Nonparametric and Semiparametric Estimation of Instrumental Variable Method

Anqi Cheng

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2019

Reading Committee:

Kwun Chuen Gary Chan, Chair

Ying Chen

Li Hsu

Program Authorized to Offer Degree:
Biostatistics

University of Washington

Abstract

Nonparametric and Semiparametric Estimation of Instrumental Variable Method

Anqi Cheng

Chair of the Supervisory Committee:
Associate Professor Kwun Chuen Gary Chan
Department of Biostatistics

The instrumental variable approach has been widely used for estimating the treatment effect in the presence of unmeasured confounding, e.g. randomized trials with noncompliance problems and observational studies. While most literature focus on the estimation of compliers averaged causal effect (CACE) nonparametrically or based on parametric assumptions, under the IV assumptions, fewer works focus on estimating distributional causal effect using IV. We study a novel monotone cumulative distribution function estimator of an outcome variable for compliers receiving treatment or control. The estimation procedures involve a weighted quantile regression and a post-estimation rearrangement adjustment. We show that the proposed estimator is consistent and develop large sample properties. Based on the asymptotic properties of the proposed estimator, a Wilcoxon-type statistic is proposed to test the equivalence of CDF for compliers receiving treatment and control. By comparing the influence function of the proposed estimator to the efficient influence function, we modify the proposed estimator and obtain a local efficient and robust estimator in the sense that when the unknown density functions are correctly specified, it reaches the semiparametric efficiency bound and when the unknown density functions are misspecified, it is still a consistent estimator. For the censoring outcomes, we propose a method to estimate quantile functions and survival functions for potential outcomes under independent censoring and noncompliance. Based on the martingale feature associated with the censoring data, we estimate quantile functions for compliers. Then using the possibly non-monotone quantile function, we construct a monotone and bounded estimator for the survival function. By using empirical process techniques, we establish asymptotic properties, including uniform consistency and weak convergence for the proposed estimators. For general observational studies with unmeasured confounding problems, we impose a no-interaction assumption proposed by Wang and Tchetgen Tchetgen (2018) and propose a new class of IV models that identify quantities of potential outcomes for the whole population. Our work complements current research on using instrumental variable method to estimate distributions of potential outcomes and infer heterogenous treatment effect for observational studies in the presence of unmeasured confounding, especially for the censoring outcomes. Simulation results, real data examples, and proofs are detailed in this dissertation.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

After five years of graduate school study at the University of Washington, today is the day for me to write this finishing touch on my dissertation. It has been a wonderful period in my lifetime, during which I learned a lot, both in how to do good research and how to be a better person. I am grateful to all the people who have supported and helped me so much throughout this period.

First, I would like to express my sincere appreciation to my adviser, Professor Gary Chan, for his continued guidance and advise throughout my research. I would also like to thank all my Dissertation Committees, Professor Ying Chen, Professor Li Hsu, and Professor Amanda Phipps. Thanks so much for the extensive help and suggestions.

Many thanks to my RA supervisors, Dr. Amalia Magaret, Dr. Janet Stanford, Dr. James Dai, and Dr. Elizabeth Blue, for their kind guidance and advice on my research and writing. It has been a great pleasure for me to collaborate with them and I have learned tremendously from them.

Thank you to all friends in the Department of Biostatistics and Statistics, for their support and help on my personal life, for their questions and academic discussions on my research. I wish you all the best with your future endeavor.

Lastly, I would like to thank my parents, whose love and guidance are with me in whatever I pursue. Most importantly, I wish to thank my lovely and supportive husband, Junjie, and our cute little fur baby, Toffy. Thanks so much for all those love and companion you offered. I will not be able to reach this far without you.

Chapter 1

INTRODUCTION

A randomized trial is a gold standard to identify the causal effect of an intervention or a treatment since the distribution of confounding variables is balanced by randomization. However, when some subjects fail to follow their assigned treatments, the balance of randomization does not hold because subjects who actually receive treatment are not comparable to subjects who actually receive control. The estimation of the causal effect based on the actual treatment received is typically not consistent with the true causal effect of treatment. The Instrumental variable approach has been widely used for estimating the treatment effect in the presence of unmeasured confounding, e.g. randomized trials with noncompliance problems. An instrumental variable (IV) is a variable which is (a) associated with treatment, (b) has no direct effect on the outcome other than through the treatment and (c) associated with the outcome without unmeasured confounders. In randomized trials with noncompliance, the association between treatment and outcome is confounded by unmeasured variables. The assignment to treatment or encouragement to take treatment can be justified as a valid instrument variable. In this setting, both the exposure and instrumental variable are binary variables. In the framework of counterfactual or potential outcome.Imbens and Angrist (1994) and (Angrist et al., 1996) proposed a set of IV assumptions, which can nonparametrically identify the causal effect of treatment for compliers, a latent subpopulation who would take treatment only when assigned to the treatment group.

While most literatures, e.g.Imbens and Rubin (1997a), Imbens and Rubin (1997b) and Cheng et al. (2009), focus on the estimation of compliers averaged causal effect (CACE) nonparametrically or based on parametric assumptions, under the IV assumptions by Imbens and Angrist (1994) and Angrist et al. (1996) , there are less works on estimating distributions of potential outcomes using IV method. When the treatment effect is a simple "location shift", CACE can capture the impact of treatment on the entire distribution. However, in most cases, the treatment effect is heterogeneous and distributional treatment effect for compliers beyond simple CACE is of fundamental interest, especially for policy-makers and social welfare comparisons. As discussed in Imbens and Rubin (1997b), they show how to estimate the cumulative distribution function (CDF) for compliers using an IV estimator, then estimate the density functions for compliers based on smoothing and obtain nonnegative density functions by truncation as a post-estimation procedure. As an extension to the IV models of Imbens and Angrist (1994) and Angrist et al. (1996), Abadie (2000, 2003) introduce a class of IV estimators of linear and nonlinear average treatment outcome models with covariates. They show that the expectation of a function of potential outcomes for compliers as a weighted expectation of observed outcomes can be identified using a weighted expectation of observed outcomes, e.g. quantile functions for compliers. Abadie (2002) proposes methods to test the global hypothesis of distributional causal effects, such as distribution equivalence and stochastic dominance, based on the estimation of CDF for compliers.

Estimating the compliers outcome distributions can also be considered as identifying latent distributions from mixture models, where data arise from a mixture of distributions. Identification and estimation problems in mixture models are common in many scientific studies. For example, mixture models are widely used in quantitative trait locus (QTL) studies to estimate the effect of a QTL predisposing a trait. Since the genotypes at the QTL is unknown, it's impossible to observe and estimate the effect of the QTL directly. Mixture models are widely used to identify the effect of the QTL using known markers. Ma and Wang (2012) characterize the complete class of consistent estimators for the latent distributions. They propose an estimator that reaches the semiparametric efficiency bound and provide an algorithm to implement the efficient estimator which requires kernel smoothing. Nonparametric maximum likelihood estimators (NPMLEs) can also be used to identify latent distributions in mixture models, however, two widely used NPMLEs are shown to be either inefficient or even inconsistent by Ma and Wang (2012).

None of the above estimators for the compliers distribution functions is guaranteed to be monotone or bounded between 0 and 1. A possible method to obtain a monotone estimator from a non-monotone estimator is weighted isotonic regression. However, in a weighted isotonic regression, it's not apparent to choose the optimal weight and the asymptotic properties of the estimator are typically unknown. Lee et al. (2016) propose a combined method of the EM algorithm and Pool-Adjacent-Violators Algorithm using a binomial likelihood to obtain monotone distribution curves for compliers. The binomial likelihood method purposely discards correlation in the full likelihood to avoid known problems in the NPMLEs, but is therefore inefficient. Moreover, the asymptotic distribution of their estimator remains unknown. To address the above problems, in Chapter 2, we propose a novel IV estimator that is guaranteed to be monotone and bounded between 0 and 1. We first estimate quantile functions of potential outcomes for compliers and apply a rearrangement operation to obtain monotone distribution functions.

Two common problems in clinical trials are censoring and noncompliance, while most of the previous literature is for continuous and uncensored outcomes, focusing on the estimation of compliers averaged causal effect nonparametrically or based on parametric assumptions (Imbens and Rubin, 1997a,b; Cheng et al., 2009). For survival outcomes, few papers have discussed randomized studies in the presence of noncompliance (Robins and Tsiatis, 1991; Joffe, 2001; Loeys and Goetghebeur, 2003; Cuzick et al., 2007). These papers propose various semiparametric models to estimate the effect of treatment on survival functions for potential failure times. Baker (1998) derived closed-form expressions for the maximum likelihood estimates (MLEs) of the hazards of compliers in the treatment and control groups for discrete survival outcomes. Baker's estimator is analogous to the standard IV estimator for a survival outcome. Building on Baker (1998)'s work, Nie et al. (2011) develop a nonparametric estimator of the survival function at a specified time for compliers based on the empirical likelihood that makes full use of the mixture structure to gain efficiency over the standard IV method. Though this nonparametric estimator is bounded between 0 and 1, the method is not efficient to obtain whole survival functions and not necessary to generate a monotone survival function. In the case of binary treatment variable and binary instrument variable (i.e. treatment assignment), we could also consider estimating the compliers outcome distributions as identifying latent distributions from observed mixture data. In the mixture data framework, Qin et al. (2014) propose a binomial likelihood approach and a combined method of the EM algorithm and Pool-Adjacent-Violators algorithm to obtain monotone distribution curves for latent distributions. However, the asymptotic distribution of their estimator remains unknown.

In Chapter 2, we mainly address the noncompliance problem in identifying the distributional causal effect of treatment in randomized trials. In Chapter 3, we further study identification and estimation problems of distributional causal effect for a time-to-event outcome subject to right censoring in randomized trials with noncompliance. Our work builds on Peng and Huang (2008), who develop a quantile regression method for survival data subject to conditionally independent censoring. Their proposal uses the martingale feature associated with the censored data. The main idea is that we first nonparametrically estimate the quantile functions for compliers. Then we apply the rearrangement operation to the possibly non-monotone quantile function and obtain cumulative distribution function or survival function for the right censoring potential outcomes for compliers, that is monotone and bounded in $[0, 1]$. Moreover, we provide a self-induced smoothing algorithm to estimate the limiting variance of the estimators.

When randomized experiments are not feasible for various reasons, observational studies are widely used to infer treatment effect. However, the treatment assignment or exposure of interest is often associated with confounding variables that are not necessarily observed in the study. Leaving out confounding variables typically leads to a biased estimate of treatment effect. A major challenge of causal inference in observational studies is to address unmeasured confounding variables. In such settings, instrumental variable (IV) methods are useful to estimate the effect of treatment with unmeasured confounding, in other words, endogenous exposure in econometrics literature. Under the standard IV assumptions proposed by Angrist et al. (1996), Abadie (2000) introduces a new class of instrumental variable (IV) estimators for linear and nonlinear treatment response models. These IV models identify local quantities of potential outcomes for the compliers, the subpopulation whose treatment status is affected by the instrumental variable. In particular, Abadie et al. (2002) focus on estimating local (conditional) quantile treatment effect (LQTE) for compliers. Chernozhukov and Hansen (2005) propose a new IV model, the instrumental variable quantile regression (IVQR) model, for estimating quantile treatment effect (QTE). The IVQR model relies on a different set of identification assumptions than the IV assumption in Angrist et al. (1996). Instead of estimating local quantile functions for compliers, the IVQR model estimates the quantile functions and QTE for the whole population. The key assumption in IVQR model is the rank similarity assumption, a condition that restricts the individual ranks in the potential outcome distributions. The two sets of IV models are generally non-nested and estimating different quantities. Wüthrich (2019) studies the relationship between the two IV models for estimating QTE and shows that there is a close connection between the IVQR model and LQTE model. The IVQR estimands for QTE are equivalent to LQTE for compliers at transformed quantile levels. Moreover, the IVQR estimand of the average treatment effect (ATE) is a convex combination of the local average treatment effect and a weighted average of LQTE. These findings do not rely on the rank similarity assumption and provide an alternate interpretation of IVQR estimands when the rank similarity assumption is violated.

Since the definition of compliers depends on the particular IV that is available, the local estimand would, in general, differ from the population-level estimand for the whole study population. To the best of our knowledge, to date, there has only been limited work focusing on population-level estimand in the context of an IV, especially for censored outcomes. Robins and Tsiatis (1991) describe a formal IV approach to analyze survival outcomes by parameterizing the exposure effect under a structural accelerated failure time model and developing G-estimation method.Loeys and Goetghebeur (2003) propose an alternative approach based on structural proportional hazard models but is more parametric and requires modeling the exposure distribution. Li et al. (2015) developed a closed-form, two-stage estimator for the causal effect in the additive hazard model, assuming linear structural equation models for the hazard function. Their method permits both continuous and discrete exposures. Tchetgen et al. (2015) independently developed a two-stage estimation in additive hazard models under a less restrictive model than Li et al. (2015). More recently, Martinussen et al. (2017) develop IV estimators under a semiparametric structural cumulative survival model without modeling exposure distribution. Their proposal is closely related to but less restrictive than the additive hazard model in Tchetgen et al. (2015) and Li et al. (2015). The above estimators model the causal effect of exposure as a location shift model on a certain scale of survival outcome (e.g. log-hazard, hazard and log-linear).

Our proposed IV model in Chapter 4 relaxes this restriction, allowing for a heterogeneous causal effect of exposure, and can accommodate adjustment for baseline covariates. In a recent paper, Wang and Tchetgen Tchetgen (2018) discuss estimating the average treatment effect (ATE) in the context of an IV. Instead of making the monotonicity assumption as in Angrist et al. (1996), they propose two alternative no-interaction assumptions involving the unobserved confounders. In Chapter 4, we propose a new class of IV models for nonlinear moment condition models for population-level estimands, generalizing Abadie's identification results for compliers in Abadie (2000, 2003). As two special cases, we show that under the new IV assumptions, the proposed estimating equations in Chapter 2 and Chapter3 identify population-level quantities in the framework of observational studies with unmeasured confounding. To study the heterogeneous treatment effect at different quantiles, Chernozhukov et al. (2015) developed a censored quantile instrumental variable (CQIV) estimator and described its properties and computation. The CQIV estimator combines Powell (1986)'s censored quantile regression to deal with censoring, with a control variable approach to incorporate endogenous exposures. However, this approach requires that the censoring time is always observed, which is not the case in most survival settings. Our proposed model allowing for random censoring complements the methodology in the context of IV for identifying distributions of right censoring potential outcomes in observational studies with unmeasured confounding.

My dissertation consists of three main Chapters about the instrumental variable method in unmeasured confounding problems. The first two chapters focus on the noncompliance problem in randomized trials. We propose IV estimators to identify the marginal distribution functions of potential outcomes for compliers. Specifically, the proposed estimators of the cumulative distribution functions are monotone and bounded between $[0, 1]$. In the third Chapter, we propose a new class of IV models that focus on identifying causal estimand for the whole population in the framework of observational studies with unmeasured confounding.

Chapter 2

MONOTONE DISTRIBUTION FUNCTION ESTIMATION IN RANDOMIZED TRIALS WITH NONCOMPLIANCE

2.1 Introduction

A randomized trial is often regarded as the gold standard to identify the causal effect of an intervention or a treatment since the distribution of measured and unmeasured confounders is balanced by randomization. However, when some subjects fail to follow their assigned treatments, the balance of randomization does not hold in general and subjects who actually receive treatment are not comparable to subjects who actually receive control. Therefore, the estimation of causal effect based on the actual treatment received is not consistent to the true causal effect of treatment. Additional assumptions to identify the effect of treatment in a meaningful way are often required. As in Imbens and Angrist (1994) and Angrist et al. (1996), under certain plausible assumptions, referred to as the instrumental variable (IV) assumptions, the causal effect of the treatment can be nonparametrically identified among compliers in the framework of the counterfactual or potential outcome. These assumptions are discussed in Section 2. As an extension to the IV model of Imbens and Angrist (1994) and Angrist et al. (1996), Abadie (2000, 2003) introduce a class of IV estimators of linear and nonlinear average treatment outcome models with covariates. Identification is shown by expressing the expectation of a function of potential outcomes for compliers as a weighted expectation of observed outcomes, where the weights can be positive or negative. While most literature (Imbens and Rubin, 1997a,b; Cheng et al., 2009) focus on the estimation of compliers averaged causal effect (CACE) nonparametrically or based on parametric assumptions, under the IV assumptions, fewer works focus on estimating distributional causal effect using IV. When the treatment effect is a simple "location shift", CACE can capture the impact of treatment on the entire distribution. However, in most cases, the treatment effect is heterogeneous and distributional treatment effect for compliers beyond simple CACE is of fundamental interest, especially for policy-makers and social welfare comparisons. As discussed in Imbens and Rubin (1997b), they show how to estimate the cumulative distribution function (CDF) for compliers using an IV estimator, then estimate the density functions for compliers based on smoothing and obtain nonnegative density functions by truncation as a post-estimation procedure. Abadie (2002) proposes methods to test the global hypothesis of distributional causal effects, such as distribution equivalence and stochastic dominance, based on the estimation of CDF for compliers. Abadie et al. (2002) develop methods for estimating the quantile treatment effects using main results in Abadie (2000), where the possibly negative weights in the target function are modified to be nonnegative based on conditional expectation which is estimated by smoothing techniques.

Estimating the compliers outcome distributions can also be considered as identifying latent distributions from mixture models, where data arise from a mixture of distributions. Identification and estimation problems in mixture models are common in many scientific studies. For example, mixture models are widely used in quantitative trait locus (QTL) studies to estimate the effect of a QTL predisposing a trait. Since the genotypes at the QTL is unknown, it's impossible to observe and estimate the effect of the QTL directly. Mixture models are widely used to identify the effect of the QTL using known markers. Ma and Wang (2012) characterize the complete class of consistent estimators for the latent distributions. They propose an estimator that reaches the semiparametric efficiency bound and provide an algorithm to implement the efficient estimator which requires kernel smoothing. Nonparametric maximum likelihood estimators (NPMLEs) can also be used to identify latent distributions in mixture models, however, two widely used NPMLEs are shown to be either inefficient or even inconsistent by Ma and Wang (2012).

None of the above estimators for the compliers distribution curves is guaranteed to be monotone or bounded between 0 and 1. A possible method that can obtain a monotone estimator from a non-monotone estimator is weighted isotonic regression. However, in a weighted isotonic regression, it's not apparent to choose the optimal weight and the asymptotic properties of the estimator are typically unknown. Lee et al. (2016) propose a binomial likelihood approach and a combined method of the EM algorithm and Pool-Adjacent-Violators Algorithm to obtain monotone distribution curves for compliers. However, the binomial likelihood method purposely discards correlation in the full likelihood to avoid known problems in the NPMLEs, so it is likely to be inefficient. We propose a novel estimator that is guaranteed to be monotone and bounded between 0 and 1. We derive the asymptotic properties and propose an extension to obtain semiparametric efficiency. The major idea is that for a possibly nonmonotone quantile function $\hat{Q}(u)$, $\hat{F}(y) = \int_0^1 I(\hat{Q}(u) \leq y)du$ is always monotone in y and bounded in $[0, 1]$. Therefore, based on this representation, our estimator is a natural CDF. It is related to the rearrangement operation from variational analysis (Hardy et al., 1952; Lorentz, 1953; Villani, 2003). Chernozhukov et al. (2009) studies the rearrangement estimator in the form of $\hat{G}(y) = \mathtt{inf}\{u \in R : \int I(\tilde{G}(v) \leq u)dv \geq y\}$, where $\tilde{G}(v)$ is the initial nonmonotone estimator. However, since $\tilde{G}(v)$ based on the conventional IV methods can be negative and larger than 1, $\hat{G}(y)$ may not be bounded in $[0, 1]$. On the other hand, our proposed CDF estimator directly rearranges a non-monotone quantile function whose domain is $[0, 1]$, and produces a CDF bounded in $[0, 1]$.

This chapter is organized as follows. In Section 2, we introduce notations, assumptions and discuss some previously established estimators. The proposed estimator, its asymptotic properties and a Wilcoxon-type test are discussed in Section 3. In Section 4, we discuss how to modify the proposed estimator to attain semiparametric efficiency. We present numerical results from several simulation experiments in Section 5. The proposed estimator is compared with the conventional IV estimator and the efficient estimator computed from the efficient score assuming known density functions. Type I error and power of the proposed Wilcoxon test are studied. Section 6 applies the proposed estimator to a randomized trial to study the efficacy of Modanifil on cocaine dependence. Some concluding remarks are given in Section 7. The proof of Theorem 1 is given in the last section.

2.2 Notation, Assumptions and Background

2.2.1 Notation

We consider a two-arm randomized trial with n subjects. Let π denote the probability of randomization to treatment, Z be the indicator of treatment assignment and its sample analog be $Z_i(i = 1, \ldots, n)$, where $Z_i = 1$ if subject i is assigned to treatment and $Z_i = 0$ if subject i is assigned to control. Let D_z be the potential treatment receive under randomization assignment z, $D = ZD_1 + (1 - Z)D_0$ be the actual treatment received and its sample analog be $D_i(i = 1, \ldots, n)$, where $D_i = 1$ if subject i receives treatment and $D_i = 0$ if subject i receives control.

Let Y_d be the potential outcome under treatment d, here we use the exclusion restriction assumption in Angrist et al. (1996) that the potential outcome is affected by treatment assignment only via treatment received, i.e., $Y_{d_z} = Y_d$. The observed outcome is $Y = ZY_1 + (1 - Z)Y_0$.

The population can be classified into four subpopulations based on their compliance status defined by the joint distribution of treatment assignment and potential treatment receipt: compliers if subjects always follow the assignment, never-takers if subjects always take the placebo, always-takers if subjects always take the treatment, defiers if subjects never follow the assignment. Let C denote the compliance classes: $C = 1$ (compliers) if $D_1 > D_0$; $C = 2$ (always-takers) if $D_1 = D_0 = 1$; $C = 0$ (never-takers) if $D_1 = D_0 = 0$; $C = 3$ (defiers) if $D_1 < D_0$. In practice, we can only observe one of D_0 and D_1, the compliance class membership for each subject is therefore unknown.

2.2.2 Assumptions

Assumption 1 *We impose standard IV assumptions similar to those in Angrist et al. (1996) and Abadie (2003):*

1. *Independence:* (Y_1, Y_0, D_1, D_0) *is jointly independent of* Z.

Table 2.1: The relationship between observed groups and latent compliance classes (C).

Z	D		C	
1	1	1 (Compliers)	or	2 (Always-takers)
1	0			0 (Never-takers)
0	0	1 (Compliers)	or	0 (Never-takers)
0	1			2 (Always-takers)

2. *Nontrivial assignment:* $\pi = P(Z = 1) \in (0, 1)$

3. *First-stage:* $E(D_1 - D_0) \neq 0$. *The causal effect of Z on D is nonzero.*

4. *Monotonicity:* $P(D_1 \geq D_0)$*=1, so there are no defiers who always receive the opposite treatment of assignment.*

Under Assumption 3 (1)-(4), a subject's compliance class membership can be partially identified based on the treatment assigned and the treatment received as in Table 3.1. In this way, we can state our problem as identifying latent distributions from mixture data. Let F_1 and F_0 be outcome CDF for compliers receiving treatment and control, F_{nt} be CDF for never-takers and F_{at} be CDF for always-takers. The observed CDF of each strata defined by Z and D can be expressed as a mixture of CDFs of latent compliance classes:

$$
\begin{aligned}
F_{Y|Z=1,D=1} &= \lambda_{c1} F_1 + \lambda_{at} F_{at} \\
F_{Y|Z=0,D=1} &= F_{at} \\
F_{Y|Z=0,D=0} &= \lambda_{c0} F_0 + \lambda_{nt} F_{nt} \\
F_{Y|Z=1,D=0} &= F_{nt}
\end{aligned}
$$

where $\lambda_{c1} = \frac{P(C=1)}{P(C=1)+P(C=2)}$,$\lambda_{c0} = \frac{P(C=1)}{P(C=1)+P(C=0)}$, $\lambda_{at} = \frac{P(C=2)}{P(C=1)+P(C=2)}$, and $\lambda_{nt} = \frac{P(C=0)}{P(C=1)+P(C=0)}$.

2.2.3 Abadie's weighted estimators

To estimate the cumulative distribution functions (CDFs) for compliers, we define the following functions of D and Z:

$$k_0 = (1-D)\frac{(1-Z)-(1-\pi)}{(1-\pi)\pi}$$
$$k_1 = D\frac{Z-\pi}{(1-\pi)\pi}$$
$$k = k_0(1-\pi) + k_1\pi = 1 - \frac{D(1-Z)}{1-\pi} - \frac{(1-D)Z}{\pi}.$$

Under Assumption 3 (1)-(4),Abadie (2003) proved the following lemma.

Lemma 2.2.1 *(Abadie) Let $g(Y, D)$ be any real function of Y, D such that $E|g(Y, D)| < \infty$,*

$$\mathrm{E}[g(Y_0, D)|C = 1] = \frac{1}{P(C=1)}E[k_0 g(Y, D)]$$
$$\mathrm{E}[g(Y_1, D)|C = 1] = \frac{1}{P(C=1)}E[k_1 g(Y, D)]$$
$$\mathrm{E}[g(Y, D)|C = 1] = \frac{1}{P(C=1)}E[k g(Y, D)].$$

This result allows us to identify functions of outcome distributions for compliers using a weighted sample. However, when $D \neq Z$, k takes negative value and is not a proper weight. Nonetheless, lemma (2.2.1) provides a simple way to estimate the CDFs of the potential outcomes for compliers. Let $g(Y, D) = I(Y \leq t)$, then CDF for compliers F_1 and F_0 can be identified as,

$$F_0(t) = \frac{1}{P(C=1)}E[k_0 I(Y \leq t)]$$
$$F_1(t) = \frac{1}{P(C=1)}E[k_1 I(Y \leq t)].$$

When $g(Y, D) = Y$, the sample counterparts of the corresponding equations are exactly the conventional instrumental variable (IV) estimators for CACE.

2.3 Estimation through quantile regression

While Abadie (2003) provides a simple way to estimate CDF for compliers, it can be non-monotone and may not be bounded in $[0,1]$ due to the negative weights. The estimator can also be very unstable in practice particularly when the instrument is weak. In this section, we introduce a new monotone CDF estimator based on a non-monotone quantile function estimator.

2.3.1 Distribution Estimation

Lemma (3.3.1) not only provides a simple way to estimate CDFs but also quantile functions for compliers. For $\theta \in (0,1)$, let $q_1(\theta)$ denote θ-quantile of Y_1 and $q_0(\theta)$ denote θ-quantile of Y_0 for compliers. There exists $\delta_\theta = (\alpha_\theta, \beta_\theta)$, such that the θ-quantile of Y given D for compliers,

$$q_\theta(Y|D, C = 1) = \alpha_\theta D + \beta_\theta,$$

where $\alpha_\theta = q_1(\theta) - q_0(\theta)$ and $\beta_\theta = q_0(\theta)$. The parameters α_θ and β_θ can be expressed as,

$$(\alpha_\theta, \beta_\theta) = \arg\min_{\alpha,\beta} \mathrm{E}(\rho_\theta(Y - \alpha D - \beta)|C = 1),$$

where $\rho_\theta(y) = y(\theta - I(y < 0))$ is often called the check function. Using Abadie's formula,

$$(\alpha_\theta, \beta_\theta) = \arg\min_{\alpha,\beta} \mathrm{E}(k\rho_\theta(Y - \alpha D - \beta)).$$

Therefore α_θ and β_θ can be estimated by minimizing,

$$\mathcal{Q}_n(\alpha, \beta) = \frac{1}{n}\sum_{i=1}^{n} k_i\rho_\theta(Y_i - \alpha D_i - \beta) = \frac{1}{n}\sum_{D_i=1} k_i\rho_\theta(Y_i - \alpha - \beta) + \frac{1}{n}\sum_{D_i=0} k_i\rho_\theta(Y_i - \beta).$$

Since $k_i = 1 - \frac{D_i(1-Z_i)}{1-\pi} - \frac{(1-D_i)Z_i}{\pi}$, when π is unknown, it can be estimated by $\hat{k}_i = 1 - \frac{D(1-Z)}{1-\hat{\pi}} - \frac{(1-D)Z}{\hat{\pi}}$, where $\hat{\pi}$ is the sample proportion of treatment assignment. Then, α_θ and β_θ are estimated by minimizing $\hat{\mathcal{Q}}_n(\alpha, \beta) = \frac{1}{n}\sum_{i=1}^{n} \hat{k}_i\rho_\theta(Y_i - \alpha D_i - \beta)$. Since $\hat{k}$ and k take negative values when D is not equal to Z, $\hat{\mathcal{Q}}_n$ and $\mathcal{Q}_n$ are typically nonconvex funtions. Optimizing nonconvex functions is not computationally efficient when the parameter space is

large. For computational considerations, Abadie et al. (2002) modifies the objective function by taking the conditional expectation of k given $U = (Y, D)$ which requires smoothing when Y is a continuous variable. As shown above, $\hat{\mathcal{Q}}_n(\alpha, \beta)$ or $\mathcal{Q}_n(\alpha, \beta)$ can be decomposed into two parts that can be minimized separately with data $D = 1$ and $D = 0$. Thus we can minimize the one-dimensional objective function by a simple grid search method.
Let $\hat{q}_1(\theta) = \hat{\alpha}_\theta + \hat{\beta}_\theta$ and $\hat{q}_0(\theta) = \hat{\beta}_\theta$. $F_1(t)$ and $F_0(t)$ can be estimated as,

$$\hat{F}_1(t) = \int_0^1 I(\hat{q}_1(u) \le t)du$$
$$\hat{F}_0(t) = \int_0^1 I(\hat{q}_0(u) \le t)du.$$

Though $\hat{q}_1$ and $\hat{q}_0$ are possibly non-monotone, $\hat{F}_1$ and $\hat{F}_0$ are nondecreasing functions and bounded in $[0, 1]$ by construction. In the next section, we derive asymptotic properties for the estimated quantile functions and CDFs.

2.3.2 Asymptotic properties and Inference

The following theorems summarize the asymptotic properties of $(\hat{\alpha}_\theta, \hat{\beta}_\theta)$ obtained from $\hat{\mathcal{Q}}_n(\alpha, \beta)$ when π is unknown or $(\tilde{\alpha}_\theta, \tilde{\beta}_\theta)$ from $\mathcal{Q}_n(\alpha, \beta)$ when π is known.

Theorem 2.3.1 *Let $\hat{\delta}_\theta = (\hat{\alpha}_\theta, \hat{\beta}_\theta)$ be solutions that minimize $\hat{\mathcal{Q}}_n(\delta_\theta)$, $\delta_\theta = (\alpha_\theta, \beta_\theta)$, $\epsilon_\theta = Y - X'\delta_\theta$, and $X = (D, 1)'$. Under Assumption 3 and regularity assumptions stated in the appendix, $\hat{\delta}_\theta$ is uniformly consistent of δ_θ for $\theta \in [0, 1]$, and $\sqrt{n}(\hat{\delta}_\theta - \delta_\theta) \Rightarrow$ to some Gaussian process $\mathcal{G}$, detailed in the appendix.*

Theorem 2.3.2 *Let $\tilde{\delta}_\theta = (\tilde{\alpha}_\theta, \tilde{\beta}_\theta)$ be solutions that minimize $\mathcal{Q}_n(\alpha, \beta)$, then $\tilde{\delta}_\theta$ is uniformly consistent of δ_θ and $\sqrt{n}(\tilde{\delta}_\theta - \delta_\theta) \Rightarrow$ to some Gaussian process $\tilde{\mathcal{G}}$, detailed in the appendix.*

Remark

For fixed θ, where π is unknown, $\sqrt{n}(\hat{\delta}_\theta - \delta_\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^n J^{-1}\psi_i(\theta) + o_p(1) \to N(\mathbf{0}, J^{-1}\Sigma_\theta J^{-1})$, and when π is known, $\sqrt{n}(\tilde{\delta}_\theta - \delta_\theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} J^{-1}\tilde{\psi}_i(\theta) + o_p(1) \to N(\mathbf{0}, J^{-1}\tilde{\Sigma}_\theta J^{-1})$, where,

$$
\begin{aligned}
J &= \mathrm{E}(f_{\epsilon_\theta|X,C=1}(0)XX'|C=1)P(C=1) \\
m(\delta_\theta) &= (\theta - I(Y - X'\delta_\theta < 0))X' \\
H(\delta_\theta) &= \mathrm{E}\left(m(\delta_\theta)\left(\frac{(1-D)Z}{\pi^2} - \frac{D(1-Z)}{(1-\pi)^2}\right)\right). \\
\psi(\theta) &= m(\delta_\theta)k + H(\delta_\theta)(Z-\pi) \\
\tilde{\psi}(\theta) &= m(\delta_\theta)k \\
\Sigma_\theta &= Var(\psi(\theta)) \\
\tilde{\Sigma}_\theta &= Var(\tilde{\psi}(\theta))
\end{aligned}
$$

It can be easily shown that $\Sigma_\theta \leq \tilde{\Sigma}_\theta$, where for symmetric matrices A and B the notation $A \leq B$ means $c'Ac \leq c'Bc$ for any nonzero vector c. Therefore, efficiency is improved when the estimated $\hat{\pi}$ is used regardless if π is known or unknown. The proof is detailed in the Appendix. The estimated θ-quantile functions for compliers $\hat{q}_1(\theta)$ and $\hat{q}_0(\theta)$, derived from $\hat{\delta}_\theta$ (or $\tilde{\delta}_\theta$) have the following asymptotic distribution:

Corollary 2.3.2.1 $\hat{q}_1(\theta) - q_1(\theta) \Rightarrow$ *some Gaussian processs and* $\hat{q}_1(\theta) - q_1(\theta) \Rightarrow$ *some Gaussian process.*

Theorem 2.3.3 $\sqrt{n}(\hat{F}_1(t) - F_1(t)) \Rightarrow$ *some Gaussian processs and* $\sqrt{n}(\hat{F}_0(t) - F_0(t)) \Rightarrow$ *some Gaussian processs.*

Remarks

At each point $\theta \in [0,1]$, $\sqrt{n}(\hat{F}_1(t) - F_1(t)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_{1i}(t) + o_p(1)$ and $\sqrt{n}(\hat{F}_0(t) - F_1(t)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \phi_{0i}(t) + o_p(1)$, where, the influence functions $\phi_1(t)$ and $\phi_0(t)$ of the proposed CDF estimators with estimated $\hat{\pi}$ are,

$$\begin{aligned}
\phi_0(t) &= -f_0(t) \cdot (0 \;\; 1) \cdot J^{-1}\psi(F_0(t)) \\
&= -f_0(t) \frac{1}{P(C=1)f_0(t)(1-\pi)} \cdot (-1 \;\; 1) \cdot \psi(F_0(t)) \\
&= \frac{1}{P(C=1)}(I(Y<t) - F_0(t))k_0 - \frac{Z-\pi}{P(C=1)(1-\pi)} \left(F_0(t) - F_{Y|C=0} \frac{P(C=0)}{\pi} \right) \\
\phi_1(t) &= -f_1(t) \cdot (1 \;\; 1) \cdot J^{-1}\psi(F_1(t)) \\
&= -f_1(t) \frac{1}{P(C=1)f_1(t)\pi} \cdot (1 \;\; 0) \cdot \psi(F_1(t)) \\
&= \frac{1}{P(C=1)}(I(Y<t) - F_1(t))k_1 + \frac{Z-\pi}{P(C=1)\pi} \left(F_1(t) - F_{Y|C=2}(t) \frac{P(C=2)}{\pi} \right).
\end{aligned}$$

Similarly, with true value of π, the influence functions $\tilde{\phi}_1(t)$ and $\tilde{\phi}_0(t)$ can be written as,

$$\begin{aligned}
\tilde{\phi}_0(t) &= -f_0(t) \cdot (0 \;\; 1) \cdot J^{-1}\tilde{\psi}(F_0(t)) = \frac{1}{P(C=1)}(I(Y<t) - F_0(t))k_0 \\
\tilde{\phi}_1(t) &= -f_1(t) \cdot (1 \;\; 1) \cdot J^{-1}\tilde{\psi}(F_1(t)) = \frac{1}{P(C=1)}(I(Y<t) - F_1(t))k_1.
\end{aligned}$$

The asymptotic variances $V_0(t)$ and $V_1(t)$ can be consistently estimated as

$$\begin{aligned}
\hat{V}_0(t) &= \frac{1}{n}\sum_{i=1}^{n} \hat{\phi}_{0i}(t)^2 \\
\hat{V}_1(t) &= \frac{1}{n}\sum_{i=1}^{n} \hat{\phi}_{1i}(t)^2,
\end{aligned}$$

where $\hat{\phi}_{0i}$ and $\hat{\phi}_{1i}$ are the influence functions of the ith observation with unknown quantities replaced by their consistent estimators.

2.3.3 Testing distribution equivalence

To evaluate the effect of treatment on the entire outcome distribution, it is often of interest to compare the equivalence of F_1 and F_0. The null hypothesis is

$$H_0 : F_1(t) = F_0(t).$$

A Wilcoxon-type statistic can be defined as

$$W_n = \sqrt{n}\left(\int \hat{F}_1 d\hat{F}_0 - \frac{1}{2}\right).$$

The following theorem gives the asymptotic distribution for the Wilcoxon-type statistic under the null hypothesis and thus provide a way to test the equivalence of F_1 and F_0.

Theorem 2.3.4 *Under the null hypothesis,*

$$W_n \rightarrow_d N(0, V),$$

where $V = \mathrm{E}\left(-\int \phi_0 dF_1 + \int \phi_1 dF_0\right)^2$ *can be consistently estimated as*

$$\hat{V} = \frac{1}{n}\sum_{i=1}^{n}\left(-\int \hat{\phi_{0i}}(t) d\hat{F}_1 + \int \hat{\phi_{1i}}(t) d\hat{F}_0\right)^2.$$

Under the alternative hypothesis that F_1 is stochastically greater or smaller than F_0, we expect the test to have decent power similar to the regular wilcoxon test.

2.4 Locally efficient estimators

In a framework of mixture models, Ma and Wang (2012) studied efficient estimation of the CDFs for the latent classes, which can be applied to our problem. Among subjects who receive treatment ($D = 1$), subjects with $Z = 1$ are mixture of always-takers and compliers with mixture probability $(\lambda_{at}, \lambda_{c1})$; subjects with $Z = 0$ are always-takers. Among subjects who receive control ($D = 0$), subjects with $Z = 0$ are mixed population of never-takers and compliers with mixture probability $(\lambda_{nt}, \lambda_{c0})$; subjects with $Z = 1$ are never-takers.

Define f_1 and f_0 be density functions for compliers, f_{nt} be density function for never-takers, and f_{at} be density function for always-takers. Let $\bar{f}_1(s) = (f_{at}(s), f_1(s))'$, $\bar{f}_0(s) = (f_{nt}(s), f_0(s))'$, $u_{11} = (\lambda_{at}, \lambda_{c1})'$, $u_{12} = (1, 0)'$, $u_{01} = (\lambda_{nt}, \lambda_{c0})'$, and $u_{02} = (1, 0)'$. The efficient influence function $\phi_{eff,0}(t)$ for compliers under control and $\phi_{eff,1}(t)$ for compliers under treatment are,

$$\phi_{eff,1}(t) = \frac{D}{\pi P(C=1)}(1(Y \le t) - K_{1,4}(t))Z + \frac{-D}{(1-\pi)P(C=1)}(1(Y \le t) - \tilde{K}_{1,4}(t))(1-Z)$$

$$\phi_{eff,0}(t) = \frac{D-1}{\pi P(C=1)}(1(Y \le t) - \tilde{K}_{0,4}(t))Z + \frac{1-D}{(1-\pi)P(C=1)}(1(Y \le t) - K_{0,4}(t))(1-Z),$$

where

$$K_1(t) = \int I(s \le t) A_1^{-1}(s) ds \left[\int A_1^{-1}(s) ds \right]^{-1} \equiv \begin{pmatrix} K_{1,1}(t) & K_{1,2}(t) \\ K_{1,3}(t) & K_{1,4}(t) \end{pmatrix}$$

$$K_0(t) = \int I(s \le t) A_0^{-1}(s) ds \left[\int A_0^{-1}(s) ds \right]^{-1} \equiv \begin{pmatrix} K_{0,1}(t) & K_{0,2}(t) \\ K_{0,3}(t) & K_{0,4}(t) \end{pmatrix}$$

$$\tilde{K}_{1,4}(t) = K_{1,4}(t) - \frac{P(C=1)}{P(C=2)} K_{1,3}(t)$$

$$\tilde{K}_{0,4}(t) = K_{0,4}(t) - \frac{P(C=1)}{P(C=0)} K_{0,3}(t)$$

$$A_1(s) = \left(\frac{u_{11} u_{11}^T}{u_{11}^T \bar{f}_1(s)} \right) P(Z=1|D=1) + \left(\frac{u_{12} u_{12}^T}{u_{12}^T \bar{f}_1(s)} \right) P(Z=0|D=1)$$

$$A_0(s) = \left(\frac{u_{01} u_{01}^T}{u_{01}^T \bar{f}_0(s)} \right) P(Z=0|D=0) + \left(\frac{u_{02} u_{02}^T}{u_{02}^T \bar{f}_0(s)} \right) P(Z=1|D=0).$$

The matrices $K_1(t)$ and $K_0(t)$ have direct relationship with the CDFs, that is, $K_1(t)(1\ 1)' = (F_{at}(t)\ F_1(t))'$ and $K_0(t)(1\ 1)' = (F_{nt}(t)\ F_0(t))'$ (Ma and Wang, 2012). We will use this fact later.

The semiparametric efficient estimator constructed from the efficient influence function involves complicated expressions and depend on unknown density functions of all compliance classes. If $\bar{f}_1(s)$ or $\bar{f}_0(s)$ is misspecified as $\bar{f}_1^*(s)$ or $\bar{f}_0^*(s)$, the influence functions are no longer mean-zero and do not produce consistent estimators. Ma and Wang (2012) suggest to robustify the influence function by constructing, $\phi_1 = \phi_{eff,1}^* - F_1 + (0\ 1)K_1^*(1\ 1)'$ and $\phi_0 = \phi_{eff,0}^* - F_0 + (0\ 1)K_0^*(1\ 1)'$, where $\phi_{eff,1}^*$, $\phi_{eff,0}^*$, K_1^* and K_0^* are defined with respect to working densities $\bar{f}_1^*$ and $\bar{f}_0^*$. The latent CDF $F_1(t)$ and $F_0(t)$ can be estimated as $\hat{F}_{eff,1}(t) = \frac{1}{n_1} \sum_{i:D_i=1} \phi_{eff,1i}^*(t) + (0\ 1)K_1^*(1\ 1)'$ and $\hat{F}_{eff,0}(t) = \frac{1}{n_0} \sum_{i:D_i=0} \phi_{eff,0i}^*(t) + (0\ 1)K_0^*(1\ 1)'$. This estimator is locally efficient in the sense that it reaches the semiparametric efficiency bound if the unknown densities are correctly specified or consistently estimated. With plugged-in estimators for unknown probabilities (i.e. $P(C=1), P(C=2), P(C=0)$ and π), some terms will be cancelled and the robustified estimator is exactly Abadie's IV estimator in our problem.

By comparing the influence functions $\phi_1(t)$ and $\phi_0(t)$ derived with true π in Section 3 to the efficient influence functions $\phi_{eff,1}(t)$ and $\phi_{eff,0}(t)$, we notice that the proposed estimator can be modified to reach the semiparametric efficiency bound through two modified quantile regression estimators. To obtain the efficient CDF estimator for F_1, we consider the quantile regression model:

$$\begin{aligned}(\alpha_{1,eff}(\theta), \beta_{1,eff}(\theta)) &= \text{arg min}_{\alpha,\beta}\text{E}(k\rho_{ZK_{1,4}+(1-Z)(K_{1,4}-\frac{P(C=1)}{P(C=2)}K_{1,3})}(Y-\alpha D-\beta)) \\ &= \text{arg min}_{\alpha,\beta}\text{E}(Z\rho_{K_{1,4}}(Y-\alpha D-\beta)+(1-Z)\rho_{\widetilde{K_{1,4}}}(Y-\alpha D-\beta)),\end{aligned}$$

where $\widetilde{K_{1,4}} = K_{1,4} - \frac{P(C=1)}{P(C=2)}K_{1,3}$. Similarly, to obtain the efficient CDF estimator for F_0, the modified quantile regression model is:

$$\begin{aligned}(\alpha_{0,eff}(\theta), \beta_{0,eff}(\theta)) &= \text{arg min}_{\alpha,\beta}\text{E}(k\rho_{(1-Z)K_{0,4}+Z(K_{0,4}-\frac{P(C=1)}{P(C=0)}K_{0,3})}(Y-\alpha D-\beta)) \\ &= \text{arg min}_{\alpha,\beta}\text{E}((1-Z)\rho_{K_{0,4}}(Y-\alpha D-\beta)+Z\rho_{\widetilde{K_{0,4}}}(Y-\alpha D-\beta)),\end{aligned}$$

where $\widetilde{K_{0,4}} = K_{0,4} - \frac{P(C=1)}{P(C=0)}K_{0,3}$.

Distribution estimators transformed from

$$\hat{q}_{1,eff}(\theta) = \hat{\alpha}_{1,eff}(\theta) + \hat{\beta}_{1,eff}(\theta)$$

and

$$\hat{q}_{0,eff}(\theta) = \hat{\beta}_{0,eff}(\theta)$$

then reach the semiparametric efficiency bound when the unknown densities are correctly specified. Note that in the modified quantile regression models, K_1 and K_0 are functions of θ, i.e. $K_1 = K_1(q_1(\theta))$ and $K_0 = K_0(q_0(\theta))$. On the other hand, K_1 and K_0 are functions of latent density functions $\bar{f}_0$ and $\bar{f}_1$, hence require working densities as in Ma and Wang (2012). However, this estimator is robust with respect to misspecification of the latent density functions. In the following, we show that when $\bar{f}_0$ and $\bar{f}_1$ are misspecified as $\bar{f}_0^*$ and $\bar{f}_1^*$, the corresponding estimators $\hat{q}^*_{1,eff}(\theta)=\hat{\alpha}^*_{1,eff}(\theta)+\hat{\beta}^*_{1,eff}(\theta)$ and $\hat{q}^*_{0,eff}(\theta)=\hat{\beta}^*_{0,eff}(\theta)$ still consistently estimate $q_1(\theta)$ and $q_0(\theta)$.

In fact, when $\bar{f}_0$ is misspecified as $\bar{f}_0^*$, $(\hat{\alpha}^*_{0,eff}(\theta), \hat{\beta}^*_{0,eff}(\theta))$ consistently estimate

$$(\alpha^*_{0,eff}(\theta), \beta^*_{0,eff}(\theta)) \equiv \arg\min_{\alpha,\beta} \mathrm{E}(k\rho_{(1-Z)K^*_{0,4}+Z(K^*_{0,4}-\frac{P(C=1)}{P(C=0)}K^*_{0,3})}(Y-\alpha D-\beta))$$

First-order conditions yield,

$$\beta^*_{0,eff}(\theta) = q_0(K^*_{0,3} + K^*_{0,4})$$

$$\alpha^*_{0,eff}(\theta) = q_1\left(\theta - \frac{1}{P(C=0)}K^*_{0,3}\right) - q_0(\theta).$$

However, $K^*_{0,3} + K^*_{0,4} = (0,\ 1)K^*_0(q^*_0(\theta))(1,\ 1)' = F^*_0(q^*_0(\theta)) = \theta$. Therefore $\hat{\beta}^*_{0,eff}(\theta)$ consistently estimates $\beta^*_{0,eff}(\theta) = q_0(\theta)$. Similarly, when f^*_1 is misspecified, $\hat{q}^*_{1,eff}(\theta) = \hat{\alpha}^*_{1,eff}(\theta) + \hat{\beta}^*_{1,eff}(\theta)$ consistently estimate $\alpha^*_{1,eff}(\theta) + \beta^*_{1,eff}(\theta) = q_1(\theta)$. Hence the modified CDF estimators are consistent though may not be efficient under misspecification.

2.5 Simulation studies

In this section, we study the finite sample performance of the proposed estimator and compare it to Abadie's IV estimator and oracle efficient estimator with the correct specification of unknown density functions. We conduct several simulation scenarios. For each scenario, we generate 1000 independent datasets. In the first simulation experiment, we evaluate the sampling error of the four estimators on estimating CDFs at a single point (the 95th quantile) and averaged discrepancy for the entire distribution (to be defined later). The outcome distributions in four latent compliance classes follow Gamma or log-normal distribution. Specifically, the means for each latent class are $E(Y|C=2) = 4$, $E(Y_1|C=1) = 2$, $E(Y_0|C=1) = 2$, $E(Y|C=0) = 1$ and variances are 1. The randomization probability is 0.5. We consider two scenarios corresponding to a strong and a weak IV. The strength of an IV is how strong the association between the IV and the actual treatment is. When the first stage F-statistic from the regression of the treatment on the IV is less than 10, the IV is commonly considered to be a weak IV (Stock et al., 2012). The compliance class status C is randomly generated as (a) strong IV: $P(C=0) = 0.25$, $P(C=1) = 0.5$ and $P(C=2) = 0.25$ (average first stage F-statistic=25, when $n = 50$); (b) weak IV: $P(C=0) = 0.45$, $P(C=1) = 0.1$ and

$P(C = 2) = 0.45$ (average first stage F-statistic=6 when $n = 500$). The efficient estimator and the modified quantile estimator are constructed using the true densities as the working density. To assess each estimator's average performance for the whole distribution, we define L_2 distance as the measurement of the discrepancy between the estimator $\hat{F}$ and the true CDF F,

$$L_2(F, \hat{F}) = \int [F(y) - \hat{F}(y)]^2 dF(y).$$

Table 2.2 summarizes the sampling bias, sampling standard deviation of point estimates and L_2 distance for the four distribution estimators under different sample sizes. Under the strong IV setting, all four estimators have a negligible bias for point estimates at the 95th quantiles. The proposed estimator has a much smaller standard deviation than the IV estimator in finite samples. In terms of averaged discrepancy (AD), the proposed estimator performs similar to efficient estimator and modified efficient estimator but much better than the IV estimator especially for $n = 50$ and $n = 100$. Though results are not shown here, we compare the estimated standard error (SE) with the averaged sampling standard deviation (SD). SE slightly overestimated the variation of the proposed estimator when the sample size is small, but when the sample size is increasing, SE is closer to SD. Under the weak IV setting, the proposed estimator performs much better than the IV estimator and performs similar to the efficient estimator and modified quantile estimator which uses the information of the true CDFs. The IV estimator is really unstable and is often unbounded in most simulation iterations when estimating the 95th quantile. We don't report the exact bias and sampling standard deviation since they are very large. The average discrepancy of the IV estimator is dramatically larger and more variable than the other three estimators.

In the second simulation experiment, under the strong IV settings, we compare CACE estimation using the proposed estimator, IV estimator and efficient estimator. Table 2.3 summarizes the sampling bias, sampling standard deviation for CACE estimates from the three distribution estimators. In the CACE estimation, we compare the proposed estimator with the IV estimator and the efficient estimator since the modified quantile estimator performs similar to the proposed estimator in finite samples. All three CACE estimators have a low bias in finite samples. The IV estimator has the highest variability in all settings. When the sample size is 200, the efficient estimator and the proposed estimator have similar efficiency.

In the third simulation experiment, we study the type I error and power of the Wilcoxon test. We also check the type I error of the Wilcoxon test for the intent-to-treat (ITT) estimator under the setting that $F_1 = F_0$. To study the power of the Wilcoxon test for the proposed estimator, the mean of the compliers outcome distribution is set as $E(Y_1|C = 1) = E(Y_0|C = 1) + \Delta$ and $E(Y_0|C = 1) = 1$, with $\Delta = 1, 1.5, 2$. For each scenario, we generate 1000 independent datasets. As shown by Table 2.4, the Wilcoxon test for the proposed estimator has a type I error slightly smaller than the nominal 0.05 level in finite samples. The Wilcoxon test for the ITT estimator does not yield correct type I errors since the CDFs for subjects assigned to treatment and control are not comparable under the null hypothesis that the CDF for compliers with treatment is the same as the CDF for compliers with control. When the sample size is 100, the power for testing a unit difference in mean outcome as Gamma distribution or Log-normal distribution is higher than 80%.

2.6 Real data example

In this section, we apply the proposed method to analyze a randomized trial of an intervention to test the efficacy of modafinil in facilitating abstinence in cocaine-dependent patients (Anderson et al., 2009). This study is a double-blind placebo-controlled study. We analyze data from 127 cocaine-dependent patients who have self-reported information on medication compliance status and daily cocaine use, where 62 patients were randomized to placebo (control group) and 65 patients to modafinil 200 mg (treatment group). Patients were required to take the drug once daily on awakening and record the time of taking the drug. These patients came to the clinic three times weekly to provide a urine sample and other research data. The primary outcome was the median of the log_{10} urine benzoylecgonine (BE) concentration at Week 3 when the percent of group retainment is about 89%. Treatment compliance was evaluated based on the percent of days medication was taken (determined

Table 2.2: Estimates of compliers CDFs F_1 and F_0 at 95th quantiles. Bias, sampling standard deviation (SD), averaged discrepancy (AD) and its sampling standard deviation (SD^{ad}), $\times 10^2$, sample size=50, 100 and 200.

	Proposed				IV				Efficient				Modified quantile			
	Bias	SD	AD	SD^{ad}	Bias	SD	AD	SD^{ad}	Bias	SD	AD	SD^{ad}	Bias	SD	AD	SD^{ad}
n=50, strong IV																
F_1^{Γ}	-5.1	11.3	2.6	2.6	1.9	25.1	4.7	19.8	-5.1	12.1	2.0	1.7	-7.1	13.2	2.1	1.9
F_0^{Γ}	-1.4	7.8	2.5	2.8	0.3	10.1	3.4	5.5	-1.0	7.3	2.6	2.4	-1.9	8.3	2.8	3.0
F_1^{LN}	-4.3	11.3	2.7	3.0	2.6	23.0	4.7	12.2	-4.4	12.2	2.1	2.0	-6.4	13.0	2.1	2.0
F_0^{LN}	-1.8	7.3	2.6	2.7	0.2	9.9	3.6	5.1	-1.6	7.3	2.8	2.4	-2.0	7.5	2.7	2.7
n=100, strong IV																
F_1^{Γ}	-3.4	8.8	1.4	1.5	0.8	14.8	1.8	2.7	-0.4	9.1	1.1	0.9	-1.1	9.3	1.1	1.0
F_0^{Γ}	-0.8	5.2	1.4	1.4	0.3	6.4	1.6	1.6	-5.1	5.2	1.4	1.2	-7.1	5.4	1.5	1.6
F_1^{LN}	-3.1	8.1	1.4	1.5	0.9	14.2	1.8	3.0	-2.5	8.5	1.0	0.9	-4.3	9.3	1.1	1.0
F_0^{LN}	-1.1	5.4	1.4	1.3	0.1	6.5	1.5	1.5	-0.1	5.3	1.4	1.2	-1.4	5.5	1.4	1.3
n=200, strong IV																
F_1^{Γ}	-1.7	6.2	0.7	0.7	-0.8	9.3	0.8	0.8	-1.1	6.2	0.5	0.6	-2.3	6.4	0.6	0.6
F_0^{Γ}	-5.1	4.0	0.7	0.7	1.9	4.6	0.7	0.7	-0.2	4.1	0.7	0.7	-0.9	4.1	0.7	0.7
F_1^{LN}	-1.9	6.6	0.7	0.8	0.6	9.9	0.8	11.4	-1.4	6.6	0.5	0.5	-2.5	6.9	0.6	0.5
F_0^{LN}	-0.7	3.9	0.7	0.6	-0.04	4.4	0.7	0.6	-0.2	3.9	0.7	0.6	-1.0	4.0	0.7	0.6
n=500, weak IV																
F_1^{Γ}	-8.0	16.4	4.7	5.9	*	*	110.1	1155.7	-9.7	20.3	2.2	2.1	-13.8	20.7	2.7	3.5
F_0^{Γ}	-4.4	14.8	5.8	6.4	*	*	35.4	242.8	-3.2	9.9	6.1	5.2	-4.5	14.9	5.8	6.4
F_1^{LN}	-7.7	15.1	4.6	5.8	*	*	88.7	667.8	-9.4	19.4	2.1	1.8	-13.4	20.6	2.6	3.7
F_0^{LN}	-4.8	15.2	5.8	6.6	*	*	50.4	494.2	-2.9	9.4	5.1	5.3	-4.9	15.2	6.8	6.6

Table 2.3: Estimates of the CACE under the strong IV setting. Bias, sampling standard deviation (SD), averged estimated standard error (SE), $\times 10^2$, sample size=50, 100 and 200.

	Proposed estimators		IV estimators		Efficient estimators	
	Bias	SD	Bias	SD	Bias	SD
			n=50			
Gamma	7.4	65.0	-7.2	92.8	10.6	56.5
Lognormal	3.0	64.0	-12.4	89.4	6.4	57.5
			n=100			
Gamma	4.6	46.6	-3.2	58.3	6.7	40.7
Lognormal	2.8	45.3	-4.3	56.7	4.2	38.8
			n=200			
Gamma	1.2	33.1	-2.8	37.7	2.5	28.4
Lognormal	1.3	34.1	-2.5	39.3	3.7	30.0

Table 2.4: Type I error of the Wilcoxon test for the proposed estimator and ITT estimator, and the power of the Wilcoxon test for the proposed estimator.

	Type I error		Power		
	Proposed	ITT	$\Delta = 1$	$\Delta = 1.5$	$\Delta = 2$
		n=50			
Gamma	4.8	31.8	63.8	83.4	93.1
Lognormal	5.6	34.6	68.8	86.9	93.1
		n=100			
Gamma	5.0	33.8	83.6	96.1	99.5
Lognormal	4.5	35.6	88.1	97.8	99.5
		n=200			
Gamma	4.2	32.6	96.4	100	100
Lognormal	4.1	32.8	98.5	99.9	100

Table 2.5: Compliance classes and sample sizes.

n	Z	D	C		
30	1	1	1 (Compliers)		
35	1	0			0 (Never-takers)
62	0	0	1 (Compliers)	or	0 (Never-takers)

by self-report at each study visit). Taking medication daily as instructed until Week 3 is considered as complying to the assigned treatment.

We apply the proposed method to estimate the CDFs for compliers in the modafinil 200 mg group and placebo group. In this study, patients in the placebo group didn't have access to the treatment medication. Patients who actually took treatment medication were compliers of which the CDF can be identified directly. The observed groups and latent compliance classes are listed in Table 2.5.

Figure 1 shows the proposed estimators and Abadie's IV estimators of CDFs for compliers. Since there are no always takers, the IV estimator for compliers with treatment overlaps with the proposed CDF estimator. For compliers in the control group, Abadie's IV estimator is not monotone and the proposed estimator smoothes the IV estimator at its wiggling regions in a data-adaptive manner, and is also guaranteed to be between 0 and 1. The IV estimator has a larger bootstrap standard error than the proposed estimator at the non-monotone regions. The bootstrap standard error for the IV estimator of $F_0(4)$ is 0.18 and the bootstrap standard error for the proposed estimator of $F_0(4)$ is 0.15. We noticed that at week 3, the estimated F_1 is stochastically greater than F_0, which indicates that among compliers the average BE concentration level in the treatment group is lower than that in the placebo group. Since the two estimated functions are not crossing, the Wilcoxon test can be used as a powerful test to test the equivalence of the two CDFs. The estimated F_1 and F_0 using the proposed method are significantly different (P-value=0.04, Wilcoxon test statistic= 2.4). At 95% significance level, we can conclude that the data provide enough

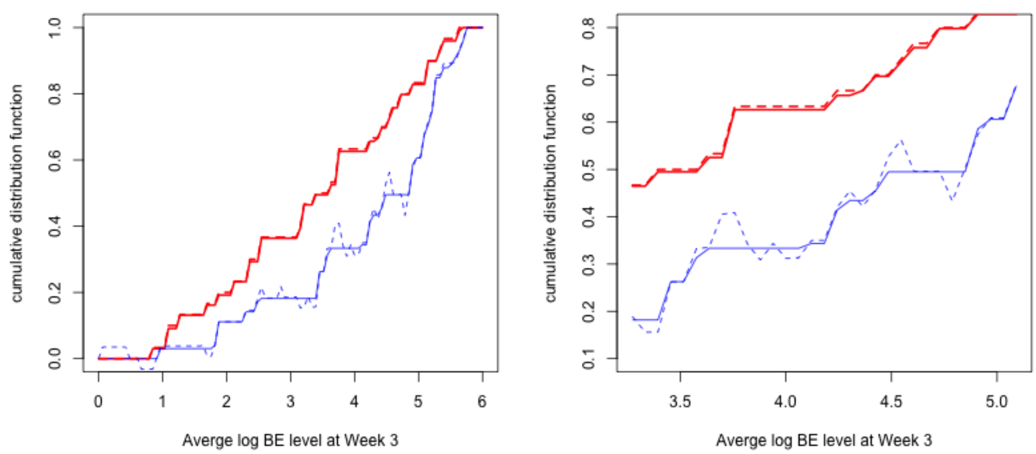


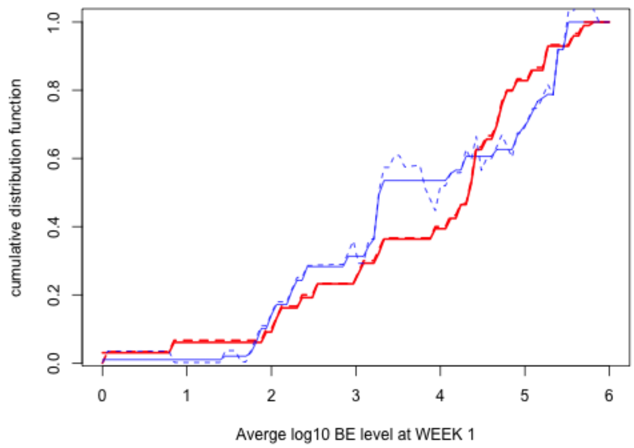
Figure 2.1: Left panel: distribution function estimates of BE level at Week 3. Right panel: a nonmonotone part of distribution function estimates of BE level at Week 3. Blue solid line is estimated CDF for compliers in placebo group using the proposed method, the red solid line is the estimated CDF for compliers in 200 mg modafinil treatment group using the proposed method. Dashed lines are the corresponding IV estimators.

information to reject the null hypothesis that F_1 and F_0 are the same.

We also examine urine BE concentration level at Week 1 and Week 2 by plotting CDF estimators for the treatment group and control group in Figure 2. The discrepancy between the two groups is increasing from Week 1 to Week 3. The estimated CDF for treated is crossing with the estimated CDF for control. The estimated CDFs are not significantly different between the treatment group and control group for controls at Week 1 (P-value=0.44) or Week 2 (P-value=0.10).

2.7 Conclusion

We study a novel monotone cumulative distribution function estimator of an outcome variable for compliers receiving treatment or control. The estimation procedures involve a weighted quantile regression and a post-estimation rearrangement adjustment. We show that the proposed estimator is $\sqrt{n}$-consistent and develop large sample properties. Based on the



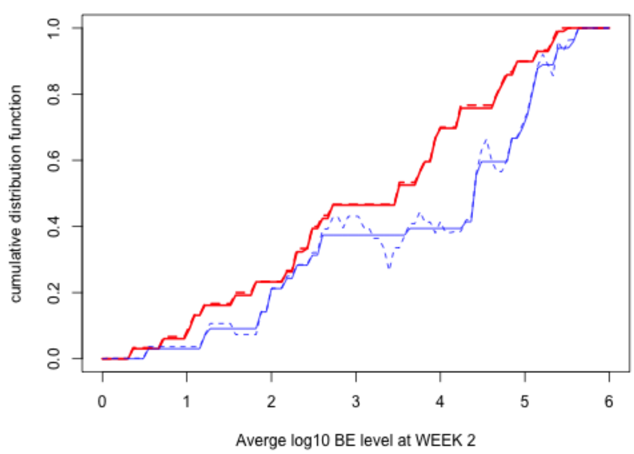


Figure 2.2: Distribution function estimates of BE level at Week 1 (left panel) and Week 2 (right panel). Blue solid lines are estimated CDFs for compliers in placebo group using the proposed method, red solid lines are the estimated CDFs for compliers in 200 mg modafinil treatment group using the proposed method. Dashed lines are the corresponding IV estimators.

asymptotic properties of the proposed estimator, a Wilcoxon-type statistic is proposed to test the equivalence of CDF for compliers receiving treatment and control. By comparing the influence function of the proposed estimator to the efficient influence function, we modify the proposed estimator and obtain a local efficient and robust estimator in the sense that when the unknown density functions are correctly specified, it reaches the semiparametric efficiency bound and when the unknown density functions are misspecified, it is still a consistent estimator. The performances of the proposed estimator, the conventional IV estimator, efficient estimator and the modified estimator under correct specification of unknown density functions are evaluated by simulation studies under different scenarios. In finite samples, the proposed estimator performs similar to the efficient estimator and the modified estimator and has a much smaller standard deviation than the IV estimator especially when the instrument variable is weak.

2.8 Proofs

2.8.1 Proof of Theorem 1 and Theorem 2

Assumption 2 *We assume the following regularity assumptions:*

1. *the data are i.i.d.;*

2. *Y_1 and Y_0 for each compliance class are continuously distributed with a compact support Ω and density functions bounded away from zero and infinity on the support.*

Let $\mathcal{Q}_0(\delta_\theta) = E(k\rho_\theta(Y - X'\delta))$. $\mathcal{Q}_0(\delta_\theta)$ is continuous and uniquely minimized at δ_θ. Before proving Theorem 1 and Theorem 2, we first give point-wise asymptotic results of $\hat{\delta}_\theta$. First, we would likt to show that $\hat{\delta}_\theta \to \delta_\theta$ in $\theta \in [0,1]$.

It suffices to show $\hat{Q}_n(\delta_\theta)$ converges in probability to $Q_0(\delta_\theta)$ and a 'well-separated' miminum condition for $Q_0(\delta_\theta)$ (Van der Vaart, 2000, Theorem 5.7). Since

$$\sup_{\delta\in\Omega,\theta\in[0,1]}|\hat{\mathcal{Q}}_n(\delta_\theta) - \mathcal{Q}_0(\delta_\theta)| \leq \sup_{\delta\in\Omega,\theta\in[0,1]}|\mathcal{Q}_n(\delta_\theta) - \mathcal{Q}_0(\delta_\theta)| + \sup_{\delta\in\Omega,\theta\in[0,1]}|\hat{\mathcal{Q}}_n(\delta) - \mathcal{Q}_n(\delta)|,$$

we can show this by bounding each term on the right hand.

In fact, the first right term $\sup_{\delta\in\Omega,\theta\in[0,1]}|\mathcal{Q}_n(\delta_\theta) - \mathcal{Q}_0(\delta_\theta)| = \sup|\frac{1}{n}\sum k_i\rho_\theta(Y_i - X_i'\delta) - E(k\rho_\theta(Y - X'\delta))| = o_p(1)$. This can be verified by showing that

$$|\rho_\theta(Y - X'\delta_1)k - \rho_\theta(Y - X'\delta_2)k| \leq max(k) \times ||X|| \times ||\delta_1 - \delta_2|| < \infty$$

for $\delta_1, \delta_2 \in \Omega$, where $X = (D,\ 1)$ and $||\cdot||$ denotes the Euclidean norm. Thus $\sup|\mathcal{Q}_n(\delta) - \mathcal{Q}_0(\delta)| = o_p(1)$ following Theorem 19.4 of Van der Vaart (2000). The second right term

$$\begin{aligned}
\sup_{\delta\in\Omega}|\hat{\mathcal{Q}}_n(\delta) - \mathcal{Q}_n(\delta)| &= \sup|\frac{1}{n}\sum \rho_\theta(Y_i - X_i'\delta)(\hat{k}_i - k_i)| \\
&= |\hat{\pi} - \pi| \times \sup|\frac{1}{n}\sum \rho_\theta(Y_i - X_i'\delta)\left(\frac{Z_i(1-D_i)}{\hat{\pi}\pi} - \frac{(1-Z_i)D_i}{(1-\hat{\pi})(1-\pi)}\right)| \\
&= |\hat{\pi} - \pi| \times \sup|E(\rho_\theta(Y - X'\delta)\left(\frac{Z(1-D)}{\pi^2} - \frac{(1-Z)D}{(1-\pi)^2}\right) + o_p(1)| \\
&= o_p(1),
\end{aligned}$$

where $\hat{\pi} = \frac{1}{n}\sum Z_i$. Thus $\sup_{\delta\in\Omega}|\hat{\mathcal{Q}}_n(\delta) - \mathcal{Q}_0(\delta)| = o_p(1)$. Hence, $\hat{\delta}_\theta$ is a consistent estimator of δ_θ.

We then prove the asymptotic distribution of $\hat{\delta}_\theta$. The first order condition writes as $\frac{\partial}{\partial\delta}\hat{\mathcal{Q}}_n(\delta) = \frac{1}{n}\sum_{i=1}^n(\theta - 1(Y_i < X_i'\delta))X_i'\hat{k}_i = 0$. Define $D(\delta) = \mathrm{E}\left((\theta - 1(Y < X'\delta))X'k\right)$ and $\hat{D}(\delta) = \mathrm{E}\left((\theta - 1(Y < X'\delta))X'\hat{k}\right)$.

$$\begin{aligned}
0 &= \sqrt{n}\frac{\partial}{\partial\delta}\hat{\mathcal{Q}}_n(\hat{\delta}_\theta) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n\left[(\theta - 1(Y_i < X_i'\hat{\delta}_\theta))X_i'\hat{k}_i - \hat{D}(\hat{\delta}_\theta)\right] + \sqrt{n}\left(\hat{D}(\hat{\delta}_\theta) - \hat{D}(\delta_\theta)\right)
\end{aligned}$$

$$
\begin{aligned}
-\sqrt{n}\hat{D}_n &\equiv \frac{1}{\sqrt{n}}\sum(\theta - 1(Y_i < X_i'\delta_\theta))X_i'\hat{k}_i \\
&= \frac{1}{\sqrt{n}}\sum(\theta - 1(Y_i < X_i'\delta_\theta))X_i'k_i + \frac{1}{\sqrt{n}}\sum(\theta - 1(Y_i < X_i'\delta_\theta))X_i'(\hat{k} - k_i) \\
&= \frac{1}{\sqrt{n}}\sum(\theta - 1(Y_i < X_i'\delta_\theta))X_i'k_i \\
&\quad + \sqrt{n}(\hat{\pi} - \pi)\frac{1}{n}\sum(\theta - 1(Y_i < X_i'\delta_\theta))X_i'\left(\frac{Z_i(1-D_i)}{\hat{\pi}\pi} - \frac{(1-Z_i)D_i}{(1-\hat{\pi})(1-\pi)}\right) \\
&= \frac{1}{\sqrt{n}}\sum(\theta - 1(Y_i < X_i'\delta_\theta))X_i'k_i \\
&\quad + \sqrt{n}(\hat{\pi} - \pi)E\left[(\theta - 1(Y < X\delta_\theta))X'\left(\frac{Z(1-D)}{\pi^2} - \frac{(1-Z)D}{(1-\pi)^2}\right)\right] + o_p(1).
\end{aligned}
$$

Then, by the central limit theorem $\sqrt{n}\hat{D}_n \to N(0, \Sigma_\theta)$, where

$$
\begin{aligned}
\Sigma_\theta &= Var(m(\delta_\theta)k + H(\delta_\theta)(Z - \pi)) \equiv Var(\psi(\theta)) \\
m(\delta_\theta) &= (\theta - 1(Y - X'\delta_\theta))X' \\
H(\delta_\theta) &= E\left[(\theta - 1(Y < X\delta_\theta))X'\left(\frac{Z(1-D)}{\pi^2} - \frac{(1-Z)D}{(1-\pi)^2}\right)\right]
\end{aligned}
$$

In fact, $\hat{\mathcal{Q}}_n(\delta)$ is linear in $\delta \in \Omega$ almost everywhere. With probability 1, $\hat{\mathcal{Q}}_n(\delta) - \hat{\mathcal{Q}}_n(\delta_\theta) - \hat{D}_n(\delta - \delta_\theta) = 0$ for $||\delta - \delta_\theta|| \leq \epsilon$. Since $\mathcal{Q}_0(\delta)$ is twice differentiable, $\mathcal{Q}_0(\delta) - \mathcal{Q}_0(\delta_\theta) = \frac{1}{2}J||\delta - \delta_\theta||^2 + o(||\delta - \delta_\theta||^2)$. Thus $\sup_{||\delta - \delta_\theta|| \leq \delta_n} |\hat{R}_n(\theta)/[1 + \sqrt{n}||\delta - \delta_\theta||]\,| \to 0$ in probability for any $\delta_n \to 0$. It follows from Theorem 7.1 of Newey and McFadden (1994) that $\sqrt{n}(\hat{\delta} - \delta) \to N(0, J^{-1}\Sigma_\theta J^{-1})$, where $J = \mathrm{E}(f_{\epsilon_\theta|X,C=1}(0)XX'|C = 1)P(C = 1)$ is the second derivative of $\mathcal{Q}_0(\delta)$ at δ_θ. Then we conclude that $\sqrt{n}(\hat{\delta} - \delta) \to N(0, J^{-1}\Sigma_\theta J^{-1})$.

When the randomization probability π is known, the asymptotic distribution of $\tilde{\delta}_\theta$ can be derived in a similar way that $\sqrt{n}(\tilde{\delta}_\theta - \delta_\theta) \to N(0, J^{-1}\tilde{\Sigma}_\theta J^{-1})$, where $\tilde{\Sigma}_\theta = Var(m(\delta_\theta)k)$.

Next we show that $\Sigma_\theta \leq \tilde{\Sigma}_\theta$.

$$
\begin{aligned}
\Sigma_\theta &= Var(m(\delta_\theta)k + H(\delta_\theta)(Z - \pi)) \\
&= Var(m(\delta_\theta)k) + Var(H(\delta_\theta)(Z - \pi)) + 2Cov(m(\delta_\theta)k, H(\delta_\theta)(Z - \pi)),
\end{aligned}
$$

The two components can be rewritten as,

$$
\begin{aligned}
Var(H(\delta_\theta)(Z - \pi)) &= H(\delta_\theta)(1 - \pi)\pi H(\delta_\theta)' \\
Cov(m(\delta_\theta)k, H(\delta_\theta)(Z - \pi)) &= E(m(\delta_\theta)kH(\delta_\theta)'(Z - \pi)) = E(m(\delta_\theta)k(Z - \pi))H(\delta_\theta)'.
\end{aligned}
$$

where

$$H(\delta_\theta) = E(m(\delta_\theta)(\frac{Z(1-D)}{\pi^2} - \frac{(1-Z)D}{(1-\pi)^2}))$$
$$= \begin{pmatrix} -(\theta - F_{Y|C=2}(\alpha+\beta))\frac{P(C=2)}{1-\pi} \\ (\theta - F_{Y|C=0}(\beta))\frac{P(C=0)}{\pi} - (\theta - F_{Y|C=2}(\alpha+\beta))\frac{P(C=2)}{1-\pi} \end{pmatrix}$$

After some algebra calculation and cancellation of terms, $E(m(\delta_\theta)k(Z-\pi))$ can be simplified as

$$E(m(\delta_\theta)k(Z-\pi)) = \begin{pmatrix} \pi P(C=2)(\theta - F_{Y|C=2}(\alpha+\beta)) \\ \pi P(C=2)(\theta - F_{Y|C=2}(\alpha+\beta)) - (1-\pi)P(C=0)(\theta - F_{Y|C=0}(\beta)) \end{pmatrix}$$
$$= -(1-\pi)\pi H(\delta_\theta).$$

Since for any nonzero vector $c \in R^2$, $cH(\delta_\theta)H(\delta_\theta)'c' \geq 0$,

$$\begin{aligned} \Sigma_\theta &= Var(m(\delta_\theta)k + H(\delta_\theta)(Z-\pi)) \\ &= Var(m(\delta_\theta)k) - (1-\pi)\pi H(\delta_\theta)H(\delta_\theta)' \\ &\leq Var(m(\delta_\theta)k) \\ &= \tilde{\Sigma}_\theta \end{aligned}$$

Therefore, plugging an $\sqrt{n}$-consistent estimate $\hat{\pi}$ of π can improve the efficiency of the distributional estimation.

We now prove uniform consistency and asymptotic normality of $\hat{\delta}(\theta) = (\hat{\alpha}(\theta), \hat{\beta}(\theta))'$, via empirical process arguments for Z-estimators. For ease of notation, we may discard θ for $\delta(\theta)$. We present results for known π, proof is similar for results with $\hat{\pi}$ and thus omitted.

Let Δ be the linear space of all quantile functions. We will use the uniform norm $\|\cdot\|_\infty$ on Δ. For vectors in Δ^2, $\|\cdot\|_\infty$ is element-wise norm. Define $\Psi_n(\delta)(\theta) = D_n(\delta, \theta) = \frac{1}{n}\sum(\theta - 1(Y_i < X_i'\delta))X_i'k_i$ with the form $\Psi_n(\delta)(\theta) = \mathbb{P}_n\psi_{\delta,\theta}$, where

$$\psi_{\delta,\theta}(Y_i, D_i, Z_i) = ((\theta - 1(Y_i < X_i'\delta))X_i'k_i)$$

and $X_i = (D_i, 1)$ for $i = 1, \ldots, n$.

Define $\Psi(\delta)(\theta) = P\psi_{\delta,\theta} = \mathrm{E}((\theta - I(Y < X'\delta))Xk)$. To show the uniform consistency of $\hat{\delta}$, we establish an identifiability condition for Ψ and uniform consistency of Ψ_n.

The identifiability condition holds if for any sequence $\{\delta_n\} \in \Delta^2$, $\|\Psi(\delta_n)\|_\infty \to \mathbf{0}$ implies $\|\delta_n - \tilde{\delta}\|_\infty \to \mathbf{0}$, where $\Psi(\tilde{\delta}) = \mathbf{0}$.

$$\begin{aligned}
\Psi(\delta_n)(\theta) &= \Psi(\delta_n)(\theta) - \Psi(\tilde{\delta})(\theta) \\
&= \mathrm{E}((\theta - I(Y < X'\delta_n))Xk) - \mathrm{E}((\theta - I(Y < X'\tilde{\delta}))Xk) \\
&= \pi_c \mathrm{E}((\theta - I(Y < X'\delta_n))X|C = 1) - \pi_c \mathrm{E}((\theta - I(Y < X'\tilde{\delta}))X|C = 1) \\
&= \pi_c(F_1(\mathbf{1}'\tilde{\delta}) - F_1(\mathbf{1}'\delta_n))\mathbf{1} + \pi_c(F_0(\mathbf{e}'\tilde{\delta}) - F_0(\mathbf{e}'\delta_n))\mathbf{e},
\end{aligned}$$

where $\mathbf{1} = (1, 1)'$, $\mathbf{e} = (0, 1)'$ and $\pi_c = P(C = 1)$.

Under the regularity assumption that density functions in each compliance class are bounded away from zero and infinity on the support, $\|\Psi(\delta_n)\|_\infty \to \mathbf{0}$ implies $\|\delta_n - \tilde{\delta}\|_\infty \to \mathbf{0}$. Thus the desired identifiability condition holds.

It is easy to show that $\mathcal{F} = \{\psi_{\delta,\theta} : \delta \in \Delta^2, \theta \in [0, 1]\}$ is a Donsker class. Note that $\{1(Y \leq X'\delta) : \delta \in \Delta^2, \theta \in [0, 1]\}$ is Donsker, and so is $\{k\}$ (trivially). We now have that $\mathcal{F}$ is Donsker by using the permanence properties of the Donsker class. Since Donsker classes are also Glivenko-Cantelli, we have $\sup_{\delta \in \Delta^2, \theta \in [0,1]} \|\Psi_n(\delta) - \Psi(\delta)\| \to \mathbf{0}$.

We now consider weak convergence of the Z-estimator $\hat{\delta}$. Note that since $\mathcal{F} = \{\psi_{\delta,\theta} : \delta \in \Delta^2, \theta \in [0, 1]\}$ is a Donsker class, we have that $\sqrt{n}(\Psi_n - \Psi)(\delta) \Rightarrow G$ for some Gaussian G. At each point $\theta in[0, 1]$, $G \sim N(\mathbf{0}, \tilde{\Sigma}_\theta)$ for known π and $G \sim N(\mathbf{0}, \Sigma_\theta)$ for unknown $\hat{\pi}$. We also have that for any $\{\delta_n\} \in \Delta$ converging uniformly to δ,

$$\begin{aligned}
\sup_{\theta \in [0,1]} P(\psi_{\delta_n,\theta} - \psi_{\delta,\theta})^2 &= (E([1(Y < X'\delta) - 1(Y < X'\delta_n)]Xk)))^2 \\
&\leq E([1(Y < X'\delta) - 1(Y < X'\delta_n)]Xk))^2 \\
&\to \mathbf{0},
\end{aligned}$$

as k is bounded. Note that $(\cdot)^2$ in this part is element-wise for vectors.

We then show that Ψ is Frechet-differentiable at δ with derivative,

$$\dot{\Psi}_\delta(h)(t) = \mathrm{E}(f_{\epsilon|X,C=1}(0)XX'|C = 1)P(C = 1)h(t),$$

for $\epsilon = Y - X'\delta$ and $h(t) \in g[0,1]^2$, all left continuous functions and with right limit. By the regularity condition that f_1 and f_0 are bounded in each strata, for all $\{h_n\} \in g[0,1]^2$ that $\|h_n\|_\infty \to 0$ and $\delta + h_n \in \Delta^2$,

$$
\begin{aligned}
&\|\Psi(\delta+h_n) - \Psi(\delta) - \dot{\Psi}_\delta(h_n)\|_\infty \\
&= \|\mathrm{E}(1(Y < X'\delta)Xk) - \mathrm{E}(1(Y < X'(\delta+h_n))Xk) - \mathrm{E}(f_{Y-X'\delta|X,C=1}(0)XX'|C=1)\pi_c h_n\|_\infty \\
&= \pi_c\|(F_{1|C=1}(\mathbf{1}'\delta) - F_{1|C=1}(\mathbf{1}'\delta + \mathbf{1}'h_n))\mathbf{1} + (F_{0|C=0}(\mathbf{e}'\delta) - F_{0|C=0}(\mathbf{e}'\delta + \mathbf{e}'h_n))\mathbf{e} \\
&\quad - \mathrm{E}(f_{\epsilon|X,C=1}(0)XX'|C=1)h_n\|_\infty \\
&= \pi_c\|f_{1|C=1}(\mathbf{1}'\delta + \eta_{n1})\mathbf{1}(\mathbf{1}'h_n) + f_{0|C=1}(\mathbf{e}'\delta + \eta_{n2})\mathbf{e}(\mathbf{e}'h_n) - f_{1|C=1}(\mathbf{1}'\delta)\mathbf{1}\mathbf{1}'h_n \\
&\quad - f_{0|C=1}(\mathbf{e}'\delta)\mathbf{e}\mathbf{e}'h_n\|_\infty \\
&\to \mathbf{0}
\end{aligned}
$$

where η_{n1} is between 0 and $\mathbf{1}'h_n$, η_{n2} is between 0 and $\mathbf{e}'h_n$. The third equality comes from expectation conditioning on $X = (D,1)'$. We have that

$$\dot{\Psi}_\delta^{-1}(a)(t) = (\mathrm{E}(f_{Y-X'\delta|X,C=1}(0)XX'|C=1)\pi_c)^{-1}a(t)$$

which is a linear operator of $a(t)$. Thus we obtain the desired weak convergence that $\sqrt{n}(\hat{\delta}(\theta) - \delta(\theta)) \Rightarrow \dot{\Psi}_\delta^{-1}(G_\theta)$, which also is Gaussian.

2.8.2 Proof of Theorem 3

Under Assumption 2, $q_1(u)$ and $q_0(u)$ are continuously differentiable functions. The CDFs $F_1(t)$ and $F_0(t)$ can be expressed as,

$$F_1(t) = \int_0^1 I(q_1(u) \le t)du$$

and

$$F_0(t) = \int_0^1 I(q_0(u) \le t)du.$$

As in Corollary 1,

$$\sqrt{n}(\hat{q}_1(\theta) - q_1(\theta)) \Rightarrow \mathbb{G}_1(\theta) \equiv \mathbf{1}'\dot{\Psi}_{\beta_1}^{-1}(G_\theta)\mathbf{1}$$

and

$$\sqrt{n}(\hat{q}_0(\theta) - q_0(\theta)) \Rightarrow \mathbb{G}_0(\theta) \equiv \mathbf{e}'\dot{\Psi}^{-1}_{\beta_1}(G_\theta)\mathbf{e}$$

Following results on Hadamard derivatives and functional limit laws in Chernozhukov et al. (2010), we have

$$\sqrt{n}(\hat{F}_1(t) - F_1(t)) \Rightarrow -f_1(t)\mathbb{G}_1(F_1(t))$$

and

$$\sqrt{n}(\hat{F}_0(t) - F_0(t)) \Rightarrow -f_0(t)\mathbb{G}_0(F_0(t)),$$

where at point t, $-f_1(t)\mathbb{G}_1(F_1(t)) \sim N(0, V_1(t))$ and $-f_0(t)\mathbb{G}_0(F_0(t)) \sim N(0, V_0(t))$

2.8.3 Proof of Theorem 4

From Theorem 3, we have

$$\sqrt{n}(\hat{F}_1(t) - F_1(t)) = \frac{1}{n}\sum_{i=1}^{n}\phi_{1i}(t) + o_p(1) \Rightarrow -f_1(t)\mathbb{G}_1(F_1(t))$$

and

$$\sqrt{n}(\hat{F}_0(t) - F_0(t)) = \frac{1}{n}\sum_{i=1}^{n}\phi_{0i}(t) + o_p(1) \Rightarrow -f_0(t)\mathbb{G}_0(F_0(t))$$

Lemma 20.10 in Van der Vaart (2000) together with the delta method implies that,

$$W_n \equiv \sqrt{n}\left(\int \hat{F}_1 d\hat{F}_0 - \int F_1 dF_0\right) = -\int \sqrt{n}(\hat{F}_0 - F_0)dF_1 + \int \sqrt{n}(\hat{F}_1 - F_1)dF_0 + o_p(1)$$

The right part can be written as the sum of independent ramdom variables, and we can then apply the central limit theorem.

$$W_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(-\int \phi_{0i}(t)dF_1 + \int \phi_{1i}(t)dF_0\right) + o_p(1) \rightarrow_d N(0, V).$$

Chapter 3

MONOTONE DISTRIBUTION FUNCTION ESTIMATION IN RANDOMIZED TRIALS WITH NONCOMPLIANCE AND RANDOM CENSORING

3.1 Introduction

In randomized trials with time to event outcomes, two common problems are censoring and noncompliance. Censoring means follow-up for a subject ends before a time-to-event outcome is observed or at a pre-specified date. Noncompliance happens when some subjects fail to follow their assigned treatments, and the balance of randomization does not hold. Subjects who actually receive treatment are not comparable to subjects who actually receive control and in general, the intent-to-treat (ITT) effect is not consistent to the causal effect of actually receiving the treatment, especially when compliance patterns differ. Additional assumptions to identify the effect of treatment in a meaningful way are often required. As in Imbens and Angrist (1994) and Angrist et al. (1996), under certain plausible assumptions, referred to as the instrumental variable (IV) assumptions, the causal effect of the treatment can be nonparametrically identified among compliers in the framework of a counterfactual or potential outcome. Based on their work, Abadie (2000) and Abadie (2003) introduce a class of IV estimators to identify the functions of potential outcomes for compliers with covariates. Identification is shown by expressing the expectation of a function of potential outcomes for compliers as a weighted expectation of observed outcomes, where the weights can be positive or negative.

Most of the early literature are for continuous, uncensored outcomes, focusing on the estimation of compliers averaged causal effect (CACE) nonparametrically or based on parametric assumptions (Imbens and Rubin, 1997a,b; Cheng et al., 2009). For survival outcomes, few papers have discussed randomized studies in the presence of noncompliance (Robins and Tsiatis, 1991; Joffe, 2001; Loeys and Goetghebeur, 2003; Cuzick et al., 2007). These papers propose various semiparametric models to estimate the effect of treatment on survival functions for potential failure times, where the effect of treatment is modeled parametrically. Baker (1998) extended the models and assumptions for discrete-time survival data and derived closed-form expressions for the effect of treatment on the hazards at a specified time. Baker's estimator is analogous to the standard IV estimator for a survival outcome. Nie et al. (2011) developed a nonparametric estimator of the survival function at a specified time for compliers based on the empirical likelihood that make use of the mixture structure to gain efficiency over the standard IV method. Though this nonparametric estimator is bounded between 0 and 1, the method is not efficient to obtain whole survival functions and not necessary to generate a monotone survival function. In the case of binary treatment variable and binary instrument variable (i.e. treatment assignment), we could also consider estimating the compliers outcome distributions as identifying latent distributions from observed mixture data. In the mixture data framework, Qin et al. (2014) propose a binomial likelihood approach and a combined method of the EM algorithm and Pool-Adjacent-Violators Algorithm to obtain monotone distribution curves for latent distributions. However, the asymptotic distribution of their estimator remains unknown.

Our work builds on Peng and Huang (2008), who develop a quantile regression method for survival data subject to conditionally independent censoring. Their proposal uses the martingale feature associated with the censored data. Based on the martingale feature associated with the potential failures subject to independent censoring, we propose a novel estimator of quantile functions for potential outcomes and construct monotone distribution functions and survival functions, that are monotone and bounded between 0 and 1. The main idea is that we first nonparametrically estimate the quantile functions for compliers. Then for the possibly nonmonotone quantile function $\hat{\beta}(u)$, we construct $\hat{F}(t) = \int_0^1 I(\hat{\beta}(u) \leq y)du$, which is monotone in y and bounded in $[0, 1]$. Based on this representation, our estimator is a natural CDF. It is related to the rearrangement operation from variational analysis

(Hardy et al. (1952),Lorentz (1953), Villani (2003)). By using empirical process techniques, we establish asymptotic properties, including uniform consistency and weak convergence for the estimated quantile functions and survival functions. Moreover, we provide a self-induced smoothing algorithm to estimate the limiting variance of the estimator.

The paper is organized as follows. In Section 2 and 3, we introduce notations, assumptions and discuss the previously established IV estimators. In Section 4, we discuss the procedure to estimate quantile functions for the censoring potential outcomes for compliers, and construct a quantile-based monotone survival function estimator. We discuss asymptotic properties of the proposed estimators in Section 5 and prove theoretical results in the appendix. In Section 6, we proposed to use a self-induced smoothing approach for variance estimation. We described the algorithm to obtain variance estimation. In Section 7, we evaluated finite-sample performance of the proposed estimators under a series of data generating models. The proposed estimator is compared with the conventional IV estimator. Section 7 applies the proposed estimator to a randomized trial that aimed at examining the effects of periodic screening on breast cancer mortality.

3.2 Assumptions and Notations

3.2.1 Notation

We consider a two-arm randomized trial with n subjects. Let π denote the probability of randomization to treatment, Z be the indicator of treatment assignment and its sample analog $Z_i(i = 1, \ldots, n)$, where $Z_i = 1$ if subject i is assigned to treatment and $Z_i = 0$ if subject i is assigned to control. Let D^z the potential treatment receive under randomization assignment z, $D = ZD^1 + (1 - Z)D^0$ the actual receive of treatment and its sample analog $D_i(i = 1, \ldots, n)$, where $D_i = 1$ if subject i receives treatment and $D_i = 0$ if subject i receives control.

Let T^d the potential time-to-event outcome under treatment d and C^d the potential censoring time under treatment d. Define $Y^d = \min(T^d, C^d)$ and $\delta^d = I(T^d \leq C^d)$. Here

Table 3.1: The relationship between observed groups and latent compliance classes (CC).

Z	D		C	
1	1	1 (Compliers)	or	2 (Always-takers)
1	0			0 (Never-takers)
0	0	1 (Compliers)	or	0 (Never-takers)
0	1			2 (Always-takers)

we assume the exclusion restriction assumption in Angrist et al. (1996) that the potential outcome is affected by treatment assignment only via treatment received, i.e., $Y^{d^z} = Y^d$. Define $Y = \min(T, C)$, and $\delta = I(T \leq C)$. The observed data consist of n iid replicates of (Y, C, δ, Z, D), denoted by $(Y_i, C_i, \delta_i, Z_i, D_i), i = 1, \ldots, n$.

The population can be classified into four subpopulations based on their compliance status defined by the joint distribution of treatment assignment and potential treatment receipt: compliers if subjects always follow the assignment, never-takers if subjects always take the placebo, always-takers if subjects always take the treatment, defiers if subjects never follow the assignment. Let CC denote the compliance classes: $CC = 1$ (compliers) if $D^1 > D^0$; $CC = 2$ (always-takers) if $D^1 = D^0 = 1$; $CC = 0$ (never-takers) if $D^1 = D^0 = 0$; $CC = 3$ (defiers) if $D^1 < D^0$. In practice, we can only observe one of D^0 and D^1, the compliance class membership for each subject is therefore unknown. Let $\pi_c = Pr(CC = 1)$, the probability of being a complier in the population.

3.2.2 Assumptions

Assumption 3 *We impose standard IV assumptions similar to those in Angrist et al. (1996) and Abadie (2003):*

1. *Independence:* $(T^1, T^0, C^1, C^0, D^1, D^0)$ *is jointly independent of* Z.

2. *Nontrivial assignment:* $\pi = P(Z = 1) \in (0, 1)$

3. *First-stage:* $E(D^1 - D^0) \neq 0$. *The causal effect of* Z *on* D *is nonzero.*

4. *Monotonicity:* $P(D^1 \geq D^0)$*=1, so there are no defiers who always receive the opposite treatment of assignment.*

Under Assumption 3 (1)-(4), a subject's compliance class membership can be partially identified based on the treatment assigned and the treatment received as in Table 3.1. The observed CDF of each strata defined by Z and D is mixture of latent compliance classes. In this way, we can state our problem as identifying latent distributions from mixture data. Let (F_c^1, S_c^1, q_c^1) and (F_c^0, S_c^0, q_c^0) cumulative distribution functions, survival functions and quantile functions for compliers receiving treatment and control, (F_{nt}, S_{nt}, q_{nt}) for never-takers and (F_{at}, S_{at}, q_{at}) for always-takers. Additionally, we assume independent censoring that C_d is independent of T_d.

3.3 Abadie's weighted IV estimators

To estimate the survival functions for compliers, we define the following functions of D and Z:

$$
\begin{aligned}
k_0 &= (1-D)\frac{(1-Z)-(1-\pi)}{(1-\pi)\pi} \\
k_1 &= D\frac{Z-\pi}{(1-\pi)\pi} \\
k &= k_0(1-\pi) + k_1\pi = 1 - \frac{D(1-Z)}{1-\pi} - \frac{(1-D)Z}{\pi}.
\end{aligned}
$$

Under Assumption 3 (1)-(4),Abadie (2003) proved the following lemma.

Lemma 3.3.1 *(Abadie) Let* $g(T, C, D)$ *be any real function of* T, C, D *such that* $E|g(T, C, D)| < \infty$*,*

$$
\begin{aligned}
\mathrm{E}[g(T^0, C^0, D)|CC = 1] &= \frac{1}{\pi_c} E[k_0 g(T, C, D)] \\
\mathrm{E}[g(T^1, C^1, D)|CC = 1] &= \frac{1}{\pi_c} E[k_1 g(T, C, D)] \\
\mathrm{E}[g(T, C, D)|CC = 1] &= \frac{1}{\pi_c} E[k g(T, C, D)].
\end{aligned}
$$

This result allows us to identify functions of outcome distributions for compliers using a weighted sample. However, when $D \neq Z$, k takes negative value and is not a proper weight. Nonetheless, lemma (3.3.1) provides a simple way to estimate the CDFs of the potential outcomes for compliers. Let $g(T, C, D) = I(T \geq t)$, then survival functions for compliers S_c^1 and S_c^0 can be identified as,

$$S_c^0(t) = \frac{1}{\pi_c} E[k_0 I(T \geq t)]$$
$$S_c^1(t) = \frac{1}{\pi_c} E[k_1 I(T \geq t)]$$

Then, the corresponding sample $\hat{S}_c^{0,IV}(t)$ and $\hat{S}_c^{1,IV}(t)$, referred as the standard IV estimators, can be written as sum of weighted estimators of observed strata-specific survival function estimators, i.e., Kaplan-Meier estimators.

3.4 Quantiled-based Monotone Estimator

To estimate the quantile curves for compliers, we define the τ-th quantile of T for compliers under treatment as $Q_c^1(\tau) = \beta_1(\tau)$ and τ-th quantile of T for compliers under control as $Q_c^0(\tau) = \beta_0(\tau)$ and $\tau \in [0, 1]$. Let T_c^1 and T_c^0 be the potential time-to-event outcomes among compliers. Define $\Lambda_c^1(t) = -log(1 - P(T_c^1 \leq t))$, $\Lambda_c^0(t) = -log(1 - P(T_c^0 \leq t))$, $N^1(t) = I(Y^1 \leq t, \delta^1 = 1)$, $N^0(t) = I(Y^0 \leq t, \delta^0 = 1)$ and $N(t) = I(Y \leq t, \delta = 1)$. Define $H(u) = -log(1 - u)$ for $u \in [0, 1]$. As discussed in Peng and Huang (2008), the stochastic property of the martingale associated with the compliers gives the following results:

$$E[N^1(\beta_1(\tau)) - \Lambda_c^1(\beta_1(\tau) \wedge Y^1)|CC = 1] = E[k_1(N(\beta_1(\tau)) - \Lambda_c^1(\beta_1(\tau) \wedge Y))] = 0 \quad (3.1)$$

$$E[N^0(\beta_0(\tau)) - \Lambda_c^0(\beta_0(\tau) \wedge Y^0)|CC = 1] = E[k_0(N(\beta_0(\tau)) - \Lambda_c^0(\beta_0(\tau) \wedge Y))] = 0 \quad (3.2)$$

where Λ_c^1 and Λ_c^0 are the cumulative hazard functions of T^1 and T^0 among compliers. The first equation in Equation 3.1 and 3.2 is direct result from Lemma 1. The relationship between the unknown cumulative hazard functions and (β_1, β_0) is,

$$\Lambda_c^1(\beta_1(\tau) \wedge Y) = H(\tau) \wedge H(F_c^1(Y)) = \int_0^\tau I(Y \geq \beta_1(u)) dH(u),$$

$$\Lambda_c^0(\beta_0(\tau) \wedge Y) = H(\tau) \wedge H(F_c^0(Y)) = \int_0^\tau I(Y^0 \geq \beta_0(u))dH(u).$$

This motivates us to consider the following estimating equations,

$$S_n^1(\beta_1, \tau) = n^{-1}\sum_{i=1}^{n} k_{1i}(N_i(\beta_1(\tau)) - \int_0^\tau I(Y_i \geq \beta_1(\tau))dH(u)) = 0 \quad (3.3)$$

$$S_n^0(\beta_0, \tau) = n^{-1}\sum_{i=1}^{n} k_{0i}(N_i(\beta_0(\tau)) - \int_0^\tau I(Y_i \geq \beta_1(\tau))dH(u)) = 0 \quad (3.4)$$

The stochastic integration representation of $S_n^1(\beta_1, \tau)$ and $S_n^0(\beta_0, \tau)$ suggests a grid-based estimation procedure for $\beta_1(\cdot)$ and $\beta_0(\cdot)$. Specifically, we define $\hat{\beta}_1(\cdot)$ and $\hat{\beta}_0(\cdot)$ as a right-continuous piecewise-constant function that jumps only on a grid, $0 = \tau_0 < \tau_1 < \ldots < \tau_{L_n} = \tau_u < 1$, where τ_u is a constant subject to identification constraints due to censoring. *More precise condition and estimate of τ_u is described in the next section.* Define $\alpha_{1i}(\tau_j) = \sum_{k=0}^{j-1} I(Y_i \geq \beta_1(\tau_k))(H(\tau_{k+1}) - H(\tau_k))$ and $\alpha_{0i}(\tau_j) = \sum_{k=0}^{j-1} I(Y_i \geq \beta_0(\tau_k))(H(\tau_{k+1}) - H(\tau_k))$, for $i = 1, \ldots, n$ and $j = 1, \ldots, L_n$. Based on (3.3) and (3.4), we propose to estimate $\hat{\beta}_1(\tau_j)$ and $\hat{\beta}_0(\tau_j)$ sequentially by solving the following estimating equations for $\beta_1(\tau_j)$ and $\beta_0(\tau_j)$

$$n^{-1}\sum_{i=1}^{n} k_{1i}(N_i(\beta_0(\tau)) - \alpha_{1i}(\tau_j)) = 0 \quad (3.5)$$

$$n^{-1}\sum_{i=1}^{n} k_{0i}(N_i(\beta_0(\tau)) - \alpha_{0i}(\tau_j)) = 0 \quad (3.6)$$

Since Equation 3.5 and 3.6 are not continuous, exact roots may not exist, and the proposed $\hat{\beta}_1(\tau_j)$ and $\hat{\beta}_0(\tau_j)(j = 1, \ldots, L_n)$ are defined as generalized solutions. The monotonicity of (3.1) and (3.2) implies that their left sides are the gradients of convex functions (Fygenson and Ritov 1994). Setting $r_1(\tau) = Y^1 - \beta_1(\tau)$ and $r_0(\tau) = Y^1 - \beta_0(\tau)$, the convex functions are

$$E\left(r_1(\tau)(\alpha_1(\tau) - I(r_1(\tau) < 0)\delta^1)|CC = 1\right) = E\left(k_1 r_1(\tau)(\alpha_1(\tau) - I(r_1(\tau) < 0)\delta)\right) \quad (3.7)$$

$$E\left(r_0(\tau)(\alpha_0(\tau) - I(r_0(\tau) < 0)\delta^0)|CC = 1\right) = E\left(k_0 r_0(\tau)(\alpha_0(\tau) - I(r_0(\tau) < 0)\delta)\right) \quad (3.8)$$

Based on equations (3.7) and (3.8), estimating $\beta_1(\tau_j)$ and $\beta_0(\tau_j)$ $(j = 1, \ldots, L_n)$ are equivalent to locating the minimizer of the following objective functions:

$$Q_1(\tau) = n^{-1/2} \sum_{i=1}^{n} k_{1i} \left(r_{1i}(\tau)(\alpha_{1i}(\tau) - I(r_{1i}(\tau) < 0)\delta_i) \right) \tag{3.9}$$

$$Q_0(\tau) = n^{-1/2} \sum_{i=1}^{n} k_{0i} \left(r_{0i}(\tau)(\alpha_{0i}(\tau) - I(r_{0i}(\tau) < 0)\delta_i) \right), \tag{3.10}$$

where, $r_{1i}(\tau) = Y_i - \beta_1(\tau)$ and $r_{0i}(\tau) = Y_i - \beta_0(\tau)$. Further examination on the objective functions 3.9 and 3.10 shows that they are piece-wise linear functions in $\beta_1(\tau)$ and $\beta_0(\tau)$ for each τ. Since the weights k_{0i} and k_{1i} can be negative values, the objective functions are not necessarily convex. In practice, we can compute the greatest convex minorant (GCM) of the piece-wise linear function and then locate the minimum point as the first slope knot changing from negative to positive. The GCM is obtained by isotonic regression of the raw slopes. We can then obtain the CDF estimators (thus survival function estimators) from,

$$\hat{F}_c^1(t) = \int_0^{\tau_u} I(\hat{\beta}_1(u) \leq t) du$$
$$\hat{F}_c^0(t) = \int_0^{\tau_u} I(\hat{\beta}_0(u) \leq t) du.$$

The integral construction naturally assures a monomotone CDF estimator that is bounded. Though $\hat{q}_c^1$ and $\hat{q}_c^0$ are possibly non-monotone, $\hat{F}_1$ and $\hat{F}_0$ are nondecreasing functions and bounded in $[0, 1]$ by construction, similar for survival function estimators $\hat{S}_1$ and $\hat{S}_0$. In the next section, we derive asymptotic properties for the estimated quantile functions and CDFs.

3.5 Asymptotic results

The derivation of asymptotic results is faciliated by the stochastic integral representation associated with the proposed estimation procedure and Z-theorem. In this section, we establish the uniform consistency and weak convergence of the proposed estimators $(\hat{\beta}_c^1, \hat{F}_c^1)$ and $(\hat{\beta}_c^0, \hat{F}_c^0)$. We first state the regularity conditions. Define $\tilde{F}_c^1(t) = Pr(Y^1 \leq t, \delta^1 = 1 | CC = 1), \tilde{f}_c^1(t) = d\tilde{F}_c^1(t)/dt, \bar{F}_c^1(t) = Pr(Y^1 \geq t | CC = 1), \bar{f}_c^1(t) = d\bar{F}_c^0(t)/dt, \tilde{F}_c^0(t) = Pr(Y^0 \leq$

$t, \delta^1 = 1|CC = 1), \tilde{f}_c^0(t) = d\tilde{F}_c^0(t)/dt, \bar{F}_c^0(t) = Pr(Y^0 \geq t|CC = 1), \bar{f}_c^0(t) = d\bar{F}_c^0(t)/dt,$. The regularity conditions are as follows:

1. $\tilde{F}_c^1(\beta_1(\tau))$ and $\tilde{F}_c^0(\beta_1(\tau))$ are Lipschitz functions of τ. $\tilde{f}_c^1(t)$, $\bar{f}_c^1(t)$,$\tilde{f}_c^0(t)$ and $\bar{f}_c^0(t)$ are bounded above uniformly in t.

2. $\bar{f}_c^1(t)\tilde{f}_c^1(t)^{-1}$ and $\bar{f}_c^0(t)\tilde{f}_c^0(t)^{-1}$ are uniformly bounded for t.

3. $f_c^1(\beta_1(\tau))Pr(C^1 \geq \beta_1(\tau)|CC = 1) > 0$ and $f_c^0(\beta_0(\tau))Pr(C^0 \geq \beta_0(\tau)|CC = 1) > 0$ for $\tau \in [0, \tau_u]$.

Conditions (1) and (2) impose mild regularity assumptions on the distributions of outcomes for compliers. Condition (3) is the key assumption that ensures the identifiability of $\beta_1(\tau), \beta_0(\tau), \tau \in [0, \tau_u]$. If f_c^1 and f_c^0 are bounded below away from 0, then Condition (3) requires that τ_u is less than the CDF of T among compliers at the upper bound of C. Alternatively, Condition (3) can be relaxed such that for T^1 and T^0 among compliers, the $\tau_u's$ are defined separately with respect to C^1 and C^0 among compliers. For the ease of notion, we use τ_u for both T^1 and T^0. We have the following theorems.

Theorem 3.5.1 *Assuming the regularity conditions hold, if* $lim_{n\to\infty} S_L = 0$*, then*

$$sup_{\tau\in[0,\tau_u]}|\hat{\beta}_c^1 - \beta_c^1| \to_p 0$$

and

$$sup_{\tau\in[0,\tau_u]}|\hat{\beta}_c^0 - \beta_c^0| \to_p 0$$

Theorem 3.5.2 *Assuming the regularity conditions hold, if* $lim_{n\to\infty} S_L = 0$*, then*

$$sup_{t\in[0,\beta_1(\tau_u)]}|\hat{F}_c^1 - F_c^1| \to_p 0$$

and

$$sup_{t\in[0,\beta_0(\tau_u)]}|\hat{F}_c^0 - F_c^0| \to_p 0$$

Theorem 3.5.3 *Assuming the regularity conditions hold, if* $\lim_{n\to\infty} S_L = 0$*, then* $\sqrt{n}(\hat{\beta}_1(\tau) - \beta_1(\tau)$ *converges weakly to a Gaussian process for* $\tau \in [0, \tau_u]$ *and* $\sqrt{n}(\hat{\beta}_0(\tau) - \beta_0(\tau)$ *converges weakly to a Gaussian process for* $\tau \in [0, \tau_u]$

Theorem 3.5.4 *Assuming the regularity conditions hold, if* $\lim_{n\to\infty} S_L = 0$*, then* $\sqrt{n}(\hat{F}_c^1(t) - F_c^1(t)$ *converges weakly to a Gaussian process for* $t \in [0, \beta_1(\tau_u)]$ *and* $\sqrt{n}(\hat{F}_c^0(t) - F_c^0(t)$ *converges weakly to a Gaussian process for* $\tau \in [0, \beta_0(\tau_u)]$

Proofs for the above theorems are described in the appendix.

3.6 Self-induced Smoothing Approach for variance estimation

Inferences about $\hat{\beta}_1(\tau)$, $\hat{\beta}_0(\tau)$, $\hat{F}_c^1(t)$ and $\hat{F}_c^0(t)$ are important in applications. However, as seen from the proof of Theorem 3 and 4, the variances of the limiting processes involve the unknown density functions. In this section, we propose to use a self-induced smoothing approach to estimate the variances of the limiting processes for $\hat{\beta}_1(\tau)$ and $\hat{\beta}_0(\tau)$. We will discuss the variance estimation for $\hat{\beta}_1(\tau)$, and method and result for $\hat{\beta}_0(\tau)$ follows similarly.

Since $\hat{\beta}_1(\tau_j)$ for $(j = 1, \ldots, L_n)$ is asymptotically normal, its difference with the true parameter value, $\hat{\beta}_1(\tau_j) - \beta_1(\tau_j)$, should be approximately a Gaussian noise $Z(\tau_j)/\sqrt{n}$, where $Z(\tau_j) \sim N(0, \sigma_j^2)$. Since we obtain $\hat{\beta}_1(\tau_j)$ sequantially for $j = 1, \ldots, L_n$, we shall estimate the asymptotic variance by approximating the joint distribution of $(\hat{\beta}_1(\tau_1), \ldots, \hat{\beta}_1(\tau_j))$. Let $\beta_1(\bar{\tau}_j) = (\beta_1(\tau_1), \ldots, \beta_1(\tau_j))$, $\bar{\sigma}_j = (\sigma_1, \ldots, \sigma_j)$ and $Z(\bar{\tau}_j) = (Z(\tau_1), \ldots, Z(\tau_j))$. The covariance matrix for the limiting distribution of $\sqrt{n}(\hat{\beta}_1(\bar{\tau}_j - \beta_1(\bar{\tau}_j))$ is Σ_j, of which $\bar{\sigma}_j^2$ are diagonal terms. Assume $Z(\bar{\tau}_j)$ is independent of data and let E_z conditional expectation given data with respect to $Z(\bar{\tau}_j)$. A self-induced smoothing objective function is defined as $\tilde{Q}_1(\beta_1(\tau_j); \beta_1(\bar{\tau}_{j-1}))) = \mathrm{E}_z Q_1(\beta_1(\tau_j) + Z(\tau_j)/\sqrt{n}; \beta_1(\bar{\tau}_{j-1}) + Z(\bar{\tau}_{j-1})/\sqrt{n})$. The self-smoothing approach using the limiting Gaussian distrbution was originally proposed by Brown and Wang (2005) for certain non-smoothing estimating functions; Jin (2014) developed a Monte Carlo version when there is no explicit form for the smoothing estimating function. The asymptotic covariance is estimated by a sandwich-type estimator of form $A^{-1}DA'^{-1}$ with A

being the Jacobian matrix of the smoothed score functions and D the covariance matrix of the smoothed score functions .

Let Φ be the standard normal distribution function, ϕ the standard normal density function, $\dot{\phi}$ the first derivative of ϕ and $\ddot{\phi}$ the first derivative of $\dot{\phi}$. It is easy to obtain an explicit form for $\tilde{Q}_1$,

$$\begin{aligned}&\tilde{Q}_1(\beta_1(\tau_j);\beta_1(\bar{\tau}_{j-1}))\\&\quad=\frac{1}{n}\sum_{i=1}^{n}k_{1i}\left[(Y_i-\beta_1(\tau_j))\sum_{k=0}^{j-1}\Phi(\sqrt{n}(Y_i-\beta_1(\tau_k))/\sigma_k)\Delta H(\tau_{k+1})\right.\\&\qquad\left.-(1-\Phi(\sqrt{n}(Y_i-\beta_1(\tau_j))/\sigma_j))\delta_i(Y_i-\beta_1(\tau_j))+\sigma_j/\sqrt{n}\phi(\sqrt{n}(Y_i-\beta_1(\tau_j))/\sigma_j)\delta_i+c_i\right],\end{aligned} \tag{3.11}$$

where the constant term c_i does not involve $\beta_1(\tau_j)$. Let $\tilde{\beta}_1(\tau_j)$ minimizing $\tilde{Q}_1(\beta_1(\tau_j);\beta_1(\bar{\tau}_{j-1}))$ denote the corresponding estimator for $\beta_1(\tau_j)$. We then differentiate the smoothed objective function $\tilde{Q}_1$ to get score function

$$\begin{aligned}\tilde{S}_1(\beta_1(\tau_j);\beta_1(\bar{\tau}_{j-1}))&=\frac{1}{n}\sum_{i=1}^{n}k_{1i}\left[-\sum_{k=0}^{j-1}\Phi(\sqrt{n}(Y_i-\beta_1(\tau_k))/\sigma_k)\Delta H(\tau_{k+1})\right.\\&\qquad+(1-\Phi(\sqrt{n}(Y_i-\beta_1(\tau_j))/\sigma_j))\delta_i\\&\qquad\left.-\phi(\sqrt{n}(Y_i-\beta_1(\tau_j))/\sigma_j)\delta_i\sqrt{n}(Y_i-\beta_1(\tau_j))/\sigma_j-\dot{\phi}(\sqrt{n}(Y_i-\beta_1(\tau_j))/\sigma_j)\delta_i\right]\\&\equiv\frac{1}{n}\sum_{i=1}^{n}\tilde{\psi}_{ij}\end{aligned} \tag{3.12}$$

Let vectors $\tilde{\psi}_{\bar{i}}=(\tilde{\psi}_{i1},\ldots,\tilde{\psi}_{ij})'$ and $\tilde{S}_1(\beta_1(\bar{\tau}_j))=\left(\tilde{S}_1(\beta_1(\tau_k);\beta_1(\bar{\tau}_{k-1})),k=1,\ldots,j\right)'$. The asymptotic covariance matrix of $\sqrt{n}\tilde{S}_1(\beta_1(\bar{\tau}_j))$ is approximated by

$$\hat{D}(\beta_1(\bar{\tau}_j),\Sigma_j)=\frac{1}{n}\sum_{i=1}^{n}\tilde{\psi}_{\bar{i}}^{\otimes 2},$$

where for a vector v, $v^{\otimes 2}=vv'$.

$$\hat{A}(\beta_1(\bar{\tau}_j),\Sigma_j)=\frac{\partial\tilde{S}_1(\beta_1(\bar{\tau}_j))}{\partial\beta_1(\bar{\tau}_j)}.$$

Let

$$\hat{V}(\beta_1(\bar{\tau}_j), \Sigma_j) = \hat{A}(\beta_1(\bar{\tau}_j), \Sigma_j)^{-1} \times \hat{D}(\beta_1(\bar{\tau}_j), \Sigma_j) \times \hat{A}(\beta_1(\bar{\tau}_j), \Sigma_j)'^{-1}.$$

By simple algebra manipulation, we should notice that $\hat{V}(\beta_1(\bar{\tau}_j), \Sigma_j)$ only involves Σ_j through $\bar{\sigma}_j$; and $\hat{A}(\beta_1(\bar{\tau}_j), \Sigma_j)$ has an embedded structure by $\hat{A}(\beta_1(\bar{\tau}_{j-1}), \Sigma_{j-1})$ and $(\beta_1(\tau_j), \sigma_j)$; and $\hat{D}(\beta_1(\bar{\tau}_j), \Sigma_j)$ has an embedded structure by $\hat{D}(\beta_1(\bar{\tau}_{j-1}), \Sigma_{j-1})$ and and $(\beta_1(\tau_j), \sigma_j)$. These embedded structure facilities our calcualtion for $\hat{V}(\beta_1(\bar{\tau}_j), \Sigma_j)$. If $\beta_1(\bar{\tau}_j)$ is the true parameter, then $\hat{V}(\beta_1(\bar{\tau}_j), \Sigma_j)$ converges to the limiting covariance matrix Σ_j. Therefore, the sandwich-type estimator implies an iterative algorithm of form $\Sigma_j^k = \hat{V}(\hat{\beta}_1(\bar{\tau}_j), \Sigma_j^{k-1})$. Specifically, we propose the following algorithm to estimate σ_j for $j = 1, \ldots, L_n$:

1. Compute the estimator $\hat{\beta}_1(\bar{\tau}_j)$ and set Σ_j^0 to be the identity matrix.

2. Update variance-covariance matrix

$$\Sigma_j^k = \hat{A}(\hat{\beta}_1(\bar{\tau}_j), \Sigma_j^{k-1})^{-1} \times \hat{D}(\hat{\beta}_1(\bar{\tau}_j), \Sigma_j^{k-1}) \times \hat{A}(\hat{\beta}_1(\bar{\tau}_j), \Sigma_j^{k-1})'^{-1}.$$

   Obtain the smoothed objective function $\tilde{Q}_1(\hat{\beta}_1(\tau_j); \hat{\beta}_1(\bar{\tau}_{j-1}))$ using the updated Σ_j^k. Minimize the smoothed objective function to get an estimator $\hat{\beta}_1(\tau_j)^k$.

3. Repeat Step 2 until $\hat{\beta}_1(\tau_j)^k$ converge.

We then use the converged $\hat{\Sigma}_{L_n}$ to approximate limiting variances of $\hat{F}_1^c(t)$ by a Monte Carlo approach. Specifically, we first generate a large number (B^*) replications $\bar{z}_i = (z(\bar{\tau}_{L_n}))_i$ $(i = 1, \ldots, B^*)$ from the multivariate norml distribution $N(0, \hat{\Sigma}_{L_n})$. Next, we obtain $\hat{F}_{1i}^c = \int_0^{\tau_u} I(\hat{\beta}_1(u) + z_i(u)/\sqrt{n} \leq t) du$ $(i = 1, \ldots, B^*)$, where $z_i(u)$ is a right-continuous piecewise-constant function defined with respect to $\bar{z}_i$. Then, the variance of $\hat{F}_1^c(t)$ is estimated as sampling variance of $\hat{F}_{1i}^c$ $(i = 1, \ldots, B^*)$.

3.7 Numerical Studies

3.7.1 Simulation results

In this section we evaluate the finite-sample performance of the proposed methods and compare with the standard IV estimator. The event times T are generated from lognormal distribution. Specifically, the means for each latent class are $E(logT^1|CC = 2) = 1$, $E(logT_1^c) = 2$, $E(logT_0^c) = 1.5$, $E(logT^0|CC = 0) = 1$ and $\{Var(logT^d|CC)\}_{d=0,1}$ are 1. We generate the G from $Bernoulli(p)$; censoring distribution C is $unif(0, 25)$ if $G = 1$ and $C = 25$ otherwise. We set p=0.8, 0 to yield 25% and 5% censoring on T. The strength of an IV is how strong the association between the IV and the actual treatment is. One factor for the strength of IV is the proportion of compliers in the study population π_c, with a higher π_c indicating a stronger IV. In the simulation studies, we consider the strong IV cases $\pi_c = 0.5$ and weak IV cases $\pi_c = 0.1$. Accordingly, the compliance class status C is randomly generated as (a) strong IV: $P(C = 0) = 0.25$, $P(C = 1) = 0.5$ and $P(C = 2) = 0.25$; (b) weak IV: $P(C = 0) = 0.45$, $P(C = 1) = 0.1$ and $P(C = 2) = 0.45$. As a intermediate case, we also evaluate the performance of the two estimators under a medium strength IV that $\pi_c = 0.2$ and $\pi_{at} = \pi_{nt} = 0.4$. Under each scenario, we generate 1000 simulated data sets of sample sizes $n = 200$ and $n = 400$. We set replication number $B^* = 500$ for the Monte Carlo scheme. We use an equally spaced grid with $S_L = 0.05$. The R function *gcmlcm* from pacakge *fdrtool* is used to computed the GCM of the objective functions.

Table 3.2 reports the empirical biases (Bias), empirical standard deviation (EmpSD), and average standard deviation estimates (EstSD) of the proposed quantile estimator $\hat{\beta}_1(\tau)$ and $\hat{\beta}_0(\tau)$ at $\tau = 0.1, 0.3, 0.5, 0.7$. The table also reports the coverage rate of the 95% confidence intervals (Cov95) constructed based on the normal approximation with EstSD. When the compliance rate is about 50 %, estimation performance improves as sample size n increases. The proposed standard error estimators and 95% confidence intervals slightly underestimate when the sample size is small but perform better as the sample size increases. Results for experiment with low compliance rate ($\pi_c = 0.2$) are not listed for small sample size $n = 200$ because of extremely high bias and standard error. The performance of the proposed estimator gets worse as τ increases, because of the sequential estimation process and high censoring rate.

Table 3.2: 25% Censoring: Estimation of quantile function for compliers on treatment group using the proposed methods, $\beta_1(\tau)$ for $\tau = 0.1, 0.3, 0.5, 0.7$. Bias, sampling standard error (EmpSD), averaged estimated standard error (EstSD) and coverage rate (Cov95).($\times 10^2$)

τ	Bias	EmpSD	EstSD	Cov95
		$\pi_c = 0.5, n = 200$		
0.1	14.8	90.8	88.1	85.6
0.3	13.8	123.1	118.5	88.7
0.5	23.3	189.4	166.12	88.1
0.7	19.7	303.6	222.3	81.8
		$\pi_c = 0.5, n = 400$		
0.1	6.7	66.5	66.1	85.2
0.3	4.7	85.4	84.4	92.5
0.5	5.3	123.6	121.5	91.1
0.7	41.5	233.5	201.2	89.0
		$\pi_c = 0.2, n = 400$		
0.1	50.8	158.8	255.3	83.8
0.3	33.3	209.9	225.9	86.7
0.5	55.3	304.5	279.4	87.6
0.7	-33.0	374.8	262.3	76.4

Table 3.3 reports Bias, EmpSD and EstSD of the proposed quantile estimator $\hat{\beta}_1(\tau)$ and $\hat{\beta}_0(\tau)$ at $\tau = 0.1, 0.3, 0.5, 0.7$ under 5% censoring rate. The table also reports the coverage rate of the 95% confidence intervals (Cov95) constructed based on the normal approximation with EstSD. In this setting, we observe similar pattern as in Table 3.2 where censoring rate is

about 25%. For cases with compliance rate $\pi_c = 0.5$, both bias and precison of the proposed estimator get better comparing to high censoring cases in Table 3.2. Results for experiment with low compliance rate ($\pi_c = 0.2$) are not listed for small sample size $n = 200$ because of extremely high bias and standard error.

Table 3.3: 5% Censoring: Estimation of quantile function for compliers on treatment group using the proposed methods, $\beta_1(\tau)$ for $\tau = 0.1, 0.3, 0.5, 0.7$. Bias, sampling standard error (EmpSD), averaged estimated standard error (EstSD) and coverage rate (Cov95).($\times 10^2$)

τ	Bias	EmpSD	EstSD	Cov95
		$\pi_c = 0.5, n = 200$		
0.1	9.5	87.1	87.1	85.6
0.3	0.6	112.5	109.2	90.8
0.5	-0.01	152.8	150.7	91.4
0.7	33.3	261.2	236.6	88.9
		$\pi_c = 0.5, n = 400$		
0.1	4.9	61.0	60.2	89.2
0.3	2.7	78.7	81.4	92.7
0.5	-0.6	110.7	112.9	92.9
0.7	17.1	182.3	176.9	93.4
		$\pi_c = 0.2, n = 400$		
0.1	36.7	137.5	255.3	84.8
0.3	12.9	182.5	209.7	87.7
0.5	3.8	233.6	245.2	87.4
0.7	33.1	374.8	350.7	88.1

Table 3.4 reports the empirical biases (Bias), empirical standard deviation (EmpSD), and average standard deviation estimates (EstSD) of the proposed estimators $\hat{S}_1^c(t)$ at $t = 0.3, 0.5, 0.7, 0.9$. As a comparison, the table also presents the empirical biases (Bias), em-

pirical standard deviation (EmpSD) of the standard IV estimators $\hat{S}^c_{1,IV}(t)$. When the compliance rate is about 50%, the proposed estimator performs similarly to the standard IV method. When the compliance rate π_c gets lower to 0.2 or 0.1, the proposed estimator is more robust than the standard IV estimator. When the compliance rate is too low, the standard IV method cannot yield reasonable point estimate. The proposed estimator tends to have higher bias and larger standard error when the compliance rate is lower. Results for $\hat{S}^c_0(t)$ and $\hat{S}^c_{0,IV}(t)$ have similar pattern as here.

3.8 Application to HIP Study

The Health Insurance Plan of Greater New York (HIP) study was a randomized trial that began in 1963 and was aimed at examining the effects of periodic screening for the early detection and treatment of breast cancer. More than 60,000 women were randomized into two groups at the beginning of the study. Women in the treatment group received an initial screening examination and three annual follow-up visits. Women in the control group received usual care. Approximately one-third of the women in the treatment group refused to screen.

A lot of literature studied this data set. Joffe (2001) used G-estimation of the accelerated failure time (AFT) model and concluded that screening increased the mean time to death from breast cancer by 22% (95% CI: 5.1%, 63.2%). This AFT model depends on a parametric model. Baker (1998) studied how screening effected medical cost and concluded that screening saved $16,000 (95% CI: $10,000, $51,000) per life year (cost-effectiveness). Baker's estimator is analogous to the standard IV estimator for a survival outcome. Nie et al. (2011) developed a nonparametric estimator based on the empirical likelihood that makes use of the mixture structure to gain efficiency over the standard IV method. They concluded that compliers who received screening have 12.3% (95% CI: 4.08%, 20.6%) higher probability to survive over 10 years. Similar to Joffe (2001) and Nie et al. (2011), we consider the first 10 years of each woman follow up, to reduce the attenuation on the screening effect by later periods when both groups received the same treatment. We also consider a limited mortality analysis by only including women whose breast cancer was diagnosed within the first 7 years of study, as in Baker (1998) and Nie et al. (2011).

We include 878 subjects in the analysis. The sample size in each compliance class is listed in Table 3.5. This was a single consent design as subjects in the control group had no access to screening In the data set, the compliance rate is about 72%. 39 subjects were lost to follow-up during the first 10-year follow-up and 542 subjects were still alive after the first 10-year follow-up. In Figure 3.8, the blue solid line is the estimated distribution function for compliers who received control, the red solid line is the estimated distribution function for compliers who received treatment and the dash lines are the corresponding IV estimator. For compliers who received control (usual care group), the standard IV estimator is not monotone but the proposed estimator smoothes out these wiggling parts, bounded between 0 and 1. For compliers who received treatment, the standard IV estimator performs similarly to the proposed estimator. We expect that our estimator performs more robust than the standard IV estimator, especially for compliers who received control. We estimate that compliers who receive control have 0.42 (95% CI: 0.35, 0.49) all-cause mortality and compliers who receive treatment have 0.28 (95% CI: 0.23, 0.33).

3.9 Conclusion

Our work builds on Peng and Huang (2008), who develop a quantile regression method for survival data subject to conditionally independent censoring. We propose a method to estimate quantile functions and survival functions for potential outcomes under independent censoring and noncompliance. Based on the martingale feature associated with the censoring data, we estimate quantile functions for compliers. Then for the possibly non-monotone quantile function $\hat{\beta}(u)$, we construct $\hat{F}(t) = \int_0^{\tau_u} I(\hat{\beta}(u) \leq t)du$, which is monotone in y and bounded in $[0, 1]$. By using empirical process techniques, we establish asymptotic properties, including uniform consistency and weak convergence for the proposed estimators. Inferences about $\hat{\beta}_1(\tau)$, $\hat{\beta}_0(\tau)$, $\hat{F}_c^1(t)$ and $\hat{F}_c^0(t)$ are important in applications. However, the fact that the limiting processes involve unknown density functions complicates variance estimation.

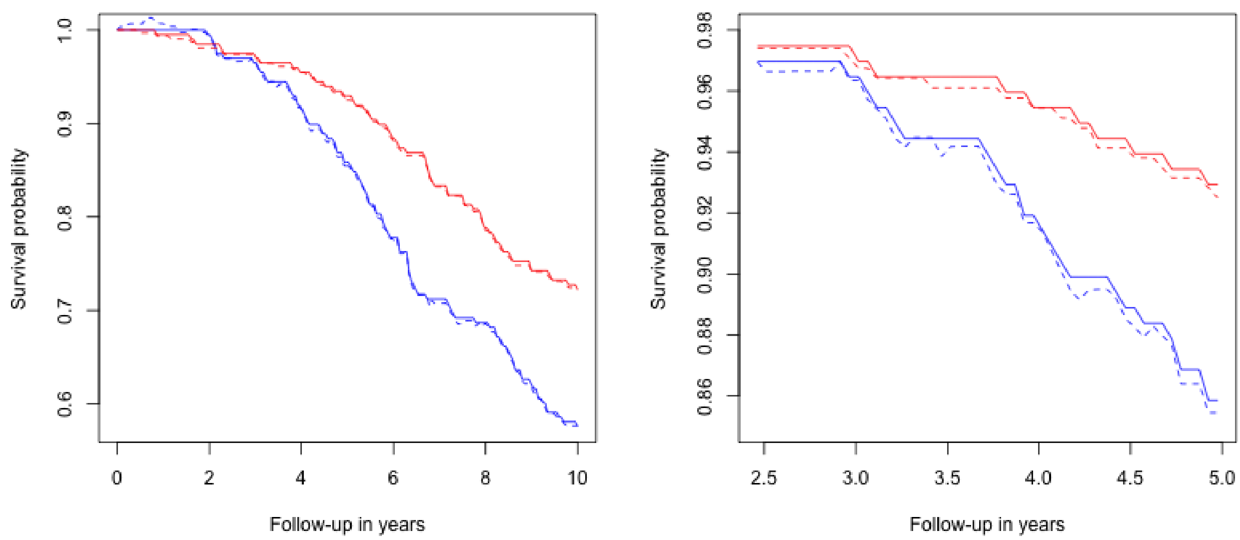


Figure 3.1: Left Panel: estimated mortality lines during the first 10-year follow-up period. Right Panel: a zoom-in area of the left panel between 2.5 to 5 follow-up years. Blue solid line: the proposed estimator for compliers who received control, blue dash line: the standard IV estimator for compliers who received control, red solid line: the proposed estimator for compliers who received treatment, red dash line: the standard IV estimator for compliers who received treatment.

To tame this, we apply a self-induced smoothing Monte-Carlo approach to estimate the variances of the limiting processes for quantiles and distribution functions. As an alternative, a multiplier bootstrapping approach is proposed to estimate variances and construct confidence intervals. In simulation studies, the proposed estimator performs more stable than the standard IV estimator, especially in the weak instrument setting. We apply our proposal to analyze the HIP study and examine the effect of periodic screening on breast cancer mortality.

3.10 Proofs

3.10.1 Proof of Theorem 1 and 2

In this section, we give proof of Theorem 1 and 2. We give the proof for $\hat{\beta}_1$ and $\hat{F}_c^1$,then proof for $\hat{\beta}_0$ and $\hat{F}_c^0$ follows similarly.

According to the estimation procedure proposed in Section 3.4,

$$S_n^1(\hat{\beta}_1, \tau_j) = n^{-1} \sum_{i=1}^{n} k_{1i}(N_i(\hat{\beta}_1(\tau_j)) - \int_0^{\tau_j} I(Y_i \geq \hat{\beta}_1(u))dH(u)) = O_p(1/n),$$

for $j = 1, \ldots, L_n$. Uniform consistency and asymptotic normality of $\hat{\beta}_1$ can be demonstrated via empirical process arguments for Z-estimators. Let Θ be the space of all quantile functions β with $\beta(0) = 0$ and restricted to $[0, \tau_u]$. We will use the uniform norm $\|\cdot\|_\infty$ on Θ. Define $\Psi_n(\beta)(\tau) = S_n^1(\beta_1, \tau)$ with the form $\Psi_n(\beta)(\tau) = \mathbb{P}_n\psi_{\beta,\tau}$, where

$$\psi_{\beta,\tau}(X_i) = k_{1i}(N_i(\beta(\tau)) - \int_0^{\tau} I(Y_i \geq \beta_(u))dH(u))$$

and $X_i = (Y_i, \delta_i, D_i, Z_i)$ for $i = 1, \ldots, n$.

Define $\Psi(\beta)(\tau) = P\psi_{\beta,\tau} = \pi_c(\tilde{F}_c^1(\beta(\tau)) - \int_0^\tau \bar{F}_c^1(\beta(u))dH(u))$. We assume that $\tau_1 < \ldots < \tau_{L_n-1}$ are equally spaced between 0 and $\tau_u = \tau_{L_n}$ with grid size S_L. Because $S_L \to 0$, we have that $\|\Psi_n(\hat{\beta}_1)\|_\infty \to_p 0$. This is true because by the definition of $\Psi_n(\hat{\beta}_1)(\tau)$ and $\Psi_n(\hat{\beta}_1)(\tau_j)$,

$$\begin{aligned}
&\sup_{\tau\in[\tau_j,\tau_{j+1}]}|\Psi_n(\hat{\beta}_1)(\tau) - \Psi_n(\hat{\beta}_1)(\tau_j)| \\
&\leq (H(\tau_{j+1}) - H(\tau_j)) \\
&\leq S_L/(1 - \tau_u)
\end{aligned}$$

$\hat{\beta}_1$ is then a Z-estimator. To show the uniform consistency of $\hat{\beta}_1$, we should establish an identifiability condition for Ψ and uniform consistency of Ψ_n. The identifiability condition holds if for any sequence $\{\beta_n\} \in \Theta$, $\|\Psi(\beta_n)\|_\infty \to 0$ implies $\|\beta_n - \beta_1\|_\infty \to 0$. $\Psi(\beta_n)(\tau) = \Psi(\beta_n)(\tau) - \Psi(\beta_1)(\tau) = \pi_c \tilde{f}_c^1(\tilde{\beta}_1(\tau))(\beta_n(\tau) - \beta_1(\tau)) - \int_0^\tau \pi_c \bar{f}_c^1(\check{\beta}_1(u))(\beta_n(u) - \beta_1(u))dH(u)$, where $\check{\beta}_1(\cdot)$ and $\tilde{\beta}_1(\cdot)$ are between $\beta_n(\cdot)$ and $\beta_1(\cdot)$. Let $\epsilon_n(\tau) = \beta_n(\tau) - \beta_1(\tau)$.By solving this integral equation, we obtain

$$\begin{aligned}\epsilon_n(\tau) &= \pi_c \tilde{f}_c^1(\tilde{\beta}_1(\tau))^{-1}(\Psi(\beta_n)(\tau) \\ &\quad + \int_0^\tau \Psi(\beta_n)(u)\bar{f}_c^1(\check{\beta}_1(u))\tilde{f}_c^1(\tilde{\beta}_1(u))^{-1} \prod_u^\tau (1 + \bar{f}_c^1(\check{\beta}_1(s))\tilde{f}_c^1(\tilde{\beta}_1(s))^{-1}dH(s))dH(u))\end{aligned}$$

where $\prod$ is the type II product integral. Then $\Psi(\beta_n)(\tau) \to 0$ uniformly over $\tau \in [0, \tau_u]$ implies $\epsilon_n(\tau) \to 0$ uniformly. Thus $\|\beta_n - \beta_1\|_\infty \to 0$ implies the desired identifiability condition.

It is easy to show that $\mathcal{F} = \{\psi_{\beta,\tau} : \beta \in \Theta, \tau \in [0, \tau_u]\}$ is a Donsker class. Note that $\{N(\beta(\tau)) = I(Y \leq \beta(\tau), \delta = 1) : \beta \in \Theta, \tau \in [0, \tau_u]\}$ is Donsker, and so is $\{k_1\}$ (trivially) and $\int_0^\tau I(Y_i \geq \beta_(u))dH(u)$ is lipschitz in τ. We now have that $\mathcal{F}$ is Donsker by using the permanence properties of the Donsker class. Since Donsker classes are also Glivenko-Cantelli, we have $\sup_{\beta\in\Theta}\|\Psi_n(\beta) - \Psi(\beta)\|_\infty \to_p 0$.

Next, we can show that the rearranged estimator $\hat{F}_c^1(t)$ constructed by $\hat{\beta}_1(\tau)$ are uniformly consistent for $\tau \in [0, \tau_u]$. Since $\sup_{\tau\in[0,\tau_u]}|\hat{\beta}_1(u) - \beta_1(u)| \to_p 0$, for sufficiently large n, we have that $\sup_{\tau\in[0,\tau_u]}|\hat{\beta}_1(u) - \beta_1(u)| \leq \epsilon$ with $\epsilon \leq \beta_1(\tau_{j+1}) - \beta_1(\tau_j)$ for $j = 1, \ldots, L_n - 1$. Let $t = \beta_1(u^*)$ and $u^* \in [\tau_k, \tau_{k+1}), k = 0, \ldots, L_n - 1$. Note that $\hat{\beta}_1(u) < t$ for $u \in [0, \tau_k)$, and $\hat{\beta}_1(u) > t$ for $u > \tau_{k+1}$. Then, we can show that $|\hat{F}_c^1(t) - F_1^c(t)|$ is bounded by $2a_n + a_n^2$

$(a_n = S_L)$ which does not depend on t.

$$\begin{aligned}
|\hat{F}_c^1(t) - F_1^c(t)| &= |\int_0^{\tau_u} I(\hat{\beta}_1(u) \le t)du - \int_0^{\tau_u} I(\beta_1(u) \le t)du| \\
&\le |\sum_{j=0}^{k-1} \Big[I(\hat{\beta}_1(\tau_j) \le \beta_1(u^*))a_n - a_n \Big] | + | I(\hat{\beta}_1(\tau_k) \le \beta_1(u^*))a_n - (u^* - \tau_k)a_n | \\
&+ |I(\hat{\beta}_1(\tau_{k+1}) \le \beta_1(u^*))a_n - 0| + |\sum_{j=k+2}^{L_n-1} \left[\int_{\tau_j}^{\tau_{j+1}} I(\hat{\beta}_1(u) \le \beta_1(u^*))du - 0 \right] | \\
&\le 0 + (a_n + a_n^2) + a_n + 0
\end{aligned}$$

Thus, $\sup_{t\in[0,\beta_1(\tau_u)]} |\hat{F}_c^1(t) - F_1^c(t)| \to_p 0$.

3.10.2 Proof of Theorem 3 and 4

We now consider weak convergence of the Z-estimator $\hat{\beta}_1$. First note that since $\sqrt{n}S_L \to 0$, we have $\|\sqrt{n}\Psi_n(\hat{\beta}_1)\|_\infty \to_p 0$. This is true because $\sqrt{n}\Psi_n(\hat{\beta}_1)(\tau_j) = o_p(1)$ and

$$\begin{aligned}
&\sup_{\tau\in[\tau_j,\tau_{j+1}]} \sqrt{n} |\Psi_n(\hat{\beta}_1)(\tau) - \Psi_n(\hat{\beta}_1)(\tau_j)| \\
&\le \sqrt{n}(H(\tau_{j+1}) - H(\tau_j)) \\
&\le \sqrt{n}S_L/(1 - \tau_u).
\end{aligned}$$

Note that since $\mathcal{F} = \{\psi_{\beta,\tau} : \beta \in \Theta, \tau \in [0, \tau_u]\}$ is a Donsker class, we easily have that $\sqrt{n}(\Psi_n - \Psi)(\beta_1) \to G$ for some tight random G. We also have that for any $\{\beta_n\} \in \Theta$ converging uniformly to β_1,

$$\begin{aligned}
\sup_{\tau\in[0,\tau_u]} P(\psi(\beta_n)(\tau) - \psi(\beta_1)(\tau))^2 &\le \sup_{\tau\in[0,\tau_u]} 2\mathrm{E}k_1^2(I(Y \le \beta_n(\tau)) - I(Y \le \beta_1(\tau)))^2\delta \\
&+ 2\mathrm{E}k_1^2 \left(\int_0^\tau I(Y \ge \beta_n(u)) - I(Y \ge \beta_1(u))dH(u) \right)^2 \\
&\to 0,
\end{aligned}$$

as k_1 and δ are bounded (trivial).

We then show that Ψ is Frechet-differentiable at β_1 with derivative

$$\dot{\Psi}_{\beta_1}(h)(t) = \pi_c \tilde{f}_c^1(\beta_1(t))h(t) - \int_0^t h(u)\pi_c \bar{f}_c^1(\beta_1(u))dH(u)$$

for $h(t) \in g[0,\tau_u]$, all left continuous functions and with right limit, $h(0) = 0$. By the regularity condition that $\dot{\tilde{f}}_1^c$ and $\dot{\bar{f}}_1^c$ are bounded uniformly by $c < \infty$, for all $\{h_n\} \in g[0,\tau_u]$ that $\|h_n\|_\infty \to 0$ and $\beta_1 + h_n \in \Theta$,

$$
\begin{aligned}
&\|\Psi(\beta_1 + h_n)(\tau) - \Psi(\beta_1)(\tau) - \dot{\Psi}_{\beta_1}(h_n)(\tau)\|_\infty \\
&= \pi_c\|\tilde{F}_c^1(\beta_1(\tau) + h_n(\tau)) - \tilde{F}_c^1(\beta_1(\tau)) - \tilde{f}_1^c(\beta_1(\tau))h_n(\tau) \\
&- \int_0^\tau \left(\bar{F}_c^1(\beta_1(u) + h_n(u)) - \bar{F}_c^1(\beta_1(u)) - h_n(u)\bar{f}_c^1(\beta_1(u))\right) dH(u)\|_\infty \\
&\le \pi_c\|\tilde{F}_c^1(\beta_1(\tau) + h_n(\tau)) - \tilde{F}_c^1(\beta_1(\tau)) - \tilde{f}_1^c(\beta_1(\tau))h_n(\tau)\|_\infty \\
&+ \pi_c\|\int_0^\tau \left(\bar{F}_c^1(\beta_1(u) + h_n(u)) - \bar{F}_c^1(\beta_1(u)) - h_n(u)\bar{f}_c^1(\beta_1(u))\right) dH(u)\|_\infty \\
&= \pi_c\|1/2\dot{\tilde{f}}_1^c(\tilde{\beta}_1(\tau))h_n(\tau)^2\|_\infty + \pi_c\|\int_0^\tau 1/2\dot{\bar{f}}_1^c(\check{\beta}_1(u))h_n(u)^2 dH(u)\|_\infty \\
&\le c\|h_n\|_\infty^2,
\end{aligned}
$$

where $\tilde{\beta}_1(\tau)$ and $\check{\beta}_1(\tau)$ are between $\beta_1(\tau)$ and $\beta_1(\tau) + h_n(\tau)$.

By solving this integral equation, we have that

$$
\begin{aligned}
\dot{\Psi}_{\beta_1}^{-1}(a)(t) &= \pi_c^{-1}\tilde{f}_1^c(\beta_1(t))^{-1}[a(t) \\
&+ \int_0^t a(u)\tilde{f}_1^c(\beta_1(u))^{-1}\bar{f}_1^c(\beta_1(u)) \times \prod_u^t \left(1 + \tilde{f}_1^c(\beta_1(s))^{-1}\bar{f}_1^c(\beta_1(s))dH(s)\right) dH(u)],
\end{aligned}
$$

which is a linear operator of $a(t)$. Thus we obtain the desired weak convergence that $\sqrt{n}(\hat{\beta}_1(\tau) - \beta_1(\tau)) \to -\dot{\Psi}_{\beta_1}^{-1}(G)$, which also is Gaussian. And by functional limiting laws (Chernozhukov et al., 2010), we have that $\sqrt{n}(\hat{F}_1(t) - F_1(t)) \to f_1^c(t)\dot{\Psi}_{\beta_1}^{-1}(G)$.

Table 3.4: 25% Censoring: Comparions between the proposed methods and the standard IV method to estimate survival function for compliers on treatment group, for $S^1(t) = 0.3, 0.5, 0.70.9$. Bias, sampling standard error (EmpSD), averaged estimated standard error (EstSD).($\times 10^2$)

	Proposed			IV	
$S^1(t)$	Bias	EmpSD	EstSD	Bias	SSE
$\pi_c = 0.5, n = 200$					
0.9	-0.20	7.20	7.30	-0.01	8.20
0.7	0.20	9.60	9.80	0.01	9.90
0.5	0.30	9.80	9.00	-0.01	9.40
0.3	0.10	9.00	6.80	-0.80	8.60
$\pi_c = 0.5, n = 400$					
0.9	0.20	5.60	5.30	-0.30	5.70
0.7	0.20	6.80	7.20	-0.01	7.10
0.5	0.01	6.60	6.80	0.20	7.00
0.3	0.40	6.00	5.40	-0.30	6.00
$\pi_c = 0.2, n = 200$					
0.9	-2.10	14.40	16.00	-0.40	23.80
0.7	0.20	20.90	18.30	0.60	25.00
0.5	0.60	21.40	17.60	-0.00	21.50
0.3	-0.60	18.50	13.20	-2.00	16.30
$\pi_c = 0.2, n = 400$					
0.9	-1.00	11.70	12.50	-0.20	16.30
0.7	1.40	15.90	15.30	0.10	17.90
0.5	1.20	15.70	14.80	0.90	15.80
0.3	0.90	13.50	10.70	-0.30	12.60
$\pi_c = 0.1, n = 400$					
0.9	-3.10	18.10	20.2	-6.2×10^{13}	1.7×10^{15}
0.7	0.60	27.20	25.8	-3.5×10^{13}	1.0×10^{15}
0.5	5.00	31.20	28.9	-1.9×10^{13}	5.4×10^{15}
0.3	6.40	32.00	29.1	-7.0×10^{13}	1.9×10^{15}

Table 3.5: Compliance classes and sample sizes.

n	Z	D	C		
308	1	1	1 (Compliers)		
122	1	0			0 (Never-takers)
448	0	0	1 (Compliers)	or	0 (Never-takers)

Chapter 4

A NEW CLASS OF IV MODEL IDENTIFYING POPULATION-LEVEL ESTIMAND IN OBSERVATIONAL STUDIES WITH UNMEASURED CONFOUNDING

4.1 Introduction

In randomized trials with noncompliance, the association between treatment and outcome is confounded by unmeasured variables and the assignment to treatment or encouragement to take treatment can be justified as a valid instrument variable. In this setting, both the exposure and instrumental variable are binary variables. In the first two projects, we use an instrumental variable approach to identify and estimate the marginal casual effect of treatment on potential outcomes for compliers who would take treatment only when assigned to the treatment group. In this project, we would like to study instrumental variable approach in a more general setting (e.g. observational studies involves non-randomized exposures).

Observational studies are widely used to infer treatment effect when randomized experiments are not feasible for various reasons. However, the treatment assignment or exposure of interest is often associated with background variables that are not necessarily observed in the study. These background variables are often called confounding variables if they are also associated with the outcome. Leaving out confounding variables typically leads to a biased estimate of treatment effect. A major challenge of causal inference in observational studies is to address unmeasured confounding variables. In such settings, instrumental variable (IV) methods are useful to estimate the effect of treatment with unmeasured confounding, in other words, endogenous exposure in econometrics literature. An instrumental variable (IV) is conventionally defined as a variable which is (a) associated with the exposure of interest or treatment, (b) affects the outcome through its effect on the treatment but is otherwise unrelated to the outcome and (c) associated with the outcome without unmeasured confounders. Under the standard IV assumptions proposed by Angrist et al. (1996), Abadie (2000) introduces a new class of instrumental variable (IV) estimators for linear and nonlinear treatment response models with covariates. These IV models identify local quantities of potential outcomes for the compliers, the subpopulation whose treatment status is affected by the instrumental variable. In particular, Abadie et al. (2002) focus on estimating local (conditional) quantile treatment effect (LQTE) for compliers. Chernozhukov and Hansen (2005) propose a new IV model, the instrumental variable quantile regression (IVQR) model, for estimating quantile treatment effect (QTE). The IVQR model relies on a different set of identification assumptions than the IV assumption in Angrist et al. (1996). Instead of estimating local quantile functions for compliers, the IVQR model estimates the quantile functions and QTE for the whole population. The key assumption in IVQR model is the rank similarity assumption, a condition that restricts the individual ranks in the potential outcome distributions. The two sets of IV models are generally non-nested and estimating different quantities. Wüthrich (2019) studies the relationship between the two IV models for estimating QTE and shows that there is a close connection between the IVQR model and LQTE model. The IVQR estimands for QTE are equivalent to LQTE for compliers at transformed quantile levels. Moreover, the IVQR estimand of the average treatment effect (ATE) is a convex combination of the local average treatment effect and a weighted average of LQTE. These findings do not rely on the rank similarity assumption and provide an alternate interpretation of IVQR estimands when the rank similarity assumption is violated. Since the definition of compliers depends on the particular IV that is available, the local estimand would, in general, differ from the population-level estimand for the whole study population. In a recent paper, Wang and Tchetgen Tchetgen (2018) discuss estimating the average treatment effect (ATE) in the context of an IV. Instead of making the monotonicity assumption as in Angrist et al. (1996), they propose two alternative no-interaction assumptions involving the unobserved confounders. As argued in the previous chapters, we are more interested in the distribution of potential outcomes other than the ATE. In this chapter, we propose a new class of IV models for nonlinear moment condition models for population-level estimands, generalizing Abadie's identification results for compliers in Abadie (2000, 2003). In the previous chapters, we proposed estimators to identify the marginal distribution of potential outcomes for compliers in the randomized trials with noncompliance. As two special cases, we show that under the new IV assumptions, the previously proposed estimating equations identify population-level quantities in the framework of observational studies with unmeasured confounding. Right censoring survival outcomes are common in epidemiological studies. However, IV methods are not as well developed for regression analysis for survival outcomes. Most literatures focus on location shift models on certain scale of survival outcome (Tchetgen et al. (2015); Li et al. (2015)). To study the heterogeneous treatment effect at different quantiles, Chernozhukov et al. (2015) developed a censored quantile instrumental variable (CQIV) estimator and described its properties and computation. The CQIV estimator combines Powell (1986)'s censored quantile regression to deal with censoring, with a control variable approach to incorporate endogenous exposures. However, this approach requires that the censoring time is always observed, which is not the case in most survival settings. Our proposed model allowing for random censoring and adjustment for baseline covariates complements methodology for identifying distributions of right censoring potential outcomes in observational studies with unmeasured confounding.

In Section 2, we describe notation and setup for the observational studies that we consider in this chapter. We relate the potential outcomes with the structural quantile function by the Skorohod representation. In Section 3, we briefly review the instrumental variable quantile regression (IVQR) model. Specifically, we introduced the IV assumptions for the IVQR model and the estimation procedures. In Section 4, we described Abadie's local IV models for identifying casual quantities for compliers. In Section 5, we discuss the no-interaction assumption and proposed a class of new IV model that estimates quantities of potential outcomes for the whole population. In Section 7, we perform numerical studies on the proposed estimators and make a comparison to the IVQR estimator for the noncensoring outcome.

4.2 Setup and Models

We consider an observational study of n subjects on an outcome y, a binary treatment variable (exposure of interest) D, and a binary instrumental variable Z. X represents a vector of predetermined variables. Let D^z the potential treatment receive under z, the observed variable $D = ZD^1 + (1 - Z)D^0$ with $D_i(i = 1, \dots, n)$, where $D_i = 1$ if subject i receives treatment and $D_i = 0$ if subject i receives control. The population can be classified into four subpopulations based on their compliance status defined by the joint distribution of treatment assignment and potential treatment receipt: compliers if subjects always follow the assignment, never-takers if subjects always take the placebo, always-takers if subjects always take the treatment, defiers if subjects never follow the assignment. Let CC denote the compliance classes: $CC = 1$ (compliers) if $D^1 > D^0$; $CC = 2$ (always-takers) if $D^1 = D^0 = 1$; $CC = 0$ (never-takers) if $D^1 = D^0 = 0$; $CC = 3$ (defiers) if $D^1 < D^0$. In practice, we can only observe one of D^0 and D^1, the compliance class membership for each subject is therefore unknown. Let Y^1 and Y^0 be potential outcome variables with and without the treatment. The core problem of causal inference is that we can only observe one of the two potential outcomes for each individual, and the observed outcome is $Y = DY^1 + (1 - D)Y^0 = Y^d$. The objective of interest is to quantify the distribution of potential outcomes Y_0 and Y_1.

We denote the quantile function of Y^d as Q_d and cumulative distribution function (CDF) as F_d, $Q_d(\tau) = \inf\{y : F_d(y) \geq \tau\}$. Conditional on the observed covariate variables $X = x$, the potential outcome Y^d can be related to the structural quantile function by the Skorohod representation.

$$Y^d = q(d, x, U^d), \text{ where } U^d \sim U(0, 1)$$

U^d is referred to as the rank variable in Chernozhukov and Hansen (2005). It determines the relative ranking of potential outcomes among individuals with similar covariates and treatment status. Thus U^d is responsible for the heterogeneity of outcomes and represents the unmeasured variables that affect potential outcomes Y^d.

Similarly, the observed outcomes Y can be expressed as $Y = q(D, X, U)$, where $Y = Y^D$ and $U = U^D$.

4.3 Instrumental variable quantile regression (IVQR) model

The IVQR model is introduced in Chernozhukov and Hansen (2004, 2005) and its inference method is further discussed in Chernozhukov and Hansen (2006). Typically, the observed covariates X is not sufficient to assure independence assumption between D and potential outcomes Y^d.

4.3.1 Assumptions

Formally, the IVQR model consists of the following main conditions.

Assumption 4 *Given a common probability space, the following conditions hold jointly with probability one:*

1. *Monotonicity: Conditioning on $X = x$, for each d, $Y^d = q(d, x, U^d)$. $q(d, x, \tau)$ is strictly increasing in τ and $U^d \sim U(0, 1)$.*

2. *Independence: Conditioning on $X = x$, U^d is independent of Z.*

3. *Selection: $D \equiv \delta(Z, X, V)$ for some unknown function δ, instrument variable Z and unmeasured random variables V.*

4. *Rank invariance or rank similarity: Conditioning on $X = x$, $Z = z$,*

   - *$\{U^d\}$ are equal to each other, or, more generally,*
   - *$\{U^d\}$ are identically distributed, conditional on V.*

The observed variables consist of Y, D, X and Z.Chernozhukov and Hansen (2005) presents a detailed discussion on the IVQR model and assumptions. In summary, Rank similarity or rank invariance is the most important assumption of the IVQR model. It implies that the potential outcomes Y^d are not truly multivariate and allows us to address the unmeasured confounding problem.

4.3.2 Estimation procedure

The main testable implication of Assumption 4 is the following conditional moment restriction (Chernozhukov and Hansen (2005)).

$$P[Y \leq q(D, X, \tau)|X, Z] = \tau \quad a.s., \tag{4.1}$$

where Z is a set of instrument variables that affects treatment status D but is independent of potential outcomes. It is equivalent to the statement that 0 is the τth quantile of random variable $Y - q(D, X, \tau)$ conditional on (X, Z):

$$0 = Q_{Y-q(D,X,\tau)}(\tau|X, Z), \quad a.s. \text{ for each } \tau \tag{4.2}$$

So the problem is to find a function $(d, x) \to q(d, x, \tau)$ such that 0 is a solution to the quantile regression of $Y - q(D, X, \tau)$ on (Z, X):

$$0 \in \arg\min_{f \in \mathcal{F}} E\rho_\tau[(Y - q(D, X, \tau) - f(Z, X))] \tag{4.3}$$

where $\mathcal{F}$ is the class of measurable functions of (X, Z) that will be restricted in applications and ρ_τ is the check function in quantile regression.

As in Chernozhukov and Hansen (2006), I consider the basic linear-in-parameter model.

$$q(D, X, \tau) = D\alpha(\tau) + X'\beta(\tau)$$

In the framework of randomized trials with noncompliance, Z is usually selected as the randomization assignment and observed covariates X are not necessary for the validity of instrument variable. In the case of a marginal model without covariates X, the linear-in parameter model is fully nonparametric.

Then the weighted quantile regression objective function is defined as,

$$Q_n(\tau, \alpha, \beta, \gamma) \equiv \frac{1}{n}\sum_{i=1}^{n} \rho_\tau(Y_i - D_i\alpha - X_i'\beta - \phi_i(\tau)'\gamma)V_i(\tau), \tag{4.4}$$

where $\phi_i(\tau) \equiv \phi(\tau, X_i, Z_i)$ is a vector of (transformations of) of observed covariates X and instruments Z, and $V_i(\tau) \equiv V(\tau, X_i, Z_i)$ is a positive weight function. In practice, let $V_i(\tau) = 1$ and form $\phi_i(\tau)$ by projection of D_i on Z_i and X_i. The procedure to solve the optimization problem of the objective function 4.4 is a two-step grid search approach described in Chernozhukov and Hansen (2006):

1. For a given τ, define a grid of α, $\{\alpha_j, j = 1, \ldots, J\}$, and run the τ-quantile regression of $Y - D\alpha_j$ on X and $\phi(\tau)$ to obtain coefficients $\hat{\beta}(\alpha_j, \tau)$ and $\hat{\gamma}(\alpha_j, \tau)$.

2. Choose a α_j that minimizes $||\hat{\gamma}(\alpha_j, \tau)||$ or the Wald statistics for testing $\gamma(\alpha_j, \tau) = 0$. The estimated coefficient is then given by $\hat{\beta}(\alpha_j, \tau)$.

The consistency and functional limit distribution of the quantile process are established in Chernozhukov and Hansen (2006). Furthermore, Chernozhukov and Hansen (2008) derive dual inference approach that allows for weak instrusment when D is weakly dependent on Z.

4.4 Local IV models

Abadie's IV model relies on different sets of assumption and identifies different quantities. Instead of identifying the potential outcome Y^d for the population, the model identifies potential outcome Y^d in the compliers, a subset of the population.

4.4.1 Assumptions

Assumption 5 *We review the standard IV assumptions similar to those in Angrist et al. (1996) and Abadie (2003) that conditioning on covariates X:*

1. *Independence: (Y^1, Y^0, D^1, D^0) is jointly independent of Z.*

2. *Nontrivial assignment: $\pi = P(Z = 1) \in (0, 1)$*

3. *First-stage:* $E(D^1 - D^0) \neq 0$. *The causal effect of* Z *on* D *is nonzero.*

4. *Monotonicity:* $P(D^1 \geq D^0)$*=1, so there are no defiers who always receive the opposite treatment of assignment.*

Under Assumption 5, the potential outcome quantile function for the compliers,

$$Q_{Y^D|CC=1}(\tau|X) = \alpha_\tau D + X'\beta_\tau$$

can be determined by the following objective function.

$$(\alpha_\tau, \beta_\tau) = \text{arg }\min_{\alpha,\beta}\text{E}(k \cdot \rho_\tau(Y - \alpha D - X'\beta)).$$

where, $\rho_\theta(y) = y(\theta - I(y < 0))$ is the usual check function in quantile regression literatures and the weights k are given by

$$k = k(D, Z, X) = 1 - \frac{D(1 - Z)}{1 - P(Z = 1|X)} - \frac{(1_D)Z}{P(Z = 1|X)}$$

Since the weight k is negative when D is not equal to Z, the sample analog of the objective function is typically nonconvex. To address this problem, Abadie et al. (2002) modify the weight k by taking its expectation conditional on the observed data (Y, D, X). This amounts to replacing k by k_v, where

$$k_v = E[k|Y, D, X] = 1 - \frac{D(1 - v(Y, D, X))}{1 - P(Z = 1|X)} - \frac{(1_D)v(Y, D, X)}{P(Z = 1|X)}$$

for $v(Y, D, X) = E[Z|Y, D, X] = P(Z = 1|Y, D, X)$. In the previous two chapters, we consider IV methods for randomized trials with noncompliance. We focus on identifying the marginal distribution of potential outcome for compliers and the objective functions do not involve observed covariates X. Instead of using the smoothing weight of k_v, We propose to use a simple grid search approach to solve the optimization problem.

4.5 A new class of IV models

Under Assumption 5, Abadie (2000) present a general result that any parameter defined as the solution to a moment condition involving (Y, D, X) is identified for compliers, using weights k. However, these estimands are for compliers, a latent subset that depends on the selection of instrument variable Z. To estimate the average treatment effect in the population, Wang and Tchetgen Tchetgen (2018) propose two no-interaction assumptions involving the unobserved confounders that allow for the identification of the average treatment effect in the population. In the following part, I show that the IV models using Abadie's weights can be generalized to identify population-level potential outcomes under a similar no-interaction assumption as in Wang and Tchetgen Tchetgen (2018). Instead of making monotonicity assumption in Assumption 5, we consider the following assumption.

Assumption 6 *(No-interaction assumption)* $\mathrm{E}(D|Z=1,U,X)-\mathrm{E}(D|Z=0,U,X)=\mathrm{E}(D|Z=1,X)-\mathrm{E}(D|Z=0,X)$, *where U are unmeasured confounding variables.*

4.5.1 Main Results

Under Assumption 5(1)-(3) and 6, we could generalize main results in Abadie (2000) to casual estimand for the population.

Theorem 4.5.1 *Let $g_\theta(Y,D,X)$ be any real function of (Y,D,X) with parameters θ such that $E|g_\theta(Y,D,X)\leq\infty|$.Given Assumption 5 and 6, $\mathrm{E}(g_\theta(Y^d,d,X)|X)=0$ for $d=0,1$ implies*

$$\mathrm{E}(k\cdot g_\theta(Y,D,X))=0$$

where $k=k(D,X,Z)=1-\frac{D(1-Z)}{1-\pi(X)}-\frac{(1-D)Z}{\pi(X)}$.

The proof is detailed in the appendix. This result enables us to identify population-level causal parameters involved in a moment condition. Theorem 4.5.1 instantly implies the following results for estimating quantile functions of potential outcomes (Y^0,Y^1) for the population.

Corollary 4.5.1.1 *Under Assumption 5(1)-(3) and 6, the quantile functions of potential outcomes, $Q_{Y^D}(\tau|X)=\alpha D_i+X_i'\beta$ can be identified from the observed data (Y,D,X). The* *estimated parameters* $(\hat{\alpha}, \hat{\beta})$ *are solution to*

$$\sum_{i=0}^{n} k_i \cdot (\tau - I(Y_i < \alpha D_i + X_i'\beta)) = 0$$

alternatively, minmized the objective function

$$\sum_{i=0}^{n} k_i \cdot \rho_\tau(Y_i - \alpha D_i - X_i'\beta)$$

.

The weights k can be replaced by the smoothing weight k_v so that an usual weighted quantile regression algorithm will solve the optimization problem. The obtained quantile functions $Q_{Y^D}(\cdot|X)$ are not necessarily monotone. As described in the previous chapters, rearrangement operation can be applied to obtain monotone quantile functions. In absence of covariates X, we describe and discuss the monotone rearranged estiamtors in the framework of randomized trials with noncompliance. Results derived in this chapter are complement to our previous frameworks by extending to observational studies and instrument variable Z is not necessary to be the causal instrument (randomization variable).

4.6 Censoring outcomes

Right censoring outcomes are common in practice. It further complicates identifying causal estimand in the presence of unmeasured confounding. Aside from literature that focuses on estimating survival functions or semiparametric survival models for compliers, there is less work on the estimand of potential outcomes on the population-level. Chernozhukov et al. (2015) develope a censored quantile instrumental variable estimator that combines Powell (1986) censored quantile regression to deal with administrative censoring. This approach requires that the censoring time is always observed, which is not the case in most survival settings. Theorem 4.5.1 implies that distribution functions of potential outcomes in the presence of random censoring can be obtained from a similar objective function as proposed in the previous chapter in the framework of randomized trials with noncompliance. Here we use T^d to denote the potential time-to-event outcome under treatment d, C^d the potential censoring time under treatment d. Define $Y^d = \min(T^d, C^d)$ and $\delta^d = I(T^d \leq C^d)$. Similiarly, for observed data, define $Y = \min(T, C)$, and $\delta = I(T \leq C)$. The observed data consist of n iid replicates of (Y, C, δ, Z, D, X), denoted by $(Y_i, C_i, \delta_i, Z_i, D_i, X_i), i = 1, \ldots, n$. In the case of random censoring that C_d is independent of T_d (given observed covariates), Theorem 4.5.1 generalizes main results in 3.5.

Corollary 4.6.0.1 *Under Assumption 5(1)-(3) and 6, the estimands in Chapter 3 are on population-level. The quantile functions of potential outcomes,* $Q_{T^D}(\tau|X) = \alpha(\tau)D_i + X_i'\beta(\tau)$ *can be identified from the observed data* (Y, C, δ, Z, D, X). *The parameters* (α, β) *are solution to*

$$E\left[k \cdot \left[N(\alpha(\tau)D_i + X_i'\beta(\tau)) - \int_0^\tau I(Y \geq \alpha(\tau)D_i + X_i'\beta(\tau))dH(u)\right]\right] = 0 \tag{4.5}$$

where, $N(t) = I(Y \leq t, \delta = 1)$ *and* $H(t) = -log(1-t)$ *for* $t \in [0, 1]$.

Specifically, we define $\hat{\alpha}(\cdot)$ and $\hat{\beta}(\cdot)$ as a right-continuous piecewise-constant function that jumps only on a grid, $0 = \tau_0 < \tau_1 < \ldots < \tau_{L_n} = \tau_u < 1$. We approximate the integral in Equation 4.5 by $\alpha_i^*(\tau_j) = \sum_{k=0}^{j-1} I(Y_i \geq (\alpha(\tau_k)D_i + X_i'\beta(\tau_k)))(H(\tau_{k+1}) - H(\tau_k))$, for $i = 1, \ldots, n$ and $j = 1, \ldots, L_n$. τ_u is a constant subject to identification constraints due to censoring. Using arguments in Chapter 3 3.4, the parameters can be obtained by minimizing

$$n^{-1/2} \sum_{i=1}^{n} k_i \left(r_i(\tau)(\alpha_i^*(\tau) - I(r_i(\tau) < 0)\delta_i)\right), \tag{4.6}$$

where,$r_i(\tau) = Y_i - (\alpha(\tau)D_i + X_i'\beta(\tau))$,$\alpha_i^*(\tau_j) = \sum_{k=0}^{j-1} I(Y_i \geq (\alpha(\tau_k)D_i + X_i'\beta(\tau_k)))(H(\tau_{k+1}) - H(\tau_k))$ The objective function is not necessarily convex, since weights k is negative when D is not equal to Z. To circumvent this, the smoothing weights k_v can be used to avoid nonconvex objective function. As discussed in Peng and Huang (2008), the smoothing objective function can be transformed to a LP problem and optimized efficiently using on-shelf programs.

4.6.1 Numerical Studies

Noncensoring Outcome

In this section, we examine the performance of the proposed estimator to IVQR estimator. We include settings where the rank similarity assumption is violated. We simulate data in settings as described below.

$$
\begin{aligned}
&U^1 \sim \text{Uniform}(0,1),\ U^0 \sim \text{Uniform}(0,1),\\
&X \sim \text{N}(0,1),\\
&p_z = \text{expit}(c_0 + c_1 \cdot X),\ Z \sim \text{Bernoulli}(p_z)\\
&P(D=1|Z,X) = X \cdot \gamma_0 + \gamma_1 Z + \gamma_2,\\
&p_d = P(D=1|Z,X,V)\\
&\quad = \text{expit}(P(D=1|Z=0,X) \cdot (1-Z) + P(D=1|Z=1,X) \cdot Z + \phi \cdot V).\\
&D \sim \text{Bernoulli}(p_d),\\
&Y^D = D \cdot X'(\beta(U^1) + \alpha(U^1)) + (1-D) \cdot X'\beta(U^0).
\end{aligned}
$$

γ_1 controls the strength of instrument variable Z. ϕ controls the influence of the unmeasured variables V on the treatment variable D. We simualte unmeasured confounding variables V in different settings to evaluate our estimator in terms of robustness to the rank similarity assumptions in Chernuzhukov (2005): (1) Balanced case $V = U^1 + U^0$; (2) Unbalanced $V = 2 * tU^1 - 0.5 * tU^0 + 1.5$, where $tU^1 \sim \Phi^{-1}(U^1)$ and $tU^0 \sim \Phi^{-1}(U^0)$. For each scenario, we generate 1000 independent datasets.

In Table 4.8.2, we set $\alpha(U) = \alpha \cdot U^2$ and $\beta(U) = \beta \cdot U$. We also set $c_0 = 0.1$,$c_1 = -0.4$, $\gamma_0 = -1$, $\gamma_1 = 1$ and $\gamma_2 = 0.5$. The parameter $\gamma_1 = 1$ and the F-statistics of regression D on Z and X is about 25, indicating a moderate IV strength. We present estimated quantile functions of Y^1 and Y^0 at $\tau = 0.25, 0.5, 0.75$, given baseline covariate $X = 0$. The IVQR estimators have larger sampling standard errors than the proposed estimator in all scenarios. We also examine the performance of the naive estimator (ITT estimator) by regressing the observed outcome Y on treatment status D and baseline covariate X. The biases of the naive estimator are larger than the IVQR estimator in all scenarios, though the sampling standard errors are mostly around 0.10. Since our primary interest is the comparison of the proposed estimator and IVQR estimator, those results are not included in the table.

Censoring Outcome

We then examine the proposed method for the censoring outcome case. We simulate data in a similar framework as in the noncensoring case. We set $V = 2 * tU^1 - 0.5 * tU^0 + 1.5$, where $tU^1 \sim \Phi^{-1}(U^1)$ and $tU^0 \sim \Phi^{-1}(U^0)$. The censoring variable $C \sim Uniform(0, c)$, where c is a constant. The potential outcomes $T^D = D \cdot (\beta(U^1) + \alpha(U^1)) + (1 - D) \cdot \beta(U^0)$. We estimate marginal quantile functions of T^1 and T^0 under approximately 20% censoring. We consider strong IV cases and weak IV cases, where the F-statistic of We consider two scenarios corresponding to a strong and a weak IV. The strength of an IV is how strong the association between the IV and the actual treatment is. When the first stage F-statistic from the regression of the treatment on the IV is less than 10, the IV is commonly considered to be a weak IV (Stock et al. (2012)). The parameter γ_1 in $P(D = 1|Z) = \gamma_1 Z + \gamma_2$ is set as (a) strong IV: $\gamma_1 = 2$ (First-stage F-statistic$\approx$,complers proportion $\approx$ 30%, for $n = 400$); (b) weak IV: $\gamma_1 = 1$ (First-stage F-statistic$\approx$6, complers proportion $\approx$ 15%, for $n = 400$). For each scenario, we generate 1000 independent datasets.

In Table 4.8.2 and Figure 4.8.2, we set $\beta(U) = \beta \cdot U$ with $\beta = 10$, $\alpha(U) = \alpha \cdot U^2$ in quadratic case and $\alpha(U) = \alpha \cdot U$ in linear case with $\alpha = 5$. Under a strong IV scenario, we report performance fo the estimated quantile functions using the proposed method at $\tau = 0.1, 0.3, 0.5, 0.7$. The proposed estimators have similar performance across scenarios of different forms of $\alpha(U)$. The sampling standard error increases as τ gets closer to 1.

In Table 4.8.2 and Figure 4.8.2, under a weak IV scenario, we report performance fo the estimated quantile functions using the proposed method at $\tau = 0.1, 0.3, 0.5, 0.7$. When the correlation between IV and treatment status D changes from strong to weak, we observe larger bias at $\tau = 0.1$; the weak IV has a larger effect on the bias of $\hat{Q}_{T^1}$ and $\Delta\hat{Q}$; all estimators have larger sampling standard error. The weak IV affects the precision of the proposed estimators, especially when τ is near 0 or 1, as shown in Figure 4.8.2 and Figure 4.8.2. We expect that the estimated lines are closer to the true lines when the censoring rate gets lower and the sample size is larger.

4.7 Conclusion

In this chapter, we describe the instrumental variable quantile regression (IVQR) model, proposed and discussed in Chernozhukov and Hansen (2005). The key (untestable) assumption is the rank similarity assumption that restricts the evolution of individual ranks of potential outcomes across treatment states. The estimation procedures is a two-step grid search algorithm. We then briefly review the IV models proposed by Abadie (2000) under the standard IV assumptions in Angrist et al. (1996). Unlike the IVQR model that estimate quantile functions for the whole population, Abadie's IV models identify the estimand of potential outcomes for compliers. Wüthrich (2019) studies the relationship between the two IV models for estimating QTE and shows that there is a close connection between the IVQR model and LQTE model. Our previous results are based on Abadie's IV models and estimate quantile functions of potential outcomes for compliers. Instead of making the monotonicity assumption as in Angrist et al. (1996), we impose a no-interaction assumption proposed by Wang and Tchetgen Tchetgen (2018) and propose a new class of IV models that identify quantities of potential outcomes for the whole population. As two special cases, we show that under the new IV assumptions, the previously proposed estimating equations identify population-level quantities in the framework of observational studies with unmeasured confounding. Our work complements current research on using instrumental variable method to estimate distributions of potential outcomes and infer heterogenous treatment effect for observational studies in the presence of unmeasured confounding, especially for the censoring outcomes. Further studies are of interest to compare our censoring regression model to other semiparametric models.

4.8 Proofs

4.8.1 Proof of Theorem 1

In this section, we give proof of Theorem 1. Let $g_\theta(Y, D, X)$ be any real function of (Y, D, X) with parameters θ such that $E|g_\theta(Y, D, X) \leq \infty|$.

$$\mathrm{E}(k \cdot g_\theta(Y, D, X)) = E[g_\theta(Y, D, X)] - E[\frac{D(1-Z)}{1-\pi(X)} g_\theta(Y, D, X)] - E[\frac{(1-D)Z}{\pi(X)} g_\theta(Y, D, X)] \tag{4.7}$$

The first term in Equation 4.7 gives

$$\begin{aligned}
\mathrm{I}_1 &= E[g_\theta(Y, D, X)] \\
&= E_X E[g_\theta(Y, D, X)|X] \\
&= E_X[E[g_\theta(Y, D, X)|X, Z=1]\pi(X)] + E_x[E[g_\theta(Y, D, X)|X, Z=0](1-\pi(X))] \\
&= E_{X,U}[E[g_\theta(Y^1, 1, X)|X, Z=1, D=1, U]\pi(X)P(D=1|X, Z=1, U)] \\
&\quad + E_{X,U}[E[g_\theta(Y^0, 0, X)|X, Z=1, D=0, U]\pi(X)P(D=0|X, Z=1, U)] \\
&\quad + E_{X,U}[E[g_\theta(Y^1, 1, X)|X, Z=0, D=1, U](1-\pi(X))P(D=1|X, Z=0, U)] \\
&\quad + E_{X,U}[E[g_\theta(Y^0, 0, X)|X, Z=0, D=0, U](1-\pi(X))P(D=0|X, Z=0, U)] \\
&\underset{(1)}{=} E_{X,U}[E[g_\theta(Y^1, 1, X)|X, U]\pi(X)P(D=1|X, Z=1, U)] \\
&\quad + E_{X,U}[E[g_\theta(Y^0, 0, X)|X, U]\pi(X)P(D=0|X, Z=1, U)] \\
&\quad + E_{X,U}[E[g_\theta(Y^1, 1, X)|X, U](1-\pi(X))P(D=1|X, Z=0, U)] \\
&\quad + E_{X,U}[E[g_\theta(Y^0, 0, X)|X, U](1-\pi(X))P(D=0|X, Z=0, U)] \\
&= E_{X,U}[E[g_\theta(Y^1, 1, X)|X, U]P(D=1|X, Z=0, U) + \pi(X)(P(D=1|X, Z=1, U) \\
&\quad - P(D=1|X, Z=0, U))] + E_{X,U}[E[g_\theta(Y^0, 0, X)|X, U]P(D=0|X, Z=0, U) \\
&\quad + \pi(X)(P(D=0|X, Z=1, U) - P(D=0|X, Z=0, U))],
\end{aligned}$$

the second term in Equation 4.7 gives

$$
\begin{aligned}
\mathrm{I}_2 &= E[\frac{D(1-Z)}{1-\pi(X)} g_\theta(Y,D,X)] \\
&= E_{X,U}[E[g_\theta(Y^1,1,X)|X,Z=0,D=1,U]P(D=1|X,Z=0,U)] \\
&= E_{X,U}[E[g_\theta(Y^1,1,X)|X,U]P(D=1|X,Z=0,U)]
\end{aligned}
$$

and the third term in Equation 4.7 gives

$$
\begin{aligned}
\mathrm{I}_3 &= E[\frac{Z(1-D)}{\pi(X)} g_\theta(Y,D,X)] \\
&= E_{X,U}[E[g_\theta(Y^0,0,X)|X,U]P(D=0|X,Z=1,U)]
\end{aligned}
$$

The Equation 4.7 can be rewritten as

$$
\begin{aligned}
&\mathrm{I}_1 - \mathrm{I}_2 - \mathrm{I}_3 \\
&= E_{X,U}[E[g_\theta(Y^1,1,X)|X,U]\pi(X)(P(D=1|X,Z=1,U) - P(D=1|X,Z=0,U))] \\
&\quad + E_{X,U}[E[g_\theta(Y^1,1,X)|X,U]\pi(X)(P(D=1|X,Z=1,U) - P(D=1|X,Z=0,U))] \\
&= E_{X,U}[E[g_\theta(Y^1,1,X)|X,U]\pi(X)(P(D=1|X,Z=1) - P(D=1|X,Z=0))] \\
&\quad + E_{X,U}[E[g_\theta(Y^1,1,X)|X,U]\pi(X)(P(D=1|X,Z=1) - P(D=1|X,Z=0))] \\
&\text{(by Assumption 6)} \\
&= E_X[E[g_\theta(Y^1,1,X)|X]\pi(X)(P(D=1|X,Z=1) - P(D=1|X,Z=0))] \\
&\quad + E_X[E[g_\theta(Y^1,1,X)|X]\pi(X)(P(D=1|X,Z=1) - P(D=1|X,Z=0))] \\
&= 0
\end{aligned}
$$

since $E[g_\theta(Y^d,d,X)|X] = 0$ for $d = 0, 1$.

4.8.2 Proofs of Corollarys

Corollary 4.5.1.1 follows Theorem 4.5.1 immediately after replacing $g_\theta(Y^d,d,X)$ with $k\cdot(\tau - I(Y^d < \alpha d + X'\beta))$. Since $E(\tau - I(Y^d < Q_{Y^d}(\tau|X))|X) = 0$ for $d \in 0, 1$, the parameters $(\alpha_\tau, \beta_\tau)$ in $Q_{Y^d}(\tau|X) = \alpha_\tau d + X'\beta_\tau$ can be identified from

$$E[k \cdot \rho_\tau(Y - \alpha_\tau D - X'\beta_\tau)].$$

Using arguments in Chapter 2, we could estimate $(\alpha_\tau, \beta_\tau)$ consistently from the observed data (Y_i, D_i, X_i, Z_i) for $i = 1, \ldots, n$. Consistency and limit theorems are established in Chapter 2. Similar arguments work for Corollary 4.6.0.1.For the censoring outcomes, we define T^d the true potential outcomes, C^d the potential censoring variables, and $Y^d = \min(T^d, C^d)$ with indicator $\delta^d = I(T^d \leq C^d)$. Let $g_\theta(Y^d, d, X) = N(Q_{Y^d}(\tau|X)) - \int_0^\tau I(Y^d \geq Q_{Y^d}(\tau|X))dH(u)$. As discussed in Peng and Huang (2008), the stochastic property of the martingale associated with the potential outcomes gives

$$E[g_\theta(Y^d, d, X)] = E[N(Q_{T^d}(\tau|X)) - \int_0^\tau I(Y^d \geq Q_{T^d}(\tau|X))dH(u)] = 0$$

where $N(t) = I(Y^d \leq t, \delta^d = 1)$ and $H(t) = -log(1-t)$.

We define $\hat{\alpha}(\cdot)$ and $\hat{\beta}(\cdot)$ as a right-continuous piecewise-constant function that jumps only on a grid, $0 = \tau_0 < \tau_1 < \ldots < \tau_{L_n} = \tau_u < 1$. We approximate the integral in Equation 4.5 by $\alpha_i^*(\tau_j) = \sum_{k=0}^{j-1} I(Y_i \geq (\alpha(\tau_k)D_i + X_i'\beta(\tau_k)))(H(\tau_{k+1}) - H(\tau_k))$, for $i = 1, \ldots, n$ and $j = 1, \ldots, L_n$. τ_u is a constant subject to identification constraints due to censoring. Using argruments in Chapter 3 3.4, the parameters can be obtained by minimizing

$$n^{-1/2} \sum_{i=1}^{n} k_i \left(r_i(\tau)(\alpha_i^*(\tau) - I(r_i(\tau) < 0)\delta_i) \right), \tag{4.8}$$

where,$r_i(\tau) = Y_i - (\alpha(\tau)D_i + X_i'\beta(\tau))$,$\alpha_i^*(\tau_j) = \sum_{k=0}^{j-1} I(Y_i \geq (\alpha(\tau_k)D_i + X_i'\beta(\tau_k)))(H(\tau_{k+1}) - H(\tau_k))$. The consistency and limit theorems of the obtained estimators are discussed in Chapter 3.

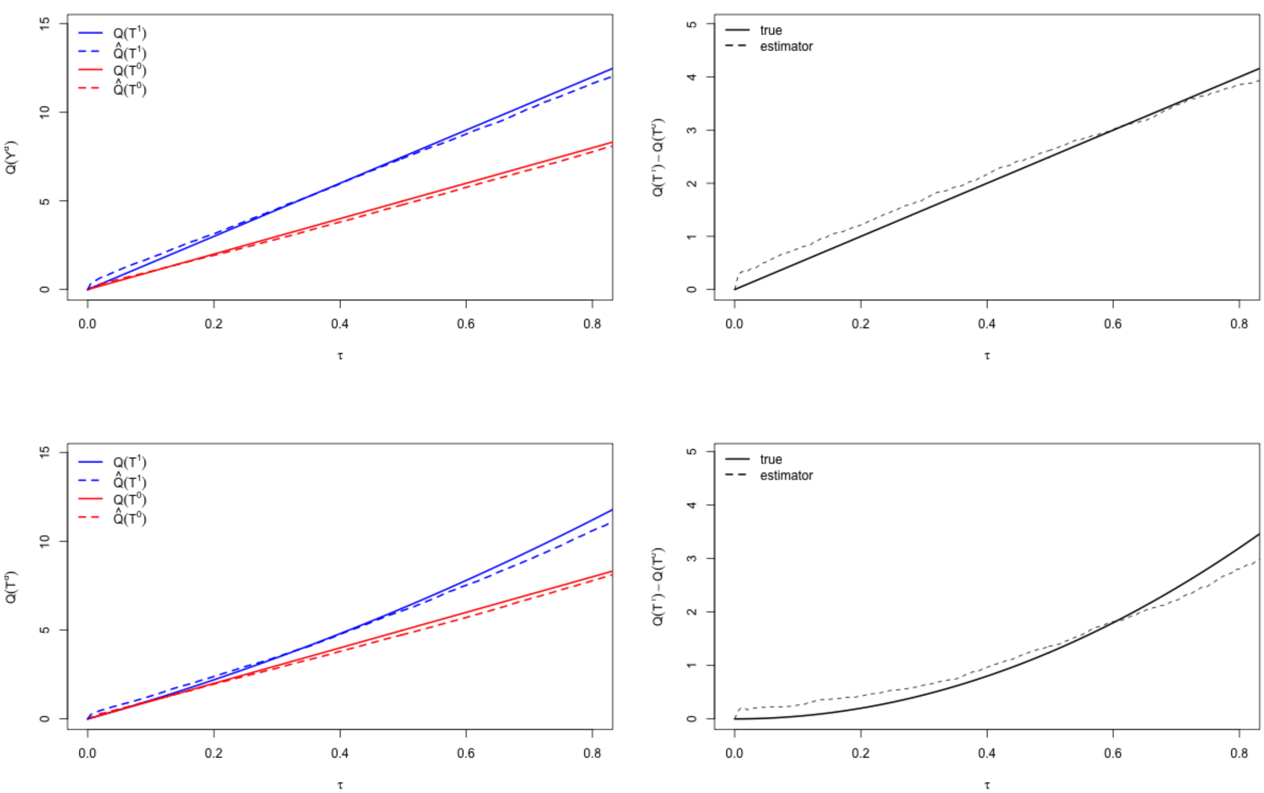


Figure 4.1: Top: plots of estimated quantile functions $\hat{Q}_{T^d}$ and true quantile functions $\hat{Q}_{T^d}$ under strong IV and linear specification of $\alpha(U)$. Bottom: plots of estimated quantile functions $\hat{Q}_{T^d}$ and true quantile functions $\hat{Q}_{T^d}$ under strong IV and quadratic specification of $\alpha(U)$

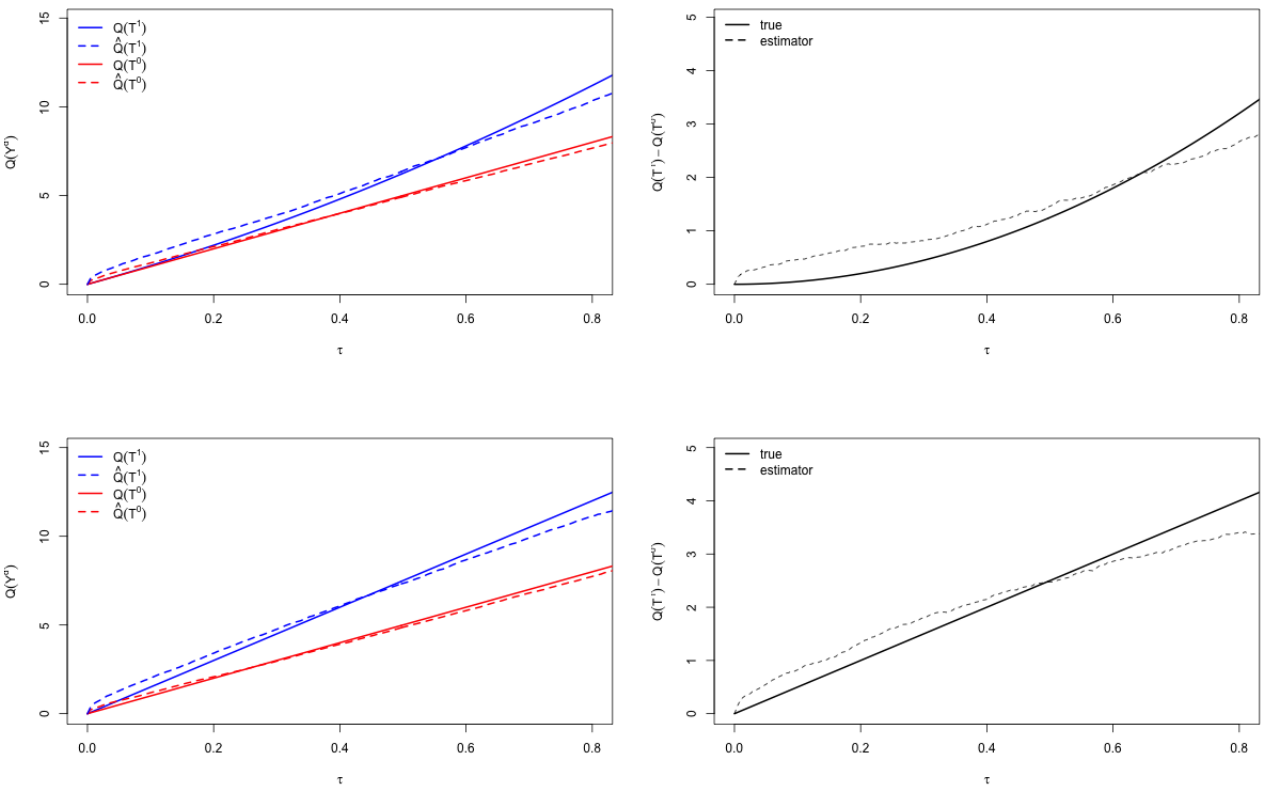


Figure 4.2: Top: plots of estimated quantile functions $\hat{Q}_{T^d}$ and true quantile functions $\hat{Q}_{T^d}$ under weak IV and linear specification of $\alpha(U)$. Bottom: tplots of estimated quantile functions $\hat{Q}_{T^d}$ and true quantile functions $\hat{Q}_{T^d}$ under weak IV and quadratic specification of $\alpha(U)$

Table 4.1: Estimation of quantile functions $Q_1(\tau|X = 0)$ and $Q_0(\tau|X = 0)$ for $\tau = 0.25, 0.5, 0.75$. Bias, sampling standard error (EmpSE), $n = 200$.

	Proposed Estimator				IVQR Estimator			
	Y_1		Y_0		Y_1		Y_0	
τ	Bias	EmpSE	Bias	EmpSE	Bias	EmpSE	Bias	EmpSE
Unbalance case								
0.25	0.09	0.27	-0.05	0.25	0.18	0.26	-0.13	0.37
0.5	0.11	0.28	0.06	0.21	0.16	0.41	0.05	0.39
0.75	0.05	0.29	0.12	0.21	0.03	0.47	0.32	0.65
Balance case								
0.25	0.05	0.26	-0.09	0.26	0.16	0.28	-0.14	0.34
0.5	0.04	0.29	0.03	0.25	0.10	0.31	0.004	0.37
0.75	-0.04	0.31	0.06	0.25	0.1	0.46	0.15	0.55

Table 4.2: 20% Censoring and strong IV: Estimation of quantile functions using the proposed method, $Q_{T^1}(\tau)$, $Q_{T^0}(\tau)$ and $\Delta Q \equiv Q_{T^1}(\tau) - Q_{T^0}(\tau)$ for $\tau = 0.1, 0.3, 0.5, 0.7$. Bias and sampling standard error (EmpSE) for $n = 400$.

	T_0		T_1		ΔQ	
τ	Bias	EmpSE	Bias	EmpSE	Bias	EmpSE
			Linear			
0.1	0.03	0.45	0.29	1.32	0.27	1.37
0.3	-0.15	0.70	0.06	1.82	0.21	1.97
0.5	-0.21	0.79	-0.09	2.05	0.12	2.21
0.7	-0.25	0.78	-0.28	2.04	-0.03	2.19
			Quadratic			
0.1	0.03	0.48	0.24	1.02	0.21	1.14
0.3	-0.14	0.74	0.04	1.58	0.18	1.78
0.5	-0.25	0.81	-0.15	1.99	0.10	2.19
0.7	-0.25	0.79	-0.47	2.23	-0.22	2.36

Table 4.3: 20% Censoring and weak IV:: Estimation of quantile functions using the proposed method, $Q_{T^1}(\tau)$, $Q_{T^0}(\tau)$ and $\Delta Q \equiv Q_{T^1}(\tau) - Q_{T^0}(\tau)$for $\tau = 0.1, 0.3, 0.5, 0.7$. Bias and sampling standard error (EmpSD) for $n = 400$.

	Q_{T^0}		Q_{T^1}		ΔQ	
τ	Bias	EmpSE	Bias	EmpSE	Bias	EmpSE
			Linear			
0.1	0.18	0.90	0.51	1.77	0.32	1.96
0.3	-0.04	1.24	0.26	2.60	0.29	2.88
0.5	-0.13	1.35	-0.17	2.89	-0.04	3.20
0.7	-0.20	1.33	-0.58	2.84	-0.39	3.21
			Quadratic			
0.1	0.20	0.89	0.62	1.72	-0.41	1.93
0.3	0.08	1.27	0.44	2.48	0.36	2.83
0.5	-0.07	1.39	0.14	2.88	0.21	3.26
0.7	-0.23	1.29	-0.42	3.14	-0.19	3.41

BIBLIOGRAPHY

Abadie, A. (2000). Semiparametric estimation of instrumental variable models for causal effects.

Abadie, A. (2002). Bootstrap tests for distributional treatment effects in instrumental variable models. *Journal of the American statistical Association*, 97(457):284–292.

Abadie, A. (2003). Semiparametric instrumental variable estimation of treatment response models. *Journal of econometrics*, 113(2):231–263.

Abadie, A., Angrist, J., and Imbens, G. (2002). Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica*, 70(1):91–117.

Anderson, A. L., Reid, M. S., Li, S.-H., Holmes, T., Shemanski, L., Slee, A., Smith, E. V., Kahn, R., Chiang, N., Vocci, F., et al. (2009). Modafinil for the treatment of cocaine dependence. *Drug and alcohol dependence*, 104(1):133–139.

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, 91(434):444–455.

Baker, S. G. (1998). Analysis of survival data from a randomized trial with all-or-none compliance: estimating the cost-effectiveness of a cancer screening program. *Journal of the American Statistical Association*, 93(443):929–934.

Baker, S. G. and Lindeman, K. S. (1994). The paired availability design: a proposal for evaluating epidural analgesia during labor. *Statistics in medicine*, 13(21):2269–2278.

Cheng, J., Small, D. S., Tan, Z., and Ten Have, T. R. (2009). Efficient nonparametric estimation of causal effects in randomized trials with noncompliance. *Biometrika*, 96(1):19–36.

Chernozhukov, V., Fernandez-Val, I., and Galichon, A. (2009). Improving point and interval estimators of monotone functions by rearrangement. *Biometrika*, page asp030.

Chernozhukov, V., Fernández-Val, I., and Galichon, A. (2010). Quantile and probability curves without crossing. *Econometrica*, 78(3):1093–1125.

Chernozhukov, V., Fernández-Val, I., and Kowalski, A. E. (2015). Quantile regression with censoring and endogeneity. *Journal of Econometrics*, 186(1):201–221.

Chernozhukov, V. and Hansen, C. (2004). The effects of 401 (k) participation on the wealth distribution: an instrumental quantile regression analysis. *Review of Economics and statistics*, 86(3):735–751.

Chernozhukov, V. and Hansen, C. (2005). An iv model of quantile treatment effects. *Econometrica*, 73(1):245–261.

Chernozhukov, V. and Hansen, C. (2006). Instrumental quantile regression inference for structural and treatment effect models. *Journal of Econometrics*, 132(2):491–525.

Chernozhukov, V. and Hansen, C. (2008). Instrumental variable quantile regression: A robust inference approach. *Journal of Econometrics*, 142(1):379–398.

Cuzick, J., Sasieni, P., Myles, J., and Tyrer, J. (2007). Estimating the effect of treatment in a proportional hazards model in the presence of non-compliance and contamination. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(4):565–588.

Fine, J. P., Zou, F., and Yandell, B. S. (2004). Nonparametric estimation of the effects of quantitative trait loci. *Biostatistics*, 5(4):501–513.

Hardy, G. H., Littlewood, J. E., and Pólya, G. (1952). *Inequalities*. Cambridge university press.

Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475.

Imbens, G. W. and Rubin, D. B. (1997a). Bayesian inference for causal effects in randomized experiments with noncompliance. *The Annals of Statistics*, pages 305–327.

Imbens, G. W. and Rubin, D. B. (1997b). Estimating outcome distributions for compliers in instrumental variables models. *The Review of Economic Studies*, 64(4):555–574.

Jin, Z., Shao, Y., and Ying, Z. (2014). A monte carlo method for variance estimation for estimators based on induced smoothing. *Biostatistics*, 16(1):179–188.

Joffe, M. M. (2001). Administrative and artificial censoring in censored regression models. *Statistics in medicine*, 20(15):2287–2304.

Lee, K., Qin, J., and Small, D. S. (2016). Nonparametric inference for distributional treatment effects in instrumental variable models. *arXiv preprint arXiv:1605.03868*.

Li, J., Fine, J., and Brookhart, A. (2015). Instrumental variable additive hazards models. *Biometrics*, 71(1):122–130.

Loeys, T. and Goetghebeur, E. (2003). A causal proportional hazards estimator for the effect of treatment actually received in a randomized trial with all-or-nothing compliance. *Biometrics*, 59(1):100–105.

Lorentz, G. (1953). An inequality for rearrangements. *The American Mathematical Monthly*, 60(3):176–179.

Ma, Y. and Wang, Y. (2012). Efficient distribution estimation for data with unobserved sub-population identifiers. *Electronic journal of statistics*, 6:710.

Martinussen, T., Vansteelandt, S., Tchetgen Tchetgen, E. J., and Zucker, D. M. (2017). Instrumental variables estimation of exposure effects on a time-to-event endpoint using structural cumulative survival models. *Biometrics*.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of econometrics*, 4:2111–2245.

Nie, H., Cheng, J., and Small, D. S. (2011). Inference for the effect of treatment on survival probability in randomized trials with noncompliance and administrative censoring. *Biometrics*, 67(4):1397–1405.

Peng, L. and Huang, Y. (2008). Survival analysis with quantile regression models. *Journal of the American Statistical Association*, 103(482):637–649.

Powell, J. L. (1984). Least absolute deviations estimation for the censored regression model. *Journal of Econometrics*, 25(3):303–325.

Powell, J. L. (1986). Censored regression quantiles. *Journal of econometrics*, 32(1):143–155.

Qin, J., Garcia, T. P., Ma, Y., Tang, M.-X., Marder, K., and Wang, Y. (2014). Combining isotonic regression and em algorithm to predict genetic risk under monotonicity constraint. *The annals of applied statistics*, 8(2):1182.

Robins, J. M. and Tsiatis, A. A. (1991). Correcting for non-compliance in randomized trials using rank preserving structural failure time models. *Communications in statistics-theory and methods*, 20(8):2609–2631.

Stock, J. H., Wright, J. H., and Yogo, M. (2012). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics*.

Tchetgen, E. J. T., Walter, S., Vansteelandt, S., Martinussen, T., and Glymour, M. (2015). Instrumental variable estimation in a survival context. *Epidemiology (Cambridge, Mass.)*, 26(3):402.

Van der Vaart, A. W. (2000). *Asymptotic statistics*, volume 3. Cambridge university press.

Villani, C. (2003). *Topics in optimal transportation*. Number 58. American Mathematical Soc.

Wang, L. and Tchetgen Tchetgen, E. (2018). Bounded, efficient and multiply robust estimation of average treatment effects using instrumental variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(3):531–550.

Wüthrich, K. (2019). A comparison of two quantile models with endogeneity. *Journal of Business & Economic Statistics*, pages 1–28.