

A Single Index Model for Censored Quantile Regression

Miao Lu

M.S., University of Virginia, United States, 2014

B.S., Zhejiang University, P.R.China, 2012

A Dissertation Presented to the Graduate Faculty
of University of Virginia in Candidacy for the Degree of
Doctor of Philosophy

Department of Statistics

University of Virginia

November, 2016

Abstract

Quantile regression, which is a valuable alternative to the commonly used Cox proportional hazards model and accelerated failure time (AFT) model in survival analysis, has been getting more attention recently due to its robustness and interpretability. By allowing nonlinear relationship between survival time and risk factors, we study a single index model for censored quantile regression, and employ the B-spline approximation for estimation. To account for censoring, we consider the redistribution-of-mass to obtain a weighted quantile regression estimator. In addition, dimension reduction approach is adopted to deal with the “curse of dimensionality”. Furthermore, we penalize the developed estimator for variable selection purpose. The proposed methods are easy to implement using the existing weighted linear quantile regression algorithm compared to available methods, and can be generalized to multiple index models. The asymptotic properties of the developed estimator are investigated and the estimator’s numerical performance is illustrated in simulation studies. We also apply the proposed methods to Boston housing data and kidney transplant study.

Acknowledgements

Getting a Ph.D. degree is not something that can be done alone. I would like to thank everyone who has guided, helped and supported me along the way.

Firstly, I want to give my deepest thanks to Professor Jianhui Zhou, my advisor, for his enduring guidance and advice. Throughout my years in the Ph.D. program, I was impressed by his immense knowledge and benefited tremendously from discussion with him. As an advisor, he is able to show me the overall picture of my research, as well as suggest every possible approach I can work on in details. He is also a good mentor with full life experience, and gave me a lot of advices on both study and life. My life at UVa would not be so smooth without his precious support and help.

I am also especially grateful to Professor Jennie Z. Ma, who supervised me during my research assistantship in PROVIDE study at school of medicine, UVa. She is very approachable and encourages me to seize every opportunity to tackle clinical problem with statistical methods. I have learned a lot from her experience and expertise, and the opportunities she provided to me built my confidence and ability. My gratitude to her is beyond words.

I also would like to acknowledge my thesis committee members (Prof. Jianhui Zhou, Prof. Jennie Z. Ma, Prof. Karen Kafadar, and Prof. Chao Du) for their advice and time devotion. Their support and guidance keep me going forward in the best direction. I also want to thank Prof. Dan Spitzner, Prof. Tingting Zhang, Prof. Jeff Holt, Prof. Daniel Keenan. Their teachings rich my understanding of statistics and broaden my view.

I am grateful to Dr. William A. Petri, Jr. and every member in the PROVIDE study. It has been an honor for me to work on this wonderful project with amazing

people. Thanks to all my colleagues, friends and classmates, and my days would be less enjoyable without their vitality and enthusiasm.

Finally, thanks to my girlfriend, Chenyi Pan and my parents, Xiajuan Wang and Junjiang Lu. Their unreserved love leads me from “a kid in a candy shop” to a man full of confidence and courage.

Contents

Abstract	ii
Acknowledgements	iii
1 Introduction	1
1.1 Motivation	1
1.2 Overview of Variable Selection	3
1.3 Overview of Single Index Model in Quantile Regression	5
1.4 Introduction to Censored Quantile Regression	7
2 Model Setup and Estimation	10
2.1 Model Setup	10
2.2 Single Index Model for Censored Quantile Regression	11
2.2.1 Estimation of Re-distribution of Mass	11
2.2.2 Dimension Reduction Approach	13
2.2.3 Estimation of the Unknown Link Function	14
2.3 Single Index Model with Variable Selection	15
2.4 Computation and Tuning	16
2.4.1 The Algorithm	16
2.4.2 Initial Value and Tuning Parameters	17

3	Asymptotics	19
3.1	Regularity Conditions	19
3.2	Asymptotic Properties	21
4	Simulation	24
4.1	Model 1: Location Model	25
4.2	Model 2: Sine-bump Model	26
4.3	Model 3: Varying Coefficient Model	31
5	Real Data Application	36
5.1	Boston Housing Data	36
5.2	Kidney Transplant Study	38
6	Conclusion and Discussion	45
7	Technical Proofs	49
7.1	Proof of Theorems	49
7.2	Lemmas	58
	Bibliography	84

Chapter 1

Introduction

1.1 Motivation

Kidney transplantation is a surgical procedure used to treat patients with end-stage renal disease. Successful kidney transplantation is associated with prolonged survival, improved quality of life and healthcare cost savings when compared to dialysis. There are two types of kidney donors, living donors and deceased donors. In 2014, 17107 kidney transplants took place in the US. Among them, 11570 came from deceased donors and 5537 came from living donors (Organ Procurement and Transplantation Network). However, the deficit between the number of the patients on the waiting list for kidney transplantation and the availability of donor organs is increasing in United States. By December 31, 2013, the kidney transplant waiting list had 86965 candidates (dialysis patients only), while there were only 17600 kidney transplants performed in 2013 (United States Renal Data System 2015 Annual Data Report). The median waiting time for a patient's first kidney transplant is 3.6 years and could be longer depending on health, compatibility and availability of organs. To combat this problem, the innovative use of the expanded criteria donor (ECD)

kidneys was introduced (Port et al., 2002). Deceased donor kidneys are classified as ECD if donors aged ≥ 60 years, or those aged between 50 and 59 years who meet at least two of the criteria: serum creatinine > 1.5 mg/dL, death due to cerebrovascular accident, or history of hypertension. In reality, some kidneys labelled as Standard Criteria Donor (SCD) may have a reduced graft survival, while some ECD kidneys might perform well. Therefore, instead of the simple dichotomized ECD/SCD criteria, quantitative models are needed to give a more precise evaluation of the donor kidney quality. An improvement over the binary ECD/SCD classification is the scoring system developed by Nyberg et al. (2003). This system included five variables: donor age, history of hypertension, creatinine clearance, cause of death and the number of HLA-mismatches. Cadaver kidneys were stratified by cumulative donor score: grade A, 0-9 points; grade B, 10-19; grade C, 20-29; and grade D, 30-39. A higher score and grade reflected poorer organ quality. However, this scoring system is not widely used in practice since it was not replicable in independent cohort (Lee and Abramowicz, 2014).

As stated in Nyberg et al. (2005), a more precise index, which takes advantage of the increasingly documented medical records including donor, recipient and transplant information, is of great value. Rao et al. (2009) proposed the Kidney Donor Risk Index (KDRI), which is a graft failure risk score that could capture donor and transplant characteristics. It consists of a multivariable Cox proportional hazards model, combined with stepwise selection. The model avoids categorization and is based on the association between ten donor/recipient characteristics and the graft survival. The KDRI is an easily applicable scoring system which provides the clinicians with a guide to objectively assess the quality of the donor kidneys. However, KDRI has some challenges. First, it does not allow heteroscedastic effect of variables for higher/lower risk (shorter/longer survival time) patients. Secondly,

the association between the graft survival and risk factors might be unknown and non-linear. Thirdly, the essential problems for traditional stepwise selection such as high bias, admirably summarized by Harrell (2001), remain for KDRI. Hence, new models and methods of creating a new index, finding informative risk factors, are in need to better match donor and recipient for improving survival time after kidney transplantation. The method developed in this thesis could take into account of the heteroscedastic effect of variables, the nonlinearity between survival and risk factors, and variable selections. Moreover, this method can be used to estimate different quantiles of the survival times and could provide a more complete picture of transplantation for physicians.

1.2 Overview of Variable Selection

High dimensional data analysis is important and has been widely used in biomedical studies. For example, tens of thousands of molecular expressions are potential predictors in microarray data; hundreds of thousands of single nucleotide polymorphisms (SNPs) are possibly associated with the clinical outcome of interest in genome-wide association study (Fan and Lv, 2010). To deal with a large number of covariates, a common biomedical approach is testing the association between each covariate and the outcome of interest through a regression model; selecting a subset of those covariates based on their significance, and using the selected covariates in a multivariable regression model. This commonly used variable selection method in biomedicine would be in great challenge when the number of covariates selected is high for large datasets. Prediction accuracy and interpretation are two main drawbacks for the above traditional regression analysis (Hastie et al., 2005). Another widely used method in variable selection is stepwise regression methods. However,

some essential problems with those methods remain and have been summarized in Harrell (2001). First, the parameter estimation tends to have higher bias. Secondly, p -values tend to be too low, due to multiple comparisons, and are difficult to correct.

Penalized regression methods, such as LASSO (Least Absolute Shrinkage and Selection Operator) in Tibshirani (1996) and SCAD (Smoothly Clipped Absolute Deviation) in Fan and Li (2001) have been developed to overcome the limitation of traditional stepwise variable selection models when the number of covariates is large in a linear regression model,

$$Y = X\beta + \epsilon, \tag{1.1}$$

where Y is an $n \times 1$ vector and X is an $n \times p$ matrix. Linear regression is a widely used method to study association between continuous outcome and covariates.

Penalized regression methods are able to accomplish variable selection and estimation simultaneously for model (1.1). The regression parameters are assumed to be sparse with some components being zero, and the nonzero components indicate which variables are informative. Relative to the maximum likelihood estimates, penalized methods shrink the estimates of regression coefficients towards zero.

A general form of the penalized least squares is

$$\frac{1}{2n} \|Y - X\beta\|^2 + \sum_{j=1}^p p_j(|\beta_j|), \tag{1.2}$$

where $p_j(|\cdot|)$ is a penalty function that can take different forms. Two examples are the L_2 penalty $p_j(|\theta|) = \lambda|\theta|^2$ resulting in ridge regression, and the LASSO has the L_1 penalty $p_j(|\theta|) = \lambda|\theta|$. A linear combination of the L_1 and L_2 penalties is the Elastic Net, which enjoys a similar sparsity representation as LASSO, while encouraging a grouping effect (Zou and Hastie, 2005). The continuous differentiable

penalty function defined by the derivative

$$p'_\lambda(\theta) = \lambda\{I(\theta \leq \lambda) + \frac{(a\lambda - \theta)_+}{(a-1)\lambda}I(\theta > \lambda)\},$$

for some $a > 2$ and $\theta > 0$, is the SCAD penalty, where $p_\lambda(0) = 0$ and $a \approx 3.7$ as suggested by Bayesian risk analysis. The Minimax Concave Penalty (MCP, Zhang 2010) is defined as

$$p'_\lambda(\theta) = \frac{(a\lambda - \theta)_+}{a},$$

which translates the flat part of the derivative of SCAD to the origin. Zou (2006) proposed a weighting scheme $w_j = |\tilde{\beta}_j|^{-\gamma}$ in the penalty function of LASSO, leading to the adaptive LASSO. The above variable selection methods for linear models can be extended to the single index models for quantile regression, which are described in the next section.

1.3 Overview of Single Index Model in Quantile Regression

Quantile regression has become a unified statistical methodology for estimating models of conditional quantile functions. By complementing the exclusive focus of least squares regression on the conditional mean, quantile regression offers an alternative strategy to examine how covariates can influence the location, scale and shape of the entire response distribution (Koenker, 2005). Quantile regression is also a valuable alternative to the commonly used Cox proportional hazards model and accelerated failure time (AFT) model used in survival analysis (Koenker and Geling, 2001). For example, in a cancer study, it is known that treatment will cause different impacts among patients in lower or upper quantiles of the survival distribution. Physicians

are interested in identifying treatment for more severe cases, i.e, lower quantiles, while Cox and AFT models estimate covariate effects on the location only. In contrast, quantile regression directly links the quantiles of patients' survival time to their demographic and genomic profiles, allowing the covariates to exhibit different impacts at different tails of the survival distribution.

The seminal work of Koenker and Bassett (1978) proposed the “check” loss function to estimate the linear quantile regression model. Since then, a variety of nonparametric methods, such as kernel based methods and spline smoothing methods, have been applied to accommodate nonlinear relationships; see for example, Yu and Jones (1998), He and Shi (1994) and Koenker et al. (1994). However, the nonparametric quantile regression with multivariable covariates is a difficult estimation problem due to the “curse of dimensionality”. Another problem of nonparametric quantile regression is that the estimated function can be difficult to visualize and interpret with multivariable covariates. To reduce the dimensionality while retaining much flexibility of a nonparametric model, a semiparametric single index model is considered in this thesis. Single index models assume that the response of interest depends on a linear combination of covariates through an unknown link function. Many widely used parametric models such as linear model can be considered as special cases of single index models. Single index models which are well motivated in both econometrics and statistics, are applied to many objective functions, including conditional mean and quantiles. For estimating the conditional mean in a single index model, many methods exist, including the backfitting algorithm (Carroll et al., 1997), the minimum average variance estimation (MAVE, Xia et al., 2002), the penalized spline estimation (Yu and Ruppert, 2002), and the profile least squares estimation (Liang et al., 2010). Moreover, Wu et al. (2010) proposed a modified version of MAVE to estimate the conditional quantile in a single index model, and

Kong and Xia (2012) modified the algorithm by adding a penalty term. Later, Alkernani and Yu (2013) and Lv et al. (2014) proposed the penalized quantile regression in a single index model. Instead of the traditional kernel based methods, Zou and Zhu (2014) proposed M-estimators for single index model using the spline based approach. Recently, Ma and He (2016) proposed the profile estimation with B-spline for both quantile estimation and inference.

1.4 Introduction to Censored Quantile Regression

Censoring happens when patients withdraw from the study, or patients haven't experienced the event when the study ends. That is to say, instead of observing the survival time T_i , we only observe the minimum of T_i and the censoring time C_i , i.e., $Y_i = \min(T_i, C_i)$. Censored quantile regression was first studied in Powell (1984) and Powell (1986) for fixed censoring, where the censoring variables C_i are known to all observations, even for those uncensored. However, censoring times are not always known in most survival analysis. In this thesis, we will focus on random censoring.

Several articles deal with quantile regression for randomly censored survival outcomes. Ying et al. (1995) proposed a semiparametric estimation procedure for a censored median regression model, assuming T_i and C_i are unconditionally independent. Assuming that C_i is independent of both x_i and T_i , Honore et al. (2002) extended Powell's approach to random censoring. Portnoy (2003) developed a novel "recursive reweighting" scheme that generalizes the Kaplan-Meier estimator, with a more relaxed assumption that T_i and C_i are independent given x_i . Peng and Huang (2008) proposed a martingale-based estimating procedure, which showed similar performance to Portnoy (2003). Both approaches in Portnoy (2003) and Peng and

Huang (2008) rely on strong global assumption, that the conditional functionals at lower quantiles are all linear when estimating the model at a given quantile level. Wang and Wang (2009) adopted a similar redistribution of mass idea for linear quantile regression, but estimated the censoring probability nonparametrically using the local Kaplan-Meier method.

To the best of our knowledge, the literature on single index model for censored quantile regression is very limited. Two unpublished manuscripts, Bücher et al. (2014) and Christou and Akritas (2016), use similar inverse-probability-weighting scheme to handle censored observations, but estimate the model using local linear smoothing and nonlinear optimization, respectively.

In this thesis, we propose the estimation of single index model for the conditional quantile function when the data is right censored. Comparing to current quantile regression for censored outcomes, our approach relaxes the strong linear assumption to the non-linear assumption with an unknown link function. As in Wang et al. (2013), we also relax the stringent assumption that T_i and C_i are unconditionally independent in Bücher et al. (2014). Our method adopts the redistribution of mass idea of Efron (1967) to account for the censoring in quantile regression. Besides, the B-spline approach is taken to estimate the link function nonparametrically, and a weighted linear quantile regression is used to estimate the parametric index. What's more, in the presence of multiple covariates, penalized methods such as LASSO (Tibshirani, 1996) and SCAD (Fan and Li, 2001) can be imposed to the developed estimator to shrink some coefficients of the single index components to zeros for selecting informative variables. Simulation studies are conducted to examine the performance of proposed method and to compare it with existing methods. The numerical outcomes show improved performance of our proposed approach across different quantile levels and censoring percentages when the number of covariates is

relatively large.

Compared to existing methods, our approach enjoys several distinctive advantages. First, our method relaxes the stringent assumption that survival time and censoring time are unconditionally independent in Bücher et al. (2014). Second, our algorithm with B-spline approximation is computationally simple and easy to implement, while Bücher et al. (2014) and Christou and Akritas (2016) estimate the model with local linear approximation and nonlinear optimization. Finally, our approach can achieve sparsity while both existing manuscripts only do estimation without variable selection.

In Section 2, the proposed model is presented and unpenalized and penalized estimators are developed, along with algorithm and computation. In Section 3, the asymptotic properties of the proposed estimators are provided. Section 4 investigates the finite sample performance of proposed methods through simulation studies, and compares the results with existing methods. Boston housing data and kidney transplant study are analyzed in Section 5 using the developed methods. The thesis concludes with some discussion of future directions in Section 6. All technical proofs are deferred to Chapter 7.

Chapter 2

Model Setup and Estimation

2.1 Model Setup

Denote T_i as an uncensored survival outcome, x_i as an observable p -dimensional covariates, and C_i as a censoring variable for the i th subject with $i = 1, \dots, n$. In the proposed single index model, the τ th conditional quantile of T_i given x_i , $Q_\tau(T_i|x_i) = \inf\{t : P(T_i \leq t|x_i) \geq \tau\}$, is expressed as

$$Q_\tau(T_i|x_i) = g_\tau(x_i^T \beta_\tau), \quad (2.1)$$

where $g_\tau(\cdot)$ is an unknown smooth link function, β_τ is a p -dimensional unknown quantile coefficient vector on the unit sphere with $\|\beta_\tau\| = 1$, and the first nonzero element of β_τ is assumed to be positive for identifiability (Lin and Kulasekera, 2007). Without loss of generality, assuming $\beta_{\tau 1} \neq 0$, the parameter space is

$$\Theta = \{\beta_\tau = (\beta_{\tau 1}, \dots, \beta_{\tau p}) : \|\beta_\tau\| = 1, \beta_{\tau 1} > 0\}.$$

In practice, we only observe (x_i, Y_i, η_i) , where $Y_i = \min(T_i, C_i)$ is the observed censored response and $\eta_i = I(T_i \leq C_i)$ is the censoring indicator. When $\eta_i = 1$ for all i , $Y_i = T_i$ are observed, and β_τ is estimated by solving the following optimization problem.

$$\hat{\beta} = \arg \min_{g \in \mathcal{H}, \beta \in \Theta} E [\rho_\tau (Y - g(X^T \beta))]$$

where $\rho_\tau(\mu) = \mu\{\tau - I(\mu < 0)\}$ is the ‘‘check’’ loss function (We omit τ in β_τ and $g_\tau(\cdot)$ for notational convenience).

Suppose $\{(x_i, Y_i), i = 1, 2, \dots, n\}$ is the sample, the minimization is as follows,

$$\hat{\beta} = \arg \min_{g \in \mathcal{H}, \beta \in \Theta} \sum_{i=1}^n \rho_\tau (Y_i - g(x_i^T \beta)) .$$

2.2 Single Index Model for Censored Quantile Regression

2.2.1 Estimation of Re-distribution of Mass

In the censored quantile regression setting, the idea is to redistribute the probability mass $P(T_i > C_i | C_i, x_i)$ of the censored cases to the right. Let $F_0(t|x) = P(T < t|x)$ denote the cumulative distribution function (cdf) of T given x . Define $\pi_{0i} = F_0(C_i|x_i)$ as the conditional probability for the i th subject not being censored. The fundamental insight of redistribution of mass is that the derivative of $\rho_\tau(T_i - g(x_i^T \beta))$ depends only on the signs of the residuals $T_i - g(x_i^T \beta)$, and the mass at C_i for some censored observations may be redistributed to any point above the data, for example, $(x_i, Y^{+\infty})$ without altering the quantile fit. The coefficient β can be estimated by

minimizing the following objective function,

$$L(\beta, g, w) = \sum_{i=1}^n \{w_i \rho_\tau(Y_i - g(x_i^T \beta)) + (1 - w_i) \rho_\tau(Y^{+\infty} - g(x_i^T \beta))\}, \quad (2.2)$$

where $Y^{+\infty}$ is sufficiently large to exceed $g(x_i^T \beta)$ for all i , and

$$w_i = \begin{cases} 1 & \eta_i = 1 \\ 0 & \eta_i = 0 \text{ and } F_0(C_i|x_i) > \tau \\ \frac{\tau - F_0(C_i|x_i)}{1 - F_0(C_i|x_i)} & \eta_i = 0 \text{ and } F_0(C_i|x_i) \leq \tau \end{cases} \quad (2.3)$$

In (2.2), the weights w_i are used to remove the estimation bias due to censoring. Note that when $C = +\infty$ (non-censored case), $F_0(C|x_i) = 1$, so $w_i = 1$ for all i . Similar assignment of the above weighting scheme can be found in Wang and Wang (2009), Portnoy and Lin (2010) and Wang et al. (2013). Estimation of quantile index at any τ depends on T_i only through the signs of the residuals, i.e. $I\{[T_i - g(x_i^T \beta)] < 0\}$. Accordingly, the weights w_i in (2.3) are assigned in the following three scenarios.

- When $\eta_i = 1$, $Y_i = T_i < C_i$ and $I\{[T_i - g(x_i^T \beta)] < 0\}$ is observed. Thus, the observation (x_i, Y_i) is assigned weight 1, and the paired “pseudo” observation $(x_i, Y^{+\infty})$ is not used.
- When $\eta_i = 0$, but $Y_i = C_i > g(x_i^T \beta)$, we immediately know $T_i - g(x_i^T \beta) > 0$, and letting $T_i = Y^{+\infty}$ gives the same estimate. Thus, weight 0 is assigned to (x_i, Y_i) , and 1 to $(x_i, Y^{+\infty})$.
- When $\eta_i = 0$, but $Y_i = C_i < g(x_i^T \beta)$, $T_i - g(x_i^T \beta)$ can be positive or negative.

However,

$$E \{ I\{[T_i - g(x_i^T \beta)] < 0\} | T_i > C_i, C_i, x_i \} = \frac{\tau - F_0(C_i | x_i)}{1 - F_0(C_i | x_i)}.$$

Thus, weight $\frac{\tau - F_0(C_i | x_i)}{1 - F_0(C_i | x_i)}$ is assigned to (x_i, Y_i) , and $1 - \frac{\tau - F_0(C_i | x_i)}{1 - F_0(C_i | x_i)}$ to $(x_i, Y^{+\infty})$.

In practice, the masses for redistribution, $1 - \pi_{0i} = 1 - F_0(C_i | x_i)$ are unknown. As in Wang and Wang (2009), we propose to use local Kaplan-Meier estimator when the dimension of x_i is low,

$$\hat{F}(t|x) = 1 - \prod_{j=1}^n \left\{ 1 - \frac{B_{nj}(x)}{\sum_{k=1}^n I(Y_k \geq Y_j) B_{nk}(x)} \right\}^{\eta_j(t)}, \quad (2.4)$$

where $\eta_j(t) = I(Y_j \leq t, \eta_j = 1)$, $B_{nk}(x) = \frac{K_q(\frac{x-x_k}{h})}{\sum_{i=1}^n K_q(\frac{x-x_i}{h})}$. In this thesis, we adopt the commonly used product kernel function $K_q(u_1, \dots, u_q) = \prod_{i=1}^q K(u_i)$, where $K(\cdot)$ is a univariate kernel function with bandwidth h .

2.2.2 Dimension Reduction Approach

The above estimator in (2.4) is only feasible when the covariates dimension p is small, due to the ‘‘curse of dimensionality’’. For large p , it can be estimated after dimension reduction as in Wang et al. (2013). Dimension reduction estimates a set of directions γ_i such that

$$T_i \perp\!\!\!\perp x_i | (x_i^T \gamma_1, \dots, x_i^T \gamma_q). \quad (2.5)$$

Censored sliced inverse regression (cSIR) in Li et al. (1999) and hazard minimum average variance estimation (hMAVE) in Xia et al. (2010) can estimate those directions with $q \ll p$ in general. With (2.5), high-dimensional x_i is replaced by its

low-dimensional projection $z_i = (z_{i1}, \dots, z_{iq})^T$ with $z_{ij} = x_i^T \gamma_j$ without losing regression information for T_i . The estimator (2.4) can then be applied to z_i to estimate w_i , given that $F_0(t|x_i) = F_0(t|z_i)$ implied by (2.5).

With the estimated weight, the objective function changes accordingly to

$$L(\beta, g, \hat{w}) = \sum_{i=1}^n \left\{ \hat{w}_i \rho_\tau(Y_i - g(x_i^T \beta)) + (1 - \hat{w}_i) \rho_\tau(Y^{+\infty} - g(x_i^T \beta)) \right\}, \quad (2.6)$$

where \hat{w}_i are the estimated redistribution of mass weights by replacing F_0 with \hat{F} in w .

2.2.3 Estimation of the Unknown Link Function

The unknown link function $g(\cdot)$ is estimated by local linear approximation in Wu et al. (2010), and Kong and Xia (2012). As in Zou and Zhu (2014), and Ma and He (2016), we adopt the B-splines approximation $g(x^T \beta) \approx B(x^T \beta) \theta$ due to faster computation. Here, θ is the B-spline coefficients. Given the spline order r and a set of quasi-uniform knots $a = t_0 < t_1 < \dots < t_s < t_{s+1} = b$, satisfying

$$\max_{0 \leq j \leq s} |t_{j+1} - t_j| / \min_{0 \leq j \leq s} |t_{j+1} - t_j| \leq M$$

for some constant $0 < M < \infty$, where s is the number of interior knots, the normalized B-spline basis functions are denoted as $B(u) = \{B_j(u)\}_{j=1}^{J_n}$ with $J_n = s + r$.

From the B-spline approximation, the objective function (2.4) can be written as

$$L(\beta, \theta, \hat{w}) = \sum_{i=1}^n \left\{ \hat{w}_i \rho_\tau(Y_i - B(x_i^T \beta) \theta) + (1 - \hat{w}_i) \rho_\tau(Y^{+\infty} - B(x_i^T \beta) \theta) \right\}. \quad (2.7)$$

The estimator of $g(u)$ at any point u is $B(u)\hat{\theta}$, where $\hat{\theta}$ is obtained by minimizing (2.7).

2.3 Single Index Model with Variable Selection

When we apply the single index model in real data analysis, many irrelevant variables may be included in the model, making interpretation and prediction less favorable. As needed, variable selection method is adopted to select informative variables. Here, we propose the penalized estimator $\hat{\beta}$ by minimizing the following adaptive LASSO (Zou, 2006) type penalized objective function

$$\begin{aligned} L_{AL}(\beta, \theta, \hat{w}) &= L(\beta, \theta, \hat{w}) + \lambda \sum_{j=1}^p \frac{|\beta_j|}{|\tilde{\beta}_j|^\delta} \\ &= \sum_{i=1}^n \left\{ \hat{w}_i \rho_\tau(Y_i - B(x_i^T \beta) \theta) + (1 - \hat{w}_i) \rho_\tau(Y^{+\infty} - B(x_i^T \beta) \theta) \right\} \\ &\quad + \lambda \sum_{j=1}^p \frac{|\beta_j|}{|\tilde{\beta}_j|^\delta}, \end{aligned} \quad (2.8)$$

where $\tilde{\beta}_j$ is the j th element of the initial consistent estimator. Other forms of penalty functions such as LASSO (Tibshirani, 1996) and SCAD (Fan and Li, 2001) can also be applied here. In our implementation, we set $\tilde{\beta}$ as the unpenalized estimator and $\delta = 2$.

2.4 Computation and Tuning

2.4.1 The Algorithm

Given the original observed data $\{(x_i, \eta_i, Y_i), i = 1, \dots, n\}$, we define n pseudo paired observations $\{(x_1, \eta_1, Y^{+\infty}), \dots, (x_n, \eta_n, Y^{+\infty})\}$. After the calculation of weights \hat{w}_i , we assign local weights \hat{w}_i to original data, while assign $1 - \hat{w}_i$ to the pseudo data. In the implementation, we set $Y^{+\infty} = 1000 \max\{|Y_1|, \dots, |Y_n|\}$. The penalized objective function can be optimized by the following iteration between $\tilde{\beta}^{(l)}$ and $\hat{\theta}^{(l)}$.

Step 1, Given $\tilde{\beta}^{(l)}$, obtain $\hat{\theta}^{(l)}$ by

$$\min_{\theta} \sum_{i=1}^n \left\{ \hat{w}_i \rho_{\tau} \left(Y_i - B(x_i^T \tilde{\beta}^{(l)}) \theta \right) + (1 - \hat{w}_i) \rho_{\tau} \left(Y^{+\infty} - B(x_i^T \tilde{\beta}^{(l)}) \theta \right) \right\},$$

Step 2, Given $\hat{\theta}^{(l)}$, $g(x_i^T \beta)$ can be approximated by $g(x_i^T \beta) \approx g(x_i^T \tilde{\beta}^{(l)}) + g'(x_i^T \tilde{\beta}^{(l)}) x_i^T (\beta - \tilde{\beta}^{(l)}) = B(x_i^T \tilde{\beta}^{(l)}) \hat{\theta}^{(l)} + B'(x_i^T \tilde{\beta}^{(l)}) \hat{\theta}^{(l)} x_i^T (\beta - \tilde{\beta}^{(l)})$. Obtain $\hat{\beta}^{(l+1)}$ by

$$\min_{\beta} \sum_{i=1}^n \left\{ \hat{w}_i \rho_{\tau} \left(Y_i - B(x_i^T \tilde{\beta}^{(l)}) \hat{\theta}^{(l)} - B'(x_i^T \tilde{\beta}^{(l)}) \hat{\theta}^{(l)} x_i^T (\beta - \tilde{\beta}^{(l)}) \right) + \right. \\ \left. (1 - \hat{w}_i) \rho_{\tau} \left(Y^{+\infty} - B(x_i^T \tilde{\beta}^{(l)}) \hat{\theta}^{(l)} - B'(x_i^T \tilde{\beta}^{(l)}) \hat{\theta}^{(l)} x_i^T (\beta - \tilde{\beta}^{(l)}) \right) \right\}, \quad (2.9)$$

where $B'(\cdot)$ contains the derivatives of the B-spline basis functions in $B(\cdot)$.

Repeat **Step 1** and **Step 2** until convergence, output $\tilde{\beta}, \hat{\theta}$.

Step 3, obtain the penalized estimator $\hat{\beta}$ by solving

$$\begin{aligned} \min_{\beta} \sum_{i=1}^n \left\{ \hat{w}_i \rho_{\tau} \left(Y_i - B(x_i^T \tilde{\beta}) \hat{\theta} - B'(x_i^T \tilde{\beta}) \hat{\theta} x_i^T (\beta - \tilde{\beta}) \right) + \right. \\ \left. (1 - \hat{w}_i) \rho_{\tau} \left(Y^{+\infty} - B(x_i^T \tilde{\beta}) \hat{\theta} - B'(x_i^T \tilde{\beta}) \hat{\theta} x_i^T (\beta - \tilde{\beta}) \right) \right\} \\ + \lambda \sum_{j=1}^p \frac{|\beta_j|}{|\tilde{\beta}_j|^{\delta}}, \end{aligned} \quad (2.10)$$

Finally, we estimate $g(u)$ with $B(u)\hat{\theta}$.

Each step in the proposed algorithm can be solved using existing functions for linear quantile regression in R package; this is advantageous to Bücher et al. (2014) and Christou and Akritas (2016), which uses time consuming local linear approximation or complicated nonlinear quantile regression algorithms. Compared with local linear approximation in Kong and Xia (2012), Wu et al. (2010) and Bücher et al. (2014), the proposed algorithm using B-spline is more than 20 times faster in computing speed in the numerical studies. Another advantage of the proposed method is to use existing penalized linear quantile estimator to select informative variables in **Step 3** for single index quantile regression model, while formulation of the problem in Bücher et al. (2014) and Christou and Akritas (2016) does not facilitate variable selection for quantiles.

2.4.2 Initial Value and Tuning Parameters

There are two main practical issues in the computation; one is the choice of initial value, another is the selection of tuning parameters. Wu et al. (2010) used average derivative estimator (ADE) of Chaudhuri et al. (1997) for uncensored data as initial value. In our numerical experience, the estimation is not sensitive to the choice of

initial values. Unless specified for comparison, we used $(1, \dots, 1) \in R^p$ as initial value.

There are two types of tuning parameters: one is bandwidth h and number of internal knots s , another is λ in the adaptive LASSO penalty. The B-spline order r is usually fixed at 3 or 4 for quadratic or cubic splines. Our numerical experience suggests that the performance of the proposed procedure is not sensitive to the choice of h and s in some reasonable range. Similarly as in Wang and Wang (2009) and Wang et al. (2013), we propose the K-fold cross validation criteria to select the optimal pair of (h, s) . In the cross validation, the data is divided into K parts randomly with roughly equal sample size. For the k th part, $k = 1, \dots, K$, we estimate $\tilde{\beta}, \hat{\theta}$ based on the rest of K-1 parts of the data, and evaluate the quantile loss $CV_k = \rho_\tau(Y_v - B(x_v^T \tilde{\beta}) \hat{\theta})$ on uncensored data in the k-th part. Here, (x_v, Y_v) are the uncensored data in the k-th part. Denoting $CV(h, s)$ as the mean of CV_1, \dots, CV_K , we select the optimal pair of (h, s) to minimize $CV(h, s)$.

After selecting (h, s) by cross validation in the unpenalized objective function, we use the commonly used Bayesian information criterion (BIC) type method in single index model to choose the optimal λ (Wang et al., 2007; Lv et al., 2014; Ma and He, 2016). The BIC-type criterion is defined as

$$BIC(\lambda) = \log \left(\frac{1}{\sum \eta_i} \sum_{i:\eta_i=1} \rho_\tau \left(y_i - B(x_i^T \hat{\beta}) \hat{\theta} \right) \right) + \frac{\log(\sum \eta_i)}{\sum \eta_i} df,$$

where $\hat{\beta}$ is the penalized estimation using λ , and df is the number of non-zero elements in $\hat{\beta}$. The selected λ minimizes the above BIC-type criterion. In our implementation, we use the 5-fold cross validation, B-spline order 3, with fourth order kernel function (Müller, 1984) $K(x) = \frac{105}{64}(1 - 5x^2 + 7x^4 - 3x^6)I(|x| \leq 1)$ in simulation study and data analysis.

Chapter 3

Asymptotics

3.1 Regularity Conditions

Before stating the main theoretical results, we define the following functions,

(i) $1 - F_0(t|x) = P(T_i > t|x)$, the survival function of T_i conditional on x .

(ii) $1 - M_0(t|x) = P(C_i > t|x)$, the survival function of C_i conditional on x .

Denote \mathcal{H}_ϑ as the collection of functions on $[a, b]$ such that the r^{th} order derivative satisfies the Holder condition of order ζ with $\vartheta = r + \zeta$. In other words, there exists a constant C_0 such that

$$|\phi^{(r)}(u_1) - \phi^{(r)}(u_2)| \leq C_0 |u_1 - u_2|^\zeta,$$

for any $\phi \in \mathcal{H}_\vartheta$, and any $a \leq u_1, u_2 \leq b$. Let $\tilde{g}(u, \beta)$ be the τ -th quantile function of T given $x^T \beta = u$. Then we have $g(x^T \beta^0) = \tilde{g}(x^T \beta^0, \beta^0)$ with the true coefficient β^0 . Also, we have $\Theta = \{\beta = (\beta_1, \dots, \beta_p)^T : \|\beta\|_2 = 1, \beta_1 > 0\}$.

To establish the consistent and asymptotic results for the developed estimators, we assume the following assumptions.

Assumption 3.1.1. *The density function of $x^T\beta$ is positive and uniformly continuous for β in a neighborhood of β^0 .*

Assumption 3.1.2. *Let $F_0(t|x)$ and $M_0(t|x)$ denote the cumulative distribution functions of T_i and C_i conditional on x , respectively. The first derivatives of $F_0(t|x)$ and $M_0(t|x)$ with respect to t , denoted as $f_0(t|x)$ and $m_0(t|x)$, are uniformly bounded with respect to t and x . The $F_0(t|x)$ and $M_0(t|x)$ have bounded (uniformly in t) second-order partial derivatives with respect to x . In addition, $\sup_t |F_0(t|x') - F_0(t|x)| = O(\|x' - x\|)$, where $\|\cdot\|$ denotes the Euclidean norm.*

Assumption 3.1.3. *The random vector x is bounded in probability, has a bounded density function, and $E(xx^T)$ is a positive definite $p \times p$ matrix.*

Assumption 3.1.4. *There exists an effective dimension reduction (EDR) direction $\gamma_j \in R^q$ such that for any $j = 1, \dots, q$, (i) $\hat{\gamma}_j - \gamma_j = O_p(n^{-1/2})$; (ii) $n^{-1/2}(\hat{\gamma}_k - \gamma_k) = n^{-1} \sum_{i=1}^n d_{ki}$, where d_{ki} are independent p -dimensional vectors with means zero and finite variances.*

Assumption 3.1.5. *The univariate kernel function $K(\cdot)$ has a compact support. It is a v -th order kernel function satisfying $\int K(u)du = 1$, $\int K^2(u)du \leq \infty$, $\int u^j K(u)du = 0$ for $j \leq v$ and $\int |u|^v K(u)du \leq \infty$, and it is Lipschitz continuous of order v , where $v \geq 2$ is an integer.*

Assumption 3.1.6. *The conditional density function of Y (either T or C) given $X = x$, $f_Y(y|x)$ satisfies the Lipschitz condition of order 1, and $f_Y(\tilde{g}(x^T\beta, \beta)|x) > 0$ for $\beta \in \Theta$.*

Assumption 3.1.7. *$\tilde{g}(u, \beta) \in \mathcal{H}_\vartheta$ for some $\vartheta > 3/2$ and any $\beta \in \Theta$.*

The above assumptions are commonly used in literature and can be satisfied in most applications. Assumption 3.1.1 was used in Wu et al. (2010) and Assumptions

3.1.2 to 3.1.5 were assumed in Wang et al. (2013). Assumption 3.1.2 is to obtain the asymptotic properties of the local Kaplan-Meier estimator. The boundedness condition on x in Assumption 3.1.3 is for technical convenience. Assumption 3.1.1 and Assumption 3.1.3 also show the boundedness of density function on $x^T\beta$. Assumption 3.1.4 indicates the root-n consistency and asymptotic properties of the estimated EDR direction. Assumption 3.1.5 shows that kernel function is v th order, which depends on the dimensionality of q . Similar as in Ma and He (2016), the Assumption 3.1.6 and Assumption 3.1.7 are used to help provide the asymptotic normality.

3.2 Asymptotic Properties

Assume C and T are independent conditional on x , and set $h = O(n^{-\alpha})$ for some α . For any positive numbers a_n and b_n , $a_n \ll b_n$ means $\frac{a_n}{b_n} = o(1)$.

Theorem 3.2.1. (*Consistency*) Under Assumptions 3.1.1-3.1.7, if $J_n \rightarrow \infty$, $J_n \ll n$ and $0 < \alpha < \min(\frac{1}{q}, \frac{1}{v})$, we have

$$\|\tilde{\beta} - \beta^0\|_2 = o_p(1)$$

Recall $J_n = s + r$, which is the dimension of the spline space for approximating $g(\cdot)$. Theorem 3.2.1 states the consistency of the unpenalized estimator $\tilde{\beta}$. To prepare for the asymptotic normality, we define

$$\Gamma = E \left\{ \left[E \left[(\tau - F_0(C|z)) I(F_0(C|z) \leq \tau) m_0(g(X^T\beta^0)) \right] \right. \right. \\ \left. \left. + E \left[(1 - M_0(T|x)) f_0(g(X^T\beta^0)) \right] \right] g'^2(X^T\beta^0) \tilde{X} \tilde{X}^T \right\},$$

$$V = E \left\{ \left(E \left[\frac{(\tau - F_0(C|z))^2}{1 - F_0(C|z)} I(F_0(C|z) \leq \tau) \rho_\tau'^2 (C - g(x^T \beta^0)) \right] \right. \right. \\ \left. \left. + E \left[(1 - M_0(T|x)) \rho_\tau'^2 (T - g_0(x^T \beta^0)) \right] \right) g'^2(X^T \beta^0) \tilde{X} \tilde{X}^T \right\},$$

$$\tilde{X} = X - E^*(X|X^T \beta^0),$$

$$E^*(X|X^T \beta^0) = \frac{E \left\{ \left[1 - M_0(g(X^T \beta^0)) \right] f_0(g(X^T \beta^0)) X \mid X^T \beta^0 \right\}}{E \left\{ \left[1 - M_0(g(X^T \beta^0)) \right] f_0(g(X^T \beta^0)) \mid X^T \beta^0 \right\}}.$$

Theorem 3.2.2. (*Asymptotic normality*) Under the Assumptions 3.1.1-3.1.7, if

$$\max \{ (n \log(n))^{\frac{1}{3\vartheta-1/2}}, n^{\frac{1}{2\vartheta+2}} \} \ll J_n \ll \frac{n^{1/4}}{\log(n)^{5/4}} \text{ and } \frac{1}{2v} < \alpha < \frac{1}{3q},$$

and $v > \frac{3q}{2}$, we have

$$n^{1/2} \{ \tilde{\beta} - \beta^0 \} \xrightarrow{D} N(0, \Gamma^+ V \Gamma^+),$$

where Γ^+ is the Moore-Penrose inverse of Γ .

The root- n consistency and asymptotic normality of the unpenalized estimator are described in Theorem 3.2.2. When the data is noncensored, i.e., $C = +\infty$, the proposed estimator has the same asymptotic variance as in Wu et al. (2010) and Ma and He (2016). Note that by definition of Γ , we have $\Gamma \beta^0 = 0$, and Γ is a singular matrix. Hence we need to use the Moore-Penrose inverse of Γ in Theorem 3.2.2.

Next we establish the asymptotic property of the penalized estimator $\hat{\beta}$. Denote $\mathcal{A} = \{j : \beta_j^0 \neq 0\}$ and $\hat{\mathcal{A}} = \{j : \hat{\beta}_j \neq 0\}$ as the true and estimated relevant sets of variables.

Theorem 3.2.3. Under Assumptions 3.1.1-3.1.7, and $\lambda/\sqrt{n} \rightarrow 0$, $\lambda n^{\frac{\delta}{2}-1} \rightarrow \infty$, we

have

$$(i). P\left(\hat{\mathcal{A}} = \mathcal{A}\right) \rightarrow 1, \text{ as } n \rightarrow \infty.$$

$$(ii). \sqrt{n}(\hat{\beta}_{\mathcal{A}}^1 - \beta_{\mathcal{A}}^1) \xrightarrow{D} N(0, \Gamma_1^+ V_1 \Gamma_1^+),$$

where $\hat{\beta}_{\mathcal{A}}^1$ denotes the subvector consisting of all the non-zero elements in $\hat{\beta}$, and $\beta_{\mathcal{A}}^1$ is the true non-zero subvector of β^0 . Γ_1 and V_1 are defined as in Γ and V in Theorem 3.2.2 with β^0 replaced by $\beta_{\mathcal{A}}^1$.

The property of penalized estimator $\hat{\beta}$ is shown in Theorem 3.2.3 with assumptions on the tuning parameter. Numerically, we can re-estimate the non-zero coefficients in $\hat{\beta}$ by minimizing the unpenalized objective function, due to the selection consistency in Theorem 3.2.1. Similar strategies are adopted in Wang et al. (2013) and Ma and He (2016).

Chapter 4

Simulation

Data in this section is generated from models commonly used in the literature. On the whole, our method performs relatively well compared with existing methods. We performed 100 repetitions for all the models with different sample sizes, censoring percentages and quantiles levels. The performance of estimation is examined by mean squared error (MSE).

In Model 1, our proposed unpenalized method is compared with the method in Bücher et al. (2014) and Christou and Akritas (2016). As a special case of proposed method, the uncensored method also achieves small mean squared error (MSE) in Model 2, in comparison with methods in Wu et al. (2010) and Ma and He (2016). In Model 3, quantile coefficient varies with quantile levels. Our penalized approach can select variables with heteroscedasticity effectively in this model, even with relatively high percent of censoring.

4.1 Model 1: Location Model

The set up for this simulation study follows Model 1 of Bücher et al. (2014) and Christou and Akritas (2016),

$$T_i = \exp(x_i^T \beta^0) + \epsilon_i, x_i = (x_{i,1}, x_{i,2}), \quad (4.1)$$

where x_{ij} is *i.i.d* from $Unif(0, 1)$, for $i = 1, \dots, n$ and $j = 1, 2$, ϵ is *i.i.d* standard exponential, $\beta^0 = (1, 2)$, and $Y_i = \min(T_i, C_i)$, with C_i from exponential distribution with parameter ς , independent of x_i and ϵ_i . Here, $\varsigma = 0.047$ yields about 25% censoring, and $\varsigma = 0.120$ yields about 50% censoring. Note that the τ th conditional of T_i given x_i is $\exp(x_i^T \beta^0) - \log(1 - \tau)$.

Table 4.1 compares the mean and standard error of the estimated β^0 by the proposed method with ones in Bücher et al. (2014) and Christou and Akritas (2016), for $n=400$ in 100 generated data sets. BGK denotes the estimator of Bücher et al. (2014), CA denotes the estimator from Christou and Akritas (2016), while CN denotes our proposed unpenalized method. Also, we rescale the first component estimate as 1, and only report the result for the second component estimation β_2 (whose true value is 2). Moreover, the BGK results are based on 10% trimming, as used in their paper due to convergence issue; while both CA and CN report untrimmed results. The comparison shows that the proposed method CN outperforms BGK in both bias and standard error (even with untrimmed results), and has comparable bias but smaller standard error compared with CA.

Figure 4.1 illustrates the performance of the estimator for the link function $\tilde{g}(u, \tilde{\beta})$ over the estimated linear combination u , where $u = x^T \tilde{\beta}$. From the comparison of “True” and “Estimated”, we can see that the estimated link function with

Table 4.1: The average and standard error (in parenthesis) of estimating $\beta_2 = 2$ for different quantiles and censoring percentage based on $n=400$, 100 runs in Model 1.

Avg. Cens	Methods	$\tau = 0.1$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.9$
25%	BGK	2.13 (0.58)	2.54 (0.23)	2.50 (0.22)	2.63 (0.26)	4.18 (2.03)
	CA	2.03 (0.24)	2.02 (0.21)	1.99 (0.21)	2.03 (0.45)	2.12 (0.74)
	CN	2.00 (0.04)	1.99 (0.06)	2.00 (0.09)	2.01 (0.16)	1.98 (0.28)
50%	BGK	2.81 (0.30)	2.58 (0.16)	2.55 (0.17)	2.71 (0.41)	217.20 (894.51)
	CA	2.01 (0.16)	2.01 (0.16)	1.97 (0.27)	2.02 (0.48)	2.19 (0.79)
	CN	1.99 (0.04)	1.99 (0.07)	1.98 (0.11)	1.98 (0.22)	1.96 (0.90)

estimated $\tilde{\beta}$ is very close to the true link function $g(x^T \beta^0) = \exp(x^T \beta^0)$ with true coefficient β^0 .

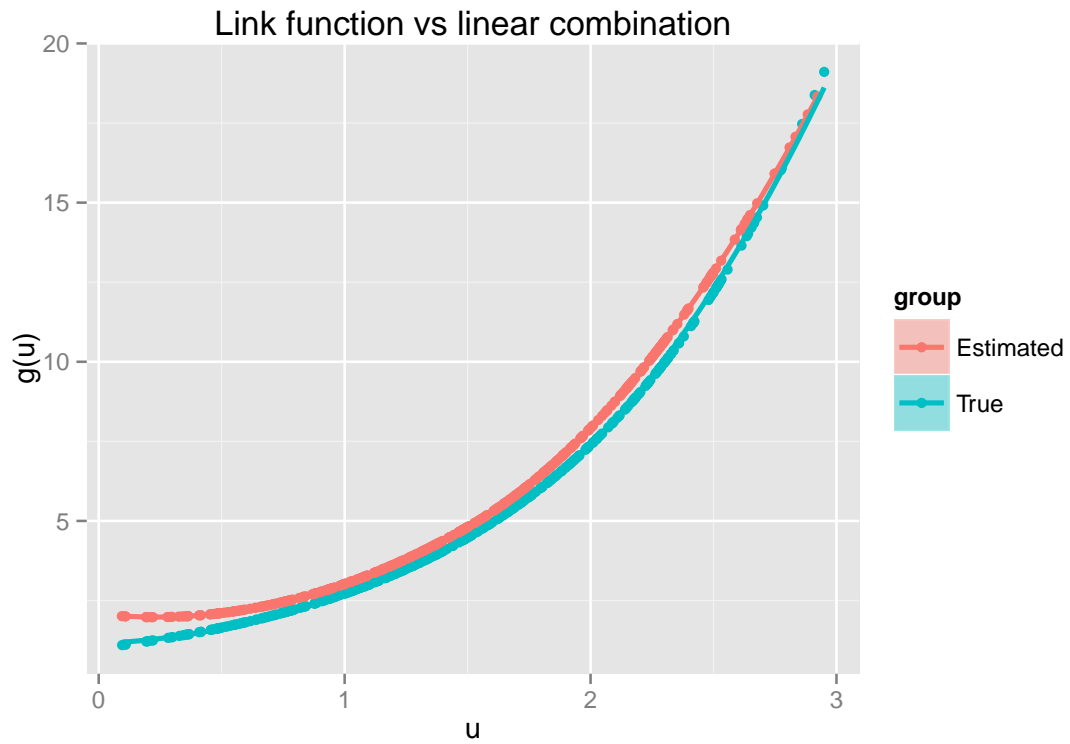
Overall, our unpenalized methods (CN) performed better than BGK and CA with less bias and variation at different quantile levels and censoring percentages. Besides, the proposed method is very robust; even for relatively high percent of censoring, high quantile levels can also be estimated with reasonable variation.

4.2 Model 2: Sine-bump Model

A sine-bump model as specified below, is considered. The same model is used for mean regression (Carroll et al, 1997), and for quantile regression without censoring (Wu et al., 2010; Ma and He, 2016),

$$T_i = \sin\left\{\frac{\pi(x_i^T \beta^0) - A}{C - A}\right\} + 0.1\epsilon_i, x_i = (x_{i,1}, \dots, x_{i,3}), \quad (4.2)$$

Figure 4.1: Comparison of link function versus linear combination for 50% censoring and $\tau = 0.25$ result in Model 1. Here, “True” means the true link function versus true linear combination ($x^T \beta^0$); while “Estimated” indicates the estimated link function over estimated linear combination ($x^T \tilde{\beta}$).



where $A = \sqrt{3}/2 - 1.645/\sqrt{12}$, $C = \sqrt{3}/2 + 1.645/\sqrt{12}$, x_{ij} are *i.i.d* from $Unif(0, 1)$, for $i = 1, \dots, n$ and $j = 1, 2, 3$.

First, we compare our methods (NC means data without censoring, CN means data with 15% censored observations) with the method in Wu et al. (2010), named WYY. As in their model, we specify the true coefficient $\beta^0 = (1, 1, 1)/\sqrt{3}$, and $\epsilon_i \sim N(0, 1)$. We observe $Y_i = \min(T_i, C_i)$, where $C_i \sim Unif(-2, 15)$ yielding about 15% censoring. From Table 4.2, we can see our method performs better than the WYY for data without censoring (NC), and results with 15% censoring (CN) is comparable to the WYY.

Table 4.2: The bias, standard error, and mean squared error (MSE) of estimation based on $n=200$, $\tau = 0.5$, 100 runs, with the initial value $(1, 2, 0)/\sqrt{5}$, true value $(1, 1, 1)/\sqrt{3}$ and $\epsilon \sim N(0, 1)$ in Model 2. Here, WYY denotes the kernel method estimation from Wu et al. (2010), NC means data without censoring, CN means 15% data with censoring.

Methods	Estimate	β_1	β_2	β_3
WYY	Bias	-0.00210	0.00509	-0.00366
	se	0.01515	0.01583	0.01577
	MSE	0.00023	0.00028	0.00026
NC	Bias	0.00051	0.00022	-0.00116
	se	0.01157	0.01395	0.01306
	MSE	0.00013	0.00019	0.00017
CN	Bias	-0.00064	0.00150	-0.00138
	se	0.01374	0.01422	0.01446
	MSE	0.00019	0.00020	0.00021

We also compared our method with the profile and backfitting non-censored methods in Ma and He (2016). In their model, the true coefficient $\beta^0 = (3, 2, 1)/\sqrt{14}$ and two distributions for ϵ are considered: t distribution with 3 degree of freedom, and Laplace distribution. Here, $C_i \sim Unif(-2, 7.5)$ yielding about 15% censoring, PR denotes the profile estimation, BA means backfitting estimator, NC means data without censoring, CN means 15% data with censoring. From Table 4.3, we can see

NC works similar as PR and is better than BA in MSE, and censored result CN is also comparable to NC, PR and BA. For Laplace distribution, BA has one outlier of estimation (0.2245, 0.2491, 0.9167), which is circled in yellow from Figure 4.2. BA tends to have outliers due to convergence issue, which is also discussed in Ma and He (2016).

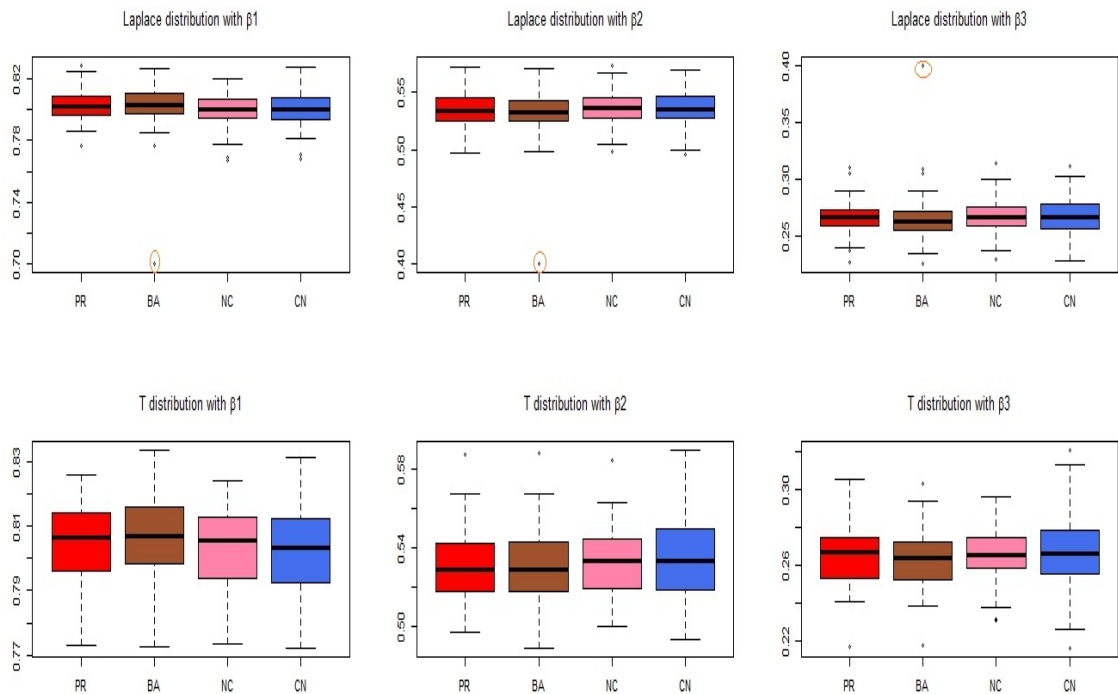
Table 4.3: The bias, standard error, and mean squared error (MSE) estimation based on $n=200$, $\tau = 0.5$, 100 simulations, with the initial value $(1, 3, 6)/\sqrt{46}$ in Model 2. Here, PR denotes the profile estimation, BA means backfitting estimator, NC denotes data without censoring, CN means 15% data with censoring.

Distribution	Methods	Estimate	β_1	β_2	β_3	
Laplace	PR	Bias	0.0012	-0.0014	-0.0018	
		se	0.0093	0.0144	0.0132	
		MSE	0.0001	0.0002	0.0002	
	BA	Bias	-0.0026	-0.0047	0.0023	
		se	0.0502	0.0320	0.0667	
		MSE	0.0025	0.0010	0.0045	
	NC	Bias	-0.0016	0.0020	-0.0001	
		se	0.0099	0.0142	0.0141	
		MSE	0.0001	0.0002	0.0002	
	CN	Bias	-0.0015	0.0020	-0.0006	
		se	0.0108	0.0157	0.0160	
		MSE	0.0001	0.0002	0.0003	
	T(3)	PR	Bias	0.0030	-0.0042	-0.0019
			se	0.0121	0.0177	0.0147
			MSE	0.0002	0.0003	0.0002
BA		Bias	0.0040	-0.0046	-0.0041	
		se	0.0125	0.0178	0.0141	
		MSE	0.0002	0.0003	0.0002	
NC		Bias	0.0011	-0.0015	-0.0015	
		se	0.0121	0.0167	0.0140	
		MSE	0.0001	0.0003	0.0002	
CN		Bias	0.0006	-0.0013	-0.0008	
		se	0.0137	0.0199	0.0177	
		MSE	0.0002	0.0004	0.0003	

The boxplot in Figure 4.2 shows that BA tends to have some outlying points for

Laplace distribution, while PR, CN and NC work quite well. The simulation results indicate for the special case without censoring, the proposed NC can work similar as PR, and is better than BA. What's more, the 15% censored result CN also work satisfactorily well in Model 2.

Figure 4.2: *Boxplots of coefficient estimates for $n=200$, $\tau = 0.5$, initial value $(1, 3, 6)/\sqrt{46}$, 100 runs, different error distributions in Model 2. Here, PR denotes the profile estimation, BA means backfitting estimator, NC denotes data without censoring, CN means 15% data with censoring. Note, the true standardized coefficients are $\beta^0 = (0.8018, 0.5345, 0.2673)$. The yellow circle is the outlier $(0.2245, 0.2491, 0.9167)$ from BA.*



4.3 Model 3: Varying Coefficient Model

Model 3 has similar setting as in Wang et al. (2013), but the linear quantile regression was extended to non linear quantile regression in the single index model.

$$T_i = \exp\{1 + 1.5x_{i1} + 0.7x_{i2} + x_{i3} - 0.5x_{i4} + (1 + \gamma x_{i4})\epsilon_i\}, x_i = (x_{i,1}, \dots, x_{i,p}), \quad (4.3)$$

where x_{ij} are *i.i.d* $\sim Unif(-1, 1)$, for $i = 1, \dots, n$ and $j = 1, \dots, p$ and ϵ are *i.i.d* standard normal with $p=10$. We set $\gamma = -0.742$, so that the quantile coefficients are $\beta_{0.25}=(1.5, 0.7, 1, 0, \dots, 0)$, and $\beta_{0.5}=(1.5, 0.7, 1, -0.5, 0, \dots, 0)$. Under this heteroscedastic model, the covariate x_{i4} has different impacts on the quantiles of the conditional distribution of T. We observe $Y_i = \min(T_i, C_i)$, where $C_i \sim Unif(0, 48)$ yielding about 15% censoring, and $C_i \sim Unif(0, 18)$ indicating about 30% censoring, and $C_i \sim Unif(0, 8.5)$ about 45% censoring. In this model, the dimension of covariates is relatively large, so we use cSIR (Li et al., 1999) to estimate the dimension reduction indexes, as well as the dimension q by the sequential chi square test. In both Table 4.4 and 4.5, we report the results with the true $q=2$ and the estimated \hat{q} by cSIR.

Table 4.4 summarizes the estimation performance in terms of MSE for different quantile levels, sample sizes and average percentages of censoring. Here, the Oracle estimator is obtained by using the proposed unpenalized method under the true model, i.e, with the first three covariates at $\tau = 0.25$ and the first four covariates at $\tau = 0.5$. We can see that the results based on the estimated \hat{q} are very similar to that based on the true q . As sample size increases, the MSEs become smaller, and closer to the Oracle estimators. As the percentage of censoring increases, the MSEs are larger. Overall, the MSEs of proposed method are relatively small, indicating a

good performance of estimation. Also, from the boxplot in Figure 4.3, the majority of estimation values are around the true coefficient, which shows good performance of estimation in heteroscedastic model.

Moreover, in Table 4.5, we can see the good performance of variable selection with relatively large OP and TP, small FP. Here, TP denotes the average number of relevant variables that are correctly selected. FP denotes the average number of irrelevant variables that are incorrectly selected. OP denotes the percentage of times that the true model is correctly selected. As n increases from 500 to 800, TP and OP increase and FP decrease. Also, as the average percentage of censoring increases, the variable selection becomes more difficult. Besides, estimation and variable selection at $\tau = 0.5$ are usually better than at $\tau = 0.25$ when censoring is well below 50%, which is widely seen in the quantile regression literature.

Figure 4.3: *Boxplots of first four coefficient estimates for $n=800$, average 15% censoring for different quantiles based on 100 runs in Model 3. Note, the true standardized coefficients for lower quantiles are $\beta_{0.25} = (0.7756, 0.3620, 0.5171, 0, 0, 0, 0, 0, 0, 0)$, and for median quantiles are $\beta_{0.5} = (0.7509, 0.3504, 0.5006, -0.2503, 0, 0, 0, 0, 0, 0)$.*

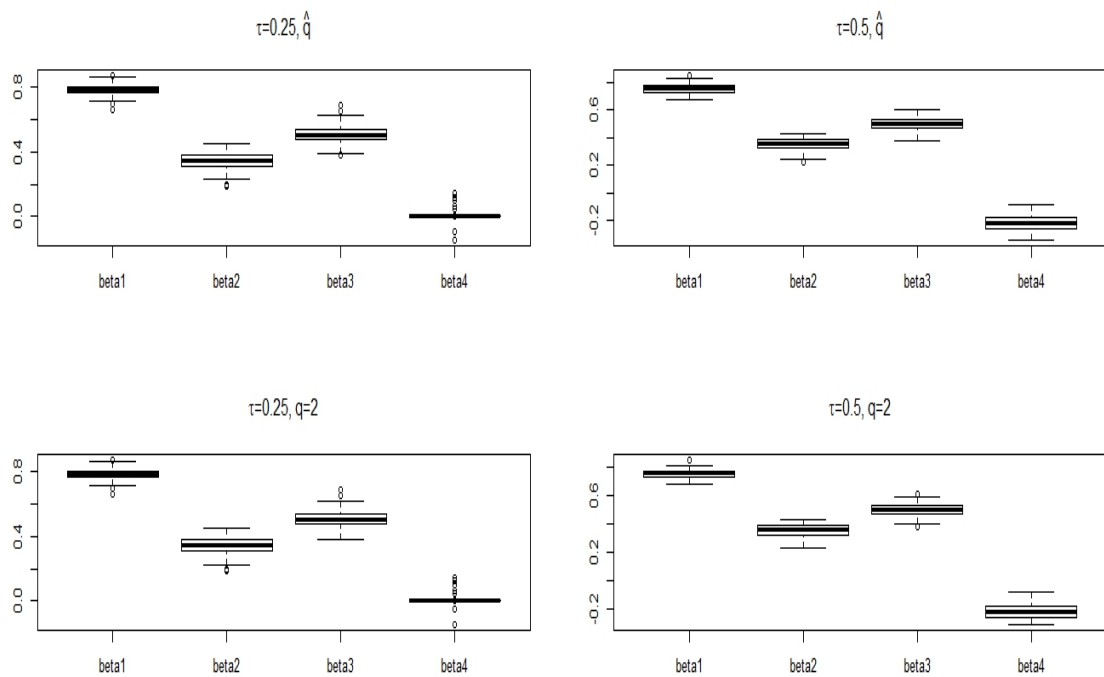


Table 4.4: The mean squared error ($\times 100$) of estimation for none zero quantile coefficients based on 100 runs in Model 3. The “Oracle” estimator is obtained by using the proposed unpenalized method under the true model. \hat{q} and $q=2$ denote the estimated/ true number of index. Note, the true standardized coefficients for lower quantiles are $\beta_{0.25} = (0.7756, 0.3620, 0.5171, 0, 0, 0, 0, 0, 0)$, and for median quantiles are $\beta_{0.5} = (0.7509, 0.3504, 0.5006, -0.2503, 0, 0, 0, 0, 0)$.

	$\tau = 0.25$			$\tau = 0.5$			
	β_1	β_2	β_3	β_1	β_2	β_3	β_4
n=500, 15% censoring							
Oracle	0.18	0.36	0.26	0.16	0.31	0.36	0.61
\hat{q}	0.28	0.59	0.45	0.24	0.55	0.43	0.84
$q=2$	0.29	0.58	0.43	0.24	0.54	0.43	1.10
n=500, 30% censoring							
Oracle	0.15	0.43	0.32	0.14	0.28	0.25	1.03
\hat{q}	0.28	0.59	0.58	0.50	0.47	0.51	2.08
$q=2$	0.26	0.67	0.53	0.24	0.44	0.36	2.17
n=500, 45% censoring							
Oracle	0.15	0.38	0.27	0.12	0.32	0.24	1.51
\hat{q}	0.77	1.10	1.05	4.13	2.09	2.57	7.72
$q=2$	0.22	0.74	0.43	0.26	0.66	0.28	4.18
n=800, 15% censoring							
Oracle	0.10	0.18	0.21	0.09	0.21	0.18	0.37
\hat{q}	0.13	0.31	0.31	0.15	0.27	0.27	0.60
$q=2$	0.13	0.32	0.31	0.14	0.30	0.26	0.58
n=800, 30% censoring							
Oracle	0.08	0.19	0.18	0.08	0.17	0.14	0.77
\hat{q}	0.17	0.37	0.37	0.16	0.28	0.20	1.74
$q=2$	0.16	0.34	0.35	0.17	0.27	0.21	2.06
n=800, 45% censoring							
Oracle	0.09	0.24	0.20	0.10	0.18	0.14	1.08
\hat{q}	0.21	0.53	0.39	1.40	1.23	2.34	4.98
$q=2$	0.20	0.52	0.36	0.20	0.28	0.18	3.58

Table 4.5: The variable selection results for Model 3 with different censoring, sample size and quantile levels. TP denotes the average number of relevant variables that are correctly selected. FP denotes the average number of irrelevant variables that are incorrectly selected. OP denotes the percentage of times that the true model is correctly selected. \hat{q} denotes the estimated number of index.

n	Avg. Cens	τ	TP(\hat{q})	FP(\hat{q})	OP(\hat{q})	TP	FP	OP
500	15%	0.25	3.00	0.48	0.71	3.00	0.51	0.70
		0.50	3.95	0.31	0.74	3.90	0.28	0.73
	30%	0.25	3.00	0.64	0.62	3.00	0.64	0.63
		0.50	3.79	0.26	0.65	3.76	0.09	0.70
	45%	0.25	2.99	1.47	0.40	3.00	0.76	0.56
		0.50	3.72	2.39	0.33	3.40	0.09	0.38
800	15%	0.25	3.00	0.31	0.77	3.00	0.29	0.78
		0.50	3.97	0.12	0.90	3.98	0.10	0.92
	30%	0.25	3.00	0.56	0.66	3.00	0.56	0.65
		0.50	3.87	0.14	0.80	3.83	0.05	0.80
	45%	0.25	3.00	0.71	0.61	3.00	0.55	0.68
		0.50	3.77	1.43	0.48	3.52	0.02	0.50

Chapter 5

Real Data Application

5.1 Boston Housing Data

We fit the single index for censored quantile regression model to the Boston housing data, available in the R package MASS. The data contains 506 observations of 14 variables. The response variable is MEDV, and the other 13 variables are predictors. A detailed description can be found in Table 5.1. There are several papers that have analyzed this data, e.g., Harrison and Rubinfeld (1978), Doksum and Samarov (1995) and Fan and Huang (2005). We follow previous studies and take the logarithmic transformations on TAX and LSTAT. One noteworthy feature is that the dependent variable is right censored. That is, the values of MEDV that are larger than 50000 have been recorded as 50000 for 16 out of 506 subjects. Thus, modeling on the conditional quantiles is more appropriate than modeling of conditional mean, see Wu et al. (2010) and Kong and Xia (2012). But both of the above methods ignore the censoring of the response.

Each conditional quantile with censoring is modeled by a single index model with variable selection. By using the dimension reduction approach, we get the estimated

Table 5.1: Description of Variables in Boston Housing Data.

Variable	Description
CRIM	per capita crime rate by town
ZN	proportion of residential land zoned for lots over 25,000 sq.ft.
INDUS	proportion of non-retail business acres per town
CHAS	Charles River dummy variable (= 1 if tract bounds river; 0 otherwise)
NOX	nitrogen oxides concentration (parts per 10 million)
RM	average number of rooms per dwelling
AGE	proportion of owner-occupied units built prior to 1940
DIS	weighted mean of distances to five Boston employment centers
RAD	index of accessibility to radial highways
TAX	full-value property-tax rate per \$10,000
PTRATIO	pupil-teacher ratio by town
BLACK	$1000(Bk - 0.63)^2$ where Bk is the proportion of blacks by town
LSTAT	lower status of the population (percent)
MEDV	median value of owner-occupied homes in \$1000

number of dimension $\hat{q} = 3$. Then the estimated 10th, 25th, 50th, 75th and 90th quantiles are listed in Table 5.2. We can see that different quantiles may have different important variable lists. For $\tau = 0.1$, CRIM, NOX, RM and LSTAT are selected, and additional DIS and PTRATIO are selected when $\tau = 0.25$. Moreover, additional CHAS, RAD and BLACK are picked for $\tau = 0.5$ than $\tau = 0.25$. For $\tau = 0.75$ and $\tau = 0.9$, they select the same sets of variables, but the coefficients are different, and even the signs of AGE are different. Overall, CRIM, NOX, RM and LSTAT are important to housing price for all quantiles, which is a consistent conclusion compared with previous studies.

As houses in higher price range are usually targeted by people with higher income, it's beneficial to study the consumption behavior of people from different income groups to examine how the coefficients of covariates vary as the quantile level changes. In this data, we can see CRIM, NOX and LSTAT have negative effects on house prices, while the influence of RM is positive. Also, for lower income

Table 5.2: Estimation and Selection Results for Different Quantiles in Boston Housing Data.

Variable	$\tau = 0.1$	$\tau = 0.25$	$\tau = 0.5$	$\tau = 0.75$	$\tau = 0.9$
CRIM	-0.056	-0.114	-0.043	-0.006	-0.006
ZN	0.000	0.000	0.000	0.001	0.002
INDUS	0.000	0.000	0.000	0.006	0.002
CHAS	0.000	0.000	0.066	0.007	0.014
NOX	-0.891	-0.757	-0.841	-0.962	-0.882
RM	0.341	0.462	0.308	0.036	0.072
AGE	0.000	0.000	0.000	-0.001	0.002
DIS	0.000	-0.159	-0.253	-0.269	-0.451
RAD	0.000	0.000	0.018	0.009	0.014
log(TAX)	0.000	0.000	0.000	0.000	0.000
PTRATIO	0.000	-0.012	-0.046	-0.022	-0.017
BLACK	0.000	0.000	0.001	0.000	0.000
log(LSTAT)	-0.295	-0.436	-0.354	-0.019	-0.115

population ($\tau = 0.1$ or $\tau = 0.25$), only few variables are taken into consideration of purchasing homes. On the other hand, for higher income population ($\tau = 0.75$ or $\tau = 0.9$), many factors are in their consideration, choosing 11 out of 13 variables. For the median income population ($\tau = 0.5$), they consider reasonable amount of variables.

5.2 Kidney Transplant Study

As mentioned in the motivation, the relentless growth in the size of the kidney transplant on waiting list accentuates the importance and urgency to develop an innovative and accurate tool to quantify the relative risk of kidney graft failure associated with various combination of donor characteristics as well as related information including transplant factors. Port et al. (2002) introduced expanded criteria donor (ECD) to relieve this problem, and it has been reported that kidney trans-

plants from ECD donors have at least 70% greater risk of graft failure than those from the lower risk Standard Criteria Donor (SCD) donors. However, the ECD and SCD dichotomy only utilizes a few donor characteristics including age, serum creatinine, history of hypertension and history of stroke.

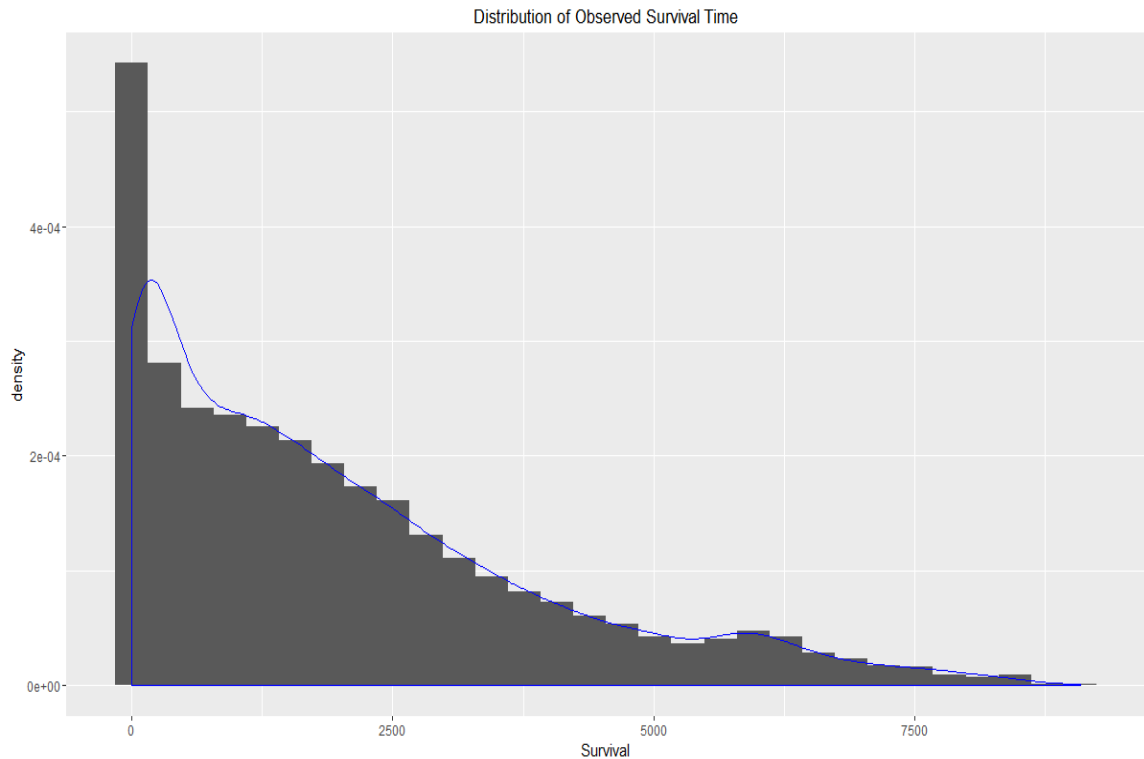
An improvement over the ECD/SCD dichotomy is the scoring system developed by Nyberg et al. (2003). The system included five variables: donor age, history of hypertension, creatinine clearance, cause of death and the number of HLA-mismatches. This scoring system is not widely used in practice since it's not replicable in independent cohort (Lee and Abramowicz, 2014).

Rao et al. (2009) proposed a continuous kidney donor risk index (KDRI) for deceased donor kidneys, combining donor and transplant variables to quantify graft failure risk. Using the data between 1995 and 2005 from the Scientific Registry of Transplant Recipients (SRTR), they analyzed 69440 first-time, kidney-only, deceased donor adult transplants by Cox regression adjusted for recipient factors. The proposed KDRI included 14 donor and transplant factors associated with graft failure. Although KDRI is a useful decision-making tool at the time of the deceased donor kidney offer, there is still room for improvement as discussed in Motivation section. For example, the risk factors may be different or have different effects on the patients with shorter survival, compared to patients with longer survival. Hence, we applied the proposed single index model for censored quantile regression, analyzing the same data obtained from the Scientific Registry of Transplant Recipients (SRTR). Moreover, random forest was performed as benchmark to evaluate the importance of selected variables.

Following the idea of Rao et al. (2009), we consider 47128 deceased donor transplants took place between October 1, 1987 and September 30, 1997 from the SRTR data by applying the same exclusion criterion. Particularly, after excluding missing

values, we are interested in a subset where Donor's age is over 50. It consists of 7321 deceased donor transplant where 3767 of them had experienced the event, the graft failure. The graft failure is defined as return to dialysis, retransplant or death. From Figure 5.1, we can tell that the distribution of observed survival time (blue line is the smoothing density curve) is quite skewed to the right. So we preferred the quantile regression since the traditional mean regression may not be appropriate.

Figure 5.1: *The Distribution of Time to Graft Failure (Days) in Kidney Transplant*



Additionally, similar sets of variables were used in our single index model as in Rao et al. (2009). Donor and transplant factors were donor age (DONAGE), race (DONRACE), gender (DONGENDER), cause of death (DONCOD), organ sharing (DONORG). Recipient factors included age (RECAGE), height (RECHGT), weight

(RECWGT), race (RECRACE), gender (RECGENDER), Hepatitis C Virus positivity (RECHCV). The summary statistics for the above factors are summarized in the Table 5.3.

Table 5.3: Summary statistics

Source	Variable	Mean (SE) or Percentage
Donor	Age	57.59 (5.84)
	Race (black)	6.08%
	Gender (female)	54.51%
	Cause of Death (stroke)	75.29%
	Organ Sharing	26.59%
Recipient	Age	45.97 (12.65)
	Height (cm)	170.40 (10.78)
	Weight (kg)	74.45 (18.10)
	Race (black)	26.97%
	Gender (female)	38.87%
	Hepatitis C Virus	2.88%

We used the dimension reduction approach (cSIR) from Li et al. (1999) to get the estimated dimension $\hat{q} = 2$. Continuous variables are standardized to have zero mean and one standard deviation. The selection results for different quantiles are in the Table 5.4. For relatively high risk patients ($\tau = 0.25$), Donor's age, race (black), cause of death (stroke), recipient's weight and race (black) are negatively associated with survival time, while Recipient's age has positive influence. But for median survival time patients ($\tau = 0.5$), cause of death (stroke) is not selected any more. The influence of Donor's race and Recipient's age is larger, while the effect of Donor's age, Recipient's weight and race are smaller. For both quantiles, Donor's gender, organ sharing, and Recipient's height, gender and Hepatitis C virus are not important in terms of survival time. We also plot the estimation of link function for both lower and median quantile levels in Figure 5.2. The shape of link function is non-linear and is different for lower and median quantile levels, which is also

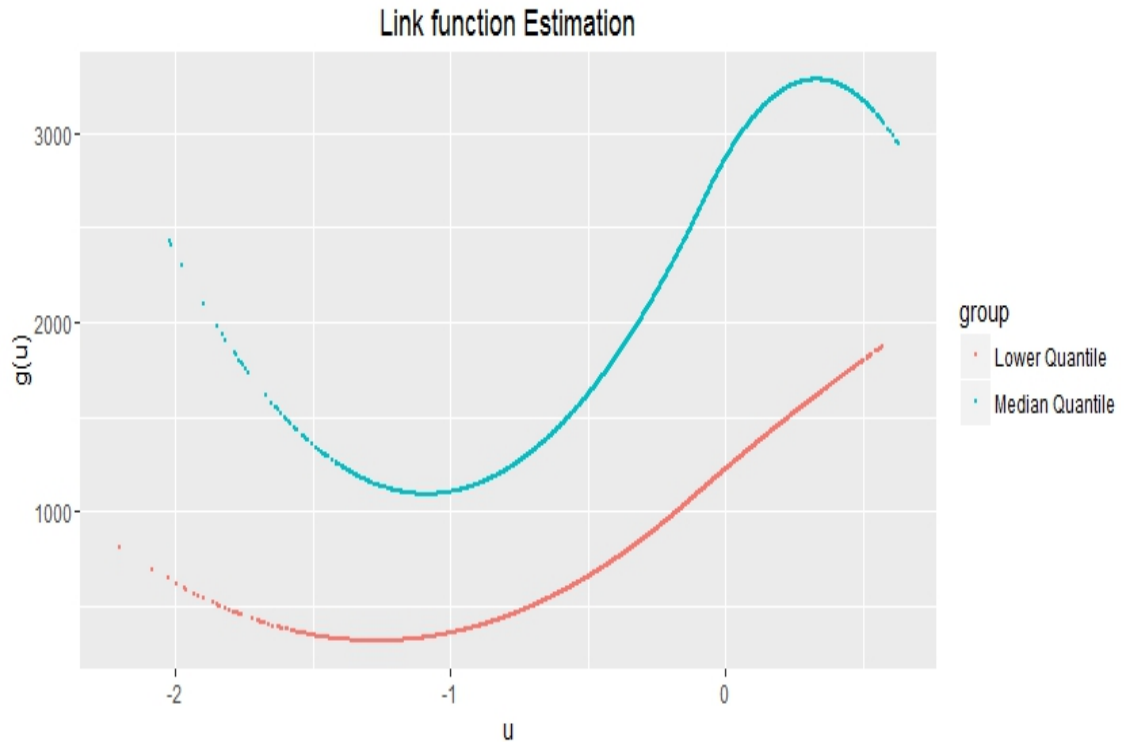
Table 5.4: The variable selection results for kidney transplant study with different quantile levels.

Source	Variable	$\tau = 0.25$	$\tau = 0.5$
Donor	Age	-0.233	-0.225
	Race (black)	-0.811	-0.869
	Gender (female)	0.000	0.000
	Cause of Death (stroke)	-0.070	0.000
	Organ Sharing	0.000	0.000
Recipient	Age	0.118	0.146
	Height (cm)	0.000	0.000
	Weight (kg)	-0.058	-0.040
	Race (black)	-0.516	-0.414
	Gender (female)	0.000	0.000
	Hepatitis C Virus	0.000	0.000

the advantages over the traditional linear models. Also, the overall trend of link function is increasing for both index of quantile levels. That is to say, higher index will indicate longer survival time.

To evaluate the importance of selected predictors, we compare our results with the variance importance scores from Random Forest methods (Breiman, 2001). In every bootstrapped sample i ($i = 1, \dots, b$, b is the number of bootstraps), we save the cases not used in the bootstrap sample (“out-of-bag”, OOB cases) and calculate an estimated OOB error e_i . After randomly permuting the values in the m^{th} of p variables (m is the number of randomly selected variables in each split) in the OOB cases, we calculate the OOB error f_i again. Then we subtract the two OOB errors $f_i - e_i$ for all the bootstrapped data. The average of this number over all bootstrapped random forest trees is the raw importance score for m^{th} variable. Ishwaran et al. (2007) extended the random forest for the survival outcome. The importance scores for all variables are summarized in the Table 5.5 and Figure 5.3. The results from Random Forest are pretty consistent with that from our approach.

Figure 5.2: *The estimation of link function for lower quantile with $\tau = 0.25$ and median quantile with $\tau = 0.5$.*

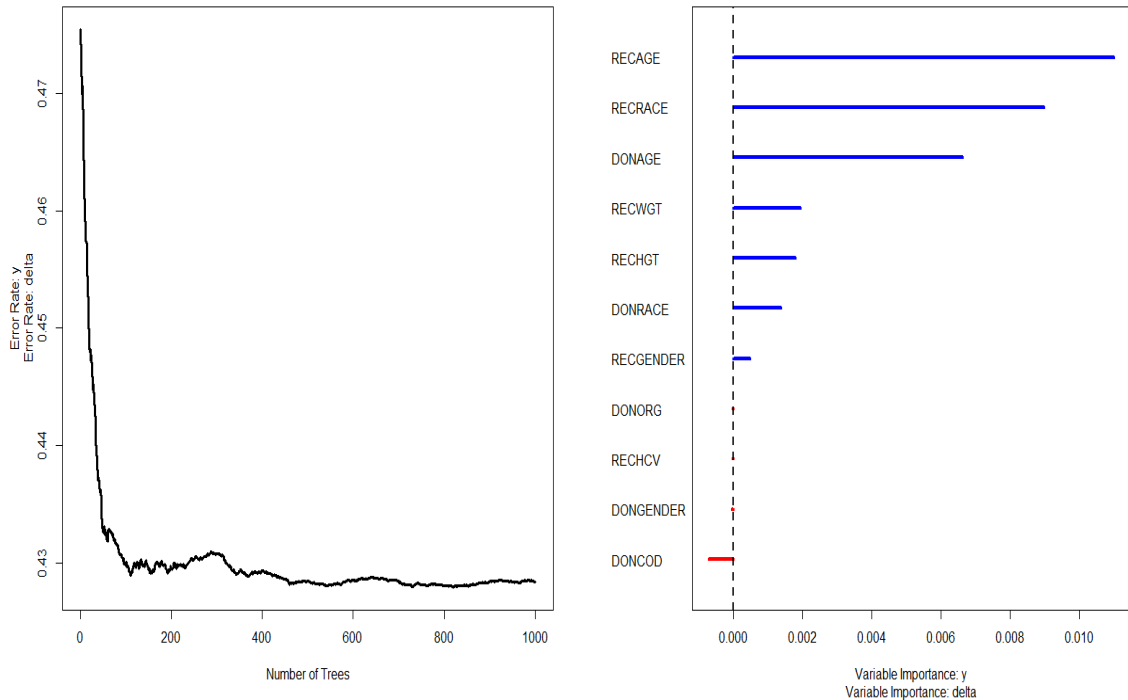


Our selected variables such as Donor age, race, and Recipient age, race and weight have the higher importance scores. Our method is favored over random forest since it can select important variables for different quantiles and indicate the direction of effect for better interpretation. Random forest, however, is a black box machine learning technique, lacking interpretation and needing shrinkage to achieve variable selection.

Table 5.5: The variable importance scores from random forest for survival outcome. The larger the score, the more important it is.

Variable	Importance	Description
RECAGE	0.0110	Recipient age
RECRACE	0.0090	Recipient race
DONAGE	0.0066	Donor age
RECWGT	0.0019	Recipient weight
RECHGT	0.0018	Recipient height
DONRACE	0.0014	Donor race
RECGENDER	0.0005	Recipient gender
DONORG	0.0000	Donor organ sharing
RECHCV	0.0000	Recipient Hepatitis c virus
DONGENDER	0.0000	Donor gender
DONCOD	-0.0007	Donor cause of death

Figure 5.3: Variable importance from random forest



Chapter 6

Conclusion and Discussion

In this thesis, we proposed a single index model for censored quantile regression that can reduce dimension and deal with censored outcomes effectively. The B-spline was used to approximate the unknown link function in the model and redistribution of mass approaches were adopted to deal with censoring. Dimension reduction approaches were adopted to reduce the dimension to alleviate “curse of dimensionality” for computing the redistributed mass. Moreover, variable selection methods, such as adaptive LASSO, were performed to select the informative variables while estimating the model simultaneously. The performance of proposed methods was evaluated by simulation studies, and compared with existing methods. Both the theoretical and numerical results show that the proposed methods perform satisfactory for censored data. Our proposed method can be transformed to a linear quantile regression problem so that many existing linear quantile regression algorithms can be used for efficient estimation; this is advantageous to Bücher et al. (2014) and Christou and Akritas (2016), which uses time consuming local linear approximation or complicated nonlinear quantile regression algorithms. Another advantage of our proposed method is that it uses the penalized linear quantile estimator to se-

lect informative variables for single index quantile regression model, while existing methods in Bücher et al. (2014) and Christou and Akritas (2016) do not facilitate variable selection for quantiles.

Some interesting future directions could be further investigated.

- The thesis focuses on the model specification and estimation with variable selection. It is still an open question to do statistical inference after variable selection in single index quantile regression model. We might borrow the ideas from the mean regression literature for statistical inference after variable selection. Wasserman and Roeder (2009) proposed a two-stage procedure for valid inference. In their method, the data is randomly divided into two equal parts: training and testing datasets. In the training data, penalized regression is used to select informative variables on the first stage. In the testing data, ordinary least squares is applied to compute standard errors and p -values for the selected variables in the first stage. A drawback of the single-split method is that the result may depend on how the data is split. To overcome this, Meinshausen et al. (2012) suggested multi-split method, which repeats the single-split multiple times, and obtains the empirical distribution of the p -values. Also, Lockhart et al. (2014) proposed the covariance test statistic to test the significance of the predictor variable that enters the current LASSO model.
- We assume that data has a fixed dimension p as n increases in the single index quantile regression model. If the number of covariates grows with sample size, challenges arise in two parts: the estimation of unknown link function $g(\cdot)$ with growing dimension, and the properties of the dimension reduction methods for quantile regression with growing dimension. Fan and Peng (2004)

proposed the variable selection procedures for parametric models via nonconcave penalized likelihood with a diverging number of parameters. The oracle property of the penalized likelihood estimators has also been established. Lam and Fan (2008) explored the profile likelihood estimation and inference for the generalized varying coefficient partially linear model with growing number of predictors. In the framework of single index model, Zhu and Zhu (2009) studied the nonconcave penalized inverse regression method with diverging number of predictors. The asymptotic normality was achieved and confidence interval can be constructed for the estimated index. In the single index quantile regression model, additional work is needed when the number of variables increases with n .

- It is also natural to extend our methods in single index models to multiple index models. That is, the $\beta_\tau \in R^{p \times d}$ can have multiple dimension with $d \geq 1$ in the model (2.3). There are a lot of work on multiple index models in mean regression, but very limited in the quantile regression. In the mean regression, there are three classes of methods available for estimating the column space of β : the inverse regression based methods, the nonparametric methods and the semiparametric methods. The origin of inverse regression is an ingenious idea of reversing the relation between the response variable and the covariates (SIR, Li et al., 1991; Duan and Li, 1991). The most original nonparametric method is the minimum average variance estimation method (MAVE, Xia et al., 2002). The semiparametric methods from Ma and Zhu (2012) involve writing down the likelihood of one observation, recognizing that the essential problem of dimension reduction is equivalent to that of parameter estimation in the presence of nuisance components, and taking advantage of the influence

function structure to avoid estimating all the nuisance components. In the quantile regression framework, we may apply our nonparametric method to the multiple index models.

Chapter 7

Technical Proofs

In this chapter, we first derive the proofs of Theorem 3.2.1 to Theorem 3.2.3 in Chapter 3. Then, we present and prove nine lemmas that are used in the proofs of three main theorems.

7.1 Proof of Theorems

Proof of Theorem 3.2.1

Denote

$$\begin{aligned} \hat{L}_{\tau n}^*(\beta) = & 1/n \sum_{i=1}^n \left\{ \frac{\tau - \hat{F}(C_i|\hat{z}_i)}{1 - \hat{F}(C_i|\hat{z}_i)} I(T_i > C_i, \hat{F}(C_i|\hat{z}_i) \leq \tau) \rho_\tau(C_i - B(X_i^T \beta) \tilde{\theta}) \right. \\ & + I(T_i \leq C_i) \rho_\tau(T_i - B(X_i^T \beta) \tilde{\theta}) \\ & \left. + \left[I(T_i > C_i) - \frac{\tau - \hat{F}(C_i|\hat{z}_i)}{1 - \hat{F}(C_i|\hat{z}_i)} I(T_i > C_i, \hat{F}(C_i|\hat{z}_i) \leq \tau) \right] \rho_\tau(Y^{+\infty} - B(X_i^T \beta) \tilde{\theta}) \right\}, \end{aligned}$$

$$\begin{aligned}
L_{\tau n}^*(\beta) = & \frac{1}{n} \sum_{i=1}^n \left\{ \frac{\tau - F_0(C_i|z_{0i})}{1 - F_0(C_i|z_{0i})} I(T_i > C_i, F_0(C_i|z_{0i}) \leq \tau) \rho_\tau \left(C_i - B(X_i^T \beta) \tilde{\theta} \right) \right. \\
& + I(T_i \leq C_i) \rho_\tau \left(T_i - B(X_i^T \beta) \tilde{\theta} \right) \\
& \left. + \left[I(T_i > C_i) - \frac{\tau - F_0(C_i|z_{0i})}{1 - F_0(C_i|z_{0i})} I(T_i > C_i, F_0(C_i|z_{0i}) \leq \tau) \right] \rho_\tau \left(Y^{+\infty} - B(X_i^T \beta) \tilde{\theta} \right) \right\},
\end{aligned}$$

$$\begin{aligned}
\tilde{L}_\tau^*(\beta) = & \frac{1}{n} \sum_{i=1}^n \left\{ \frac{\tau - F_0(C_i|z_{0i})}{1 - F_0(C_i|z_{0i})} I(T_i > C_i, F_0(C_i|z_{0i}) \leq \tau) \rho_\tau \left(C_i - \tilde{g}(X_i^T \beta, \beta) \right) \right. \\
& + I(T_i \leq C_i) \rho_\tau \left(T_i - \tilde{g}(X_i^T \beta, \beta) \right) \\
& \left. + \left[I(T_i > C_i) - \frac{\tau - F_0(C_i|z_{0i})}{1 - F_0(C_i|z_{0i})} I(T_i > C_i, F_0(C_i|z_{0i}) \leq \tau) \right] \rho_\tau \left(Y^{+\infty} - \tilde{g}(X_i^T \beta, \beta) \right) \right\},
\end{aligned}$$

and

$$\begin{aligned}
L_\tau^*(\beta) = & E \left\{ \frac{\tau - F_0(C|z_0)}{1 - F_0(C|z_0)} I(T > C, F_0(C|z_0) \leq \tau) \rho_\tau \left(C - \tilde{g}(X^T \beta, \beta) \right) \right. \\
& + I(T \leq C) \rho_\tau \left(T - \tilde{g}(X^T \beta, \beta) \right) \\
& \left. + \left[I(T > C) - \frac{\tau - F_0(C|z_0)}{1 - F_0(C|z_0)} I(T > C, F_0(C|z_0) \leq \tau) \right] \rho_\tau \left(Y^{+\infty} - \tilde{g}(X^T \beta, \beta) \right) \right\}.
\end{aligned}$$

Assume $\tilde{\beta}$ is the minimizer of $\hat{L}_{\tau n}^*(\beta)$, then $P\{\hat{L}_{\tau n}^*(\tilde{\beta}) \leq \hat{L}_{\tau n}^*(\beta^0)\} = 1$. Then for any open set $S(\beta^0)$ that includes β^0 ,

$$\begin{aligned}
& P\{\hat{L}_{\tau n}^*(\tilde{\beta}) \leq \hat{L}_{\tau n}^*(\beta^0)\} \\
= & P\{\hat{L}_{\tau n}^*(\tilde{\beta}) \leq \hat{L}_{\tau n}^*(\beta^0), \text{ and } \tilde{\beta}_\tau \in S(\beta^0)\} + P\{\hat{L}_{\tau n}^*(\tilde{\beta}_\tau) \leq \hat{L}_{\tau n}^*(\beta^0), \text{ and } \tilde{\beta}_\tau \in \Theta \setminus S(\beta^0)\} \\
\leq & P\{\tilde{\beta}_\tau \in S(\beta^0)\} + P\{\inf_{\beta \in \Theta \setminus S(\beta^0)} \hat{L}_{\tau n}^*(\beta) \leq \hat{L}_{\tau n}^*(\beta^0)\}.
\end{aligned}$$

If we can prove that

$$P\{\inf_{\beta \in \Theta \setminus S(\beta^0)} \hat{L}_{\tau n}^*(\beta) \leq \hat{L}_{\tau n}^*(\beta^0)\} \rightarrow 0,$$

then $\tilde{\beta}$ must be in any open set $S(\beta^0)$, so that $\tilde{\beta}$ is a consistent estimator of β^0 with probability approaching 1.

$$\begin{aligned} & P\left\{\inf_{\beta \in \Theta \setminus S(\beta^0)} \hat{L}_{\tau n}^*(\beta) \leq \hat{L}_{\tau n}^*(\beta^0)\right\} \\ &= P\left\{\inf_{\beta \in \Theta \setminus S(\beta^0)} \left[\hat{L}_{\tau n}^*(\beta) - L_{\tau n}^*(\beta) + L_{\tau n}^*(\beta) - \tilde{L}_{\tau}^*(\beta) + \tilde{L}_{\tau}^*(\beta) - L_{\tau}^*(\beta) + L_{\tau}^*(\beta)\right] \leq \hat{L}_{\tau n}^*(\beta^0)\right\} \\ &\leq P\left\{\inf_{\beta \in \Theta \setminus S(\beta^0)} [\hat{L}_{\tau n}^*(\beta) - L_{\tau n}^*(\beta)] + \inf_{\beta \in \Theta \setminus S(\beta^0)} [L_{\tau n}^*(\beta) - \tilde{L}_{\tau}^*(\beta)] \right. \\ &\quad \left. + \inf_{\beta \in \Theta \setminus S(\beta^0)} [\tilde{L}_{\tau}^*(\beta) - L_{\tau}^*(\beta)] + L_{\tau}^*(\beta^0) - L_{\tau n}^*(\beta^0) \leq L_{\tau}^*(\beta^0) - \inf_{\beta \in \Theta \setminus S(\beta^0)} L_{\tau}^*(\beta)\right\} \\ &\leq P\left\{\sup_{\beta \in \Theta \setminus S(\beta^0)} |\hat{L}_{\tau n}^*(\beta) - L_{\tau n}^*(\beta)| + \sup_{\beta \in \Theta \setminus S(\beta^0)} |L_{\tau n}^*(\beta) - \tilde{L}_{\tau}^*(\beta)| \right. \\ &\quad \left. + \sup_{\beta \in \Theta \setminus S(\beta^0)} |\tilde{L}_{\tau}^*(\beta) - L_{\tau}^*(\beta)| + |L_{\tau}^*(\beta^0) - L_{\tau n}^*(\beta^0)| \geq \inf_{\beta \in \Theta \setminus S(\beta^0)} L_{\tau}^*(\beta) - L_{\tau}^*(\beta^0)\right\} \end{aligned}$$

Since β^0 is the unique minimizer of $L_{\tau}^*(\beta)$, then for every open set $S(\beta^0)$, there exists $\epsilon > 0$ such that $\inf_{\beta \in \Theta \setminus S(\beta^0)} L_{\tau}^*(\beta) - L_{\tau}^*(\beta^0) > \epsilon$. It's sufficient to prove, for every $\epsilon > 0$,

- (i). $P\left\{\sup_{\beta \in \Theta \setminus S(\beta^0)} |\hat{L}_{\tau n}^*(\beta) - L_{\tau n}^*(\beta)| > \epsilon\right\} \rightarrow 0$,
- (ii). $P\left\{\sup_{\beta \in \Theta \setminus S(\beta^0)} |L_{\tau n}^*(\beta) - \tilde{L}_{\tau}^*(\beta)| > \epsilon\right\} \rightarrow 0$,
- (iii). $P\left\{\sup_{\beta \in \Theta \setminus S(\beta^0)} |\tilde{L}_{\tau}^*(\beta) - L_{\tau}^*(\beta)| > \epsilon\right\} \rightarrow 0$,
- (iv). $P\left\{|L_{\tau}^*(\beta^0) - L_{\tau n}^*(\beta^0)| > \epsilon\right\} \rightarrow 0$.

To verify (i), we have the following,

$$\begin{aligned} & \hat{L}_{\tau n}^*(\beta) - L_{\tau n}^*(\beta) \\ = & 1/n \sum_{i=1}^n \left[\frac{\tau - \hat{F}(C|\hat{z})}{1 - \hat{F}(C|\hat{z})} I(\hat{F}(C|\hat{z}) \leq \tau, T_i > C_i) - \frac{\tau - F_0(C|z_0)}{1 - F_0(C|z_0)} I(F_0(C|z_0) \leq \tau, T_i > C_i) \right] \\ & \left[\rho_\tau(C_i - B(X_i^T \beta) \tilde{\theta}) - \rho_\tau(Y^{+\infty} - B(X_i^T \beta) \tilde{\theta}) \right]. \end{aligned}$$

For

$$\begin{aligned} & \left| \left(\frac{\tau - \hat{F}}{1 - \hat{F}} I(\hat{F} \leq \tau) - \frac{\tau - F_0}{1 - F_0} I(F_0 \leq \tau) \right) \right| \\ = & I(\hat{F} \leq \tau, F_0 \leq \tau) \frac{(1 - \tau)|\hat{F} - F_0|}{(1 - \hat{F})(1 - F_0)} + I(\hat{F} < \tau < F_0) \frac{1 - \tau}{1 - F_0} + I(F_0 < \tau < \hat{F}) \frac{1 - \tau}{1 - \hat{F}} \\ \leq & \frac{|\hat{F} - F_0|}{1 - \tau} + I(\hat{F} < \tau < F_0) + I(F_0 < \tau < \hat{F}) \end{aligned}$$

then we have

$$E \left[\sup_{\|\hat{F} - F_0\|_{\mathcal{H}} \leq \epsilon_n} I(F_0 < \tau < \hat{F}) \right] \leq P(F_0 < \tau < F_0 + \epsilon_n) = M(F_0^{-1}(\tau)) - M(F_0^{-1}(\tau - \epsilon_n)) \leq c \cdot \epsilon_n.$$

Together with Lemma 7.2.9, we can have the following,

$$|\hat{L}_{\tau n}^*(\beta) - L_{\tau n}^*(\beta)| \leq O_p \left(\left\{ \frac{\log n}{nh^q} \right\}^{1/2} + h^v \right).$$

Under the condition that $0 < \alpha < \frac{1}{q}$ when $h = O(n^{-\alpha})$, then

$$\sup_{\beta \in \Theta \setminus \mathcal{S}(\beta^0)} |\hat{L}_{\tau n}^*(\beta) - L_{\tau n}^*(\beta)| \leq o_p(1).$$

To prove (ii), from Lemma 7.2.4,

$$\begin{aligned}
& \sup_{\beta \in \Theta \setminus S(\beta^0)} |L_{\tau n}^*(\beta) - \tilde{L}_\tau^*(\beta)| \\
&= \sup_{\beta \in \Theta \setminus S(\beta^0)} \left| \frac{1}{n} \sum_{i=1}^n \left[\rho_\tau(T_i - \tilde{g}_n(X_i^T \beta, \beta)) - \rho_\tau(T_i - \tilde{g}(X_i^T \beta, \beta)) \right] I(T_i \leq C_i) \right. \\
&\quad + \frac{\tau - F_0(C_i | z_{0i})}{1 - F_0(C_i | z_{0i})} I(T_i > C_i, F_0(C_i | z_{0i})) \left[\rho_\tau(C_i - \tilde{g}_n(X_i^T \beta, \beta)) - \rho_\tau(C_i - \tilde{g}(X_i^T \beta, \beta)) \right] \\
&\quad \left. + \left[I(T_i > C_i) - \frac{\tau - F_0(C_i | z_{0i})}{1 - F_0(C_i | z_{0i})} I(T_i > C_i, F_0(C_i | z_{0i})) \right] \tau \left[\tilde{g}_n(X_i^T \beta, \beta) - \tilde{g}(X_i^T \beta, \beta) \right] \right| \\
&\leq \sup_{\beta \in \Theta \setminus S(\beta^0)} \frac{3}{n} \sum_{i=1}^n \left| \tilde{g}_n(X_i^T \beta, \beta) - \tilde{g}(X_i^T \beta, \beta) \right| = o_p(1).
\end{aligned}$$

The proof of (iii),

$$P \left\{ \sup_{\beta \in \Theta \setminus S(\beta^0)} |\tilde{L}_\tau^*(\beta) - L_\tau^*(\beta)| > \epsilon \right\} \rightarrow 0$$

is followed by the uniform consistency theorem in Andrew (1987).

To get (iv), by Lemma 7.2.4, we have $L_{\tau n}^*(\beta^0) - \tilde{L}_{\tau n}^*(\beta^0) = o_p(1)$. Also by the weak law of large number, $\tilde{L}_\tau^*(\beta^0) - L_{\tau n}^*(\beta^0) = o_p(1)$. Again, from (i), under the condition that $0 < \alpha < \min(\frac{1}{q}, \frac{1}{v})$, we know, $|\hat{L}_{\tau n}^*(\beta) - L_{\tau n}^*(\beta)| \leq o_p(1)$. Then

$$|L_\tau^*(\beta^0) - \hat{L}_{\tau n}^*(\beta^0)| \leq |\hat{L}_{\tau n}^*(\beta^0) - L_{\tau n}^*(\beta^0)| + |L_{\tau n}^*(\beta^0) - \tilde{L}_\tau^*(\beta^0)| + |\tilde{L}_\tau^*(\beta^0) - L_\tau^*(\beta^0)| = o_p(1).$$

Thus, the Theorem 3.2.1 is obtained.

Proof of Theorem 3.2.2:

Since we know from Lemma 7.2.2,

$$\sqrt{n} \{ \tilde{\beta}_{F_0, \gamma_0} - \beta^0 \} \xrightarrow{D} N(0, \Gamma^+ V \Gamma^+),$$

then if we can prove

$$\sqrt{n} \left(\tilde{\beta}_{F_0, \gamma_0} - \tilde{\beta} \right) = o_p(1),$$

then it's complete.

From Lemma 7.2.2, we know $\hat{b}_{\tau, n} = \tilde{\beta}_{F_0, \gamma_0} - \beta^0$, and $\hat{b}_{\tau, n} - \tilde{b}_{\tau, n} = o_p(n^{-1/2})$. Then similarly, then we have $\hat{b}_{\tau, n}^* - \tilde{b}_{\tau, n}^* = o_p(n^{-1/2})$, if $\hat{b}_{\tau, n}^* = \tilde{\beta} - \beta^0$. Note that

$$\tilde{b}_{\tau, n} = -(\Gamma_n + \xi_n I)^{-1} \Omega_n,$$

$$\tilde{b}_{\tau, n}^* = -(\hat{\Gamma}_n + \xi_n I)^{-1} \hat{\Omega}_n,$$

where

$$\begin{aligned} \Omega_n = & -\frac{1}{n} \sum_{i=1}^n \left\{ \frac{\tau - F_0(C_i | z_{0i})}{1 - F_0(C_i | z_{0i})} I(T_i > C_i, F_0(C_i | z_{0i}) \leq \tau) \rho'_\tau(C_i - g(x_i^T \beta^0)) \right. \\ & + I(T_i \leq C_i) \rho'_\tau(T_i - g(x_i^T \beta^0)) \\ & \left. + \left[E_T(M(t|x)) - E_C(1 - F_0(C_i | z_{0i})) \frac{\tau - F_0(C_i | z_{0i})}{1 - F_0(C_i | z_{0i})} \right] \tau \right\} g'(x_i^T \beta^0) \tilde{x}_i, \end{aligned}$$

$$\begin{aligned} \hat{\Omega}_n = & -\frac{1}{n} \sum_{i=1}^n \left\{ \frac{\tau - \hat{F}(C_i | \hat{z}_i)}{1 - \hat{F}(C_i | \hat{z}_i)} I(T_i > C_i, \hat{F}(C_i | \hat{z}_i) \leq \tau) \rho'_\tau(C_i - g(x_i^T \beta^0)) \right. \\ & + I(T_i \leq C_i) \rho'_\tau(T_i - g(x_i^T \beta^0)) \\ & \left. + \left[E_T(M(t|x)) - E_C(1 - F_0(C_i | z_{0i})) \frac{\tau - \hat{F}(C_i | \hat{z}_i)}{1 - \hat{F}(C_i | \hat{z}_i)} \right] \tau \right\} g'(x_i^T \beta^0) \tilde{x}_i, \end{aligned}$$

$$\begin{aligned} \Gamma_n = & \frac{1}{n} \sum_{i=1}^n \left\{ E_T \left[(1 - M(T_i|x)) f_0(g(x_i^T \beta^0)) \right] \right. \\ & \left. + E_C \left[(\tau - F_0(C_i | z_{0i})) I(F_0(C_i | z_{0i}) \leq \tau) m_0(g(x_i^T \beta^0)) \right] \right\} [g'(x_i^T \beta^0)]^2 \tilde{x}_i \tilde{x}_i^T, \end{aligned}$$

$$\begin{aligned}\hat{\Gamma}_n &= \frac{1}{n} \sum_{i=1}^n \left\{ E_T \left[(1 - M(T_i|x)) f_0(g(x_i^T \beta^0)) \right] \right. \\ &\quad \left. + E_C \left[(1 - F_0(C_i|z_{0i})) \frac{\tau - \hat{F}(C_i|\hat{z}_i)}{1 - \hat{F}(C_i|\hat{z}_i)} I(\hat{F}(C_i|\hat{z}_i) \leq \tau) m_0(g(x_i^T \beta^0)) \right] \right\} [g'(x_i^T \beta^0)]^2 \tilde{x}_i \tilde{x}_i^T,\end{aligned}$$

The following is to prove that $\tilde{b}_{\tau,n} - \tilde{b}_{\tau,n}^* = o_p(n^{-1/2})$, then the proof is trivial.

$$\begin{aligned}\tilde{b}_{\tau,n} - \tilde{b}_{\tau,n}^* &= -(\Gamma_n + \xi_n I)^{-1} \Omega_n + (\hat{\Gamma}_n + \xi_n I)^{-1} \hat{\Omega}_n \\ &= -(\Gamma_n + \xi_n I)^{-1} \Omega_n + (\Gamma_n + \xi_n I)^{-1} \hat{\Omega}_n - (\Gamma_n + \xi_n I)^{-1} \hat{\Omega}_n + (\hat{\Gamma}_n + \xi_n I)^{-1} \hat{\Omega}_n \\ &= (\Gamma_n + \xi_n I)^{-1} (\hat{\Omega}_n - \Omega_n) + [(\hat{\Gamma}_n + \xi_n I)^{-1} - (\Gamma_n + \xi_n I)^{-1}] \hat{\Omega}_n\end{aligned}$$

We get the above rate by the following,

$$\begin{aligned}&\hat{\Omega}_n - \Omega_n \\ &= \frac{1}{n} \sum_{i=1}^n g'(x_i^T \beta^0) \tilde{x}_i \left\{ E_C (1 - F_0(C_i|z_{0i})) \left[\frac{\tau - F_0(C_i|z_{0i})}{1 - F_0(C_i|z_{0i})} - \frac{\tau - \hat{F}(C_i|\hat{z}_i)}{1 - \hat{F}(C_i|\hat{z}_i)} \right] \tau + \right. \\ &\quad \left[\frac{\tau - F_0(C_i|z_{0i})}{1 - F_0(C_i|z_{0i})} I(T_i > C_i, F_0(C_i|z_{0i}) \leq \tau) \right. \\ &\quad \left. \left. - \frac{\tau - \hat{F}(C_i|\hat{z}_i)}{1 - \hat{F}(C_i|\hat{z}_i)} I(T_i > C_i, \hat{F}(C_i|\hat{z}_i) \leq \tau) \right] \rho'_\tau(C_i - g(x_i^T \beta^0)) \right\} \\ &= \frac{1}{n} \sum_{i=1}^n \left\{ \left[\frac{\tau - F_0}{1 - F_0} I(T_i > C_i, F_0 \leq \tau) - \frac{\tau - \hat{F}}{1 - \hat{F}} I(T_i > C_i, \hat{F} \leq \tau) \right] \rho'_\tau(C_i - g(x_i^T \beta^0)) + \right. \\ &\quad \left. E_C \left[\frac{\hat{F} - F_0}{1 - \hat{F}} \right] \tau (1 - \tau) \right\} g'(x_i^T \beta^0) \tilde{x}_i\end{aligned}$$

Using the same technique in Theorem 3.2.1, we have

$$\begin{aligned}
& \left| \left(\frac{\tau - \hat{F}}{1 - \hat{F}} I(\hat{F} \leq \tau) - \frac{\tau - F_0}{1 - F_0} I(F_0 \leq \tau) \right) \right| \\
&= I(\hat{F} \leq \tau, F_0 \leq \tau) \frac{(1 - \tau)|\hat{F} - F_0|}{(1 - \hat{F})(1 - F_0)} + I(\hat{F} < \tau < F_0) \frac{1 - \tau}{1 - F_0} + I(F_0 < \tau < \hat{F}) \frac{1 - \tau}{1 - \hat{F}} \\
&\leq \frac{|\hat{F} - F_0|}{1 - \tau} + I(\hat{F} < \tau < F_0) + I(F_0 < \tau < \hat{F}),
\end{aligned}$$

Here, we insert the linear representation of $\hat{F} - F_0$ in Lemma 7.2.9 (ii) and applying the Taylor extension, we obtain $\hat{\Omega}_n - \Omega_n = \frac{1}{n} \sum_{i=1}^n \phi_i + O_p(\{\frac{\log n}{nh^q}\}^{3/4} + h^v)$ where $\phi_i =$, and ξ is defined in Lemma 7.2.9. Here ϕ_i are independent random variables with mean zero. Thus, under the condition that $\frac{1}{2v} < \alpha < \frac{1}{3q}$, we have

$$\hat{\Omega}_n - \Omega_n = O_p(\{\frac{\log n}{nh^q}\}^{3/4} + h^v) = O_p(n^{-1/2}).$$

By using the techniques that

$$[(\hat{\Gamma}_n + \xi_n I)^{-1} - (\Gamma_n + \xi_n I)^{-1}] = -(\hat{\Gamma}_n + \xi_n I)^{-1}(\hat{\Gamma}_n - \Gamma_n)(\Gamma_n + \xi_n I)^{-1}$$

And we know that

$$\begin{aligned}
\hat{\Gamma}_n - \Gamma_n &= \frac{1}{n} \sum_{i=1}^n \left\{ E_C \left[(1 - F_0) \left(\frac{\tau - F_0}{1 - F_0} I(T_i > C_i, F_0 \leq \tau) - \frac{\tau - \hat{F}}{1 - \hat{F}} I(T_i > C_i, \hat{F} \leq \tau) \right) \right. \right. \\
&\quad \left. \left. m_0(g(x_i^T \beta^0)) \right] \right\} [g'(x_i^T \beta^0)]^2 \tilde{x}_i \tilde{x}_i^T \\
&= O_p(n^{-1/2}).
\end{aligned}$$

Also, we know $\hat{\Gamma}_n, \Gamma_n, \hat{\Omega}_n, \Omega_n$ are all $O_p(1)$. Thus, we can prove

$$\tilde{b}_{\tau,n} - \tilde{b}_{\tau,n}^* = o_p(n^{-1/2}).$$

So the proof of Theorem 3.2.2 is complete.

Proof of Theorem 3.2.3:

Let $\hat{u} = \sqrt{n}(\hat{\beta} - \beta^0)$, and from the proof of Theorem 3.2.2, \hat{u} minimizes the following

$$\begin{aligned} \hat{Q}_n(u) &= \frac{1}{2} u^T \hat{\Gamma}_n u + \hat{\Omega}_n^T u + o_p(1) \\ &\quad + \sum_{j=1}^p \frac{\lambda_n}{\sqrt{n} |\tilde{\beta}_j|^r} \sqrt{n} \left(|\beta_j^* + \frac{u_j}{\sqrt{n}}| - |\beta_j^*| \right), \end{aligned}$$

where $\hat{\Gamma}_n$ and $\hat{\Omega}_n$ defined in the proof of Theorem 3.2.2.

If $\beta_j^* \neq 0$, then $|\tilde{\beta}_j|^r \xrightarrow{P} |\beta_j^*|^r$, and $\sqrt{n} \left(|\beta_j^* + \frac{u_j}{\sqrt{n}}| - |\beta_j^*| \right) \xrightarrow{P} u_j \text{sgn}(\beta_j^*)$. By Slutsky's Theorem, and $\lambda_n/\sqrt{n} \rightarrow 0$, we have $\frac{\lambda_n}{\sqrt{n} |\tilde{\beta}_j|^r} \sqrt{n} \left(|\beta_j^* + \frac{u_j}{\sqrt{n}}| - |\beta_j^*| \right) \xrightarrow{P} 0$.

If $\beta_j^* = 0$, then $\sqrt{n} \left(|\beta_j^* + \frac{u_j}{\sqrt{n}}| - |\beta_j^*| \right) = |u_j|$, and $\sqrt{n} \tilde{\beta}_j = O_p(1)$,

we have $\frac{\lambda_n}{\sqrt{n} |\tilde{\beta}_j|^r} = \frac{\lambda_n}{\sqrt{n}} n^{r/2} |\sqrt{n} \tilde{\beta}_j|^{-r} \rightarrow \infty$. Thus, by Slutsky's Theorem,

$$\hat{Q}_n(u) \rightarrow L(u) = \begin{cases} \frac{1}{2} u^T \hat{\Gamma}_n u + \hat{\Omega}_n^T u, & \text{if } u_j = 0, \forall j \notin \mathcal{A} \\ \infty, & \text{otherwise} \end{cases}$$

Since $\hat{Q}_n(u)$ is convex and $L(u)$ has a unique minimizer. Following the epi-convergence results of Geyer (1994) and Knight and Fu (2000), we can prove the asymptotic normality.

Now we prove the consistency part. $\forall j \in \mathcal{A}$, the asymptotic normality indicates that $\hat{\beta}_j \xrightarrow{P} \beta_j^*$, then $P(j \in \hat{\mathcal{A}}) \rightarrow 1$. Hence it suffices to show that $\forall j \notin \mathcal{A}, P(j \in \hat{\mathcal{A}}) \rightarrow 0$. Follow the similar proof in Theorem 2 in Wang et al. (2013), suppose there exists a $k \in \mathcal{A}^c$, such that $|\hat{\beta}_k| \neq 0$. Denote $\hat{\beta}^*$ as a vector constructed by replacing $\hat{\beta}_k$ with 0 in $\hat{\beta}$. Note that $|\rho_\tau(a) - \rho_\tau(b)| \leq |a - b|$ and

$0 \leq w_i \leq 1$, then for large enough n , we have

$$\begin{aligned}
L(\hat{\beta}, \hat{w}) - L(\hat{\beta}^*, \hat{w}) &= \sum_i \left\{ \hat{w}_i \left[\rho_\tau(Y_i - B(x_i^T \hat{\beta})\theta) - \rho_\tau(Y_i - B(x_i^T \hat{\beta}^*)\theta) \right] \right. \\
&\quad \left. + (1 - \hat{w}_i) \left[\rho_\tau(Y^{+\infty} - B(x_i^T \hat{\beta})\theta) - \rho_\tau(Y^{+\infty} - B(x_i^T \hat{\beta}^*)\theta) \right] \right\} \\
&\quad + \lambda_n \frac{\hat{\beta}_k}{|\hat{\beta}_k|^r} \\
&\geq -2 \sum_i \theta \|x_i\| \cdot |\hat{\beta}_k| + \lambda_n \frac{\hat{\beta}_k}{|\hat{\beta}_k|^r} > 0,
\end{aligned}$$

where the last inequality holds as $\sum_j \sum_i \|x_{ij}\| \leq 2 \sum_j \sum_i \|x\| = O_p(n)$ by Assumption 3, and $\frac{\lambda_n}{n|\hat{\beta}_k|^r} \geq n^{\frac{r}{2}-1} \lambda_n \rightarrow \infty$. It contradicts the fact that $L(\hat{\beta}, \hat{w}) \leq L(\hat{\beta}^*, \hat{w})$. Then $\forall j \notin \mathcal{A}, P(j \in \hat{\mathcal{A}}) \rightarrow 0$, thus we have $P(\hat{\mathcal{A}} = \mathcal{A}) \rightarrow 1$, as $n \rightarrow \infty$.

7.2 Lemmas

The following nine lemmas are used in the proof of main theorems.

Lemma 7.2.1. *Under assumptions 3.1.1-3.1.7, if $J_n \rightarrow \infty, J_n \ll n$, we have*

$$\sqrt{n}\{\tilde{\beta}_{F_0, \gamma_0} - \beta^0\} \xrightarrow{D} N(0, \Gamma^+ V \Gamma^+),$$

where Γ and V are defined in Theorem 3.2.2, and $\tilde{\beta}_{F_0, \gamma_0}$ is the estimator of $\tilde{\beta}$ with substituting true F_0 and γ_0 .

Proof of Lemma 7.2.1:

Let

$$\begin{aligned} \tilde{L}_{\tau n}^*(\beta) = & \frac{1}{n} \sum_{i=1}^n \left\{ \frac{\tau - F_0(C_i|z_{0i})}{1 - F_0(C_i|z_{0i})} I(T_i > C_i, F_0(C_i|z_{0i}) \leq \tau) \rho_\tau \left(C_i - B(X_i^T \beta) \tilde{\theta} \right) \right. \\ & \left. + I(T_i \leq C_i) \rho_\tau \left(T_i - B(X_i^T \beta) \tilde{\theta} \right) \right. \\ & \left. + \left[I(T_i > C_i) - \frac{\tau - F_0(C_i|z_{0i})}{1 - F_0(C_i|z_{0i})} I(T_i > C_i, F_0(C_i|z_{0i}) \leq \tau) \right] \rho_\tau \left(Y^{+\infty} - B(X_i^T \beta) \tilde{\theta} \right) \right\}, \end{aligned}$$

where $\tilde{\beta}_\tau$ is the minimizer of $E\{L_{\tau n}(\theta, \beta)\}$. Denote $D_{\tau n}(b) = L_{\tau n}^*(\beta^0 + b) - L_{\tau n}^*(\beta^0)$, and $\tilde{D}_{\tau n}(b) = \tilde{L}_{\tau n}^*(\beta^0 + b) - \tilde{L}_{\tau n}^*(\beta^0)$, and $\hat{\beta}$ be the minimizer of $L_{\tau n}^*(\beta)$. First of all, standardization doesn't change the minimization. That's to say, if we set $\tilde{\beta}_\tau^* = \frac{\hat{\beta}_\tau}{\|\hat{\beta}_\tau\|_2}$, then $\tilde{\beta}_\tau^*$ is the minimizer of $L_{\tau n}^*(\beta)$. In the following, we will show that $\sqrt{n}\|\tilde{\beta}_\tau^* - \beta^0\|_2 = O_p(1)$ and $\sqrt{n}(\tilde{\beta}_\tau^* - \beta^0)$ is asymptotic normal.

STEP 1, Decompose $\tilde{\beta}_\tau^* = a_{\tau n}\beta^0 + s_{\tau n}^*$ such that $\eta \perp \beta^0$, $\|\eta\|_2 = 1$, $a_{\tau n}^2 + s_{\tau n}^{*2} = 1$. Note that $\tilde{\beta}_\tau^*$ is a consistent estimator of β^0 , then $|a_{\tau n} - 1| = o_p(1)$, $s_{\tau n}^* = o_p(1)$, $a_{\tau n} = \sqrt{1 - s_{\tau n}^{*2}}$. Then $a_{\tau n}^{-1}\hat{\beta}_\tau = \beta^0 + s_{\tau n}\eta$, where $s_{\tau n} = a_{\tau n}^{-1}s_{\tau n}^*$. Next, we will show $s_{\tau n} = O_p(n^{-1/2})$, which implies that $s_{\tau n}^*O_p(n^{-1/2})$ and $1 - a_{\tau n} = 1 - \sqrt{1 - s_{\tau n}^{*2}} = \sqrt{s_{\tau n}^{*2}}1 + \sqrt{1 - s_{\tau n}^{*2}} = O(s_{\tau n}^{*2}) = O_p(n^{-1})$, then

$$\|\tilde{\beta}_\tau - \beta^0\|_2 = \sqrt{(1 - a_{\tau n})^2 + s_{\tau n}^{*2}} = O_p(n^{-1/2}).$$

Denote $\hat{\beta}_\tau = a_{\tau n}^{-1}\tilde{\beta}_\tau$, then $\hat{b}_\tau = \hat{\beta}_\tau - \beta^0 = s_{\tau n}\eta$, so \hat{b}_τ minimizes $D_{\tau n}(b)$. Write $c_n(b) = (1 + \sqrt{n}\|b\|_2)^{-1}\sqrt{nb}$,

$$\begin{aligned} \Omega_n = & -\frac{1}{n} \sum_{i=1}^n \left\{ \frac{\tau - F_0(C_i|z_{0i})}{1 - F_0(C_i|z_{0i})} I(T_i > C_i, F_0(C_i|z_{0i}) \leq \tau) \rho'_\tau(C_i - g(x_i^T \beta^0)) \right. \\ & \left. + I(T_i \leq C_i) \rho'_\tau(T_i - g(x_i^T \beta^0)) \right. \\ & \left. + \left[E_T(M(t|x)) - E_C(1 - F_0(C_i|z_{0i})) \frac{\tau - F_0(C_i|z_{0i})}{1 - F_0(C_i|z_{0i})} \right] \tau \right\} g'(x_i^T \beta^0) \tilde{x}_i, \end{aligned}$$

$$\begin{aligned} \Gamma_n &= \frac{1}{n} \sum_{i=1}^n \left\{ E_T \left[(1 - M(T_i|x)) f_0(g(x_i^T \beta^0)) \right] \right. \\ &\quad \left. + E_C \left[(\tau - F_0(C_i|z_{0i})) I(F_0(C_i|z_{0i}) \leq \tau) m_0(g(x_i^T \beta^0)) \right] \right\} [g'(x_i^T \beta^0)]^2 \tilde{x}_i \tilde{x}_i^T, \end{aligned}$$

Apply the weak law of large numbers to $c_n^T(\hat{b}_{\tau n}) \sqrt{n} \Omega_n$, we have

$$c_n^T(\hat{b}_{\tau n}) \sqrt{n} \Omega_n = O_p \left\{ \left[c_n^T(\hat{b}_{\tau n}) E(\tilde{X} \tilde{X}^T) c_n^T(\hat{b}_{\tau n}) \right]^{1/2} \right\} = O_p \left\{ \|c_n^T(\hat{b}_{\tau n})\|_2 \right\},$$

since $E(\tilde{X} \tilde{X}^T)$ is bounded from infinity by that X is distributed on a compact set.

From Lemma 7.2.8 and Lemma 7.2.9, we have

$$D_{\tau n}(\hat{b}_{\tau n}) = \Omega_n \hat{b}_{\tau n} + \frac{1}{2} \hat{b}_{\tau n}^T \Gamma_n \hat{b}_{\tau n} + o_p(n^{-1/2} \|\hat{b}_{\tau n}\|_2) + O_p((\log(n)) \frac{J_n}{n} \|\hat{b}_{\tau n}\|_2^{1/2}) + o_p(n^{-1}).$$

Since $2n^{-1/2} \|\hat{b}_{\tau n}\|_2 \leq n^{-1} + \|\hat{b}_{\tau n}\|_2^2$, then

$$\begin{aligned} 2 \log(n) \frac{J_n}{n} \|\hat{b}_{\tau n}\|_2^{1/2} &\leq \|\hat{b}_{\tau n}\|_2 J_n^2 / n [\log(n)]^{2+2/10} + n^{-1} [\log(n)]^{-2/10} \\ &\leq 0.5 \|\hat{b}_{\tau n}\|_2^2 J_n^4 n^{-1} [\log(n)]^{4+6/10} + 0.5 n^{-1} [\log(n)]^{-2/10} + n^{-1} [\log(n)]^{-5/4} \\ &= o(1) \|\hat{b}_{\tau n}\|_2^2 + o(n^{-1}). \end{aligned}$$

By the assumption $J_n \ll n^{1/4} [\log(n)]^{-5/4}$, then

$$D_{\tau n}(\hat{b}_{\tau n}) = \Omega_n \hat{b}_{\tau n} + \frac{1}{2} \hat{b}_{\tau n}^T \Gamma_n \hat{b}_{\tau n} + o_p(1) \|\hat{b}_{\tau n}\|_2^2 + o_p(n^{-1}).$$

By definition of $\hat{b}_{\tau n}$, we have $D_{\tau n}(\hat{b}_{\tau n}) \leq 0$. Multiply both sides of above by $n(1 +$

$\sqrt{n}||\hat{b}_{\tau n}||_2^{-2}$), we have

$$c_n^T(\hat{b}_{\tau n})\sqrt{n}\Omega_n(1 + \sqrt{n}||\hat{b}_{\tau n}||_2)^{-1} + \frac{1}{2}c_n^T(\hat{b}_{\tau n})\Omega_n c_n(\hat{b}_{\tau n}) + o_p(1) + o_p(1)(1 + \sqrt{n}||\hat{b}_{\tau n}||_2^{-2}) \leq 0.$$

Let n_k be a subsequence that $\sqrt{n_k}s_{\tau n_k} \rightarrow \infty$, then we have $||c_{n_k}(\hat{b}_{\tau n_k})||_2 \asymp 1$, so that $c_{n_k}^T(\hat{b}_{\tau n_k})\sqrt{n_k}\Omega_{n_k}(1 + \sqrt{n_k}||\hat{b}_{\tau n_k}||_2)^{-1} = o_p(1)$. By condition (C2) and (C3), we have

$$c_{n_k}^T(\hat{b}_{\tau n_k})\Gamma c_{n_k}(\hat{b}_{\tau n_k}) = o_p(1).$$

Since $\hat{b}_{\tau n} = s_{\tau n}\eta$, then $c_{n_k}^T(\hat{b}_{\tau n_k})\Gamma c_{n_k}(\hat{b}_{\tau n_k}) = (1 + \sqrt{n_k}|s_{\tau, n_k}|)^{-2}n_k s_{\tau n_k}^2 \eta^T \Gamma \eta > 0$. It's a contradiction. Thus, $\sqrt{n}|s_{\tau n}| = O_p(1)$, which indicates $s_{\tau n} = O_p(n^{-1/2})$. So, $\sqrt{n}||\tilde{\beta}_{\tau}^* - \beta^0||_2 = O_p(1)$.

STEP 2 , Since $||\hat{b}_{\tau n}||_2 = O_p(n^{-1/2})$, then we can write $D_{\tau n}(\hat{b}_{\tau n})$ as

$$D_{\tau n}(\hat{b}_{\tau n}) = \hat{b}_{\tau n}^T \Omega_n + \frac{1}{2}\hat{b}_{\tau n}^T (\Gamma_n + \xi_n I) \hat{b}_{\tau n} + o_p(n^{-1}),$$

for any $\xi_n = o(1)$ and $\xi_n \notin \sigma(\Gamma_n)$, where $\sigma(\Gamma_n)$ is the spectrum of a square matrix Γ_n . Then $(\Gamma_n + \xi_n I)^{-1}$ exists. Define

$$D_{\tau n}^*(b) = b^T \Omega_n + \frac{1}{2}b^T (\Gamma_n + \xi_n I) b,$$

then $\tilde{b}_{\tau n} = -(\Gamma_n + \xi_n I)^{-1} \Omega_n$ minimizes $D_{\tau n}^*(b)$. Moreover, $||\tilde{b}_{\tau n}||_2 = O_p(n^{-1/2})$. Let $\alpha_{1n}, \dots, \alpha_{rn}$ be the distinct eigenvalues of Γ_n , then $\Gamma_n = \sum_{i=1}^r \alpha_{in} s_{in}$, where s_{in} are the eigenvectors of Γ_n . Then

$$\lim_{x \rightarrow \infty} (\Gamma_n + \xi_n I)^{-1} = \lim_{x \rightarrow \infty} \sum_{i=1}^r (\alpha_{in} + \xi_n)^{-1} s_{in} = \lim_{x \rightarrow \infty} \sum_{i=1}^r \alpha_{in}^{-1} s_{in} = \lim_{x \rightarrow \infty} \Gamma_n^+ = \Gamma^+.$$

By central limit theorem, as $n \rightarrow \infty$, then $\sqrt{n}\tilde{b}_{\tau n} \rightarrow N(0, \Gamma^+ V \Gamma^+)$. The Γ and V are defined in Theorem 3.2.2.

Recall that $a_{\tau n}^{-1}\tilde{\beta}_\tau = \beta^0 + \hat{b}_{\tau n}$ and $1 - a_{\tau n} = O_p(n^{-1})$, if we can prove that

$$\|\hat{b}_{\tau n} - \tilde{b}_{\tau n}\|_2 = o_p(n^{-1/2}),$$

then by Slutsky's Theorem, we will have $\sqrt{n}(\tilde{\beta}_\tau - \beta^0) \rightarrow N(0, \Gamma^+ V \Gamma^+)$. So the following is to prove that $\|\hat{b}_{\tau n} - \tilde{b}_{\tau n}\|_2 = o_p(n^{-1/2})$. If there exists a subsequence n_k such that $\hat{b}_{\tau n_k} - \tilde{b}_{\tau n_k} = v_k n_k^{-1/2} u_{n_k}$, where $v_k \asymp 1$ and u_{n_k} is a unit vector. Since $D_{\tau n_k}^*(b)$ is convex, then for $0 < l < v_k$ and $l \asymp v_k$, we have

$$(1 - \frac{l}{v_k})D_{\tau n_k}^*(\tilde{b}_{\tau n_k}) + \frac{l}{v_k}D_{\tau n_k}^*(\hat{b}_{\tau n_k}) \geq D_{\tau n_k}^*(\tilde{b}_{\tau n_k} + \frac{l}{v_k}(\hat{b}_{\tau n_k})) = D_{\tau n_k}^*(\tilde{b}_{\tau n_k} + l n_k^{-1/2} u_{n_k}),$$

then

$$\frac{l}{v_k}[D_{\tau n_k}^*(\hat{b}_{\tau n_k}) - D_{\tau n_k}^*(\tilde{b}_{\tau n_k})] \geq D_{\tau n_k}^*(\tilde{b}_{\tau n_k} + l n_k^{-1/2} u_{n_k}) - D_{\tau n_k}^*(\tilde{b}_{\tau n_k}).$$

Since $D_{\tau n_k}^*(\hat{b}_{\tau n_k}) - D_{\tau n_k}^*(\tilde{b}_{\tau n_k}) = D_{\tau n_k}(\hat{b}_{\tau n_k}) - D_{\tau n_k}(\tilde{b}_{\tau n_k}) + o_p(n_k^{-1})$, then

$$\begin{aligned} n_k [D_{\tau n_k}^*(\hat{b}_{\tau n_k}) - D_{\tau n_k}^*(\tilde{b}_{\tau n_k})] &= n_k [D_{\tau n_k}(\hat{b}_{\tau n_k}) - D_{\tau n_k}(\tilde{b}_{\tau n_k})] + o(1) \\ &\geq n_k \frac{v_k}{l} [D_{\tau n_k}^*(\tilde{b}_{\tau n_k} + l n_k^{-1/2} u_{n_k}) - D_{\tau n_k}^*(\tilde{b}_{\tau n_k})] + o(1) \\ &= \frac{v_k}{l} n_k (l n_k^{1/2} u_{n_k})^T (\Gamma_{n_k} + \xi_{n_k} I) (l n_k^{1/2} u_{n_k}) + o(1) \\ &\geq C v_k l u_{n_k}^T \Gamma_{n_k} u_{n_k}, \end{aligned}$$

with probability approaching 1, where C is some constant. Similarly, decompose $u_{n_k} = h_{1n_k} \beta^0 + h_{2n_k} \eta$, where η is unit vector and orthogonal to β^0 . Then $\beta^{0T} u_{n_k} =$

h_{1n_k} . Since $\hat{b}_{\tau n_k} = s_{\tau n} \eta$, then

$$\beta^{0T} u_{n_k} = v_k - 1n_k 1/2v_k n^{-1/2} \beta^{0T} u_{n_k} = v_k - 1n_k 1/2 \beta^{0T} (\hat{b}_{\tau n_k} - \tilde{b}_{\tau n_k}) = -v_k^{-1} n_k^{1/2} \beta^{0T} \tilde{b}_{\tau n_k} = h_{1n_k}.$$

Also, we have with probability approaching 1, $Var(n_k^{1/2} \beta^{0T} \tilde{b}_{\tau n_k}) \rightarrow \beta^{0T} \Gamma^+ V \Gamma^+ \beta^0$.

Since $\Gamma \beta^0 = 0$, β^0 is an eigenvector of Γ and orthogonal to other eigenvectors of Γ , then $\Gamma^+ \beta^0 = 0$. Therefore, $n_k^{1/2} \beta^{0T} \tilde{b}_{\tau n_k} = o_p(1)$, which implies that $h_{1n_k} = o_p(1)$.

Then $h_{2n_k}^2 = 1 - o_p(1)$, hence with probability approaching 1,

$$n_k \left[D_{\tau n_k}^* (\hat{b}_{\tau n_k}) - D_{\tau n_k}^* (\tilde{b}_{\tau n_k}) \right] \geq C v_k l u_{n_k}^T \Gamma_{n_k} u_{n_k} = C v_k l h_{2n_k}^2 \eta^T \Gamma_{n_k} \eta > 0$$

This contradicts with the fact that $\hat{b}_{\tau n_k}$ minimizes $D_{\tau n_k}(b)$, thus $\|\hat{b}_{\tau n_k} - \tilde{b}_{\tau n_k}\|_2 = o_p(n^{-1/2})$.

Lemma 7.2.2. *Under Assumption 3.1.1, and $(\log n)J_n \ll n$, then for $\beta \in \Theta$,*

$$\sup_{1 \leq j, j' \leq J_n} \left| 1/n \sum_{i=1}^n B_j(X_i^T \beta) B_{j'}(X_i^T \beta) - E [B_j(X_i^T \beta) B_{j'}(X_i^T \beta)] \right| = O_{a.s.} \left(\sqrt{\log(n)/(nJ_n)} \right)$$

Lemma 7.2.2 can be proved by Bernstein's inequality given in Bosq (1998).

Lemma 7.2.3. *Under Assumption 3.1.1, there exists constants $0 < c_1 \leq C_1 < \infty$, such that for sufficiently large n and for any j, j' , for $\beta \in \Theta$,*

$$c_1 J_n^{-1} \leq \lambda_{\min} [E \{B(X_i^T \beta) B(X_i^T \beta)^T\}] \leq \lambda_{\max} [E \{B(X_i^T \beta) B(X_i^T \beta)^T\}] \leq C_1 J_n^{-1},$$

where λ_{\min} and λ_{\max} denote the minimal and maximal eigenvalues of matrix.

Lemma 7.2.3 can be derived from Theorem 5.4.2 from DeVore and Lorentz (1993).

Lemma 7.2.4. *Under Assumptions 3.1.6 and 3.1.7, if $J_n \rightarrow \infty$, and $J_n^{2l+1} \ll n$, then for $l = 0, 1$,*

$$\sup_{\beta \in \Theta} \left\{ \frac{1}{n} \sum_{i=1}^n [\tilde{g}_n^{(l)}(X_i^T \beta, \beta) - \tilde{g}^{(l)}(X_i^T \beta, \beta)]^2 \right\} = O_p(J_n^{-2(\zeta-l)} + J_n^{2l+1}/n)$$

Proof of Lemma 7.2.4:

First, by Corollary of Schumaker (1981), for any given $\beta \in \Theta$, there exists $\theta_\tau^0(\beta) \in R^{J_n}$ such that $g_\tau^0(\mu, \beta) = B(\mu)^T \theta_\tau^0(\beta) \in \mathcal{H}_\tau$ and

$$\sup_{\mu \in [a_0, b_0]} |R_\tau(\mu, \beta)| = \sup_{\mu \in [a_0, b_0]} |\tilde{g}_\tau(\mu, \beta) - g_\tau^0(\mu, \beta)| \leq |\tilde{C}_m(\beta)| J_n^{-r},$$

where the continuous function $\tilde{C}_m(\beta)$ depends on the m and C_0 in the Holder condition. Let $Y_i^*(\beta) = Y_i - \tilde{g}_\tau(X_i^T \beta, \beta) + R_\tau(X_i^T \beta, \beta)$, where $R_\tau(\mu, \beta) = \tilde{g}_\tau(\mu, \beta) - g_\tau^0(\mu, \beta)$. Then $L_{\tau n}(\theta_\tau, \beta)$ can be rewritten as

$$L_{\tau n}(\theta_\tau, \beta) = \frac{1}{n} \sum_{i=1}^n \left\{ \rho_\tau \left(Y_i^*(\beta) - B(X_i^T \beta)^T (\theta_\tau - \theta_\tau^0(\beta)) \right) \right\}.$$

Define $D_{\tau n}(\theta_\tau, \beta) = \frac{1}{n} \sum_{i=1}^n D_{\tau n, i}(\theta_\tau, \beta) = L_{\tau n}(\theta_\tau, \beta) - \frac{1}{n} \sum_{i=1}^n \rho_\tau(Y_i^*(\beta))$, and $\tilde{g}_{\tau n}^{(l)}(\mu, \beta) = B^{(j)}(\mu)^T \tilde{\theta}_\tau(\beta)$, with $\tilde{\theta}_\tau(\beta)$ obtained by minimizing $L_{\tau n}(\theta, \beta)$. From Lemma 3.2 of He and Shi (1994), for $L_n \asymp J_n n^{-1/2} + J_n^{-r+1/2}$, we have

$$P \left\{ \sup_{\|\theta_\tau - \theta_\tau^0(\beta)\|_2 = L_n} \frac{J_n}{n} \left| D_{\tau n}(\theta_\tau, \beta) - E \left(D_{\tau n}(\theta_\tau, \beta) | U(\beta) \right) + \frac{1}{n} \sum_{i=1}^n B(X_i^T \beta)^T (\theta_\tau - \theta_\tau^0(\beta)) \right. \right. \\ \left. \left. + \left\{ \frac{\tau - F_0(C_i | z_{0i})}{1 - F_0(C_i | z_{0i})} I(T_i > C_i, F_0(C_i | z_{0i}) \leq \tau) (\tau - I(C_i < \tilde{g}(X_i^T \beta, \beta))) \right. \right. \right. \\ \left. \left. \left. + I(T_i \leq C_i) [\tau - I(T_i < \tilde{g}(X_i^T \beta, \beta))] + \left[I(T_i > C_i) - \frac{\tau - F_0(C_i | z_{0i})}{1 - F_0(C_i | z_{0i})} \right] \tau \right\} \right| \geq \epsilon \right\} \rightarrow 0$$

Secondly, for any $\epsilon > 0$, there is a sufficiently large $C = C_\epsilon$ such that

$$P \left\{ J_n^{-1} \inf_{\|\theta_\tau - \theta_\tau^0(\beta)\|_2 = L_n} E \left(D_{\tau n}(\theta_\tau, \beta) | U(\beta) \right) - C \frac{1}{n} \sum_{i=1}^n \left| B(X_i^T \beta)^T (\theta_\tau - \theta_\tau^0(\beta)) \right. \right. \\ \left. \left. \left\{ \frac{\tau - F_0(C_i | z_{0i})}{1 - F_0(C_i | z_{0i})} I(T_i > C_i, F_0(C_i | z_{0i}) \leq \tau) (\tau - I(C_i < \tilde{g}(X_i^T \beta, \beta))) \right. \right. \right. \\ \left. \left. \left. + I(T_i \leq C_i) [\tau - I(T_i < \tilde{g}(X_i^T \beta, \beta))] + \left[I(T_i > C_i) - \frac{\tau - F_0(C_i | z_{0i})}{1 - F_0(C_i | z_{0i})} \tau \right] \right\} \right| > 1 \right\} > 1 - \epsilon$$

By Bernstein's inequality, we have

$$\sup_{\beta \in \Theta} \frac{1}{n} \sum_{i=1}^n \left| B(X_i^T \beta)^T (\theta_\tau - \theta_\tau^0(\beta)) \left\{ \frac{\tau - F_0(C_i | z_{0i})}{1 - F_0(C_i | z_{0i})} I(T_i > C_i, F_0(C_i | z_{0i}) \leq \tau) \right. \right. \\ \left. \left. (\tau - I(C_i < \tilde{g}(X_i^T \beta, \beta))) + I(T_i \leq C_i) [\tau - I(T_i < \tilde{g}(X_i^T \beta, \beta))] \right. \right. \\ \left. \left. + \left[I(T_i > C_i) - \frac{\tau - F_0(C_i | z_{0i})}{1 - F_0(C_i | z_{0i})} \tau \right] \right\} \right| \\ = o_p(1)$$

Then we have

$$P \left\{ \inf_{\|\theta_\tau - \theta_\tau^0(\beta)\|_2 = L_n} D_{\tau n}(\theta_\tau, \beta) > 0, \text{ for all } \beta \in \Theta \right\} > 1 - \epsilon.$$

By definition of $D_{\tau n}(\theta_\tau, \beta)$, Corollary 25 from Eggleston (1958), then

$$P \left\{ \inf_{\|\theta_\tau - \theta_\tau^0(\beta)\|_2 \geq L_n} \frac{1}{n} \sum_{i=1}^n \left\{ \rho_\tau \left(Y_i^*(\beta) - B(X_i^T \beta)^T (\theta_\tau - \theta_\tau^0(\beta)) \right) \right\} \right. \\ \left. > n^{-1} \sum_{i=1}^n \rho_\tau \left(Y_i^*(\beta) \right), \text{ for all } \beta \in \Theta \right\} > 1 - \epsilon$$

It implies that $P \left\{ \sup_{\theta_\tau \in \Theta} \|\theta_\tau - \theta_\tau^0(\beta)\|_2 \leq L_n \right\} > 1 - \epsilon$. Then by Lemma 7.2.2 and

Lemma 7.2.3, and above statement, we have with probability approaching 1,

$$\begin{aligned}
& \sup_{\beta \in \Theta} \left\{ \frac{1}{n} \sum_{i=1}^n [\tilde{g}_{\tau n}(X_i^T \beta, \beta) - \tilde{g}_{\tau}(X_i^T \beta, \beta)]^2 \right\} \\
& \leq 2 \sup_{\beta \in \Theta} \left\{ \frac{1}{n} \sum_{i=1}^n [\tilde{g}_{\tau n}(X_i^T \beta, \beta) - g_{\tau}^0(X_i^T \beta, \beta)]^2 \right\} \\
& \quad + 2 \sup_{\beta \in \Theta} \left\{ \frac{1}{n} \sum_{i=1}^n [g_{\tau}^0(X_i^T \beta, \beta) - \tilde{g}_{\tau}(X_i^T \beta, \beta)]^2 \right\} \\
& \asymp J_n^{-1} \sup_{\beta \in \Theta} \|\theta_{\tau} - \theta_{\tau}^0(\beta)\|_2^2 + O(J_n^{-2r}) \\
& = O_p(J_n^{-2(r-l)} + J_n^{2l+1}/n)
\end{aligned}$$

And the result for $l = 1$ follows from above and using the techniques in Zhou and Wolfe (2000) for spline derivative estimation.

Lemma 7.2.5. *Under the same assumptions of Theorem 3.2.2, then,*

$$\begin{aligned}
& \sup_{1 \leq i \leq n} \left| B(X_i^T \beta^0)^T \frac{\partial \tilde{\theta}_{\tau}(\beta^0)}{\partial \beta} + [g_{\tau}^{(1)}(X_i^T \beta^0) E(X_i^T | X_i^T \beta^0)] \right| \\
& = O_p(J_n^{1-\zeta} + J_n^{-1} + (\log(n))^{1/2} J_n^{1/2} n^{-1/2})
\end{aligned}$$

Proof of Lemma 7.2.5:

Since $\tilde{\theta}_{\tau}(\beta)$ is the minimizer of $E(L_{\tau n}^*(\theta_{\tau}, \beta) | X)$, where

$$\begin{aligned}
& L_{\tau n}^*(\theta_{\tau}, \beta) \\
& = \frac{1}{n} \sum_{i=1}^n \left\{ I(T_i \leq C_i) \left[\rho_{\tau}(T_i - B(X_i^T \beta) \theta_{\tau}) - \rho_{\tau}(T_i) \right] \right. \\
& \quad + \frac{\tau - F_0}{1 - F_0} I(T_i > C_i, F_0 \leq \tau) \left[\rho_{\tau}(C_i - B(X_i^T \beta) \theta_{\tau}) - \rho_{\tau}(C_i) \right] \\
& \quad \left. - \left[I(T_i > C_i) - \frac{\tau - F_0}{1 - F_0} I(T_i > C_i, F_0 \leq \tau) \right] \tau B(X_i^T \beta) \theta_{\tau} \right\}
\end{aligned}$$

With the identity given in Knight (1998) that

$$\rho_\tau(s-t) - \rho_\tau(s) = t\{I(s \leq 0) - \tau\} + \int_0^t \{I(s \leq x) - I(s \leq 0)\}dx,$$

we have

$$\begin{aligned} & L_{\tau n}^*(\theta_\tau, \beta) \\ &= \frac{1}{n} \sum_{i=1}^n \left\{ I(T_i \leq C_i) \left[\int_0^{B(X_i^T \beta) \theta_\tau} I(T_i \leq t) dt - \tau B(X_i^T \beta) \theta_\tau \right] \right. \\ &\quad \left. + \frac{\tau - F_0}{1 - F_0} I(T_i > C_i, F_0 \leq \tau) \left[\int_0^{B(X_i^T \beta) \theta_\tau} I(C_i \leq t) dt - \tau B(X_i^T \beta) \theta_\tau \right] \right. \\ &\quad \left. - \left[I(T_i > C_i) - \frac{\tau - F_0}{1 - F_0} I(T_i > C_i, F_0 \leq \tau) \right] \tau B(X_i^T \beta) \theta_\tau \right\} \\ &= \frac{1}{n} \sum_{i=1}^n \left\{ I(T_i \leq C_i) \int_0^{B(X_i^T \beta) \theta_\tau} I(T_i \leq t) dt \right. \\ &\quad \left. + \frac{\tau - F_0}{1 - F_0} I(T_i > C_i, F_0 \leq \tau) \int_0^{B(X_i^T \beta) \theta_\tau} I(C_i \leq t) dt \right. \\ &\quad \left. - \tau B(X_i^T \beta) \theta_\tau \right\} \end{aligned}$$

Then we have

$$\begin{aligned} & E\{L_{\tau n}^*(\theta_\tau, \beta) | X\} \\ &= \frac{1}{n} \sum_{i=1}^n \left\{ \int_0^{B(X_i^T \beta) \theta_\tau} \int_{-\infty}^t \left[(1 - M_0(u | X_i)) f_0(u | X_i) + (\tau - F_0(u | X_i))^+ m_0(u | X_i) \right] du dt \right. \\ &\quad \left. - \tau B(X_i^T \beta) \theta_\tau \right\} \end{aligned}$$

Since that $\tilde{\theta}_\tau(\beta)$ is the minimizer of $E\{L_{\tau n}^*(\theta_\tau, \beta)|X\}$ for given β , then

$$\begin{aligned} 0 &= \frac{\partial E\{L_{\tau n}^*(\theta_\tau, \beta)|X\}}{\partial \theta_\tau} \Big|_{\theta_\tau = \tilde{\theta}_\tau(\beta)} \\ &= \frac{1}{n} \sum_{i=1}^n B(X_i^T \beta) \left\{ \int_{-\infty}^{B(X_i^T \beta)\theta_\tau} \left[(1 - M_0(u|X_i))f_0(u|X_i) + (\tau - F_0(u|X_i))^+ m_0(u|X_i) \right] du - \tau \right\} \end{aligned}$$

Since $\frac{\partial E\{L_{\tau n}^*(\theta_\tau, \beta)|X\}}{\partial \theta_\tau}$ is differentiable to θ_τ and β , and

$$\begin{aligned} &\frac{\partial^2 E\{L_{\tau n}^*(\theta_\tau, \beta)|X\}}{\partial \theta_\tau \partial \theta_\tau^T} \Big|_{\theta_\tau = \tilde{\theta}_\tau(\beta)} \\ &= \frac{1}{n} \sum_{i=1}^n B(X_i^T \beta) B^T(X_i^T \beta) \left[(1 - M_0(u|X_i))f_0(u|X_i) + (\tau - F_0(u|X_i))^+ m_0(u|X_i) \right] \Big|_{B(X_i^T \beta)\theta_\tau} \\ &\neq 0, \end{aligned}$$

by Implicit Function Theorem, we can find the $\frac{\partial \tilde{\theta}_\tau(\beta^0)}{\partial \beta^T}$ through the following equation.

$$\begin{aligned} 0 &= \frac{\partial^2 E\{L_{\tau n}^*(\tilde{\theta}_\tau(\beta^0), \beta^0)|X\}}{\partial \theta_\tau \partial \beta^T} \\ &= \Delta_1 + \Delta_2 + \Delta_3 \frac{\partial \tilde{\theta}_\tau(\beta^0)}{\partial \beta^T}, \end{aligned}$$

where

$$\begin{aligned}\Delta_1 &= \frac{1}{n} B^{(1)}(X_i^T \beta^0) X_i^T \cdot \\ &\quad \sum_{i=1}^n \left\{ \int_{-\infty}^{B(X_i^T \beta^0) \tilde{\theta}_\tau(\beta^0)} \left[(1 - M_0(u|X_i)) f_0(u|X_i) + (\tau - F_0(u|X_i))^+ m_0(u|X_i) \right] du - \tau \right\}, \\ \Delta_2 &= B^{(1)}(X_i^T \beta^0) \tilde{\theta}_\tau(\beta^0) X_i^T \cdot \\ &\quad \frac{1}{n} \sum_{i=1}^n B(X_i^T \beta^0) \left[(1 - M_0(u|X_i)) f_0(u|X_i) + (\tau - F_0(u|X_i))^+ m_0(u|X_i) \right] \Big|_{u=B(X_i^T \beta^0) \tilde{\theta}_\tau(\beta^0)}, \\ \Delta_3 &= B^T(X_i^T \beta^0) \cdot \\ &\quad \frac{1}{n} \sum_{i=1}^n B(X_i^T \beta^0) \left[(1 - M_0(u|X_i)) f_0(u|X_i) + (\tau - F_0(u|X_i))^+ m_0(u|X_i) \right] \Big|_{u=B(X_i^T \beta^0) \tilde{\theta}_\tau(\beta^0)}.\end{aligned}$$

Then we have

$$\frac{\partial \tilde{\theta}_\tau(\beta^0)}{\partial \beta^T} = -\Delta_3^{-1} (\Delta_1 + \Delta_2).$$

Since we have that

$$\begin{aligned}\sup_{u \in [a_0, b_0]} \left| B(u) \tilde{\theta}_\tau(\beta^0) - g_\tau(u) \right| &= O_p(J_n^{-r}), \\ \sup_{u \in [a_0, b_0]} \left| B^{(1)}(u) \tilde{\theta}_\tau(\beta^0) - g_\tau^{(1)}(u) \right| &= O_p(J_n^{1-r})\end{aligned}$$

then

$$\begin{aligned}&\sup_{1 \leq i \leq n} \left\{ \int_{-\infty}^{B(X_i^T \beta^0) \tilde{\theta}_\tau(\beta^0)} \left[(1 - M_0(u|X_i)) f_0(u|X_i) + (\tau - F_0(u|X_i))^+ m_0(u|X_i) \right] du - \tau \right\} \\ &= O_p(J_n^{-r}).\end{aligned}$$

Moreover, by the Bernstein's inequality, we can prove that

$$\begin{aligned} \sup_{j,k} \frac{1}{n} \sum_{i=1}^n \left| B^{(1)}(X_i^T \beta^0) X_{ik}^T \right| &= O_p(1), \\ \sup_{j,k} \frac{1}{n} \sum_{i=1}^n \left| B(X_i^T \beta^0) X_{ik}^T \right| &= O_p(J_n^{-1}). \end{aligned}$$

Denote that

$$\begin{aligned} \Delta_4 &= g_\tau^{(1)}(X_i^T \beta^0) X_i^T. \\ &\frac{1}{n} \sum_{i=1}^n B(X_i^T \beta^0) \left[(1 - M_0(u|X_i)) f_0(u|X_i) + (\tau - F_0(u|X_i))^+ m_0(u|X_i) \right] \Big|_{u=g_\tau(X_i^T \beta^0)}, \\ \Delta_5 &= B^T(X_i^T \beta^0). \\ &\frac{1}{n} \sum_{i=1}^n B(X_i^T \beta^0) \left[(1 - M_0(u|X_i)) f_0(u|X_i) + (\tau - F_0(u|X_i))^+ m_0(u|X_i) \right] \Big|_{u=g_\tau(X_i^T \beta^0)} \end{aligned}$$

The above results imply that

$$\begin{aligned} |\Delta_1| &\leq O_p(J_n^{-r}), \\ |\Delta_2 - \Delta_4| &\leq O_p(J_n^{-r}). \end{aligned}$$

Hence, we have

$$|\Delta_1 + \Delta_2 - \Delta_4| = O_p(J_n^{-r}).$$

Therefore, by the above results, we have the following,

$$\begin{aligned} \left| \frac{\partial \tilde{\theta}_\tau(\beta^0)}{\partial \beta^T} - \Delta_5^{-1} \Delta_4 \right| &\leq \|\Delta_3\|_\infty^{-1} \cdot |\Delta_1 + \Delta_2 - \Delta_4| + \|\Delta_3^{-1} - \Delta_5^{-1}\|_\infty \cdot |\Delta_4| \\ &= O_p(J_n) O_p(J_n^{-r}) + O_p(J_n^2) O_p(J_n^{-r}) O_p(J_n^{-1}) = O_p(J_n^{1-r}). \end{aligned}$$

Define $\hat{E}^* \left(g_\tau^{(1)}(X_i^T \beta^0) X_i^T | X_i^T \beta^0 \right) = B^T(X_i^T \beta^0) [\Delta_4^{-1} \Delta_5]$, which is the B-spline estimator of $g_\tau^{(1)}(X_i^T \beta^0) E^* \left(X_i^T | X_i^T \beta^0 \right)$, where

$$E^* \left(X_i^T | X_i^T \beta^0 \right) = \frac{E \left\{ \left[1 - M_0(g_\tau(X_i^T \beta^0)) \right] f_0(g_\tau(X_i^T \beta^0)) X_i | X_i^T \beta^0 \right\}}{E \left\{ \left[1 - M_0(g_\tau(X_i^T \beta^0)) \right] f_0(g_\tau(X_i^T \beta^0)) | X_i^T \beta^0 \right\}}.$$

By Lemmas 7.2.2 and 7.2.3, one has

$$\lambda_{max} \left[\frac{1}{n} \sum_{i=1}^n f_0(g_\tau(X_i^T \beta^0) | X_i) B(X_i^T \beta^0) B(X_i^T \beta^0)^T \right]^{-1} = O_p(J_n),$$

where $\lambda_{max}(A)$ means the largest eigenvalue of the matrix A. Hence,

$$\sup_{1 \leq i \leq n} \left| B(X_i^T \beta^0)^T \frac{\partial \tilde{\theta}_\tau(\beta^0)}{\partial \beta^T} - \left\{ -\hat{E}^* \left(g_\tau^{(1)}(X_i^T \beta^0) X_i^T | X_i^T \beta^0 \right) \right\} \right|^{-1} = O_p(J_n^{1-r}).$$

Moreover, by Assumption 7.2.8, we have

$$\begin{aligned} \sup_{1 \leq i \leq n} \left| \hat{E}^* \left(g_\tau^{(1)}(X_i^T \beta^0) X_i^T | X_i^T \beta^0 \right) - g_\tau^{(1)}(X_i^T \beta^0) \hat{E}^* \left(X_i^T | X_i^T \beta^0 \right) \right| \\ = O_p(J_n^{-1} + (\log n)^{1/2} J_n^{1/2} n^{-1/2}). \end{aligned}$$

Combine above two, we can prove the Lemma 7.2.5.

Lemma 7.2.6. *Under the same assumptions of Theorem 3.2.2, we have,*

$$\sup_{\beta \in \Theta} \|\tilde{\theta}_\tau(\beta) - \theta_\tau^0(\beta)\|_2 = O_p(J_n^{1/2-r})$$

Proof of Lemma 7.2.6

Since $\tilde{\theta}_\tau(\beta)$ minimizes $E\{L_{\tau n}(\theta_\tau, \beta)|X\}$, then for any $\beta \in \Phi$, we have

$$E\{L_{\tau n}(\tilde{\theta}_\tau(\beta), \beta)|X\} \leq E\{L_{\tau n}(\theta_\tau(\beta), \beta)|X\}.$$

That's to say,

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n E \left\{ I(T_i \leq C_i) \left[\rho_\tau(T_i - B(X_i^T \beta) \tilde{\theta}_\tau(\beta)) \right] \right. \\ & \quad + \frac{\tau - F_0}{1 - F_0} I(T_i > C_i, F_0 \leq \tau) \left[\rho_\tau(C_i - B(X_i^T \beta) \tilde{\theta}_\tau(\beta)) \right] \\ & \quad \left. + \left[I(T_i > C_i) - \frac{\tau - F_0}{1 - F_0} I(T_i > C_i, F_0 \leq \tau) \right] \rho_\tau(Y^{+\infty} - B(X_i^T \beta) \tilde{\theta}_\tau(\beta)) \right\} \\ & \leq \frac{1}{n} \sum_{i=1}^n E \left\{ I(T_i \leq C_i) \left[\rho_\tau(T_i - B(X_i^T \beta) \theta_\tau^0(\beta)) \right] \right. \\ & \quad + \frac{\tau - F_0}{1 - F_0} I(T_i > C_i, F_0 \leq \tau) \left[\rho_\tau(C_i - B(X_i^T \beta) \theta_\tau^0(\beta)) \right] \\ & \quad \left. + \left[I(T_i > C_i) - \frac{\tau - F_0}{1 - F_0} I(T_i > C_i, F_0 \leq \tau) \right] \rho_\tau(Y^{+\infty} - B(X_i^T \beta) \theta_\tau^0(\beta)) \right\} \end{aligned}$$

By the definition of $\tilde{g}(X_i^T \beta, \beta)$, we can have

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n E \left\{ I(T_i \leq C_i) \left[\rho_\tau(T_i - \tilde{g}_\tau(X_i^T \beta, \beta)) \right] \right. \\ & \quad + \frac{\tau - F_0}{1 - F_0} I(T_i > C_i, F_0 \leq \tau) \left[\rho_\tau(C_i - \tilde{g}_\tau(X_i^T \beta, \beta)) \right] \\ & \quad \left. + \left[I(T_i > C_i) - \frac{\tau - F_0}{1 - F_0} I(T_i > C_i, F_0 \leq \tau) \right] \rho_\tau(Y^{+\infty} - \tilde{g}_\tau(X_i^T \beta, \beta)) \right\} \\ & \leq \frac{1}{n} \sum_{i=1}^n E \left\{ I(T_i \leq C_i) \left[\rho_\tau(T_i - B(X_i^T \beta) \tilde{\theta}_\tau(\beta)) \right] \right. \\ & \quad + \frac{\tau - F_0}{1 - F_0} I(T_i > C_i, F_0 \leq \tau) \left[\rho_\tau(C_i - B(X_i^T \beta) \tilde{\theta}_\tau(\beta)) \right] \\ & \quad \left. + \left[I(T_i > C_i) - \frac{\tau - F_0}{1 - F_0} I(T_i > C_i, F_0 \leq \tau) \right] \rho_\tau(Y^{+\infty} - B(X_i^T \beta) \tilde{\theta}_\tau(\beta)) \right\}. \end{aligned}$$

By the above result and $\sup_{\beta \in \Theta} \frac{1}{n} \sum_{i=1}^n E \left\{ \tilde{g}_\tau(X_i^T \beta, \beta) - B(X_i^T \beta) \theta_\tau^0(\beta) \right\}^2 \leq C J_n^{-2r}$, then for some constant $0 < C < \infty$,

$$\sup_{\beta \in \Theta} \frac{1}{n} \sum_{i=1}^n E \left\{ \tilde{g}_\tau(X_i^T \beta, \beta) - B(X_i^T \beta) \tilde{\theta}_\tau(\beta) \right\}^2 \leq C J_n^{-2r}.$$

Therefore,

$$\sup_{\beta \in \Theta} \frac{1}{n} \sum_{i=1}^n \left\{ B(X_i^T \beta) \theta_\tau^0(\beta) - B(X_i^T \beta) \tilde{\theta}_\tau(\beta) \right\}^2 \leq O_p(J_n^{-2r}).$$

By the above result and Lemma 7.2.3, with probability 1 that

$$c_1 J_n^{-1} \leq \lambda_{\min} \left[\frac{1}{n} \sum_{i=1}^n \{ B(X_i^T \beta) B(X_i^T \beta)^T \} \right] \leq \lambda_{\max} \left[\frac{1}{n} \sum_{i=1}^n \{ B(X_i^T \beta) B(X_i^T \beta)^T \} \right] \leq C_1 J_n^{-1},$$

we have

$$\sup_{\beta \in \Theta} \|\tilde{\theta}_\tau(\beta) - \theta_\tau^0(\beta)\|_2 = O_p(J_n^{1/2-r})$$

Lemma 7.2.7. *Under the same assumptions of Theorem 3.2.2, we have,*

$$\sup_{b \in N_0} \left| \|b\|_2^{-1} \left\{ \tilde{D}_{\tau n}(b) - \left[\Omega_n b + \frac{1}{2} b^T (\Gamma_n + \xi_1 I) b \right] \right\} \right| = o_p(n^{-1/2})$$

where $\xi_1 = o_p(1)$ uniformly in $b \in N_0$.

Proof of Lemma 7.2.7

First, define $N_0 = \{b \in R^p : \|b\|_2 \leq \xi_n\}$ for any sequence of positive numbers $\xi_n = o(1)$. For the proof of following two lemmas, the arguments involving $o(\cdot)$, $o_p(\cdot)$, $O(\cdot)$ and $O_p(\cdot)$ hold in $b \in N_0$ uniformly. Let $T_i^* = T_i - \tilde{g}_n(x_i^T \beta^0, \beta^0)$ and $C_i^* = C_i - \tilde{g}_n(x_i^T \beta^0, \beta^0)$, where $\tilde{g}_n(x_i^T \beta^0, \beta^0) = B(x_i^T \beta^0) \tilde{\theta}_\tau(\beta^0)$.

Define

$$\begin{aligned}
\psi_{\tau n}(b) &= \frac{1}{n} \sum_{i=1}^n \psi_{\tau n, i}(b) \\
&= \frac{1}{n} \sum_{i=1}^n \left\{ \tilde{D}_{\tau n, i}(b) - E\left[\tilde{D}_{\tau n, i}(b) \mid X\right] \right. \\
&\quad + bI(T_i \leq C_i) \left[\rho_{\tau}^{(1)}(T_i^*) \frac{\partial \tilde{g}_n(X_i^T \beta^0, \beta^0)}{\partial \beta^T} - E\left(\rho_{\tau}^{(1)}(T_i^*) \frac{\partial \tilde{g}_n(X_i^T \beta^0, \beta^0)}{\partial \beta^T} \mid X\right) \right] \\
&\quad \left. + b \frac{\tau - F_0}{1 - F_0} I(T_i \geq C_i, F_0 \leq \tau) \left[\rho_{\tau}^{(1)}(C_i^*) \frac{\partial \tilde{g}_n(X_i^T \beta^0, \beta^0)}{\partial \beta^T} - E\left(\rho_{\tau}^{(1)}(C_i^*) \frac{\partial \tilde{g}_n(X_i^T \beta^0, \beta^0)}{\partial \beta^T} \mid X\right) \right] \right\}.
\end{aligned}$$

Here, $\tilde{D}_{\tau n}(b) = \tilde{L}_{\tau n}^*(\beta^0 + b) - \tilde{L}_{\tau n}^*(\beta^0)$. First, we want to show that

$$(i) \sup_{b \in N_0} (\|b\|_2^{-1} |\psi_{\tau n}(b)|) = o_p(n^{-1/2}),$$

$$\begin{aligned}
(ii) \sup_{b \in N_0} \|b\|_2^{-1} &\left| \frac{1}{n} \sum_{i=1}^n E\left[\tilde{D}_{\tau n, i}(b) \mid X\right] - \frac{1}{2} b^T \Gamma_n b \right. \\
&\quad - \frac{1}{n} \sum_{i=1}^n b I(T_i \leq C_i) \left[F_0(\tilde{g}_n(X_i^T \beta^0, \beta^0) \mid X) - \tau \right] \left[\frac{\partial \tilde{g}_n(X_i^T \beta^0, \beta^0)}{\partial \beta^T} \right] \\
&\quad \left. - \frac{1}{n} \sum_{i=1}^n b \frac{\tau - F_0}{1 - F_0} I(T_i \geq C_i, F_0 \leq \tau) \left[M_0(\tilde{g}_n(X_i^T \beta^0, \beta^0) \mid X) - \tau \right] \left[\frac{\partial \tilde{g}_n(X_i^T \beta^0, \beta^0)}{\partial \beta^T} \right] \right| = \xi_1 = o(1)
\end{aligned}$$

$$\begin{aligned}
(iii) \sup_{b \in N_0} &\left\{ \|b\|_2^{-1} \left| \frac{1}{n} \sum_{i=1}^n b \left\{ I(T_i \leq C_i) \rho_{\tau}^{(1)}(T_i^*) \left[\frac{\partial \tilde{g}_n(X_i^T \beta^0, \beta^0)}{\partial \beta^T} \right] \right. \right. \right. \\
&\quad \left. \left. + \frac{\tau - F_0}{1 - F_0} I(T_i \geq C_i, F_0 \leq \tau) \rho_{\tau}^{(1)}(C_i^*) \left[\frac{\partial \tilde{g}_n(X_i^T \beta^0, \beta^0)}{\partial \beta^T} \right] - T_n b \right\} \right\} = d_n = o_p(n^{-1/2})
\end{aligned}$$

Combine the above three results (i), (ii) and (iii), then

$$\sup_{b \in N_0} \left\| \|b\|_2^{-1} \left\{ \tilde{D}_{\tau n}(b) - \left[\Omega_n b + \frac{1}{2} b^T (\Gamma_n + \xi_1 I) b \right] \right\} \right\| \leq \sup_{b \in N_0} (\|b\|_2^{-1} |\psi_{\tau n}(b)|) + d_n = o_p(n^{-1/2})$$

Hence, the Lemma 7.2.7 is proved. First, let us prove (i). Let

$$r_i(\beta^0, b) = \tilde{g}_n(X_i^T(\beta^0 + b), \beta^0 + b) - \tilde{g}_n(X_i^T\beta^0, \beta^0),$$

by Assumptions, we have

$$\sup_X \left| r_i(\beta^0, b) - \left[\frac{\partial \tilde{g}_n(X_i^T\beta^0, \beta^0)}{\partial \beta^T} \right] b \right| \leq c_0 \|b\|_2^2.$$

We also know that $\sup_{1 \leq i \leq n} \left| B^{(1)}(X_i^T\beta^0)^T \tilde{\theta}_\tau(\beta^0) - g_n^{(1)}(X_i^T\beta^0) \right| = O_p(J_n^{1-r})$, from Lemma 7.2.5, we can have

$$\begin{aligned} & \sup_{1 \leq i \leq n} \left| \frac{\partial \tilde{g}_n(X_i^T\beta^0, \beta^0)}{\partial \beta^T} - g_n^{(1)}(X_i^T\beta^0) \tilde{X}_i^T \right| \\ &= \sup_{1 \leq i \leq n} \left| B^{(1)}(X_i^T\beta^0)^T \tilde{\theta}_\tau(\beta^0) X_i^T + B(X_i^T\beta^0)^T \frac{\partial \tilde{\theta}_\tau(\beta^0)}{\partial \beta^T} \right. \\ & \quad \left. - g_n^{(1)}(X_i^T\beta^0) X_i^T + g_n^{(1)}(X_i^T\beta^0) E^*(X_i^T | X_i^T\beta^0) \right| \\ &\leq \sup_{1 \leq i \leq n} \left| B^{(1)}(X_i^T\beta^0)^T \tilde{\theta}_\tau(\beta^0) - g_n^{(1)}(X_i^T\beta^0) \right| |X_i| \\ & \quad + \sup_{1 \leq i \leq n} \left| B(X_i^T\beta^0)^T \frac{\partial \tilde{\theta}_\tau(\beta^0)}{\partial \beta^T} + g_n^{(1)}(X_i^T\beta^0) E^*(X_i^T | X_i^T\beta^0) \right| \\ &= O_p((\log n)^{1/2} J_n^{1/2} n^{-1/2} + J_n^{1-r} + N_n^{-1}). \end{aligned}$$

By the above two results, then with probability approaching 1,

$$\begin{aligned} & \sup_{1 \leq i \leq n} \left| r_i(\beta^0, b) - g_n^{(1)}(X_i^T\beta^0) \tilde{X}_i^T b \right| \\ &\leq \sup_{1 \leq i \leq n} \left| \left[\frac{\partial \tilde{g}_n(X_i^T\beta^0, \beta^0)}{\partial \beta^T} \right] b - g_n^{(1)}(X_i^T\beta^0) \tilde{X}_i^T b \right| + c_0 \|b\|_2^2 \\ &\leq c_1 ((\log n)^{1/2} J_n^{1/2} n^{-1/2} + J_n^{1-r} + N_n^{-1}) \|b\|_2 + c_0 \|b\|_2^2, \end{aligned}$$

for constant $c_1 > 0$. Then for constant $c_2 > 0$, we have with probability approaching 1,

$$\sup_{1 \leq i \leq n} |r_i(\beta^0, b)| \leq c_2 \|b\|_2.$$

Denote

$$\begin{aligned} \Delta_{n,i}(b) &= \frac{\tau - F_0}{1 - F_0} I(T_i \geq C_i, F_0 \leq \tau) \left[\rho_\tau(C_i - B(X_i^T(\beta^0 + b)))^T \tilde{\theta}_\tau(\beta^0 + b) \right. \\ &\quad \left. - \rho_\tau(C_i - B(X_i^T \beta^0, \beta^0))^T \tilde{\theta}_\tau(\beta^0) + \rho_\tau^{(1)}(C_i^*) \frac{\partial \tilde{g}_n(X_i^T \beta^0, \beta^0)}{\partial \beta^T} \right] \\ &+ I(T_i \leq C_i) \left[\rho_\tau(T_i - B(X_i^T(\beta^0 + b)))^T \tilde{\theta}_\tau(\beta^0 + b) \right. \\ &\quad \left. - \rho_\tau(T_i - B(X_i^T \beta^0, \beta^0))^T \tilde{\theta}_\tau(\beta^0) + \rho_\tau^{(1)}(T_i^*) \frac{\partial \tilde{g}_n(X_i^T \beta^0, \beta^0)}{\partial \beta^T} \right] \\ &+ \tau \left[I(T_i \leq C_i) - \frac{\tau - F_0}{1 - F_0} I(T_i \geq C_i, F_0 \leq \tau) \right] \\ &\quad \left[B(X_i^T(\beta^0 + b))^T \tilde{\theta}_\tau(\beta^0 + b) - B(X_i^T \beta^0, \beta^0)^T \tilde{\theta}_\tau(\beta^0) \right] \end{aligned}$$

Then by definition of $\phi_{\tau n,i}(b)$, we know $\phi_{\tau n,i}(b) = \Delta_{n,i}(b) - E(\Delta_{n,i}(b)|X)$. With probability approaching 1, we have

$$\sup_{1 \leq i \leq n} E \left\{ \left[\|b\|_2^{-1} \Delta_{n,i}(b) \right]^2 \middle| X \right\} \leq c_3 \|b\|_2 \leq c_3 \delta_n,$$

for constant $c_3 > 0$. So, we can have the following,

$$\begin{aligned}
& \sup_{1 \leq i \leq n} E \left\{ \left[\|b\|_2^{-1} \phi_{\tau n, i}(b) \right]^2 \middle| X \right\} \\
&= \sup_{1 \leq i \leq n} E \left\{ \left[\|b\|_2^{-1} \left(\Delta_{n, i}(b) - E(\Delta_{n, i}(b) | X) \right) \right]^2 \middle| X \right\} \\
&\leq 2 \sup_{1 \leq i \leq n} E \left\{ \left[\|b\|_2^{-1} \Delta_{n, i}(b) \right]^2 \middle| X \right\} + 2 \sup_{1 \leq i \leq n} \left| E \left[\|b\|_2^{-1} \Delta_{n, i}(b) \middle| X \right] \right|^2 \\
&\leq 2 \sup_{1 \leq i \leq n} E \left\{ \left[\|b\|_2^{-1} \Delta_{n, i}(b) \right]^2 \middle| X \right\} + 2 \sup_{1 \leq i \leq n} \left| E \left[\|b\|_2^{-1} \Delta_{n, i}(b) \middle| X \right] \right|^2 \\
&\leq 4 \sup_{1 \leq i \leq n} E \left\{ \left[\|b\|_2^{-1} \Delta_{n, i}(b) \right]^2 \middle| X \right\} \leq c_4 \delta_n
\end{aligned}$$

Also, $E[\phi_{\tau n, i}(b) | X] = 0$. Then for $i \neq j$,

$$E \left[\phi_{\tau n, i}(b) \phi_{\tau n, j}(b) \middle| X \right] = E \left\{ \left[\Delta_{n, i}(b) - E(\Delta_{n, i}(b) | X) \right] \left[\Delta_{n, j}(b) - E(\Delta_{n, j}(b) | X) \right] \middle| X \right\} = 0$$

Also, for some constant $c_5 > 0$, we have $\sup_{1 \leq i \leq n} \|b\|_2^{-1} |\Delta_{n, i}(b)| \leq (1 + \tau) c_0 \|b\|_2 + 2 \|b\|_2^{-1} \sup_{1 \leq i \leq n} |r_i(\beta^0, b)| \leq c_5$. By Bernstein's inequality in Bosq (1998), for $0 < \zeta < 1/2$, there exists positive constant c_6, c_7 , such that $P \left\{ \|b\|_2^{-1} |\phi_{\tau n}(b)| > c_6 n^{1/2} \xi_n^{1/2 - \zeta} / n \right\} \leq 2 \exp \left\{ -c_7 \min(\xi_n^{-2\zeta}, n^{1/2} \xi_n^{1/2 - \zeta}) \right\}$. Then,

$$\sup_{1 \leq i \leq n} \|b\|_2^{-1} |\phi_{\tau n}(b)| = O_p(n^{-1/2} \xi_n^{1/2 - \zeta}) = o_p(n^{-1/2}).$$

Secondly, let's prove (ii). From the identify in Knight (1998), we have

$$\begin{aligned}
& \frac{1}{n} \sum_{i=1}^n E\{\tilde{D}_{\tau n, i}(b)|X\} \\
&= \frac{1}{n} \sum_{i=1}^n \left\{ \frac{\tau - F_0}{1 - F_0} I(T_i \geq C_i, F_0 \leq \tau) \left[\int_0^{r_i(\beta^0, b)} F_0 \left\{ \tilde{g}_{\tau n}(X_i^T \beta^0, \beta^0) + t|X_t \right\} dt - \frac{\tau}{n} \sum_{i=1}^n r_i(\beta^0, b) \right] \right. \\
&\quad + I(T_i \leq C_i) \left[\int_0^{r_i(\beta^0, b)} M_0 \left\{ \tilde{g}_{\tau n}(X_i^T \beta^0, \beta^0) + t|X_i \right\} dt - \frac{\tau}{n} \sum_{i=1}^n r_i(\beta^0, b) \right] \\
&\quad - \tau \left[I(T_i \leq C_i) - \frac{\tau - F_0}{1 - F_0} I(T_i \geq C_i, F_0 \leq \tau) \right] \\
&\quad \left. \left[B(X_i^T(\beta^0 + b))^T \tilde{\theta}_\tau(\beta^0 + b) - B(X_i^T \beta^0, \beta^0)^T \tilde{\theta}_\tau(\beta^0) \right] \right\}
\end{aligned}$$

By the Lipschitz condition of $f_0(t|X_i)$ and $m_0(t|X_i)$, with probability approaching 1, for positive constants c_8, c_9 , we have

$$\begin{aligned}
& \sup_{b \in N_0} \|b\|_2^{-2} \sup_{1 \leq i \leq n} \left| \int_0^{r_i(\beta^0, b)} F_0 \left\{ \tilde{g}_{\tau n}(X_i^T \beta^0, \beta^0) + t|X_i \right\} dt \right. \\
&\quad \left. - \left[F_0 \left\{ \tilde{g}_{\tau n}(X_i^T \beta^0, \beta^0)|X_i \right\} r_i(\beta^0, b) + \frac{1}{2} f_0 \left\{ \tilde{g}_{\tau n}(X_i^T \beta^0, \beta^0)|X_i \right\} r_i^2(\beta^0, b) \right] \right| \leq c_8 \xi_n \\
& \sup_{b \in N_0} \|b\|_2^{-2} \sup_{1 \leq i \leq n} \left| \int_0^{r_i(\beta^0, b)} M_0 \left\{ \tilde{g}_{\tau n}(X_i^T \beta^0, \beta^0) + t|X_i \right\} dt \right. \\
&\quad \left. - \left[M_0 \left\{ \tilde{g}_{\tau n}(X_i^T \beta^0, \beta^0)|X_i \right\} r_i(\beta^0, b) + \frac{1}{2} m_0 \left\{ \tilde{g}_{\tau n}(X_i^T \beta^0, \beta^0)|X_i \right\} r_i^2(\beta^0, b) \right] \right| \leq c_9 \xi_n
\end{aligned}$$

Then we have

$$\begin{aligned}
& \sup_{b \in N_0} \|b\|_2^{-2} \frac{1}{n} \left| \sum_{i=1}^n E\{\tilde{D}_{\tau n, i}(b) | X\} \right. \\
& \quad - \frac{1}{n} \sum_{i=1}^n I(Ti \leq C_i) \left\{ r_i(\beta^0, b) \left[F_0 \left\{ \tilde{g}_{\tau n}(X_i^T \beta^0, \beta^0) | X_i \right\} - \tau \right] + \frac{1}{2} f_0 \left\{ \tilde{g}_{\tau n}(X_i^T \beta^0, \beta^0) | X_i \right\} r_i^2(\beta^0, b) \right\} \\
& \quad - \frac{1}{n} \sum_{i=1}^n \frac{1}{n} \sum_{i=1}^n \left\{ \frac{\tau - F_0}{1 - F_0} I(Ti \geq C_i, F_0 \leq \tau) \left\{ r_i(\beta^0, b) \left[M_0 \left\{ \tilde{g}_{\tau n}(X_i^T \beta^0, \beta^0) | X_i \right\} - \tau \right] \right. \right. \\
& \quad \left. \left. + \frac{1}{2} m_0 \left\{ \tilde{g}_{\tau n}(X_i^T \beta^0, \beta^0) | X_i \right\} r_i^2(\beta^0, b) \right\} \right\} \left| \leq c_{10} \xi_n
\end{aligned}$$

By the result that $\sup_X \left| r_i(\beta^0, b) - \left[\frac{\partial \tilde{g}_n(X_i^T \beta^0, \beta^0)}{\partial \beta^T} \right] b \right| \leq c_0 \|b\|_2^2$, and $\sup_{1 \leq i \leq n} |\tilde{g}_{\tau n}(X_i^T \beta^0, \beta^0) - g_{\tau}(X_i^T \beta^0)| = O_p(J_n^{-r})$, with probability approaching 1,

$$\sup_{b \in N_0} \|b\|_2^{-2} \left| \sup_{1 \leq i \leq n} \left[r_i(\beta^0, b) - \frac{\partial \tilde{g}_{\tau n}(X_i^T \beta^0, \beta^0)}{\partial \beta^T} \right] \left[F_0 \left\{ \tilde{g}_{\tau n}(X_i^T \beta^0, \beta^0) | X_i \right\} - \tau \right] \right| \leq c_{11} J_n^{-r}$$

and

$$\sup_{b \in N_0} \|b\|_2^{-2} \left| \sup_{1 \leq i \leq n} \left[r_i(\beta^0, b) - \frac{\partial \tilde{g}_{\tau n}(X_i^T \beta^0, \beta^0)}{\partial \beta^T} \right] \left[M_0 \left\{ \tilde{g}_{\tau n}(X_i^T \beta^0, \beta^0) | X_i \right\} - \tau \right] \right| \leq c_{12} J_n^{-r}$$

Also,

$$\begin{aligned}
& \sup_{b \in N_0} \|b\|_2^{-2} \frac{1}{n} \left| \sum_{i=1}^n I(Ti \leq C_i) \frac{1}{2} \left[f_0 \left\{ \tilde{g}_{\tau n}(X_i^T \beta^0, \beta^0) | X_i \right\} r_i^2(\beta^0, b) \right. \right. \\
& \quad - f_0(g_{\tau}(X_i^T \beta^0)) [g_{\tau}^{(1)}(X_i^T \beta^0)]^2 b^T \tilde{X}_i \tilde{X}_i^T b \left. \right] - \frac{1}{n} \sum_{i=1}^n \frac{1}{n} \sum_{i=1}^n \frac{1}{2} \left[f_0 \left\{ \tilde{g}_{\tau n}(X_i^T \beta^0, \beta^0) | X_i \right\} r_i^2(\beta^0, b) \right. \\
& \quad \left. - m_0(g_{\tau}(X_i^T \beta^0)) [g_{\tau}^{(1)}(X_i^T \beta^0)]^2 b^T \tilde{X}_i \tilde{X}_i^T b \right] + \frac{1}{2} m_0 \left\{ \tilde{g}_{\tau n}(X_i^T \beta^0, \beta^0) | X_i \right\} r_i^2(\beta^0, b) \left| \right. \\
& \leq c_{13} J_n^{-r} + c_{14} \left((\log n)^{1/2} J_n^{1/2} n^{-1/2} + J_n^{1-r} + J_n^{-1} \right) + c_{15} \xi_n
\end{aligned}$$

Thus, the ξ_1 from (ii) would be $c_{10}\xi_n + c_{11}J_n^{-r} + c_{12}J_n^{-r} + c_{13}J_n^{-r} + c_{14}\left((\log n)^{1/2}J_n^{1/2}n^{-1/2} + J_n^{1-r} + J_n^{-1}\right) + c_{15}\xi_n = o(1)$.

Thirdly, we want to show (iii), which is $d_n = o_p(n^{-1/2})$. Decompose

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \rho_\tau^{(1)}(T_i^*) \frac{\partial \tilde{g}_{\tau n}(X_i^T \beta^0, \beta^0)}{\partial \beta^T} b - \frac{1}{n} \sum_{i=1}^n \left[\tau - I(T_i < g_\tau(X_i^T \beta^0)) \right] g_\tau^{(1)}(X_i^T \beta^0) \tilde{X}_i^T b \\ &= (\Pi_1 + \Pi_2 + \Pi_3)^T b \end{aligned}$$

where

$$\Pi_1 = \frac{1}{n} \sum_{i=1}^n \left[\tau - I(T_i < g_\tau(X_i^T \beta^0)) \right] \left[\frac{\partial \tilde{g}_{\tau n}(X_i^T \beta^0, \beta^0)}{\partial \beta^T} - g_\tau^{(1)}(X_i^T \beta^0) \tilde{X}_i \right],$$

$$\Pi_2 = \frac{1}{n} \sum_{i=1}^n \left\{ \rho_\tau^{(1)}(T_i^*) - \left[\tau - I(T_i < g_\tau(X_i^T \beta^0)) \right] \right\} g_\tau^{(1)}(X_i^T \beta^0) \tilde{X}_i,$$

$$\Pi_3 = \frac{1}{n} \sum_{i=1}^n \left\{ \rho_\tau^{(1)}(T_i^*) - \left[\tau - I(T_i < g_\tau(X_i^T \beta^0)) \right] \right\} \left[\frac{\partial \tilde{g}_{\tau n}(X_i^T \beta^0, \beta^0)}{\partial \beta^T} - g_\tau^{(1)}(X_i^T \beta^0) \tilde{X}_i \right].$$

Similarly, decompose

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \rho_\tau^{(1)}(C_i^*) \frac{\partial \tilde{g}_{\tau n}(X_i^T \beta^0, \beta^0)}{\partial \beta^T} b - \frac{1}{n} \sum_{i=1}^n \left[\tau - I(C_i < g_\tau(X_i^T \beta^0)) \right] g_\tau^{(1)}(X_i^T \beta^0) \tilde{X}_i^T b \\ &= (\Pi_4 + \Pi_5 + \Pi_6)^T b \end{aligned}$$

where

$$\Pi_4 = \frac{1}{n} \sum_{i=1}^n \left[\tau - I(C_i < g_\tau(X_i^T \beta^0)) \right] \left[\frac{\partial \tilde{g}_{\tau n}(X_i^T \beta^0, \beta^0)}{\partial \beta^T} - g_\tau^{(1)}(X_i^T \beta^0) \tilde{X}_i \right],$$

$$\Pi_5 = \frac{1}{n} \sum_{i=1}^n \left\{ \rho_\tau^{(1)}(C_i^*) - \left[\tau - I(C_i < g_\tau(X_i^T \beta^0)) \right] \right\} g_\tau^{(1)}(X_i^T \beta^0) \tilde{X}_i,$$

$$\Pi_6 = \frac{1}{n} \sum_{i=1}^n \left\{ \rho_\tau^{(1)}(C_i^*) - \left[\tau - I(C_i < g_\tau(X_i^T \beta^0)) \right] \right\} \left[\frac{\partial \tilde{g}_{\tau n}(X_i^T \beta^0, \beta^0)}{\partial \beta^T} - g_\tau^{(1)}(X_i^T \beta^0) \tilde{X}_i \right].$$

It's sufficient to show that $\mathbf{P}_j = o_p(n^{-1/2})$ for $j = 1, \dots, 6$. The proof can be found in Ma and He (2016), Page S.15 - S.17. Then

$$\begin{aligned} d_n &\leq \left| \sup_{b \in N_0} \|b\|_2^{-1} I(T \leq C) (\Pi_1 + \Pi_2 + \Pi_3)^T b \right| \\ &\quad + \left| \sup_{b \in N_0} \|b\|_2^{-1} \frac{\tau - F_0}{1 - F_0} I(T \geq C, F_0 \leq \tau) (\Pi_4 + \Pi_5 + \Pi_6)^T b \right| \\ &\leq \|\Pi_1 + \Pi_2 + \Pi_3\|_2 + \|\Pi_4 + \Pi_5 + \Pi_6\|_2 \\ &= o_p(n^{1/2}) \end{aligned}$$

Thus, the Lemma 7.2.7 is proved.

Lemma 7.2.8. *Under the conditions of Theorem 3.2.2, we have,*

$$\sup_{b \in N_0} \left| D_{\tau n}(b) - \tilde{D}_{\tau n}(b) + \xi_2 \|b\|_2 + \xi_3 \|b\|_2^{1/2} \right| = o_p(n^{-1})$$

where $\xi_2 = o_p(n^{1/2})$ and $\xi_3 = O_p((\log n)J_n/n)$ uniformly in $b \in N_0$.

Proof of Lemma 7.2.8

The result of Lemma 7.2.8 will be proved if we can show the followings,

$$(i) \sup_{b \in N_0} \left| E\{D_{\tau n}(b)|X\} - E\{\tilde{D}_{\tau n}(b)|X\} + \delta_2 \|b\|_2 + \delta_3 \|b\|_2^{1/2} \right| = o_p(n^{-1})$$

where $\delta_2 = o_p(n^{-1/2})$ and $\delta_3 = O_p((\log(n))J_n n^{-1})$ uniformly in $b \in N_0$, and

$$(ii) \sup_{b \in N_0} \left| D_{\tau n}(b) - \tilde{D}_{\tau n}(b) - E\{D_{\tau n}(b)|X\} + E\{\tilde{D}_{\tau n}(b)|X\} + \delta_2^* \|b\|_2 + \delta_3^* \|b\|_2^{1/2} \right| = o_p(n^{-1})$$

where $\delta_2^* = o_p(n^{-1/2})$ and $\delta_3^* = O_p((\log(n))J_n n^{-1})$ uniformly in $b \in N_0$. We will

show the result of (i), while the (ii) can be obtained similarly. Since $\tilde{\theta}_\tau(\beta)$ is the minimizer of $E\{L_{\tau n}(\theta, \beta)|X\}$, then set $\frac{\partial E\{L_{\tau n}(\theta, \beta)|X\}}{\partial \theta} = 0$. We have the following,

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n E \left\{ \frac{\tau - F_0}{1 - F_0} I(T \geq C, F_0 \leq \tau) \rho_\tau^{(1)}(C_i - B(X_i^T \beta))^T \tilde{\theta}_\tau(\beta) \right. \\ & \quad \left. + I(T_i \leq C_i) \rho_\tau^{(1)}(T_i - B(X_i^T \beta))^T \tilde{\theta}_\tau(\beta) \right. \\ & \quad \left. + \left[I(T_i > C_i) - \frac{\tau - F_0}{1 - F_0} I(T \geq C, F_0 \leq \tau) \right] B(X_i^T \beta) \right\} = 0 \end{aligned}$$

Then by Taylor expansion,

$$\begin{aligned} & E\{L_{\tau n}(\theta, \beta)|X\} - E\{L_{\tau n}(\tilde{\theta}_\tau(\beta), \beta)|X\} \\ &= -\frac{1}{n} \sum_{i=1}^n \frac{\tau - F_0}{1 - F_0} I(T \geq C, F_0 \leq \tau) \left\{ E \left[\rho_\tau^{(1)}(C_i - B(X_i^T \beta))^T \tilde{\theta}_\tau(\beta) B(X_i^T \beta)^T (\theta - \tilde{\theta}_\tau(\beta)) \right] \right\} - \\ & \quad \frac{1}{2} m_0 \left(B(X_i^T \beta)^T \theta_\tau^*(\beta) | X_i \right) (\theta - \tilde{\theta}_\tau(\beta))^T B(X_i^T \beta) B(X_i^T \beta)^T (\theta - \tilde{\theta}_\tau(\beta)) \left\} \right. \\ & \quad \left. - \frac{1}{n} \sum_{i=1}^n I(T_i \leq C_i) \left\{ E \left[\rho_\tau^{(1)}(T_i - B(X_i^T \beta))^T \tilde{\theta}_\tau(\beta) B(X_i^T \beta)^T (\theta - \tilde{\theta}_\tau(\beta)) \right] \right\} - \right. \\ & \quad \left. \frac{1}{2} f_0 \left(B(X_i^T \beta)^T \theta_\tau^*(\beta) | X_i \right) (\theta - \tilde{\theta}_\tau(\beta))^T B(X_i^T \beta) B(X_i^T \beta)^T (\theta - \tilde{\theta}_\tau(\beta)) \right\} \\ &= \frac{1}{2n} \sum_{i=1}^n \left\{ I(T_i \leq C_i) f_0 \left(B(X_i^T \beta)^T \theta_\tau^*(\beta) | X_i \right) (\theta - \tilde{\theta}_\tau(\beta))^T B(X_i^T \beta) B(X_i^T \beta)^T (\theta - \tilde{\theta}_\tau(\beta)) + \right. \\ & \quad \left. \frac{\tau - F_0}{1 - F_0} I(T \geq C, F_0 \leq \tau) m_0 \left(B(X_i^T \beta)^T \theta_\tau^*(\beta) | X_i \right) (\theta - \tilde{\theta}_\tau(\beta))^T B(X_i^T \beta) B(X_i^T \beta)^T (\theta - \tilde{\theta}_\tau(\beta)) \right\} \end{aligned}$$

where $\theta_\tau^*(\beta)$ is between θ and $\tilde{\theta}_\tau(\beta)$. By the Lipschitz continuity of $f_0(\cdot|X_i)$ and $m_0(\cdot|X_i)$, then

$$\begin{aligned} & \left| f_0 \left(B(X_i^T \beta)^T \theta_\tau^*(\beta) | X_i \right) - f_0 \left(\tilde{g}_\tau(X_i^T \beta, \beta) \theta_\tau^*(\beta) | X_i \right) \right| \\ & \leq c_1 \left(\|B(X_i^T \beta)\|_2 \|\theta^0(\beta) - \tilde{\theta}(\beta)\|_2 \right) + R_1(X_i^T \beta, \beta) \end{aligned}$$

and,

$$\begin{aligned} & \left| m_0 \left(B(X_i^T \beta)^T \theta_\tau^*(\beta) | X_i \right) - m_0 \left(\tilde{g}_\tau(X_i^T \beta, \beta) \theta_\tau^*(\beta) | X_i \right) \right| \\ & \leq c_2 \left(\|B(X_i^T \beta)\|_2 \|\theta^0(\beta) - \tilde{\theta}(\beta)\|_2 \right) + R_2(X_i^T \beta, \beta) \end{aligned}$$

for some positive constants c_1, c_2 . Hence for $L_n \asymp J_n n^{-1/2}$,

$$\begin{aligned} & \sup_{\beta \in \Theta, \|\tilde{\theta}_\tau(\beta) - \theta\|_2 \leq L_n} \left| E\{L_{\tau n}(\theta, \beta) | X\} - E\{L_{\tau n}(\tilde{\theta}_\tau(\beta), \beta) | X\} - \right. \\ & \left. \frac{1}{2n} \sum_{i=1}^n \left\{ I(T_i \leq C_i) f_0 \left(B(X_i^T \beta)^T \theta_\tau^*(\beta) | X_i \right) (\theta - \tilde{\theta}_\tau(\beta))^T B(X_i^T \beta) B(X_i^T \beta)^T (\theta - \tilde{\theta}_\tau(\beta)) - \right. \right. \\ & \left. \left. \frac{\tau - F_0}{1 - F_0} I(T \geq C, F_0 \leq \tau) m_0 \left(B(X_i^T \beta)^T \theta_\tau^*(\beta) | X_i \right) (\theta - \tilde{\theta}_\tau(\beta))^T B(X_i^T \beta) B(X_i^T \beta)^T (\theta - \tilde{\theta}_\tau(\beta)) \right\} \right| \\ & = O \left(J_n n^{-1} (J_n n^{-1/2} + J_n^{-r}) \right) = o(n^{-1}) \end{aligned}$$

Moreover, by Lemma 7.2.6, we know

$$\sup_{\beta \in \Theta} \|\tilde{\theta}_\tau(\beta) - \tilde{\theta}_\tau(\beta)\|_2 = O_p(J_n n^{-1/2}).$$

Since $L_{\tau n}(\tilde{\theta}_\tau(\beta), \beta) = \tilde{L}_{\tau n}^*(\beta)$ and $L_{\tau n}(\tilde{\theta}_\tau(\beta), \beta) = L_{\tau n}^*(\beta)$, then

$$\sup_{\beta \in \Theta} \left| E[L_{\tau n}^*(\beta) | X] - E[\tilde{L}_{\tau n}^*(\beta) | X] - \frac{1}{2} \Pi_n(\beta) \right| = o_p(n^{-1}).$$

Since $D_{\tau n}(b) = L_{\tau n}^*(\beta^0 + b) - L_{\tau n}^*(\beta^0)$ and $\tilde{D}_{\tau n}(b) = \tilde{L}_{\tau n}^*(\beta^0 + b) - \tilde{L}_{\tau n}^*(\beta^0)$, then we have

$$\sup_{\beta \in \Theta} \left| E[D_{\tau n}(b) | X] - E[\tilde{D}_{\tau n}(b) | X] - \frac{1}{2} [\Pi_n(\beta^0 + b) - \Pi_n(\beta^0)] \right| = o_p(n^{-1}).$$

Follow the procedure in Ma and He (2016) Page S.19 - S.20, we can show that

$$\Pi_n(\beta^0 + b) - \Pi_n(\beta^0) = o_p(n^{1/2})\|b\|_2 + O_p((\log n)J_n/n)\|b\|_2^{1/2}.$$

Thus, the proof of Lemma 7.2.8 is complete.

Lemma 7.2.9. *Suppose Assumptions 3.1.1-3.1.5 hold, then for any $q \geq 1$,*

(i)

$$\|\hat{F} - F_0\|_{\mathcal{H}} = \sup_t \sup_x |\hat{F}(t|\hat{x}) - F_0(t|x)| = \sup_t \sup_z |\hat{F}(t|\hat{z}) - F_0(t|z_0)| = O_p\left(\left\{\frac{\log n}{nh^q}\right\}^{1/2} + h^v\right),$$

(ii)

$$\hat{F}(t|z) - F_0(t|z) = \sum_{j=1}^n B_{nj}(z)\xi(Y_j, \eta_j, t, z) + O_p\left(\left\{\frac{\log n}{nh^q}\right\}^{3/4} + h^v\right), a.s.$$

where,

$$\begin{aligned} \xi(Y, \eta, t, z) = \{1 - F_0(t|z)\} & \left[- \int_0^{\min(Y,t)} f_0(s|z)\{1 - F_0(s|z)\}^{-2}\{1 - G(s|z)\}^{-1} ds \right. \\ & \left. + I(Y \leq t, \eta = 1)\{1 - F_0(Y|z)\}^{-1}\{1 - G(Y|z)\}^{-1} \right]. \end{aligned}$$

The Lemma 7.2.9 comes from Lemma 1 of Wang et al. (2013).

Bibliography

- [1] Alkenani, A., and Yu, K. (2013). Penalized single-index quantile regression. *International Journal of Statistics and Probability*, **2**, 12.
- [2] Breiman, L. (2001). Random forests. *Machine Learning*, **45**, 5-32.
- [3] Bücher, A., El Ghouch, A., and Van Keilegom, I. (2014). Single index quantile regression models for censored data. *Manuscript*.
- [4] Carroll, R. J., Fan, J., Gijbels, I., and Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association*, **92**, 477-489.
- [5] Chaudhuri, P., Doksum, K., and Samarov, A. (1997). On average derivative quantile regression. *The Annals of Statistics*, **25**, 715-744.
- [6] Christou, E. and Akritas, M. (2016). Single-index quantile regression for censored data. *Manuscript*.
- [7] Doksum, K., and Samarov, A. (1995). Nonparametric estimation of global functionals and a measure of the explanatory power of covariates in regression. *The Annals of Statistics*, 1443-1473.

- [8] Duan, N. and Li, K.C. (1991). Slicing regression: a link-free regression method. *The Annals of Statistics*, 505-530.
- [9] Efron, B. (1967). The two sample problem with censored data. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 4, No. University of California Press, Berkeley, CA, pp. 831-853). Prentice-Hall Engewood Cliffs, NJ.
- [10] Fan, J., and Huang, T. (2005). Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli*, **11**, 1031-1057.
- [11] Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, **96**, 1348-1360.
- [12] Fan, J., and Lv, J. (2010). A selective overview of variable selection in high dimensional feature space. *Statistica Sinica*, **20**, 101.
- [13] Fan, J. and Peng, H. (2004). Nonconcave penalized likelihood with a diverging number of parameters. *The Annals of Statistics*, **32**, 928-961.
- [14] Harrell, F.E. (2001). Regression Modeling Strategies. *Springer, New York*.
- [15] Hardle, W., Hall, P., and Ichimura, H. (1993). Optimal smoothing in single-index models. *The Annals of Statistics*, **21**, 157-178.
- [16] Harrison, D., and Rubinfeld, D. L. (1978). Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management*, **5**, 81-102.

- [17] Hastie, T., Tibshirani, R., Friedman, J., & Franklin, J. (2005). The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, **27**, 83-85.
- [18] He, X. and Shi, P. (1994). Convergence rate of b-spline estimators of nonparametric conditional quantile functions. *Journal of Nonparametric Statistics*, **3**, 299-308.
- [19] Honore, B., Khan, S., and Powell, J. L. (2002). Quantile regression under random censoring. *Journal of Econometrics*, **109**, 67-105.
- [20] Ishwaran, H., Kogalur, U. B., Gorodeski, E. Z., Minn, A. J., and Lauer, M. S. (2010). High-dimensional variable selection for survival data. *Journal of the American Statistical Association*, **105**, 205-217.
- [21] Klein, R. W., and Spady, R. H. (1993). An efficient semiparametric estimator for binary response models. *Econometrica: Journal of the Econometric Society*, 387-421.
- [22] Koenker, R. (2005). Quantile Regression. *Cambridge University Press*, Number **38**.
- [23] Koenker, R. and Bassett Jr, G. (1978). Regression quantiles. *Econometrica: Journal of the Econometric Society*, 33-50.
- [24] Koenker, R. and Geling, O. (2001). Reappraising medfly longevity: a quantile regression survival analysis. *Journal of the American Statistical Association*, **96**, 458-468.
- [25] Koenker, R., Ng, P., and Portnoy, S. (1994). Quantile Smoothing Splines. *Biometrika*, **81**, 673-680.

- [26] Kong, E. and Xia, Y. (2012). A single-index quantile regression model and its estimation. *Econometric Theory*, **28**, 730-768.
- [27] Lam, C. and Fan, J. (2008). Profile-kernel likelihood inference with diverging number of parameters. *The Annals of Statistics*, **36**, 2232.
- [28] Lee, A.P. and Abramowicz, D. (2014). Is the kidney donor risk index a step forward in the assessment of deceased donor kidney quality? *Nephrology Dialysis Transplantation*, 304.
- [29] Li, K.C. (1991). Sliced inverse regression for dimension reduction. *Journal of the American Statistical Association*, **86**, 316-327.
- [30] Li, K. C., Wang, J. L., and Chen, C. H. (1999). Dimension reduction for censored regression data. *The Annals of Statistics*, **27**, 1-23.
- [31] Liang, H., Liu, X., Li, R., and Tsai, C. L. (2010). Estimation and testing for partially linear single-index models. *The Annals of Statistics*, **38**, 3811.
- [32] Lin, W. and Kulasekera, K. B. (2007). Identifiability of single-index models and additive-index models. *Biometrika*, **94**, 496-501.
- [33] Lockhart, R., Taylor, J., Tibshirani, R.J. and Tibshirani, R. (2014). A significance test for the lasso. *The Annals of Statistics*, **42**, 413.
- [34] Lv, Y., Zhang, R., Zhao, W., and Liu, J. (2014). Quantile regression and variable selection for the single-index model. *Journal of Applied Statistics*, **41**, 1565-1577.
- [35] Ma, S. and He, X. (2016). Inference for Single-index Quantile Regression Models with Profile Optimization. *The Annals of Statistics*, **44**, 1234-1268.

- [36] Ma, Y. and Zhu, L. (2012). A semiparametric approach to dimension reduction. *Journal of the American Statistical Association*, **107**, 1681-79.
- [37] Meinshausen, N., Meier, L. and Bühlmann, P. (2012). P-values for high-dimensional regression. *Journal of the American Statistical Association*.
- [38] Müller, H.G. (1984). Smooth optimum kernel estimators of densities, regression curves and modes. *The Annals of Statistics*, , 766-774.
- [39] Nyberg, S. L., Baskin-Bey, E. S., Kremers, W., Prieto, M., Henry, M. L., and Stegall, M. D. (2005). Improving the prediction of donor kidney quality: deceased donor score and resistive indices. *Transplantation*, **80**, 925-929.
- [40] Nyberg, S.L., Matas, A.J., Kremers, W.K., Thostenson, J.D., Larson, T.S., Prieto, M., Ishitani, M.B., Sterioff, S. and Stegall, M.D., (2003). Improved scoring system to assess adult donors for cadaver renal transplantation. *American Journal of Transplantation*, **3**, 715-721.
- [41] Peng, L. and Huang, Y. (2008). Survival analysis with quantile regression models. *Journal of the American Statistical Association*, **103**, 637-649.
- [42] Port, F., Bragg-Gresham, J., Metzger, R., Dykstra, D., Gillespie, B., Young, E., Delmonico, F., Wynn, J., Merion, R., Wolfe, R., and Held, P. (2002). Donor characteristics associated with reduced graft survival: an approach to expanding the pool of kidney donors. *Transplantation*, **74**, 1281-1286.
- [43] Portnoy, S. (2003). Censored regression quantiles. *Journal of the American Statistical Association*, **98**, 1001-1012.
- [44] Portnoy, S., and Lin, G. (2010). Asymptotics for censored regression quantiles. *Journal of Nonparametric Statistics*, **22**, 115-130.

- [45] Powell, J. L. (1984). Least absolute deviations estimation for the censored regression model. *Journal of Econometrics*, **25**, 303-325.
- [46] Powell, J. L. (1986). Censored regression quantiles. *Journal of Econometrics*, **32**, 143-155.
- [47] Rao, P., Schaubel, D., Guidinger, M., Andreoni, K., Wolfe, R., Merion, R., Port F., and Sung, R. (2009). A comprehensive risk quantification score for deceased donor kidneys: the kidney donor risk index. *Transplantation*, **88**, 231-236.
- [48] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 267-288.
- [49] Wang, H. and Wang, L. (2009). Locally weighted censored quantile regression. *Journal of the American Statistical Association*, **104**, 1117-1128.
- [50] Wang, H., Zhou, J., and Li, Y. (2013). Variable selection for censored quantile regression. *Statistica Sinica*, **23**, 145-167.
- [51] Wasserman, L. and Roeder, K. (2009). High dimensional variable selection. *The Annals of Statistics*, **37**, 2178.
- [52] Wu, T. Z., Yu, K., and Yu, Y. (2010). Single-index quantile regression. *Journal of Multivariate Analysis*. **101**, 1607-1621.
- [53] Xia, Y., Tong, H., Li, W. K., and Zhu, L. X. (2002). An adaptive estimation of dimension reduction space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **64**, 363-410.
- [54] Xia, Y., Zhang, D., and Xu, J. (2010). Dimension reduction and semiparametric estimation of survival models. *Journal of the American Statistical Association*, **105**, 278-290.

- [55] 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.
- [56] Yu, K. and Jones, M. C. (1998). Local linear quantile regression. *Journal of the American Statistical Association*, **93**, 228-237.
- [57] Yu, Y., and Ruppert, D. (2002). Penalized spline estimation for partially linear single-index models. *Journal of the American Statistical Association*, **97**, 1042-1054.
- [58] Zhang, C. H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 894-942.
- [59] Zhu, L.P. and Zhu, L.X. (2009). Nonconcave penalized inverse regression in single-index models with high dimensional predictors. *Journal of Multivariate Analysis*, **100**, 862-875.
- [60] Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, **101**, 1418-1429.
- [61] Zou, H., and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **67**, 301-320.
- [62] Zou, Q. and Zhu, Z. (2014). M-estimators for single-index model using B-spline. *Metrika*, **77**, 225-246.
- [63] Organ Procurement and Transplantation Network (OPTN). <https://optn.transplant.hrsa.gov>

- [64] United States Renal Data System 2015 Annual Data Report. https://www.usrds.org/2015/view/v2_07.aspx