 + \lambda_{\theta_0}^\top g(\boldsymbol{Z}_i; \boldsymbol{\theta}_0)\right\}}, \quad i = 1, 2, \ldots, n. \tag{4.4}$$

The Lagrange multiplier, $\widehat{\lambda}_{\theta_0}$ is the solution to the equation

$$\sum_{i=1}^n \frac{g(\boldsymbol{Z}_i; \boldsymbol{\theta}_0)}{n\left\{1 + \lambda_{\theta_0}^\top g(\boldsymbol{Z}_i; \boldsymbol{\theta}_0)\right\}} = 0.$$

The estimated $p_i(\cdot)$'s are used as the weights (ω_i) in (4.2) for the weighted censored quantile regression.

In some cases, $\boldsymbol{\theta}$ may not be available. We can use an estimate of $\boldsymbol{\theta}$, say $\widehat{\boldsymbol{\theta}}_A$ obtained from previous studies. Using this $\widehat{\boldsymbol{\theta}}_A$, the new probabilities will be

$$p_i(\widehat{\boldsymbol{\theta}}_A) = \frac{1}{n\left\{1 + \lambda_{\widehat{\theta}_A}^\top g(\boldsymbol{Z}_i; \widehat{\boldsymbol{\theta}}_A)\right\}}, \quad i = 1, 2, \ldots, n. \tag{4.5}$$

The Lagrange multiplier, $\widehat{\lambda}_{\widehat{\theta}_A}$ is the solution to the equation

$$\sum_{i=1}^n \frac{g(\boldsymbol{Z}_i; \widehat{\boldsymbol{\theta}}_A)}{n\left\{1 + \lambda_{\widehat{\theta}_A}^\top g(\boldsymbol{Z}_i; \widehat{\boldsymbol{\theta}}_A)\right\}} = 0.$$

Chen and Qin [1993] and Qin and Lawless [1994] showed that for a random sample,

Y_i, and $p_i(\cdot)$'s are estimated using (4.4), $\widetilde{F}_n(y) = \sum_{i=1}^{n} p_i \mathbb{I}(Y_i \leq y)$ has smaller variance than the empirical distribution function, $\widehat{F}_n(y) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(Y_i \leq y)$. As a result, with Bahadur representation [Bahadur, 1966; Kiefer, 1967], for a given τ $(0 < \tau < 1)$, the quantile estimate, $\widetilde{F}_n^{-1}(\tau)$ has smaller variance than $\widehat{F}_n^{-1}(\tau)$ (See Kuk and Mak [1989] and Rao et al. [1990]). Hence our proposed method is expected to be more efficient than the censored quantile regression ignoring the auxiliary information.

4.3 Estimation of Weighted Censored Quantile Regression Parameters

Define the distribution function of T_i conditional on the p-vector covariate, $\boldsymbol{X}_i$ as $F_{T_i}(t \mid \boldsymbol{X}_i) = Pr(T_i \leq t \mid \boldsymbol{X}_i)$. Let $\Lambda_{T_i}(t \mid \boldsymbol{X}_i) = -\log\{1 - Pr(T_i \leq t \mid \boldsymbol{X}_i)\}$, $\mathbb{N}_i(t) = \mathbb{I}(Y_i \leq t, \delta_i = 1)$, and $\mathbb{M}_i(t) = \mathbb{N}_i(t) - \Lambda_{T_i}(t \wedge Y_i \mid \boldsymbol{X}_i)$. Here $\Lambda_{T_i}(\cdot \mid \boldsymbol{X}_i)$ is the cumulative hazard function conditional on $\boldsymbol{X}_i$, $\mathbb{N}_i(t)$ is the counting process and $\mathbb{M}_i(t)$ is the martingale process associated with $\mathbb{N}_i(t)$ [Fleming and Harrington, 2011]. We consider an extension of Peng and Huang [2008] censored quantile regression estimation procedure. Assuming that p_i's are known and $E\{p_i \ \mathbb{M}_i(t) \mid \boldsymbol{X}_i\} = \mathbf{0}$ (by the martingale property) for $t \geq 0$, we have

$$E\left\{\sqrt{n}\sum_{i=1}^{n} p_i \boldsymbol{X}_i \left(\mathbb{N}_i\left(e^{\boldsymbol{X}_i^\top \boldsymbol{\beta}_0(\tau)}\right) - \Lambda_T\left[e^{\boldsymbol{X}_i^\top \boldsymbol{\beta}_0(\tau)} \wedge Y_i \,\middle|\, \boldsymbol{X}_i\right]\right)\right\} = \mathbf{0}, \qquad (4.6)$$

where $\boldsymbol{\beta}_0(\tau)$ denotes the true $\boldsymbol{\beta}(\tau)$ in (4.2) for a given quantile, τ.

Our weighted censored quantile regression estimating equation is

$$\sqrt{n}\ S_n(\boldsymbol{\beta}, \tau) = \mathbf{0}, \qquad (4.7)$$

where

$$S_n(\boldsymbol{\beta}, \tau) = \sum_{i=1}^{n} p_i \boldsymbol{X}_i \left\{\mathbb{N}_i\left(e^{\boldsymbol{X}_i^\top \boldsymbol{\beta}(\tau)}\right) - \int_0^\tau \mathbb{I}\left[Y_i \geq e^{\boldsymbol{X}_i^\top \boldsymbol{\beta}(u)}\right] dH(u)\right\}.$$

Here p_i's are defined in (4.4) and $H(u) = -\log(1-u)$ for $0 \leq u < 1$. Let $s(\boldsymbol{\beta}, \tau) = E\{S_n(\boldsymbol{\beta}, \tau)\}$. The martingale property of $\mathbb{M}(\cdot)$ gives $s(\boldsymbol{\beta}_0, \tau) = \mathbf{0}$. For a particular quantile, τ_k and an estimator of $\boldsymbol{\beta}_0(\tau_k)$, $\widehat{\boldsymbol{\beta}}(\tau_k)$ is a right-continuous step function which jumps only on a grid, $\mathbb{S}_L = \{0 = \tau_0 < \tau_1 < \cdots < \tau_L = \tau_U < 1\}$. Here L depends on the sample size, n. The size of $\mathbb{S}_L$ is defined as $\|\mathbb{S}_L\| = \max_k(\tau_k - \tau_{k-1})$.

For different quantiles, $\tau_0, \tau_1, \ldots, \tau_L$ $(0 = \tau_0 < \tau_1 < \cdots < \tau_L < 1)$, based on (4.7), we can obtain $\widehat{\boldsymbol{\beta}}(\tau_k)(k = 1, 2, \ldots, L)$ by recursively solving the following monotone estimating equation for $\boldsymbol{\beta}(\tau_k)$:

$$\sqrt{n}\sum_{i=1}^{n} p_i \boldsymbol{X}_i \left\{ \mathbb{N}_i\left(e^{\boldsymbol{X}_i^\top \boldsymbol{\beta}(\tau_k)}\right) - \sum_{r=0}^{k-1} \mathbb{I}\left[Y_i \geq e^{\boldsymbol{X}_i^\top \widehat{\boldsymbol{\beta}}(\tau_r)}\right] \{H(\tau_{r+1}) - H(\tau_r)\} \right\} = \mathbf{0}. \tag{4.8}$$

Similar to previous chapters, we define the estimators, $\widehat{\boldsymbol{\beta}}(\tau_k)$ as the generalized solutions [Fygenson and Ritov, 1994] because equation (4.8) is not continuous and the solution may not be unique.

4.3.1 Asymptotic Theory

We derived the asymptotic properties of the EL based weighted censored quantile regression estimators using the approach of Peng and Huang [2008]. Now we prove the uniform consistency adnd weak Gaussian convergence of the proposed weighted censored quantile regression estimator, $\widehat{\boldsymbol{\beta}}(\cdot)$.

Define $F(t \mid \boldsymbol{X}) = Pr(Y \leq t \mid \boldsymbol{X})$, $\overline{F}(t \mid \boldsymbol{X}) = Pr(Y > t \mid \boldsymbol{X})$, $\widetilde{F}(t \mid \boldsymbol{X}) = Pr(Y \leq t, \delta = 1 \mid \boldsymbol{X})$, $\overline{f}(y \mid \boldsymbol{X}) = -f(y \mid \boldsymbol{X}) = -dF(y \mid \boldsymbol{X})/dy$ and $\widetilde{f}(y \mid \boldsymbol{X}) = d\widetilde{F}(y \mid \boldsymbol{X})/dy$. (For a vector h, $h^{\otimes 2} = hh^T$, $h^{(l)} = l^{th}$ component of h, $\|h\|$ is the Euclidean norm of h.)

Define $\boldsymbol{W}_i = \lambda_{\theta_0}^\top g(\boldsymbol{Z}_i; \boldsymbol{\theta}_0)\boldsymbol{X}_i,\ i = 1, 2, \ldots, n$ as a p-vector.

Regularity Conditions:

R.1 The observations, $\boldsymbol{Z}_i,\ i = 1, 2, \ldots, n$ are iid observations from some distribution. Without loss of generality, we assume that $(Y_i, \delta_i, \boldsymbol{X}_{di}^\top)^\top \subset \boldsymbol{Z}_i$ for all $i = 1, 2, \ldots, n$.

R.2.1: There exists $\boldsymbol{\theta}_0$ such that $E\{g(\boldsymbol{Z}_i; \boldsymbol{\theta}_0)\} = 0$, the matrix $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0) = E\big\{g(\boldsymbol{Z}_i; \boldsymbol{\theta}_0)g(\boldsymbol{Z}_i; \boldsymbol{\theta}_0)^\top\big\}$ is positive definite, $\dfrac{\partial g(\boldsymbol{z}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$ is continuous in the neighborhood of $\boldsymbol{\theta}_0$. The matrix $E\left\{\dfrac{\partial g(\boldsymbol{Z}; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\}$ is of full rank. Furthermore, there exist functions $H_{lj}(\boldsymbol{z})$ such that for $\boldsymbol{\theta}$ in the neighborhood of $\boldsymbol{\theta}_0$:

(a) $\dfrac{\partial g_l(\boldsymbol{z}; \boldsymbol{\theta})}{\partial \theta_j} \leq H_{lj}(\boldsymbol{z})$,

(b) For a constant C, $E\{H_{lj}^2(\boldsymbol{Z})\} \leq C < \infty$ for $l = 1, \ldots q$ and $j = 1, \ldots d$.

R.2.2: $\max_i \|\boldsymbol{X}_i\|^2 = o(\sqrt{n})$ and $\max_i \|\boldsymbol{X}_i Y_{iG}\| = o(\sqrt{n})$, a.s.

R.3: $\sup_i \|\boldsymbol{X}_i\| < \infty$ and $\sup_i \|\boldsymbol{W}_i\| < \infty$.

R.4: (a) Each component of $E\left[\boldsymbol{X}\mathbb{N}\left(e^{\boldsymbol{X}^\top \boldsymbol{\beta}_0(\tau)}\right)\right]$ is a Lipschitz function of τ.

(b) $\widetilde{f}(t \mid \mathbf{x})$ and $f(t \mid \mathbf{x})$ are bounded above uniformly in t and $\mathbf{x}$.

R.5: (a) $\widetilde{f}\left(e^{\boldsymbol{X}^\top \boldsymbol{b}} \,\middle|\, \boldsymbol{X}\right) > 0$ for all $\boldsymbol{b} \in \mathscr{B}(d_0)$.

(b) To have the positive definiteness, $E\{\boldsymbol{X}^{\otimes 2}\} > 0$.

(c) Let $\boldsymbol{\mu}(\boldsymbol{b}) = E\left[\boldsymbol{X}\mathbb{N}\left(e^{\boldsymbol{X}^\top \boldsymbol{b}}\right)\right]$. For $d > 0$, define $\mathscr{B}(d) = \{\boldsymbol{b} \in \Re^p : \inf_{\tau \in (0,\tau_U]} \left\|\boldsymbol{\mu}(\boldsymbol{b}) - \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)\}\right\| \leq d\}$. Each component of $E\left[\boldsymbol{X}^{\otimes 2}\ \overline{f}\left(e^{\boldsymbol{X}^\top \boldsymbol{b}} \,\middle|\, \boldsymbol{X}\right) e^{\boldsymbol{X}^\top \boldsymbol{b}}\right] \times \left(E\left[\boldsymbol{X}^{\otimes 2}\ \widetilde{f}\left(e^{\boldsymbol{X}^\top \boldsymbol{b}} \,\middle|\, \boldsymbol{X}\right) e^{\boldsymbol{X}^\top \boldsymbol{b}}\right]\right)^{-1}$ is uniformly bounded in $\boldsymbol{b} \in \mathscr{B}(d_0)$; $\mathscr{B}(d_0)$ is a neighborhood containing $\{\boldsymbol{\beta}_0(\tau), \tau \in (0, \tau_U]\}$.

R.6: For any $\nu \in (0, \tau_U]$, $\inf_{\tau \in [\nu, \tau_U]}$ eigmin $E\left[\boldsymbol{X}^{\otimes 2}\ \widetilde{f}\left(e^{\boldsymbol{X}^\top \boldsymbol{\beta}_0(\tau)} \,\middle|\, \boldsymbol{X}\right)\ e^{\boldsymbol{X}^\top \boldsymbol{\beta}_0(\tau)}\right] > 0$, where eigmin$(\cdot)$ denotes the minimum eigenvalue of a matrix.

Condition **R.1** implies that $\boldsymbol{Z}_i$ may contain extra variables other than $(Y_i, \boldsymbol{X}_{di}^\top)^\top$ for censored quantile regression. This provides wide acceptability for our proposed method by including more general auxiliary information. For example, in our NCCTG data analysis (Section 4.4.4, Page 82), we considered only the continuous variables for the auxiliary information. The standard error was reduced not only for the parameter estimates corresponding to the continuous variables, but also was reduced for the parameter estimates corresponding to the other variables.

Theorem 4.3.1. *Assuming that the regularity conditions* ***R.1-R.6*** *hold, if* $\lim_{n\to\infty} \|\mathbb{S}_L\| = 0$*, then* $\sup_{\tau \in [\nu, \tau_U]} \|\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\| \xrightarrow{\Pr} 0$*, where* $0 < \nu < \tau_U$*.*

Theorem 4.3.2. *Assuming that the regularity conditions* ***R.1-R.6*** *hold, if* $\lim_{n\to\infty} n^{1/2} \|\mathbb{S}_L\| = 0$*, then* $n^{1/2}\{\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\}$ *weakly converges to a zero-mean Gaussian process for* $\tau \in [\nu, \tau_U]$*, where* $0 < \nu < \tau_U$*.*

To prove *Theorems* 4.3.1 and 4.3.2, we need to show that $\max_{1 \leq i \leq n} |\lambda_{\theta_0}^\top g(\boldsymbol{Z}_i; \boldsymbol{\theta}_0)| = o_p(1)$. We consider two different types of $g(\boldsymbol{Z}_i; \boldsymbol{\theta})$. First, $g(\boldsymbol{Z}_i; \boldsymbol{\theta})$ does not contain the censored observations, as given in (4.10). The second, $g(\boldsymbol{Z}_i;\boldsymbol{\theta})$, contains the censored observations, as given in (4.14).

In the case of uncensored observations, by Owen [1991] and *Lemma* 11.2 of Owen [2001], we have $\max_{1\le i\le n} ||g(\boldsymbol{Z}_i;\boldsymbol{\theta}_0)|| = o_p(\sqrt{n})$. By *Lemma* 1 of Tang and Leng [2012], we have under the regularity condition **R.2.1**; the λ_{θ_0} in (4.4) satisfies $\|\lambda_{\theta_0}\| = O_p\left(\frac{1}{\sqrt{n}}\right)$. So,

$$\max_{1\le i\le n}\left|\lambda_{\theta_0}^\top g(\boldsymbol{Z}_i;\boldsymbol{\theta}_0)\right| = O_p\left(\frac{1}{\sqrt{n}}\right) o_p\left(\sqrt{n}\right) = o_p(1). \tag{4.9}$$

Under the condition **R.2.2**; Qin and Jing [2001] proved $\max_{1\le i\le n}|\lambda_{\theta_0}^\top g(\boldsymbol{Z}_i;\boldsymbol{\theta}_0)| = o_p(1)$ for the $g(\cdot)$ with censored observations.

Now following Owen [2001], using Taylor's series expansion of the weights, p_i's defined in (4.4) can be rewritten as,

$$\begin{aligned}
p_i(\boldsymbol{\theta}_0) &= \frac{1}{n\left\{1+\lambda_{\theta_0}^\top g(\boldsymbol{Z}_i;\boldsymbol{\theta}_0)\right\}}\\
&= \frac{1}{n}\left[1-\lambda_{\theta_0}^\top g(\boldsymbol{Z}_i;\boldsymbol{\theta}_0)\{1+o_p(1)\}\right]\\
&= \frac{1}{n}\left[1-\lambda_{\theta_0}^\top g(\boldsymbol{Z}_i;\boldsymbol{\theta}_0)\right] + o_p\left(\frac{1}{n}\right);\ i=1,2,\dots n.
\end{aligned}$$

Now we rewrite the $S_n(\boldsymbol{\beta},\tau)$ as

$$\begin{aligned}
S_n(\boldsymbol{\beta},\tau) &= \frac{1}{n}\sum_{i=1}^n \left[1-\lambda_{\theta_0}^\top g(\boldsymbol{Z}_i;\boldsymbol{\theta}_0)\right]\boldsymbol{X}_i\Bigg\{\mathbb{N}_i\left(e^{\boldsymbol{X}_i^\top\boldsymbol{\beta}(\tau)}\right)\\
&\qquad - \int_0^\tau \mathbb{I}\left[Y_i \ge e^{\boldsymbol{X}_i^\top\boldsymbol{\beta}(u)}\right] dH(u)\Bigg\} + o_p\left(\frac{1}{n}\right)\\
&= \frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i\left\{\mathbb{N}_i\left(e^{\boldsymbol{X}_i^\top\boldsymbol{\beta}(\tau)}\right) - \int_0^\tau \mathbb{I}\left[Y_i \ge e^{\boldsymbol{X}_i^\top\boldsymbol{\beta}(u)}\right] dH(u)\right\}\\
&\quad - \frac{1}{n}\sum_{i=1}^n \lambda_{\theta_0}^\top g(\boldsymbol{Z}_i;\boldsymbol{\theta}_0)\boldsymbol{X}_i\left\{\mathbb{N}_i\left(e^{\boldsymbol{X}_i^\top\boldsymbol{\beta}(\tau)}\right) - \int_0^\tau \mathbb{I}\left[Y_i \ge e^{\boldsymbol{X}_i^\top\boldsymbol{\beta}(u)}\right] dH(u)\right\}\\
&\qquad + o_p\left(\frac{1}{n}\right)\\
&= \frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_i\left\{\mathbb{N}_i\left(e^{\boldsymbol{X}_i^\top\boldsymbol{\beta}(\tau)}\right) - \int_0^\tau \mathbb{I}\left[Y_i \ge e^{\boldsymbol{X}_i^\top\boldsymbol{\beta}(u)}\right] dH(u)\right\}\\
&\quad - \frac{1}{n}\sum_{i=1}^n \boldsymbol{W}_i\left\{\mathbb{N}_i\left(e^{\boldsymbol{X}_i^\top\boldsymbol{\beta}(\tau)}\right) - \int_0^\tau \mathbb{I}\left[Y_i \ge e^{\boldsymbol{X}_i^\top\boldsymbol{\beta}(u)}\right] dH(u)\right\} + o_p\left(\frac{1}{n}\right).
\end{aligned}$$

Asymptotically, by (4.9), we have $\|\boldsymbol{W}_i\| = o_p(1);\ i = 1, 2, \ldots, n$. So,

$$S_n(\boldsymbol{\beta}, \tau) = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i \left\{ \mathbb{N}_i\left(e^{\boldsymbol{X}_i^\top \boldsymbol{\beta}(\tau)}\right) - \int_0^\tau \mathbb{I}\left[Y_i \geq e^{\boldsymbol{X}_i^\top \boldsymbol{\beta}(u)}\right] dH(u) \right\} + o_p\left(\frac{1}{n}\right).$$

Asymptotically our estimating function, $S_n(\boldsymbol{\beta}, \tau)$ is the same as Peng and Huang [2008]. Following the similar arguments of Peng and Huang [2008], the proofs of Theorems 4.3.1 and 4.3.2 are straightforward, so we ignore the remaining proof.

4.4 Numerical Analysis

We conduct extensive simulation studies to compare the performance between our proposed EL based weighted censored quantile regression estimator and the standard censored quantile regression estimator. For our simulation, we use similar models discussed in Tang and Leng [2012].

The simulation models used to generate the logarithmic event time (T_r) and logarithmic censoring time (C_r) for the r^{th} ($r = 1, 2, \ldots, N$) subject are given in Table 4.1 under four Cases (i)-(iv).

Cases	Models	Error Distribution
(i)	$T_r = \theta_0 + \theta_1 x_{1r} + \theta_2 x_{2r} + u_r,$ $C_r = \gamma_0 + \gamma_1 x_{1r} + \gamma_2 x_{2r} + v_r.$	$u_r,\ v_r \sim N(0, 1)$
(ii)	$T_r = \theta_0 + \theta_1 x_{1r} + \theta_2 x_{2r} + u_r,$ $C_r = \gamma_0 + \gamma_1 x_{1r} + \gamma_2 x_{2r} + v_r.$	$u_r,\ v_r \sim t(3)$
(iii)	$T_r = \theta_0 + \theta_1 x_{1r} + \theta_2 x_{2r} + (\pi_0 + \pi_0 x_{1r} + \pi_2 x_{2r}) u_r,$ $C_r = \gamma_0 + \gamma_1 x_{1r} + \gamma_2 x_{2r} + (\pi_0 + \pi_0 x_{1r} + \pi_2 x_{2r}) v_r.$	$u_r,\ v_r \sim N(0, 1)$
(iv)	$T_r = \theta_0 + \theta_1 x_{1r} + \theta_2 x_{2r} + (\pi_0 + \pi_0 x_{1r} + \pi_2 x_{2r}) u_r,$ $C_r = \gamma_0 + \gamma_1 x_{1r} + \gamma_2 x_{2r} + (\pi_0 + \pi_0 x_{1r} + \pi_2 x_{2r}) v_r.$	$u_r,\ v_r \sim t(3)$

Table 4.1: Four simulation models to generate event and censoring times

In Cases (i) and (ii), event times and censoring times are generated from the homoscedastic models and in Cases (iii) and (iv), we considered heteroscedastic models to examine the efficiency gain of our proposed method over the standard censored quantile regression. We set the parameter values as $\boldsymbol{\theta}^\top = (0, -1, 0.2)$, $\boldsymbol{\pi}^\top = (0.3, -0.1, 0.1)$ and selected $\boldsymbol{\gamma}^\top$ to maintain approximately 30% of the censoring proportion in each case. We generated explanatory variables from zero mean bivariate normal distribution with covariance, $\Sigma = \begin{bmatrix} 1 & \sigma_{x_1,x_2} \\ \sigma_{x_1,x_2} & 1 \end{bmatrix}$ and $\sigma_{x_1,x_2} = 0$ & 0.5.

We considered two different ways to compute the EL based probability weights. In numerical study - I, we compute p_i's based on the auxiliary information related to the failure time, T_i, whereas in numerical study - II, p_i's are computed using the observed survival time, $Y_i = \min(T_i, C_i)$. In numerical study -II, we employ the synthetic variable approach [Koul et al., 1981; Qin and Jing, 2001; Li and Wang, 2003] to compute the EL based data driven probability weights.

4.4.1 Numerical Study - I

4.4.1.A Auxiliary information based on both x_1 and x_2

To compute p_i's, first we need to have a known population parameter, $\boldsymbol{\theta}$, or its estimate. We considered a linear relation between T and $\boldsymbol{X} = (X_1, X_2)$ with slopes (θ_1 and θ_2) and intercept (θ_0) as the auxiliary information. We estimated $\boldsymbol{\theta}$ using the standard linear regression (least square) based on a large, finite population with size, $N = 10000$. We need to generate censoring times as well to compute the event indicator, $\delta_i = \mathbb{I}(T_i \leq C_i)$ and survival time, $Y_i = \min(T_i, C_i)$ to estimate the censored quantile regression parameters. To fit the weighted censored quantile regression model given in (4.2), we generated another n observations $\{y_i, \boldsymbol{x}_i\}_{i=1}^n$ with $n \ll N$, using the same models given in Table 4.1. We considered the sample sizes, $n = 100$ and 200 and three quantiles, $\tau = 0.25,\ 0.5,\ 0.75$. For our proposed method, we estimated p_i's using the estimating function, $g(t_i, \boldsymbol{x}_i; \boldsymbol{\theta})$ defined based on the normal equations of the linear least squares method [Owen, 1991].

$$g_i(\boldsymbol{z}_i; \boldsymbol{\theta}) = g(t_i, \boldsymbol{x}_i; \boldsymbol{\theta}) = \boldsymbol{x}_i(t_i - \boldsymbol{x}_i^\top \widehat{\boldsymbol{\theta}}),\ i = 1, 2, \ldots, n. \tag{4.10}$$

For a given quantile, τ, the true value of the censored quantile regression parameters $\boldsymbol{\beta}_0(\tau)$ are estimated from the population of size, $N = 10000$. In general, under a linear model assumption, the true value of the censored quantile regression slope parameters are the same as the $\boldsymbol{\theta}$ (i.e, $\beta_1(\tau) = \theta_1, \beta_2(\tau) = \theta_2$). But for the intercept, it is $\beta_0(\tau) = \theta_0 + F^{-1}(\tau)$, where F is the error distribution.

a) Independent covariates: In this case, we generated the covariates assuming $\sigma_{x_1,x_2} = 0$. We conducted 1000 simulations and computed mean bias, standard error

(SE) and 95% coverage probability (CP) of the model parameter estimates for different sample sizes using 250 bootstrap samples. We compared the performance of our proposed method (CQR-EL1) with the standard censored quantile regression (CQR) model. We present the simulation results in Tables 4.2 to 4.5 respectively for Cases (i)-(iv) with $\sigma_{x_1,x_2} = 0$.

	n		**CQR**			**CQR-EL1**		
		$\tau \rightarrow$	0.25	0.50	0.75	0.25	0.50	0.75
Bias	100	β_0.	0.0042	0.0170	0.0647	0.0027	0.0180	0.0771
		β_1.	0.0029	0.0035	0.0094	-0.0014	-0.0048	0.0030
		β_2.	-0.0049	-0.0141	-0.0100	-0.0047	-0.0124	-0.0171
	200	β_0	0.0218	0.0298	0.0501	0.0199	0.0322	0.0635
		β_1	0.0016	0.0026	0.0057	0.0008	0.0028	0.0048
		β_2	-0.0020	-0.0032	-0.0078	-0.0010	0.0001	-0.0071
SE	100	β_0.	0.1449	0.1404	0.2268	0.1103	0.1086	0.2110
		β_1.	0.1533	0.1515	0.2141	0.1159	0.1109	0.2000
		β_2.	0.1519	0.1525	0.2198	0.1149	0.1109	0.2082
	200	β_0	0.0973	0.0929	0.1292	0.0720	0.0703	0.1221
		β_1	0.1040	0.1029	0.1341	0.0746	0.0718	0.1173
		β_2	0.1041	0.1027	0.1354	0.0752	0.0717	0.1177
CP	100	β_0.	93.3	93.4	95.7	95.8	96.6	97.0
		β_1.	94.7	95.8	96.5	95.1	96.2	97.9
		β_2.	96.0	96.3	96.4	96.4	96.4	96.9
	200	β_0	92.3	91.9	92.7	92.7	92.5	94.8
		β_1	94.5	96.2	95.0	95.0	95.5	96.9
		β_2	93.6	95.0	95.2	94.2	94.9	95.8

Table 4.2: Bias, SE and CP of regression parameters for Case (i) model with independent covariates ($\sigma_{x_1,x_2} = 0$)

	n	$\tau \rightarrow$	CQR			CQR-EL1		
			0.25	0.50	0.75	0.25	0.50	0.75
Bias	100	β_0.	0.0105	0.0288	0.1088	0.0119	0.0270	0.1062
		β_1.	0.0063	0.0214	0.0169	0.0005	0.0102	0.0066
		β_2.	0.0164	0.0096	-0.0170	0.0152	0.0079	-0.0184
	200	β_0	0.0267	0.0355	0.0821	0.0276	0.0340	0.0760
		β_1	0.0006	-0.0032	0.0050	0.0042	0.0032	0.0024
		β_2	0.0112	0.0025	0.0051	0.0029	-0.0038	-0.0057
SE	100	β_0.	0.1871	0.1538	0.2980	0.1522	0.1304	0.2914
		β_1.	0.1946	0.1664	0.2698	0.1555	0.1318	0.2480
		β_2.	0.1955	0.1676	0.2733	0.1556	0.1327	0.2543
	200	β_0	0.1235	0.1029	0.1621	0.0998	0.0871	0.1556
		β_1	0.1301	0.1146	0.1663	0.1010	0.0893	0.1473
		β_2	0.1315	0.1149	0.1671	0.1023	0.0897	0.1465
CP	100	β_0.	95.5	93.1	94.7	96.2	94.8	97.2
		β_1.	95.6	93.5	96.4	95.7	95.6	97.8
		β_2.	95.9	95.4	96.4	96.0	95.0	97.2
	200	β_0	93.1	91.2	94.0	93.0	93.8	95.7
		β_1	95.0	95.5	95.4	94.8	95.5	96.2
		β_2	95.5	95.7	95.5	95.0	95.2	96.3

Table 4.3: Bias, SE and CP of regression parameters for Case (ii) model with independent covariates ($\sigma_{x_1,x_2} = 0$)

	n	$\tau \rightarrow$	**CQR**			**CQR-EL1**		
			0.25	0.50	0.75	0.25	0.50	0.75
Bias	100	β_0.	0.0062	0.0088	0.0224	0.0055	0.0085	0.0254
		β_1.	0.0042	0.0051	0.0076	0.0034	0.0016	0.0057
		β_2.	-0.0038	-0.0039	-0.0069	-0.0013	0.0003	-0.0010
	200	β_0	0.0064	0.0072	0.0167	0.0064	0.0089	0.0195
		β_1	0.0012	0.0038	0.0033	-0.0006	-0.0003	-0.0014
		β_2	-0.0015	-0.0031	-0.0017	-0.0004	0.0002	0.0023
SE	100	β_0.	0.0472	0.0466	0.0767	0.0349	0.0338	0.0737
		β_1.	0.0566	0.0570	0.0796	0.0424	0.0411	0.0708
		β_2.	0.0567	0.0575	0.0807	0.0425	0.0418	0.0720
	200	β_0	0.0313	0.0301	0.0402	0.0225	0.0213	0.0345
		β_1	0.0371	0.0377	0.0489	0.0276	0.0267	0.0402
		β_2	0.0367	0.0376	0.0488	0.0270	0.0267	0.0401
CP	100	β_0.	94.4	95.0	96.1	94.3	96.0	97.1
		β_1.	95.0	95.2	95.5	95.2	95.3	97.4
		β_2.	96.6	96.7	97.3	95.4	96.6	98.0
	200	β_0	94.1	93.4	94.9	93.2	94.0	94.1
		β_1	94.0	94.9	96.0	93.0	95.1	95.9
		β_2	94.6	95.0	95.3	94.4	95.3	94.8

Table 4.4: Bias, SE and CP of regression parameters for Case (iii) model with independent covariates ($\sigma_{x_1,x_2} = 0$)

	n	$\tau \rightarrow$	CQR 0.25	CQR 0.50	CQR 0.75	CQR-EL1 0.25	CQR-EL1 0.50	CQR-EL1 0.75
Bias	100	β_0.	0.0066	0.0097	0.0364	0.0048	0.0076	0.0273
		β_1.	0.0031	0.0039	0.0041	0.0026	0.0043	0.0036
		β_2.	0.0008	-0.0009	-0.0018	0.0008	-0.0035	-0.0028
	200	β_0	0.0083	0.0089	0.0243	0.0100	0.0103	0.0258
		β_1	-0.0020	0.0016	0.0017	-0.0022	-0.0008	-0.0018
		β_2	0.0008	-0.0012	-0.0031	0.0026	0.0012	0.0004
SE	100	β_0.	0.0600	0.0507	0.1103	0.0466	0.0407	0.1038
		β_1.	0.0667	0.0592	0.0993	0.0514	0.0468	0.0885
		β_2.	0.0677	0.0600	0.1014	0.0525	0.0470	0.0921
	200	β_0	0.0395	0.0327	0.0521	0.0305	0.0260	0.0464
		β_1	0.0429	0.0386	0.0568	0.0331	0.0298	0.0491
		β_2	0.0429	0.0389	0.0580	0.0331	0.0301	0.0501
CP	100	β_0.	93.5	95.0	97.7	94.7	95.5	97.8
		β_1.	95.6	96.6	97.0	96.0	96.3	97.3
		β_2.	96.0	96.2	97.3	95.8	96.7	97.0
	200	β_0	93.0	93.9	94.9	93.5	93.4	94.1
		β_1	95.6	95.8	94.7	94.5	95.2	95.4
		β_2	94.5	95.9	95.5	94.5	96.0	95.2

Table 4.5: Bias, SE and CP of regression parameters for Case (iv) model with independent covariates ($\sigma_{x_1,x_2} = 0$)

From Tables 4.2-4.5, we see that our proposed estimator has approximately zero bias. A comparison of SE of CQR-EL1 with CQR indicates that the SE of CQR-EL1 reduces remarkably for all the parameters irrespective of any quantile. For example, we consider the scenario of $n = 100$ and $\tau = 0.25$ for comparison purposes throughout this section. From Table 4.2, for CQR, SE of $\widehat{\beta}_1$ is 0.1533 and for CQR-EL1, SE of $\widehat{\beta}_1$ is reduced to 0.1159. When the sample size is increased to 200, SE of $\widehat{\beta}_1$ of our proposed method further is reduced to 0.0746. If we compare the CP of our proposed method with the nominal level of 95%, CQR-EL1 provides approximately 95% coverage and becomes more stable when the sample size increases. Similar conclusions can be reached for Case (ii) (results are in Table 4.3) even though we considered heavy tailed distribution for the failure time compared to Case (i). For example, SE of $\widehat{\beta}_1$ using CQR is 0.1946, whereas it is only 0.1555 for the CQR-EL1 based estimate. We also observed that SE is comparatively high in Case (ii) compared to Case(i).

In Cases (iii) and (iv), the error depends on the covariates. Simulation results for these Cases (Tables 4.4 and 4.5) are almost similar to the cases where error is independent of covariates. For example, in Case (iii) (Table 4.4), SE of $\widehat{\beta}_1$ is 0.0566 and 0.0424 for CQR and CQR-EL1 respectively. Similarly, in Case (iv) (Table 4.5), SE of $\widehat{\beta}_1$ is 0.0667 and 0.0514 for CQR and CQR-EL1 respectively. Here, we could also see a slight increase in the SE of estimates for Case (iv) because of the heavy tailed distribution assumption for the failure time compared to Case (iii).

b) Dependent covariates: Next we consider the effect of correlation between the covariates regarding the efficiency of our proposed estimators, and generated covariates with $\sigma_{x_1,x_2} = 0.5$. We present the simulation results in Tables 4.6 to 4.9 respectively for Cases (i)-(iv).

	n	$\tau \rightarrow$	**CQR** 0.25	0.50	0.75	**CQR-EL1** 0.25	0.50	0.75
Bias	100	β_0.	0.0120	0.0179	0.0588	0.0101	0.0261	0.0743
		β_1.	0.0006	0.0016	0.0134	0.0046	-0.0025	0.0101
		β_2.	0.0001	-0.0047	-0.0146	0.0025	-0.0004	-0.0168
	200	β_0	0.0268	0.0284	0.0487	0.0235	0.0280	0.0559
		β_1	-0.0032	-0.0001	-0.0037	0.0018	0.0050	0.0004
		β_2	-0.0006	-0.0072	-0.0030	-0.0007	-0.0049	-0.0058
SE	100	β_0.	0.1438	0.1386	0.2225	0.1093	0.1079	0.2230
		β_1.	0.1788	0.1759	0.2493	0.1324	0.1286	0.2324
		β_2.	0.1768	0.1749	0.2543	0.1325	0.1276	0.2362
	200	β_0	0.0972	0.0922	0.1273	0.0726	0.0699	0.1204
		β_1	0.1197	0.1193	0.1543	0.0865	0.0835	0.1337
		β_2	0.1203	0.1193	0.1553	0.0871	0.0834	0.1348
CP	100	β_0.	94.0	93.4	95.4	95.7	96.1	97.4
		β_1.	95.4	96.7	95.7	95.8	97.0	98.1
		β_2.	95.9	96.4	96.3	96.8	96.5	97.9
	200	β_0	93.3	92.1	94.6	93.1	93.8	96.2
		β_1	94.6	94.4	94.7	94.0	94.5	94.6
		β_2	95.0	94.4	95.5	95.8	94.5	96.2

Table 4.6: Bias, SE and CP of regression parameters for Case (i) model with dependent covariates ($\sigma_{x_1,x_2} = 0.5$)

	n	$\tau \rightarrow$	CQR			CQR-EL1		
			0.25	0.50	0.75	0.25	0.50	0.75
Bias	100	β_0.	0.0092	0.0301	0.1157	0.0197	0.0306	0.1136
		β_1.	0.0241	0.0040	-0.0053	0.0257	0.0050	0.0109
		β_2.	-0.0140	-0.0102	-0.0016	-0.0158	-0.0043	-0.0186
	200	β_0	0.0264	0.0258	0.0605	0.0286	0.0249	0.0622
		β_1	0.0027	0.0004	0.0034	0.0093	0.0008	0.0045
		β_2	-0.0010	-0.0017	-0.0066	-0.0076	-0.0006	0.0003
SE	100	β_0.	0.1868	0.1530	0.2943	0.1507	0.1300	0.2909
		β_1.	0.2261	0.1970	0.3164	0.1777	0.1534	0.2872
		β_2.	0.2261	0.1962	0.3163	0.1766	0.1542	0.2941
	200	β_0	0.1228	0.1007	0.1619	0.0995	0.0859	0.1581
		β_1	0.1495	0.1307	0.1938	0.1159	0.1014	0.1709
		β_2	0.1497	0.1305	0.1960	0.1164	0.1018	0.1731
CP	100	β_0.	94.7	93.8	95.9	95.5	95.4	97.8
		β_1.	95.7	96.6	96.6	95.7	95.4	97.1
		β_2.	96.1	95.5	97.2	95.6	96.3	98.1
	200	β_0	91.7	92.9	93.4	93.0	94.3	95.5
		β_1	96.4	96.2	96.4	96.3	95.1	95.8
		β_2	95.2	95.0	96.0	94.2	96.0	96.8

Table 4.7: Bias, SE and CP of regression parameters for Case (ii) model with dependent covariates ($\sigma_{x_1,x_2} = 0.5$)

	n	$\tau \rightarrow$	**CQR** 0.25	0.50	0.75	**CQR-EL1** 0.25	0.50	0.75
Bias	100	β_0.	0.0067	0.0104	0.0202	0.0051	0.0090	0.0228
		β_1.	0.0037	0.0040	0.0091	0.0020	0.0016	0.0041
		β_2.	-0.0013	-0.0048	-0.0105	-0.0012	-0.0006	-0.0037
	200	β_0	0.0073	0.0092	0.0182	0.0066	0.0093	0.0184
		β_1	0.0010	0.0025	0.0030	-0.0006	-0.0009	-0.0015
		β_2	-0.0006	-0.0021	-0.0041	0.0007	0.0012	0.0005
SE	100	β_0.	0.0458	0.0440	0.0770	0.0341	0.0325	0.0760
		β_1.	0.0604	0.0607	0.0877	0.0457	0.0449	0.0814
		β_2.	0.0604	0.0610	0.0894	0.0454	0.0449	0.0830
	200	β_0	0.0308	0.0293	0.0400	0.0222	0.0213	0.0358
		β_1	0.0398	0.0409	0.0547	0.0292	0.0293	0.0463
		β_2	0.0396	0.0411	0.0549	0.0290	0.0291	0.0469
CP	100	β_0.	94.6	93.9	96.2	95.7	94.9	97.8
		β_1.	96.6	95.9	97.1	96.0	96.6	97.9
		β_2.	96.7	96.1	97.2	95.9	95.9	97.9
	200	β_0	94.1	92.8	93.8	94.3	93.6	95.3
		β_1	95.8	95.1	95.5	95.6	95.1	95.4
		β_2	95.0	94.3	93.9	94.8	95.2	96.1

Table 4.8: Bias, SE and CP of regression parameters for Case (iii) model with dependent covariates ($\sigma_{x_1,x_2} = 0.5$)

	n	$\tau \rightarrow$	CQR 0.25	CQR 0.50	CQR 0.75	CQR-EL1 0.25	CQR-EL1 0.50	CQR-EL1 0.75
Bias	100	β_0.	0.0068	0.0122	0.0332	0.0071	0.0124	0.0312
		β_1.	-0.0006	0.0045	0.0115	-0.0036	-0.0016	0.0060
		β_2.	-0.0000	-0.0045	-0.0118	0.0020	0.0014	-0.0048
	200	β_0	0.0075	0.0083	0.0226	0.0107	0.0108	0.0261
		β_1	-0.0010	0.0013	0.0034	-0.0042	-0.0017	-0.0016
		β_2	0.0014	-0.0003	-0.0026	0.0040	0.0025	0.0015
SE	100	β_0.	0.0581	0.0488	0.1093	0.0454	0.0393	0.1035
		β_1.	0.0723	0.0655	0.1118	0.0557	0.0508	0.1013
		β_2.	0.0726	0.0661	0.1144	0.0553	0.0510	0.1035
	200	β_0	0.0384	0.0316	0.0518	0.0301	0.0259	0.0473
		β_1	0.0477	0.0422	0.0644	0.0368	0.0330	0.0561
		β_2	0.0470	0.0427	0.0645	0.0362	0.0333	0.0564
CP	100	β_0.	94.3	93.0	97.1	95.6	95.5	98.8
		β_1.	95.3	96.6	96.5	95.8	95.7	98.0
		β_2.	96.4	95.7	97.3	95.4	95.6	97.9
	200	β_0	93.8	92.4	95.3	93.3	92.8	94.2
		β_1	94.4	94.7	95.4	95.1	95.3	95.4
		β_2	94.3	96.2	96.7	95.6	95.3	96.0

Table 4.9: Bias, SE and CP of regression parameters for Case (iv) model with dependent covariates ($\sigma_{x_1,x_2} = 0.5$)

We presented the simulation results with the correlated covariates in Tables 4.6 to 4.9 for Cases (i)-(iv) respectively. For our proposed method, we observed similar results as those for the uncorrelated covariates such as negligible bias, smaller SE for all the parameter estimates including intercept compared to CQR and approximately 95% coverage probability. For the simulation results in Tables 4.2 to 4.9, we utilized the auxiliary information in relation to both x_1 and x_2 in the censored quantile regression estimator, which resulted in a considerable reduction in the standard error of $\widehat{\beta}_0$, $\widehat{\beta}_1$ and $\widehat{\beta}_2$, as compared with that of the standard censored quantile regression.

4.4.1.B Auxiliary information based on x_1

Now consider a more practical scenario when only partial information is available; i.e., we assume that we only have the information about θ_0 and θ_1.

To consider this scenario, we assume a linear relationship between T and X_1 with slope (θ_1) and intercept (θ_0), estimated using the standard linear regression based on the finite population size of $N = 10000$. For the simulation studies, we estimated p_i's using the estimating function, $g(t_i, x_{1i}; \boldsymbol{\theta})$. We repeated our simulations 1000 times and computed the mean bias. We used a 250 bootstrap sample to estimate the standard error (SE) and 95% coverage probability (CP). The summaries of these studies are given in Tables 4.10 to 4.13 for the uncorrelated covariates and in Tables 4.14 to 4.17 for the correlated covariates.

	n		**CQR**			**CQR-EL1**		
		$\tau \rightarrow$	0.25	0.50	0.75	0.25	0.50	0.75
Bias	100	β_0.	0.0096	0.0191	0.0594	0.0134	0.0272	0.0737
		β_1.	0.0038	0.0063	0.0162	0.0020	0.0026	0.0070
		β_2.	0.0035	-0.0008	-0.0102	0.0068	0.0039	-0.0026
	200	β_0	0.0227	0.0267	0.0543	0.0224	0.0314	0.0588
		β_1	-0.0011	-0.0005	0.0019	0.0016	-0.0009	0.0060
		β_2	0.0012	-0.0032	-0.0034	0.0022	-0.0012	-0.0014
SE	100	β_0.	0.1437	0.1394	0.2205	0.1146	0.1127	0.2149
		β_1.	0.1526	0.1517	0.2064	0.1191	0.1159	0.1919
		β_2.	0.1536	0.1544	0.2186	0.1542	0.1555	0.2209
	200	β_0	0.0982	0.0914	0.1276	0.0758	0.0719	0.1207
		β_1	0.1035	0.1011	0.1351	0.0780	0.0738	0.1172
		β_2	0.1062	0.1023	0.1351	0.1067	0.1025	0.1376
CP	100	β_0.	92.6	93.5	95.6	95.7	96.1	96.7
		β_1.	95.5	95.5	96.9	97.0	97.7	97.9
		β_2.	95.7	96.2	96.6	94.8	94.9	97.1
	200	β_0	94.0	93.4	93.9	94.6	93.9	95.3
		β_1	95.1	95.2	95.2	95.1	95.3	96.2
		β_2	95.4	94.3	94.1	95.9	94.0	94.8

Table 4.10: Bias, SE and CP of regression parameters for Case (i) model with independent covariates ($\sigma_{x_1,x_2} = 0$)

	n	$\tau \rightarrow$	CQR 0.25	CQR 0.50	CQR 0.75	CQR-EL1 0.25	CQR-EL1 0.50	CQR-EL1 0.75
Bias	100	β_0.	0.0105	0.0353	0.0918	0.0217	0.0422	0.1076
		β_1.	-0.0030	0.0051	0.0164	-0.0047	0.0039	0.0175
		β_2.	0.0092	0.0046	-0.0032	0.0056	0.0075	-0.0115
	200	β_0	0.0237	0.0272	0.0708	0.0305	0.0329	0.0781
		β_1	0.0019	0.0028	0.0081	0.0019	0.0011	0.0110
		β_2	-0.0017	-0.0026	-0.0019	-0.0000	0.0008	0.0014
SE	100	β_0.	0.1837	0.1542	0.2913	0.1524	0.1326	0.2874
		β_1.	0.1927	0.1657	0.2589	0.1539	0.1342	0.2414
		β_2.	0.1934	0.1669	0.2687	0.1954	0.1689	0.2790
	200	β_0	0.1235	0.1007	0.1667	0.1009	0.0866	0.1607
		β_1	0.1298	0.1126	0.1688	0.1014	0.0896	0.1515
		β_2	0.1304	0.1125	0.1696	0.1292	0.1137	0.1730
CP	100	β_0.	94.2	93.9	95.3	94.6	94.6	96.6
		β_1.	95.8	95.1	95.7	97.0	96.0	97.7
		β_2.	95.8	94.3	95.5	95.1	94.6	96.2
	200	β_0	94.0	91.9	94.0	93.6	93.7	94.7
		β_1	95.1	95.9	95.1	95.4	95.9	96.5
		β_2	93.4	95.8	95.0	94.4	95.3	94.4

Table 4.11: Bias, SE and CP of regression parameters for Case (ii) model with independent covariates ($\sigma_{x_1,x_2} = 0$)

	n	$\tau \rightarrow$	CQR 0.25	CQR 0.50	CQR 0.75	CQR-EL1 0.25	CQR-EL1 0.50	CQR-EL1 0.75
Bias	100	β_0.	0.0062	0.0088	0.0224	0.0078	0.0113	0.0280
		β_1.	0.0042	0.0051	0.0076	0.0045	0.0028	0.0077
		β_2.	-0.0038	-0.0039	-0.0069	-0.0011	0.0010	-0.0024
	200	β_0	0.0095	0.0111	0.0198	0.0081	0.0112	0.0200
		β_1	0.0007	0.0018	0.0023	0.0013	0.0003	0.0011
		β_2	0.0011	-0.0016	-0.0006	0.0009	-0.0001	0.0020
SE	100	β_0.	0.0472	0.0466	0.0767	0.0394	0.0380	0.0751
		β_1.	0.0566	0.0570	0.0796	0.0464	0.0458	0.0743
		β_2.	0.0567	0.0575	0.0807	0.0543	0.0547	0.0822
	200	β_0	0.0317	0.0302	0.0403	0.0257	0.0239	0.0361
		β_1	0.0371	0.0379	0.0492	0.0300	0.0299	0.0430
		β_2	0.0373	0.0372	0.0490	0.0351	0.0350	0.0478
CP	100	β_0.	94.4	95.0	96.1	94.3	95.1	97.0
		β_1.	95.0	95.2	95.5	96.0	96.2	97.7
		β_2.	96.6	96.7	97.3	95.6	96.5	97.2
	200	β_0	93.9	92.5	93.6	93.6	93.8	93.3
		β_1	95.4	94.4	95.3	95.8	94.5	96.2
		β_2	94.4	94.9	96.6	94.1	95.1	95.7

Table 4.12: Bias, SE and CP of regression parameters for Case (iii) model with independent covariates ($\sigma_{x_1,x_2} = 0$)

	n	$\tau \rightarrow$	**CQR** 0.25	0.50	0.75	**CQR-EL1** 0.25	0.50	0.75
Bias	100	β_0.	0.0026	0.0108	0.0334	0.0066	0.0133	0.0356
		β_1.	0.0043	0.0027	0.0111	0.0012	0.0001	0.0072
		β_2.	-0.0010	-0.0014	-0.0086	0.0036	0.0013	-0.0071
	200	β_0	0.0095	0.0125	0.0232	0.0111	0.0137	0.0249
		β_1	-0.0007	0.0002	0.0020	-0.0013	-0.0002	0.0006
		β_2	0.0011	0.0011	0.0009	0.0030	0.0026	0.0024
SE	100	β_0.	0.0594	0.0508	0.1093	0.0491	0.0428	0.1036
		β_1.	0.0668	0.0600	0.0964	0.0543	0.0486	0.0875
		β_2.	0.0663	0.0598	0.0996	0.0625	0.0567	0.0999
	200	β_0	0.0397	0.0329	0.0514	0.0325	0.0272	0.0470
		β_1	0.0429	0.0383	0.0567	0.0348	0.0312	0.0506
		β_2	0.0432	0.0389	0.0573	0.0409	0.0368	0.0560
CP	100	β_0.	94.1	94.0	96.9	94.5	95.3	98.2
		β_1.	96.4	96.7	97.5	96.2	95.5	98.2
		β_2.	96.3	97.0	96.5	95.5	97.1	96.8
	200	β_0	93.4	91.8	94.7	93.7	93.0	94.3
		β_1	95.8	96.5	95.5	95.7	94.1	96.4
		β_2	95.5	95.7	94.9	94.3	95.2	95.3

Table 4.13: Bias, SE and CP of regression parameters for Case (iv) model with independent covariates ($\sigma_{x_1,x_2} = 0$)

For Case (i) with the independent covariates, we see from Table 4.10 that our proposed method has approximately zero bias for all the parameters even with the partial auxiliary information. Both methods provided approximately 95% coverage probability and when the sample size increases, the coverage probability attains its nominal level of 95%. Since we have used the auxiliary information in relation to x_1 only, the reduction in SE is observed only for $\widehat{\beta}_0$ and $\widehat{\beta}_1$, not for $\widehat{\beta}_2$. For example, we consider the scenario $n = 100$ and $\tau = 0.25$ for comparison. SE of $\widehat{\beta}_1$ for CQR is 0.1526 and for CQR-EL1, it reduces to 0.1191. But for $\widehat{\beta}_2$, SE of both CQR and CQR-EL1 are 0.1536 and 0.1542 respectively. The standard error of $\widehat{\beta}_2$ is almost the same for both methods. In comparison with Case (i), we considered a heavy tailed distribution for the failure time in Case (ii) (results are in Table 4.11) and we noticed a slight increase in SE for both our proposed method and CQR.

The simulation results for the models with the error term depending on the covariates (Cases (iii) & (iv)) are provided in Tables 4.12 and 4.13 respectively. As mentioned above, there is a considerable reduction in SE for $\widehat{\beta}_0$ and $\widehat{\beta}_1$ because we use the population information in relation to x_1. For example, the SE of $\widehat{\beta}_1$ with $n = 100$ and $\tau = 0.25$ (from Table 4.12) for CQR and CQR-EL1 methods are 0.0566 and 0.0464 respectively. Since the errors depend on the covariates, the SE of $\widehat{\beta}_2$ shows a slight reduction from 0.0567 to 0.0543.

	n		**CQR**			**CQR-EL1**		
		$\tau \rightarrow$	0.25	0.50	0.75	0.25	0.50	0.75
Bias	100	β_0.	0.0230	0.0372	0.0725	0.0145	0.0339	0.0777
		β_1.	-0.0045	0.0001	-0.0009	-0.0049	-0.0057	0.0004
		β_2.	0.0039	0.0025	-0.0042	0.0050	0.0100	-0.0001
	200	β_0	0.0233	0.0283	0.0488	0.0248	0.0314	0.0556
		β_1	0.0035	0.0028	-0.0033	0.0010	0.0019	-0.0079
		β_2	-0.0012	0.0002	0.0033	-0.0026	0.0024	0.0091
SE	100	β_0.	0.1441	0.1407	0.2251	0.1136	0.1125	0.2192
		β_1.	0.1787	0.1800	0.2483	0.1493	0.1480	0.2360
		β_2.	0.1794	0.1807	0.2549	0.1799	0.1823	0.2564
	200	β_0	0.0976	0.0911	0.1269	0.0751	0.0714	0.1199
		β_1	0.1205	0.1176	0.1559	0.0978	0.0944	0.1415
		β_2	0.1223	0.1185	0.1562	0.1229	0.1186	0.1588
CP	100	β_0.	94.7	93.3	95.5	96.4	96.3	97.6
		β_1.	94.8	95.4	95.0	96.5	95.8	96.9
		β_2.	94.4	94.7	96.5	95.0	94.6	96.4
	200	β_0	91.9	92.0	92.8	93.6	94.5	94.5
		β_1	94.8	95.0	94.4	95.1	94.2	95.6
		β_2	93.6	94.5	95.3	94.1	94.1	96.0

Table 4.14: Bias, SE and CP of regression parameters for Case (i) model with dependent covariates ($\sigma_{x_1,x_2} = 0.5$)

	n		CQR			CQR-EL1		
		$\tau \rightarrow$	0.25	0.50	0.75	0.25	0.50	0.75
Bias	100	β_0.	0.0150	0.0321	0.0935	0.0181	0.0384	0.0993
		β_1.	0.0060	0.0035	0.0019	0.0053	0.0127	0.0088
		β_2.	-0.0092	-0.0037	0.0007	-0.0022	-0.0013	-0.0019
	200	β_0	0.0241	0.0268	0.0754	0.0254	0.0278	0.0714
		β_1	-0.0055	-0.0047	-0.0025	-0.0031	-0.0041	-0.0028
		β_2	-0.0000	0.0035	0.0067	-0.0004	0.0061	0.0078
SE	100	β_0.	0.1830	0.1542	0.2937	0.1514	0.1325	0.2871
		β_1.	0.2248	0.1978	0.3138	0.1936	0.1710	0.2976
		β_2.	0.2280	0.1985	0.3218	0.2299	0.2011	0.3303
	200	β_0	0.1214	0.1010	0.1657	0.0995	0.0866	0.1593
		β_1	0.1493	0.1303	0.1957	0.1260	0.1106	0.1794
		β_2	0.1492	0.1317	0.1969	0.1489	0.1317	0.1990
CP	100	β_0.	94.4	93.2	96.3	94.5	94.5	97.3
		β_1.	96.0	95.5	96.2	96.4	96.2	97.3
		β_2.	94.9	95.2	96.0	94.5	95.1	95.1
	200	β_0	92.1	91.4	93.9	92.6	93.0	95.9
		β_1	95.8	95.3	95.3	94.8	94.2	96.3
		β_2	94.9	95.2	95.7	95.3	94.2	95.8

Table 4.15: Bias, SE and CP of regression parameters for Case (ii) model with dependent covariates ($\sigma_{x_1,x_2} = 0.5$)

	n	$\tau \rightarrow$	**CQR** 0.25	0.50	0.75	**CQR-EL1** 0.25	0.50	0.75
Bias	100	β_0.	0.0043	0.0076	0.0218	0.0072	0.0114	0.0239
		β_1.	-0.0002	0.0035	0.0118	-0.0023	-0.0007	0.0071
		β_2.	0.0026	-0.0026	-0.0090	0.0053	0.0013	-0.0044
	200	β_0	0.0076	0.0104	0.0173	0.0071	0.0097	0.0170
		β_1	-0.0004	-0.0024	0.0010	-0.0001	-0.0029	-0.0010
		β_2	-0.0005	0.0034	-0.0008	0.0002	0.0039	0.0016
SE	100	β_0.	0.0456	0.0441	0.0774	0.0377	0.0357	0.0771
		β_1.	0.0601	0.0607	0.0876	0.0515	0.0507	0.0838
		β_2.	0.0606	0.0612	0.0884	0.0580	0.0587	0.0905
	200	β_0	0.0305	0.0290	0.0399	0.0248	0.0230	0.0361
		β_1	0.0400	0.0410	0.0545	0.0337	0.0338	0.0493
		β_2	0.0401	0.0413	0.0547	0.0382	0.0389	0.0538
CP	100	β_0.	95.1	95.0	97.7	95.4	96.1	98.2
		β_1.	96.6	96.6	96.5	95.7	96.2	96.5
		β_2.	95.8	95.3	96.7	95.9	95.4	96.8
	200	β_0	91.6	91.8	94.1	94.8	93.1	95.2
		β_1	95.4	95.8	95.7	95.1	94.6	95.5
		β_2	94.6	94.7	94.4	95.0	94.7	94.3

Table 4.16: Bias, SE and CP of regression parameters for Case (iii) model with dependent covariates ($\sigma_{x_1,x_2} = 0.5$)

	n	$\tau \rightarrow$	**CQR** 0.25	0.50	0.75	**CQR-EL1** 0.25	0.50	0.75
Bias	100	β_0.	0.0042	0.0110	0.0382	0.0045	0.0108	0.0364
		β_1.	0.0016	0.0041	0.0109	-0.0008	0.0030	0.0082
		β_2.	-0.0002	-0.0032	-0.0119	0.0029	-0.0009	-0.0112
	200	β_0	0.0083	0.0100	0.0244	0.0087	0.0106	0.0246
		β_1	-0.0020	0.0017	0.0031	-0.0014	0.0005	0.0017
		β_2	0.0017	0.0000	-0.0030	0.0020	0.0014	0.0003
SE	100	β_0.	0.0595	0.0498	0.1099	0.0493	0.0422	0.1068
		β_1.	0.0735	0.0663	0.1134	0.0622	0.0570	0.1077
		β_2.	0.0747	0.0668	0.1147	0.0708	0.0634	0.1172
	200	β_0	0.0383	0.0319	0.0517	0.0316	0.0269	0.0479
		β_1	0.0471	0.0426	0.0654	0.0400	0.0367	0.0600
		β_2	0.0475	0.0424	0.0643	0.0451	0.0406	0.0633
CP	100	β_0.	95.4	95.0	97.2	95.9	95.4	98.1
		β_1.	95.7	96.4	96.9	96.8	97.2	97.3
		β_2.	96.0	96.3	96.9	96.6	96.1	96.8
	200	β_0	93.6	93.0	94.3	94.2	93.5	95.4
		β_1	95.7	95.3	95.4	94.5	95.1	95.0
		β_2	96.1	95.9	95.3	95.7	95.5	95.3

Table 4.17: Bias, SE and CP of regression parameters for Case (iv) model with dependent covariates ($\sigma_{x_1,x_2} = 0.5$)

From Tables 4.14 - 4.17, we see that the estimators have negligible bias and provide approximately 95% coverage probability for all the parameters. The standard error of $\widehat{\beta}_0$ and $\widehat{\beta}_1$ reduced considerably irrespective of whether the error term is depending on the covariates. But the standard error of $\widehat{\beta}_2$ remains the same as for CQR when the error term is independent of the covariates (Tables 4.14 and 4.15) and is slightly reduced when the error depend on the covariates (Tables 4.16 and 4.17). A comparison of the results in Tables 4.10 - 4.13 leads to the conclusion that the correlation between the covariates does not have much influence on the parameter estimates. These simulation studies show that if the population information about the relationship between T and the covariates is available, our proposed EL based weighted censored quantile regression has a remarkable efficiency gain compared to the standard censored quantile regression method.

4.4.2 Numerical Study - II

In most of the survival data with random right censoring, the observed data are the triplet $\{Y = \min(T, C), \boldsymbol{X}, \delta\}$. In this section, we consider a linear relationship between the survival time (Y) and the covariates as the auxiliary information. Here we cannot use the EL estimating function, $g(\cdot)$ defined in (4.10) because of the censoring. There are other methods available in the literature which take care of the right censoring in the linear regression.

Koul et al. [1981] introduced a synthetic data approach by transforming the survival time, Y_r to a synthetic variable, $\widetilde{Y}_r$ as

$$\widetilde{Y}_r = \frac{\delta_r Y_r}{1 - G(Y_r)};\ r = 1, 2, \ldots, N, \tag{4.11}$$

where δ_r is the censoring indicator and $G(\cdot)$ is the distribution of the censoring time. $E(\widetilde{Y} \mid \boldsymbol{X}) = E(Y \mid \boldsymbol{X})$ if C is independent of both $\boldsymbol{X}$ and Y. When $G(\cdot)$ is unknown, we can replace it with its Kaplan-Meier estimate. The estimator of $G(\cdot)$ using the Kaplan-Meier [Kaplan and Meier, 1958] estimator is

$$1 - \widehat{G}_N(t) = \prod_{r=1}^{N} \left(\frac{N - r}{N - r + 1} \right)^{\mathbb{I}(Y_{(r)} \leq t, \delta_{(r)} = 0)}, \tag{4.12}$$

where $Y_{(r)}$s are ordered and the corresponding censoring indicator is $\delta_{(r)}$. We can estimate $\boldsymbol{\theta}$ as

$$\widetilde{\boldsymbol{\theta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \widetilde{Y}_r. \tag{4.13}$$

Qin and Jing [2001] and Li and Wang [2003] independently provided the estimating function to compute the EL based data driven probabilities as

$$g_i(\boldsymbol{z}_i; \widetilde{\boldsymbol{\theta}}) = g(y_i, \boldsymbol{x}_i, \delta_i; \widetilde{\boldsymbol{\theta}}) = \boldsymbol{x}_i(\widetilde{y}_i - \boldsymbol{x}_i^\top \widetilde{\boldsymbol{\theta}}),\ i = 1, 2, \ldots, n. \tag{4.14}$$

We can compute the $\widetilde{y}_i$ and $\widehat{G}_n(t)$ using the sample analogues of (4.11) and (4.12) respectively.

4.4.2.A Auxiliary information based on both x_1 and x_2

To compute p_i's, we consider a linear relation between Y and $\boldsymbol{X} = (X_1, X_2)$ with slopes (θ_1 and θ_2) and intercept (θ_0). We estimate $\boldsymbol{\theta}$ using (4.13) based on a large, finite population with size, $N = 10000$. To fit the weighted censored quantile regression model given in (4.2), we generate another n observations $\{y_i, \boldsymbol{x}_i\}_{i=1}^n$ with $n \ll N$ using the same models given in Table 4.1. For our proposed method, we estimate p_i's using the estimating function, $g(y_i, \boldsymbol{x}_i, \delta_i; \widetilde{\boldsymbol{\theta}})$ given in (4.14).

Similar to numerical study - I, we present the results based on 1000 simulations and report the bias, standard error (SE) and empirical coverage probability (CP) for the nominal level of 95% based on 250 bootstrap samples. We provide the summary of the simulation results for this study in Tables E.1 to E.8 (Appendix E). In Tables E.1 to E.4, we present the simulation results for the models with the uncorrelated covariates and in Tables E.5 to E.8, the simulation results are for the correlated covariates.

Similar to the population information related to T (numerical study - I), conclusions are almost similar for both correlated and uncorrelated covariates. Our proposed method (CQR-EL2) provides unbiased estimates irrespective of any sample size and quantile. If we consider the coverage probability, both CQR and CQR-EL2 provide approximately 95% coverage. For any quantile, there is a reduction in the standard error of CQR-EL2 parameter estimates compared to CQR parameter estimates. If we consider Case (i) as a basic model, CQR-EL2 with Case (ii) has reasonably higher SE along with CQR because of the heavy tailed distribution of the observed survival time. When the error depended on the covariates (Cases (iii) & (iv)), the SE of CQR-EL2 reduced considerably.

4.4.2.B Auxiliary information based on x_1

The results in Tables E.9 to E.16 are based on partial population information. Now the weights, p_i's are computed using the estimating function, $g(y_i, x_{1i}, \delta_i; \widetilde{\boldsymbol{\theta}})$. Similar to previous simulation settings, we considered the uncorrelated covariates models and reported results in Tables E.9 to E.12, and the correlated covariates models with results reported in Tables E.13 to E.16.

In numerical study-I, we have a slight reduction in SE of $\widehat{\beta}_2$ using heteroscedastic models for EQR-EL1. But using the estimating function, $g(y_i, x_{1i}, \delta_i; \widetilde{\boldsymbol{\theta}})$ (EQR-EL2), does not reduce the SE of $\widehat{\beta}_2$ under heteroscedastic models. Since we utilized only partial population information in relation to X_1, the standard error of $\widehat{\beta}_0$ and $\widehat{\beta}_1$ reduced for CQR-EL2 compared to CQR. The standard error of $\widehat{\beta}_2$ was not changed.

Our simulation studies reveal that auxiliary information greatly enhances the efficiency of estimation, if the population information related to both X_1 and X_2 is available. If the population information is only related to X_1, the efficiency gain is limited to β_0 and β_1 only. However, under heteroscedastic models, the efficiency of estimating β_2 slightly improved in numerical study - I, but not in numerical study - II.

4.4.3 Other Choices of $g(\cdot)$

If the auxiliary information is in the form of a linear relationship between Y and $\boldsymbol{X}$, there are other EL estimating functions, $g(\cdot)$'s, available for the computation of p_i's from the right censored data in the literature.

Zhou and Li [2008] introduced the censored EL using the Buckley-James [see Buckley and James, 1979; Ritov, 1990] estimating function. Let $e_i(\boldsymbol{b}) = Y_i - \boldsymbol{X}_i^\top \boldsymbol{b}$ and let $\boldsymbol{b}$ be the candidate estimator of $\boldsymbol{\theta}$. The ordered $e_i(\boldsymbol{b})$'s are denoted as $e_{(i)}(\boldsymbol{b})$; $i = 1, 2, \ldots, n$ and the corresponding covariates and censoring indicator are $\boldsymbol{X}_{(i)}$ and $\delta_{(i)}$ respectively. To compute the EL based data driven probability weights, the estimating function is

$$g_{(i)}(\boldsymbol{z}_i; \widehat{\boldsymbol{\theta}}) = g_{(i)}(y_i, \boldsymbol{x}_i, \delta_i; \widehat{\boldsymbol{\theta}}) = \delta_{(i)} e_{(i)}(\boldsymbol{b}) \frac{\sum_{j<i} M[j,i] \boldsymbol{X}_{(j)}}{n\omega_i}, \quad i = 1, 2, \ldots, n, \tag{4.15}$$

where $\omega_i = \sum_j M[j,i]/n$, M is a upper triangular matrix with order n and its elements are defined as

$$M[i,j] = \begin{cases} 0 & j \leq i \text{ and } \delta_{(i)} = 0 \\ \dfrac{\Delta \widehat{F}(e_{(j)})}{1 - \widehat{F}(e_{(i)})} & j > i \text{ and } \delta_{(i)} = 0 \\ 1 & j = i \text{ and } \delta_{(i)} = 1 \\ 0 & j \neq i \text{ and } \delta_{(i)} = 1. \end{cases}$$

Here $\widehat{F}(t)$ denotes the Kaplan–Meier estimator of $F(t)$ based on the sample $(e_{(i)}(\boldsymbol{b}), \delta_{(i)})$. This method forces the $p_i = 0$ for censored observations [Zhou, 2005, 2015]. This method is implemented in the "emplike" package in R software. Here the Buckley-James estimate, $\widehat{\boldsymbol{\theta}} = \boldsymbol{b}$, can be used to compute the $g_i(\cdot)$'s and then p_i's. The computation of p_i's is based on the modified EM algorithm proposed by Zhou [2005]. By forcing $p_i = 0$ for censored observations, we will get biased censored quantile regression estimates; hence, we avoided this method in the comparison.

Fang et al. [2013] proposed an empirical likelihood based on the Buckley-James estimating function which provides non-zero p_i's irrespective of censored or uncensored failure times. The estimating function is

$$g_i(\boldsymbol{z}_i;\widehat{\boldsymbol{\theta}}) = g_i(y_i,\boldsymbol{x}_i,\delta_i;\widehat{\boldsymbol{\theta}}) = (\boldsymbol{X}_i - \boldsymbol{\mu}_X)\left\{\delta_i e_i(\boldsymbol{b}) + (1-\delta_i)\frac{\int_{e_i(\boldsymbol{b})}^{\infty} t dF(t)}{1-F(e_i(\boldsymbol{b}))}\right\},\ i=1,2,\ldots,n, \tag{4.16}$$

where $\boldsymbol{X}_i$ is the covariates, $E(\boldsymbol{X}) = \boldsymbol{\mu}_X$, F is the error distribution (known), $e_i(\boldsymbol{b}) = Y_i - \boldsymbol{X}_i^\top \boldsymbol{b}$ and the censoring indicator is δ_i. However, they mentioned that even though it is easy to compute probabilities, the method is not as efficient as the method of Zhou and Li [2008]. We omitted this approach also from the performance analysis.

4.4.4 NCCTG Lung Cancer Study

The North Central Cancer Treatment Group (NCCTG) was initiated by a group of physicians from the north central region of the United States of America and the Mayo Clinic in Rochester, Minnesota. This study was conducted by NCCTG to determine whether the conclusions from the patient-completed questionnaire and those already obtained by the patient's physician were independent or not [Loprinzi et al., 1994]. They used the performance scores (ECOG and Karnofsky) to assess the patient's daily activities. The dataset is available in the "survival" package of R software with readings of 228 patients. Because of the incompleteness of the some of the variables, we had to limit the dataset to 167 observations. For the illustration of our proposed method, we changed our focus to identify the effect of following covariates over the observed survival time at different quantiles. We considered 'age', patient's age in years; 'sex', (Male=1 Female=2); 'ph.ecog', ECOG performance score measured by physician (0=good 5=dead); 'meal.cal', calories consumed at meals and 'wt.loss', weight loss in the last six months as the covariates. After removing the incomplete patient readings, the available ECOG scores were 0,1 and 2 only. We defined two dummy categorical variables for 'ph.ecog' as follows.

$$\text{ecog1} = \begin{cases} 1, & \text{if ph.ecog=1} \\ 0, & \text{Otherwise} \end{cases}$$

$$\text{ecog2} = \begin{cases} 1, & \text{if ph.ecog=2} \\ 0, & \text{Otherwise} \end{cases}$$

To demonstrate the usefulness of our proposed method, we randomly selected part

(100 observations) of the complete data (167 observations) by considering it to be the data available from the previous study. We assumed that there exists a linear relation between the logarithm of the observed survival time and all the continuous explanatory variables (age, meal.cal and wt.loss) as the available auxiliary information. We estimated the $\boldsymbol{\theta} = (\theta_0, \theta_{\text{age}}, \theta_{\text{meal}}, \theta_{\text{wt}})$ by the least square method based on 100 observations where the response is the synthetic variable defined by (4.11). Then we computed the EL based data driven probability weights for the present study data points (67 observations). After computing the weights, we estimated the weighted censored quantile regression parameters using Peng and Huang [2008] method with all the covariates mentioned at the beginning of this Section. For the present study data, the censoring proportion is 0.283. Interestingly, we estimated the regression parameters using CQR up to the 86^{th} quantile, where as we could estimate to the 90^{th} quantile using CQR-EL2. Along with the estimates for the quantiles, $\tau = 0.25, 0.5, 0.75$, we report standard error (SE) and 95% confidence limits using 250 bootstrap samples as well in Table 4.18.

		CQR			CQR-EL2		
	$\tau \rightarrow$	0.25	0.50	0.75	0.25	0.50	0.75
$\hat{\beta}$	Intercept	5.4777	4.2651	5.5380	4.7531	4.1729	6.4258
	Age	-0.0168	0.0179	0.0040	-0.0047	0.0202	-0.0032
	Sex	0.7201	0.6180	0.4181	0.7606	0.6638	0.3651
	ECOG1	-0.7059	-0.5449	-0.2029	-0.5701	-0.5355	-0.2884
	ECOG2	-0.8677	-0.9402	-0.8336	-1.1584	-1.0612	-1.0192
	MealCal	0.0004	0.0001	0.0001	0.0004	0.0001	-0.0000
	WtLoss	-0.0007	-0.0084	-0.0023	-0.0023	-0.0100	-0.0135
SE	Intercept	1.9235	1.4314	1.7494	1.6628	1.4149	1.4666
	Age	0.0277	0.0188	0.0225	0.0256	0.0184	0.0176
	Sex	0.5610	0.3389	0.3716	0.5374	0.3317	0.2809
	ECOG1	0.6521	0.3436	0.3375	0.6498	0.3493	0.2434
	ECOG2	1.0317	0.5410	0.6061	0.9336	0.5413	0.3879
	MealCal	0.0009	0.0006	0.0008	0.0009	0.0006	0.0005
	WtLoss	0.0181	0.0128	0.0231	0.0157	0.0124	0.0100
CI	Intercept	(1.6,9.14)	(2.38,8)	(2.08,8.94)	(1.79,8.31)	(2.32,7.87)	(3.14,8.89)
	Age	(-0.07,0.04)	(-0.04,0.04)	(-0.04,0.05)	(-0.06,0.04)	(-0.03,0.04)	(-0.03,0.04)
	Sex	(-0.45,1.74)	(0,1.33)	(-0.13,1.33)	(-0.39,1.71)	(-0.04,1.27)	(-0.07,1.03)
	ECOG1	(-1.75,0.81)	(-1.15,0.2)	(-0.97,0.35)	(-1.86,0.69)	(-1.18,0.19)	(-0.78,0.18)
	ECOG2	(-2.88,1.16)	(-2,0.12)	(-2.11,0.26)	(-2.83,0.83)	(-2.13,-0.01)	(-1.73,-0.21)
	WtLoss	(-0.04,0.03)	(-0.03,0.02)	(-0.05,0.04)	(-0.04,0.02)	(-0.03,0.01)	(-0.04,0)

Table 4.18: Estimates, SE and 95% CI for regression parameters of NCCTG lung cancer data

From Table 4.18, we see that the standard error of the estimates of all the continuous variable parameters and the intercept reduced considerably because we considered the auxiliary information related to them. For the remaining variables, a reduction of standard error can also be seen, even though we did not consider any auxiliary information related to them. In the censored quantile regression with the EL based data driven probability weights, we see narrower 95% confidence limits for all the variables compared to those using the standard censored quantile regression.

4.5 Summary

In this chapter, we proposed an effective use of auxiliary information to improve the efficiency of the censored quantile regression estimator. We developed a methodology to transform the population information available from previous clinical trials or from some existing facts into non-parametric empirical likelihood based data driven probabilities. We developed the EL based data driven probability computation for both known and unknown cases of prior information regarding population parameters. Then we applied these probabilities as the weights into Peng and Huang [2008] censored quantile regression model. Our proposed method is efficient compared to standard censored quantile regression and provides consistent estimators of regression coefficients with asymptotic normality.

The standard error of the parameter estimates based on our proposed methods (CQR-EL1 and CQR-EL2) is lower than the standard method (CQR) when we use all the covariates for computing the EL based data driven probability weights. Our proposed weighted censored quantile regression method provides almost the same coverage probability compared to the nominal level. In the case of heteroscedastic models, even the use of the auxiliary information regarding a subset of population parameters improved the efficiency of the estimates of all the parameters by using CQR-EL1. But in CQR-EL2, the efficiency improvement was limited to the corresponding subset of variables and intercept. In homoscedastic models, the use of auxiliary information regarding a subset of population parameters improved the efficiency only for that particular subset of parameters and the intercept in both CQR-EL1 and CQR-EL2. In the real data analysis, we observed that our proposed method provides more efficient quantile estimates and narrower confidence limits compared to the standard censored quantile regression.

Chapter 5

Concluding Remarks

Quantile regression, developed by Koenker and Basset [1978], is an emerging area in both statistics and economics. It models the conditional quantiles of the response variable. Quantile regression provides a global assessment of the covariate effect on the response at properly selected quantile levels.

Powell [1984, 1986] developed a censored quantile regression model for the cases when all the censoring times are fixed. Among the major contributions to the field of censored quantile regression under random censoring are those of Portnoy [2003] and Peng and Huang [2008].

The severe censoring could force the large failure times to be unobserved and cause an identifiability problem in the parameter estimation for the extreme quantiles of the failure time. To overcome this problem, it is not always a practical choice to wait until the larger failure times are observed because of the restrictions of the study duration. In this thesis, we proposed three methods to tackle this problem and improve the efficiency of the censored quantile regression estimators using auxiliary information.

In epidemiology studies, exposure assessment is solely based on the questionnaire. The questionnaire responses could be inaccurate and might cause significant estimation bias in the analysis. Because of the restrictions of the study time, the budgetary issues or due to other limitations, the accurate measurements of the key exposure might sometimes be limited to a subcohort (validation sample). If we use only these accurate key exposure readings available from this subcohort in the censored quantile regression model under heavy right censoring, it could result in an identification problem for the higher quantiles of the failure times because of a relatively small sample size. If we ignore this accurately measured key exposure, it could result in a serious information loss. We proposed two methods to handle this problem, considering both the surrogate/auxiliary covariate and the accurately measured main exposure available through a validation sample.

In the first method, we proposed a regression calibration based approach to the censored quantile regression model. We assumed that there exists a linear association between the accurately measured covariate and its surrogate/auxiliary covariate and other available covariates. First we predicted the unobserved covariate in the non-validation sample using the regression calibration method with the help of the auxiliary covariate and other available covariates. In the next step, we combined the accurately measured covariate readings from the validation sample with the predicted key exposures in the non-validation sample to estimate the censored quantile regression parameters. We developed a new estimating function based on Peng and Huang [2008] censored quantile regression estimating function. We also provided its asymptotic properties such as consistency and the asymptotic normality of the estimators. In the simulation study, we compared our proposed method with the results based solely on the validation sample and the completely known main exposure scenario. The standard error of the parameter estimates of our proposed method is always smaller than the one using only the validation sample, irrespective of the value of σ^2_{ε} and the quantile level. When the σ^2_{ε} is small, our proposed method and the 'complete' case have almost the same standard error. Our proposed method provided asymptotically unbiased estimates and the coverage probability of their confidence intervals is almost equal to the nominal level. Under heavy censoring, we observed that the validation sample approach fails to provide regression estimates for high quantiles when the sample size is low. As an illustration, we applied our proposed method to PBC data [Fleming and Harrington, 2011] by predicting the unobserved copper content in urine values. In application, we should use only the auxiliary covariate which has a strong linear relationship with the accurately measured covariate.

We developed the second method for the scenario, for use when we are unsure about the nature of the association between the accurately measured covariate and its auxiliary covariate when the other covariates are present. Instead of the regression calibration based approach, we used the non-parametric kernel smoothing method to predict the unobserved main exposure in the non-validation sample. We developed another new estimating function based on Peng and Huang [2008] censored quantile regression estimating function and investigated its large sample properties. From the simulation study and the Colorado Plateau uranium miners cohort data analysis, we arrived at similar conclusions as with the regression calibration based approach. We applied our proposed method to PBC data as well, for illustration.

It is possible to have unstable estimates when the dimension of the kernel goes beyond 2. If the kernel dimension is more than 2, we suggest using the regression calibration based approach. In general, we have to be very cautious when the validation sample size is very small compared to the sample size of the entire study cohort. It could affect the prediction of the unobserved key exposure in the non-validation sample.

We introduced an empirical likelihood [Owen, 2001] based weighted censored quantile regression model as our third method to improve the efficiency of the parameter estimates. When we have prior information regarding the target population parameters from previous studies or from the existing facts, we can convert this auxiliary information into empirical likelihood based data driven probabilities and apply them as the weights into censored quantile regression. Similar to our other proposed methods, we developed a new estimating equation based on Peng and Huang [2008] model and investigated the asymptotic properties of the estimator. In our first simulation study, we assumed the linear relationship between the failure time and the covariates as the auxiliary information. We used empirical likelihood (Owen [1991]) approach for the linear model to compute the probability weights. In the second simulation study, we replaced the failure time by the observed survival time in the auxiliary information, which is a more realistic scenario. We used empirical likelihood approach of Qin and Jing [2001] and Li and Wang [2003] for the right censored linear regression model based on the synthetic variable [Koul et al., 1981] to compute the probability weights. From these simulation studies, we arrived at the following conclusions. Compared to the standard censored quantile regression, using our proposed method, the efficiency enhanced only for the censored quantile regression parameter associated with the covariates which are used in both the auxiliary information and in the censored quantile regression model, including the intercept. The standard error of the weighted censored quantile regression parameter estimates associated with the covariates which are not a part of the auxiliary information remained the same as that for the standard censored quantile regression. But in the first simulation study, the standard error of all the parameter estimates reduced for the heteroscedastic censored quantile regression models, even with partial auxiliary information. In the application of an EL based weighted censored quantile regression to the NCCTG lung cancer study, the standard error reduced for all the parameter estimates with partial auxiliary information. Using our proposed method, we could identify the censored quantile regression parameters at more extreme quantile levels which failed while 1using the standard censored quantile regression.

5.1 Future Work

Quantile regression [Koenker and Basset, 1978; Koenker, 2005] models, with properly chosen quantiles, provide a global assessment of the covariate effect on the response. We proposed regression calibration and kernel smoothing based approaches in censored quantile regression for the continuous predictor variables. We would like to develop methods for discrete and categorical covariates. The theoretical justification could be less challenging than the one with the continuous covariates.

We proposed an empirical likelihood [Owen, 2001] based weighted censored quantile regression model using auxiliary information. When a subcohort has an accurately measured covariate and its auxiliary covariate is available throughout the cohort along with the information regarding the parameters of the target population from previous studies, we could combine the EL based weighted censored quantile regression model with the regression calibration and non-parametric kernel smoothing approaches.

We are also planning more research using the other choices of EL estimating function, $g(\cdot)$, to compute the probability weights when the relationship between the observed survival time and the covariates is present as auxiliary information.

A lasso based variable selection for censored quantile regression model is discussed by Wang, Zhou and Li [2013]. We would like to develop an EL based variable selection for censored quantile regression.

Another interesting area for future work is the joint modeling of survival data and longitudinal data using censored quantile regression. This could be a study based on the combination of both the quantiles of survival time and the conditional mean of longitudinal data.

Koenker and Machado [1999] proposed a goodness of fit test for quantile regression. We would like to extend it to censored quantile regression. It will be an analogue of coefficient of determination, R^2 in linear models. The test statistic, $R^2(\tau)$, will be calculated based on the minimum of the $\sum_{i=1}^{n} \rho_\tau(Y_i - \min\{C_i, \boldsymbol{X}_i{}^\top \boldsymbol{\beta}\})$ under restricted and unrestricted models.

References

Aalen, O. O. [1978]. Nonparametric inference for a family of counting processes, *The Annals of Statistics* **6**: 701–726.

Alexander, K. S. [1984]. Probability inequalities for empirical processes and a law of the iterated logarithm, *The Annals of Statistics* **12**: 1041–1067.

Andersen, P. K., Borgan, O., Gill, R. D. and Keiding, N. [1993]. *Statistical Models Based on Counting Processes*, York: Springer-Verlag.

Angrist, J., Chernozhukov, V. and Fernandez-Val, I. [2006]. Quantile regression under mis-specification, with an application to the u.s. wage structure, *Econometrica* **74**: 539–563.

Bahadur, R. R. [1966]. A note on quantiles in large samples, *The Annals of Mathematical Statistics* **37**: 577–580.

Barrodale, I. and Roberts, F. D. K. [1974]. Solution of an overdetermined system of equations in the ℓ_1 norm, *Communications of the ACM* **17**: 319–320.

Buchinsky, M. and Hahn, J. [1998]. An alternative estimator for censored quantile regression, *Econometrica* **66**: 653–671.

Buckley, J. and James, I. [1979]. Linear regression with censored data, *Biometrika* **66**: 429–436.

Cade, B. and Guo, Q. [2000]. Estimating effects of constraints on plant performance with regression quantiles, *Oikos* **91**: 245–254.

Carey, J. R., Liedo, P., Orozco, D. and Vaupel, J. W. [1992]. Slowing of mortality rates at older ages in large medfly cohorts, *Science* **258**: 457–461.

Carroll, R. J., Ruppert, D., Stefanski, L. A. and Crainiceanu, C. [2006]. *Measurement Error in Nonlinear Models: A Modern Perspective*, 2 edn, Chapman & Hall/CRC.

Chen, J. and Qin, J. [1993]. Empirical likelihood estimation for finite populations and the effective usage of auxiliary information, *Biometrika* **80**: 107–116.

Chernozhukov, V., Fernandez-Val, I. and Kowalski, A. E. [2015]. Quantile regression with censoring and endogeneity, *Journal of Econometrics* **186**: 201–221.

Chernozhukov, V. and Hong, H. [2002]. Three-step censored quantile regression and extramarital affairs, *Journal of the American Statistical Association* **97**: 872–882.

Chesher, A. [2001]. Parameter approximations for quantile regressions with measurement error.
URL: *http://cemmap.ifs.org.uk/wps/cwp0102.pdf*

Cox, D. and Oakes, D. [1984]. *Analysis of Survival Data*, New York: Chapman & Hall.

Cox, D. R. [1972]. Regression models and life-tables, *Journal of the Royal Statistical Society, Series B* **34**: 187–220.

Fan, Z. and Wang, X. [2009]. Marginal hazards model for multivariate failure time data with auxiliary covariates, *Journal Of Nonparametric Statistics* **21**: 771–786.

Fang, K.-T., Li, G., Lu, X. and Qin, H. [2013]. An empirical likelihood method for semiparametric linear regression with right censored data, *Computational and Mathematical Methods in Medicine* **2013**: 1–9.

Fitzenberger, B. [1997]. A guide to censored quantile regressions, *Handbook of Statistics: Robust Inference* **15**: 405–437. eds. G. S. Maddala and C. R. Rao, Amsterdam: North-Holland.

Fitzenberger, B. and Winker, P. [2007]. Improving the computation of censored quantile regressions, *Computational Statistics & Data Analysis* **52**: 88–108.

Fleming, T. R. and Harrington, D. P. [2011]. *Counting Processes and Survival Analysis*, John Wiley & Sons, Inc.

Fygenson, M. and Ritov, Y. [1994]. Monotone estimating equations for censored data, *The Annals of Statistics* **22**: 732–746.

Gill, R. D. and Johansen, S. [1990]. A survey of product-integration with a view toward application in survival analysis, *The Annals of Statistics* **18**: 1501–1555.

Granville, K. and Fan, Z. [2014]. Buckley-James estimator of AFT models with auxiliary covariates, *PLOS ONE* **9**(8): e104817. doi:10.1371/journal.pone.0104817.

He, X. and Liang, H. [2000]. Quantile regression estimates for a class of linear and partially linear errors-in-variables models, *Statistica Sinica* **10**: 129–140.

Honore, B., Khan, S. and Powell, J. [2002]. Quantile regression under random censoring, *Journal Of Econometrics* **109**: 67–105.

Kalbfleisch, J. D. and Prentice, R. L. [2002]. *The Statistical Analysis of Failure Time Data*, 2 edn, John Wiley & Sons, Inc.

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

Kiefer, J. [1967]. On bahadur's representation of sample quantiles, *The Annals of Mathematical Statistics* **38**: 1323–1342.

Klein, J. P. and Moeschberger, M. L. [2003]. *Survival Analysis: Techniques for Censored and Truncated Data*, Springer Science & Business Media.

Koenker, R. [2005]. *Quantile Regression*, Cambridge.

Koenker, R. [2008]. Censored quantile regression redux, *Journal Of Statistical Software* **27**(06): 1–25.

Koenker, R. and Basset, G. [1978]. Regression quantiles, *Econometrica* **46**: 33–50.

Koenker, R. and Geling, O. [2001]. Reappraising medfly longevity: A quantile regression survival analysis, *Journal Of the American Statistical Association* **96**: 458–468.

Koenker, R. and Machado, J. A. F. [1999]. Goodness of fit and related inference processes for quantile regression, *Journal of the American Statistical Association* **94**: 1296–1310.

Koul, H., Susarla, V. and Ryzin, J. V. [1981]. Regression analysis with randomly right-censored data, *The Annals of Statistics* **9**: 1276–1288.

Kuk, A. Y. C. and Mak, T. K. [1989]. Median estimation in the presence of auxiliary information, *Journal of Royal Statistical Society, Series B* **51**: 261–269.

Lai, T. L. and Ying, Z. [1988]. Stochastic integrals of empirical-type processes with applications to censored regression, *Journal of Multivariate Analysis* **27**: 334–358.

Langholz, B. and Goldstein, L. [1996]. Risk set sampling in epidemiological cohort studies, *Statistical Sciences* **11**: 35–53.

Leng, C. and Tong, X. [2013]. A quantile regression estimator for censored data, *Bernoulli* **19**: 344–361.

Li, G. and Wang, Q.-H. [2003]. Empirical likelihood regression analysis for right censored data, *Statistica Sinica* **13**: 51–68.

Lindgren, A. [1997]. Quantile regression with censored data using generalized L_1 minimization., *Computational Statistics & Data Analysis* **23**: 509–524.

Liu, Y., Yuan, Z., Cai, J. and Zhou, H. [2012]. Marginal hazard regression for correlated failure time data with auxiliary covariates, *Lifetime Data Analysis* **18**: 116–138.

Liu, Y., Zhou, H. and Cai, J. [2009]. Estimated pseudo-partial-likelihood method for correlated failure time data with auxiliary covariates, *Biometrics* **65**: 1184–1193.

Loprinzi, C. L., Laurie, J. A., Wieand, H. S., Krook, J. E., Novotny, P. J., Kugler, J. W., Bartel, J., Law, M., Bateman, M., Klatt, N. E., Dose, A. M., Etzell, P. S., Nelimark, R. A., Mailliard, J. A. and Moertel, C. G. [1994]. Prospective evaluation of prognostic variables from patient-completed questionnaires, *Journal of Clinical Oncology* **12**: 601–607. North Central Cancer Treatment Group.

Lubin, J., Boice, J., Edling, C., Hornung, R., Howe, G., Hunz, E., Kusiak, R., Morrison, H., Radford, E., Samet, J., Tirmarche, M., Woodward, A., Xiang, Y. and Pierce, D. [1994]. *Radon and lung cancer risk: a joint analysis of 11 underground miners studies*, Vol. 94 - 3644 of *National Institutes of Health Publication*, U.S. Dept. of Health and Human Services, Public Health Service, National Institutes of Health.

Ma, Y. and Yin, G. [2011]. Censored quantile regression with covariate measurement errors, *Statistica Sinica* **21**: 949–971.

McKeague, I., Subramanian, S. and Sun, Y. [2001]. Median regression and the missing information principle, *Journal of Nonparametric Statistics* **13**: 709–727.

Montes-Rojas, G. [2011]. Quantile regression with classical additive measurement errors, *Economics Bulletin* **31**: 2863–2868.

Müller, U. U. and Keilegom, I. V. [2014]. Efficient quantile regression with auxiliary information, *Contemporary Developments in Statistical Theory*, Vol. 68 of *Springer*

Proceedings in Mathematics & Statistics, Springer International Publishing, pp. 365–374.

Nadaraya, E. A. [1964]. On estimating regression, *Theory of Probability and its Applications* **9**: 141–142.

Nelson, W. B. [1972]. Theory and applications of hazard plotting for censored survival data, *Technometrics* **14**: 945–965.

Neocleous, T. and Portnoy, S. [2009]. Partially linear censored quantile regression, *Lifetime Data Analysis* **15**: 357–378.

Neocleous, T., Vanden Branden, K. and Portnoy, S. [2006]. Correction to Censored Regression Quantiles by S. Portnoy, 98 (2003), 1001-1012, *Journal of the American Statistical Association* **101**: 860–861. Correction.

Newey, W. K. and Powell, J. L. [1990]. Efficient estimation of linear and type I censored regression models under conditional quantile restrictions, *Econometric Theory* **6**: 295–317.

Owen, A. [1991]. Empirical likelihood for linear models, *The Annals of Statistics* **19**: 1725–1747.

Owen, A. [1998]. Empirical likelihood ratio confidence interval for a single functional, *Biometrika* **75**: 237–249.

Owen, A. [2001]. *Empirical Likelihood*, Chapman & Hall/CRC.

Peng, L. and Huang, Y. [2008]. Survival analysis with quantile regression models, *Journal Of the American Statistical Association* **103**: 637–649.

Pepe, M. S., Self, S. G. and Prentice, R. L. [1989]. Further results on covariate measurement errors in cohort studies with time to response data, *Statistics in Medicine* **8**: 1167–1178.

Portnoy, S. [2003]. Censored regression quantiles, *Journal Of the American Statistical Association* **98**: 1001–1012.

Portnoy, S. and Lin, G. [2010]. Asymptotics for censored regression quantiles, *Journal of Nonparametric Statistics* **22**: 115–130.

Powell, J. [1984]. Least absolute deviations estimation for the censored regression model, *Journal Of Econometrics* **25**: 303–325.

Powell, J. [1986]. Censored regression quantiles, *Journal Of Econometrics* **32**: 143–155.

Prentice, R. L. [1978]. Linear rank tests with right censored data, *Biometrika* **65**: 167–180.

Prentice, R. L. [1982]. Covariate measurement errors and parameter estimation in a failure time regression model, *Biometrika* **69**: 331–342.

Qin, G. and Jing, B.-Y. [2001]. Empirical likelihood for censored linear regression, *Scandinavian Journal of Statistics* **28**: 661–673.

Qin, J. and Lawless, J. [1994]. Empirical likelihood and general estimating equations, *The Annals of Statistics* **22**: 300–325.

Rao, J. N. K., Kovar, J. G. and Mantel, H. J. [1990]. On estimating distribution functions and quantiles from survey data using auxiliary information, *Biometrika* **77**: 365–375.

Ritov, Y. [1990]. Estimation in a linear regression model with censored data, *The Annals of Statistics* **18**: 303–328.

Römisch, W. [2005]. Delta method, infinite dimensional, *Encyclopedia of Statistical Sciences 3*, 2 edn, New York: Wiley, pp. 1575–1583.

Schennach, S. [2008]. Quantile regression with mismeasured covariates, *Econometric Theory* **24**: 1010–1043.

Silverman, B. W. [1986]. *Density Estimation for Statistics and Data Analysis*, Vol. 26 of *Monographs on Statistics and Applied Probability*, London: Chapman & Hall/CRC.

Tang, C. and Leng, C. [2012]. An empirical likelihood approach to quantile regression with auxiliary information, *Statistics & Probability Letters* **82**: 29–36.

Tao, T. [2016]. Generalized solutions, `http://www.math.ucla.edu/~tao/preprints/generalized_solutions.pdf`. [Online; accessed on August 20, 2016].

Tobin, J. [1958]. Estimation of relationships for limited dependent variables, *Econometrica* **26**: 24–36.

Tsiatis, A. A. [1990]. Estimating regression parameters using linear rank tests for censored data, *The Annals of Statistics* **18**: 354–372.

van der Vaart, A. W. and Wellner, J. A. [1996]. *Weak Convergence and Empirical Processes: With Applications to Statistics*, New York: Springer-Verlag.

Wagener, J., Volgushev, S. and Dette, H. [2012]. The quantile process under random censoring, *Mathematical Methods of Statistics* **21**: 127–141.

Wang, C. Y., Hsu, L., Feng, Z. D. and Prentice, R. L. [1997]. Regression calibration in failure time regression, *Biometrics* **53**: 131–145.

Wang, H. J. and Fygenson, M. [2009]. Inference for censored quantile regression models in longitudinal studies, *The annals of statistics* **37**: 756–781.

Wang, H. J., Stefanski, L. A. and Zhu, Z. [2012]. Corrected-loss estimation for quantile regression with covariate measurement errors, *Biometrika* **99**: 405–421.

Wang, H. J. and Wang, L. [2009]. Locally weighted censored quantile regression, *Journal of the American Statistical Association* **104**: 1117–1128.

Wang, H. J., Zhou, J. and Li, Y. [2013]. Variable selection for censored quantile regresion, *Statistica Sinica* **23**: 145–167.

Watson, G. S. [1964]. Smooth regression analysis, *Sankhya: The Indian Journal of Statistics, Series A* **26**: 359–372.

Wei, L. J., Ying, Z. and Lin, D. Y. [1990]. Linear regression analysis of censored survival data based on rank tests, *Biometrika* **77**: 845–851.

Wei, Y. and Carroll, R. J. [2009]. Quantile regression with measurement error, *Journal of the American Statistical Association* **104**: 1129–1143.

Wu, Y., Ma, Y. and Yin, G. [2015]. Smoothed and corrected score approach to censored quantile regression with measurement errors, *Journal of the American Statistical Association* **110**: 1670–1683.

Wu, Y. and Yin, G. [2013]. Cure rate quantile regression for censored data with a survival fraction, *Journal of the American Statistical Association* **108**: 1517–1531.

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

Yin, G. and Cai, J. [2005]. Quantile regression models with multivariate failure time data, *Biometrics* **61**: 151–161.

Yin, G., Zeng, D. and Li, H. [2008]. Power-transformed linear quantile regression with censored data, *Journal of the American Statistical Association* **103**: 1214–1224.

Yin, G., Zeng, D. and Li, H. [2014]. Censored quantile regression with varying coefficients, *Statistica Sinica* **24**: 855–870.

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

Yu, M. and Nan, B. [2010]. Regression calibration in semiparametric accelerated failure time models, *Biometrics* **66**: 405–414.

Zhou, H. and Pepe, M. [1995]. Auxiliary covariate data in failure time regression, *Biometrika* **82**: 139–149.

Zhou, H. and Wang, C. [2000]. Failure time regression with continuous covariates measured with error, *Journal Of Royal Statistical Society. Series B (Statistical Methodology)* **62**: 657–665.

Zhou, M. [2005]. Empirical likelihood ratio with arbitrarily censored/truncated data by em algorithm, *Journal of Computational and Graphical Statistics* **14**: 643–656.

Zhou, M. [2015]. *Empirical Likelihood Method in Survival Analysis*, Vol. 79 of *Chapman & Hall/CRC Biostatistics Series*, CRC Press.

Zhou, M. and Li, G. [2008]. Empirical likelihood analysis of the Buckley–James estimator, *Journal of Multivariate Analysis* **99**: 649–664.

Zou, H. and Yuan, M. [2008]. Composite quantile regression and the oracle model selection theory, *Annals of Statistics* **36**: 1108–1126.

Appendix A

Proof of Theorem 2.2.1

Define,

- $\boldsymbol{\mu}(\boldsymbol{b}) = E\left[\boldsymbol{X}\mathbb{N}\left(e^{\boldsymbol{X}^\top \boldsymbol{b}}\right)\right]$, $\boldsymbol{B}(\boldsymbol{b}) = E\left[\boldsymbol{X}^{\otimes 2}\ \widetilde{f}\left(e^{\boldsymbol{X}^\top \boldsymbol{b}} \,\middle|\, \boldsymbol{X}\right)\quad e^{\boldsymbol{X}^\top \boldsymbol{b}}\right]$,
  $\boldsymbol{\Gamma}_{m_v}(\boldsymbol{b}) = \dfrac{1}{m_v}\sum_{j\in\mathbb{V}} \boldsymbol{X}_j \mathbb{N}_j\left(e^{\boldsymbol{X}_j^\top \boldsymbol{b}}\right) - \boldsymbol{\mu}(\boldsymbol{b})$.

- $\overline{\boldsymbol{\mu}}(\boldsymbol{b}) = E\left[\boldsymbol{Z}\mathbb{N}\left(e^{\boldsymbol{Z}^\top \boldsymbol{b}}\right)\right]$, $\overline{\boldsymbol{B}}(\boldsymbol{b}) = E\left[\boldsymbol{Z}^{\otimes 2}\ \widetilde{f}\left(e^{\boldsymbol{Z}^\top \boldsymbol{b}} \,\middle|\, \boldsymbol{Z}\right)\quad e^{\boldsymbol{Z}^\top \boldsymbol{b}}\right]$,
  $\overline{\boldsymbol{\Gamma}}_{m_n}(\boldsymbol{b}) = \dfrac{1}{m_n}\sum_{l\in\overline{\mathbb{V}}} \boldsymbol{Z}_l \mathbb{N}_l\left(e^{\boldsymbol{Z}_l^\top \boldsymbol{b}}\right) - \overline{\boldsymbol{\mu}}(\boldsymbol{b})$.

- $\boldsymbol{\mu}^*(\boldsymbol{b}) = E\left[\boldsymbol{X}\mathbb{I}\left(Y \geq e^{\boldsymbol{X}^\top \boldsymbol{b}}\right)\right]$, $\boldsymbol{B}^*(\boldsymbol{b}) = E\left[\boldsymbol{X}^{\otimes 2}\ \overline{f}\left(e^{\boldsymbol{X}^\top \boldsymbol{b}} \,\middle|\, \boldsymbol{X}\right)\quad e^{\boldsymbol{X}^\top \boldsymbol{b}}\right]$,
  $\boldsymbol{\Gamma}^*_{m_v}(\boldsymbol{b}) = \dfrac{1}{m_v}\sum_{j\in\mathbb{V}} \boldsymbol{X}_j \mathbb{I}\left[Y_j \geq e^{\boldsymbol{X}_j^\top \boldsymbol{b}}\right] - \boldsymbol{\mu}^*(\boldsymbol{b})$.

- $\overline{\boldsymbol{\mu}}^*(\boldsymbol{b}) = E\left[\boldsymbol{Z}\mathbb{I}\left(Y \geq e^{\boldsymbol{Z}^\top \boldsymbol{b}}\right)\right]$, $\overline{\boldsymbol{B}}^*(\boldsymbol{b}) = E\left[\boldsymbol{Z}^{\otimes 2}\ \overline{f}\left(e^{\boldsymbol{Z}^\top \boldsymbol{b}} \,\middle|\, \boldsymbol{Z}\right)\quad e^{\boldsymbol{Z}^\top \boldsymbol{b}}\right]$,
  $\overline{\boldsymbol{\Gamma}}^*_{m_n}(\boldsymbol{b}) = \dfrac{1}{m_n}\sum_{l\in\overline{\mathbb{V}}} \boldsymbol{Z}_l \mathbb{I}\left[Y_l \geq e^{\boldsymbol{Z}_l^\top \boldsymbol{b}}\right] - \overline{\boldsymbol{\mu}}^*(\boldsymbol{b})$.

Assume that $\tau_1 < \cdots < \tau_{L-1}$ are equally spaced between 0 and τ_U. Let $a_n = \left\|\mathbb{S}_L\right\|$ and $b_n = a_n/(1-\tau_U)$; then $L = \tau_U/a_n$. It is clear that $0 < H(\tau_k) - H(\tau_{k-1}) \leq b_n$ for $k = 1, 2, \ldots, L$.

For $d > 0$, define $\mathscr{B}(d) = \{\boldsymbol{b} \in \Re^p : \inf_{\tau\in(0,\tau_U]} \left\|\rho\ [\boldsymbol{\mu}(\boldsymbol{b}) - \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)\}] + (1-\rho)\ [\overline{\boldsymbol{\mu}}(\boldsymbol{b}) - \overline{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\}]\right\| \leq d\}$. Let $\boldsymbol{\alpha}_0(\tau) = \rho\ \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)\} + (1-\rho)\ \overline{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\}$, $\widehat{\boldsymbol{\alpha}}(\tau) = \rho\ \boldsymbol{\mu}\{\widehat{\boldsymbol{\beta}}(\tau)\} + (1-\rho)\ \overline{\boldsymbol{\mu}}\{\widehat{\boldsymbol{\beta}}(\tau)\}$ and $\mathscr{A}(d) = \{\rho\ \boldsymbol{\mu}(\boldsymbol{b}) + (1-\rho)\ \overline{\boldsymbol{\mu}}(\boldsymbol{b}) : \boldsymbol{b} \in \mathscr{B}(d)\}$.

Let $\boldsymbol{b}$ and $\boldsymbol{b}' \in \mathscr{B}(d_0)$ such that $\rho\ \boldsymbol{\mu}(\boldsymbol{b}) + (1-\rho)\ \overline{\boldsymbol{\mu}}(\boldsymbol{b}) = \rho\ \boldsymbol{\mu}(\boldsymbol{b}') + (1-\rho)\ \overline{\boldsymbol{\mu}}(\boldsymbol{b}')$, then

$$0 = (\boldsymbol{b} - \boldsymbol{b}^{'})^{\top}\{\rho\ \boldsymbol{\mu}(\boldsymbol{b}) + (1-\rho)\ \overline{\boldsymbol{\mu}}(\boldsymbol{b}) - \rho\ \boldsymbol{\mu}(\boldsymbol{b}^{'}) - (1-\rho)\ \overline{\boldsymbol{\mu}}(\boldsymbol{b}^{'})\}$$
$$= \rho\ E\left\{\left(\boldsymbol{X}^{\top}\boldsymbol{b} - \boldsymbol{X}^{\top}\boldsymbol{b}^{'}\right)\left[\widetilde{F}\left(e^{\boldsymbol{X}^{\top}\boldsymbol{b}}\,\middle|\,\boldsymbol{X}\right) - \widetilde{F}\left(e^{\boldsymbol{X}^{\top}\boldsymbol{b}^{'}}\,\middle|\,\boldsymbol{X}\right)\right]\right\}$$
$$+ (1-\rho)\ E\left\{\left[\boldsymbol{Z}^{\top}\boldsymbol{b} - \boldsymbol{Z}^{\top}\boldsymbol{b}^{'}\right]\left[\widetilde{F}\left(e^{\boldsymbol{Z}^{\top}\boldsymbol{b}}\,\middle|\,\boldsymbol{Z}\right) - \widetilde{F}\left(e^{\boldsymbol{Z}^{\top}\boldsymbol{b}^{'}}\,\middle|\,\boldsymbol{Z}\right)\right]\right\}.$$

By condition **C5(a)**, the above equation holds if and only if $\boldsymbol{X}^{\top}\boldsymbol{b} = \boldsymbol{X}^{\top}\boldsymbol{b}^{'}$ and $\boldsymbol{Z}^{\top}\boldsymbol{b} = \boldsymbol{Z}^{\top}\boldsymbol{b}^{'}$ with probability 1.

By the positive definiteness of $E[\boldsymbol{X}^{\otimes 2}]$ and $E[\boldsymbol{Z}^{\otimes 2}]$, it is clear that $\boldsymbol{b} = \boldsymbol{b}^{'}$. So there exists an inverse function $\boldsymbol{\eta}$, from $\mathscr{A}(d_0)$ to $\mathscr{B}(d_0)$ such that $\boldsymbol{\eta}\{\rho\ \boldsymbol{\mu}(\boldsymbol{b}) + (1-\rho)\ \overline{\boldsymbol{\mu}}(\boldsymbol{b})\} = \boldsymbol{b}$ for any $\boldsymbol{b} \in \mathscr{B}(d_0)$. Now we conclude that under the condition **C5**, $\rho\ \boldsymbol{\mu} + (1-\rho)\ \overline{\boldsymbol{\mu}}$ is also a one to one mapping from $\mathscr{B}(d_0)$ to $\mathscr{A}(d_0)$.

According to our estimating procedure, $\rho\ \Omega^{\mathbb{V}}_{m_v}(\widehat{\boldsymbol{\beta}}, \tau_k) + (1-\rho)\ \Omega^{\overline{\mathbb{V}}}_{m_n}(\widehat{\boldsymbol{\beta}}, \tau_k) + O_p\left(\frac{\mathbf{1}}{\sqrt{n}}\right) + o_{(0,\tau_U])}(\mathbf{1}) = 0$, which implies

$$\frac{\rho}{m_v}\sum_{j\in\mathbb{V}}\boldsymbol{X}_j\mathbb{N}_j\left(e^{\boldsymbol{X}_j^{\top}\widehat{\boldsymbol{\beta}}(\tau_k)}\right) + \frac{1-\rho}{m_n}\sum_{l\in\overline{\mathbb{V}}}\boldsymbol{Z}_l\ \mathbb{N}_l\left(e^{\boldsymbol{Z}_l^{\top}\widehat{\boldsymbol{\beta}}(\tau_k)}\right)$$
$$= \frac{\rho}{m_v}\sum_{j\in\mathbb{V}}\int_0^{\tau_k}\boldsymbol{X}_j\mathbb{I}\left[Y_j \geq e^{\boldsymbol{X}_j^{\top}\widehat{\boldsymbol{\beta}}(u)}\right]dH(u)$$
$$+ \frac{1-\rho}{m_n}\sum_{l\in\overline{\mathbb{V}}}\int_0^{\tau_k}\boldsymbol{Z}_l\mathbb{I}\left[Y_l \geq e^{\boldsymbol{Z}_l^{\top}\widehat{\boldsymbol{\beta}}(u)}\right]dH(u) + O_p\left(\frac{\mathbf{1}}{\sqrt{n}}\right) + o_{(0,\tau_U])}(\mathbf{1}).$$

Simple algebra leads to

$$\frac{\rho}{m_v}\sum_{j\in\mathbb{V}}\boldsymbol{X}_j\mathbb{N}_j\left(e^{\boldsymbol{X}_j^{\top}\widehat{\boldsymbol{\beta}}(\tau_k)}\right) + \frac{1-\rho}{m_n}\sum_{l\in\overline{\mathbb{V}}}\boldsymbol{Z}_l\ \mathbb{N}_l\left(e^{\boldsymbol{Z}_l^{\top}\widehat{\boldsymbol{\beta}}(\tau_k)}\right)$$
$$- \rho\ E\left[\boldsymbol{X}\mathbb{N}\left(e^{\boldsymbol{X}^{\top}\boldsymbol{\beta}_0(\tau_k)}\right)\right] - (1-\rho)\ E\left[\boldsymbol{Z}\mathbb{N}\left(e^{\boldsymbol{Z}^{\top}\boldsymbol{\beta}_0(\tau_k)}\right)\right]$$
$$= \rho\left[\boldsymbol{\Gamma}_{m_v}\left\{\widehat{\boldsymbol{\beta}}(\tau_k)\right\} + \boldsymbol{\mu}\left\{\widehat{\boldsymbol{\beta}}(\tau_k)\right\} - \boldsymbol{\mu}\left\{\boldsymbol{\beta}_0(\tau_k)\right\}\right]$$
$$+ (1-\rho)\left[\overline{\boldsymbol{\Gamma}}_{m_n}\left\{\widehat{\boldsymbol{\beta}}(\tau_k)\right\} + \overline{\boldsymbol{\mu}}\left\{\widehat{\boldsymbol{\beta}}(\tau_k)\right\} - \overline{\boldsymbol{\mu}}\left\{\boldsymbol{\beta}_0(\tau_k)\right\}\right]$$

and

$$\frac{\rho}{m_v}\sum_{j\in\mathbb{V}}\int_0^{\tau_k}\boldsymbol{X}_j\mathbb{I}\left[Y_j \geq e^{\boldsymbol{X}_j^{\top}\widehat{\boldsymbol{\beta}}(u)}\right]dH(u)$$

$$+ \frac{1-\rho}{m_n} \sum_{l \in \overline{\mathbb{V}}} \int_0^{\tau_k} \boldsymbol{Z}_l \mathbb{I} \left[Y_l \geq e^{\boldsymbol{Z}_l^\top \widehat{\boldsymbol{\beta}}(u)} \right] dH(u)$$

$$- \rho \; E \left[\boldsymbol{X} \int_0^{\tau_k} \mathbb{I} \left[Y \geq e^{\boldsymbol{X}^\top \boldsymbol{\beta}_0(u)} \right] dH(u) \right]$$

$$- (1-\rho) \; E \left[\boldsymbol{Z} \int_0^{\tau_k} \mathbb{I} \left[Y \geq e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(u)} \right] dH(u) \right]$$

$$= \rho \left[\int_0^{\tau_k} \boldsymbol{\Gamma}_{m_v}^* \left\{ \widehat{\boldsymbol{\beta}}(u) \right\} dH(u) + \int_0^{\tau_k} \left[\boldsymbol{\mu}^* \left\{ \widehat{\boldsymbol{\beta}}(u) \right\} - \boldsymbol{\mu}^* \left\{ \boldsymbol{\beta}_0(u) \right\} \right] dH(u) \right]$$

$$+ (1-\rho) \left[\int_0^{\tau_k} \overline{\boldsymbol{\Gamma}}_{m_n}^* \left\{ \widehat{\boldsymbol{\beta}}(u) \right\} dH(u) + \int_0^{\tau_k} \left[\overline{\boldsymbol{\mu}}^* \left\{ \widehat{\boldsymbol{\beta}}(u) \right\} - \overline{\boldsymbol{\mu}}^* \left\{ \boldsymbol{\beta}_0(u) \right\} \right] dH(u) \right].$$

By martingale property,

$$E \left[\boldsymbol{X} \mathbb{N} \left(e^{\boldsymbol{X}^\top \boldsymbol{\beta}_0(\tau_k)} \right) \right] = E \left[\boldsymbol{X} \int_0^{\tau_k} \mathbb{I} \left[Y \geq e^{\boldsymbol{X}^\top \boldsymbol{\beta}_0(u)} \right] dH(u) \right] \text{ and}$$

$$E \left[\boldsymbol{Z} \mathbb{N} \left(e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau_k)} \right) \right] = E \left[\boldsymbol{Z} \int_0^{\tau_k} \mathbb{I} \left[Y \geq e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(u)} \right] dH(u) \right].$$

Then combining previous two equations,

$$\rho \left[\boldsymbol{\mu} \left\{ \widehat{\boldsymbol{\beta}}(\tau_k) \right\} - \boldsymbol{\mu} \left\{ \boldsymbol{\beta}_0(\tau_k) \right\} \right] + (1-\rho) \left[\overline{\boldsymbol{\mu}} \left\{ \widehat{\boldsymbol{\beta}}(\tau_k) \right\} - \overline{\boldsymbol{\mu}} \left\{ \boldsymbol{\beta}_0(\tau_k) \right\} \right]$$

$$= -\rho \; \boldsymbol{\Gamma}_{m_v} \left\{ \widehat{\boldsymbol{\beta}}(\tau_k) \right\} + \rho \int_0^{\tau_k} \boldsymbol{\Gamma}_{m_v}^* \left\{ \widehat{\boldsymbol{\beta}}(u) \right\} dH(u)$$

$$+ \rho \sum_{r=1}^{k} \int_{\tau_{r-1}}^{\tau_r} \left[\boldsymbol{\mu}^* \left\{ \widehat{\boldsymbol{\beta}}(u) \right\} - \boldsymbol{\mu}^* \left\{ \boldsymbol{\beta}_0(u) \right\} \right] dH(u)$$

$$- (1-\rho) \; \overline{\boldsymbol{\Gamma}}_{m_n} \left\{ \widehat{\boldsymbol{\beta}}(\tau_k) \right\} + (1-\rho) \int_0^{\tau_k} \overline{\boldsymbol{\Gamma}}_{m_n}^* \left\{ \widehat{\boldsymbol{\beta}}(u) \right\} dH(u)$$

$$+ (1-\rho) \sum_{r=1}^{k} \int_{\tau_{r-1}}^{\tau_r} \left[\overline{\boldsymbol{\mu}}^* \left\{ \widehat{\boldsymbol{\beta}}(u) \right\} - \overline{\boldsymbol{\mu}}^* \left\{ \boldsymbol{\beta}_0(u) \right\} \right] dH(u) + O_p \left(\frac{\mathbf{1}}{\sqrt{n}} \right) + o_{(0,\tau_U])}(\mathbf{1})$$

$$= \rho \left[-\boldsymbol{\Gamma}_{m_v} \left\{ \widehat{\boldsymbol{\beta}}(\tau_k) \right\} + \int_0^{\tau_k} \boldsymbol{\Gamma}_{m_v}^* \left\{ \widehat{\boldsymbol{\beta}}(u) \right\} dH(u) \right]$$

$$- (1-\rho) \left[\overline{\boldsymbol{\Gamma}}_{m_n} \left\{ \widehat{\boldsymbol{\beta}}(\tau_k) \right\} + \int_0^{\tau_k} \overline{\boldsymbol{\Gamma}}_{m_n}^* \left\{ \widehat{\boldsymbol{\beta}}(u) \right\} dH(u) \right]$$

$$+ \sum_{r=1}^{k} \int_{\tau_{r-1}}^{\tau_r} \Bigg(\rho \; \left[\boldsymbol{\mu}^* \left\{ \widehat{\boldsymbol{\beta}}(u) \right\} - \boldsymbol{\mu}^* \left\{ \boldsymbol{\beta}_0(u) \right\} \right]$$

$$+ (1-\rho) \left[\overline{\boldsymbol{\mu}}^* \left\{ \widehat{\boldsymbol{\beta}}(u) \right\} - \overline{\boldsymbol{\mu}}^* \left\{ \boldsymbol{\beta}_0(u) \right\} \right] \Bigg) dH(u)$$

$$+ O_p\left(\frac{\mathbf{1}}{\sqrt{n}}\right) + o_{(0,\tau_U])}(\mathbf{1}). \tag{A.1}$$

Consider

$$G_1 = \left\{\boldsymbol{X}_j \mathbb{I}\left[Y_j \le e^{\boldsymbol{X}_j^\top \boldsymbol{b}}\right]\delta_j : \boldsymbol{b} \in \Re^p\right\}, \quad G_2 = \left\{\boldsymbol{X}_j \mathbb{I}\left[Y_j \ge e^{\boldsymbol{X}_j^\top \boldsymbol{b}}\right] : \boldsymbol{b} \in \Re^p\right\},$$
$$\overline{G}_1 = \left\{\boldsymbol{Z}_l \mathbb{I}\left[Y_l \le e^{\boldsymbol{Z}_l^\top \boldsymbol{b}}\right]\delta_l : \boldsymbol{b} \in \Re^p\right\} \text{ and } \overline{G}_2 = \left\{\boldsymbol{Z}_l \mathbb{I}\left[Y_l \ge e^{\boldsymbol{Z}_l^\top \boldsymbol{b}}\right] : \boldsymbol{b} \in \Re^p\right\}.$$

Since the class of indicator functions of polytopes in $\Re^p$ is Glivenko Cantelli and $\boldsymbol{X}_j$ and $\boldsymbol{Z}_l$ are bounded, so here all the G_1, G_2, $\overline{G}_1$ and $\overline{G}_2$ are Glivenko Cantelli [van der Vaart and Wellner, 1996]. So $\sup_{\boldsymbol{b}\in\Re^p} \|\boldsymbol{\Gamma}_{m_v}(\boldsymbol{b})\| \xrightarrow{a.s} 0$, $\sup_{\boldsymbol{b}\in\Re^p} \|\boldsymbol{\Gamma}^*_{m_v}(\boldsymbol{b})\| \xrightarrow{a.s} 0$, $\sup_{\boldsymbol{b}\in\Re^p} \|\overline{\boldsymbol{\Gamma}}_{m_n}(\boldsymbol{b})\| \xrightarrow{a.s} 0$ and $\sup_{\boldsymbol{b}\in\Re^p} \|\overline{\boldsymbol{\Gamma}}^*_{m_n}(\boldsymbol{b})\| \xrightarrow{a.s} 0$ (Glivenko Cantelli theorem). Then, for any given C_1 and $\overline{C}_1$ (> 0) and for sufficiently large m_v and n, $\sup_k \left\| -\boldsymbol{\Gamma}_{m_v}\{\widehat{\boldsymbol{\beta}}(\tau_k)\} + \int_0^{\tau_k} \boldsymbol{\Gamma}^*_{m_v}\{\widehat{\boldsymbol{\beta}}(u)\}dH(u)\right\| < C_1$ and $\sup_k \left\| -\overline{\boldsymbol{\Gamma}}_{m_n}\{\widehat{\boldsymbol{\beta}}(\tau_k)\} + \int_0^{\tau_k} \overline{\boldsymbol{\Gamma}}^*_{m_n}\{\widehat{\boldsymbol{\beta}}(u)\}dH(u)\right\| < \overline{C}_1$ with probability 1.

There exists $C_2 > 0$ such that $\sup_i \|\boldsymbol{X}_i\| < C_2$ and $\sup_i \|\boldsymbol{Z}_i\| < C_2$ (by **C3**).

For some constants C_3 and $\overline{C}_3$ (> 0), $\|\boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau) - \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau')\}\| \le C_3|\tau - \tau'|$ and $\|\overline{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau) - \overline{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau')\}\| \le \overline{C_3}|\tau - \tau'|$ (by **C4(a)**) for any τ, $\tau' \in (0, \tau_U]$.

There exists C_4 and $\overline{C}_4$ (> 0) such that $\|\{B(\boldsymbol{b})\}^{-1}B^*(\boldsymbol{b})\mathbf{y}\| \le C_4\|\mathbf{y}\|$; $\boldsymbol{b} \in \mathscr{B}(d_0)$ and $\|\{\overline{B}(\boldsymbol{b})\}^{-1}\overline{B}^*(\boldsymbol{b})\mathbf{y}\| \le \overline{C}_4\|\mathbf{y}\|$; $\boldsymbol{b} \in \mathscr{B}(d_0)$ for any $\mathbf{y} \in \Re^p$ (by **C5(c)**).

Define $\mathscr{C}_1 = \rho\ C_1 + (1-\rho)\ \overline{C}_1$, $\mathscr{C}_2 = C_2$, $\mathscr{C}_3 = \rho\ C_3 + (1-\rho)\ \overline{C}_3$ and $\mathscr{C}_4 = \rho\ C_4 + (1-\rho)\ \overline{C}_4$.

For given n, define a sequence $\{\mathcal{E}_u\}_{u=0}^{L-1}$, where $\mathcal{E}_0 = \mathscr{C}_3 a_n$, $\mathcal{E}_1 = \mathscr{C}_1 + \mathscr{C}_2(1/n) + \mathscr{C}_3 a_n + \mathcal{E}_0 \mathscr{C}_4 b_n$ and $\mathcal{E}_u = \mathscr{C}_1 + \mathscr{C}_2(1/n) + \mathscr{C}_3 a_n + \left(\sum_{r=0}^{u-1} \mathcal{E}_r\right)\mathscr{C}_4 b_n$ for $u = 2, 3, \ldots, L-1$. By the definition of $\mathcal{E}_u$, $\mathcal{E}_u - \mathcal{E}_{u-1} = \mathcal{E}_{u-1}\mathscr{C}_4 b_n$, hence $\mathcal{E}_u = (1 + \mathscr{C}_4 b_n)^{u-1}\mathcal{E}_1$.

Given that $\lim_{n\to\infty} a_n = 0$ and $L = \tau_U/a_n$, implies that $\lim_{n\to\infty}(1+\mathscr{C}_4 b_n)^{L-1} = \exp\{\mathscr{C}_4\tau_U/(1-\tau_U)\}$. Since $\mathcal{E}_u$ is increasing with u, and for some N_0 such that $n \ge N_0$, we can choose sufficiently small $\mathscr{C}_1$ so that $\mathcal{E}_u \le 2\exp\{\tau_U/(1-\tau_U)\}\mathscr{C}_1 \le d_0$ for all $u = 0, 1, \ldots, L-1$.

Next we prove that

$$\sup_{\tau_u \le \tau < \tau_{u+1}} \left\|\rho\left[\boldsymbol{\mu}\{\widehat{\boldsymbol{\beta}}(\tau)\} - \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)\}\right] + (1-\rho)\left[\overline{\boldsymbol{\mu}}\{\widehat{\boldsymbol{\beta}}(\tau)\} - \overline{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\}\right]\right\| < \mathcal{E}_u,$$

for $u = 0, 1, \ldots, L-1$.

Considering $n \geq N_0$, and by the definition of $\widehat{\boldsymbol{\beta}}(\tau)$,

$$\begin{aligned}\sup_{\tau_0 \leq \tau < \tau_1} &\left\| \rho \left[\boldsymbol{\mu}\{\widehat{\boldsymbol{\beta}}(\tau)\} - \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)\} \right] + (1-\rho) \left[\overline{\boldsymbol{\mu}}\{\widehat{\boldsymbol{\beta}}(\tau)\} - \overline{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\} \right] \right\| \\ &= \sup_{\tau_0 \leq \tau < \tau_1} \{ \| \rho \ \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)\} + (1-\rho) \ \overline{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\} \| \} \\ &\leq \mathscr{C}_3 a_n = \mathcal{E}_0.\end{aligned}$$

Considering (A.1) with $k = 1$, for $\tau \in [\tau_0, \tau_1)$,

$$\begin{aligned}&\left\| \rho \left[\boldsymbol{\mu}^* \left\{ \widehat{\boldsymbol{\beta}}(\tau_k) \right\} - \boldsymbol{\mu}^* \left\{ \boldsymbol{\beta}_0(\tau_k) \right\} \right] + (1-\rho) \left[\overline{\boldsymbol{\mu}}^* \left\{ \widehat{\boldsymbol{\beta}}(\tau_k) \right\} - \overline{\boldsymbol{\mu}}^* \left\{ \boldsymbol{\beta}_0(\tau_k) \right\} \right] \right\| \\ &\quad = \left\| \rho \Big[\boldsymbol{\mu}^* \left[\boldsymbol{\eta} \left\{ \widehat{\boldsymbol{\alpha}}(\tau) \right\} \right] - \boldsymbol{\mu}^* \left[\boldsymbol{\eta} \left\{ \boldsymbol{\alpha}_0(\tau) \right\} \right] \Big] + (1-\rho) \Big[\overline{\boldsymbol{\mu}}^* \left[\boldsymbol{\eta} \left\{ \widehat{\boldsymbol{\alpha}}(\tau) \right\} \right] - \overline{\boldsymbol{\mu}}^* \left[\boldsymbol{\eta} \left\{ \boldsymbol{\alpha}_0(\tau) \right\} \right] \Big] \right\| \\ &\quad = \Big\| \rho \left(\boldsymbol{B} \left[\boldsymbol{\eta} \left\{ \breve{\boldsymbol{\alpha}}(\tau) \right\} \right] \right)^{-1} \boldsymbol{B}^* \left[\boldsymbol{\eta} \left\{ \breve{\boldsymbol{\alpha}}(\tau) \right\} \right] \left\{ \widehat{\boldsymbol{\alpha}}(\tau) - \boldsymbol{\alpha}_0(\tau) \right\} \\ &\qquad\qquad + (1-\rho) \left(\overline{\boldsymbol{B}} \left[\boldsymbol{\eta} \left\{ \breve{\boldsymbol{\alpha}}(\tau) \right\} \right] \right)^{-1} \overline{\boldsymbol{B}}^* \left[\boldsymbol{\eta} \left\{ \breve{\boldsymbol{\alpha}}(\tau) \right\} \right] \left\{ \widehat{\boldsymbol{\alpha}}(\tau) - \boldsymbol{\alpha}_0(\tau) \right\} \Big\| \\ &\quad \leq \mathscr{C}_4 \mathcal{E}_0,\end{aligned}$$

where $\breve{\boldsymbol{\alpha}}(\tau)$ is between $\widehat{\boldsymbol{\alpha}}(\tau)$ and $\boldsymbol{\alpha}_0(\tau)$. So using conditions defined earlier, the norm of the right hand side of (A.1) is not bigger than $\mathscr{C}_1 + \mathcal{E}_0 \mathscr{C}_4 b_n + \mathscr{C}_2(1/n)$; So,

$$\begin{aligned}\sup_{\tau_1 \leq \tau < \tau_2} &\left\| \rho \left[\boldsymbol{\mu}\{\widehat{\boldsymbol{\beta}}(\tau)\} - \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)\} \right] + (1-\rho) \left[\overline{\boldsymbol{\mu}}\{\widehat{\boldsymbol{\beta}}(\tau)\} - \overline{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\} \right] \right\| \\ &\leq \left\| \rho \left[\boldsymbol{\mu}\{\widehat{\boldsymbol{\beta}}(\tau_1)\} - \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau_1)\} \right] + (1-\rho) \left[\overline{\boldsymbol{\mu}}\{\widehat{\boldsymbol{\beta}}(\tau_1)\} - \overline{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau_1)\} \right] \right\| \\ &\quad + \sup_{\tau_1 \leq \tau < \tau_2} \left\| \rho \left[\boldsymbol{\mu}\{\widehat{\boldsymbol{\beta}}(\tau_1)\} - \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)\} \right] + (1-\rho) \left[\overline{\boldsymbol{\mu}}\{\widehat{\boldsymbol{\beta}}(\tau_1)\} - \overline{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\} \right] \right\| \\ &\leq \mathscr{C}_1 + \frac{\mathscr{C}_2}{n} + \mathscr{C}_3 a_n + \mathcal{E}_0 \mathscr{C}_4 b_n = \mathcal{E}_1.\end{aligned}$$

Using similar approach, we can arrive $\widehat{\boldsymbol{\beta}}(\tau_u) \in \mathscr{B}(d_0)$ and

$$\sup_{\tau_u \leq \tau < \tau_{u+1}} \left\| \rho \left[\boldsymbol{\mu}\{\widehat{\boldsymbol{\beta}}(\tau)\} - \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)\} \right] + (1-\rho) \left[\overline{\boldsymbol{\mu}}\{\widehat{\boldsymbol{\beta}}(\tau)\} - \overline{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\} \right] \right\| \leq \mathcal{E}_u,$$

for all $u = 2, 3, \ldots, L-1$. As n increases, $a_n \to 0$ and $\mathscr{C}_1$ can become arbitrarily small, which implies that

$$\sup_{0 < \tau < \tau_U} \left\| \rho \left[\boldsymbol{\mu}\{\widehat{\boldsymbol{\beta}}(\tau)\} - \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)\} \right] + (1-\rho) \left[\overline{\boldsymbol{\mu}}\{\widehat{\boldsymbol{\beta}}(\tau)\} - \overline{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\} \right] \right\| \xrightarrow{\text{Pr}} 0.$$

Using Taylor series expansion of $\boldsymbol{\eta}\{\widehat{\boldsymbol{\alpha}}(\tau)\}$ at $\boldsymbol{\alpha}_0(\tau)$ for $\tau \in [\nu, \tau_U]$, from condition **C6**, we arrive that

$$\begin{aligned}\left\|\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\right\| &\leq \Big\|\rho \left(\boldsymbol{B}\left\{\boldsymbol{\beta}_0(\tau)\right\}\right)^{-1}\left\{\widehat{\boldsymbol{\alpha}}(\tau) - \boldsymbol{\alpha}_0(\tau)\right\} \\ &\qquad + (1-\rho)\left(\overline{\boldsymbol{B}}\left\{\boldsymbol{\beta}_0(\tau)\right\}\right)^{-1}\left\{\widehat{\boldsymbol{\alpha}}(\tau) - \boldsymbol{\alpha}_0(\tau)\right\}\Big\| + \|\epsilon_n^*(\tau)\| \\ &\leq \mathscr{C}_6 \left\|\widehat{\boldsymbol{\alpha}}(\tau) - \boldsymbol{\alpha}_0(\tau)\right\| + \left\|\epsilon_n^*(\tau)\right\|,\end{aligned}$$

where $\mathscr{C}_6(> 0)$ is independent of τ and $\sup_{\nu\leq\tau\leq\tau_U} \|\epsilon_n^*(\tau)\| \xrightarrow{\Pr} 0$. Hence the consistency proof.

Appendix B

Proof of Theorem 2.2.2

Lemma B.1. *For any sequence,* $\left\{\widetilde{\boldsymbol{\beta}}_n(\mathcal{T}),\ \mathcal{T} \in (0, \mathcal{T}_U]\right\}_{n=1}^{\infty}$, *we have*

$$\sup_{\mathcal{T}\in(0,\mathcal{T}_U]} \left\| \frac{\sqrt{\rho}}{\sqrt{m_v}} \sum_{j\in\mathbb{V}} \boldsymbol{X}_j \left[\mathbb{N}_j \left(e^{\boldsymbol{X}_j^\top \widetilde{\boldsymbol{\beta}}_n(\mathcal{T})} \right) - \mathbb{N}_j \left(e^{\boldsymbol{X}_j^\top \boldsymbol{\beta}_0(\mathcal{T})} \right) \right] \right.$$

$$+ \frac{\sqrt{(1-\rho)}}{\sqrt{m_n}} \sum_{l\in\overline{\mathbb{V}}} \boldsymbol{Z}_l \left[\mathbb{N}_l \left(e^{\boldsymbol{Z}_l^\top \widetilde{\boldsymbol{\beta}}_n(\mathcal{T})} \right) - \mathbb{N}_l \left(e^{\boldsymbol{Z}_l^\top \boldsymbol{\beta}_0(\mathcal{T})} \right) \right]$$

$$\left. - \sqrt{\rho\ m_v} \left[\boldsymbol{\mu}\{\widetilde{\boldsymbol{\beta}}_n(\mathcal{T})\} - \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\mathcal{T})\} \right] - \sqrt{(1-\rho)m_n} \left[\overline{\boldsymbol{\mu}}\{\widetilde{\boldsymbol{\beta}}_n(\mathcal{T})\} - \overline{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\mathcal{T})\} \right] \right\| \xrightarrow{\text{Pr}} 0,$$

if

$$\sup_{\mathcal{T}\in(0,\mathcal{T}_U]} \left\| \rho \left[\boldsymbol{\mu}\{\widetilde{\boldsymbol{\beta}}_n(\mathcal{T})\} - \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\mathcal{T})\} \right] + (1-\rho) \left[\overline{\boldsymbol{\mu}}\{\widetilde{\boldsymbol{\beta}}_n(\mathcal{T})\} - \overline{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\mathcal{T})\} \right] \right\| \xrightarrow{\text{Pr}} 0.$$

***Proof of Lemma* B.1**: Define $\mu_1(\boldsymbol{b}) = E\left[\mathbb{N}\left(e^{\boldsymbol{X}^\top \boldsymbol{b}}\right)\right]$, $\overline{\mu}_1(\boldsymbol{b}) = E\left[\mathbb{N}\left(e^{\boldsymbol{Z}^\top \boldsymbol{b}}\right)\right]$, and

$$\sigma_d^2(\boldsymbol{b}) = \text{Var}\left\{ \sqrt{\rho} \left[\mathbb{N}\left(e^{\boldsymbol{X}^\top \boldsymbol{b}}\right) - \mathbb{N}\left(e^{\boldsymbol{X}^\top \boldsymbol{\beta}_0(\mathcal{T})}\right) - \mu_1\{\boldsymbol{b}\} + \mu_1\{\boldsymbol{\beta}_0(\mathcal{T})\} \right] \right.$$

$$\left. + \sqrt{(1-\rho)} \left[\mathbb{N}\left(e^{\boldsymbol{Z}^\top \boldsymbol{b}}\right) - \mathbb{N}\left(e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\mathcal{T})}\right) - \overline{\mu}_1\{\boldsymbol{b}\} + \overline{\mu}_1\{\boldsymbol{\beta}_0(\mathcal{T})\} \right] \right\}.$$

Provided $\boldsymbol{X}$ and $\boldsymbol{Z}$ are bounded and errors are independent, it suffices to prove that $\sigma_d^2\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} \xrightarrow{\text{Pr}} 0$, by following the arguments provided in Alexander [1984] and Lai and Ying [1988]. If $\widetilde{\boldsymbol{\beta}}_n(\tau)$ is fixed,

$$
\begin{aligned}
&\sigma_d^2\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} \\
&\quad = \mathrm{Var}\Bigg\{\sqrt{\rho}\left[\mathbb{N}\left(e^{\boldsymbol{X}^\top\widetilde{\boldsymbol{\beta}}_n(\tau)}\right) - \mathbb{N}\left(e^{\boldsymbol{X}^\top\boldsymbol{\beta}_0(\tau)}\right) - \mu_1\{\widetilde{\boldsymbol{\beta}}_n(\tau)\} + \mu_1\{\boldsymbol{\beta}_0(\tau)\}\right] \\
&\qquad + \sqrt{(1-\rho)}\left[\mathbb{N}\left(e^{\boldsymbol{Z}^\top\widetilde{\boldsymbol{\beta}}_n(\tau)}\right) - \mathbb{N}\left(e^{\boldsymbol{Z}^\top\boldsymbol{\beta}_0(\tau)}\right) - \overline{\mu}_1\{\widetilde{\boldsymbol{\beta}}_n(\tau)\} + \overline{\mu}_1\{\boldsymbol{\beta}_0(\tau)\}\right]\Bigg\} \\
&\quad = \rho\ \mathrm{Var}\left[\mathbb{N}\left(e^{\boldsymbol{X}^\top\widetilde{\boldsymbol{\beta}}_n(\tau)}\right) - \mathbb{N}\left(e^{\boldsymbol{X}^\top\boldsymbol{\beta}_0(\tau)}\right) - \mu_1\{\widetilde{\boldsymbol{\beta}}_n(\tau)\} + \mu_1\{\boldsymbol{\beta}_0(\tau)\}\right] \\
&\qquad + (1-\rho)\ \mathrm{Var}\left[\mathbb{N}\left(e^{\boldsymbol{Z}^\top\widetilde{\boldsymbol{\beta}}_n(\tau)}\right) - \mathbb{N}\left(e^{\boldsymbol{Z}^\top\boldsymbol{\beta}_0(\tau)}\right) - \overline{\mu}_1\{\widetilde{\boldsymbol{\beta}}_n(\tau)\} + \overline{\mu}_1\{\boldsymbol{\beta}_0(\tau)\}\right] \\
&\qquad + 2\sqrt{\rho(1-\rho)}\ \mathrm{Cov}\Bigg[\mathbb{N}\left(e^{\boldsymbol{X}^\top\widetilde{\boldsymbol{\beta}}_n(\tau)}\right) - \mathbb{N}\left(e^{\boldsymbol{X}^\top\boldsymbol{\beta}_0(\tau)}\right) - \mu_1\{\widetilde{\boldsymbol{\beta}}_n(\tau)\} + \mu_1\{\boldsymbol{\beta}_0(\tau)\}, \\
&\qquad\qquad \mathbb{N}\left(e^{\boldsymbol{Z}^\top\widetilde{\boldsymbol{\beta}}_n(\tau)}\right) - \mathbb{N}\left(e^{\boldsymbol{Z}^\top\boldsymbol{\beta}_0(\tau)}\right) - \overline{\mu}_1\{\widetilde{\boldsymbol{\beta}}_n(\tau)\} + \overline{\mu}_1\{\boldsymbol{\beta}_0(\tau)\}\Bigg] \\
&\quad \le \rho\ \mathrm{Var}\left[\mathbb{N}\left(e^{\boldsymbol{X}^\top\widetilde{\boldsymbol{\beta}}_n(\tau)}\right) - \mathbb{N}\left(e^{\boldsymbol{X}^\top\boldsymbol{\beta}_0(\tau)}\right) - \mu_1\{\widetilde{\boldsymbol{\beta}}_n(\tau)\} + \mu_1\{\boldsymbol{\beta}_0(\tau)\}\right] \\
&\qquad + (1-\rho)\ \mathrm{Var}\left[\mathbb{N}\left(e^{\boldsymbol{Z}^\top\widetilde{\boldsymbol{\beta}}_n(\tau)}\right) - \mathbb{N}\left(e^{\boldsymbol{Z}^\top\boldsymbol{\beta}_0(\tau)}\right) - \overline{\mu}_1\{\widetilde{\boldsymbol{\beta}}_n(\tau)\} + \overline{\mu}_1\{\boldsymbol{\beta}_0(\tau)\}\right] \\
&\qquad + 2\sqrt{\rho\ \mathrm{Var}\left[\mathbb{N}\left(e^{\boldsymbol{X}^\top\widetilde{\boldsymbol{\beta}}_n(\tau)}\right) - \mathbb{N}\left(e^{\boldsymbol{X}^\top\boldsymbol{\beta}_0(\tau)}\right) - \mu_1\{\widetilde{\boldsymbol{\beta}}_n(\tau)\} + \mu_1\{\boldsymbol{\beta}_0(\tau)\}\right]} \\
&\qquad \times \sqrt{(1-\rho)\ \mathrm{Var}\left[\mathbb{N}\left(e^{\boldsymbol{Z}^\top\widetilde{\boldsymbol{\beta}}_n(\tau)}\right) - \mathbb{N}\left(e^{\boldsymbol{Z}^\top\boldsymbol{\beta}_0(\tau)}\right) - \overline{\mu}_1\{\widetilde{\boldsymbol{\beta}}_n(\tau)\} + \overline{\mu}_1\{\boldsymbol{\beta}_0(\tau)\}\right]} \\
&\quad = \rho\ \sigma_{1d}^2\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} + (1-\rho)\ \sigma_{2d}^2\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} + 2\sqrt{\rho(1-\rho)\ \sigma_{1d}^2\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\}\sigma_{2d}^2\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\}}.
\end{aligned}
$$

Following the arguments given in Appendix B of Peng and Huang [2008], we can show that $\sigma_{1d}^2\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} \xrightarrow{\Pr} 0$ and $\sigma_{2d}^2\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} \xrightarrow{\Pr} 0$. This completes the proof of $\sigma_d^2\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} \xrightarrow{\Pr} 0$ and *Lemma* B.1.

Proof of Theorem 2.2.2

From the proofs of *Theorem* 2.2.1 and *Lemma* B.1, we have

$$
\sup_{\tau\in(0,\tau_U]}\left\|\frac{\sqrt{\rho}}{\sqrt{m_v}}\sum_{j\in\mathbb{V}}\boldsymbol{X}_j\left[\mathbb{N}_j\left(e^{\boldsymbol{X}_j^\top\widehat{\boldsymbol{\beta}}(\tau)}\right) - \mathbb{N}_j\left(e^{\boldsymbol{X}_j^\top\boldsymbol{\beta}_0(\tau)}\right)\right]\right.
$$

$$+ \frac{\sqrt{(1-\rho)}}{\sqrt{m_n}} \sum_{l \in \overline{\mathbb{V}}} \boldsymbol{Z}_l \left[\mathbb{N}_l \left(e^{\boldsymbol{Z}_l^\top \widehat{\boldsymbol{\beta}}(\tau)} \right) - \mathbb{N}_l \left(e^{\boldsymbol{Z}_l^\top \boldsymbol{\beta}_0(\tau)} \right) \right]$$
$$- \sqrt{\rho\, m_v} \left[\boldsymbol{\mu}\{\widehat{\boldsymbol{\beta}}(\tau)\} - \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)\} \right] - \sqrt{(1-\rho) m_n} \left[\overline{\boldsymbol{\mu}}\{\widehat{\boldsymbol{\beta}}(\tau)\} - \overline{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\} \right] \Bigg\| \xrightarrow{\Pr} 0, \tag{B.1}$$

Similarly

$$\sup_{\tau \in (0,\tau_U]} \Bigg\| \frac{\sqrt{\rho}}{\sqrt{m_v}} \sum_{j \in \mathbb{V}} \boldsymbol{X}_j \left(\mathbb{I}\left[Y_j \geq e^{\boldsymbol{X}_j^\top \widehat{\boldsymbol{\beta}}(\tau)} \right] - \mathbb{I}\left[Y_j \geq e^{\boldsymbol{X}_j^\top \boldsymbol{\beta}_0(\tau)} \right] \right)$$
$$+ \frac{\sqrt{(1-\rho)}}{\sqrt{m_n}} \sum_{l \in \overline{\mathbb{V}}} \boldsymbol{Z}_l \left(\mathbb{I}\left[Y_l \geq e^{\boldsymbol{Z}_l^\top \widehat{\boldsymbol{\beta}}(\tau)} \right] - \mathbb{I}\left[Y_l \geq e^{\boldsymbol{Z}_l^\top \boldsymbol{\beta}_0(\tau)} \right] \right)$$
$$- \sqrt{\rho\, m_v} \left[\boldsymbol{\mu}^*\{\widehat{\boldsymbol{\beta}}(\tau)\} - \boldsymbol{\mu}^*\{\boldsymbol{\beta}_0(\tau)\} \right] - \sqrt{(1-\rho) m_n} \left[\overline{\boldsymbol{\mu}}^*\{\widehat{\boldsymbol{\beta}}(\tau)\} - \overline{\boldsymbol{\mu}}^*\{\boldsymbol{\beta}_0(\tau)\} \right] \Bigg\| \xrightarrow{\Pr} 0. \tag{B.2}$$

$\sqrt{n}\; S_n(\widehat{\boldsymbol{\beta}}, \tau) = o_{(0,\tau_U]}(1)$, a. s. because $\sqrt{n}\; \|\mathbb{S}_L\| \to 0$. This is true because, by the definition of $S_n(\widehat{\boldsymbol{\beta}}, \tau)$,

$$\sup_{\tau \in [\tau_k, \tau_{k+1}]} \sqrt{n}\; \left\| S_n(\widehat{\boldsymbol{\beta}}, \tau) - S_n(\widehat{\boldsymbol{\beta}}, \tau_k) \right\| \leq \sqrt{n}\; \mathscr{C}_2\; \{H(\tau_{k+1}) - H(\tau_k)\}$$
$$\leq \sqrt{n}\; \mathscr{C}_2\; a_n/(1-\tau_U).$$

Given that $\rho\, \boldsymbol{\mu}\{\widehat{\boldsymbol{\beta}}(\tau)\} + (1-\rho)\, \overline{\boldsymbol{\mu}}\{\widehat{\boldsymbol{\beta}}(\tau)\}$ uniformly converges in probability to $\rho\, \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)\} + (1-\rho)\, \overline{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\}$ for $\tau \in (0, \tau_U]$, by (B.1) and (B.2),

$$-\sqrt{n}\; S_n(\boldsymbol{\beta}_0, \tau)$$
$$= \sqrt{\rho\, m_v} \left[\boldsymbol{\mu}\left\{ \widehat{\boldsymbol{\beta}}(\tau) \right\} - \boldsymbol{\mu}\left\{ \boldsymbol{\beta}_0(\tau) \right\} \right] + \sqrt{(1-\rho)\; m_n} \left[\overline{\boldsymbol{\mu}}\left\{ \widehat{\boldsymbol{\beta}}(\tau) \right\} - \overline{\boldsymbol{\mu}}\left\{ \boldsymbol{\beta}_0(\tau) \right\} \right]$$
$$- \int_0^\tau \left(\sqrt{\rho\, m_v} \left[\boldsymbol{\mu}^*\left\{ \widehat{\boldsymbol{\beta}}(u) \right\} - \boldsymbol{\mu}^*\left\{ \boldsymbol{\beta}_0(u) \right\} \right] \right.$$
$$\left. + \sqrt{(1-\rho)\; m_n} \left[\overline{\boldsymbol{\mu}}^*\left\{ \widehat{\boldsymbol{\beta}}(u) \right\} - \overline{\boldsymbol{\mu}}^*\left\{ \boldsymbol{\beta}_0(u) \right\} \right] \right) dH(u) + O_p\left(\frac{\mathbf{1}}{\sqrt{n}} \right) + o_{(0,\tau_U])}(\mathbf{1})$$
$$= \sqrt{\rho\, m_v} \left[\boldsymbol{\mu}\left\{ \widehat{\boldsymbol{\beta}}(\tau) \right\} - \boldsymbol{\mu}\left\{ \boldsymbol{\beta}_0(\tau) \right\} \right] + \sqrt{(1-\rho)\; m_n} \left[\overline{\boldsymbol{\mu}}\left\{ \widehat{\boldsymbol{\beta}}(\tau) \right\} - \overline{\boldsymbol{\mu}}\left\{ \boldsymbol{\beta}_0(\tau) \right\} \right]$$
$$- \int_0^\tau \Bigg\{ \Bigg[\sqrt{\rho\, m_v}\; \boldsymbol{B}^*\left\{ \boldsymbol{\beta}_0(u) \right\} \left(\boldsymbol{B}\left\{ \boldsymbol{\beta}_0(u) \right\} \right)^{-1} + \sqrt{(1-\rho)\; m_n}\; \overline{\boldsymbol{B}}^*\left\{ \boldsymbol{\beta}_0(u) \right\} \left(\overline{\boldsymbol{B}}\left\{ \boldsymbol{\beta}_0(u) \right\} \right)^{-1}$$

$$+ O_p\left(\frac{\mathbf{1}}{\sqrt{n}}\right) + o_{(0,\tau_U])}(\mathbf{1})\Bigg]$$

$$\times\left(\sqrt{\rho\, m_v}\left[\boldsymbol{\mu}\left\{\widehat{\boldsymbol{\beta}}(u)\right\} - \boldsymbol{\mu}\left\{\boldsymbol{\beta}_0(u)\right\}\right] + \sqrt{(1-\rho)\, m_n}\left[\overline{\boldsymbol{\mu}}\left\{\widehat{\boldsymbol{\beta}}(u)\right\} - \overline{\boldsymbol{\mu}}\left\{\boldsymbol{\beta}_0(u)\right\}\right]\right)\Bigg\}dH(u)$$

$$+ O_p\left(\frac{\mathbf{1}}{\sqrt{n}}\right) + o_{(0,\tau_U])}(\mathbf{1}).$$

Here $\sqrt{n}\, S_n(\boldsymbol{\beta}_0,\tau) = 0$ can be viewed as a stochastic differential equation for $\sqrt{\rho\, m_v}\ [\boldsymbol{\mu}\{\widehat{\boldsymbol{\beta}}(\tau)\} - \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)\}] + \sqrt{(1-\rho)\, m_n}\ [\overline{\boldsymbol{\mu}}\{\widehat{\boldsymbol{\beta}}(\tau)\} - \overline{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\}]$, and using the production integration theory [Gill and Johansen, 1990; Andersen et al., 1993], we have

$$\sqrt{\rho\, m_v}\left[\boldsymbol{\mu}\left\{\widehat{\boldsymbol{\beta}}(\tau)\right\} - \boldsymbol{\mu}\left\{\boldsymbol{\beta}_0(\tau)\right\}\right] + \sqrt{(1-\rho)\, m_n}\left[\overline{\boldsymbol{\mu}}\left\{\widehat{\boldsymbol{\beta}}(\tau)\right\} - \overline{\boldsymbol{\mu}}\left\{\boldsymbol{\beta}_0(\tau)\right\}\right]$$
$$= \boldsymbol{\phi}\{-\sqrt{n}\, S_n(\boldsymbol{\beta}_0,\tau)\} + O_p\left(\frac{\mathbf{1}}{\sqrt{n}}\right) + o_{(0,\tau_U])}(\mathbf{1}), \qquad \text{(B.3)}$$

where $\boldsymbol{\phi}$ is a map from $\mathscr{F}$ to $\mathscr{F}$ such that for $\boldsymbol{\gamma} \in \mathscr{F}$,

$$\boldsymbol{\phi}(\boldsymbol{\gamma})(\tau) = \int_0^{\tau} \mathcal{I}(s,\tau)d\boldsymbol{\gamma}(s),$$

with

$$\mathcal{I}(s,t) = \boldsymbol{\pi}_{u\in(s,t]}\Bigg\{\mathbf{I}_p + \Bigg[\sqrt{\rho\, m_v}\ \boldsymbol{B}^*\left\{\boldsymbol{\beta}_0(u)\right\}\left(\boldsymbol{B}\left\{\boldsymbol{\beta}_0(u)\right\}\right)^{-1}$$
$$+ \sqrt{(1-\rho)\, m_n}\ \overline{\boldsymbol{B}}^*\left\{\boldsymbol{\beta}_0(u)\right\}\left(\overline{\boldsymbol{B}}\left\{\boldsymbol{\beta}_0(u)\right\}\right)^{-1}\Bigg]dH(u)\Bigg\}$$

and $\mathscr{F} = \{\boldsymbol{\gamma} : [0,\tau_U] \to \Re^p,\ \boldsymbol{\gamma}$ is left-continuous with right limit, $\boldsymbol{\gamma}(0) = 0\}$.

Consider that

$$\left\{\rho\ \boldsymbol{X}_j\mathbb{N}_j\left(e^{\boldsymbol{X}_j^\top\boldsymbol{\beta}_0(\tau)}\right) + (1-\rho)\ \boldsymbol{Z}_l\mathbb{N}_l\left(e^{\boldsymbol{Z}_l^\top\boldsymbol{\beta}_0(\tau)}\right); \tau \in [0,\tau_U]\right\}$$

is a VC-class [van der Vaart and Wellner, 1996] and

$$\int_0^{\tau}\left(\rho\ \boldsymbol{X}_j\mathbb{I}\left[Y_j \geq e^{\boldsymbol{X}_j^\top\boldsymbol{\beta}_0(u)}\right] + (1-\rho)\ \boldsymbol{Z}_l\mathbb{I}\left[Y_l \geq e^{\boldsymbol{Z}_l^\top\boldsymbol{\beta}_0(u)}\right]\right)dH(u)$$

is Lipschitz in τ, and by the permanence properties of the Donsker class we have that

$$\Big\{\rho\ \boldsymbol{X}_j \mathbb{N}_j\Big(e^{\boldsymbol{X}_j^\top \boldsymbol{\beta}_0(\tau)}\Big) + (1-\rho)\ \boldsymbol{Z}_l \mathbb{N}_l\Big(e^{\boldsymbol{Z}_l^\top \boldsymbol{\beta}_0(\tau)}\Big)$$
$$-\int_0^\tau \Big(\rho\ \boldsymbol{X}_j \mathbb{I}\Big[Y_j \geq e^{\boldsymbol{X}_j^\top \boldsymbol{\beta}_0(u)}\Big] + (1-\rho)\ \boldsymbol{Z}_l \mathbb{I}\Big[Y_l \geq e^{\boldsymbol{Z}_l^\top \boldsymbol{\beta}_0(u)}\Big]\Big) dH(u), \tau \in [\nu, \tau_U]\Big\}$$

is a Donsker class. By the Donsker theorem, $-\sqrt{n}\ S_n(\boldsymbol{\beta}_0, \tau)$ converges weakly to a tight Gaussian process, $\boldsymbol{G}(\tau)$, with mean 0 and covariance $\boldsymbol{\Sigma}(s,t)$ for $\tau \in [0, \tau_U]$, where $\boldsymbol{\Sigma}(s,t) = E\{\boldsymbol{\iota}_{\widehat{j}}(s)\boldsymbol{\iota}_{\widehat{j}}(t)^\top\} + E\{\boldsymbol{\iota}_{\widehat{l}}(s)\boldsymbol{\iota}_{\widehat{l}}(t)^\top\}$ with

$$\boldsymbol{\iota}_{\widehat{j}}(\tau) = \rho\ \boldsymbol{X}_{\widehat{j}} \mathbb{N}_{\widehat{j}}\Big(e^{\boldsymbol{X}_{\widehat{j}}^\top \boldsymbol{\beta}_0(\tau)}\Big) - \int_0^\tau \rho\ \boldsymbol{X}_{\widehat{j}} \mathbb{I}\Big[Y_{\widehat{j}} \geq e^{\boldsymbol{X}_{\widehat{j}}^\top \boldsymbol{\beta}_0(u)}\Big] dH(u)$$

and

$$\boldsymbol{\iota}_{\widehat{l}}(\tau) = (1-\rho)\ \boldsymbol{Z}_{\widehat{l}} \mathbb{N}_{\widehat{l}}\Big(e^{\boldsymbol{Z}_{\widehat{l}}^\top \boldsymbol{\beta}_0(\tau)}\Big) - \int_0^\tau (1-\rho)\ \boldsymbol{Z}_{\widehat{l}} \mathbb{I}\Big[Y_{\widehat{l}} \geq e^{\boldsymbol{Z}_{\widehat{l}}^\top \boldsymbol{\beta}_0(u)}\Big] dH(u).$$

$\boldsymbol{\phi}\{\boldsymbol{G}(\tau)\}$ for $\tau \in (0, \tau_U]$ is also Gaussian process because $\boldsymbol{\phi}$ is a linear operator [Römisch, 2005]. $\rho\ (\boldsymbol{B}\{\boldsymbol{\beta}_0(\tau)\})^{-1} + (1-\rho)\ (\overline{\boldsymbol{B}}\{\boldsymbol{\beta}_0(\tau)\})^{-1}$ is bounded uniformly for $\tau \in [\nu, \tau_U]$ (**C6**). Applying the Taylor expansion to $\boldsymbol{\eta}[\rho\ \boldsymbol{\mu}\{\widehat{\boldsymbol{\beta}}(\tau)\} + (1-\rho)\ \overline{\boldsymbol{\mu}}\{\widehat{\boldsymbol{\beta}}(\tau)\}] - \boldsymbol{\eta}[\rho\ \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)\} + (1-\rho)\ \overline{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\}]$ and by the continuous mapping theorem, we have, for $\tau \in [\nu, \tau_U]$, $\sqrt{n}\{\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\}$ converges weakly to $[\rho\ (\boldsymbol{B}\{\boldsymbol{\beta}_0(\tau)\})^{-1} + (1-\rho)\ (\overline{\boldsymbol{B}}\{\boldsymbol{\beta}_0(\tau)\})^{-1}]\boldsymbol{\phi}\{\boldsymbol{G}(\tau)\}$, which is Gaussian.

Appendix C

Proof of Theorem 3.2.1

Define,

- $D_{m_v}^{\mathbb{V}}(\widehat{\boldsymbol{\beta}}, \tau_k) = \dfrac{1}{m_v} \sum_{j \in \mathbb{V}} \left\{ \boldsymbol{X}_j \mathbb{M}_j \left\{ \tau_k, \boldsymbol{X}_j, \widehat{\boldsymbol{\beta}}(\tau_k) \right\} + \varrho \left[\boldsymbol{X}_j \mathbb{W}_{qM\overline{\mathbb{V}}} \left(\mathbb{M}_l \left\{ \tau_k, \boldsymbol{Z}_l, \widehat{\boldsymbol{\beta}}(\tau_k) \right\} \right) - \right.\right.$
  $\left.\left. \mathbb{W}_{qM\overline{\mathbb{V}}} \left(\boldsymbol{Z}_l \ \mathbb{M}_l \left\{ \tau_k, \boldsymbol{Z}_l, \widehat{\boldsymbol{\beta}}(\tau_k) \right\} \right) \right] \right\}.$

- $\boldsymbol{\mu}(\boldsymbol{b}) = E \left[\boldsymbol{X} \mathbb{N} \left(e^{\boldsymbol{X}^\top \boldsymbol{b}} \right) \right]$, $\boldsymbol{B}(\boldsymbol{b}) = E \left[\boldsymbol{X}^{\otimes 2} \ \widetilde{f} \left(e^{\boldsymbol{X}^\top \boldsymbol{b}} \,\middle|\, \boldsymbol{X} \right) \ \ e^{\boldsymbol{X}^\top \boldsymbol{b}} \right]$,
  $\boldsymbol{\Gamma}_{m_v}(\boldsymbol{b}) = \dfrac{1}{m_v} \sum_{j \in \mathbb{V}} \boldsymbol{X}_j \mathbb{N}_j \left(e^{\boldsymbol{X}_j^\top \boldsymbol{b}} \right) - \boldsymbol{\mu}(\boldsymbol{b}).$

- $\widetilde{\boldsymbol{\mu}}(\boldsymbol{b}) = E \left[\boldsymbol{X} \ \mathbb{W}_{qM\overline{\mathbb{V}}} \left\{ \mathbb{N} \left(e^{\boldsymbol{Z}^\top \boldsymbol{b}} \right) \right\} - \mathbb{W}_{qM\overline{\mathbb{V}}} \left\{ \boldsymbol{Z} \mathbb{N} \left(e^{\boldsymbol{Z}^\top \boldsymbol{b}} \right) \right\} \right]$,
  $\widetilde{\boldsymbol{B}}(\boldsymbol{b}) = E \left[\boldsymbol{X}^{\otimes 2} \ \mathbb{W}_{qM\overline{\mathbb{V}}} \left\{ \widetilde{f} \left(e^{\boldsymbol{Z}^\top \boldsymbol{b}} \,\middle|\, \boldsymbol{Z} \right) \ e^{\boldsymbol{Z}^\top \boldsymbol{b}} \right\} - \mathbb{W}_{qM\overline{\mathbb{V}}} \left\{ \boldsymbol{Z}^{\otimes 2} \ \ \widetilde{f} \left(e^{\boldsymbol{Z}^\top \boldsymbol{b}} \,\middle|\, \boldsymbol{Z} \right) \ e^{\boldsymbol{Z}^\top \boldsymbol{b}} \right\} \right]$,
  $\widetilde{\boldsymbol{\Gamma}}_{m_v}(\boldsymbol{b}) = \dfrac{1}{m_v} \sum_{j \in \mathbb{V}} \left[\boldsymbol{X}_j \mathbb{W}_{qM\overline{\mathbb{V}}} \left\{ \mathbb{N} \left(e^{\boldsymbol{Z}^\top \boldsymbol{b}} \right) \right\} - \mathbb{W}_{qM\overline{\mathbb{V}}} \left\{ \boldsymbol{Z} \mathbb{N} \left(e^{\boldsymbol{Z}^\top \boldsymbol{b}} \right) \right\} \right] - \widetilde{\boldsymbol{\mu}}(\boldsymbol{b}).$

- $\overline{\boldsymbol{\mu}}(\boldsymbol{b}) = E \left[\boldsymbol{Z} \mathbb{N} \left(e^{\boldsymbol{Z}^\top \boldsymbol{b}} \right) \right]$, $\overline{\boldsymbol{B}}(\boldsymbol{b}) = E \left[\boldsymbol{Z}^{\otimes 2} \ \widetilde{f} \left(e^{\boldsymbol{Z}^\top \boldsymbol{b}} \,\middle|\, \boldsymbol{Z} \right) \ \ e^{\boldsymbol{Z}^\top \boldsymbol{b}} \right]$,
  $\overline{\boldsymbol{\Gamma}}_{m_n}(\boldsymbol{b}) = \dfrac{1}{m_n} \sum_{l \in \overline{\mathbb{V}}} \boldsymbol{Z}_l \mathbb{N}_l \left(e^{\boldsymbol{Z}_l^\top \boldsymbol{b}} \right) - \overline{\boldsymbol{\mu}}(\boldsymbol{b}).$

- $\boldsymbol{\mu}^*(\boldsymbol{b}) = E \left[\boldsymbol{X} \mathbb{I} \left(Y \geq e^{\boldsymbol{X}^\top \boldsymbol{b}} \right) \right]$, $\boldsymbol{B}^*(\boldsymbol{b}) = E \left[\boldsymbol{X}^{\otimes 2} \ \overline{f} \left(e^{\boldsymbol{X}^\top \boldsymbol{b}} \,\middle|\, \boldsymbol{X} \right) \ \ e^{\boldsymbol{X}^\top \boldsymbol{b}} \right]$,
  $\boldsymbol{\Gamma}_{m_v}^*(\boldsymbol{b}) = \dfrac{1}{m_v} \sum_{j \in \mathbb{V}} \boldsymbol{X}_j \mathbb{I} \left[Y_j \geq e^{\boldsymbol{X}_j^\top \boldsymbol{b}} \right] - \boldsymbol{\mu}^*(\boldsymbol{b}).$

- $\widetilde{\boldsymbol{\mu}}^*(\boldsymbol{b}) = E \left[\boldsymbol{X} \ \mathbb{W}_{qM\overline{\mathbb{V}}} \left\{ \mathbb{I} \left[Y \geq e^{\boldsymbol{Z}^\top \boldsymbol{b}} \right] \right\} - \mathbb{W}_{qM\overline{\mathbb{V}}} \left\{ \boldsymbol{Z} \mathbb{I} \left[Y \geq e^{\boldsymbol{Z}^\top \boldsymbol{b}} \right] \right\} \right]$,
  $\widetilde{\boldsymbol{B}}^*(\boldsymbol{b}) = E \left[\boldsymbol{X}^{\otimes 2} \ \mathbb{W}_{qM\overline{\mathbb{V}}} \left\{ \overline{f} \left(e^{\boldsymbol{Z}^\top \boldsymbol{b}} \,\middle|\, \boldsymbol{Z} \right) \ e^{\boldsymbol{Z}^\top \boldsymbol{b}} \right\} - \mathbb{W}_{qM\overline{\mathbb{V}}} \left\{ \boldsymbol{Z}^{\otimes 2} \ \ \overline{f} \left(e^{\boldsymbol{Z}^\top \boldsymbol{b}} \,\middle|\, \boldsymbol{Z} \right) \ e^{\boldsymbol{Z}^\top \boldsymbol{b}} \right\} \right]$,

$$\widetilde{\boldsymbol{\Gamma}}^*_{m_v}(\boldsymbol{b}) = \frac{1}{m_v}\sum_{j\in\mathbb{V}}\left[\boldsymbol{X}_j\mathbb{W}_{qM\mathbb{V}}\left\{\mathbb{I}\left[Y\geq e^{\boldsymbol{Z}^\top\boldsymbol{b}}\right]\right\} - \mathbb{W}_{qM\mathbb{V}}\left\{\boldsymbol{Z}\mathbb{I}\left[Y\geq e^{\boldsymbol{Z}^\top\boldsymbol{b}}\right]\right\}\right] - \widetilde{\boldsymbol{\mu}}^*(\boldsymbol{b}).$$

- $\overline{\boldsymbol{\mu}}^*(\boldsymbol{b}) = E\left[\boldsymbol{Z}\mathbb{I}\left(Y\geq e^{\boldsymbol{Z}^\top\boldsymbol{b}}\right)\right]$, $\overline{\boldsymbol{B}}^*(\boldsymbol{b}) = E\left[\boldsymbol{Z}^{\otimes 2}\ \overline{f}\left(e^{\boldsymbol{Z}^\top\boldsymbol{b}}\,\middle|\,\boldsymbol{Z}\right)\quad e^{\boldsymbol{Z}^\top\boldsymbol{b}}\right]$,
  $\overline{\boldsymbol{\Gamma}}^*_{m_n}(\boldsymbol{b}) = \dfrac{1}{m_n}\displaystyle\sum_{l\in\overline{\mathbb{V}}}\boldsymbol{Z}_l\mathbb{I}\left[Y_l\geq e^{\boldsymbol{Z}_l^\top\boldsymbol{b}}\right] - \overline{\boldsymbol{\mu}}^*(\boldsymbol{b})$.

Assume that $\tau_1 < \cdots < \tau_{L-1}$ are equally spaced between 0 and τ_U. Let $a_n = \left\|\mathbb{S}_L\right\|$ and $b_n = a_n/(1-\tau_U)$; then $L = \tau_U/a_n$. It is clear that $0 < H(\tau_k) - H(\tau_{k-1}) \leq b_n$ for $k = 1, 2, \ldots, L$.

For $d > 0$, define $\mathscr{B}(d) = \{\boldsymbol{b}\in\Re^p : \inf_{\tau\in(0,\tau_U]}\left\|\rho\left[\boldsymbol{\mu}(\boldsymbol{b}) - \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)\}\right] + \rho\ \varrho[\widetilde{\boldsymbol{\mu}}(\boldsymbol{b}) - \widetilde{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\}] + (1-\rho)\left[\overline{\boldsymbol{\mu}}(\boldsymbol{b}) - \overline{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\}\right]\right\| \leq d\}$. Let $\boldsymbol{\alpha}_0(\tau) = \rho\ \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)\} + \rho\ \varrho\ \widetilde{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\} + (1-\rho)\ \overline{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\}$, $\widehat{\boldsymbol{\alpha}}(\tau) = \rho\ \boldsymbol{\mu}\{\widehat{\boldsymbol{\beta}}(\tau)\} + \rho\ \varrho\ \widetilde{\boldsymbol{\mu}}\{\widehat{\boldsymbol{\beta}}(\tau)\} + (1-\rho)\ \overline{\boldsymbol{\mu}}\{\widehat{\boldsymbol{\beta}}(\tau)\}$ and $\mathscr{A}(d) = \rho\ \boldsymbol{\mu}(\boldsymbol{b}) + \rho\ \varrho\ \widetilde{\boldsymbol{\mu}}(\boldsymbol{b}) + (1-\rho)\ \overline{\boldsymbol{\mu}}(\boldsymbol{b}) : \boldsymbol{b}\in\mathscr{B}(d)\}$.

Let $\boldsymbol{b}$ and $\boldsymbol{b}^{'}\in\mathscr{B}(d_0)$ such that $\rho\ \boldsymbol{\mu}(\boldsymbol{b}) + \rho\ \varrho\ \widetilde{\boldsymbol{\mu}}(\boldsymbol{b}) + (1-\rho)\ \overline{\boldsymbol{\mu}}(\boldsymbol{b}) = \rho\ \boldsymbol{\mu}(\boldsymbol{b}^{'}) + \rho\ \varrho\ \widetilde{\boldsymbol{\mu}}(\boldsymbol{b}^{'}) + (1-\rho)\ \overline{\boldsymbol{\mu}}(\boldsymbol{b}^{'})$, then

$$\begin{aligned}
0 &= (\boldsymbol{b}-\boldsymbol{b}^{'})^\top\{\rho\ \boldsymbol{\mu}(\boldsymbol{b}) + \rho\ \varrho\ \widetilde{\boldsymbol{\mu}}(\boldsymbol{b}) + (1-\rho)\ \overline{\boldsymbol{\mu}}(\boldsymbol{b}) - \rho\ \boldsymbol{\mu}(\boldsymbol{b}^{'}) - \rho\ \varrho\ \widetilde{\boldsymbol{\mu}}(\boldsymbol{b}^{'}) - (1-\rho)\ \overline{\boldsymbol{\mu}}(\boldsymbol{b}^{'})\}\\
&= \rho\ E\left\{\left(\boldsymbol{X}^\top\boldsymbol{b} - \boldsymbol{X}^\top\boldsymbol{b}^{'}\right)\left[\widetilde{F}\left(e^{\boldsymbol{X}^\top\boldsymbol{b}}\,\middle|\,\boldsymbol{X}\right) - \widetilde{F}\left(e^{\boldsymbol{X}^\top\boldsymbol{b}^{'}}\,\middle|\,\boldsymbol{X}\right)\right]\right\}\\
&\quad + \rho\ \varrho\ E\left\{\left(\boldsymbol{X}^\top\boldsymbol{b} - \boldsymbol{X}^\top\boldsymbol{b}^{'}\right)\left[\mathbb{W}_{qM\mathbb{V}}\left\{\widetilde{F}\left(e^{\boldsymbol{Z}^\top\boldsymbol{b}}\,\middle|\,\boldsymbol{Z}\right)\right\} - \mathbb{W}_{qM\mathbb{V}}\left\{\widetilde{F}\left(e^{\boldsymbol{Z}^\top\boldsymbol{b}^{'}}\,\middle|\,\boldsymbol{Z}\right)\right\}\right]\right.\\
&\quad - \left[\mathbb{W}_{qM\mathbb{V}}\left\{\left(\boldsymbol{Z}^\top\boldsymbol{b} - \boldsymbol{Z}^\top\boldsymbol{b}^{'}\right)\widetilde{F}\left(e^{\boldsymbol{Z}^\top\boldsymbol{b}}\,\middle|\,\boldsymbol{Z}\right)\right\}\right.\\
&\qquad\qquad\left.\left. - \mathbb{W}_{qM\mathbb{V}}\left\{\left(\boldsymbol{Z}^\top\boldsymbol{b} - \boldsymbol{Z}^\top\boldsymbol{b}^{'}\right)\widetilde{F}\left(e^{\boldsymbol{Z}^\top\boldsymbol{b}^{'}}\,\middle|\,\boldsymbol{Z}\right)\right\}\right]\right\}\\
&\quad + (1-\rho)\ E\left\{\left(\boldsymbol{Z}^\top\boldsymbol{b} - \boldsymbol{Z}^\top\boldsymbol{b}^{'}\right)\left[\widetilde{F}\left(e^{\boldsymbol{Z}^\top\boldsymbol{b}}\,\middle|\,\boldsymbol{Z}\right) - \widetilde{F}\left(e^{\boldsymbol{Z}^\top\boldsymbol{b}^{'}}\,\middle|\,\boldsymbol{Z}\right)\right]\right\}.
\end{aligned}$$

By condition **R3(a)**, the above equation holds if and only if $\boldsymbol{X}^\top\boldsymbol{b} = \boldsymbol{X}^\top\boldsymbol{b}^{'}$ and $\boldsymbol{Z}^\top\boldsymbol{b} = \boldsymbol{Z}^\top\boldsymbol{b}^{'}$ with probability 1.

By the positive definiteness of $E(\boldsymbol{X}^{\otimes 2})$ and $E(\boldsymbol{Z}^{\otimes 2})$ it is clear that $\boldsymbol{b} = \boldsymbol{b}^{'}$. So there exists an inverse function $\boldsymbol{\eta}$, from $\mathscr{A}(d_0)$ to $\mathscr{B}(d_0)$ such that $\boldsymbol{\eta}\{\rho\ \boldsymbol{\mu}(\boldsymbol{b}) + \rho\ \varrho\ \widetilde{\boldsymbol{\mu}}(\boldsymbol{b}) + (1-\rho)\ \overline{\boldsymbol{\mu}}(\boldsymbol{b})\} = \boldsymbol{b}$ for any $\boldsymbol{b}\in\mathscr{B}(d_0)$. Now we conclude that under the condition **R3**, $\rho\ \boldsymbol{\mu} + \rho\ \varrho\ \widetilde{\boldsymbol{\mu}} + (1-\rho)\ \overline{\boldsymbol{\mu}}$ is also a one to one mapping from $\mathscr{B}(d_0)$ to $\mathscr{A}(d_0)$.

According to our estimating procedure, $\rho\ D^{\mathbb{V}}_{m_v}(\widehat{\boldsymbol{\beta}}, \tau_k) + (1-\rho)\ \Omega^{\overline{\mathbb{V}}}_{m_n}(\widehat{\boldsymbol{\beta}}, \tau_k) + O_p\left(\dfrac{\mathbf{1}}{\sqrt{n}}\right) +$

$o_{(0,\tau_U])}(\mathbf{1}) = 0$, which implies

$$
\begin{aligned}
&\frac{\rho}{m_v}\sum_{j\in\mathbb{V}}\boldsymbol{X}_j\mathbb{N}_j\left(e^{\boldsymbol{X}_j^\top\widehat{\boldsymbol{\beta}}(\tau_k)}\right)+\frac{1-\rho}{m_n}\sum_{l\in\overline{\mathbb{V}}}\boldsymbol{Z}_l\ \mathbb{N}_l\left(e^{\boldsymbol{Z}_l^\top\widehat{\boldsymbol{\beta}}(\tau_k)}\right)\\
&\qquad+\frac{\rho}{m_v}\ \varrho\sum_{j\in\mathbb{V}}\left[\boldsymbol{X}_j\mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\mathbb{N}\left(e^{\boldsymbol{Z}^\top\widehat{\boldsymbol{\beta}}(\tau_k)}\right)\right\}-\mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\boldsymbol{Z}\ \mathbb{N}\left(e^{\boldsymbol{Z}^\top\widehat{\boldsymbol{\beta}}(\tau_k)}\right)\right\}\right]\\
&=\frac{\rho}{m_v}\sum_{j\in\mathbb{V}}\int_0^{\tau_k}\boldsymbol{X}_j\mathbb{I}\left[Y_j\geq e^{\boldsymbol{X}_j^\top\widehat{\boldsymbol{\beta}}(u)}\right]dH(u)+\frac{1-\rho}{m_n}\sum_{l\in\overline{\mathbb{V}}}\int_0^{\tau_k}\boldsymbol{Z}_l\mathbb{I}\left[Y_l\geq e^{\boldsymbol{Z}_l^\top\widehat{\boldsymbol{\beta}}(u)}\right]dH(u)\\
&\qquad+\frac{\rho}{m_v}\ \varrho\sum_{j\in\mathbb{V}}\int_0^{\tau_k}\left[\boldsymbol{X}_j\mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\mathbb{I}\left[Y\geq e^{\boldsymbol{Z}^\top\widehat{\boldsymbol{\beta}}(u)}\right]\right\}\right.\\
&\qquad\qquad\left.-\mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\boldsymbol{Z}\ \mathbb{I}\left[Y\geq e^{\boldsymbol{Z}^\top\widehat{\boldsymbol{\beta}}(u)}\right]\right\}\right]dH(u)+O_p\left(\frac{\mathbf{1}}{\sqrt{n}}\right)+o_{(0,\tau_U])}(\mathbf{1}).
\end{aligned}
$$

Simple algebra leads to

$$
\begin{aligned}
&\frac{\rho}{m_v}\sum_{j\in\mathbb{V}}\boldsymbol{X}_j\mathbb{N}_j\left(e^{\boldsymbol{X}_j^\top\widehat{\boldsymbol{\beta}}(\tau_k)}\right)+\frac{1-\rho}{m_n}\sum_{l\in\overline{\mathbb{V}}}\boldsymbol{Z}_l\ \mathbb{N}_l\left(e^{\boldsymbol{Z}_l^\top\widehat{\boldsymbol{\beta}}(\tau_k)}\right)\\
&\qquad+\frac{\rho}{m_v}\ \varrho\sum_{j\in\mathbb{V}}\left[\boldsymbol{X}_j\mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\mathbb{N}\left(e^{\boldsymbol{Z}^\top\widehat{\boldsymbol{\beta}}(\tau_k)}\right)\right\}-\mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\boldsymbol{Z}\ \mathbb{N}\left(e^{\boldsymbol{Z}^\top\widehat{\boldsymbol{\beta}}(\tau_k)}\right)\right\}\right]\\
&\qquad-\frac{\rho}{m_v}\ E\left[\boldsymbol{X}\mathbb{N}\left(e^{\boldsymbol{X}^\top\boldsymbol{\beta}_0(\tau_k)}\right)\right]-\frac{1-\rho}{m_n}E\left[\boldsymbol{Z}\mathbb{N}\left(e^{\boldsymbol{Z}^\top\boldsymbol{\beta}_0(\tau_k)}\right)\right]\\
&\qquad-\frac{\rho}{m_v}\ \varrho\ E\left[\boldsymbol{X}\ \mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\mathbb{N}\left(e^{\boldsymbol{Z}^\top\boldsymbol{\beta}_0(\tau_k)}\right)\right\}-\mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\boldsymbol{Z}\mathbb{N}\left(e^{\boldsymbol{Z}^\top\boldsymbol{\beta}_0(\tau_k)}\right)\right\}\right]\\
&=\rho\left[\boldsymbol{\Gamma}_{m_v}\left\{\widehat{\boldsymbol{\beta}}(\tau_k)\right\}+\boldsymbol{\mu}\left\{\widehat{\boldsymbol{\beta}}(\tau_k)\right\}-\boldsymbol{\mu}\left\{\boldsymbol{\beta}_0(\tau_k)\right\}\right]\\
&\qquad+\rho\ \varrho\left[\widetilde{\boldsymbol{\Gamma}}_{m_v}\left\{\widehat{\boldsymbol{\beta}}(\tau_k)\right\}+\widetilde{\boldsymbol{\mu}}\left\{\widehat{\boldsymbol{\beta}}(\tau_k)\right\}-\widetilde{\boldsymbol{\mu}}\left\{\boldsymbol{\beta}_0(\tau_k)\right\}\right]\\
&\qquad+(1-\rho)\left[\overline{\boldsymbol{\Gamma}}_{m_n}\left\{\widehat{\boldsymbol{\beta}}(\tau_k)\right\}+\overline{\boldsymbol{\mu}}\left\{\widehat{\boldsymbol{\beta}}(\tau_k)\right\}-\overline{\boldsymbol{\mu}}\left\{\boldsymbol{\beta}_0(\tau_k)\right\}\right]
\end{aligned}
$$

and

$$
\begin{aligned}
&\frac{\rho}{m_v}\sum_{j\in\mathbb{V}}\int_0^{\tau_k}\boldsymbol{X}_j\mathbb{I}\left[Y_j\geq e^{\boldsymbol{X}_j^\top\widehat{\boldsymbol{\beta}}(u)}\right]dH(u)+\frac{1-\rho}{m_n}\sum_{l\in\overline{\mathbb{V}}}\int_0^{\tau_k}\boldsymbol{Z}_l\mathbb{I}\left[Y_l\geq e^{\boldsymbol{Z}_l^\top\widehat{\boldsymbol{\beta}}(u)}\right]dH(u)\\
&\qquad+\frac{\rho}{m_v}\ \varrho\sum_{j\in\mathbb{V}}\int_0^{\tau_k}\left[\boldsymbol{X}_j\mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\mathbb{I}\left[Y\geq e^{\boldsymbol{Z}^\top\widehat{\boldsymbol{\beta}}(u)}\right]\right\}\right.\\
&\qquad\qquad\left.-\mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\boldsymbol{Z}\ \mathbb{I}\left[Y\geq e^{\boldsymbol{Z}^\top\widehat{\boldsymbol{\beta}}(u)}\right]\right\}\right]dH(u)
\end{aligned}
$$

$$
\begin{aligned}
&- \frac{\rho}{m_v}\, E\left[\boldsymbol{X} \int_0^{\tau_k} \mathbb{I}\left[Y \geq e^{\boldsymbol{X}^\top \boldsymbol{\beta}_0(u)}\right] dH(u)\right] \\
&- \frac{1-\rho}{m_n}\, E\left[\boldsymbol{Z} \int_0^{\tau_k} \mathbb{I}\left[Y \geq e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(u)}\right] dH(u)\right] \\
&- \frac{\rho}{m_v}\, \varrho\, E\Bigg[\int_0^{\tau_k} \Bigg(\boldsymbol{X}\ \mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\mathbb{I}\left[Y \geq e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(u)}\right]\right\} \\
&\qquad - \mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\boldsymbol{Z}\ \mathbb{I}\left[Y \geq e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(u)}\right]\right\}\Bigg) dH(u)\Bigg] \\
= \rho &\left\{\int_0^{\tau_k} \boldsymbol{\Gamma}^*_{m_v}\left\{\widehat{\boldsymbol{\beta}}(u)\right\} dH(u) + \int_0^{\tau_k} \left[\boldsymbol{\mu}^*\left\{\widehat{\boldsymbol{\beta}}(u)\right\} - \boldsymbol{\mu}^*\left\{\boldsymbol{\beta}_0(u)\right\}\right] dH(u)\right\} \\
&+ \rho\ \varrho \left\{\int_0^{\tau_k} \widetilde{\boldsymbol{\Gamma}}^*_{m_v}\left\{\widehat{\boldsymbol{\beta}}(u)\right\} dH(u) + \int_0^{\tau_k} \left[\widetilde{\boldsymbol{\mu}}^*\left\{\widehat{\boldsymbol{\beta}}(u)\right\} - \widetilde{\boldsymbol{\mu}}^*\left\{\boldsymbol{\beta}_0(u)\right\}\right] dH(u)\right\} \\
&+ (1-\rho)\left\{\int_0^{\tau_k} \overline{\boldsymbol{\Gamma}}^*_{m_n}\left\{\widehat{\boldsymbol{\beta}}(u)\right\} dH(u) + \int_0^{\tau_k} \left[\overline{\boldsymbol{\mu}}^*\left\{\widehat{\boldsymbol{\beta}}(u)\right\} - \overline{\boldsymbol{\mu}}^*\left\{\boldsymbol{\beta}_0(u)\right\}\right] dH(u)\right\}.
\end{aligned}
$$

By martingale property,

$$
\begin{aligned}
E\left[\boldsymbol{X}\mathbb{N}\left(e^{\boldsymbol{X}^\top \boldsymbol{\beta}_0(\tau_k)}\right)\right] &= E\left[\boldsymbol{X} \int_0^{\tau_k} \mathbb{I}\left[Y \geq e^{\boldsymbol{X}^\top \boldsymbol{\beta}_0(u)}\right] dH(u)\right], \\
E\left[\boldsymbol{Z}\mathbb{N}\left(e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau_k)}\right)\right] &= E\left[\boldsymbol{Z} \int_0^{\tau_k} \mathbb{I}\left[Y \geq e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(u)}\right] dH(u)\right]
\end{aligned}
$$

and by the estimating equation property provided in Peng and Huang [2008, Sec. 2] and martingale property,

$$
\begin{aligned}
E&\left[\boldsymbol{X}\ \mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\mathbb{N}\left(e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau_k)}\right)\right\} - \mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\boldsymbol{Z}\mathbb{N}\left(e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau_k)}\right)\right\}\right] \\
&= E\Bigg[\int_0^{\tau_k} \Bigg(\boldsymbol{X}\ \mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\mathbb{I}\left[Y \geq e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(u)}\right]\right\} \\
&\qquad - \mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\boldsymbol{Z}\ \mathbb{I}\left[Y \geq e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(u)}\right]\right\}\Bigg) dH(u)\Bigg].
\end{aligned}
$$

Then combining previous two equations,

$$
\begin{aligned}
\rho&\left[\boldsymbol{\mu}\left\{\widehat{\boldsymbol{\beta}}(\tau_k)\right\} - \boldsymbol{\mu}\left\{\boldsymbol{\beta}_0(\tau_k)\right\}\right] + \rho\ \varrho\left[\widetilde{\boldsymbol{\mu}}\left\{\widehat{\boldsymbol{\beta}}(\tau_k)\right\} - \widetilde{\boldsymbol{\mu}}\left\{\boldsymbol{\beta}_0(\tau_k)\right\}\right] \\
&\qquad + (1-\rho)\left[\overline{\boldsymbol{\mu}}\left\{\widehat{\boldsymbol{\beta}}(\tau_k)\right\} - \overline{\boldsymbol{\mu}}\left\{\boldsymbol{\beta}_0(\tau_k)\right\}\right] \\
&= \rho\Bigg[- \boldsymbol{\Gamma}_{m_v}\left\{\widehat{\boldsymbol{\beta}}(\tau_k)\right\} + \int_0^{\tau_k} \boldsymbol{\Gamma}^*_{m_v}\left\{\widehat{\boldsymbol{\beta}}(u)\right\} dH(u)
\end{aligned}
$$

$$
\begin{aligned}
&\quad + \sum_{r=1}^{k} \int_{\tau_{r-1}}^{\tau_r} \left[\boldsymbol{\mu}^* \left\{\widehat{\boldsymbol{\beta}}(u)\right\} - \boldsymbol{\mu}^* \left\{\boldsymbol{\beta}_0(u)\right\}\right] dH(u)\Bigg] \\
&+ \rho\ \varrho \Bigg[-\widetilde{\boldsymbol{\Gamma}}_{m_v} \left\{\widehat{\boldsymbol{\beta}}(\tau_k)\right\} + \int_0^{\tau_k} \widetilde{\boldsymbol{\Gamma}}^*_{m_v} \left\{\widehat{\boldsymbol{\beta}}(u)\right\} dH(u) \\
&\quad + \sum_{r=1}^{k} \int_{\tau_{r-1}}^{\tau_r} \left[\widetilde{\boldsymbol{\mu}}^* \left\{\widehat{\boldsymbol{\beta}}(u)\right\} - \widetilde{\boldsymbol{\mu}}^* \left\{\boldsymbol{\beta}_0(u)\right\}\right] dH(u)\Bigg] \\
&+ (1-\rho) \Bigg[- \widetilde{\boldsymbol{\Gamma}}_{m_n} \left\{\widehat{\boldsymbol{\beta}}(\tau_k)\right\} + \int_0^{\tau_k} \widetilde{\boldsymbol{\Gamma}}^*_{m_n} \left\{\widehat{\boldsymbol{\beta}}(u)\right\} dH(u) \\
&\quad + \sum_{r=1}^{k} \int_{\tau_{r-1}}^{\tau_r} \left[\widetilde{\boldsymbol{\mu}}^* \left\{\widehat{\boldsymbol{\beta}}(u)\right\} - \widetilde{\boldsymbol{\mu}}^* \left\{\boldsymbol{\beta}_0(u)\right\}\right] dH(u)\Bigg] + O_p\left(\frac{\mathbf{1}}{\sqrt{n}}\right) + o_{(0,\tau_U])}(\mathbf{1}) \\
= \rho &\left[-\boldsymbol{\Gamma}_{m_v} \left\{\widehat{\boldsymbol{\beta}}(\tau_k)\right\} + \int_0^{\tau_k} \boldsymbol{\Gamma}^*_{m_v} \left\{\widehat{\boldsymbol{\beta}}(u)\right\} dH(u)\right] \\
&+ \rho\ \varrho \left[-\widetilde{\boldsymbol{\Gamma}}_{m_v} \left\{\widehat{\boldsymbol{\beta}}(\tau_k)\right\} + \int_0^{\tau_k} \widetilde{\boldsymbol{\Gamma}}^*_{m_v} \left\{\boldsymbol{\beta}_0(u)\right\} dH(u)\right] \\
&+ (1-\rho) \left[-\overline{\boldsymbol{\Gamma}}_{m_n} \left\{\widehat{\boldsymbol{\beta}}(\tau_k)\right\} + \int_0^{\tau_k} \overline{\boldsymbol{\Gamma}}^*_{m_n} \left\{\widehat{\boldsymbol{\beta}}(u)\right\} dH(u)\right] \\
&+ \sum_{r=1}^{k} \int_{\tau_{r-1}}^{\tau_r} \Bigg(\rho \left[\boldsymbol{\mu}^* \left\{\widehat{\boldsymbol{\beta}}(u)\right\} - \boldsymbol{\mu}^* \left\{\boldsymbol{\beta}_0(u)\right\}\right] + \rho\ \varrho \left[\widetilde{\boldsymbol{\mu}}^* \left\{\widehat{\boldsymbol{\beta}}(u)\right\} - \widetilde{\boldsymbol{\mu}}^* \left\{\boldsymbol{\beta}_0(u)\right\}\right] \\
&+ (1-\rho) \left[\overline{\boldsymbol{\mu}}^* \left\{\widehat{\boldsymbol{\beta}}(u)\right\} - \overline{\boldsymbol{\mu}}^* \left\{\boldsymbol{\beta}_0(u)\right\}\right] \Bigg) dH(u) + O_p\left(\frac{\mathbf{1}}{\sqrt{n}}\right) + o_{(0,\tau_U])}(\mathbf{1}).
\end{aligned}
\tag{C.1}
$$

Consider

$$
\begin{aligned}
G_1 &= \left\{\boldsymbol{X}_j \mathbb{I}\left[Y_j \le e^{\boldsymbol{X}_j^\top \boldsymbol{b}}\right] \delta_j : \boldsymbol{b} \in \Re^p\right\}, \quad G_2 = \left\{\boldsymbol{X}_j \mathbb{I}\left[Y_j \ge e^{\boldsymbol{X}_j^\top \boldsymbol{b}}\right] : \boldsymbol{b} \in \Re^p\right\}, \\
\overline{G}_1 &= \left\{\boldsymbol{Z}_l \mathbb{I}\left[Y_l \le e^{\boldsymbol{Z}_l^\top \boldsymbol{b}}\right] \delta_l : \boldsymbol{b} \in \Re^p\right\}, \quad \overline{G}_2 = \left\{\boldsymbol{Z}_l \mathbb{I}\left[Y_l \ge e^{\boldsymbol{Z}_l^\top \boldsymbol{b}}\right] : \boldsymbol{b} \in \Re^p\right\}, \\
\widetilde{G}_1 &= \left\{\left(\boldsymbol{X}_j \mathbb{W}_{qM\mathbb{V}}\left\{\mathbb{I}\left[Y \le e^{\boldsymbol{Z}^\top \boldsymbol{b}}\right]\right\} - \mathbb{W}_{qM\mathbb{V}}\left\{\boldsymbol{Z}\ \mathbb{I}\left[Y \le e^{\boldsymbol{Z}^\top \boldsymbol{b}}\right]\right\}\right) \delta_j : \boldsymbol{b} \in \Re^p\right\}, \\
\widetilde{G}_2 &= \left\{\boldsymbol{X}_j\ \mathbb{W}_{qM\mathbb{V}}\left\{\mathbb{I}\left[Y \le e^{\boldsymbol{Z}^\top \boldsymbol{b}}\right]\right\} - \mathbb{W}_{qM\mathbb{V}}\left\{\boldsymbol{Z}\ \mathbb{I}\left[Y \le e^{\boldsymbol{Z}^\top \boldsymbol{b}}\right]\right\} : \boldsymbol{b} \in \Re^p\right\}.
\end{aligned}
$$

Since the class of indicator functions of polytopes in $\Re^p$ is Glivenko Cantelli and $\boldsymbol{X}_j$ and $\boldsymbol{Z}_l$ are bounded, so here all the G_1, G_2, $\widetilde{G}_1$, $\widetilde{G}_2$, $\overline{G}_1$ and $\overline{G}_2$ are Glivenko Cantelli [van der Vaart and Wellner, 1996]. So $\sup_{\boldsymbol{b}\in\Re^p} \|\boldsymbol{\Gamma}_{m_v}(\boldsymbol{b})\| \xrightarrow{a.s} 0$, $\sup_{\boldsymbol{b}\in\Re^p} \|\boldsymbol{\Gamma}^*_{m_v}(\boldsymbol{b})\| \xrightarrow{a.s} 0$, $\sup_{\boldsymbol{b}\in\Re^p} \|\widetilde{\boldsymbol{\Gamma}}_{m_v}(\boldsymbol{b})\| \xrightarrow{a.s} 0$, $\sup_{\boldsymbol{b}\in\Re^p} \|\widetilde{\boldsymbol{\Gamma}}^*_{m_v}(\boldsymbol{b})\| \xrightarrow{a.s} 0$, $\sup_{\boldsymbol{b}\in\Re^p} \|\overline{\boldsymbol{\Gamma}}_{m_n}(\boldsymbol{b})\| \xrightarrow{a.s} 0$ and $\sup_{\boldsymbol{b}\in\Re^p} \|\overline{\boldsymbol{\Gamma}}^*_{m_n}(\boldsymbol{b})\| \xrightarrow{a.s}$

0 (Glivenko Cantelli theorem). Then, for any given C_1, $\widetilde{C}_1$ and $\overline{C}_1$ (> 0) and for sufficiently large m_v and n, $\sup_k \left\| -\mathbf{\Gamma}_{m_v}\{\widehat{\boldsymbol{\beta}}(\tau_k)\} + \int_0^{\tau_k} \mathbf{\Gamma}^*_{m_v}\{\widehat{\boldsymbol{\beta}}(u)\}dH(u) \right\| < C_1$, $\sup_k \left\| -\widetilde{\mathbf{\Gamma}}_{m_v}\{\widehat{\boldsymbol{\beta}}(\tau_k)\} + \int_0^{\tau_k} \widetilde{\mathbf{\Gamma}}^*_{m_v}\{\widehat{\boldsymbol{\beta}}(u)\}dH(u) \right\| < \widetilde{C}_1$ and $\sup_k \left\| -\overline{\mathbf{\Gamma}}_{m_n}\{\widehat{\boldsymbol{\beta}}(\tau_k)\} + \int_0^{\tau_k} \overline{\mathbf{\Gamma}}^*_{m_n}\{\widehat{\boldsymbol{\beta}}(u)\}dH(u) \right\| < \overline{C}_1$ with probability 1.

There exists $C_2 > 0$ such that $\sup_i \|\boldsymbol{X}_i\| < C_2$ and $\sup_i \|\boldsymbol{Z}_i\| < C_2$ (by **R1**).

For some constants C_3, $\widetilde{C}_3$ and $\overline{C}_3$ (> 0), $\|\boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau) - \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau')\}\| \leq C_3|\tau - \tau'|$, $\|\widetilde{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau) - \widetilde{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau')\}\| \leq \widetilde{C_3}|\tau - \tau'|$ and $\|\overline{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau) - \overline{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau')\}\| \leq \overline{C_3}|\tau - \tau'|$ (by **R2(a)**) for any τ, $\tau' \in (0, \tau_U]$.

There exists C_4, $\widetilde{C}_4$ and $\overline{C}_4$ (> 0) such that $\|\{B(\boldsymbol{b})\}^{-1}B^*(\boldsymbol{b})\mathbf{y}\| \leq C_4\|\mathbf{y}\|$; $\boldsymbol{b} \in \mathscr{B}(d_0)$, $\|\{\widetilde{B}(\boldsymbol{b})\}^{-1}\widetilde{B}^*(\boldsymbol{b})\mathbf{y}\| \leq \widetilde{C}_4\|\mathbf{y}\|$; $\boldsymbol{b} \in \mathscr{B}(d_0)$ and $\|\{\overline{B}(\boldsymbol{b})\}^{-1}\overline{B}^*(\boldsymbol{b})\mathbf{y}\| \leq \overline{C}_4\|\mathbf{y}\|$; $\boldsymbol{b} \in \mathscr{B}(d_0)$ for any $\mathbf{y} \in \Re^p$ (by **R3(c)**).

Define $\mathscr{C}_1 = \rho\ C_1 + \rho\ \varrho\ \widetilde{C}_1 + (1-\rho)\ \overline{C}_1$, $\mathscr{C}_2 = C_2$, $\mathscr{C}_3 = \rho\ C_3 + \rho\ \varrho\ \widetilde{C}_3 + (1-\rho)\ \overline{C}_3$ and $\mathscr{C}_4 = \rho\ C_4 + \rho\ \varrho\ \widetilde{C}_4 + (1-\rho)\ \overline{C}_4$.

For given n, define a sequence $\{\mathcal{E}_u\}_{u=0}^{L-1}$, where $\mathcal{E}_0 = \mathscr{C}_3 a_n$, $\mathcal{E}_1 = \mathscr{C}_1 + \mathscr{C}_2(1/n) + \mathscr{C}_3 a_n + \mathcal{E}_0 \mathscr{C}_4 b_n$ and $\mathcal{E}_u = \mathscr{C}_1 + \mathscr{C}_2(1/n) + \mathscr{C}_3 a_n + \left(\sum_{r=0}^{u-1} \mathcal{E}_r\right) \mathscr{C}_4 b_n$ for $u = 2, 3, \ldots, L-1$. By the definition of $\mathcal{E}_u$, $\mathcal{E}_u - \mathcal{E}_{u-1} = \mathcal{E}_{u-1}\mathscr{C}_4 b_n$, hence $\mathcal{E}_u = (1 + \mathscr{C}_4 b_n)^{u-1}\mathcal{E}_1$.

Given that $\lim_{n\to\infty} a_n = 0$ and $L = \tau_U/a_n$, implies that $\lim_{n\to\infty}(1+\mathscr{C}_4 b_n)^{L-1} = \exp\{\mathscr{C}_4\tau_U/(1-\tau_U)\}$. Since $\mathcal{E}_u$ is increasing with u, and for some N_0 such that $n \geq N_0$, we can choose sufficiently small $\mathscr{C}_1$ so that $\mathcal{E}_u \leq 2\exp\{\tau_U/(1-\tau_U)\}\mathscr{C}_1 \leq d_0$ for all $u = 0, 1, \ldots, L-1$.

Next we prove that $\sup_{\tau_u \leq \tau < \tau_{u+1}} \Big\| \rho[\boldsymbol{\mu}\{\widehat{\boldsymbol{\beta}}(\tau)\} - \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)\}] + \rho\ \varrho[\widetilde{\boldsymbol{\mu}}\{\widehat{\boldsymbol{\beta}}(\tau)\} - \widetilde{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\}] + (1-\rho)[\overline{\boldsymbol{\mu}}\{\widehat{\boldsymbol{\beta}}(\tau)\} - \overline{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\}] \Big\| < \mathcal{E}_u$, $u = 0, 1, \ldots, L-1$.

Considering $n \geq N_0$, and by the definition of $\widehat{\boldsymbol{\beta}}(\tau)$,

$$
\begin{aligned}
\sup_{\tau_0 \leq \tau < \tau_1} \Big\| \rho \left[\boldsymbol{\mu}\{\widehat{\boldsymbol{\beta}}(\tau)\} - \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)\}\right] &+ \rho\ \varrho \left[\widetilde{\boldsymbol{\mu}}\{\widehat{\boldsymbol{\beta}}(\tau)\} - \widetilde{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\}\right] \\
&+ (1-\rho) \left[\overline{\boldsymbol{\mu}}\{\widehat{\boldsymbol{\beta}}(\tau)\} - \overline{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\}\right] \Big\| \\
= \sup_{\tau_0 \leq \tau < \tau_1} \|\rho\ \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)\} &+ \rho\ \varrho\ \widetilde{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\} + (1-\rho)\ \overline{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\}\| \leq \mathscr{C}_3 a_n = \mathcal{E}_0.
\end{aligned}
$$

Considering (C.1) with $k = 1$, for $\tau \in [\tau_0, \tau_1)$,

$$
\Big\| \rho \left[\boldsymbol{\mu}^* \left\{\widehat{\boldsymbol{\beta}}(\tau_k)\right\} - \boldsymbol{\mu}^* \left\{\boldsymbol{\beta}_0(\tau_k)\right\}\right] + \rho\ \varrho \left[\widetilde{\boldsymbol{\mu}}^* \left\{\widehat{\boldsymbol{\beta}}(\tau_k)\right\} - \widetilde{\boldsymbol{\mu}}^* \left\{\boldsymbol{\beta}_0(\tau_k)\right\}\right]
$$

$$+ (1-\rho)\left[\widetilde{\boldsymbol{\mu}}^*\left\{\widehat{\boldsymbol{\beta}}(\tau_k)\right\} - \widetilde{\boldsymbol{\mu}}^*\left\{\boldsymbol{\beta}_0(\tau_k)\right\}\right]\Bigg\|$$

$$= \Bigg\|\rho\Big[\boldsymbol{\mu}^*\left(\boldsymbol{\eta}\left\{\widehat{\boldsymbol{\alpha}}(\tau)\right\}\right) - \boldsymbol{\mu}^*\left(\boldsymbol{\eta}\left\{\boldsymbol{\alpha}_0(\tau)\right\}\right)\Big] + \rho\ \varrho\Big[\widetilde{\boldsymbol{\mu}}^*\left(\boldsymbol{\eta}\left\{\widehat{\boldsymbol{\alpha}}(\tau)\right\}\right) - \widetilde{\boldsymbol{\mu}}^*\left(\boldsymbol{\eta}\left\{\boldsymbol{\alpha}_0(\tau)\right\}\right)\Big]$$

$$+ (1-\rho)\Big[\overline{\boldsymbol{\mu}}^*\left(\boldsymbol{\eta}\left\{\widehat{\boldsymbol{\alpha}}(\tau)\right\}\right) - \overline{\boldsymbol{\mu}}^*\left(\boldsymbol{\eta}\left\{\boldsymbol{\alpha}_0(\tau)\right\}\right)\Big]\Bigg\|$$

$$= \Bigg\|\rho\left(\boldsymbol{B}\left[\boldsymbol{\eta}\left\{\breve{\boldsymbol{\alpha}}(\tau)\right\}\right]\right)^{-1}\boldsymbol{B}^*\left[\boldsymbol{\eta}\left\{\breve{\boldsymbol{\alpha}}(\tau)\right\}\right]\left\{\widehat{\boldsymbol{\alpha}}(\tau) - \boldsymbol{\alpha}_0(\tau)\right\}$$

$$+ \rho\ \varrho\left(\widetilde{\boldsymbol{B}}\left[\boldsymbol{\eta}\left\{\breve{\boldsymbol{\alpha}}(\tau)\right\}\right]\right)^{-1}\widetilde{\boldsymbol{B}}^*\left[\boldsymbol{\eta}\left\{\breve{\boldsymbol{\alpha}}(\tau)\right\}\right]\left\{\widehat{\boldsymbol{\alpha}}(\tau) - \boldsymbol{\alpha}_0(\tau)\right\}$$

$$+ (1-\rho)\left(\overline{\boldsymbol{B}}\left[\boldsymbol{\eta}\left\{\breve{\boldsymbol{\alpha}}(\tau)\right\}\right]\right)^{-1}\overline{\boldsymbol{B}}^*\left[\boldsymbol{\eta}\left\{\breve{\boldsymbol{\alpha}}(\tau)\right\}\right]\left\{\widehat{\boldsymbol{\alpha}}(\tau) - \boldsymbol{\alpha}_0(\tau)\right\}\Bigg\|$$

$$\leq \mathscr{C}_4\mathcal{E}_0,$$

where $\breve{\boldsymbol{\alpha}}(\tau)$ is between $\widehat{\boldsymbol{\alpha}}(\tau)$ and $\boldsymbol{\alpha}_0(\tau)$. So using conditions defined earlier, the norm of the right hand side of (C.1) is not bigger than $\mathscr{C}_1 + \mathcal{E}_0\mathscr{C}_4 b_n + \mathscr{C}_2(1/n)$; So,

$$\sup_{\tau_1\leq\tau<\tau_2}\Bigg\|\rho\left[\boldsymbol{\mu}\{\widehat{\boldsymbol{\beta}}(\tau)\} - \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)\}\right] + \rho\ \varrho\left[\widetilde{\boldsymbol{\mu}}\{\widehat{\boldsymbol{\beta}}(\tau)\} - \widetilde{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\}\right]$$

$$+ (1-\rho)\left[\overline{\boldsymbol{\mu}}\{\widehat{\boldsymbol{\beta}}(\tau)\} - \overline{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\}\right]\Bigg\|$$

$$\leq \Bigg\|\rho\left[\boldsymbol{\mu}\{\widehat{\boldsymbol{\beta}}(\tau_1)\} - \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau_1)\}\right] + \rho\ \varrho\left[\widetilde{\boldsymbol{\mu}}\{\widehat{\boldsymbol{\beta}}(\tau_1)\} - \widetilde{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau_1)\}\right]$$

$$+ (1-\rho)\left[\overline{\boldsymbol{\mu}}\{\widehat{\boldsymbol{\beta}}(\tau_1)\} - \overline{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau_1)\}\right]\Bigg\|$$

$$+ \sup_{\tau_1\leq\tau<\tau_2}\Bigg\|\rho\left[\boldsymbol{\mu}\{\widehat{\boldsymbol{\beta}}(\tau_1)\} - \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)\}\right] + \rho\ \varrho\left[\widetilde{\boldsymbol{\mu}}\{\widehat{\boldsymbol{\beta}}(\tau_1)\} - \widetilde{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\}\right]$$

$$+ (1-\rho)\left[\overline{\boldsymbol{\mu}}\{\widehat{\boldsymbol{\beta}}(\tau_1)\} - \overline{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\}\right]\Bigg\|$$

$$\leq \mathscr{C}_1 + \mathscr{C}_2\frac{1}{n} + \mathscr{C}_3 a_n + \mathcal{E}_0\mathscr{C}_4 b_n = \mathcal{E}_1.$$

Using similar approach, we can arrive $\widehat{\boldsymbol{\beta}}(\tau_u) \in \mathscr{B}(d_0)$ and

$$\sup_{\tau_u\leq\tau<\tau_{u+1}}\Bigg\|\rho\left[\boldsymbol{\mu}\{\widehat{\boldsymbol{\beta}}(\tau)\} - \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)\}\right] + \rho\ \varrho\left[\widetilde{\boldsymbol{\mu}}\{\widehat{\boldsymbol{\beta}}(\tau)\} - \widetilde{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\}\right]$$

$$+ (1-\rho)\left[\overline{\boldsymbol{\mu}}\{\widehat{\boldsymbol{\beta}}(\tau)\} - \overline{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\}\right]\Bigg\| \leq \mathcal{E}_u,$$

for all $u = 2, 3, \ldots, L-1$. As n increases, $a_n \to 0$ and $\mathscr{C}_1$ can become arbitrarily small, which implies that

$$\sup_{0<\tau<\tau_U}\Bigg\|\rho\left[\boldsymbol{\mu}\{\widehat{\boldsymbol{\beta}}(\tau)\} - \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)\}\right] + \rho\ \varrho\left[\widetilde{\boldsymbol{\mu}}\{\widehat{\boldsymbol{\beta}}(\tau)\} - \widetilde{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\}\right]$$

$$+ (1-\rho)\left[\overline{\boldsymbol{\mu}}\{\widehat{\boldsymbol{\beta}}(\tau)\} - \overline{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\}\right]\Big\| \xrightarrow{\text{Pr}} 0.$$

Using Taylor series expansion of $\boldsymbol{\eta}\{\widehat{\boldsymbol{\alpha}}(\tau)\}$ at $\boldsymbol{\alpha}_0(\tau)$ for $\tau \in [\nu, \tau_U]$, from condition **R4**, we arrive that

$$\begin{aligned}
\left\|\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\right\| \leq{}& \Big\|\rho\left(\boldsymbol{B}\left\{\boldsymbol{\beta}_0(\tau)\right\}\right)^{-1}\left\{\widehat{\boldsymbol{\alpha}}(\tau) - \boldsymbol{\alpha}_0(\tau)\right\} \\
&+ \rho\ \varrho\left(\widetilde{\boldsymbol{B}}\left\{\boldsymbol{\beta}_0(\tau)\right\}\right)^{-1}\left\{\widehat{\boldsymbol{\alpha}}(\tau) - \boldsymbol{\alpha}_0(\tau)\right\} \\
&+ (1-\rho)\left(\overline{\boldsymbol{B}}\left\{\boldsymbol{\beta}_0(\tau)\right\}\right)^{-1}\left\{\widehat{\boldsymbol{\alpha}}(\tau) - \boldsymbol{\alpha}_0(\tau)\right\}\Big\| + \left\|\epsilon_n^*(\tau)\right\| \\
\leq{}& \mathscr{C}_6\left\|\widehat{\boldsymbol{\alpha}}(\tau) - \boldsymbol{\alpha}_0(\tau)\right\| + \left\|\epsilon_n^*(\tau)\right\|,
\end{aligned}$$

where $\mathscr{C}_6(>0)$ is independent of τ and $\sup_{\nu\leq\tau\leq\tau_U}\|\epsilon_n^*(\tau)\| \xrightarrow{\text{Pr}} 0$. Hence the consistency proof.

Appendix D

Proof of Theorem 3.2.2

Lemma D.1. *For any sequence,* $\left\{\widetilde{\boldsymbol{\beta}}_n(\tau),\ \tau \in (0, \tau_U]\right\}_{n=1}^{\infty}$*, we have*

$$
\begin{aligned}
\sup_{\tau \in (0,\tau_U]} \Bigg\| & \frac{\sqrt{\rho}}{\sqrt{m_v}} \sum_{j \in \mathbb{V}} \boldsymbol{X}_j \left[\mathbb{N}_j \left(e^{\boldsymbol{X}_j^{\top} \widetilde{\boldsymbol{\beta}}_n(\tau)} \right) - \mathbb{N}_j \left(e^{\boldsymbol{X}_j^{\top} \boldsymbol{\beta}_0(\tau)} \right) \right] \\
& \qquad - \sqrt{\rho\ m_v} \left[\boldsymbol{\mu} \left\{ \widetilde{\boldsymbol{\beta}}_n(\tau) \right\} - \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)\} \right] \\
& + \frac{\sqrt{\rho}}{\sqrt{m_v}}\ \varrho \sum_{j \in \mathbb{V}} \left[\boldsymbol{X}_j \mathbb{W}_{qM\overline{\mathbb{V}}} \left\{ \mathbb{N} \left(e^{\boldsymbol{Z}^{\top} \widetilde{\boldsymbol{\beta}}_n(\tau)} \right) \right\} - \mathbb{W}_{qM\overline{\mathbb{V}}} \left\{ \boldsymbol{Z}\ \mathbb{N} \left(e^{\boldsymbol{Z}^{\top} \widetilde{\boldsymbol{\beta}}_n(\tau)} \right) \right\} \right. \\
& \qquad \left. - \boldsymbol{X}_j \mathbb{W}_{qM\overline{\mathbb{V}}} \left\{ \mathbb{N} \left(e^{\boldsymbol{Z}^{\top} \boldsymbol{\beta}_0(\tau)} \right) \right\} + \mathbb{W}_{qM\overline{\mathbb{V}}} \left\{ \boldsymbol{Z}\ \mathbb{N} \left(e^{\boldsymbol{Z}^{\top} \boldsymbol{\beta}_0(\tau)} \right) \right\} \right] \\
& \qquad - \sqrt{\rho\ m_v}\ \varrho \left[\widetilde{\boldsymbol{\mu}} \left\{ \widetilde{\boldsymbol{\beta}}_n(\tau) \right\} - \widetilde{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\} \right] \\
& + \frac{\sqrt{(1-\rho)}}{\sqrt{m_n}} \sum_{l \in \overline{\mathbb{V}}} \boldsymbol{Z}_l \left[\mathbb{N}_l \left(e^{\boldsymbol{Z}_l^{\top} \widetilde{\boldsymbol{\beta}}_n(\tau)} \right) - \mathbb{N}_l \left(e^{\boldsymbol{Z}_l^{\top} \boldsymbol{\beta}_0(\tau)} \right) \right] \\
& \qquad - \sqrt{(1-\rho)\ m_n} \left[\overline{\boldsymbol{\mu}} \left\{ \widetilde{\boldsymbol{\beta}}_n(\tau) \right\} + \overline{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\} \right] \Bigg\| \xrightarrow{\Pr} 0,
\end{aligned}
$$

if

$$
\begin{aligned}
\sup_{\tau \in (0,\tau_U]} \Bigg\| & \rho \left[\boldsymbol{\mu}\{\widetilde{\boldsymbol{\beta}}_n(\tau)\} - \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)\} \right] + \rho\ \varrho[\widetilde{\boldsymbol{\mu}}\{\widetilde{\boldsymbol{\beta}}_n(\tau)\} - \widetilde{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\}] \\
& \qquad + (1-\rho) \left[\overline{\boldsymbol{\mu}}\{\widetilde{\boldsymbol{\beta}}_n(\tau)\} - \overline{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\} \right] \Bigg\| \xrightarrow{\Pr} 0.
\end{aligned}
$$

***Proof of Lemma* D.1**: Define $\mu_1(\boldsymbol{b}) = E\left[\mathbb{N}\left(e^{\boldsymbol{X}^{\top}\boldsymbol{b}}\right)\right]$, $\overline{\mu}_1(\boldsymbol{b}) = E\left[\mathbb{N}\left(e^{\boldsymbol{Z}^{\top}\boldsymbol{b}}\right)\right]$,

$$\widetilde{\mu}_1(\boldsymbol{b}) = E\Big[\mathbb{W}_{qM\overline{\mathbb{V}}}\Big\{\mathbb{N}\Big(e^{\boldsymbol{Z}^\top \boldsymbol{b}}\Big)\Big\}\Big] \text{ and}$$

$$\begin{aligned}
\sigma_d^2(\boldsymbol{b}) = \operatorname{Var}\Bigg\{ & \sqrt{\rho}\left[\mathbb{N}\left(e^{\boldsymbol{X}^\top \boldsymbol{b}}\right) - \mathbb{N}\left(e^{\boldsymbol{X}^\top \boldsymbol{\beta}_0(\tau)}\right) - \mu_1(\boldsymbol{b}) + \mu_1\{\boldsymbol{\beta}_0(\tau)\}\right] \\
& + \sqrt{\rho}\ \varrho\Big[\mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\mathbb{N}\left(e^{\boldsymbol{Z}^\top \boldsymbol{b}}\right)\right\} - \widetilde{\mu}_1(\boldsymbol{b}) \\
& \qquad - \mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\mathbb{N}\left(e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau)}\right)\right\} + \widetilde{\mu}_1\{\boldsymbol{\beta}_0(\tau)\}\Big] \\
& + \sqrt{(1-\rho)}\ \left[\mathbb{N}\left(e^{\boldsymbol{Z}^\top \boldsymbol{b}}\right) - \mathbb{N}\left(e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau)}\right) - \overline{\mu}_1(\boldsymbol{b}) + \overline{\mu}_1\{\boldsymbol{\beta}_0(\tau)\}\right]\Bigg\}.
\end{aligned}$$

Provided $\boldsymbol{X}$ and $\boldsymbol{Z}$ are bounded, it suffices to prove that $\sigma_d^2\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} \xrightarrow{\text{Pr}} 0$ [Alexander, 1984; Lai and Ying, 1988]. For a given $\widetilde{\boldsymbol{\beta}}_n(\tau)$,

$$\begin{aligned}
& \sigma_d^2\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} \\
& = \operatorname{Var}\Bigg\{\sqrt{\rho}\left[\mathbb{N}\left(e^{\boldsymbol{X}^\top \widetilde{\boldsymbol{\beta}}_n(\tau)}\right) - \mathbb{N}\left(e^{\boldsymbol{X}^\top \boldsymbol{\beta}_0(\tau)}\right) - \mu_1\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} + \mu_1\{\boldsymbol{\beta}_0(\tau)\}\right] \\
& \quad + \sqrt{\rho}\ \varrho\Big[\mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\mathbb{N}\left(e^{\boldsymbol{Z}^\top \widetilde{\boldsymbol{\beta}}_n(\tau)}\right)\right\} - \widetilde{\mu}_1\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} \\
& \qquad - \mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\mathbb{N}\left(e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau)}\right)\right\} + \widetilde{\mu}_1\{\boldsymbol{\beta}_0(\tau)\}\Big] \\
& \quad + \sqrt{(1-\rho)}\ \left[\mathbb{N}\left(e^{\boldsymbol{Z}^\top \widetilde{\boldsymbol{\beta}}_n(\tau)}\right) - \mathbb{N}\left(e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau)}\right) - \overline{\mu}_1\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} + \overline{\mu}_1\{\boldsymbol{\beta}_0(\tau)\}\right]\Bigg\} \\
& = \rho\ \operatorname{Var}\left[\mathbb{N}\left(e^{\boldsymbol{X}^\top \widetilde{\boldsymbol{\beta}}_n(\tau)}\right) - \mathbb{N}\left(e^{\boldsymbol{X}^\top \boldsymbol{\beta}_0(\tau)}\right) - \mu_1\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} + \mu_1\{\boldsymbol{\beta}_0(\tau)\}\right] \\
& \quad + \rho\ \varrho^2\ \operatorname{Var}\Big[\mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\mathbb{N}\left(e^{\boldsymbol{Z}^\top \widetilde{\boldsymbol{\beta}}_n(\tau)}\right)\right\} - \widetilde{\mu}_1\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} \\
& \qquad - \mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\mathbb{N}\left(e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau)}\right)\right\} + \widetilde{\mu}_1\{\boldsymbol{\beta}_0(\tau)\}\Big] \\
& \quad + (1-\rho)\ \operatorname{Var}\left[\mathbb{N}\left(e^{\boldsymbol{Z}^\top \widetilde{\boldsymbol{\beta}}_n(\tau)}\right) - \mathbb{N}\left(e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau)}\right) - \overline{\mu}_1\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} + \overline{\mu}_1\{\boldsymbol{\beta}_0(\tau)\}\right] \\
& \quad + 2\sqrt{\rho}\ \varrho\ \operatorname{Cov}\Big[\mathbb{N}\left(e^{\boldsymbol{X}^\top \widetilde{\boldsymbol{\beta}}_n(\tau)}\right) - \mathbb{N}\left(e^{\boldsymbol{X}^\top \boldsymbol{\beta}_0(\tau)}\right) - \mu_1\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} + \mu_1\{\boldsymbol{\beta}_0(\tau)\}, \\
& \qquad \mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\mathbb{N}\left(e^{\boldsymbol{Z}^\top \widetilde{\boldsymbol{\beta}}_n(\tau)}\right)\right\} - \widetilde{\mu}_1\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} \\
& \qquad - \mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\mathbb{N}\left(e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau)}\right)\right\} + \widetilde{\mu}_1\{\boldsymbol{\beta}_0(\tau)\}\Big]
\end{aligned}$$

$$
\begin{aligned}
&+ 2\sqrt{(1-\rho)}\ \varrho\ \mathrm{Cov}\Big[\mathbb{N}\left(e^{\boldsymbol{Z}^\top \widetilde{\boldsymbol{\beta}}_n(\tau)}\right) - \mathbb{N}\left(e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau)}\right) - \overline{\mu}_1\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} + \overline{\mu}_1\{\boldsymbol{\beta}_0(\tau)\},\\
&\qquad \mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\mathbb{N}\left(e^{\boldsymbol{Z}^\top \widetilde{\boldsymbol{\beta}}_n(\tau)}\right)\right\} - \widetilde{\mu}_1\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\}\\
&\qquad\qquad - \mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\mathbb{N}\left(e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau)}\right)\right\} + \widetilde{\mu}_1\{\boldsymbol{\beta}_0(\tau)\}\Big]\\
&+ 2\sqrt{\rho(1-\rho)}\ \mathrm{Cov}\Big[\mathbb{N}\left(e^{\boldsymbol{Z}^\top \widetilde{\boldsymbol{\beta}}_n(\tau)}\right) - \mathbb{N}\left(e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau)}\right) - \overline{\mu}_1\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} + \overline{\mu}_1\{\boldsymbol{\beta}_0(\tau)\},\\
&\qquad \mathbb{N}\left(e^{\boldsymbol{X}^\top \widetilde{\boldsymbol{\beta}}_n(\tau)}\right) - \mathbb{N}\left(e^{\boldsymbol{X}^\top \boldsymbol{\beta}_0(\tau)}\right) - \mu_1\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} + \mu_1\{\boldsymbol{\beta}_0(\tau)\}\Big]\\
&\le \rho\ \mathrm{Var}\left[\mathbb{N}\left(e^{\boldsymbol{X}^\top \widetilde{\boldsymbol{\beta}}_n(\tau)}\right) - \mathbb{N}\left(e^{\boldsymbol{X}^\top \boldsymbol{\beta}_0(\tau)}\right) - \mu_1\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} + \mu_1\{\boldsymbol{\beta}_0(\tau)\}\right]\\
&+ \rho\ \varrho^2\ \mathrm{Var}\Big[\mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\mathbb{N}\left(e^{\boldsymbol{Z}^\top \widetilde{\boldsymbol{\beta}}_n(\tau)}\right)\right\} - \widetilde{\mu}_1\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\}\\
&\qquad\qquad - \mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\mathbb{N}\left(e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau)}\right)\right\} + \widetilde{\mu}_1\{\boldsymbol{\beta}_0(\tau)\}\Big]\\
&+ (1-\rho)\ \mathrm{Var}\left[\mathbb{N}\left(e^{\boldsymbol{Z}^\top \widetilde{\boldsymbol{\beta}}_n(\tau)}\right) - \mathbb{N}\left(e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau)}\right) - \overline{\mu}_1\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} + \overline{\mu}_1\{\boldsymbol{\beta}_0(\tau)\}\right]\\
&+ 2\sqrt{\rho}\ \varrho\ \sqrt{\mathrm{Var}\left[\mathbb{N}\left(e^{\boldsymbol{X}^\top \widetilde{\boldsymbol{\beta}}_n(\tau)}\right) - \mathbb{N}\left(e^{\boldsymbol{X}^\top \boldsymbol{\beta}_0(\tau)}\right) - \mu_1\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} + \mu_1\{\boldsymbol{\beta}_0(\tau)\}\right]}\\
&\quad \times \sqrt{\mathrm{Var}\Big[\mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\mathbb{N}\left(e^{\boldsymbol{Z}^\top \widetilde{\boldsymbol{\beta}}_n(\tau)}\right)\right\} - \widetilde{\mu}_1\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} - \mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\mathbb{N}\left(e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau)}\right)\right\} + \widetilde{\mu}_1\{\boldsymbol{\beta}_0(\tau)\}\Big]}\\
&+ 2\sqrt{(1-\rho)}\ \varrho\ \sqrt{\mathrm{Var}\left[\mathbb{N}\left(e^{\boldsymbol{Z}^\top \widetilde{\boldsymbol{\beta}}_n(\tau)}\right) - \mathbb{N}\left(e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau)}\right) - \overline{\mu}_1\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} + \overline{\mu}_1\{\boldsymbol{\beta}_0(\tau)\}\right]}\\
&\quad \times \sqrt{\mathrm{Var}\Big[\mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\mathbb{N}\left(e^{\boldsymbol{Z}^\top \widetilde{\boldsymbol{\beta}}_n(\tau)}\right)\right\} - \widetilde{\mu}_1\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} - \mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\mathbb{N}\left(e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau)}\right)\right\} + \widetilde{\mu}_1\{\boldsymbol{\beta}_0(\tau)\}\Big]}\\
&+ 2\sqrt{\rho(1-\rho)}\ \sqrt{\mathrm{Var}\left[\mathbb{N}\left(e^{\boldsymbol{Z}^\top \widetilde{\boldsymbol{\beta}}_n(\tau)}\right) - \mathbb{N}\left(e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau)}\right) - \overline{\mu}_1\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} + \overline{\mu}_1\{\boldsymbol{\beta}_0(\tau)\}\right]}\\
&\quad \times \sqrt{\mathrm{Var}\left[\mathbb{N}\left(e^{\boldsymbol{X}^\top \widetilde{\boldsymbol{\beta}}_n(\tau)}\right) - \mathbb{N}\left(e^{\boldsymbol{X}^\top \boldsymbol{\beta}_0(\tau)}\right) - \mu_1\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} + \mu_1\{\boldsymbol{\beta}_0(\tau)\}\right]}\\
&= \rho\ \sigma^2_{1d}\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} + \rho\ \varrho^2\ \sigma^2_{2d}\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} + (1-\rho)\ \sigma^2_{3d}\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\}
\end{aligned}
$$

$$+ 2\varrho \sqrt{\rho \ \sigma^2_{1d}\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\}\sigma^2_{2d}\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\}} + 2\varrho \sqrt{(1-\rho) \ \sigma^2_{3d}\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\}\sigma^2_{2d}\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\}}$$
$$+ 2\sqrt{\rho(1-\rho) \ \sigma^2_{3d}\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\}\sigma^2_{1d}\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\}}.$$

Following the arguments in Peng and Huang [2008], we can show that $\sigma^2_{1d}\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} \xrightarrow{\text{Pr}} 0$ and $\sigma^2_{3d}\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} \xrightarrow{\text{Pr}} 0$. To prove that $\sigma^2_{2d}\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} \xrightarrow{\text{Pr}} 0$, we will use the similar arguments provided in Peng and Huang [2008].

Since $\widetilde{\mu}_1\{\boldsymbol{\beta}_0(0)\} = 0$ and $\widetilde{\mu}_1\{\boldsymbol{\beta}_0(\tau)\}$ is Lipschitz-continuous in τ, for any $\vartheta > 0$, we can find some ν_ϑ such that $\sup_{\tau\in(0,\nu_\vartheta)} \|\widetilde{\mu}_1\{\boldsymbol{\beta}_0(\tau)\}\| \le \vartheta/8$. Because $\sup_{\tau\in(0,\tau_U]} \|\widetilde{\mu}_1\{\widetilde{\boldsymbol{\beta}}_n(\tau)\} - \widetilde{\mu}_1\{\boldsymbol{\beta}_0(\tau)\}\| \xrightarrow{\text{Pr}} 0$, for any $\zeta > 0$, there exists $N_{\vartheta,\zeta,1} > 0$ such that for $n \ge N_{\vartheta,\zeta,1}$,

$$\Pr\left(\sup_{\tau\in(0,\tau_U]}\left\|\widetilde{\mu}_1\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} - \widetilde{\mu}_1\{\boldsymbol{\beta}_0(\tau)\}\right\| > \vartheta/8\right) < \zeta/3.$$

Consider the case where $\sup_{\tau\in(0,\tau_U]} \|\widetilde{\mu}_1\{\widetilde{\boldsymbol{\beta}}_n(\tau)\} - \widetilde{\mu}_1\{\boldsymbol{\beta}_0(\tau)\}\| < \vartheta/8$. First, we have $\sup_{\tau\in(0,\nu_\vartheta)} \|\widetilde{\mu}_1\{\widetilde{\boldsymbol{\beta}}_n(\tau)\}\| \le \vartheta/4$. Note that, for a given $\widetilde{\boldsymbol{\beta}}_n(\tau)$,

$$\begin{aligned}
\sigma^2_{2d}\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} &\le E\left(\mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\mathbb{I}\left[Y \le e^{\boldsymbol{Z}^\top\widetilde{\boldsymbol{\beta}}_n(\tau)}\right]\right\} - \mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\mathbb{I}\left[Y \le e^{\boldsymbol{Z}^\top\boldsymbol{\beta}_0(\tau)}\right]\right\}\right)^2 \\
&\le E\left(\mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\mathbb{I}\left[Y \le e^{\boldsymbol{Z}^\top\widetilde{\boldsymbol{\beta}}_n(\tau)}\right]\right\} + \mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\mathbb{I}\left[Y \le e^{\boldsymbol{Z}^\top\boldsymbol{\beta}_0(\tau)}\right]\right\}\right) \\
&= \widetilde{\mu}_1\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} + \widetilde{\mu}_1\left\{\boldsymbol{\beta}_0(\tau)\right\};
\end{aligned}$$

therefore $\sigma^2_{2d}\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} \le \vartheta/2$.

For any $\nu \in (0, \tau_U)$, there exists a $N_{\vartheta,\zeta,2}$ such that for $n \ge N_{\vartheta,\zeta,2}$ and given that $\sup_{\tau\in[\nu,\tau_U]} \|\widetilde{\boldsymbol{\beta}}_n(\tau) - \boldsymbol{\beta}_0(\tau)\| \xrightarrow{\text{Pr}} 0$,

$$\Pr\left(\sup_{\tau\in[\nu_\vartheta,\tau_U]}\left\|\widetilde{\boldsymbol{\beta}}_n(\tau) - \boldsymbol{\beta}_0(\tau)\right\| > \vartheta^*\right) < \zeta/3,$$

where ϑ^* satisfies

$$\sup_{\tau\in(0,\tau_U],\ \mathbf{X}\in\mathbb{X}} e^{\mathbf{X}^\top\boldsymbol{\beta}_0(\tau)}\ e^{(p\ \mathscr{C}_2\ \vartheta^*)}(p\ \mathscr{C}_2\ \vartheta^*)\mathscr{C}_7 \le \vartheta/2$$

and

$$\sup_{\tau\in(0,\tau_U],\ \mathbf{Z}\in\mathbb{Z}} e^{\mathbf{Z}^\top \boldsymbol{\beta}_0(\tau)}\ e^{(p\ \mathscr{C}_2\ \vartheta^*)}(p\ \mathscr{C}_2\ \vartheta^*)\mathscr{C}_7 \leq \vartheta/2.$$

Here $\mathscr{X}$ and $\mathscr{Z}$ are covariate spaces related to validation and non-validation samples and $\mathscr{C}_7$ is the uniform upper bound for $\widetilde{f}(t|\mathbf{x})$ and $\widetilde{f}(t|\mathbf{z})$. If $\|\widetilde{\boldsymbol{\beta}}_n(\tau)-\boldsymbol{\beta}_0(\tau)\| \leq \vartheta^*$, then it is easy to see that

$$\left\| e^{\mathbf{x}^\top \widetilde{\boldsymbol{\beta}}_n(\tau)} - e^{\mathbf{x}^\top \boldsymbol{\beta}_0(\tau)} \right\| \leq \sup_{\tau\in(0,\tau_U],\ \mathbf{X}\in\mathscr{X}} e^{\mathbf{x}^\top \boldsymbol{\beta}_0(\tau)}\ e^{(p\ \mathscr{C}_2\ \vartheta^*)}$$

and

$$\left\| e^{\mathbf{z}^\top \widetilde{\boldsymbol{\beta}}_n(\tau)} - e^{\mathbf{z}^\top \boldsymbol{\beta}_0(\tau)} \right\| \leq \sup_{\tau\in(0,\tau_U],\ \mathbf{Z}\in\mathscr{Z}} e^{\mathbf{z}^\top \boldsymbol{\beta}_0(\tau)}\ e^{(p\ \mathscr{C}_2\ \vartheta^*)},$$

and thus $\sigma^2_{2d}\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} \leq \vartheta/2$ for all $\tau \in [\nu_\vartheta, \tau_U]$.

It then follows that for $n \geq \max(N_{\vartheta,\zeta,1}, N_{\vartheta,\zeta,2})$,

$$\begin{aligned}\Pr\left(\sup_{\tau\in(0,\tau_U]} \sigma^2_{2d}\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} > \vartheta\right) &\leq \Pr\left(\sup_{\tau\in(0,\tau_U]} \left\|\widetilde{\mu}_1\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} - \widetilde{\mu}_1\left\{\boldsymbol{\beta}_0(\tau)\right\}\right\| > \vartheta/8\right) \\ &\quad + \Pr\left(\sup_{\tau\in[\nu_\vartheta,\tau_U]} \left\|\widetilde{\boldsymbol{\beta}}_n(\tau) - \boldsymbol{\beta}_0(\tau)\right\| > \vartheta^*\right) \\ &< \zeta.\end{aligned}$$

This completes the proof of $\sigma^2_d\left\{\widetilde{\boldsymbol{\beta}}_n(\tau)\right\} \xrightarrow{\Pr} 0$ and *Lemma* D.1.

Proof of Theorem 3.2.2

From the proofs of *Theorem* 3.2.1 and *Lemma* D.1, we have

$$\begin{aligned}\sup_{\tau\in(0,\tau_U]} \Bigg\| &\frac{\sqrt{\rho}}{\sqrt{m_v}} \sum_{j\in\mathbb{V}} \boldsymbol{X}_j \left[\mathbb{N}_j\left(e^{\boldsymbol{X}_j^\top \widehat{\boldsymbol{\beta}}(\tau)}\right) - \mathbb{N}_j\left(e^{\boldsymbol{X}_j^\top \boldsymbol{\beta}_0(\tau)}\right)\right] \\ &- \sqrt{\rho\ m_v}\left[\boldsymbol{\mu}\left\{\widehat{\boldsymbol{\beta}}(\tau)\right\} - \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)\}\right] \\ &+ \frac{\sqrt{\rho}}{\sqrt{m_v}}\ \varrho \sum_{j\in\mathbb{V}} \Bigg[\boldsymbol{X}_j \mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\mathbb{N}\left(e^{\boldsymbol{Z}^\top \widehat{\boldsymbol{\beta}}(\tau)}\right)\right\} - \mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\boldsymbol{Z}\ \mathbb{N}\left(e^{\boldsymbol{Z}^\top \widehat{\boldsymbol{\beta}}(\tau)}\right)\right\} \\ &- \boldsymbol{X}_j \mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\mathbb{N}\left(e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau)}\right)\right\} + \mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\boldsymbol{Z}\ \mathbb{N}\left(e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau)}\right)\right\}\Bigg] \\ &- \sqrt{\rho\ m_v}\ \varrho\left[\widetilde{\boldsymbol{\mu}}\left\{\widehat{\boldsymbol{\beta}}(\tau)\right\} - \widetilde{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\}\right]\end{aligned}$$

$$+\frac{\sqrt{(1-\rho)}}{\sqrt{m_n}}\sum_{l\in\overline{\mathbb{V}}}\boldsymbol{Z}_l\left[\mathbb{N}_l\left(e^{\boldsymbol{Z}_l^\top\widehat{\boldsymbol{\beta}}(\tau)}\right)-\mathbb{N}_l\left(e^{\boldsymbol{Z}_l^\top\boldsymbol{\beta}_0(\tau)}\right)\right]$$

$$-\sqrt{(1-\rho)\ m_n}\left[\overline{\boldsymbol{\mu}}\left\{\widehat{\boldsymbol{\beta}}(\tau)\right\}+\overline{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\}\right]\Bigg\|\xrightarrow{\text{Pr}}0, \qquad \text{(D.1)}$$

Similarly we can get

$$\sup_{\tau\in(0,\tau_U]}\Bigg\|\frac{\sqrt{\rho}}{\sqrt{m_v}}\sum_{j\in\mathbb{V}}\boldsymbol{X}_j\left(\mathbb{I}\left[Y_j\geq e^{\boldsymbol{X}_j^\top\widehat{\boldsymbol{\beta}}(\tau)}\right]-\mathbb{I}\left[Y_j\geq e^{\boldsymbol{X}_j^\top\boldsymbol{\beta}_0(\tau)}\right]\right)$$

$$+\frac{\sqrt{\rho}}{\sqrt{m_v}}\ \varrho\sum_{j\in\mathbb{V}}\Bigg(\boldsymbol{X}_j\mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\mathbb{I}\left[Y\geq e^{\boldsymbol{X}^\top\widehat{\boldsymbol{\beta}}(\tau)}\right]\right\}$$

$$-\mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\boldsymbol{Z}\ \mathbb{I}\left[Y\geq e^{\boldsymbol{X}^\top\widehat{\boldsymbol{\beta}}(\tau)}\right]\right\}-\boldsymbol{X}_j\mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\mathbb{I}\left[Y\geq e^{\boldsymbol{Z}^\top\boldsymbol{\beta}_0(\tau)}\right]\right\}$$

$$+\mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\boldsymbol{Z}\ \mathbb{I}\left[Y\geq e^{\boldsymbol{Z}^\top\boldsymbol{\beta}_0(\tau)}\right]\right\}\Bigg)$$

$$+\frac{\sqrt{(1-\rho)}}{\sqrt{m_n}}\sum_{l\in\overline{\mathbb{V}}}\boldsymbol{Z}_l\left(\mathbb{I}\left[Y_l\geq e^{\boldsymbol{Z}_l^\top\widehat{\boldsymbol{\beta}}(\tau)}\right]-\mathbb{I}\left[Y_l\geq e^{\boldsymbol{Z}_l^\top\boldsymbol{\beta}_0(\tau)}\right]\right)$$

$$-\sqrt{\rho\ m_v}\left[\boldsymbol{\mu}^*\left\{\widehat{\boldsymbol{\beta}}(\tau)\right\}-\boldsymbol{\mu}^*\{\boldsymbol{\beta}_0(\tau)\}\right]-\sqrt{\rho\ m_v}\ \varrho\left[\widetilde{\boldsymbol{\mu}}^*\{\widehat{\boldsymbol{\beta}}(\tau)\}-\widetilde{\boldsymbol{\mu}}^*\{\boldsymbol{\beta}_0(\tau)\}\right]$$

$$-\sqrt{(1-\rho)\ m_n}\left[\overline{\boldsymbol{\mu}}^*\left\{\widehat{\boldsymbol{\beta}}(\tau)\right\}+\overline{\boldsymbol{\mu}}^*\{\boldsymbol{\beta}_0(\tau)\}\right]\Bigg\|\xrightarrow{\text{Pr}}0. \qquad \text{(D.2)}$$

$\sqrt{n}\ S_n(\widehat{\boldsymbol{\beta}},\tau)=o_{(0,\tau_U]}(1)$, a.s. because $\sqrt{n}\ \|\mathbb{S}_L\|\to 0$. This is true because, by the definition of $S_n(\widehat{\boldsymbol{\beta}},\tau)$,

$$\sup_{\tau\in[\tau_k,\tau_{k+1}]}\sqrt{n}\ \left\|S_n(\widehat{\boldsymbol{\beta}},\tau)-S_n(\widehat{\boldsymbol{\beta}},\tau_k)\right\|\leq\sqrt{n}\ \mathscr{C}_2\{H(\tau_{k+1})-H(\tau_k)\}$$

$$\leq\sqrt{n}\ \mathscr{C}_2\ a_n/(1-\tau_U).$$

Given that $\rho\ \boldsymbol{\mu}\{\widehat{\boldsymbol{\beta}}(\tau)\}+\rho\ \varrho\ \widetilde{\boldsymbol{\mu}}\{\widehat{\boldsymbol{\beta}}(\tau)\}+(1-\rho)\ \overline{\boldsymbol{\mu}}\{\widehat{\boldsymbol{\beta}}(\tau)\}$ uniformly converges in probability to $\rho\ \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)\}+\rho\ \varrho\ \widetilde{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\}+(1-\rho)\ \overline{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\}$ for $\tau\in(0,\tau_U]$, by (D.1) and (D.2),

$$-\sqrt{n}\ S_n(\boldsymbol{\beta}_0,\tau)$$

$$=\sqrt{\rho\ m_v}\left[\boldsymbol{\mu}\left\{\widehat{\boldsymbol{\beta}}(\tau)\right\}-\boldsymbol{\mu}\left\{\boldsymbol{\beta}_0(\tau)\right\}\right]+\sqrt{\rho\ m_v}\ \varrho\left[\widetilde{\boldsymbol{\mu}}\left\{\widehat{\boldsymbol{\beta}}(\tau)\right\}-\widetilde{\boldsymbol{\mu}}\left\{\boldsymbol{\beta}_0(\tau)\right\}\right]$$

$$+\sqrt{(1-\rho)\ m_n}\left[\overline{\boldsymbol{\mu}}\left\{\widehat{\boldsymbol{\beta}}(\tau)\right\}-\overline{\boldsymbol{\mu}}\left\{\boldsymbol{\beta}_0(\tau)\right\}\right]$$

$$
\begin{aligned}
&-\int_0^{\mathcal{T}} \Bigg(\sqrt{\rho\, m_v}\Big[\boldsymbol{\mu}^*\left\{\widehat{\boldsymbol{\beta}}(u)\right\} - \boldsymbol{\mu}^*\left\{\boldsymbol{\beta}_0(u)\right\}\Big] + \sqrt{\rho\, m_v}\ \varrho\left[\widetilde{\boldsymbol{\mu}}^*\left\{\widehat{\boldsymbol{\beta}}(u)\right\} - \widetilde{\boldsymbol{\mu}}^*\left\{\boldsymbol{\beta}_0(u)\right\}\right] \\
&\qquad + \sqrt{(1-\rho)\ m_n}\Big[\overline{\boldsymbol{\mu}}^*\left\{\widehat{\boldsymbol{\beta}}(u)\right\} - \overline{\boldsymbol{\mu}}^*\left\{\boldsymbol{\beta}_0(u)\right\}\Big]\Bigg) dH(u) + O_p\left(\frac{\mathbf{1}}{\sqrt{n}}\right) + o_{(0,\tau_U])}(\mathbf{1}) \\
&= \sqrt{\rho\, m_v}\Big[\boldsymbol{\mu}\left\{\widehat{\boldsymbol{\beta}}(\mathcal{T})\right\} - \boldsymbol{\mu}\left\{\boldsymbol{\beta}_0(\mathcal{T})\right\}\Big] + \sqrt{\rho\, m_v}\ \varrho\left[\widetilde{\boldsymbol{\mu}}\left\{\widehat{\boldsymbol{\beta}}(\mathcal{T})\right\} - \widetilde{\boldsymbol{\mu}}\left\{\boldsymbol{\beta}_0(\mathcal{T})\right\}\right] \\
&\qquad + \sqrt{(1-\rho)\ m_n}\Big[\overline{\boldsymbol{\mu}}\left\{\widehat{\boldsymbol{\beta}}(\mathcal{T})\right\} - \overline{\boldsymbol{\mu}}\left\{\boldsymbol{\beta}_0(\mathcal{T})\right\}\Big] \\
&\quad -\int_0^{\mathcal{T}} \Bigg\{\Big[\sqrt{\rho\, m_v}\ \boldsymbol{B}^*\left\{\boldsymbol{\beta}_0(u)\right\}\left(\boldsymbol{B}\left\{\boldsymbol{\beta}_0(u)\right\}\right)^{-1} + \sqrt{\rho\, m_v}\ \varrho\ \widetilde{\boldsymbol{B}}^*\left\{\boldsymbol{\beta}_0(u)\right\}\left(\widetilde{\boldsymbol{B}}\left\{\boldsymbol{\beta}_0(u)\right\}\right)^{-1} \\
&\qquad + \sqrt{(1-\rho)\ m_n}\ \overline{\boldsymbol{B}}^*\left\{\boldsymbol{\beta}_0(u)\right\}\left(\overline{\boldsymbol{B}}\left\{\boldsymbol{\beta}_0(u)\right\}\right)^{-1} + O_p\left(\frac{\mathbf{1}}{\sqrt{n}}\right) + o_{(0,\tau_U]}(\mathbf{1})\Big] \\
&\qquad \times \Bigg(\sqrt{\rho\, m_v}\Big[\boldsymbol{\mu}\left\{\widehat{\boldsymbol{\beta}}(u)\right\} - \boldsymbol{\mu}\left\{\boldsymbol{\beta}_0(u)\right\}\Big] + \sqrt{\rho\, m_v}\ \varrho\left[\widetilde{\boldsymbol{\mu}}\left\{\widehat{\boldsymbol{\beta}}(u)\right\} - \widetilde{\boldsymbol{\mu}}\left\{\boldsymbol{\beta}_0(u)\right\}\right] \\
&\qquad + \sqrt{(1-\rho)\ m_n}\Big[\overline{\boldsymbol{\mu}}\left\{\widehat{\boldsymbol{\beta}}(u)\right\} - \overline{\boldsymbol{\mu}}\left\{\boldsymbol{\beta}_0(u)\right\}\Big]\Bigg)\Bigg\} dH(u) \\
&\qquad + O_p\left(\frac{\mathbf{1}}{\sqrt{n}}\right) + o_{(0,\tau_U])}(\mathbf{1}).
\end{aligned}
$$

$\sqrt{n}\, S_n(\boldsymbol{\beta}_0, \tau) = 0$ can be viewed as a stochastic differential equation for $\sqrt{\rho\, m_v}\ [\boldsymbol{\mu}\{\widehat{\boldsymbol{\beta}}(\tau)\} - \boldsymbol{\mu}\{\boldsymbol{\beta}_0(\tau)\}] + \sqrt{\rho\, m_v}\ \varrho\ [\widetilde{\boldsymbol{\mu}}\{\widehat{\boldsymbol{\beta}}(\tau)\} - \widetilde{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\}] + \sqrt{(1-\rho)\ m_n}\ [\overline{\boldsymbol{\mu}}\{\widehat{\boldsymbol{\beta}}(\tau)\} - \overline{\boldsymbol{\mu}}\{\boldsymbol{\beta}_0(\tau)\}]$, and using the production integration theory (Gill and Johansen 1990; Andersen et al. 1998, II.6), we get

$$
\begin{aligned}
&\sqrt{\rho\, m_v}\Big[\boldsymbol{\mu}\left\{\widehat{\boldsymbol{\beta}}(\mathcal{T})\right\} - \boldsymbol{\mu}\left\{\boldsymbol{\beta}_0(\mathcal{T})\right\}\Big] + \sqrt{\rho\, m_v}\ \varrho\left[\widetilde{\boldsymbol{\mu}}\left\{\widehat{\boldsymbol{\beta}}(\mathcal{T})\right\} - \widetilde{\boldsymbol{\mu}}\left\{\boldsymbol{\beta}_0(\mathcal{T})\right\}\right] \\
&\qquad + \sqrt{(1-\rho)\ m_n}\Big[\overline{\boldsymbol{\mu}}\left\{\widehat{\boldsymbol{\beta}}(\mathcal{T})\right\} - \overline{\boldsymbol{\mu}}\left\{\boldsymbol{\beta}_0(\mathcal{T})\right\}\Big] \\
&= \boldsymbol{\phi}\{-\sqrt{n}\ S_n(\boldsymbol{\beta}_0, \mathcal{T})\} + O_p\left(\frac{\mathbf{1}}{\sqrt{n}}\right) + o_{(0,\tau_U])}(\mathbf{1}), \qquad \text{(D.3)}
\end{aligned}
$$

where $\boldsymbol{\phi}$ is a map from $\mathscr{F}$ to $\mathscr{F}$ such that for $\boldsymbol{\gamma} \in \mathscr{F}$,

$$
\boldsymbol{\phi}(\boldsymbol{\gamma})(\mathcal{T}) = \int_0^{\mathcal{T}} \mathcal{I}(s, \tau) d\boldsymbol{\gamma}(s),
$$

with

$$
\mathcal{I}(s,t) = \boldsymbol{\pi}_{u\in(s,t]}\Big\{\mathbf{I}_p + \Big[\sqrt{\rho\, m_v}\ \boldsymbol{B}^*\left\{\boldsymbol{\beta}_0(u)\right\}\left(\boldsymbol{B}\left\{\boldsymbol{\beta}_0(u)\right\}\right)^{-1}
$$

$$+ \sqrt{\rho\ m_v}\ \varrho\ \widetilde{\boldsymbol{B}}^*\left\{\boldsymbol{\beta}_0(u)\right\}\left(\widetilde{\boldsymbol{B}}\left\{\boldsymbol{\beta}_0(u)\right\}\right)^{-1}$$
$$\left. + \sqrt{(1-\rho)\ m_n}\ \overline{\boldsymbol{B}}^*\left\{\boldsymbol{\beta}_0(u)\right\}\left(\overline{\boldsymbol{B}}\left\{\boldsymbol{\beta}_0(u)\right\}\right)^{-1}\right] dH(u)\right\}$$

and $\mathscr{F} = \{\boldsymbol{\gamma} : [0, \tau_U] \to \Re^p,\ \boldsymbol{\gamma}$ is left-continuous with right limit, $\boldsymbol{\gamma}(0) = 0\}$.

By considering that

$$\left\{\rho\ \boldsymbol{X}_j \mathbb{N}_j\left(e^{\boldsymbol{X}_j^\top \boldsymbol{\beta}_0(\tau)}\right) + (1-\rho)\ \boldsymbol{Z}_l \mathbb{N}_l\left(e^{\boldsymbol{Z}_l^\top \boldsymbol{\beta}_0(\tau)}\right)\right.$$
$$\left. + \rho\ \varrho\left[\boldsymbol{X}_j \mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\mathbb{N}\left(e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau)}\right)\right\} - \mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\boldsymbol{Z}\mathbb{N}\left(e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau)}\right)\right\}\right], \tau \in [0, \tau_U]\right\}$$

is a VC-class [van der Vaart and Wellner, 1996] and

$$\int_0^\tau \left(\rho\ \boldsymbol{X}_j \mathbb{I}\left[Y_j \geq e^{\boldsymbol{X}_j^\top \boldsymbol{\beta}_0(u)}\right] + (1-\rho)\ \boldsymbol{Z}_l \mathbb{I}\left[Y_l \geq e^{\boldsymbol{Z}_l^\top \boldsymbol{\beta}_0(u)}\right]\right.$$
$$\left. + \rho\ \varrho\left[\boldsymbol{X}_j \mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\mathbb{I}\left[Y \geq e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(u)}\right]\right\} - \mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\boldsymbol{Z}\mathbb{I}\left[Y \geq e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(u)}\right]\right\}\right]\right) dH(u)$$

is Lipschitz in τ, and by using the permanence properties of the Donsker class we can tell that

$$\left\{\rho\ \boldsymbol{X}_j \mathbb{N}_j\left(e^{\boldsymbol{X}_j^\top \boldsymbol{\beta}_0(\tau)}\right) + (1-\rho)\ \boldsymbol{Z}_l \mathbb{N}_l\left(e^{\boldsymbol{Z}_l^\top \boldsymbol{\beta}_0(\tau)}\right)\right.$$
$$+ \rho\ \varrho\left[\boldsymbol{X}_j \mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\mathbb{N}\left(e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau)}\right)\right\} - \mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\boldsymbol{Z}\mathbb{N}\left(e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau)}\right)\right\}\right]$$
$$- \int_0^\tau \left(\rho\ \varrho\left[\boldsymbol{X}_j \mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\mathbb{I}\left[Y \geq e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(u)}\right]\right\} - \mathbb{W}_{qM\overline{\mathbb{V}}}\left\{\boldsymbol{Z}\mathbb{I}\left[Y \geq e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(u)}\right]\right\}\right]\right.$$
$$\left.\left. + \rho\ \boldsymbol{X}_j \mathbb{I}\left[Y_j \geq e^{\boldsymbol{X}_j^\top \boldsymbol{\beta}_0(u)}\right] + (1-\rho)\ \boldsymbol{Z}_l \mathbb{I}\left[Y_l \geq e^{\boldsymbol{Z}_l^\top \boldsymbol{\beta}_0(u)}\right]\right) dH(u), \tau \in [\nu, \tau_U]\right\}$$

is a Donsker class. By the Donsker theorem, $-\sqrt{n}\ S_n(\boldsymbol{\beta}_0, \tau)$ converges weakly to a tight Gaussian process, $\boldsymbol{G}(\tau)$, with mean 0 and covariance $\boldsymbol{\Sigma}(s,t)$ for $\tau \in [0, \tau_U]$, where $\boldsymbol{\Sigma}(s,t) = E\{\boldsymbol{\iota}_{\widehat{j}}(s)\boldsymbol{\iota}_{\widehat{j}}(t)^\top\} + E\{\boldsymbol{\iota}_{\widehat{l}}(s)\boldsymbol{\iota}_{\widehat{l}}(t)^\top\}$ with

$$\boldsymbol{\iota}_{\widehat{j}}(\tau) = \rho\ \boldsymbol{X}_{\widehat{j}} \mathbb{N}_{\widehat{j}}\left(e^{\boldsymbol{X}_{\widehat{j}}^\top \boldsymbol{\beta}_0(\tau)}\right)$$

$$
\begin{aligned}
&+ \rho\ \varrho \left[\boldsymbol{X}_{\hat{j}} \mathbb{W}_{qM\overline{\mathbb{V}}} \left\{ \mathbb{N}\left(e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau)} \right) \right\} - \mathbb{W}_{qM\overline{\mathbb{V}}} \left\{ \boldsymbol{Z}\ \mathbb{N}\left(e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(\tau)} \right) \right\} \right] \\
&- \rho \int_0^\tau \left(\boldsymbol{X}_{\hat{j}} \mathbb{I}\left[Y_{\hat{j}} \geq e^{\boldsymbol{X}_{\hat{j}}^\top \boldsymbol{\beta}_0(u)} \right] \right. \\
&\left. + \varrho \left[\boldsymbol{X}_{\hat{j}} \mathbb{W}_{qM\overline{\mathbb{V}}} \left\{ \mathbb{I}\left[Y \geq e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(u)} \right] \right\} - \mathbb{W}_{qM\overline{\mathbb{V}}} \left\{ \boldsymbol{Z}\ \mathbb{I}\left[Y \geq e^{\boldsymbol{Z}^\top \boldsymbol{\beta}_0(u)} \right] \right\} \right] \right) dH(u)
\end{aligned}
$$

and

$$
\boldsymbol{\iota}_{\hat{l}}(\tau) = (1-\rho)\ \boldsymbol{Z}_{\hat{l}} \mathbb{N}_{\hat{l}} \left(e^{\boldsymbol{Z}_{\hat{l}}^\top \boldsymbol{\beta}_0(\tau)} \right) - \int_0^\tau (1-\rho)\ \boldsymbol{Z}_{\hat{l}} \mathbb{I}\left[Y_{\hat{l}} \geq e^{\boldsymbol{Z}_{\hat{l}}^\top \boldsymbol{\beta}_0(u)} \right] dH(u).
$$

$\boldsymbol{\phi}\{\boldsymbol{G}(\tau)\}$ for $\tau \in (0, \tau_U]$ is also Gaussian process because $\boldsymbol{\phi}$ is a linear operator (Römisch 2005). $\rho\ (\boldsymbol{B}\{\boldsymbol{\beta}_0(\tau)\})^{-1} + \rho\ \varrho\ (\widetilde{\boldsymbol{B}}\{\boldsymbol{\beta}_0(\tau)\})^{-1} + (1-\rho)\ (\overline{\boldsymbol{B}}\{\boldsymbol{\beta}_0(\tau)\})^{-1}$ is bounded uniformly for $\tau \in [\nu, \tau_U]$ (by **R4**). Applying the Taylor expansion technique to $\boldsymbol{\eta}[\rho\ \boldsymbol{\mu}(\widehat{\boldsymbol{\beta}}(\tau)) + \rho\ \varrho\ \widetilde{\boldsymbol{\mu}}(\widehat{\boldsymbol{\beta}}(\tau)) + (1-\rho)\ \overline{\boldsymbol{\mu}}(\widehat{\boldsymbol{\beta}}(\tau))] - \boldsymbol{\eta}[\rho\ \boldsymbol{\mu}(\boldsymbol{\beta}_0(\tau)) + \rho\ \varrho\ \widetilde{\boldsymbol{\mu}}(\boldsymbol{\beta}_0(\tau)) + (1-\rho)\ \overline{\boldsymbol{\mu}}(\boldsymbol{\beta}_0(\tau))]$ and the continuous mapping theorem, we get that for $\tau \in [\nu, \tau_U]$, $\sqrt{n}\ \{\widehat{\boldsymbol{\beta}}(\tau) - \boldsymbol{\beta}_0(\tau)\}$ converges weakly to $[\rho\ (\boldsymbol{B}\{\boldsymbol{\beta}_0(\tau)\})^{-1} + \rho\ \varrho\ (\widetilde{\boldsymbol{B}}\{\boldsymbol{\beta}_0(\tau)\})^{-1} + (1-\rho)\ (\overline{\boldsymbol{B}}\{\boldsymbol{\beta}_0(\tau)\})^{-1}]\boldsymbol{\phi}\{\boldsymbol{G}(\tau)\}$, which is also Gaussian.

Appendix E

Simulation Result Summary for Numerical Study - II in Chapter 4

	n	$\tau \rightarrow$	**CQR**			**CQR-EL2**		
			0.25	0.50	0.75	0.25	0.50	0.75
Bias	100	β_0.	0.0042	0.0170	0.0647	0.0217	0.0275	0.0720
		β_1.	0.0029	0.0035	0.0094	-0.0491	-0.0411	-0.0090
		β_2.	-0.0049	-0.0141	-0.0100	0.0116	-0.0029	-0.0194
	200	β_0	0.0218	0.0298	0.0501	0.0220	0.0323	0.0562
		β_1	0.0016	0.0026	0.0057	-0.0295	-0.0273	-0.0119
		β_2	-0.0020	-0.0032	-0.0078	0.0034	0.0053	-0.0011
SE	100	β_0.	0.1449	0.1404	0.2268	0.1273	0.1233	0.2160
		β_1.	0.1533	0.1515	0.2141	0.1475	0.1416	0.2075
		β_2.	0.1519	0.1525	0.2198	0.1416	0.1414	0.2162
	200	β_0	0.0973	0.0929	0.1292	0.0840	0.0798	0.1239
		β_1	0.1040	0.1029	0.1341	0.0970	0.0921	0.1278
		β_2	0.1041	0.1027	0.1354	0.0957	0.0936	0.1304
CP	100	β_0.	93.3	93.4	95.7	94.3	96.1	96.8
		β_1.	94.7	95.8	96.5	94.6	96.1	96.9
		β_2.	96.0	96.3	96.4	95.4	95.4	97.4
	200	β_0	92.3	91.9	92.7	92.9	92.3	94.3
		β_1	94.5	96.2	95.0	95.3	95.3	94.8
		β_2	93.6	95.0	95.2	93.5	94.9	95.9

Table E.1: Bias, SE and CP of regression parameters for Case (i) model with independent covariates ($\sigma_{x_1,x_2} = 0$)

	n	$\tau \rightarrow$	**CQR** 0.25	0.50	0.75	**CQR-EL2** 0.25	0.50	0.75
Bias	100	β_0.	0.0105	0.0288	0.1088	0.0306	0.0461	0.1139
		β_1.	0.0063	0.0214	0.0169	-0.0841	-0.0503	-0.0216
		β_2.	0.0164	0.0096	-0.0170	0.0329	0.0260	-0.0094
	200	β_0	0.0267	0.0355	0.0821	0.0419	0.0508	0.0921
		β_1	0.0006	-0.0032	0.0050	-0.0022	-0.0010	-0.0188
		β_2	0.0112	0.0025	0.0051	0.0251	0.0137	0.0133
SE	100	β_0.	0.1871	0.1538	0.2980	0.1619	0.1379	0.2768
		β_1.	0.1946	0.1664	0.2698	0.1863	0.1595	0.2548
		β_2.	0.1955	0.1676	0.2733	0.1787	0.1549	0.2632
	200	β_0	0.1235	0.1029	0.1621	0.1048	0.0900	0.1551
		β_1	0.1301	0.1146	0.1663	0.1214	0.1052	0.1575
		β_2	0.1315	0.1149	0.1671	0.1185	0.1044	0.1606
CP	100	β_0.	95.5	93.1	94.7	95.9	94.2	97.5
		β_1.	95.6	93.5	96.4	94.8	93.3	96.7
		β_2.	95.9	95.4	96.4	94.2	94.2	96.3
	200	β_0	93.1	91.2	94.0	93.5	93.0	94.7
		β_1	95.0	95.5	95.4	94.5	94.0	94.9
		β_2	95.5	95.7	95.5	94.8	94.5	95.4

Table E.2: Bias, SE and CP of regression parameters for Case (ii) model with independent covariates ($\sigma_{x_1,x_2} = 0$)

	n	$\tau \rightarrow$	**CQR** 0.25	0.50	0.75	**CQR-EL2** 0.25	0.50	0.75
Bias	100	β_0.	0.0062	0.0088	0.0224	0.0127	0.0146	0.0302
		β_1.	0.0042	0.0051	0.0076	-0.0071	-0.0043	0.0021
		β_2.	-0.0038	-0.0039	-0.0069	0.0018	0.0017	-0.0040
	200	β_0	0.0064	0.0072	0.0167	0.0094	0.0105	0.0197
		β_1	0.0012	0.0038	0.0033	-0.0042	-0.0026	-0.0007
		β_2	-0.0015	-0.0031	-0.0017	0.0009	-0.0003	0.0015
SE	100	β_0.	0.0472	0.0466	0.0767	0.0448	0.0445	0.0801
		β_1.	0.0566	0.0570	0.0796	0.0541	0.0549	0.0830
		β_2.	0.0567	0.0575	0.0807	0.0538	0.0558	0.0833
	200	β_0	0.0313	0.0301	0.0402	0.0292	0.0283	0.0396
		β_1	0.0371	0.0377	0.0489	0.0348	0.0356	0.0484
		β_2	0.0367	0.0376	0.0488	0.0344	0.0359	0.0488
CP	100	β_0.	94.4	95.0	96.1	93.9	94.7	96.9
		β_1.	95.0	95.2	95.5	94.6	94.7	96.3
		β_2.	96.6	96.7	97.3	95.8	96.4	97.3
	200	β_0	94.1	93.4	94.9	93.9	93.8	94.9
		β_1	94.0	94.9	96.0	94.1	94.3	95.0
		β_2	94.6	95.0	95.3	94.0	95.4	94.3

Table E.3: Bias, SE and CP of regression parameters for Case (iii) model with independent covariates ($\sigma_{x_1,x_2} = 0$)

	n	$\tau \rightarrow$	**CQR**			**CQR-EL2**		
			0.25	0.50	0.75	0.25	0.50	0.75
Bias	100	β_0.	0.0066	0.0097	0.0364	0.0189	0.0169	0.0419
		β_1.	0.0031	0.0039	0.0041	-0.0138	-0.0073	-0.0000
		β_2.	0.0008	-0.0009	-0.0018	0.0074	0.0060	0.0024
	200	β_0	0.0083	0.0089	0.0243	0.0124	0.0119	0.0273
		β_1	-0.0020	0.0016	0.0017	-0.0097	-0.0051	-0.0032
		β_2	0.0008	-0.0012	-0.0031	0.0019	0.0004	-0.0020
SE	100	β_0.	0.0600	0.0507	0.1103	0.0548	0.0486	0.1159
		β_1.	0.0667	0.0592	0.0993	0.0618	0.0581	0.1018
		β_2.	0.0677	0.0600	0.1014	0.0616	0.0578	0.1066
	200	β_0	0.0395	0.0327	0.0521	0.0359	0.0304	0.0516
		β_1	0.0429	0.0386	0.0568	0.0397	0.0364	0.0558
		β_2	0.0429	0.0389	0.0580	0.0397	0.0368	0.0579
CP	100	β_0.	93.5	95.0	97.7	92.9	95.2	97.6
		β_1.	95.6	96.6	97.0	94.2	95.5	97.4
		β_2.	96.0	96.2	97.3	96.3	97.0	97.6
	200	β_0	93.0	93.9	94.9	93.3	94.2	95.8
		β_1	95.6	95.8	94.7	94.0	95.5	95.2
		β_2	94.5	95.9	95.5	94.9	96.0	94.7

Table E.4: Bias, SE and CP of regression parameters for Case (iv) model with independent covariates ($\sigma_{x_1,x_2} = 0$)

	n	$\tau \rightarrow$	CQR 0.25	CQR 0.50	CQR 0.75	CQR-EL2 0.25	CQR-EL2 0.50	CQR-EL2 0.75
Bias	100	β_0.	0.0120	0.0179	0.0588	0.0203	0.0284	0.0663
		β_1.	0.0006	0.0016	0.0134	-0.0679	-0.0632	-0.0172
		β_2.	0.0001	-0.0047	-0.0146	0.0184	0.0133	-0.0053
	200	β_0	0.0268	0.0284	0.0487	0.0238	0.0297	0.0518
		β_1	-0.0032	-0.0001	-0.0037	-0.0371	-0.0351	-0.0243
		β_2	-0.0006	-0.0072	-0.0030	0.0089	0.0039	0.0023
SE	100	β_0.	0.1438	0.1386	0.2225	0.1274	0.1221	0.2098
		β_1.	0.1788	0.1759	0.2493	0.1763	0.1671	0.2416
		β_2.	0.1768	0.1749	0.2543	0.1686	0.1664	0.2491
	200	β_0	0.0972	0.0922	0.1273	0.0840	0.0789	0.1209
		β_1	0.1197	0.1193	0.1543	0.1124	0.1091	0.1470
		β_2	0.1203	0.1193	0.1553	0.1118	0.1096	0.1510
CP	100	β_0.	94.0	93.4	95.4	94.6	96.5	97.5
		β_1.	95.4	96.7	95.7	94.7	95.0	97.2
		β_2.	95.9	96.4	96.3	94.8	94.9	97.0
	200	β_0	93.3	92.1	94.6	95.0	94.7	96.1
		β_1	94.6	94.4	94.7	94.4	94.5	94.4
		β_2	95.0	94.4	95.5	94.8	94.2	95.0

Table E.5: Bias, SE and CP of regression parameters for Case (i) model with dependent covariates ($\sigma_{x_1,x_2} = 0.5$)

	n	$\tau \rightarrow$	**CQR** 0.25	0.50	0.75	**CQR-EL2** 0.25	0.50	0.75
Bias	100	β_0.	0.0092	0.0301	0.1157	0.0525	0.0623	0.1337
		β_1.	0.0241	0.0040	-0.0053	-0.0826	-0.0711	-0.0348
		β_2.	-0.0140	-0.0102	-0.0016	0.0182	0.0182	0.0039
	200	β_0	0.0264	0.0258	0.0605	0.0498	0.0451	0.0825
		β_1	0.0027	0.0004	0.0034	-0.0411	-0.0436	-0.0263
		β_2	-0.0010	-0.0017	-0.0066	0.0120	0.0168	0.0119
SE	100	β_0.	0.1868	0.1530	0.2943	0.1618	0.1391	0.2699
		β_1.	0.2261	0.1970	0.3164	0.2172	0.1912	0.2958
		β_2.	0.2261	0.1962	0.3163	0.2081	0.1843	0.3035
	200	β_0	0.1228	0.1007	0.1619	0.1061	0.0894	0.1565
		β_1	0.1495	0.1307	0.1938	0.1416	0.1211	0.1851
		β_2	0.1497	0.1305	0.1960	0.1376	0.1194	0.1892
CP	100	β_0.	94.7	93.8	95.9	94.3	94.4	96.5
		β_1.	95.7	96.6	96.6	94.9	95.5	95.9
		β_2.	96.1	95.5	97.2	94.2	96.1	96.6
	200	β_0	91.7	92.9	93.4	93.1	93.9	94.3
		β_1	96.4	96.2	96.4	94.7	95.3	94.9
		β_2	95.2	95.0	96.0	94.5	94.7	96.3

Table E.6: Bias, SE and CP of regression parameters for Case (ii) model with dependent covariates ($\sigma_{x_1,x_2} = 0.5$)

	n	$\tau \rightarrow$	CQR 0.25	CQR 0.50	CQR 0.75	CQR-EL2 0.25	CQR-EL2 0.50	CQR-EL2 0.75
Bias	100	β_0.	0.0067	0.0104	0.0202	0.0123	0.0161	0.0252
		β_1.	0.0037	0.0040	0.0091	-0.0071	-0.0062	0.0030
		β_2.	-0.0013	-0.0048	-0.0105	0.0017	0.0010	-0.0058
	200	β_0	0.0073	0.0092	0.0182	0.0096	0.0107	0.0194
		β_1	0.0010	0.0025	0.0030	-0.0041	-0.0019	0.0000
		β_2	-0.0006	-0.0021	-0.0041	0.0005	-0.0009	-0.0019
SE	100	β_0.	0.0458	0.0440	0.0770	0.0439	0.0431	0.0802
		β_1.	0.0604	0.0607	0.0877	0.0592	0.0608	0.0917
		β_2.	0.0604	0.0610	0.0894	0.0587	0.0613	0.0932
	200	β_0	0.0308	0.0293	0.0400	0.0290	0.0278	0.0396
		β_1	0.0398	0.0409	0.0547	0.0381	0.0393	0.0544
		β_2	0.0396	0.0411	0.0549	0.0380	0.0396	0.0550
CP	100	β_0.	94.6	93.9	96.2	94.0	94.7	98.0
		β_1.	96.6	95.9	97.1	96.0	96.4	97.1
		β_2.	96.7	96.1	97.2	95.7	96.0	97.2
	200	β_0	94.1	92.8	93.8	93.9	94.2	94.1
		β_1	95.8	95.1	95.5	94.7	94.6	94.9
		β_2	95.0	94.3	93.9	93.8	94.1	93.9

Table E.7: Bias, SE and CP of regression parameters for Case (iii) model with dependent covariates ($\sigma_{x_1,x_2} = 0.5$)

	n	$\tau \rightarrow$	CQR 0.25	CQR 0.50	CQR 0.75	CQR-EL2 0.25	CQR-EL2 0.50	CQR-EL2 0.75
Bias	100	β_0.	0.0068	0.0122	0.0332	0.0104	0.0155	0.0341
		β_1.	-0.0006	0.0045	0.0115	-0.0159	-0.0081	0.0050
		β_2.	-0.0000	-0.0045	-0.0118	-0.0010	-0.0029	-0.0118
	200	β_0	0.0075	0.0083	0.0226	0.0097	0.0099	0.0228
		β_1	-0.0010	0.0013	0.0034	-0.0092	-0.0053	0.0002
		β_2	0.0014	-0.0003	-0.0026	0.0013	0.0011	-0.0021
SE	100	β_0.	0.0581	0.0488	0.1093	0.0539	0.0465	0.1084
		β_1.	0.0723	0.0655	0.1118	0.0705	0.0644	0.1121
		β_2.	0.0726	0.0661	0.1144	0.0694	0.0647	0.1152
	200	β_0	0.0384	0.0316	0.0518	0.0353	0.0297	0.0509
		β_1	0.0477	0.0422	0.0644	0.0451	0.0402	0.0637
		β_2	0.0470	0.0427	0.0645	0.0443	0.0409	0.0646
CP	100	β_0.	94.3	93.0	97.1	94.1	94.6	98.0
		β_1.	95.3	96.6	96.5	94.7	94.8	98.4
		β_2.	96.4	95.7	97.3	95.9	96.5	97.2
	200	β_0	93.8	92.4	95.3	94.5	94.2	95.3
		β_1	94.4	94.7	95.4	94.4	94.5	95.9
		β_2	94.3	96.2	96.7	94.3	95.9	96.2

Table E.8: Bias, SE and CP of regression parameters for Case (iv) model with dependent covariates ($\sigma_{x_1,x_2} = 0.5$)

	n	$\tau \rightarrow$	**CQR** 0.25	0.50	0.75	**CQR-EL2** 0.25	0.50	0.75
Bias	100	β_0.	0.0096	0.0191	0.0594	0.0174	0.0273	0.0685
		β_1.	0.0038	0.0063	0.0162	-0.0381	-0.0394	-0.0177
		β_2.	0.0035	-0.0008	-0.0102	0.0072	0.0033	-0.0060
	200	β_0	0.0227	0.0267	0.0543	0.0217	0.0281	0.0540
		β_1	-0.0011	-0.0005	0.0019	-0.0240	-0.0234	-0.0154
		β_2	0.0012	-0.0032	-0.0034	0.0043	0.0006	-0.0009
SE	100	β_0.	0.1437	0.1394	0.2205	0.1277	0.1216	0.2113
		β_1.	0.1526	0.1517	0.2064	0.1459	0.1398	0.2002
		β_2.	0.1536	0.1544	0.2186	0.1555	0.1569	0.2205
	200	β_0	0.0982	0.0914	0.1276	0.0852	0.0790	0.1221
		β_1	0.1035	0.1011	0.1351	0.0958	0.0914	0.1281
		β_2	0.1062	0.1023	0.1351	0.1066	0.1031	0.1360
CP	100	β_0.	92.6	93.5	95.6	95.4	95.9	97.3
		β_1.	95.5	95.5	96.9	95.0	94.9	96.2
		β_2.	95.7	96.2	96.6	95.0	94.6	96.8
	200	β_0	94.0	93.4	93.9	94.8	95.4	95.5
		β_1	95.1	95.2	95.2	94.2	94.9	96.3
		β_2	95.4	94.3	94.1	95.0	94.7	94.2

Table E.9: Bias, SE and CP of regression parameters for Case (i) model with independent covariates ($\sigma_{x_1,x_2} = 0$)

	n	$\tau \rightarrow$	**CQR** 0.25	0.50	0.75	**CQR-EL2** 0.25	0.50	0.75
Bias	100	β_0.	0.0105	0.0353	0.0918	0.0113	0.0359	0.0920
		β_1.	-0.0030	0.0051	0.0164	-0.0647	-0.0458	-0.0199
		β_2.	0.0092	0.0046	-0.0032	0.0132	0.0084	0.0001
	200	β_0	0.0237	0.0272	0.0708	0.0208	0.0274	0.0702
		β_1	0.0019	0.0028	0.0081	-0.0411	-0.0343	-0.0206
		β_2	-0.0017	-0.0026	-0.0019	0.0017	0.0040	0.0035
SE	100	β_0.	0.1837	0.1542	0.2913	0.1610	0.1368	0.2742
		β_1.	0.1927	0.1657	0.2589	0.1811	0.1535	0.2477
		β_2.	0.1934	0.1669	0.2687	0.1955	0.1679	0.2686
	200	β_0	0.1235	0.1007	0.1667	0.1075	0.0891	0.1571
		β_1	0.1298	0.1126	0.1688	0.1208	0.1024	0.1582
		β_2	0.1304	0.1125	0.1696	0.1312	0.1136	0.1699
CP	100	β_0.	94.2	93.9	95.3	96.0	95.2	97.0
		β_1.	95.8	95.1	95.7	94.2	94.4	96.3
		β_2.	95.8	94.3	95.5	94.4	95.0	96.5
	200	β_0	94.0	91.9	94.0	93.8	93.5	94.7
		β_1	95.1	95.9	95.1	94.7	94.8	94.2
		β_2	93.4	95.8	95.0	93.9	95.5	94.7

Table E.10: Bias, SE and CP of regression parameters for Case (ii) model with independent covariates ($\sigma_{x_1,x_2} = 0$)

	n	$\tau \rightarrow$	**CQR** 0.25	0.50	0.75	**CQR-EL2** 0.25	0.50	0.75
Bias	100	β_0.	0.0062	0.0088	0.0224	0.0120	0.0139	0.0297
		β_1.	0.0042	0.0051	0.0076	-0.0054	-0.0045	0.0021
		β_2.	-0.0038	-0.0039	-0.0069	0.0008	0.0011	-0.0047
	200	β_0	0.0095	0.0111	0.0198	0.0116	0.0134	0.0221
		β_1	0.0007	0.0018	0.0023	-0.0035	-0.0022	-0.0010
		β_2	0.0011	-0.0016	-0.0006	0.0028	0.0004	0.0015
SE	100	β_0.	0.0472	0.0466	0.0767	0.0444	0.0441	0.0763
		β_1.	0.0566	0.0570	0.0796	0.0537	0.0541	0.0784
		β_2.	0.0567	0.0575	0.0807	0.0561	0.0581	0.0817
	200	β_0	0.0317	0.0302	0.0403	0.0297	0.0284	0.0396
		β_1	0.0371	0.0379	0.0492	0.0352	0.0360	0.0486
		β_2	0.0373	0.0372	0.0490	0.0365	0.0368	0.0496
CP	100	β_0.	94.4	95.0	96.1	94.0	95.2	96.6
		β_1.	95.0	95.2	95.5	95.8	95.0	96.3
		β_2.	96.6	96.7	97.3	96.2	96.7	97.2
	200	β_0	93.9	92.5	93.6	94.4	93.4	93.9
		β_1	95.4	94.4	95.3	95.1	94.4	96.0
		β_2	94.4	94.9	96.6	94.2	95.1	96.2

Table E.11: Bias, SE and CP of regression parameters for Case (iii) model with independent covariates ($\sigma_{x_1,x_2} = 0$)

	n	$\tau \rightarrow$	CQR 0.25	CQR 0.50	CQR 0.75	CQR-EL2 0.25	CQR-EL2 0.50	CQR-EL2 0.75
Bias	100	β_0.	0.0026	0.0108	0.0334	0.0102	0.0160	0.0389
		β_1.	0.0043	0.0027	0.0111	-0.0105	-0.0086	0.0027
		β_2.	-0.0010	-0.0014	-0.0086	0.0043	0.0027	-0.0041
	200	β_0	0.0095	0.0125	0.0232	0.0121	0.0145	0.0254
		β_1	-0.0007	0.0002	0.0020	-0.0081	-0.0061	-0.0032
		β_2	0.0011	0.0011	0.0009	0.0026	0.0029	0.0025
SE	100	β_0.	0.0594	0.0508	0.1093	0.0539	0.0473	0.1085
		β_1.	0.0668	0.0600	0.0964	0.0616	0.0563	0.0942
		β_2.	0.0663	0.0598	0.0996	0.0642	0.0595	0.1005
	200	β_0	0.0397	0.0329	0.0514	0.0364	0.0305	0.0501
		β_1	0.0429	0.0383	0.0567	0.0402	0.0360	0.0554
		β_2	0.0432	0.0389	0.0573	0.0420	0.0381	0.0574
CP	100	β_0.	94.1	94.0	96.9	94.2	95.4	97.9
		β_1.	96.4	96.7	97.5	94.8	94.9	97.4
		β_2.	96.3	97.0	96.5	95.3	96.1	97.0
	200	β_0	93.4	91.8	94.7	94.0	93.7	94.7
		β_1	95.8	96.5	95.5	94.2	95.1	95.6
		β_2	95.5	95.7	94.9	95.0	94.6	95.5

Table E.12: Bias, SE and CP of regression parameters for Case (iv) model with independent covariates ($\sigma_{x_1,x_2} = 0$)

	n	$\tau \rightarrow$	**CQR** 0.25	0.50	0.75	**CQR-EL2** 0.25	0.50	0.75
Bias	100	β_0.	0.0230	0.0372	0.0725	0.0142	0.0308	0.0690
		β_1.	-0.0045	0.0001	-0.0009	-0.0582	-0.0516	-0.0314
		β_2.	0.0039	0.0025	-0.0042	0.0072	0.0096	0.0033
	200	β_0	0.0233	0.0283	0.0488	0.0189	0.0245	0.0446
		β_1	0.0035	0.0028	-0.0033	-0.0311	-0.0319	-0.0276
		β_2	-0.0012	0.0002	0.0033	0.0010	0.0028	0.0071
SE	100	β_0.	0.1441	0.1407	0.2251	0.1274	0.1229	0.2126
		β_1.	0.1787	0.1800	0.2483	0.1740	0.1712	0.2421
		β_2.	0.1794	0.1807	0.2549	0.1801	0.1829	0.2563
	200	β_0	0.0976	0.0911	0.1269	0.0856	0.0787	0.1194
		β_1	0.1205	0.1176	0.1559	0.1153	0.1097	0.1499
		β_2	0.1223	0.1185	0.1562	0.1236	0.1192	0.1569
CP	100	β_0.	94.7	93.3	95.5	96.1	96.4	96.9
		β_1.	94.8	95.4	95.0	93.5	94.3	95.6
		β_2.	94.4	94.7	96.5	95.4	95.4	96.3
	200	β_0	91.9	92.0	92.8	94.8	94.4	94.6
		β_1	94.8	95.0	94.4	94.1	94.7	94.6
		β_2	93.6	94.5	95.3	94.8	95.1	95.2

Table E.13: Bias, SE and CP of regression parameters for Case (i) model with dependent covariates ($\sigma_{x_1,x_2} = 0.5$)

	n	$\tau \rightarrow$	CQR			CQR-EL2		
			0.25	0.50	0.75	0.25	0.50	0.75
Bias	100	β_0.	0.0150	0.0321	0.0935	0.0128	0.0369	0.0920
		β_1.	0.0060	0.0035	0.0019	-0.0707	-0.0568	-0.0356
		β_2.	-0.0092	-0.0037	0.0007	0.0008	0.0039	-0.0030
	200	β_0	0.0241	0.0268	0.0754	0.0267	0.0294	0.0758
		β_1	-0.0055	-0.0047	-0.0025	-0.0454	-0.0374	-0.0246
		β_2	-0.0000	0.0035	0.0067	0.0030	0.0081	0.0070
SE	100	β_0.	0.1830	0.1542	0.2937	0.1615	0.1375	0.2743
		β_1.	0.2248	0.1978	0.3138	0.2197	0.1893	0.2978
		β_2.	0.2280	0.1985	0.3218	0.2298	0.2005	0.3174
	200	β_0	0.1214	0.1010	0.1657	0.1055	0.0891	0.1563
		β_1	0.1493	0.1303	0.1957	0.1418	0.1229	0.1861
		β_2	0.1492	0.1317	0.1969	0.1509	0.1327	0.1960
CP	100	β_0.	94.4	93.2	96.3	95.4	94.9	97.4
		β_1.	96.0	95.5	96.2	95.7	95.5	95.6
		β_2.	94.9	95.2	96.0	94.8	95.1	96.1
	200	β_0	92.1	91.4	93.9	92.9	92.9	95.3
		β_1	95.8	95.3	95.3	94.9	94.8	96.1
		β_2	94.9	95.2	95.7	94.9	94.4	95.4

Table E.14: Bias, SE and CP of regression parameters for Case (ii) model with dependent covariates ($\sigma_{x_1,x_2} = 0.5$)

	n	$\tau \rightarrow$	CQR			CQR-EL2		
			0.25	0.50	0.75	0.25	0.50	0.75
Bias	100	β_0.	0.0043	0.0076	0.0218	0.0097	0.0108	0.0208
		β_1.	-0.0002	0.0035	0.0118	-0.0113	-0.0060	0.0038
		β_2.	0.0026	-0.0026	-0.0090	0.0075	0.0018	-0.0049
	200	β_0	0.0076	0.0104	0.0173	0.0094	0.0114	0.0180
		β_1	-0.0004	-0.0024	0.0010	-0.0056	-0.0072	-0.0019
		β_2	-0.0005	0.0034	-0.0008	0.0018	0.0048	0.0007
SE	100	β_0.	0.0456	0.0441	0.0774	0.0431	0.0421	0.0798
		β_1.	0.0601	0.0607	0.0876	0.0588	0.0597	0.0894
		β_2.	0.0606	0.0612	0.0884	0.0605	0.0623	0.0900
	200	β_0	0.0305	0.0290	0.0399	0.0288	0.0276	0.0398
		β_1	0.0400	0.0410	0.0545	0.0387	0.0396	0.0541
		β_2	0.0401	0.0413	0.0547	0.0396	0.0408	0.0552
CP	100	β_0.	95.1	95.0	97.7	94.4	95.1	98.4
		β_1.	96.6	96.6	96.5	96.2	96.2	96.9
		β_2.	95.8	95.3	96.7	95.7	95.9	96.7
	200	β_0	91.6	91.8	94.1	93.7	93.2	94.6
		β_1	95.4	95.8	95.7	94.4	94.9	95.8
		β_2	94.6	94.7	94.4	94.9	94.0	94.4

Table E.15: Bias, SE and CP of regression parameters for Case (iii) model with dependent covariates ($\sigma_{x_1,x_2} = 0.5$)

	n	$\tau \rightarrow$	CQR 0.25	CQR 0.50	CQR 0.75	CQR-EL2 0.25	CQR-EL2 0.50	CQR-EL2 0.75
Bias	100	β_0.	0.0042	0.0110	0.0382	0.0098	0.0147	0.0391
		β_1.	0.0016	0.0041	0.0109	-0.0150	-0.0080	0.0028
		β_2.	-0.0002	-0.0032	-0.0119	0.0049	0.0016	-0.0110
	200	β_0	0.0083	0.0100	0.0244	0.0094	0.0102	0.0234
		β_1	-0.0020	0.0017	0.0031	-0.0106	-0.0057	-0.0015
		β_2	0.0017	0.0000	-0.0030	0.0031	0.0019	-0.0014
SE	100	β_0.	0.0595	0.0498	0.1099	0.0541	0.0471	0.1067
		β_1.	0.0735	0.0663	0.1134	0.0717	0.0655	0.1109
		β_2.	0.0747	0.0668	0.1147	0.0734	0.0672	0.1164
	200	β_0	0.0383	0.0319	0.0517	0.0353	0.0299	0.0507
		β_1	0.0471	0.0426	0.0654	0.0454	0.0413	0.0638
		β_2	0.0475	0.0424	0.0643	0.0466	0.0421	0.0643
CP	100	β_0.	95.4	95.0	97.2	94.5	96.1	97.6
		β_1.	95.7	96.4	96.9	95.1	96.3	96.6
		β_2.	96.0	96.3	96.9	95.7	95.9	96.7
	200	β_0	93.6	93.0	94.3	94.5	94.5	95.7
		β_1	95.7	95.3	95.4	94.1	95.4	94.7
		β_2	96.1	95.9	95.3	96.0	96.4	95.0

Table E.16: Bias, SE and CP of regression parameters for Case (iv) model with dependent covariates ($\sigma_{x_1,x_2} = 0.5$)